THE TELECOMMUNICATIONS HANDBOOK

ENGINEERING GUIDELINES FOR FIXED, MOBILE AND SATELLITE SYSTEMS

EDITED BY JYRKI T. J. PENTTINEN

WILEY

The Telecommunications Handbook

The Telecommunications Handbook

Engineering Guidelines for Fixed, Mobile and Satellite Systems

Edited by

JYRKI T. J. PENTTINEN

WILEY

This edition first published 2015

Registered office
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom

For details of our global editorial offices, for customer services and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com.

Library of Congress Cataloging-in-Publication Data

The telecommunications handbook : engineering guidelines for fixed, mobile, and satellite systems / edited by: Jyrki T.J. Penttinen.
pages cm
Includes bibliographical references and index.
1. Telecommunication–Handbooks, manuals, etc. I. Penttinen, Jyrki T. J., editor.
TK5101.T35533 2015
621.384–dc23

2014022270

A catalogue record for this book is available from the British Library.

ISBN: 9781119944881

Set in 10/12pt Times by Aptara Inc., New Delhi, India

1 2015

Contents

Preface

The evolution of mobile telecommunications is breathtaking. It is also an excellent indicator of technical advances in general – as computers and processors evolve, there is impact on telecommunications solutions with an ever-growing need for capacity and data rates. Knowing that mobile communications were still only utilized by a small group of privileged people back in the 1980s, it is fascinating to realize the current speed of the development of telecommunications networks and devices both technically as well as for business opportunities. It is actually hard to find any other business area which has changed the lifestyle of so many in such a short time period. Presently, the majority of global population has mobile phone whilst the utilization of Internet is growing exponentially – all this only within a couple of decades! Who would want to return any more to the era prior to emails and mobile phones?

The speed of this evolution has also generated challenges. Systems are becoming more and more complicated, and it is very hard to establish a complete picture of telecommunications technologies and systems achievement. There are many new technology areas that need to be learned and taken into account in realistic network deployments, such as security and advanced network planning methods. Furthermore, there is no longer a single concept of fixed and once-and-for-all learning of some areas of telecommunications as new solutions require constant upgrading of knowledge.

The updated understanding about the wider aspects of current and future systems is important for many professionals and higher-level decision-makers because there are increasingly interdependencies along with the evolvement of systems and services. One example is the inclusion of 2G, 3G, 4G, local connectivity and location-based services into smart devices, so knowledge about the respective possibilities as well as limitations of the solutions is essential for service providers, device manufacturers, network architects and planners, and many more professionals. Another example is the efficient planning of the transition from older telecom systems as new systems start taking place. The optimal solution might not be simply a matter of ramping down the previous system to offer maximum capacity for the new one. Instead, utilizing the optimized intermediate solutions for spectral efficient gradual handing over the capacity offered between networks might save huge amount of money for operators. One concrete solution for the gradual lowering of GSM spectrum is the VAMOS terminals and base stations which serve a sufficient number of users within a narrower spectrum whilst pre-4G LTE and actual 4G LTE-Advanced may have greater capacity.

It is thus soon highly recommendable for telecom engineers to also start learning 5G! Currently, it is on the design table but as 2020 approaches, more professionals with updated knowledge are needed. It is a matter of maintaining relevant knowledge for efficient working as the understanding of functionality and end-to-end performance of the complete set of systems gives great assets to optimize user experiences.

This *Telecommunications Handbook* aims to tackle the need prior to the concretization of 5G. It is a well-known fact that systems evolve so fast that literature tends to become outdated at the moment of publication. Nevertheless, the basics of the relevant systems are valid for the long term, and the presentation of the complete set within one book is justified, especially when the information is useful for a variety of professional profiles in order to understand the interdependencies of the systems. This book is meant for experienced professionals who are seeking updated information about systems outside their own special area, and also for persons not familiar with practical telecommunications systems, for example, in technical universities and institutes. The main focus of this book is to combine the information needed in both practical and academic environments:

seasoned professionals can get easy access to telecom theories, and students can obtain realistic views of the practicalities of the systems.

Gradually, as systems evolve, new aspects require new editions, but I totally believe that this book will not be outdated too soon – whilst the systems remain in the markets, their basis as described in this publication will remain relevant. I also maintain updated information via the website www.tlt.fi which contains clarifications as well as extra information, to complement the contents not only of this book but my previous publications *The DVB-H Handbook* and *The LTE/SAE Deployment Handbook*, all published by John Wiley & Sons, Ltd.

I hope you find this *Telecommunications Handbook* useful in your work and studies and I would very much appreciate any feedback via my personal email address: jyrki.penttinen@hotmail.com.

Jyrki Penttinen
Morristown, NJ, USA

Acknowledgements

This book is based on many experiences from real-world projects, results of academic studies, and other investigations in telecommunications field. It also references the research, development and technical project results over a long period of time of many professionals I have worked with in Europe, North America and Latin America, interfacing with telecom companies, governmental organizations and educational institutions. I believe these activities have formed a useful knowledge base for summarizing telecommunications topics in book format. I would thus like to express my special thanks to all my good colleagues at TeliaSonera Finland and Yoigo (Xfera) Spain, standardization groups of ETSI, 3GPP and DVB-H, Aalto University School of Electrical Engineering, United Nations Development Program, Inter-American Development Bank, Finnish Information Society Centre, European Commission, Giesecke & Devrient and organizations of the Nokia umbrella – to mention only some – for the friendliest cooperation whilst I have worked with my employers or as a consultant via my company Finesstel Ltd.

The collection of a complete telecom summary into a single book is without doubt a challenging task for presenting relevant topics in balance, in a compact yet sufficiently deep manner. I acknowledge that our contributor team succeeded in this job excellently by sacrificing valuable personal time with the understanding attitude of the families and significant ones. I appreciate the dedication of the team higher that can be expressed by words.

I am also most thankful for the professional support of Wiley. This book would not be a reality without the planning and organization skills, and gentle yet persistent track keeping of Wiley. I thank very much Liz Wingett, Sandra Grayson, Clarissa Lim, Mark Hammond, and all others from the Wiley team who have worked on this project, as well as Baljinder Kaur at Aptara. I would like to acknowledge especially the hard work of Martin Noble in language editing.

I also give my warmest thanks to the Association of Finnish Non-Fiction Writers for their support.

Finally, I would like to give my special thanks to Elva, Stephanie, Carolyne, Miguel, Katriina, Pertti and all my close family for their support and understanding during the writing.

Jyrki Penttinen

Abbreviations

1G	First Generation of Mobile Communications
1PPS	One Pulse per Second
1xEV-DO	Evolution-Data Optimized (CDMA)
1xRTT	One times Radio Transmission Technology (CDMA)
2G	Second Generation of Mobile Communications
3G	Third Generation of Mobile Communications
3GGP2	American 3rd Generation Partnership Project
3GPP	3rd Generation Partnership Project
4G	Fourth Generation of Mobile Communications
A/D	Analog to Digital
A2DP	Advanced Audio Distribution Profile
AAA	Authentication, Authorization and Accounting
AAS	Active Antenna System
ABMF	Account Balance Management Function
ABS	Advanced Base Stations
AC	Admission Control
AC	Authentication Center (CDMA)
ACeS	Asia Cellular Satellite
ACK	Acknowledge
ACTS	Advanced Communications Technology Satellite
AD	Area Director
ADM	Add/Drop Multiplexer
ADMF	Administration Function
ADSL	Asymmetric Digital Subscriber Line
AEHF	Advanced Extremely High Frequency
AES	Advanced Encryption Standard
AF	Application Function
AGA	Air-Ground-Air
AGC	Automatic Gain Control
AGCH	Access Grant Channel (GSM)
AI	air interface (TETRA)
AICH	Acquisition Indicator Channel (UMTS)
AIE	Air Interface Encryption
AIN	Advanced IN
AKA	Authentication and Key Agreement
ALA	Automatic Link Adaptation
ALC	Asynchronous Layered Coding
AM	Amplitude Modulation
AMC	Adaptive Modulation and Coding
AMI	Alternate Mark Inversion
AMPS	Advanced Mobile Phone Systems
AMR	Adaptive Multirate codec
AMS	Advanced Mobile Stations
AM-VSB	AM Vestigial Side Band
ANDSF	Access Network Discovery and Selection Function
ANR	Automatic Neighbor Relation
ANSI	American National Standards Institute
AP	Access Point
APDU	Application Data Unit
ARIB	Association of Radio Industries and Businesses (Japan)
ARP	Auto Radio Phone
ARPU	Average Return per User
ARQ	Automatic Retransmission on reQuest

ARS	Advanced Relay Stations
AS SMC	Authentication Server Security Mode Command
AS	Access Stratum
AS	Application Server
AS	Authentication Server
ASA	Authorized Shared Access (also: LSA)
ASCII	American Standard Code for Information Interchange
ASI	Adjacent Satellite Interface
ASK	Amplitude Shift Keying
ASME	Access Security Management Entity
ASN	Access Service Network
ASN.1	Abstract Syntax Notation One
ASN-GW	ASN Gateway
AT	AT command (attention)
ATIS	Alliance for Telecommunications Industry Solutions
ATM	Asynchronous Transfer Mode
ATSC	Advanced Television Standards Committee
ATT	Attribute Profile
AuC	Authentication Centre
AVRCP	Audio/Video Remote Control Profile
AWGN	Additional White Gaussian Noise
B6ZS	Bipolar with Six-Zero Substitution
B8ZS	Bipolar with Eight-Zero Substitution
BAN	Body Area Network
BARG	Billing, Accounting and Roaming Group
BCC	Base Station Color Code
BCC	Binary Convolutional Coding
BCCH	Broadcast Control Channel (GSM)
BCH	Broadcast Channel (GSM; UMTS)
BCS	Block Check Sum
BD	Billing Domain
BECN	Backward Explicit Congestion Notification
BER	Bit Error Rate
BFSK	Binary Frequency Shift Keying
BG	Border Gateway
BGAN	Broadband Global Area Network
BICC	Bearer Independent Call Control
BIP	Basic Imaging Profile
B-ISDN	Broadband ISDN
BITS	Building Integrated Time Source
BLEP	Block Error Probability
BLER	Block Error Rate
bmcoforum	Broadcast Mobile Convergence Forum
BM-SC	Broadcast / Multicast Service Centre
BPL	Broadband over Power Lines
BPP	Basic Printing Profile
BPSK	Binary Phase Shift Keying
BS	Base Station (CDMA)
BSC	Base Station Controller
BSIC	Base Station Identity Code
BSS	Base Station Subsystem
BSS	Broadcast Satellite Service
BSSAP	Base Station Subsystem Application Part
BSSMAP	BSS Management Application Part
BTS	Base Transceiver Station
C/R	Command/Response
CA	Carrier Aggregation
CA	Certification Authority
CAI	Computer-Assisted Instruction
CAP	Carrier-less Amplitude and Phase (modulation)
CAPEX	Capital Expenditure
CAT	Carrier Ethernet Transport
CATV	Cable Television
CB	Cell Broadcast
CBCH	Cell Broadcast Channel (GSM)
CBI	Computer-Based Instruction
CBR	Constant Bit Rate
CBT	Computer-Based Training
CC	Component Carrier
CC	Congestion Control
CC	Content of Communication

CCCH	Common Control Channel (GSM)
CCH	Control Channel (GSM)
CCSA	China Communications Standards Association
CDD	Cyclic Delay Diversity
CDF	Charging Data Function
CDF	Cumulative Distribution Function
CDG	CDMA Development Group
CDM	Code Division Multiplexing
CDMA	Code Division Multiple Access
CDPD	Cellular Digital Packet Data
CDR	Charging Data Record
CEIR	Central EIR
CENELEC	Comitée Européen de Normalisation Electrotechique (European Committee for Electrotechnical Standardization)
CEPT	European Conference of Postal and Telecommunications Administrations
CERP	Comité Européen de Réglementation Postale
CET	Carrier Ethernet Transport
CGF	Charging Gateway Function
CGI	Cell Global Identification
CI	Cell Identity
CIF	Common Intermediate Format
CIM	Consumer Instant Messaging
CIP	Common ISDN Access Profile
CLP	Cell Loss Priority
CMAS	Commercial Mobile Alert System
CMIP	Common Management Information Protocol
CMP	Certificate Management Protocol
CN	Core Network
CoC	Content of Communications
COM	Circle Optimized Modulation
CoMP	Coordinated Multipoint Transmission
CP	Control Physical channel
CP	Control Plane
CP	Cyclic Prefix
CPC	Continuous Packet Connectivity
CPFSK	Continuous Phase Frequency Shift Keying
CPICH	Common Pilot Channel (UMTS)
CPS	Characters per Second
CQI	Channel Quality Indicator (UMTS)
CR	Cognitive Radio
C-RAN	Centralized RAN
CRC	Cyclic Redundancy Check
CRNC	Controlling RNC
CS	Circuit Switched
CS	Control Segment
CSC	Canadian Broadcasting Corporation
CSCF	Call State Control Function
CSD	Circuit Switched Data
CSG	Closed Subscriber Group
CSI	Channel State Information
CSMA	Carrier Sense Multiple Access
CSMA/CA	Collision Sense Multiple Access with Collision Avoidance
CSMA/CD	Carrier Sense Multiple Access with Collision Detection
CSN	Connectivity Service Network
CSTA	Computer-Supported Telecommunications Applications
CTF	Charging Trigger Function
CTIA	Cellular Telecommunications Industry Association
CTP	Cordless Telephony Profile
CW	Continuous Wave
D/A	Digital to Analog
DAB	Digital Audio Broadcasting
DAE	Digital Agenda for Europe
DARPA	Defense Advanced Research Projects Agency
DAS	Distributed Antenna System
DBS	Direct Broadcast Service
DC	Direct Current
DC	Dual Carrier
DCCA	Diameter Credit Control Application

DCCH	Dedicated Control Channel (GSM)
DCH	Dedicated Channel (UMTS)
DCH	Dedicated Transport Channel
DC-HSDPA	Dual Cell HSDPA
DCS	Dynamic Cell Selection
DDI	Direct Dialing In
DE	Discard Eligibility (indicator)
DECT	Digital Enhanced Cordless Telephony
DFCA	Dynamic Frequency and Channel Allocation
DFS	Dynamic Frequency Selection
DFT	Discrete Fourier Transform
DGNA	Dynamic Group Number Assignment
DHR	Dual Half Rate
DIP	Device ID Profile
DL DPCCH	Downlink Dedicated Physical Control Channel (UMTS)
DL DPDCH	Downlink Dedicated Physical Data Channel (UMTS)
DL	Downlink
DLDC	Downlink Dual Carrier
DLL	Data Link Layer
DM	Demodulation
DMH	Data Message Handler (CDMA)
DMO	Direct Mode Operation (TETRA)
DMT	Discrete Multitone (line coding)
DoD	Department of Defense
DoS	Denial of Service
DPDCH	Dedicated Physical Data Channel (UMTS)
DPI	Deep Packet Inspection
DRNC	Drift RNC
DRX	Discontinuous Reception
DSCH	Downlink Shared Channel (UMTS)
DSCS	Defense Satellite Communications System
DSL	Digital Subscriber Line
DSP	Digital Signal Processor
DSRC	Dedicated Short-Range Communications
DSS	Digital Subscriber Signaling
DSSS	Direct Sequence Spread Spectrum
DTAP	Direct Transfer Application Part
DTE	Data Terminal
DTH	Direct To Home
DTMF	Dual Tone Multi Frequency
DTV	Digital Television
DUA	DPNSS/DASS2 User Adaptation
DUN	Dial-Up Networking Profile
DVB	Digital Video Broadcasting
DVB-C	Digital Video Broadcasting, Cable
DVB-CBMS	Digital Video Broadcasting, Convergence of Broadcasting and Mobile Service
DVB-H	Digital Video Broadcasting, Handheld
DVB-IPDC	DVB, IP Datacast
DVB-NGH	DVB-H, Next Generation
DVB-S	Digital Video Broadcasting, Satellite
DVB-T	Digital Video Broadcasting, Terrestrial
DXC	Digital Cross-Connect
EA	Extended Address
E-AGCH	Enhanced Absolute Grant Channel (UMTS)
EAI	Enterprise Application Integration
EAMR	Enhanced Adaptive Multi Rate (voice codec)
EAP	Extensible Authentication Protocol
EAPoL	EAP over LAN
EAP-TLS	EAP, Transport Layer Security
EAP-TTLS	EAP, Tunneled Transport Layer Security
EAS	Emergency Alert System
EBU	European Broadcasting Union
EC	European Commission
ECCH	Extended Control Channel
ECO	European Communications Office

E-CSCF	Emergency Call State Control Function
ECTRA	European Committee for Regulatory Telecommunications Affairs
EDCA	Enhanced Distributed Channel Access
E-DCH	Enhanced DCH (UMTS)
EDGE	Enhanced Data Rates for GSM/Global Evolution
E-DPCCH	Enhanced Dedicated Physical Control Channel (UMTS)
E-DPDCH	Enhanced Dedicated Physical Data Channel (UMTS)
EEO	Extremely Elliptical Orbit
EG	ETSI Guides
EGNOS	European Geostationary Navigation Overlay Service
E-GSM	Extended GSM
E-HICH	Enhanced HARQ Indicator Channel (UMTS)
EHS	Electro Hypersensitivity
EHSD	Enhanced High Speed Data
eICIC	Enhanced Intercell Interference Coordination
EIR	Equipment Identity Register
EIRP	Effective Isotropic Radiated Power
eMBM	Enhanced MBMS
EMC	Electro Magnetic Compatibility
EN	European Norm
EOC	Edge of Coverage
EOL	End of Life
EPC	Evolved Packet Core
EPP	EBU Partnership Program
ERC	European Radiocommunications Committee
E-RGCH	Enhanced Relative Grant Channel (UMTS)
ERO	European Radiocommunications Office
ERP	Effective Radiated Power
ES	Energy Saving
ES	ETSI Standard
ESA	European Space Agency
ESC	Engineering Service Circuit
eSE	Embedded Secure Element
ESG	Electronic Service Guide
ESMC	Ethernet Synchronization Messaging Channel
ETO	European Telecommunications Office
ETSI	European Telecommunications Standards Institute
ETWS	Earthquake and Tsunami Warning System
EU	European Union
E-UTRAN	Evolved Universal Terrestrial Radio Access Network
EVM	Error Vector Magnitude
F	Noise Factor
FAC	Final Assembly Code
FACCH	Fast Associated Control Channel (GSM)
FACH	Forward Access Channel (UMTS)
FAX	Fax Profile
FB	FleetBroadband
FBR	Fixed Bit Rate
FCC	Federal Communications Commission
FCCH	Frequency Correction Channel (GSM)
FCS	Frame Check Sequence
FDD	Frequency Division Duplex
FDDI	Fiber Distributed Data Interface
FDM	Frequency Division Multiplex
FDMA	Frequency Division Multiple Access
FDPS	Frequency-Domain Packet Scheduling
FDR	Frame Delay Range
FDT	File Delivery Table
FEC	Forward Error Correction
FECN	Forward Backward Explicit Congestion Notification
FEMA	Federal Emergency Management Agency
FER	Frame Error Rate
FFSK	Fast Frequency Shift Keying
FFT	Fast Fourier Transform

FGW	Femto Gateway
FH	Frequency Hopping
FIFO	First In First Out
FLUTE	File Transport over Unidirectional Transport
FM	Frequency Modulation
FNO	Fixed Network Operators
FPC	Fractional Power Control
FR	Frame Relay
FSK	Frequency Shift Keying
FSTD	Frequency Switched Transmit Diversity
FTA	Free-To-Air
FTD	Frame Transfer Delay
FTP	File Transfer Profile
FTP	File Transfer Protocol
GAN	Global Area Network
GAP	Generic Access Profile
GATT	Generic Attribute Profile
GAVDP	Generic Audio/Video Distribution Profile
GCC	Ground Control Center
GCN	GALILEO Communications Network
GCR	Group Call Register
GEO	Geostationary Earth Orbit
GERAN	GSM EDGE Radio Access Network
GFC	General Flow Control
GGSN	Gateway GPRS Support Node
GISFI	Global ICT Standardization Forum for India
GMSC	Gateway MSC
GMSK	Gaussian Minimum Shift Keying
GMT	Greenwich Meridian Time
GOEP	Generic Object Exchange Profile
GoS	Grade of Service
GPRS	General Packet Radio Service
GPS	Global Positioning System
GR	GPRS Register
GRX	GPRS Roaming Exchange
GS	Group Specifications
GSM	Global System for Mobile Communication
GSMA	GSM Association
GSM-R	GSM Railways
GSN	GPRS Support Node
GSO	Geosynchronous orbit
GSPS	Global Satellite Phone Service
GSS	GALILEO Sensor Stations
GT	Global Title
GTP	GPRS Tunneling Protocol
GTRF	GALILEO Terrestrial Reference Frame
GUS	GALILEO Uplink Station
GUTI	Global Unique Temporary Identity
GWSC	Gateway Switching Centre (also: TSC)
HA	Home Agent
HAC	Hearing Aid Compatibility
HARQ	Hybrid Automatic Repeat Request (UMTS)
HARQ	Hybrid Automatic Retransmission on reQuest
HCCA	HCF Controlled Access
HCF	Hybrid Coordination Function
HCRP	Hard Copy Cable Replacement Profile
HDB3	High Density Bipolar 3
HDLC	High-Level Data Link Control
HDP	Health Device Profile
HDSL	High bit rate Digital Subscriber Line
HDTV	High Definition TV
HEC	Header Error Control
HeNB	Home Evolved NodeB
HEO	Highly Elliptical Orbit
HetNet	Heterogeneous Network
HF	High Frequency
HFP	Hands-Free Profile
HICH	HARQ Indicator Channel (UMTS)
HID	Human Interface Device Profile
HLC	Home Location Center (CDMA)
HLR	Home Location Register
HO	Handover
HOM	Higher Order Modulation
HON	Handover Number
HP	High Precision

HPLMN	Home Public Land Mobile Network
HR	Half Rate
HS	European Harmonized Standard
HS	Headset
HS2.0	HotSpot 2.0
HSCSD	High Speed Circuit Switched Data
HSD	High Speed Data (TETRA)
HSDPA	High Speed Downlink Packet Access
HS-DPCCH	High Speed Dedicated Physical Control Channel (UMTS)
HS-DSCH	High Speed Downlink Shared Channel (UMTS)
HSP	Headset Profile
HSPA	High Speed Packet Access
HS-PDSCH	High Speed Physical Downlink Shared Channel (UMTS)
HSS	Home Subscriber Server
HS-SCCH	High Speed Shared Control Channel (UMTS)
HSUPA	High Speed Uplink Packet Access
HTTP	Hypertext Transfer Protocol
I	Interoperability
IADB	Inter-American Development Bank
IANA	Internet Assigned Numbers Authority
IARC	International Agency for Research of Cancer
I-BCF	Interconnection Bearer Control Function
IBT	Internet-Based Training
ICI	Inter-Carrier Interference
ICP	Intercom Profile
ICS	IMS Centralized Services
ICS	Industrial Control System
I-CSCF	Interrogating Call State Control Function
ICT	Information and Communication Technologies
ICU	Infocommunication Services Market Participants Union
IDEA	International Data Encryption Algorithm
iDEN	Integrated Digital Enhanced Networks
IDFT	Inverse Discrete Fourier Transform
IDM	Identity Management
IDSL	ISDN Digital Subscriber Line
IEC	International Electrotechnical Commission
IEEE	Institute of Electrical and Electronics Engineers
IESG	Internet Engineering Steering Group
IETF	Internet Engineering Task Force
IF	Intermediate Frequency
IFFT	Inverse Fast Fourier Transform
IFRB	International Frequency Registration Board
IM	Instant Messaging
IMEI	International Mobile Equipment Identity
IMEISV	IMEI Software Version Number
IMPP	Instant Messaging and Presence Protocol
IMS	IP Multimedia Subsystem
IMSI	International Mobile Subscriber Identity
IMS-MGW	IMS-Media Gateway
IM-SSF	IP Multimedia – Service Switching Function
IMT-A	International Mobile Telecommunications-Advanced
IMT-MC	IMT Multicarrier
IN	Intelligent Network
INAP	Intelligent Network Application Protocol
InH	Indoor Hotspot
IP SCP	IP Service Control Point
IP	Internet Protocol
IPDC	IP Datacast
IPE	IP Encapsulator
IPI	IP Interworking
IPR	Intellectual Property Rights

IPSec	IP Security
IP-SM-GW	IP Short Message Gateway
IPv4	IP version 4
IPv6	IP version 6
IPXIP	Packet Exchange
IR	Infrared
IRD	Integrated Receiver and Decoder
IrDA	Infrared Data Association
IREG	Interworking & Roaming Expert Group
IRI	Intercept Related Information
IRNSS	Indian Regional Navigational Satellite System
IS	Interim Standard
ISC	International Switching Centre
ISDB-T	Integrated Services Digital Broadcasting, Terrestrial (Japan)
ISDN	Integrated Services Digital Network
ISI	Intersymbol Interference
ISI	Intersystem Interface
ISLAN	Integrated Services LAN (also: isoEthernet)
ISM	Industrial Scientific Medical
ISO	International Standardization Organization
ISOC	Internet Society
ISP	Internet Service Provider
ISUP	ISDN User Part
ITS	Intelligent Transportation System
ITSO	International Telecommunications Satellite Organization
ITU	International Telecommunication Union
ITU-D	ITU development of the telecommunications area
ITU-R	The Radio communication Sector of the International Telecommunication Union
ITU-T	ITU standardization of telecommunications area
IUA	ISDN Q.921-User Adaptation Layer
IWF	Interworking Functions
JAIN	Java APIs for Integrated Networks
JAXA	Japan Aerospace Exploration Agency
JSLEE	JAIN Service Logic Execution Environments
JT	Joint Transmission
KDF	Key Derivation Function
KORA	Korea Radio Station Management Agency
KPI	Key Performance Indicator
LA	Location Area
LAC	Location Area Code
LAI	Location Area Identification
LAN	Local Area Network
LAP	LAN Access Profile
LAP	Link Access Procedure
LAPB	Link Access Protocol Balanced
LAPD	Link Access Protocol in D channel
LAPDm	Link Access Protocol on the modified D channel
LAPF	Link Access Procedure for Frame mode bearer services
LAPM	Link Access Procedure for Modems
LBS	Location Based Service
LBT	L Band Transceiver
LCR	Low Chip Rate
LCS	Location Service framework
LCT	Layered Coding Transport
LDPC	Low-Density Parity Check
LEA	Law Enforcement Agencies
LEMF	Law Enforcement Monitoring Facilities
LEO	Low Earth Orbit
LI	Legal Interception
LIG	Legal Interception Gateway
LLC	Logical Link Control
LNA	Low-Noise Amplifier
LNB	Low-Noise Block down-converter
LOS	Line of Sight
LPD	Link Protocol Discriminator
LRF	Location Retrieval Function

LSA	Licensed Shared Access (also: ASA)
LSB	Least Significant Bit
LSC	Local Switching Centre
LSP	Label Switch Path
LSP	Locally Significant Part
LTE	Long Term Evolution
LTE-A	Long Term Evolution Advanced
M2M	Machine to Machine
M2PA	MTP2 Peer-to-peer user Adaptation layer
M2UA	SS7 Message Transfer Part 2 (MTP2) User Adaptation layer
M3UA	SS7 Message Transfer Part 3 (MTP3) User Adaptation layer
MA	Mobile Allocation
MAC	Medium Access Control
MAIO	Mobile Allocation Index Offset
MAN	Metropolitan Area Networks
MAP	Membership Approval Procedure
MAP	Message Access Profile
MAP	Mobile Application Part
MATV	Master Antenna Television
MAU	Multistation Access Unit
MBMS	Multimedia Broadcast Multicast Service
MBSFN	Multicast Broadcast Single Frequency Network
MBWA	Mobile Broadband Wireless Access
MCC	Mobile Country Code
MCCH	Multicast Control Channel
MC-HSDPA	Multicarrier HSDPA
MCPC	Multiple Channels Per Carrier
MCS	Master Control Station
MCS	Modulation and Coding Scheme
MC-TD-SCDMA	Multicarrier Time-Division Synchronous-Code-Division Multiple Access
ME id	Mobile Equipment Identifier
ME	Mobile Equipment
MEF	Metro Ethernet Forum
MELPe	Mixed Excitation Liner Predictive, enhanced (TETRA; voice codec)
MEO	Medium Earth Orbit
MER	Modulation Error Rate
MFN	Multi Frequency Network
MFSK	Multiple Frequency Shift Keying
MG	Media Gateway
MGC	Media Gateway Controller
MGCF	Media Gateway Control Function
MGW	Media Gateway
MHA	Mast Head Amplifier (also: TMA)
MIB	Management Information Base
MIH	Media Independent Handover
Milstar	Military Strategic and Tactical Relay
MIM	Machine Identification Modules
MIM	Mobile Instant Messaging
MIMO	Multiple-Input, Multiple-Output
MISO	Multiple-Input, Single-Output
MLB	Mobility Load Balancing
MM	Mobility Management
MME	Mobility Management Element
MME	Mobility Management Entity
MMI	Man-Machine Interface
MMS	Multimedia Messaging Service
MMSE	Minimum Mean Square Error
MMTel	Multimedia Telephony
MMUSIC	Multiparty Multimedia Session Control
MNC	Mobile Network Code
MNO	Mobile Network Operator
MOC	Mobile Originated Call
MOS	Mean Opinion Score
MO-SBD	Mobile Originated SBD
MoU	Memorandum of Understanding
MPDS	Mobile Packet Data Service
MPLS	Multi Protocol Label Switching
MRC	Maximum Ratio Combining
MRF	Media Resource Function

MRFC	Media Resource Function Controller
MRFP	Media Resource Function Processor
MRO	Mobility Robustness Optimization
MS	Mobile Station
MSB	Most Significant Bit
MSC	Mobile services Switching Centre
MSISDN	Mobile Subscriber ISDN number
MSK	Minimum Shift Keying
MSRN	Mobile Station Roaming Number
MSS	Mobile Satellite Services
MSS	Mobile Subscriber Station
MTBF	Mean Time Between Failures
MTC	Mobile Terminated Call
MTIE	Maximum Time Interval Error
MTP	Message Transfer Part
MTSAT	Multifunctional Transport Satellite
MT-SBD	Mobile Terminated SBD
MTTF	Mean Time To Failure
MTTR	Mean Time To Repair
MUD	Multiuser Detection
MU-MIMO	Multiuser MIMO
MUX	Multiplexer
MVNO	Mobile Virtual Network Operator
MWBA	Mobile Wireless Broadband Access
MWC	Mobile World Congress
NAB	National Association of Broadcasters
NACK	Negative Acknowledge
NAP	Network Access Provider
NAS SMC	NAS Security Mode Command
NAS	Non Access Stratum
NASA	National Aeronautics and Space Administration
NAT	Network Address Translation
NATO	North Atlantic Treaty Organization
NCC	Network Color Code
NCTA	National Cable Television Association
NDEF	NFC Data Exchange Format
NDS	Network Domain Security
NE Id	Network Element Identifier
NF	Noise Figure
NFC Forum	Near Field Communication Forum
NFC	Near Field Communication
NFV	Network Functions Virtualization
NGMN	Next Generation Mobile Networks (Alliance)
NGN	Next Generation Network
NICAM	Near Instantaneous Compounding Audio Multiplexing
NIR	Non-Ionizing Radiation
N-ISDN	Narrowband ISDN
NMT	Nordic Mobile Telephone
NNI	Network-Network Interface
NRT	Neighbor Relation Tables
nrtPS	Nonreal Time Polling Service
NRZ	Nonreturn to Zero
NSC	National Science Foundation
NSP	Network Service Provider
NSS	Network and Switching Sub-system
NT	Network Terminator
NT	Nontransparent
NTIA	National Telecommunications and Information Administration
NTSC	National Television Standards Committee
NUDET	Nuclear Detonation
NUP	National User Part
OAM	Operation and Maintenance
OAP	One-step Approval Procedure
OBEX	OBject EXchange
OCC	Orthogonal Cover Codes
OCF	Online Charging Function
OCS	Operation Control Segment
OCX	Operation Control System
ODM	Original Device Manufacturer
OEM	Original Equipment Manufacturer

OFCOM	Office of Communications (before: OFTEL)
OFDM	Orthogonal Frequency Division Multiplex
OFDMA	Orthogonal Frequency Division Multiple Access
OFTEL	The Office of Telecommunications (United Kingdom; nowadays OFCOM)
OIPF	Open IPTV Forum
OLPC	Open Loop Power Control
OMA	Open Mobile Alliance
OMC	Operations and Maintenance Center
OMS	Operations and Management System
OOK	On-Off Keying
OP	Organizational Partners
OPEX	Operation Expenditure
OPP	Object Push Profile
OQPSK	Offset Quarternary Phase Shift Keying
OS	Open Services
OSC	Orthogonal Subchannel
OSI	Open Systems Interconnection
OSS	Operations Subsystem
OTA	Over the Air
P2P	Peer-to-Peer
PAD	Packet Assembly/Disassembly
PAGCH	Paging and Access Grant Channel (GSM)
PAL	Phase Alternating Line
PAN	Personal Area Network
PAN	Personal Area Networking Profile
PAPR	Peak-to-Average Power Ratio
PBA	Phone Book Access Profile (also: PBAP)
PBAP / PBA	Phone Book Access Profile (also: PBA)
PBCH	Physical Broadcast Channel (LTE)
PBX	Private Branch Exchange
PCC	Policy and Charging Control
PCC	Primary Component Carrier
PCCPCH	Primary Common Control Physical Channel (UMTS)
PCEF	Policy and Charging Enforcement Point
PCFICH	Physical Control Format Indicator Channel (LTE)
PCG	Project Coordination Group
PCH	Paging Channel (GSM; UMTS)
PCI	Peripheral Component Interconnect
PCI	Physical Cell Identifier
PCI	Precoding Control Information
PCM	Pulse Code Modulation
PCN	Personal Communications Network
PCRF	Policy and Charging Resource Function
P-CSCF	Proxy Call State Control Function
PCU	Packet Control Unit
PDCCH	Physical Downlink Control Channel (LTE)
PDCP	Packet Data Convergence Protocol
PDH	Plesiochronous Digital Hierarchy
PDM	Polarization Division Multiplex
PDN GW	Packet Data Network Gateway
PDN	Packet Data Network
PDSCH	Physical Downlink Shared CHannel (LTE)
PDU	Protocol Data Unit
PEI	Peripheral Equipment Interface
PER	Packet Error Rate
P-GW	Packet Data Network Gateway (LTE)
PHICH	Physical HARQ Indicator Channel (LTE)
PI	Paging Indication
PICH	Paging Indicator Channel (UMTS)
PKM	Privacy Key Management
PLMN	Public Land Mobile Network
PMI	Precoding Matrix Indicator
PMR	Private Mobile Radio
PMR	Professional Mobile Radio

PNT	Positioning, Navigation and Time
PoC AS	Push to Talk Application Server
POM	Power Optimized Modulation
POTS	Plain Old Telephone System
POW	Power Optimized Modulation
PP	Precautionary Principle
PPDR	Public Protection and Disaster Relief
PRACH	Physical Random Access Channel (UMTS; LTE)
PRB	Physical Resource Block
PRC	Primary Reference Clock
PRD	Permanent Reference Document
PRN	Pseudo-Random Noise
PS	Presence Server
PSAP	Public Safety Answering Point
P-SCH	Primary Synchronization Channel
PSI	Program Specific Information
PSK	Phase Shift Keying
PSS	Primary Synchronization Signal
PSTN	Public Switching Telephone Network
PTI	Payload Type Identifier
PTM	Point-to-Multipoint
PTM-SC	PTM Service Center
PTP	Point-to-Point
PTP	Precision Timing Protocol
PTT	Post, Telephone and Telegraph Administration
PTT	Push To Talk (also: PoC)
PUCCH	Physical Uplink Control Channel (LTE)
PUSCH	Physical Uplink Shared Channel (LTE)
PVC	Permanent Virtual Circuit
QAM	Quadrature Amplitude Modulation
QC	Quad Carrier
QoE	Quality of Experience
QoS	Quality of Service
QPSK	Quadrature Phase Shift Keying
R	Receive

R1BS	Revision 1 Base Station
R1MS	Revision 1 Mobile Station
RA	Registration Authority
RA	Routing Area
RACH	Random Access Channel (GSM; UMTS)
RAN	Radio Access Network
RAPA	Radio Promotion Association (Korea)
RAT	Radio Access Technology
RCS	Rich Communication Suite
RDS	Radio Data System
RF	Radio Frequency
RF	Rating Function
RFID	Radio Frequency ID
R-GSM	GSM, Railways
RI	Rank Indicator
RLF	Radio Link Failure
RLM	Radio Link Monitoring
RMa	Rural Macro
RN	Relay Node
RNC	Radio Network Controller
RNTI	Radio Network Temporary Identifiers
ROI	Return of Investments
RoT	Rise over Thermal
RPR	Resilient Packet Ring
RRC	Radio Resource Control
RRM	Radio Resource Management
RS	Uplink Reference Signal (LTE)
rSAP	Remote SIM Access Profile
RSRP	Reference Signal Received Power
RSRQ	Reference Signal Received Quality
RSZI	Regional Subscription Zone Identity
RTCP	Real-time Transport Control Protocol
RTD	NFC Record Type Definition
RTP	Real-time Transport Protocol
RTSP	Real-time Streaming Protocol
RWG	Regulatory Working Group
RZ	Return to Zero
S	Send
SA	Services and system aspects
SA	System Architecture

SACCH	Slow Associated Control Channel (GSM)
SAE	System Architecture Evolution
SAIC	Single Antenna Interference Cancellation
SAN	Satellite Access Node
SAP	SIM Access Profile
SAPI	Service Access Point Identifier
SAR	Specific Absorption Rate
Satcoms	Satellite Communications
SB	SwiftBroadband
SBD	Short Burst Data
SCC AS	Service Centralization and Continuity Application Server
SCC	Secondary Component Carrier
SCCPCH	Secondary Common Control Physical Channel (UMTS)
SCENIHR	Scientific Committee for Emerging and Newly Identified Health Risks
SCF	Service Control Function
SC-FDM	Single Carrier Frequency Division Multiplexing
SC-FDMA	Single Carrier Frequency Division Multiple Access
SCH	Synchronization channel (GSM; UMTS)
SCIM	Service Control Interaction Management
SCP	Service Control Point
SCPC	Single Channel per Carrier
S-CPICH	Secondary Common Pilot Indicator Channel
S-CSCF	Serving Call State Control Function
SCTP	Stream Control Transmission Protocol
SDAP	Service Discovery Application Profile
SDCCH	Standalone Dedicated Control Channel (GSM)
SDH	Synchronous Digital Hierarchy
SDM	Spatial Division Multiplex
SDN	Software Defined Networking
SDP	Session Description Protocol
SDR	Software Defined Radio
SDSL	Symmetric Digital Subscriber Line
SDTV	Standard Definition TV
SDU	Service Data Units
SE	Secure Element
SEG	Security Gateway
SEL	Spectral Efficiency Loss
SF	Spreading Factor
SFD	Saturated Flux Density
SFN	Single Frequency Network
SFPG	Security and Fraud Prevention Group
SG	Signaling Gateway
SGI	Short Guard Interval
SGSN	Serving GPRS Support Node
S-GW	Serving Gateway (LTE)
SI	Service Information
SIGTRAN	Signaling Transport
SIM	SIM Access Profile
SIM	Subscriber Identity Module
SIMO	Single-Input, Multiple-Output
SIMPLE	Session Initiation Protocol Instant Messaging and Presence Leveraging Extensions
SINR	Signal to Interference and Noise Ratio
SIP	Session Initiation Protocol
SIR	Signal-to-Interference Ratio (also: Carrier per Interference, C/I)
SIS	Signal in Space (GALILEO)
SISO	Single-Input, Single-Output
SIWF	Shared IWF
SLA	Service Level Agreement
SLF	Subscription Locator Function
SMATV	Satellite Master Antenna Television System
SMG	Special Mobile Group
SMP	Standards Making Process
SMPTE	Society of Motion Picture and Television Engineers
SMS	Short Message Service
SM-SC	Short Message Service Center
SMTP	Simple Mail Transfer Protocol
SNMP	Simple Network Management Protocol

SNR	Serial Number
SNR	Signal-to-Noise Ratio (also: Carrier per Noise, C/N)
SNRi	Input Signal-to-Noise Ratio
SNRo	Output Signal-to-Noise Ratio
SOCC	Satellite Operations Control Center
S-OFDMA	Scalable-Orthogonal Frequency Division Multiple Access
SoL	Civilian Safety of Life
SON	Self-Organizing/Optimizing Network
SONET	Synchronous Optical Network
SP	Spare Number
SP	Standard Precision
SPC	Signalling Point Code
SPC	Stored Program Control
SPDU	Session Packet Data Unit
SPP	Serial Port Profile
SR	Special Report
SRI	Scheduling Request Indicator (LTE)
SRNC	Serving Radio Network Controller
SRS	Sounding Reference Signal (LTE)
SRVCC	Single Radio Voice Call Continuity
SS	Space Segment
SS	Subscriber Station
SS	Synchronization Signal
SS7	Common Channel signaling system number seven
S-SCH	Secondary Synchronization Channel
SSID	Service Set Identifier
SSO	Semi-Synchronous Orbit
SSP	Service Switching Point
SSPA	Solid State Power Amplifier
SSS	Secondary Synchronization Signal
STA	Station
STBC	Space-Time Block Coding
STM	Synchronous Transfer Mode
STP	Shielded Twisted Pair
STTD	Space Time Transmit Diversity
SUA	Signalling Connection Control Part User Adaptation Layer
SV	Space Vehicle
SVC	Switched Virtual Circuit
SYNCH	Synchronization Profile
T	Terminal
T	Transparent
TA	Terminal Adapter
TA	Timing Advance (GSM)
TA	Tracking Area (LTE)
TAC	Type Approval Code
TACS	Total Access Communications Systems
T-ADS	Terminating Access Domain Selection
TAG	Technical Advisory Group
TAI	Tracking Area Identifier (LTE)
TAP	Two-step Approval Procedure
TAPI	Telephony Application Programming Interface
TAS	Telephony Application Server
TAU	Tracking Area Update (LTE)
TBS	Transport Block Size
TC	Technical Committee
TCH	Traffic Channel (GSM)
TCH/F	Full rate Traffic Channel (GSM)
TCH/H	Half rate Traffic Channel (GSM)
TCP	Transfer Control Protocol
TCP/IP	Transfer Control Protocol / Internet Protocol
TDD	Telecommunication Device for the deaf persons
TDD	Time Division Duplex
TDEV	Time Deviation
TDM	Time Division Multiplex
TDMA	Time Division Multiple Access
TD-SCDMA	Time Division Synchronous CDMA
TDU	Transport Data Unit
TE	Terminal Equipment
TEA	TETRA Encryption Algorithm
TEDS	TETRA Enhanced Data Service
TEL	Technology-Enhanced Learning

TETRA	Terrestrial Trunked Radio
TFCI	Transport Format Channel Indicator
TFO	Tandem Free Operation (also: TrFO)
THIG	Topology Hiding
TIA	Telecommunications Industry Association (North America)
TIPHON	Telecommunications and IP Harmonization on Networks
TKIP	Temporal Key Integrity Protocol
TM	Terminal Multiplexer
TMA	Tower Mounted Amplifier (also: MHA)
TMO	Trunked Mode Operation (TETRA)
TMSI	Temporal Mobile Subscriber Identity
TOC	Train Operating Companies
ToD	Time of Day
ToP	Timing over Packet
TP	Traffic Physical channel
TPC	Transmit Power Control
TPS	Transmission Parameter Signaling
TR	Technical Recommendation
TR	Technical Report
TRAU	Transcoder / Rate Adapter Unit
TrFO	Transcoder Free Operation (also: TFO)
TrGW	Transition Gateway
TRX	Transceiver
TS	Technical Specification
TS	Timeslot (also: TSL)
TS	Transport Stream
TSC	Technical Steering Committee
TSC	Technical Subcommittee
TSC	Transit Switching Centre (also: GWSC)
TSG	Technical Specification Group
TSL	Timeslot (also: TS)
TSM	Trusted Service Manager
TSS	Telecommunications Standardization Sector
TTA	Telecommunications Technology Association
TTAC	Tracking, Telemetry And Control
TTC	Telecommunications Technology Association (Japan)
TTI	Transmission Time Interval
TTT	Time-To-Trigger
TTY	TeleType device
TU	Typical Urban
TUP	Telephone User Part
TVRO	Television Receive terminals
TWG	Technical Working Group
TxBF	Transmit Beam Forming
UA	User Agent (SIP)
UDP	Unstructured Datagram Protocol
UDP	User Datagram Protocol
UE	User Equipment
UHF	Ultra High Frequency
UI	User Interface
UICC	Universal Integrated Circuit Card
UL DPCCH	Uplink Dedicated Physical Control Channel (UMTS)
UL DPDCH	Uplink Dedicated Physical Data Channel (UMTS)
UL HS-DPCCH	Uplink High Speed Dedicated Physical Control Channel (UMTS)
UL	Uplink
UMa	Urban Macro
UMi	Urban Micro
UMTS	Universal Mobile Telecommunication System
UN	United Nations
UNDP	United Nations Development Program
UNI	User to Network Interface
UP	User Plane
UPnP	Universal Plug and Play
URA	UTRAN Registration Area
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
URS	User-specific Reference Symbols
US	User Segment
USAT	Ultra Small Aperture Terminals

USB	Universal Serial Bus
USB-IF	USB Implementers Forum
USF	Uplink State Flag
USIM	Universal Subscriber Identity Module
USNC	US National Committee
USNDS	US Nuclear Detonation Detection System
UTP	Unshielded Twisted Pair
UTRA	Universal Terrestrial Radio Access
UTRAN	UMTS Terrestrial Radio Access Network
UWB	Ultra Wide Band
V5UA	V5.2-User Adaptation Layer
VBR	Variable Bit Rate
VC	Virtual Circuit
VCI	Virtual Channel Identifier
VDP	Video Distribution Profile
VDSL	Very high speed Digital Subscriber Line
VGCS	Voice Group Call Service
VGP	Vehicle Gateway Platform
VHF	Very High Frequency
VLC	Visible Light Communication
VLE	Virtual Learning Environment
VLR	Visitor Location Register
V-MIMO	Virtual-MIMO
VNC	Virtual Network Computing
VoBB	Voice over Broadband
VOD	Video-on-Demand
VoIP	Voice over IP
VPI	Virtual Path Identifier
VPN	Virtual Private Network
VRRA MAC	Variable Rate Reservation Access, Medium Access Protocol
VSAT	Very Small Aperture Terminal
WAN	Wide Area Networks
WAPB	Wireless Application Protocol Bearer
WB-AMR	Wideband Adaptive Multi Rate
WBT	Web-Based Training
WCDMA	Wideband Code Division Multiple Access
WDM	Wavelength Division Multiplexing
WEP	Wired Equivalent Privacy
WG	Working Group
WHO	World Health Organization
WI	Work Item
WiMAX	Worldwide Interoperability for Microwave Access
WLAN	Wireless Local Area Network
WLL	Wireless Local Loop
WMAN	Wireless Metropolitan Area Networks
WPA	Wi-Fi Protected Access
WPA2	Wi-Fi Protected Access 2
WRAN	Wireless Regional Area Network
WRC	World Radiocommunication Conference (previously WARC, World Administrative Radio Conference)
WRIX	Wireless Roaming Intermediary Exchange
WRIX-d	WRIX Data Clearing
WRIX-f	WRIX Financial Settlement
WRIX-i	WRIX Interconnect
WRP	Wireless Roaming Proxy
XDMS	XML Document Management Server
XOR	Exclusive OR
ZC	Zone Code
ZF	Zero-Forcing

List of Contributors

Dr Jyrki Penttinen has worked in technical management and expert positions with TeliaSonera Finland, Xfera Spain, Nokia / NSN Mexico, Spain and USA, and as a freelancer and telecom consultant via Finesstel Ltd in Europe and Latin America. His focus areas have included GSM, UMTS and LTE cellular and digital mobile TV, radio and core network planning, optimization, standardization and research. Currently, he works in mobile security area as Program Manager with Giesecke & Devrient America Inc. with special interest in M2M and MVNO environment. He is also active lecturer and contributor to technical publications. Dr Penttinen obtained his MSc (EE), LicSc (Tech) and DSc (Tech) degrees at Aalto University, Espoo, Finland, in 1994, 1999 and 2011, respectively. In his free time, he enjoys writing books (presented in www.tlt.fi), practicing archery and communicating via Morse code at his ham radio stations.

Dr Mohmmad Anas has over 9 years of systems design engineering and product management experience working on 4G/LTE wireless networks at Nokia Networks and consumer electronics and wearable devices at Flextronics. His original contribution has resulted in several patent applications on 3GPP LTE capacity and coverage improvements. He has received a PhD for his novel contribution on uplink radio resource management for QoS provisioning in LTE in 2009 from Aalborg University, Denmark. His current interests are in the area of Internet of Things.

Dr Francesco Davide Calabrese is a wireless system specialist with more than 8 years of experience from both, academy and industry. He earned his PhD from the University of Aalborg with a dissertation on Scheduling and Link Adaptation for the Uplink of SC-FDMA Systems. For a few years his main interests remained within the design and optimization of L2 and L3 algorithms for WCDMA and LTE until the last couple of years, when his focus has shifted toward the design of self-learning system algorithms based on novel Machine Learning techniques.

Mr Ryszard Dokuczal received his Master's degree in Microsystems Electronics and Photonics from Wroclaw's University of Technology, Poland. He joined NSN (now Nokia Networks) in 2008 and since then has been involved in several projects focused on network dimensioning, higher sectorization, baseband dimensioning and testing RU20/RU30 features. In 2012 he joined 3GPP standardization team and he is a RAN1 delegate focusing on 3G aspects (UL MIMO, HetNet, FEUL).

Dr Jacek Góra received his Master of Science in Microsystems Engineering from the Wrocław University of Technology in 2008. He obtained his doctoral degree in Communication Technologies from the Poznań University of Technology in 2013. Since 2009 he has been with the Nokia Networks (previously called NSN) where he works as a radio research engineer. He is the author of around 40 technical publications.

Mr Jukka Hongisto is a Solution Architect at Mobile Broadband with Nokia. Starting in circuit switched data services in mobile networks, Jukka later held various positions as Packet Core specification, standardization and competence manager. He then moved on to Packet Core system product management, where he was in charge of Evolve Packet Core (EPC) network architecture planning and standardization for mobile packet networks. Jukka has long experience from 2G/3G/LTE and IP technologies. Since 2007 he has been focused on the mobile voice evolution, targeting industry wide solutions like Voice over LTE – VoLTE. He has worked as a packet switched voice Solution Architect by managing voice-related requirements to products, introducing solutions to mobile operators, doing patents for new solutions, helping standardization and has held seminars for related topics. You can follow Jukka at https://twitter.com/JukkaHongisto.

Mr Tero Jalkanen has worked for TeliaSonera since 2001 as Senior R&D Specialist. He has been involved in the development of various IP based mobile technologies since graduating in 1998, especially the inter-operator (roaming, interworking, interconnection) related aspects of networks as well as services. He has had standardization experience since 1999 with WAP Forum (later OMA) as well as a number of "semi-official" international industry interest groups. He has been involved with GSM Association since 2001 in various working groups, task forces and projects. His expertise includes IMS, SIP, SIP-I, IM, Presence, PoC, Video Share, Image Share, ENUM, WLAN and MMS. He has been heavily involved in the development of GRX and IPX as well as RCS. Tero holds around 20 patents and has the position of Rapporteur for multiple GSM Association Permanent Reference Documents.

Mr Juha Kallio lives in Finland and currently works for Nokia. Juha has been in the field of telecommunications for 20 years and has held various positions in SW development, product specification and architectural design.

Mr Ilkka Keisala has worked with mobile operator TeliaSonera (earlier Sonera) since 1998. He has worked in different R&D activities during this time, especially related to Wi-Fi and 3G areas. These activities also include GSMA driven projects. Currently his focus areas are M2M and Mobile ID. Mr Keisala obtained his BSc (tech) at Metropolia, University of Applied Science, Espoo, Finland in 1999. He previously worked as an air traffic controller and a commercial pilot.

Dr Jaroslaw Lachowski is an experienced researcher, technical and project leader. Throughout his personal career, Jaroslaw has focused on radio access research in HSPA, HSPA enhancements (LTHE), LTE & LTE-A technologies. His interest lies predominately in topics such as Self Organizing Networks (SON), HetNet, Interference Management, Mobility, Traffic Steering and HSDPA Multiple Antennas. He is a recognized expert in those areas with innovative track record proven through IPR generation, patents, scientific publications and various 3GPP and book contributions (e.g., LTE Self-Organizing Networks (SON): Network Management Automation for Operational Efficiency).

Dr Patrick Marsch received his Dipl.Ing and Dr.Ing degrees from Technische Universität Dresden, Germany, in 2004 and 2010, respectively. He was the technical project coordinator of the project EASY-C, where the world's largest research test beds for LTE-Advanced were established. After heading a research group at TU Dresden, Germany, he is now a research team manager within Nokia Networks, Wrocław, Poland. He has (co-)authored 50+ journal and conference papers, has received three best paper awards, been editor of or contributor to several books and has been awarded the Philipp Reis Prize for pioneering research in the field of Coordinated Multi-Point (CoMP).

Mr Michał Maternia received his Master of Optical Telecommunications degree from Wroclaw's Technical University, Poland. He started his career in NSN in 2006 where he has been involved in multiple research projects focusing on system level aspects of 3G, 4G and beyond 4G. His research area ranges from mobility aspects through deployment research and interference management. He is now a senior radio research engineer in Nokia Networks, Wrocław, Poland, and is leading a Multi-RAT/Multi-Layer work package in 5G project METIS.

Dr Guillaume Monghal graduated in 2005 from French engineering school Telecom Sud-Paris and obtained a Master's degree in Mobile communications at Aalborg University. He further pursued his education at Aalborg University and in cooperation with Nokia's network division with a PhD study, which concluded with graduation in June 2009 and then a post doc. Since May 2010, Guillaume Monghal has been working at Intel

Mobile Communications as a wireless specialist, participating in the development of wireless connectivity products.

Mr Olli Ramula is a Project Management Professional (PMP) and has worked for 20 years in mobile telecommunications. His work experience includes network deployment, network maintenance and R&D.

Mr Jouko Rautio served in a Finnish Air Force radar unit and then studied at the Oulu University, Finland and did his Master's thesis on antenna measurements. Having worked in several positions in Telecom Finland (later Sonera Corporation and currently a part of TeliaSonera), mainly in radio network development for mobile telephone services, he has specialized in the EMF area. He has been issued one patent of radio technology and has contributed to two books and written for various magazines. Mr Rautio has been a lecturer in several courses on antenna, radio network and EMF topics and was the Vice Chairman of the trade association EMF Advisory Board in 2001–13. In autumn 2013 he started working on sustainability issues.

Mr Marcin Rybakowski received his Master's degree in Electronics and Telecommunication in 2003 with specialization in Mobile Telecommunication at Wroclaw's University of Technology, Poland. He worked after graduation for Becker Avionics in Poland and Fujitsu Laboratories in Japan as RF Design and Test Engineer. He joined Siemens (now Nokia Networks) in 2006 and since then has been involved in features verification of Base Station and Active Antenna Systems for WCDMA (HSPA) networks. In 2012 he joined Radio Research team and was involved in HSPA research and 3GPP standards development with the focus on Smartphone Signaling, Machine to Machine Communication and Heterogeneous Networks. He is now focusing on millimeter wave deployment research and propagation modeling for future 5G systems.

Mr Szymon Stefański is a team leader in the Research Department at DATAX in Wrocław, Poland. He received a degree in electronic and telecommunication from the Wrocław University of Technology, Poland. He is currently working towards his PhD degree in telecommunication at Poznań University of Technology. His research interests are focused on self-optimization in telecommunication systems.

Mr Luis Angel Maestro Ruiz de Temiño graduated in Mobile Communications from Miguel Hernandez University, Elche, Spain, in July 2007. From September 2007 to January 2009 he worked as a research assistant in the Electronic Systems Department of Aalborg University, where he collaborated with Nokia, Networks Business Unit in the development of solutions for upcoming LTE-A systems. In February 2009, he joined Nokia where he has held different positions. Currently he is heading the Smart Labs team that focuses on studying the interaction between smartphones, mobile operating systems and applications with LTE and UMTS radio networks.

Mr Ali Yaver received his Master's degree in Wireless Systems in 2007 from the Royal Institute of Technology (KTH) in Stockholm. He joined NSN in 2009 and since then he has been involved in several 3G and 4G projects conducting back office research for 3GPP standardization. Currently he is part of the NSN's 5G research program. His research interests include mobility, radio resource management and multi-cell connectivity.

Other contributors to this book:

Mr. Maciej Januszewski
Mr. Damian Kolmas
Mr. Michal Panek
Mr. Krystian Safjan
Mr. Stanisław Strzyż
Ms. Agnieszka Szufarska
Mr. Pertti Virtanen

1

Introduction

Jyrki T. J. Penttinen

1.1 General

This chapter provides an introduction to the contents of the book. It also includes high level information about telecommunications, and instructions on how to utilize the modular structure of the chapters in an efficient way.

The main idea of the book is to combine the theoretical and practical aspects of the complete telecommunication fields, including fixed, mobile, satellite, broadcast and special systems, which is clarified in this chapter.

This Telecommunications Handbook describes principles and details of all the major and modern telecommunications systems that are relevant for the industry and to end-users, and gives useful information about usage, architectures, functioning, planning, construction, measurements and optimization. The book describes applications, equipment, radio, transport and core networks of the selected systems. The book focuses on practical descriptions and gives useful tips for the planning, setup, measurements, optimization, utilization and feasible options. In general, the book will help readers to understand the complete telecommunications field in a practical way.

The contents include the introduction of each technology, evolution path, feasibility, utilization, motivation, importance, solution, network architecture, and technical functioning of the systems. This includes signaling, coding, different modes for channel delivery and security of core and radio system as well as the planning of the core and radio networks. There are system-specific field test measurement guidelines, hands-on network planning advice and suggestions for the parameter adjustments included in several sections of the book. The book also describes the most probable future systems.

The Telecommunications Handbook: Engineering Guidelines for Fixed, Mobile and Satellite Systems, First Edition.
Edited by Jyrki T. J. Penttinen.

1.2 Short History of Telecommunications

1.2.1 The Beginning

The initiation of the actual telecommunications as we understand the term has a long precedence and history, from the era of fire and smoke signals in the most primitive yet functional format in order to deliver simple messages between two different physical locations. Claude Chappe was one of the pioneers who brought the optical signaling techniques to a new level by introducing a method that was based on the different positions of wooden signaling poles 1792 [1].

The finding of ways to handle electricity finally opened the new era of telecommunications as it provided the necessary means to deliver messages over long distances without the limitations of the line-of-sight that previous optical methods required. The characteristics of copper as telecommunication line conductor were well understood in the nineteenth century [2]. The most concrete application of this era was Morse code in the 1800s, which is still utilized actively by radio amateurs, or hams, all over the globe, although its importance in commercial communications has practically disappeared and it is utilized merely as a backup support in limited environments when other systems fail. Table 1.1 shows the original Morse codes that are still utilized in the ham community in addition to other transmission modes.

Voice services took their first steps soon after, and Alexander Graham Bell patented the fixed telephone in 1876. Regardless of official recognition, there were also other inventors like Elisha Gray brainstorming on the same topic, which was a concrete sign that people realized the importance of telecommunications.

The relevance of early experiments by radio amateurs cannot be underestimated. Radio and television broadcasting as we know it today benefited greatly from the experiments that radio amateurs carried out. After hobby-based activities, broadcasting was taken over by governments as the importance of communications started to become clear. Nevertheless, radio amateurs still continue with the experiments of the old and new transmission modes of wireless communications. The radio amateurs or "ham radio" community enjoys the amateur radio hobby in such a way that licensed participants operate communications equipment with a deep appreciation of the radio art [3]. Connections are typically confirmed via QSL card, i.e., cards for the acknowledgement of radio amateur connections, which nowadays can also have electronic form. Figure 1.1 shows an example of QSL card of radio amateurs which is used for acknowledging 2-way connections. Today, amateur radio activity is a mix of fun, public service, and convenience.

Amateurs have a basic knowledge of radio technology and operating principles, and pass an examination for the regulators' license to operate on radio frequencies in the amateur bands. As an example, FCC is responsible for radio amateur licensing in the USA. As soon as the candidate passes the exams that include radio technologies, communications procedures and regulations, the candidate is awarded a license to operate

Table 1.1 *The Morse code table*

A	.-	K	-.-	U	..-	0	-----
B	-...	L	.-..	V	...-	1	.----
C	-.-.	M	--	W	.--	2	..---
D	-..	N	-.	X	-..-	3	...--
E	.	O	---	Y	-.--	4	....-
F	..-.	P	.--.	Z	--..	5	.....
G	--.	Q	--.-	Dot	.-.-.-	6	-....
H	....	R	.-.	=	-...-	7	--...
I	..	S	...	Error	...-.	8	---..
J	.---	T	-	End	...-.-	9	----.

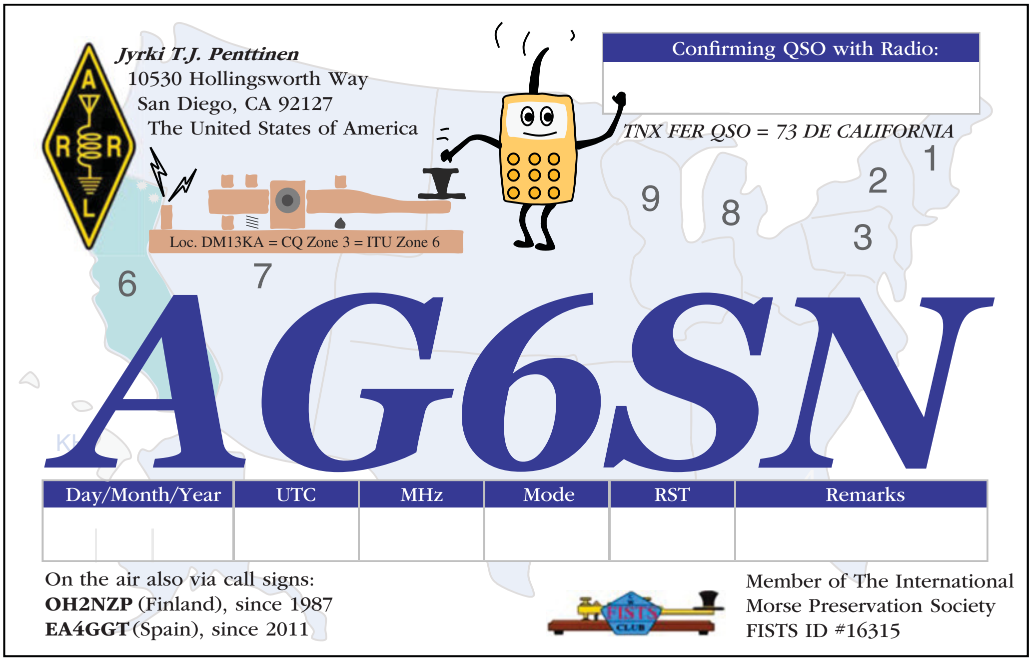

Figure 1.1 *The QSL card is still utilized in amateur radio communications for confirming the connections.*

in the frequencies with the modes and technical limitations the respective class dictates. It should be noted that earlier the passing of the Morse code exam was also required for part of the classes, but it is not included in the official exam any more. Nevertheless, operating in continuous wave (CW) via Morse code is still one of the popular modes today.

The regulators, following the national and international principles, reserve the radio amateur frequency blocks for use by hams. There are various bands for the use of radio amateurs in almost all of the practically usable frequencies, beginning with the low frequency bands of 160 m (1.8 MHz band) up to the mm-bands near the limit of the ITU (International Telecommunications Union) radio frequency allocation tables. This arrangement provides an excellent opportunity to experiment with the practical radio wave propagation by utilizing various, analog and digital modes for voice, data and video transmission. Introductions for the ham radio can be found in Ref. [4].

The amateur radio, as well as the commercial, special and other types of radio stations, cellular and broadcast operators, and all the other entities that need a license or permit to send the radio signals over the air, are aligned via the national and international rules. The highest entity that dictates the utilization of the radio frequencies is ITU, and the national regulators and other related entities plan jointly the overall rules for the utilization, as well as the more specific limits for allowed power levels, frequency boundaries, and technologies that may possibly utilize per band. This joint alignment is to agree common rules, to avoid interferences between technologies, operators and countries. The great challenge of this work is to find an optimal solution so that as many services can be utilized as possible at a global level. The problem is that there have been countless solutions appearing in different useful bands over time since the invention of communications over the radio interface.

Due to the complexity of the different regional services on the radio frequencies, the world has been divided into three ITU regions as presented in Figure 1.2. The regions have slightly different divisions for frequency

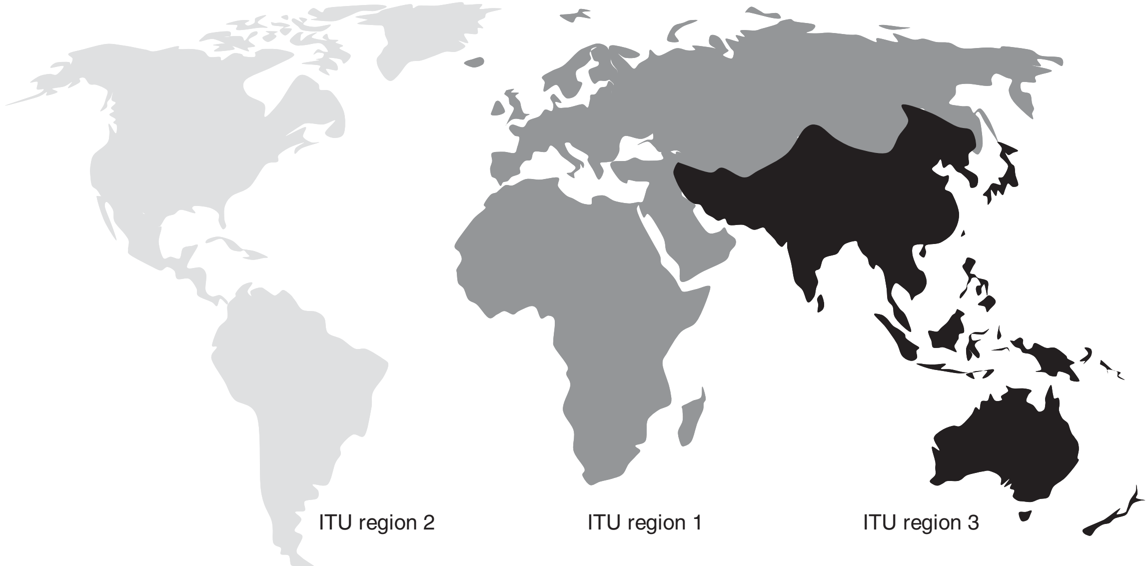

Figure 1.2 *ITU regions.*

utilization and different allowable power limits. This ITU division into three regions is valid for all radio communications from radio amateur activities to commercial, military and scientific mobile communications.

1.2.2 Analog Telephony Era

Early analog systems were based on manual connection of the voice calls. These exchange switchboards included the connection matrix which was handled by personnel. The crossbar switch was developed for automatic voice call delivery. The concept was most popular from 1950 to 1980. Most of the modern telephone switches at that time period were based on some variant of the crossbar switching system as presented in Figure 1.4. Along with the more advanced technology, the relay solutions took place as shown in Figure 1.3.

1.2.3 Wireless Era

The wireless telephony systems appeared in the markets in the beginning of 1970s. One of the early systems in the commercial and publicly available pioneers was ARP, Auto Radio Phone in Finland that was opened for the public 1971 [5]. It was still a manually operated system, and can be interpreted to represent the pre-1G systems. The automatic analog mobile systems falls into the category of 1G-systems and they appeared in the commercial markets as of the beginning of 1980s. Nordic Mobile Telephone (NMT) was one of the examples of this era, together with various similar systems in the USA and Europe appearing in the VHF/UHF bands. 1G-systems started to pave the way for the wireless era, and stayed in the markets for several decades until they were closed down typically at the beginning of 2000 due to the more advanced and spectral efficient variants of the next generation.

The common factor for 2G is digital functionality. GSM (Global System for Mobile communications), defined by ETSI/3GPP, is one of the most widespread and popular systems in the history of wireless telecommunications so far. GSM can be expected to stay in the markets and still evolve, even if the first

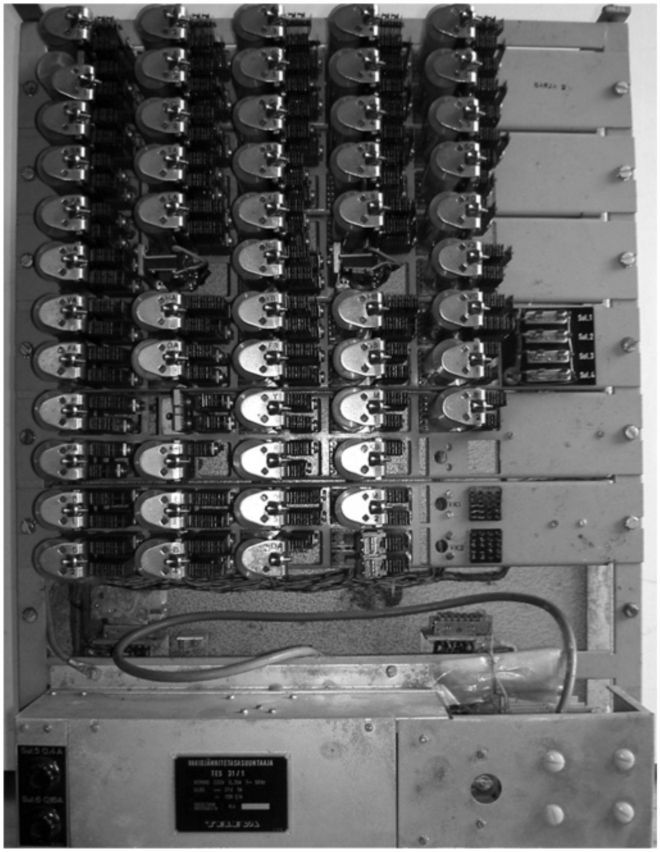

Figure 1.3 An example of small-scale telephone exchange based on mechanical switches. This TELEVA KAU 2/5 model was produced in Finland and it was able to deliver 5 internal and 2 external calls based on the relay arrays.

networks already appeared at the beginning of 1990s. In the US market, 2G is based on CDMA (1x), whilst GSM utilizes TDMA.

3G represents more spectral efficient systems, which are more clearly multimedia capable. Wide band CDMA-based UMTS (Universal Mobile Telecommunications System) paved the way for this era, and there have been various other systems. The evolution has brought, for example, HSPA (High Speed Packet Access), providing considerably higher data rates compared with the first 3G networks that were launched at the beginning of the 2000s.

The mobile system generations have appeared around once per decade, and 4G is no exception. The pre-version of the fully equipped 4G is, for example, LTE (Long Term Evolution) which will evolve towards ITU-defined 4G requirement compliance by the introduction of LTE-Advanced.

1.3 The Telecommunications Scene

1.3.1 Current Information Sources

Even if various telecommunications handbooks exist in the market, they are somewhat limited; they typically have a theoretical approach and are weighted either to radio or core IP networks. Furthermore, the complete

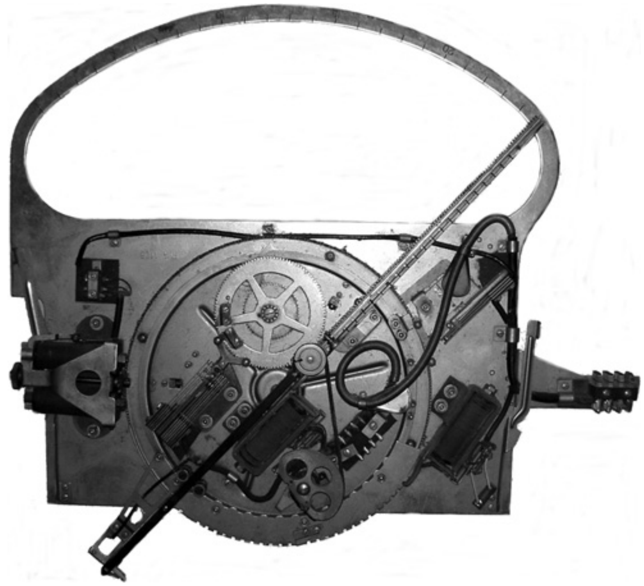

Figure 1.4 An electromechanical crossbar selector for analog telephone exchanges. This element is able to handle 25 angle positions, and the connector of the arm can be adjusted to 22 different lengths, that is, the element could deliver 25 × 22 = 550 phone calls. The presented equipment was still utilized at Telecom Finland in 1970s. Data published by European Union.

telecom field is evolving so rapidly that part of the contents of earlier books is outdated. In addition, existing books tend to give overall presentations of the systems at a somewhat higher level than is really needed in the industry and educational institutes, and the most practical point of view is missing. There is thus a lack of the practical, yet sufficiently in-depth description of the modern systems. This book aims to give a complete picture as well as practical details – with plenty of examples from operational networks – in order to be used as a handbook and a centralized source of guidelines in studies of the complete field, and in the planning and operation of networks. The book also aims to act as a bridge between telecommunications education at universities and institutes and practical knowledge and skills needed in the work of the telecommunications industry.

The weight of the book is in current and near future networks, and the most up-to-date functionalities of each network are described at a sufficiently deep level for deployment purposes. The description and planning of 3GPP mobile communications systems, GSM, UMTS and LTE are emphasized in this book.

The book offers guidelines for the ever-developing telecommunications area, with in-sight into the most relevant telecom systems and is thus useful in global operations of telecommunication systems.

1.3.2 Telecommunications Market

The telecommunications market is without doubt one of the most important at the global level. Both fixed and mobile communications create the base for major part of the world's population for voice and data services. The importance of data services has been increasing greatly, and the general developmental trend is for usage to be going towards all IP, and towards all-mobile. It is evident that the role of mobile communications has already taken over from fixed line communications in several countries for the last few years. According to

Ref. [6] a growing number of subscribers have replaced fixed voice telephony lines with mobile voice service, or voice over IP-types of service. According to the statistics in Ref. [6], mobile voice traffic surpassed fixed voice traffic for the first time in 2009 with 52% of total traffic. This evolution can be seen clearly from another statistic claiming that only 9% of European households have a fixed telephone access but no mobile telephone access.

According to ITU data released in June 2012, there were almost 6 billion mobile communications subscriptions in 2011 [7]. This refers to a global mobile subscription penetration of approximately 86%. Even if it can be assumed that many users have two or more subscriptions, this figure indicates that the growth of mobile communications has been substantial in recent years and the main growth is derived from developing countries.

Back in 1990s, it was very rare for a single country to have over 100% mobile subscription penetration. By the end of 2011, there were already more than 100 countries that had reached this milestone. Especially for mobile broadband, there were more than 1 billion subscriptions at a global level in 2011. As a comparison with the fixed broadband, there were 590 million subscriptions at a global level active in 2011.

After decades of quite constant and relatively peaceful development of the circuit switched voice service, and still during the initial steps of the circuit switched data of fixed networks until 1990s, the transition towards IP based networks and services has triggered a major change in the telecommunications area. The data, including voice service, multimedia, audio, video, messaging and file transfer, is delivered by default via the principles of IP networks and services. It can be estimated that IP with its developed variant IPv6 will take over communications worldwide. IPv6 is rapidly advancing in many areas of the world, especially China and Japan. Activities are deployed based on IPv6, with IP address portability. It is logical that this transition happens in a parallel way both in fixed and mobile systems in such a way that the convergence of services, seamless continuum of connections and transparent transfer of data is developing in great leaps.

It can be claimed that the current worldwide trend indicates that the industry value is shifting towards mobile devices. As the content, applications and services are going mobile, the prices and margins of traditional access services are declining. There is consumer need, and also national and international support for pawing the way for this evolution. As an example, the European Union wants to ensure comprehensive availability and take-up of fast and ultra-fast Internet as it is noted to be one of the major building blocks of the Digital Agenda for Europe (DAE). EU has thus reasoned that in order to take advantage of the benefits of a sustainable economic and social environment, it is essential that advanced broadband networks and applications are available to all European consumers of the IT services, including both private consumers as well as business users. For this reason, the DAE and the European growth strategy for the next decade, called Europe 2020, is officially committed to reaching goals for high-speed data enablers. Concretely, this means that basic broadband networks have aimed to be available to all EU citizens by 2013. According to the plan, half of European households should be able to subscribe to at least 100 Mb/s by the year 2020, whereas 30 Mb/s is planned to be available to all Europeans [6].

In order to estimate the evolution of fixed broadband communications, source [6] has noted that at the end of 2010, DSL access was available to 95.3% of the EU population. It was growing slightly, having been at 94.4% in 2009. As a comparison, DSL coverage in rural areas was at 82.5% compared to the whole rural population. This means that there are still 23.5 million EU citizens in total, from which 18 million located to rural areas, without the possibility of having a fixed access broadband network. On the other hand, the plan of the United States, through its National Broadband Plan, is to provide availability of broadband networks with the minimum available today, 4 Mb/s download and 1 Mb/s upload, to all US citizens.

The near future plans of the EU include national goals. Amongst the challenging goals of various countries – for example, Italy, Portugal and Finland represent the advanced end – the strategy for 2009–15 includes an extension of the aim of universal service to broadband connections to be increased in such a way that by the end of 2015, 100 Mbit/s broadband networks should be available for all.

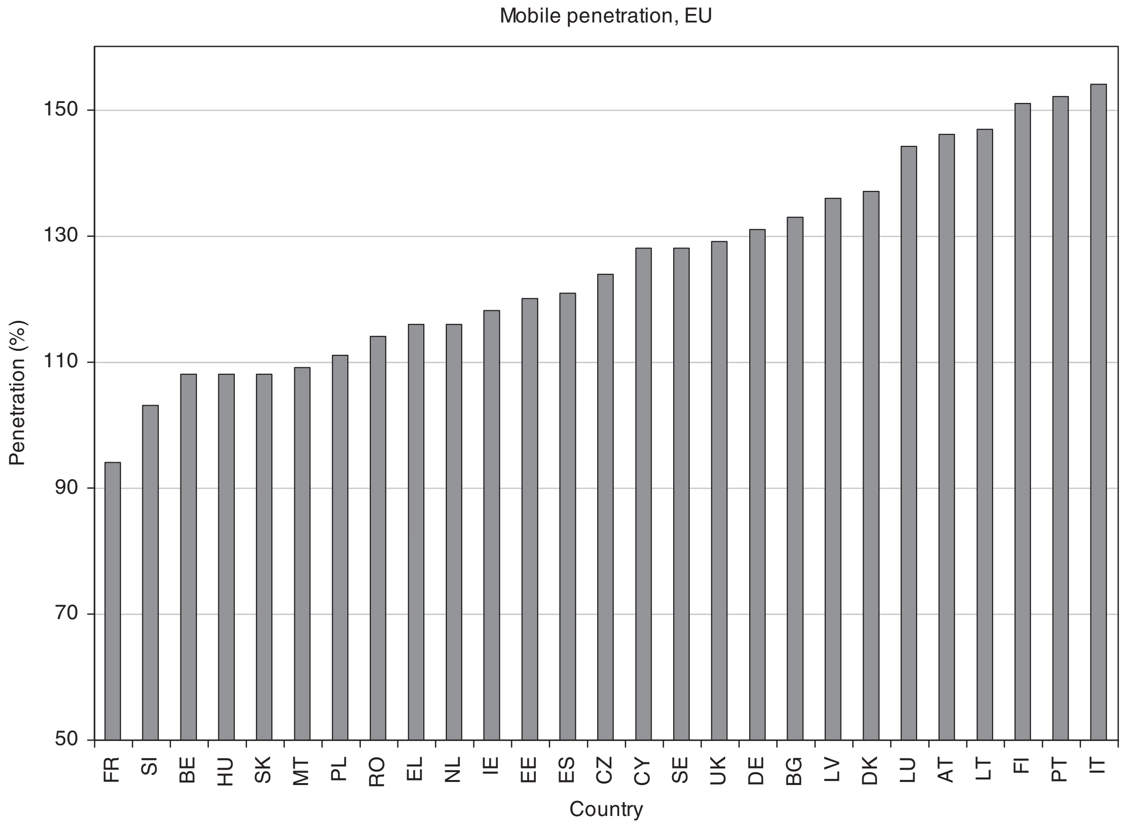

Figure 1.5 *The mobile subscriber penetration in EU countries, 2010 [6]. The average mobile subscription penetration over the whole EU is 125%. Data published by GSM Association.*

Along with fixed networks, there are also demanding goals for mobile systems in the EU, especially to satisfy the expected, very fast increase of the mobile data services. Prior to LTE deployments, operators have been concentrating on the enhancement of previous solutions based on High Speed Packet Access (HSPA) networks. In Europe, 3G availability was 90% for the population by 2010. Even the evolution of HSPA already provides considerably higher mobile data rates than ever before; especially LTE, and the developed variant LTE-Advance, will boost data throughput into the next level. The practical data rates are thus competitive with DSL products, leaving fixed network solutions as being perhaps less attractive – depending logically on the pricing strategies of data services.

Nevertheless, source [6] has noted that there are still large variations in the national level of EU for data utilization and mobile subscription penetration. Figure 1.5 summarizes information found in the source [6] and presents the histogram of the member states for the penetration. It can be seen that Italy has the highest penetration. This can be explained at least partially via a very high ratio of prepaid subscriptions. Other high-penetration countries are Portugal, Finland, Lithuania, Austria and Luxembourg. At the other extreme is France, which is the only EU country with the penetration below 100%. This is probably due to the high number of postpaid subscriptions.

In addition to the EU which is the leader of active SIM penetration (i.e., the same individual may have one or more SIM subscriptions), as well as unique subscriptions, the world market for mobile communications is

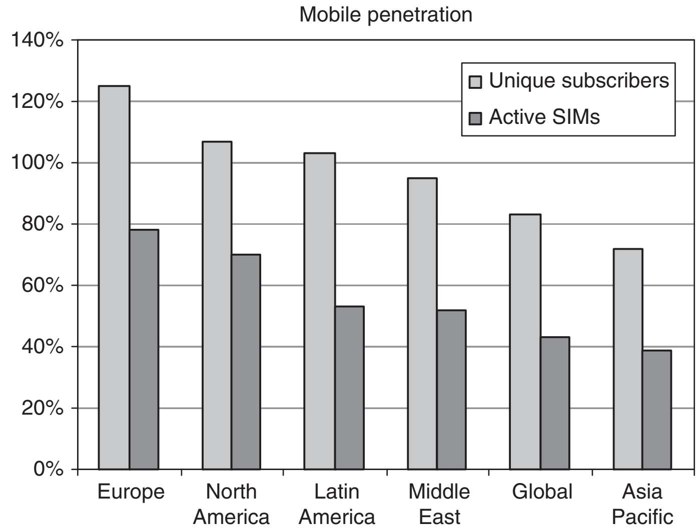

Figure 1.6 *World's mobile subscription statistics [8].*

also growing. Figure 1.6 shows global statistics. At the same time, the effective price per minute is steadily lowering. As some examples of Ref. [8], about 20 euro cent cost of mobile call minute has lowered to around 10–13 euro cent level, for example, for Germany, Italy, UK, France and Spain.

Along with an increase in the use of mobile communications for both voice and data traffic, the prices of services has constantly decreased. Based on measurements of the OECD (baskets comparison), mobile telephony prices have decreased for all consumer patterns [6]. As an example, the low-usage category index was 13.1 in 2006 and dropped to 9.1 by 2010. Equally, the mid-usage category prices moved from 24.0 to 15.3, and high-usage category from 41.5 to 24.2 between the years of 2006 and 2010. This has been a positive trend for end-users whist the ARPU (Average Return per User) has decreased, making the competition between mobile operators tougher. In any case, the trend is such that mobile broadband is the most important item for the revenue growth of mobile operators in the EU.

1.3.3 Effect of Video Services

It can be estimated that in modern telecommunications, there are major effects from video services which are going to be the global driver for data traffic growth on fixed and mobile networks, resulting in the increase of user-generated video transmission. The increase of audio and video is fast and strong due to the transition of "traditional" circuit switched voice services to high-speed packet data networks – which is increasingly the public Internet. VoIP is one of the solutions to facilitate this transition, and makes voice calls just one of the many data services over the Internet. Compared to previous circuit switched solutions, this transition provides much richer calls and by default, there is a set of options users can select to enhance the voice call experience, for example, by including text, special characters, photos, video, audio and other accompanying information to the same call. This transition not only happens for the public use of the tele services, but increasingly also for the company use. In addition to technological benefits making the content delivery richer, the important benefit of VoIP-based calls is reduced cost.

Video and, in general, multimedia contents that have been possible to consume via fixed networks have driven users to wish for similar on-demand access to content via mobile devices. This has not been very

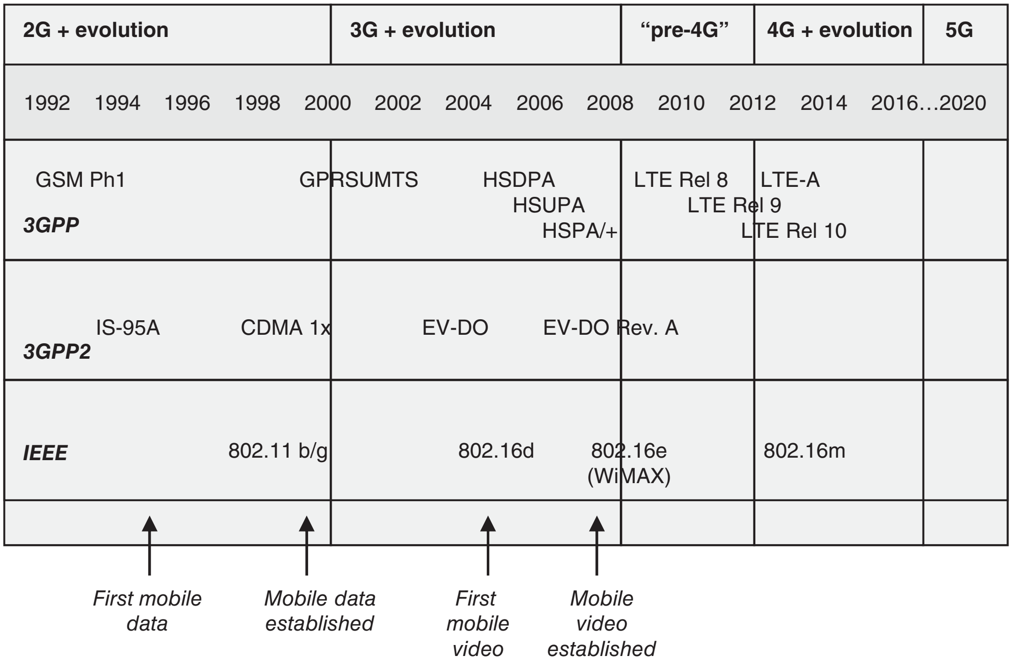

Figure 1.7 *The basic principle of the states in the cellular systems development where video delivery is suitable via mobile communications systems.*

logical yet via the network solution up to basic 3G mobile systems, due to the limitations of data rates and delays of data transmission, and due to economical aspects. LTE is also now a suitable enabler for multimedia contents.

Nevertheless, if the consumption of the video services increases substantially, the capacity problem may be present in the hot spot areas of the cells as the connections are established point-to-point. In any case, LTE allows for significantly higher capacity at a lower cost per bit, which results in highly improved commercial delivery of video services.

Figure 1.7 presents the high-level development of mobile communications with those technologies that are suitable for more demanding video contents.

Along with evolved 2G and 3G technologies, including GPRS of GSM and 1xRTT of CDMA, and HSPA of 3G, the first data services started to take off. Nevertheless, these services were merely basic ones like simple web browsing and email. In fact, it was possible to utilize video services already via 1G-systems [9], but this type of solution was not very practical prior to the deployment of evolved 2G stage and 3G networks that provided sufficiently high bit rates for a fluent video experience. Also the high cost of data delivery and limited functionality of mobile devices meant that the time was not yet ready for serious video service utilization.

Increased capacity demand is constant for voice and data services as well as for the resulting signaling traffic. This also requires increased capacity for transport networks.

One of the modern phenomena is that ever more audiovisual content is available for on-demand streaming, as a result of hugely increased activities of user generated content. There are various types of media for this new method of content production, including TV network programming and applications or services like YouTube. The popularity of the latter has exploded Internet traffic. According to Ref. [10], users of YouTube stream in average 50 videos per month. Furthermore, from 2006 onwards, YouTube in the US consumed more capacity than the entire World Wide Web in 2000. When streaming this type of content via mobile networks instead of fixed networks, the quality has been notably lower due to the capacity restrictions as the format has been the same for all the use cases.

An important factor in video consumption is screen size and resolution. Other relevant parameters are frames rate and bit rate. Table 1.2 summarizes some of the most relevant use cases for different resolutions.

The available throughput for LTE, and especially LTE-Advanced, is clearly higher than any other previous mobile communications solution can offer. Nevertheless, there are situations where the network is in danger of saturating due to excess of simultaneous users.

The service delivered for single users requires dedicated resources from the network. LTE is suitable for high-speed and high-capacity transmissions, so the bit rates that are required by the logical screen sizes of mobile devices can be handled. The problems might arise when multiple users are sharing the resources of the same site, which also increases signaling load.

One of the solutions for this type of situation is the use of MBMS (Multimedia Broadcast Multicast Service) for LTE as presented in Figure 1.8, that is, enhanced MBMS (eMBMS), which is defined first time in the 3GPP Release 9. The previously defined version of MBMS, according to the Release 6 specifications, has not been successful in practice as the demand for PTP (Point to Point) voice and data has required all the capacity. Now, eMBMS may provide a suitable platform for user experiences like television streaming as well as enhanced and interactive contents streaming along with the possibility to show, for example, common URLs for the customers, and move to single PTP connections per user. Some use cases for the eMBMS service could be football stadiums where repetition of the goals can be observed via eMBMS television service. The absolute benefit of this broadcast type of service is that the operator needs to define only the broadcast channels without variations of the capacity as all the users under the eMBMS cell receive the same contents (or, selecting the contents via multiple channels) without utilizing uplink capacity. The obvious disadvantage of this solution is that content must be prescheduled which is comparable with fixed network television programs.

The MBSFN (Multicast Broadcast Single Frequency Network) can be utilized in order to optimize the broadcast mode. It makes bit-synchronizing transmission via a number of eNBs in such a way that the contents can be transmitted at the same time and frequency. The benefit of SFN is that there may be SFN gain in the reception of the signal, according to the principles of OFDM technology.

1.3.4 Network Scalability

Fixed networks have been experiencing a sudden and drastic change recently, developing from the original circuit switched analog era towards the new era of the “all IP” concept. This provides totally new methods and possibilities to deliver data – from where the voice service is just one of the countless data service set. This, in turn, increases awareness of the new services and at the same time as consumers get used to the increasing offering of novelty applications, the data and signaling load of the networks increase. Thanks to the adaptive principles of IP, the new networks can be scaled up to meet with the capacity demand more easily than was possible in the old-fashioned analog networks.

Wireless networks are part of the complete telecommunications machinery, forming an important additional radio interface to the complete end-to-end chain. As radio technologies develop fast, the core and transport networks must be able to follow overall development with increased data delivery and lowered delays in the data delivery. This union of radio and fixed telecommunications networks is quite interesting as it seems to generate a kind of joint competition which assures further development of the systems. Thus, along with fast development of the radio interface and increase of the data rates, there must be solutions for the core network. One logical solution, in addition to the capacity increase, is to convert the core, or transport network as IP based. This solution handles data streams according to packet data definitions and IP protocols. The transition

Table 1.2 *Estimated required display resolutions in order to present high quality video*

Type	Display (inch)	Resolution	Pixels	Viewing distance / m	Human eye max dpi at distance	HQ video requirement, resolution	MPEG4 bit rate, typical kb/s
Smart device	2.5–3	QVGA	320 x 240	0.3–0.5	200	QVGA	0.2–0.4
Multimedia device	3.3–5	HVGA	480 x 320	0.3–0.5	200	HVGA	0.5–1
PMP	4–7	WVGA	800 x 480	0.3 (handheld) –0.8 (table)	120	VGA	0.8–1.5
Ultra mobile PC	7–9	WVGA / WSVGA	800 x 480 / 1024 x 600	0.3 (handheld) –0.8 (table)	120	SD 480/576i	1–2
Laptop	12–17	WXVGA / WUXGA	1280 x 600 / 1920 x 12 900	0.5 (lap) –0.8 (table)	80	HD 720i	3–4
Small HD TV	<32	HD 720p	720p	2–3	50	HD720p	6–8
Large HD TV	>32	HD 1020p	1020p	>3	30	HD1080p	12–16

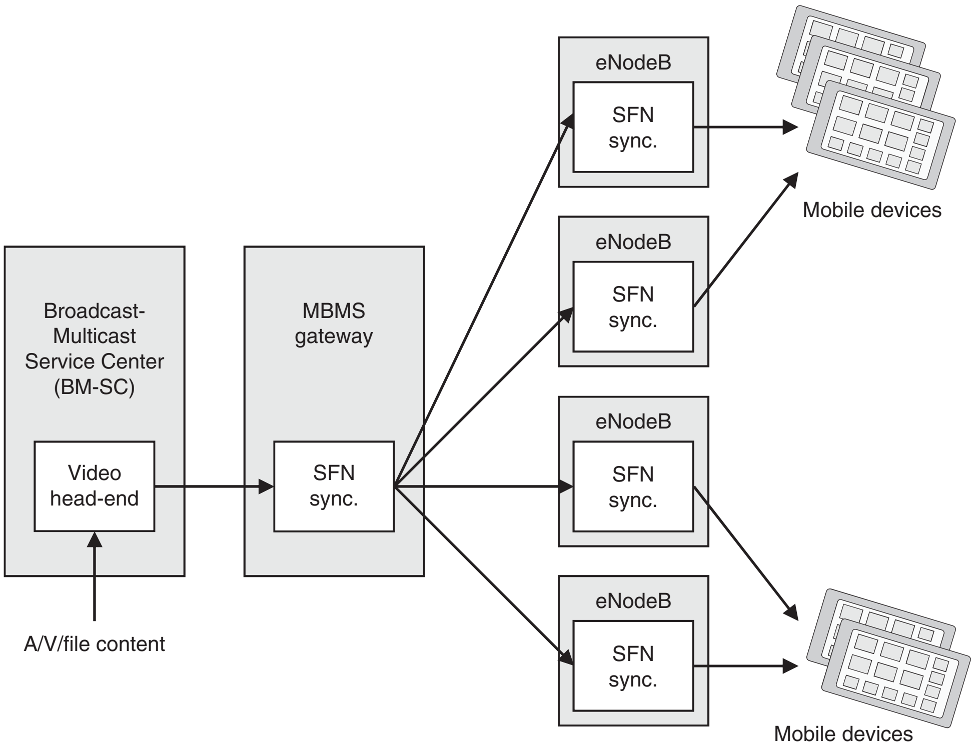

Figure 1.8 *The architectural principle of the Single Frequency Network option of eMBMS.*

towards this solution can be gradual, by first introducing adapters to connect circuit switched elements to IP transport, and finally offering end-to-end IP connectivity.

The handling of increased data traffic via mobile communications networks can also be done in such a way that excess of data is offloaded via some other available network, and merged in the receiving end in a seamless manner. One of the solutions along this development path is the Wi-Fi Offloading concept. The device may be able to detect automatically when the cellular network is experiencing capacity limitations and, possibly available Wi-Fi networks in the same area where the user has access rights. As an example, an operator can operate both cellular and Wi-Fi networks, which makes the handover between these networks logical.

1.3.5 How to Handle Increased Smartphone Signaling

Along with the introduction of smartphones, the signaling load in mobile communications networks has increased rapidly. This created a concrete issue for cellular network operators as of 2010. Prior to that period, the real effect of smartphones was not understood, and the impact on network capacity utilization was not clear. As soon as the popularity of smart devices grew, users especially in the USA and Europe, started to indicate the reduction of the quality of voice connections and data service of mobile networks. In addition to the connection-related problems, there started to be clear indications about poor battery duration of smart devices. These indications appeared when average smartphones generated about a sixth of the traffic of typical laptop-generated data traffic, which was confusing as for the destruction of the cellular network quality.

Previously cellular networks have maintained the inactive handset in Idle mode. This has been due to the fact that the device consumes only a low amount of power in Idle mode. In case of data call initiation, a mobile device entering active mode with a cellular network may generate about 30 signaling messages prior to the activation of the transmission of data. Furthermore, the procedure to return the device to Idle mode generates signaling regardless of whether the device was in fact transmitting or receiving data or small-scale data messaging. This also happens when only background signaling is transferred by, for example, active application. The signaling may take on average up to two seconds in order to start the actual data transfer.

A solution for this – assuming active messaging is taking place – is that the network instructs the handheld device to stay in Active mode for a longer time in order to deliver the forthcoming messages without this initial signaling. By applying this solution, an example of the Nokia Siemens Networks network with Cell_PCH enabled shows that it generated an average of 40 percent fewer signaling from high signaling-generating applications compared to cases where Cell_PCH is not enabled. An additional benefit of this solution, as the results of this case show, is that the respective battery life was noted to increase by up to 30 percent.

This example shows the importance of the design of the complete chain, along with new innovations in networks. The effects on network capacity, for example, may be extremely difficult to estimate without an in-depth analysis prior to the deployment of the new equipment such as smartphone devices. Another practical example of signaling load increase has been the modification of the standard TCP/IP behavior by adjusting signaling message timers in such a way that the user data device (which can be smartphone, USB-based data stick or any other data device) did not return to Idle mode. These keep-alive messages, or so-called heartbeat signaling, increase the signaling load by increasing the keep-alive messaging, for example, every 1–2 seconds, although they reduce the above mentioned, relatively heavy initial signaling in order to make the transition from the terminal’s idle mode to active mode. In order to correctly dimension the network, it is extremely important that all network components, including the user devices, behave as standardized [11].

The development of the smartphone market has been fast since 2010, although devices capable of utilizing data-based applications in cellular networks have already been in the markets for a relatively long time. As the amount of versatile applications increases heavily, the utilization of resources of operators might be challenging to ensure that applications do not cause uncontrolled problems to the network load via heavy side-signaling. It would thus be important to ensure the application developer community understands the common guidelines for the traffic generation. There are also possibilities of creating automated application testing methodology for measuring the signaling load in typical circumstances so that the behavior of applications may be adjusted prior to the commercial phase.

1.3.6 Effects of Online Video

Streaming of the Video-on-Demand (VOD) over the Internet is considered to be a major step in the evolution path of the delivery of media content to the users. Already for various years, satellite and cable companies have offered the possibility to utilized VOD services. Also video rental services have appeared into markets along with the technology platform enablers of the fixed network. Typically, VOD systems provide the users the possibility to browse, select and preview in on-demand way media content which is a part of complete content repository simply by utilizing a home TV set.

Nevertheless, there have been only few large-scale studies of the effects of the VOD. According to the measurements and analysis of Ref. [12], diurnal patterns in user access patterns were noted in China. There was a very high inactivity period also noted which is suitable for maintenance and upgrade operations, in this case morning hours from 5 am to 8 am as it minimizes the impacts on users. For systems deployed in the USA or Europe, the source suspects that there might be more even distribution of accessing the VOD services during the morning hours. It is interesting to note that a relatively high proportion (70%) of sessions was terminated in the first 20 minutes. This means that system caches can maximize the effectiveness of these transmissions by storing a major part of the capacity, especially in the beginning of the movie segments. Furthermore, an important observation was that especially the initial segments of popular movies should be prioritized over the latter segments in any caching scheme. It is also clear that in order to fully optimize the capacity utilization, VOD systems need to be aware of the time-varying user interest patterns. This can be done by partitioning videos into segments and by taking their time index in replica and cache management in an intelligent way based on the historical data of the videos.

Source [12] also has found out that the applying of commonly recognized Pareto Principle of 80–20 rule for the skew of user interest distributions is not necessarily close to the realistic user profiles. The source had studied the situation by providing the opportunity to choose from a wide selection of videos, and has noted that user requests are spread more widely than predicted by the Pareto principle. The results show that 10% of the most popular objects account for approximately 60% of all accesses while 23% of the objects account for 80% of the accesses. According to the analysis, this is a more moderate result than the frequently referred to 80/20 or 90/10 rule. The source has concluded that this moderate Pareto principle is relevant to VOD systems with relatively large libraries. For this reason, VOD systems are likely to require larger than expected caches in order to achieve the same hit rates predicted by the traditional Pareto Principle.

The scientific studies show that the VOD systems can be optimized in order to estimate more accurately the probable need for capacity, and to enhance the rate of estimating beforehand the popularity of the contents. The basic question is thus to seek for optimal balance for caching which provides a fluent user experience, but without exceeding the pre-transferred data because if the customer switches the channel, all the previously transmitted data will be useless and merely wastes network resources.

In a detailed network capacity optimization, the seasonal and short-term variations can be investigated in order to dynamically switch the offered capacity via different network technologies by taking into account, for example, the energy consumption of each system. The idea would be to balance between the power levels per technology, for example, in such a way that the number of elements and thus sites can be minimized. This would be important for selecting the most environmental-friendly solutions, to reduce the CO_2 levels. Depending on the network manufacturer's supported set of functionalities, it could be possible to, for example, switch on and off a part of the transmitters of certain technologies depending on the actual and estimated need for delivered traffic. In the most detailed network optimization, it could be possible to even combine the probability values of daily utilization based on historical utilization data and weather forecasts in order to fine-tune the offered capacity in advance.

1.4 The Focus of the Book

The primary audience for the book is technical personnel of telecom operators, equipment and terminal manufacturers. The book is also designed for professors and students of technical universities and institutes, as well as professionals in standardization bodies, theoretical and hands-on training courses, and telecommunications consultancy personnel. The book should be highly useful also for service providers and consultancy companies, and other technical, marketing and solutions selling personnel.

The structure of the book is modular. The modules apply to the overall structure of the book, as well as to the structure of each chapter, giving both overall descriptions of the architectures and functionality of the typical use cases, as well as deeper and practical guidelines for telecom professionals. The book contains introductory modules of the systems that are suitable for initial studies of the technology. The latter modules are suitable for the experienced professionals who would benefit from the practical descriptions of the physical core and radio architecture, functionality, network planning, end-to-end performance measurements, physical network construction and optimization of the system.

The book is thus aimed to technical personnel. The book gives a complete picture of the field via practical descriptions and case examples, that is, each topic are handled sufficiently deeply in order to understand the principles and details. The book is thus suitable as a complete study material in telecom courses, and as a centralized information source for the industry. Furthermore, the book clearly indicates the references where the more detailed information can be obtained.

The book is a useful set of guidelines for diverse practical situations in telecom networks planning, deployment and multivendor interworking plans, where practical multisystem descriptions are needed in a compact and centralized form. A typical situation where the book would give added value is in hands-on work in telecom projects. In this way, for example, a radio engineer would understand the bottlenecks and possibilities of transmission in order to dimension correctly the radio interface, and vice versa. In addition, this book gives realistic points and messages from the operational field for the academic environment which would help in adjusting the study items into realistic paths.

1.5 Instructions for Reading the Book Contents

The chapters present the selected topics individually. Each chapter contains key aspects such as (when applicable per system): functional blocks and interfaces, protocol layers, technical description of the hardware and software, high-level circuit diagrams of the most important parts of the modules, functionality and performance of the elements, planning, optimization, use cases, challenges, solutions to potential problems, and so on. One of the main ideas of the book is that the chapters can be read independently from each other without any specific order (when relevant, there are cross-references), the general theory part applying to the rest of the chapters. The book aims to provide an appropriate balance between the new systems which will enter the mainstream versus the legacy systems they replace partially or totally.

The contents are divided in a general theory section (I) and in the following modules: (II) Fixed telecommunications; (III) Mobile communications; (IV) Space communications, and (V) Other and special communications. In addition, the contents include a common module for the planning and management of telecommunication networks (VI). The more detailed content per chapter is the following:

Chapter	Description
	I – GENERAL MODULE
1	Introduction • Overall description of the telecom field, status and advances, future paths, world's markets and their regional differences. • Description and analysis of current worldwide trend which indicates that the industry value is shifting towards mobile devices, content, applications and services whilst prices and margins of traditional access services are declining. • Description and analysis of the effects of video services which are going to be the major driver of data traffic growth on fixed and mobile networks, resulting in the increase in user-generated video transmission. • Explanation how the networks can be scaled in order to ensure required user experience and capacity across fixed and mobile platforms and devices along new usage trends. Description of new solutions and their effects to support the advanced services. • Analysis of smartphone markets and effects of increasing signaling load. • Contents description of the book.

Chapter	Description
2	Standardization and regulation • The most important institutes and their roles, documents and information sources. • Introduction for fixed and mobile standardization development. • Description of challenges, for example, in spectrum requirements, differing availabilities and usefulness of different bands at national, regional and global level. • ITU, ETSI, IEEE, CEPT, T1, ARIB and TTC. • 3GPP and 3GPP2. • Broadcast standardization. • Satellite systems standardization. • Other standardization organizations. • Industry forums, including GSM Association, UMTS Forum, WiMAX forum, BMCO forum, and Global Mobile Suppliers Association. • Guideline for finding and interpreting standards. • Regulators (the role, challenges, work procedures).
3	Telecommunications principles • Telecommunications terminology. • Evolution of the systems, applications and needs. • Spectrum allocations, ITU regions and principles, regional aspects, and frequency regulation. • Interference aspects and shielding. • Near future and longer term systems. • Spectral efficiency and optimal utilization. • Physical aspects, including principles of radio interface and radio links, electrical wires, optical fibers.
	II – FIXED TELECOMMUNICATIONS MODULE
4	Wired environment: Protocols • OSI, in theory and in practice. • TCP/IP and UDP. • Error recovery. • LAP protocol family. • Signaling.
5	Connectivity and payment • Principles of the coding theories. • Modern coding. • Examples of the methods.
6	Fixed telecommunications networks • Architecture, functionality, synchronization, requirements, performance, backhaul requirements, implications at element and interface level. • Plain old public telephone system (POTS) and ISDN. • Intelligent network, weight in evolution. • IP Multimedia Subsystem. • Transport solutions, weight in IP transition. • Cloud computing. • Other networks.

(*continued*)

Chapter	Description
7	Data networks • Architecture, functionality, synchronization, requirements, performance. • IP networks (IPv4, IPv6). • TDM, ATM and current development towards IP. • Frame relay. • LAN and evolved versions. • Other data networks (packet/circuit switched).
8	Telecommunications networks applications • Practical guideline for the feasible applications, effect of different users as function of the traffic load, quality of service and application type. • Voice services. • Messaging services. • Audio/video and multimedia services. • Text services.
9	Transmission networks • Principle, functionality, performance, dimensioning, applicability. • Physical transmission systems, including electrical wires, optical, radio links, and LAN/WAN/MAN variants. • Coding techniques. • PCM. • PDH, SDH. • Carrier Ethernet.
	III – MOBILE COMMUNICATIONS MODULE
10	Wireless environment: Modulation • Challenges of information transfer. • Analog modulation methods. • Digital modulation methods. • The mathematical base of modulations. • From analog to digital modulation.
11	3GPP mobile communications: GSM • Architecture, core and radio systems and interface description (elements, modulation, coding, modes, functionality), functionality and performance, frequencies, terminals, planning, dimensioning and optimization of radio, transport and core, site solutions (physical construction, towers, antenna systems), interworking, applications, practical use cases. • Release 1, 2, 2+ and evolution. • Release 97 up to actual. • Evolved version of GSM (voice and data). • Special GSM solutions (GSM-R, machine-to-machine communications, energy saving functionalities, smartphone signaling optimizations features). • Advanced GSM features on the fluent 3G/4G deployment and refarming.
12	3GPP mobile communications: WCDMA and HSPA • Architecture, core and radio systems and interface description (elements, modulation, coding, modes, functionality), functionality and performance, frequencies, terminals, planning, dimensioning and optimization of radio, transport and core, site solutions (physical construction, towers, antenna systems), interworking, applications, practical use cases. • Release 99, 4/5, 6, 7 (relevant aspects for today's network operation). • Release 8 and 9, functionality and modern features.

Chapter	Description
13	3GPP mobile communications: LTE/SAE and LTE-A • Architecture, core and radio systems and interface description (elements, modulation, coding, modes, functionality), functionality and performance, frequencies, terminals, planning, dimensioning and optimization of radio, transport and core, site solutions (physical construction, towers, antenna systems), interworking, applications, practical use cases. • Release 8 and 9 (evolved UTRAN / LTE and evolved core / EPC and their options). • Release 10 (LTE-Advanced, including carrier aggregation, self-optimizing networks, MIMO, adaptive antenna systems). • Future releases.
14	Wireless LAN and evolution • Architecture, functionality and performance, planning and optimization, applications, use cases. • IEEE networks (WLAN, WiMAX, etc.). • Evolved versions.
15	Terrestrial broadcast networks • Architecture, core and radio systems and interface description (elements, modulation, coding, modes, functionality), functionality and performance, frequencies, terminals, planning, dimensioning and optimization of radio, transport and core, site solutions (physical construction, towers, antenna systems), interworking, applications, practical use cases. • Television broadcast systems (European, American, Asian systems). • Radio broadcasting systems (European, American, Asian systems). • Mobile broadcast systems (DVB family, American and Asian systems). • Broadcast via mobile networks (3GPP, American and Asian solutions).
	IV – SPACE COMMUNICATIONS
16	Satellite systems: communications • Architecture, core and radio systems and interface description (elements, modulation, coding, modes, functionality), functionality and performance, frequencies, terminals, planning, dimensioning and optimization of radio, transport and core, site solutions (physical construction, towers, antenna systems), interworking, applications, practical use cases. • 2-way communications. • Broadcast systems.
17	Satellite systems: location services and telemetry • Architecture, core and radio systems and interface description (elements, modulation, coding, modes, functionality), functionality and performance, frequencies, terminals, planning, dimensioning and optimization of radio, transport and core, site solutions (physical construction, towers, antenna systems), interworking, applications, practical use cases. • Positioning systems: GPS, GALILEO, GLONASS. • Other systems.
	V – OTHER AND SPECIAL COMMUNICATIONS
18	Special networks: TETRA • Architecture, core and radio systems and interface description (elements, modulation, coding, modes, functionality), functionality and performance, frequencies, terminals, planning, dimensioning and optimization of radio, transport and core, site solutions (physical construction, towers, antenna systems), interworking, applications, practical use cases. • Principles, role, applicability. • Architecture, functionality. • Trunking, direct mode. • Data services. • Applications.

(*continued*)

Chapter	Description
	VI – PLANNING AND MANAGEMENT OF TELECOMMUNICATION NETWORKS
19	Security aspects of telecommunications: mobile networks (Case: LTE) • Basic principles of the protection (legislatures, technical challenges and possibilities in mobile environment). • Identification of attack types and preparation. • Protection methods (traffic analysis, legal interception, network element and interface protection, user and network authentication / authorization, USIM security. • Charging safety issues, Charging Data Record protection. • Fraud prevention, fraud monitoring. • Use cases. • Recommendations.
20	Planning of modern 2G
21	Planning of evolved 3G
22	Planning of mobile TV
23	Planning of core networks
24	Radiation safety and health aspects
25	Deployment and transition between telecommunications systems • Network deployment cases for the most probable transition scenarios (e.g., evolution of GSM / 3G / other systems towards LTE). • Impacts, dimensioning, rehoming, challenges and how to cope with them.
26	Wireless network measurements • Radio interface measurement types. • Modulation, performance, error rate and other relevant measurements of modern radio networks. • Interpretation of the measurements. • Postprocessing and analysis. • Measurement equipment for the radio and core networks. • Measurement principles and measurement points. • Measurement types and protocol layer analysis. • Performance monitoring; the network statistics, key performance indicators KPI and counters for LTE and SAE. • Interpretation of the system specific error rates, modulation error rates, vector error rates and other error indicators, effect of the parameter setting on the error rates. • Guidelines for the result postprocessing and analysis, general measurement methods • Measurement use cases, and results from real environments.

References

1. Penttinen, J. (2003) *Todo Sobre Los Teléfonos Móviles; Libro Dirigido A Los Usuarios de Telefonía Móvil*, 1st edn in Spanish, Finesstel Ltd, 195 p. ISBN 952−99096−0-8.
2. Eugenio Agacino y Martinez, D. (1899) *Cartilla de Electricidad Práctica*, 4th edn, Naranjo, Encuadernador, Cadiz.
3. Radio amateur information: http://www.arrl.org/what-is-ham-radio (last accessed 17 July 2014).
4. American Radio Relay League: www.arrl.org (last accessed 17 July 2014).
5. Jokela, Leszczynski, Paile, *et al*. Radiation safety of handheld mobile phones and base stations. STUK-A161. Helsinki 1999. 76 pp. http://www.stuk.fi/julkaisut/stuk-a/a161_1.pdf (last accessed 17 July 2014).
6. Pillar 4. Fast and ultra fast Internet access. Digital Agenda Scoreboard 2011, European Union. http://ec.europa.eu/information_society/digital-agenda/scoreboard/docs/pillar/broadband.pdf (last accessed 17 July 2014).

7. ITU statistics: http://www.itu.int/ITU-D/ict/statistics/ (last accessed 17 July 2014).
8. Mobile Economy Europe 2013 of GSMA: http://gsma.com/newsroom/wp-content/uploads/2013/12/GSMA_ME_Europe_Report_2013.pdf (last accessed 17 July 2014).
9. Björklund, D., Rautio, J., Penttinen, J. (1994) NMTImage. DMR (Digital Mobile Radio) Conference, June.
10. Opportunity and impact of video on LTE Networks. White paper, Motorola. 20 p.
11. Penttinen, J.T.J. (2011) *The LTE-SAE Deployment Handbook*, John Wiley & Sons, Ltd.
12. Understanding User Behavior in Large-Scale Video-on-Demand Systems: https://www.cs.ucsb.edu/~ravenben/publications/pdf/vod-eurosys06.pdf (last accessed 17 July 2014).
13. Nokia Siemens Networks. Solving the mystery of network signaling. http://nsn.com/news-events/publications/unite-magazine-issue-8/solving-the-mystery-of-network-signaling (last accessed 17 July 2014).
14. Crossbar exchange: http://www.thg.org.uk/ (last accessed 23 August 2014).
15. Telecom history: http://www.ericssonhistory.com (last accessed 23 August 2014).

2

Standardization and Regulation

Jyrki T. J. Penttinen

2.1 Introduction

Without commonly agreed rules for communications within and between telecommunications systems, with functionalities and interfaces, the end-to-end calls between devices of operators and countries would not be possible. Standards are thus designed at all levels both nationally and internationally. In addition to systems, within wider communities like the European Union, there are also processes to unify definitions in order to provide fluent end-to-end functionality and performance.

This chapter gives an introduction to the most important institutes and their roles in telecommunications standardization and definition of new solutions, including guidelines on how to identify information sources and how to seek specific information.

This chapter gives an overall view as well as sufficiently deep details to understand the focus and aim of each group involved in the definitions of common rules for fixed and mobile system development. This chapter also gives practical points of view about challenges, for example, in spectrum requirements, differing availabilities and usefulness of different bands at national, regional and global level.

2.2 Standardization Bodies

2.2.1 ITU

ITU (International Telecommunications Union) is a worldwide organization that takes care of the global requirements of telecommunications. ITU belongs to the United Nations (UN). ITU is specialized for information and communication technologies (ICT).

The tasks of ITU include the allocation of global radio spectrum and satellite orbits, development of technical standards. In addition, ITU paves the way to enhance global access to ICTs to underserved communities. In general, the high-level focus of ITU is to facilitate people's right to communicate, which enhances different fundamental functions of the human lifestyle, including emergency services, water supplies, power networks,

The Telecommunications Handbook: Engineering Guidelines for Fixed, Mobile and Satellite Systems, First Edition.
Edited by Jyrki T. J. Penttinen.

health care, education, government services, financial markets, transportation systems and environmental management, via the developed communications [1]. ITU has both public and private sector membership, including 193 Member States, ICT regulators, leading academic institutions and about 700 private companies.

The ITU organization is divided into three main parts:

- ITU-T: Standardization of the telecommunications area.
- ITU-R: Standardization of radio area.
- ITU-D: Development of the telecommunications area.

The work at ITU is typically executed in study groups of experts. The study groups have a certain, limited focus in order to guarantee correct competence of participants. The task of the study groups is to plan frameworks providing optimum functioning of telecommunications services. The concrete result of the work is to form technical standards or guidelines, that is, Recommendations of ITU.

The fundamental principle of ITU is to work independently from commercial interests. This can be noted by the methods of work of group experts, which logically represent sometimes certain commercial interests. Neutral and democratic decision making is assured within ITU, so the groups provide a neutral platform for global consensus. ITU offers thus an efficient service to telecommunications industry, which is in fact the main driver for social and economic development.

2.2.1.1 ITU-T

The Telecommunication Standardization Sector of ITU (ITU-T) produces standards, that is, "Recommendations," which function as an essential building block for the operation of modern ICT networks. ITU standards provide a compatible way to communicate at a global level, and thus, for example, the interoperability of the public telephone network functions everywhere, as well as the utilization of Internet, is equally possible in a compatible way in every spot of the globe. This is facilitated by common transport protocols and compatible voice and video compression, among other aspects of ICT, via ITU standards. ITU produces actively new standards in all the areas of telecommunications.

2.2.1.2 ITU-R

The Radiocommunication Sector of ITU (ITU-R) coordinates an important set of radiocommunication services. ITU-R also manages globally the radio-frequency spectrum and satellite orbits. The development of new radio network systems has resulted in a growing demand of spectrum, and the allocation of the limited resources is managed at the highest level by ITU-R. One of the working methods of ITU-R is the organization of conferences and study group activities which executes high-level planning of mobile broadband communications and broadcasting technologies. The importance of these activities is growing, along with the introduction of new systems in the commercial environment.

As an example, for the UMTS definitions, the most important group of ITU-R has been ITU-R TG 8/1. The work of this group – along with other technical areas – has been related to radio regulations that are handled at the WRC (World Radio Conference) approximately every two years.

At the moment, the most important focus area of ITU-R is the development of International Mobile Telecommunications-Advanced (IMT-Advanced) systems. The definition of IMT-A is that they are mobile systems including new capabilities to the previously defined requirement set of IMT-2000. These further developed systems provide access to various telecommunication services via both mobile and fixed networks. The common characteristic of these systems is the packet switched technique, which has been one of the most fundamental changes in the history of previous circuit switched techniques of telecommunications systems.

Systems belonging in the set of IMT-Advanced systems are compliant with a variety of mobility scenarios, with applications and a range of data rates that are aligned with user and service demands in multiple user environments. The special focus of IMT-Advanced is to provide capabilities for high quality multimedia applications which can be done via various services and platforms. In general, the requirements of IMT-Advanced have been designed in order to provide significant improvements for the performance and quality of service of the systems.

A complete list of ITU-R recommendations for the IMT-Advanced can be found in Ref. [2].

2.2.1.3 ITU-D

The Telecommunication Development Sector of ITU (ITU-D) is active in various major initiatives globally related to functions such as its ITU Connect events or Connect a School, Connect a Community. In addition, ITU actively publishes ICT statistics of the field.

There are various subgroups in ITU-T that produce, for example, IMT-2000 related requirements.

2.2.1.4 ITU Recommendations

A complete recommendation set of ITU-T can be found in source [3] and the set of recommendations of ITU-R is located in Ref. [4].

Table 2.1 presents the series of ITU-T.

The recommendations of ITU-T are numbered by the prefix shown above, followed by a decimal point and an integer number. An example of recommendations is V.42bis which was adopted – in addition to the modems – also in the early stage GPRS network as an internal data compression method.

In general, the tasks of ITU-T include the creation of the highest level standards 7 requirements of the third and fourth generation mobile communications systems.

The division of the recommendations of ITU-R is presented in Table 2.2.

2.2.2 ETSI

ETSI (The European Telecommunications Standards Institute) produces globally applicable standards for ICT including fixed, mobile, radio, broadcast, Internet, aeronautical and other areas.

In the area of mobile communications, one of the most significant tasks of ETSI has been to create the first GSM and UMTS standards. The GSM standardization work was executed in various groups under the name SMG (Special Mobile Group). Since 1999, 3GPP (Third Generation Partnership Project) has taken over almost all the areas of the further development of GSM and UMTS. Among various other major activities, another important task of ETSI was the acceptance of DVB-H (Digital Video Broadcasting, Hand-held).

The European Union recognizes ETSI as an official European standards organization, and one of the aims of ETSI is to provide access to European markets. ETSI has goals to produce high quality standards with low time-to-market. In order to achieve these goals, ETSI collaborates with research bodies. ETSI is also active in significant complementary areas, including interoperability of the systems. ETSI provides facilities for the standardization work in the main location, Sofia Antipolis in France.

ETSI clusters represent different ICT standardization activities. The clusters are basically indicating the major component of a global ICT architecture. ETSI is executing standardization work under all the clusters in various Technical Committees (TC) and Working Groups (WG). In addition, the work of a certain TC may be represented in several clusters. Figure 2.1 presents the ETSI clusters.

The ETSI standards creation process is defined in Ref. [5]. It is applied in the new proposal for a standard topic. The initial need for a common standard is indicated by an ETSI member, or typically also jointly by a

Table 2.1 The ITU-T recommendations

Series	Description
A	Organization of the work of ITU-T
B	Means of expression: definitions, symbols, classification
C	General telecommunication statistics
D	General tariff principles
E	Overall network operation, telephone service, service operation and human factors
F	Nontelephone telecommunication services
G	Transmission systems and media, digital systems and networks
H	Audiovisual and multimedia systems
I	Integrated services digital network
J	Cable networks and transmission of television, sound program and other multimedia signals
K	Protection against interference
L	Construction, installation and protection of cables and other elements of outside plant
M	Telecommunication management, including TMN and network maintenance
N	Maintenance: international sound program and television transmission circuits
O	Specifications of measuring equipment
P	Terminals and subjective and objective assessment methods
Q	Switching and signaling
R	Telegraph transmission
S	Telegraph services terminal equipment
T	Terminals for telematic services
U	Telegraph switching
V	Data communication over the telephone network
X	Data networks, open system communications and security
Y	Global information infrastructure, Internet protocol aspects and next-generation networks
Z	Languages and general software aspects for telecommunication systems

Table 2.2 The classification of ITU-R standards

Series	Description
BO	Satellite delivery
BR	Recording for production, archival and playout; film for television
BS	Broadcasting service (sound)
BT	Broadcasting service (television)
F	Fixed service
M	Mobile, radio determination, amateur and related satellite services
P	Radio wave propagation
RA	Radio astronomy
RS	Remote sensing systems
S	Fixed-satellite service
SA	Space applications and meteorology
SF	Frequency sharing and coordination between fixed-satellite and fixed service systems
SM	Spectrum management
SNG	Satellite news gathering
TF	Time signals and frequency standards emissions
V	Vocabulary and related subjects

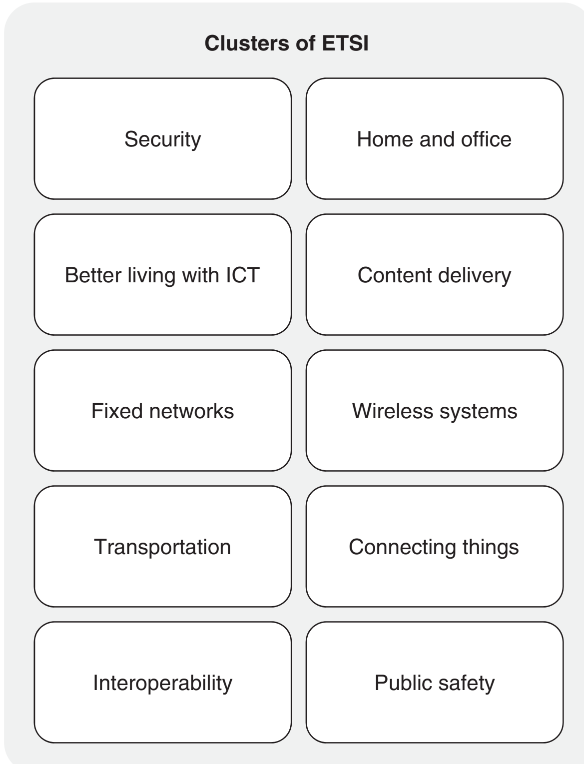

Figure 2.1 *The clusters of ETSI.*

group of members. This triggers a concrete proposal, which is handled in a technical committee that is related to the initiation. At this point, the proposal is called as a work item (WI).

Technically, this new work item is documented in a form which is visible to all at ETSI. It shows the ETSI deliverable type, title, scope and schedule of the work item, and the person responsible for drafting the standard. The person is called a rapporteur. The form also lists the ETSI members who support the work. After the item has been visible during a certain time period, and no objections have been raised, the work item is considered adopted by the membership, and the actual work of the standard drafting can be initiated. The items typically require various meetings with subject matter experts, and the time for the standardization in this phase can take from some months up to more than one year. After the drafting has been finalized and agreed in a working group, the standard then enters into the appropriate approval process according to its deliverable type.

The aim of this ETSI Standards Making Process (SMP) is to identify a market need and make sure that its compatibility is taken into account in the development. The participating bodies in the process are ETSI Technical Organization, the ETSI members and the ETSI National Standards Organizations. The process is based on the real market need and the standardization of the respective solutions based on democratic decision making. The process is defined in the ETSI Directives.

SMP consists of identifying needs for standardization, mapping a suitable technical committee for the standardization work flow, identification, definition, approval and adoption of work items, drafting and editing the contents, and finally the actual publication. The work is done by applying high quality of work and contents, making sure the work advances on time.

SMP is applied to different types of ETSI documents, which are Technical Specifications (TS), Technical Reports (TR), Group Specifications (GS) and Special Reports (SR). Following the ETSI Technical Working Procedure, the draft specification that has been approved and adopted by the TC, is then submitted to the ETSI Secretariat, which publishes the TS, TR or SR. The draft that is approved by TC goes to Membership Approval Procedure (MAP), which applies to ETSI Guides (EG) and ETSI Standards (ES).

There are also One-step Approval Procedure (OAP) and Two-step Approval Procedure (TAP) that are utilized for the creation of European Norms (EN) and European Harmonized Standards (HS).

2.2.3 IEEE

IEEE provides a wide range of publications and standards that make the exchange of technical knowledge and information possible among technology professionals [6]. The contents of these outcomes is delivered through the IEEE Xplore Digital Library. One of the most important areas of IEEE standards is the series of IEEE 802. These standards can be accessed easiest via [7], which contains the following, most relevant documentations (names have copyright):

- IEEE 802: Overview & Architecture.
- IEEE 802.1: Bridging & Management.
- IEEE 802.2: Logical Link Control.
- IEEE 802.3: Ethernet.
- IEEE 802.11: Wireless LANs.
- IEEE 802.15: Wireless PANs.
- IEEE 802.16: Broadband Wireless MANs.
- IEEE 802.17: Resilient Packet Rings.
- IEEE 802.20: Mobile Broadband Wireless Access.
- IEEE 802.21: Media Independent Handover Services.
- IEEE 802.22: Wireless Regional Area Networks.

2.2.4 IETF

2.2.4.1 Structure

The Internet Engineering Task Force (IETF) is an international and open community of network operators, vendors, designers, and researchers. The task of IETF is to be actively involved in the evolution of the Internet architecture and to assure the Internet is and will function smoothly. The outcome of the work of IETF is documented as recommendations, which are numbered as RFC *n*, where n is an integer number referring to a specific area of the definition. The list of the documents can be found in Ref. [8], and the specific documents can be searched in Ref. [9].

IETF is organized into working groups, each focusing to a certain technical area. The way of work of IETF is widely based on email lists and exchanging comments thus on line. This method greatly differs from the more formal type of meetings that, for example, ETSI has. Nevertheless, IETF also organizes meetings, but only three times per year.

The more specific organization of IETF groups is based on Area Directors (AD). They are members of the Internet Engineering Steering Group (IESG). The General AD chairs the IESG and IETF, and is an ex-officio member of the IAB, the Internet Architecture Board, which in turn gives input about the architectural oversight. As IAB works at a high level, some of its tasks are related also to the adjudication of appeals in case there are complains about the decisions of IESG, assuring democratic ways of working. IAB and IESG have been mandated by Internet Society (ISOC) for these situations.

In addition, there are tasks like assignment of parameter values for Internet protocols, which is taken care of in a centralized way by the Internet Assigned Numbers Authority (IANA), chartered by the Internet Society (ISOC).

For more information on the work and structure of IETF, useful guidelines can be found in Ref. [10], available in IEFT web pages, and the IETF documents can be found in Ref. [11].

2.2.4.2 IETF Standards Creation Process

The basic idea of the process for the IETF standards creation can be found in Ref. [12]. It should be noted though that it has been amended several times since its publication. In practice, the intellectual property rules can currently be found in Refs. [13, 14]. As there are also other updates for the process, methods of working can be found in any case in the updated version of Ref. [12].

A high-level summary of Internet standards creation is as follows:

- First, the specification has a development period. During this time, it is reviewed by the Internet community in an iterative way. The experience of the reviewers plays a major role at this stage. This is the most challenging stage as the aim of the work is to provide a high quality specification that is respected by the wide community, taking into account the different interests of the participating entities. The definitions should also be sufficiently clear in order to be understood by all those affected.
- After the reviews, the specification proposal is adopted as a Standard by the appropriate body.
- After the adaptation, the standard is ready for publishing.

The basic principle of IETF is to guarantee sufficient time for the participants to comment the development – yet without jeopardizing the overall goals of the implementation timelines. This is logically challenging since the network is evolving rapidly. Nevertheless, in order to assure the high quality of the standard, there is typically a prior implementation and testing period. It is interesting to note that thanks to this phase, for example, the IPv6 has been tested – but this happened under the name IPv5 which was already reserved as a title to this testing phase. This explains why the actual implementation is called IPv6 instead of IPv5.

2.2.5 CEPT

2.2.5.1 General

The European Conference of Postal and Telecommunications Administrations (CEPT) has been initiated in 1959, when 19 countries formed the first setup. It soon expanded as new members joined, and currently CEPT has 48 members, covering nearly all Europe. At the initial phase, the members were monopoly-holding postal and telecommunications administrations, having cooperation on commercial, operational, regulatory and technical standardization issues.

2.2.5.2 Formation of the Organizations

The current setup contains the actual CEPT organization [15], the Electronic Communications Committee [16], the European Communications Office [17], the Committee for ITU Policy [18], and the European Committee for Postal Regulations [19].

The relationship between European Telecommunications Standards Institute (ETSI) and CEPT was formed in 1988 when CEPT created ETSI. ETSI then took over the standardization of telecommunication items. Furthermore, in 1992 the postal and telecommunications operators created their own organizations which were named as Post Europe and ETNO. This was the time to separate internationally in Europe the postal and telecommunications operations from policy-making and regulatory functions. As a result, CEPT took the responsibilities related to the policy-makers and regulators.

Since 1995, the role of CEPT has been to offer its members the possibility to, for example, establish a European forum for discussions on sovereign and regulatory issues in the field of post and telecommunications issues, provide assistance among members with regard to the settlement of sovereign/regulatory issues, and exert an influence on the goals and priorities in the field of European Post and Telecommunications through common positions. There are many other areas also where ETSI participates, from shaping the areas coming under its responsibilities, increasing cooperation with Eastern and Central European countries, and influencing developments within ITU and UPU (Universal Postal Union) in accordance with European goals.

CEPT has the following committees: CERP (Comité européen de Réglementation Postale) for postal matters, as well as ERC (European Radiocommunications Committee) and ECTRA (European Committee for Regulatory Telecommunications Affairs) for electronic communications. These committees harmonize activities and adopt recommendations via the working groups and project teams.

The European Radiocommunications Committee established 1991 the European Radiocommunications Office (ERO). Its aim is to support the activities of the committee and execute studies internally and for the European Commission (EC). The development further resulted in ECTRA establishing the European Telecommunications Office (ETO) in 1994, which has same type of focus. Then, ERO and ETO have been merged in practice in 2001 as the European Communications Office (ECO) which finally had official organizational status in 2009.

Finally, the two committees that previously handled separately the radiocommunications and telecommunications topics were replaced by the Electronic Communications Committee.

2.2.6 T1

T1 stands for Accredited Standards Committee on Telecommunications. T1 develops telecommunications standards, definitions and technical reports for the needs of the USA. T1 has 6 technical subcommittees (TSC), which are administred by T1 Advisory Group (T1AG). As an example, the relevant group as for the GSM development is T1P1 (Wireless/Mobile Services and Systems), which is further divided into five subgroups: T1P1.1 for International Wireless/Mobile Standards Coordination, T1P1.2 for Personal Communications Service Descriptions and Network Architectures, T1P1.3 for Personal Advanced Communications Systems (PACS), T1P1.5 for PCS 1900, and T1P1.6 for CDMA/TDMA.

The other T1 groups are: T1A1 (Performance & Signal Processing), T1E1 (Interfaces, Power & Protection for Networks), T1M1 (Internetwork Operations, Administration, Maintenance, & Provisioning), T1S1 (Services, Architectures & Signaling), and T1X1 (Digital Hierarchy & Synchronization).

2.2.7 ANSI

The high level aim of the American National Standards Institute (ANSI) is to strengthen the position of the USA in the global economy. It also takes care of the important safety and health issues of consumers, as well as the protection of the environment [20].

ANSI is involved in the creation of norms and guidelines of various different areas. Some examples of the big variety of standards include acoustical devices, construction equipment, dairy and livestock production, as well as energy distribution. In addition, ANSI is involved in the accrediting programs for enhancing conformance to standards. Some examples of these activities are the quality standard ISO 9000 and environmental standard ISO 14 000 management systems.

Although ANSI is not involved directly in the development of American National Standards (ANSs), it provides a neutral venue for the American parties, in order to facilitate the actual working for achieving common agreements. The process is thus about creation of voluntary standards in an open manner. One part of the transparent way of work is guaranteed by granting access to the standards process and appeals mechanism for all the parties that are directly or materially affected by a standard under development.

One of the tasks of ANSI is to promote the use of the US standards in the international environment. On the other hand, it also facilitates the adoption of international standards for the US environment in case these are seen to be suitable for the needs of the user community. Importantly, ANSI is the sole representative of the USA of the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC), the latter connection created via the US National Committee (USNC). ANSI is, in fact, a founding member of the ISO and has an important weight in its work.

At the national level, ANSI provides a forum for over 200 ANSI-accredited standards developers. These organizations present both private and public sectors, and their task is to develop voluntary national consensus standards and American National Standards (ANS) in cooperation. For being a member of ANSI accreditation, standards developers must follow the compliance of the requirements or procedures via the "ANSI Essential Requirements: Due process requirements for American National Standards" [21]. Among other guidelines for the processes and standards policies, it also includes the ANSI patent policy, with the inclusion of Patents in American National Standards.

ANSI offers an online library with open documents, as well as restricted documents for only members [22].

2.2.8 ARIB

ARIB (Association of Radio Industries and Businesses) is the Japanese standardization body that has strongly developed the third generation definitions of mobile communications, and interacts in cooperation with other bodies developing further stages of mobile communications.

2.2.9 TTC

TTC (Telecommunications Technology Association) is another Japanese standardization body that has been actively developing mobile communications systems. The most important goals of ARIB are to execute investigation, research and development, as well as consultation of utilization of radio spectrum as seen by industry development. ARIB also promotes realization and popularization of new radio systems in the field of telecommunications and broadcasting.

Some of the activities of ARIB, related to the radio spectrum of telecommunications and broadcast are:

- Investigation, research and development on utilization of radio spectrum. Some examples of activities are: investigation on new uses of radio spectrum for telecommunications and broadcasting, and survey trends of demands, technology and radio industries. One of the important work groups is the IMT-2000 (International Mobile Telecommunications-2000) Study Committee which carries out technical studies on IMT-2000 and the systems beyond IMT-2000, with the aim to contribute to International Telecommunication Union standardization activities for these systems and to cooperate with other international standardization organizations.

Table 2.3 The joint work bodies of TTC

Organization	Title
TTA (Telecommunications Technology Association)	Memorandum for mutual cooperation between ARIB and TTA. Since 1996.
RAPA (Korea Radio Promotion Association)	Memorandum for mutual cooperation between ARIB and RAPA. Since 1996.
3GPP (3rd Generation Partnership Project)	Third Generation Partnership Project Agreement. Since 1998.
3GGP2 (American 3rd Generation Partnership Project)	Third Generation Partnership Project Agreement for 3GPP2. Since 1999.
CCSA (China Communications Standards Association), TTA, TTC (Telecommunication Technology Committee)	Memorandum of Understanding for mutual cooperation among CCSA, ARIB, TTC and TTA. Since 2002.
KORA (Korea Radio Station Management Agency)	Memorandum of Understanding between KORA and ARIB. Since 2005.
SMPTE (Society of Motion Picture and Television Engineers)	Memorandum of Understanding between ARIB and SMPTE. Since 2007.
ICU (Infocommunication Services Market Participants Union)	Memorandum of Understanding between ARIB and ICU. Since 2009.
GISFI (Global ICT Standardization Forum for India)	Letter of Intent between GISFI and ARIB. Since 2009.
ITU (International Telecommunication Union), ARIB, CCSA, TTA, TTC	Memorandum of Understanding between ITU, ARIB, CCSA, TTA and TTC. Since 2011.
ETSI (European Telecommunications Standards Institute)	Cooperation Agreement between ETSI and ARIB. Since 2011.

- Consultation and publications on utilization of radio spectrum. An example of these activities is the radio link design for fixed microwave circuits and satellite circuits, estimation of radio interference from existing radio stations and selection of best available frequency.
- Developing of technical standards for radio systems. The standards for telecommunications and broadcasting radio systems are created via ARIB Standards (ARIB STD). These standards act as a complimentary element for the mandatory requirements of Ministry of Public Management, Home Affairs, Posts and Telecommunications regulations.
- Correspondence, coordination and cooperation with international organizations. ARIB actively cooperates with other organizations through the mutual agreements.

For the international cooperation, Table 2.3 summarizes the joint work bodies.

2.2.9.1 TTC Standards

TTC is contributing to the telecommunications area. The TTC Standard Summary can be found via the Internet in Ref. [23]. This website provides an introduction of the TTC Standards that are produced and approved by the Technical Assembly (TA). In the numbering system of TTC, the standards based on the ITU-T recommendations are indicated with the identification of "JT," and ISO standards are indicated as "JS." The third category is indicated by "JF," and they are related to the standards that are globally recognized by international forums. On this Internet page, the own standards of TTC would be indicated as "JJ," but at this point they are not available without a fee because they have not been defined as international standards.

The standards areas of TTC include the following: Internetwork Transmission, ISDN Internetwork Signaling, IP Signaling, Packet and Frame Relay, Mobile Communications, Intercarrier Signaling Interface, ISDN User-Network Interface, B-ISDN, PBX, LAN, MHS/OSI, Telematics, Infrared Communications, Network Management, NGN, Voice Coding, Video Coding/AV Communications, IPTV, and Home Network.

2.2.10 3GPP

3GPP (Third Generation Partnership Project) was established by several standardization bodies in 1998. Its original aim has been to develop and enhance 3rd generation mobile communications systems. The original cooperating parties within 3GPP were: ARIB, ETSI, T1, TTA and TTC. In addition to the setup of 3GPP, there is also a group for taking care of the American variant, in 3GPP2. Ever since there have been more participants joining in the activities of 3GPP.

3GPP is formed by a main level group called PGC, project coordination group, and a total of 5 subgroups which are:

- Services and system aspects (SA)
- Radio access network (RAN)
- Core network (CN)
- Terminals (T)
- GSM EDGE Radio Access Network (GERAN).

Especially as a continuum to the previous GSM related work which was done in ETSI, the GERAN took over almost all of the GSM activities in July 2000. It continues the GSM development of both ETSI and T1P1, including the evolved data services like GPRS and EDGE.

3GPP unites participating telecommunications standards bodies, which are officially called Organizational Partners. The aim of the 3GPP work is to provide to the members with a productive environment to create Reports and Specifications defining 3GPP technologies.

The initial aim of 3GPP was to release Technical Specifications and Technical Reports for 3G Mobile System that was based on the evolution of GSM radio and core networks. The radio definitions are based on Universal Terrestrial Radio Access (UTRA) which includes both Frequency Division Duplex (FDD) and Time Division Duplex (TDD) modes.

As has been the case in an accelerating speed with already previous setups like ETSI, the technologies are constantly evolving and forming commercial mobile communications system generations. The original aim of 3GPP was to develop the previous GSM system towards the following, 3rd generation era. The 3rd generation has already been developed for a relatively long time. The new era was initiated via the definitions of Release 8 3GPP specifications including LTE, which can be considered as one of the pre-4G representatives of the mobile communications systems. Along with this development, together with EPC, forming the complete base for the 4G defined as of the Release 10 specifications, the development of LTE has become the most important focus of 3GPP. From 3GPP Release 10 onwards, the LTE/SAE defined by 3GPP is compliant with the requirements of ITU-R for the 4G, that is, IMT-Advanced systems beyond 3G. Along with the new era, the 3GPP standard provides the peak data rates of 100 Mbit/s for high mobility and 1 Gbit/s for low mobility communication, which is one of the key milestones of the 4G as seen via ITU-R.

3GPP shares information about development openly via Internet pages (www.3gpp.org). It is straightforward to find further details of the developed features and current work items. A logical starting point for further information seeking is the Work Plan of 3GPP, which gives an overall introduction to current works. For each 3GPP release, there is also an overview document.

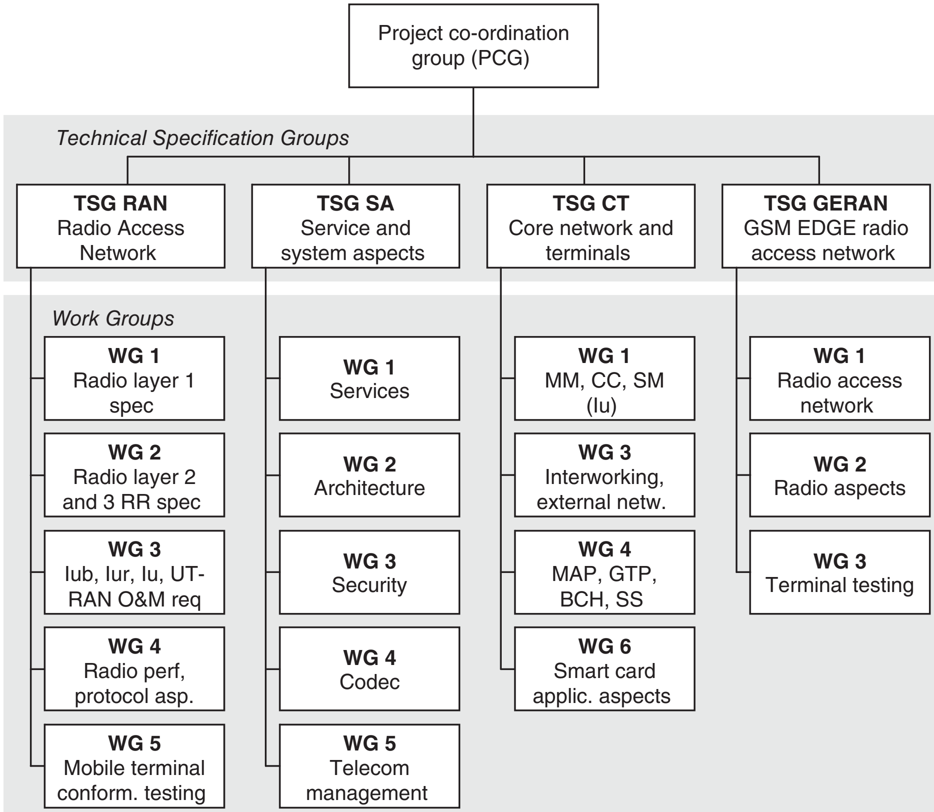

Figure 2.2 *3GPP standardization is done in Technical Specification Groups (TSG) that includes Radio Access Networks (RAN), System Architecture (SA), Core Networks and Terminals (CT) and GSM EDGE Radio Access Network (GERAN). The final standards are approved or rejected in plenary meetings.*

Since the early days of GSM specification via ETSI, the standardization work is now ongoing via 3GPP in such a way that there is a regular set of new standard releases. The most relevant specifications are updated up to four times a year.

The organization and work methods of 3GPP are democratic. Solutions are thus proposed and evaluated in a transparent way, and proposals are finally selected as official work items based on voting in subgroups and in the quarterly TSG plenary meetings of 3G. The TSG GERAN meets five times a year though.

Figure 2.2 shows the current setup of the standardization groups of 3GPP. The RAN groups have defined the LTE specifications whilst SA has made respective definitions for the SAE.

All relevant definitions of GSM, UMTS and LTE/SAE can be found in the 3GPP web page [24]. The 3GPP standards and specifications numbering consist of the main level item, and the latter part indicates the more specific technical area for each Release. A practical way to seek for the definitions is to investigate first the overall specification numbering scheme. The 3GPP specifications are divided into thematic topics that are indicated by the first number of the specification that can be technical specification (TS), or Technical recommendation (TR) according to Table 2.4.

The specification title thus consists of the main technical area and sub-classification (after the decimal point), including the version number and the date when that specific version has been formalized in the 3GPP

Table 2.4 The 3GPP specification division

Item	Specification/recommendation numbering: GSM specific, original ETSI nr/Release 4 and beyond	Specification/recommendation numbering: 3G and GSM, Release 99 and later
Requirements	01 / 41	21
Service aspects of stage 1	02 / 42	22
Technical realization of stage 2	03 / 43	23
Signaling protocols of stage 3, UE – network	04 / 44	24
Radio aspects	05 / 45	25
Voice codecs	06 / 46	26
Data	07	27
Signaling protocols of stage 3, RSS – CN	08 / 48	28
Signaling protocols of stage 3, intrafixed-network	09 / 49	29
Program management	10 / 50	30
Subscriber Identity Module (SIM), Universal Subscriber Identity Module (USIM), IC cards, test specifications	11 / 51	31
Operations, Administration, Maintenance, and Provisioning (OAM&P) and charging	12 / 52	32
Access requirements and test specifications	13	
Security aspects	—	33
UE, SIM and USIM test specifications	11	34
Security algorithms	Not public	35
Evolved UTRA (LTE) and LTE-Advanced radio technology	—	36
Multiple radio access technology aspects	—	37

acceptance processes. The front page of 3GPP LTE/SAE specification consists of the information as seen in the example of Figure 2.3.

As for the most relevant current topic of 3GPP, LTE and SAE, the Release 8 is the first one where the respective definitions are found. For LTE/SAE, the evolution includes the following releases: Release 8 for the base for LTE, Release 9 for relatively minor additions of LTE, Release 10 for the base for LTE-Advanced, that is, 4G, and Release 11 and beyond for the enhancements for LTE-Advanced.

The publication of Release 8 and Release 9 provides to the telecommunications industry a logical evolution towards the next generation mobile networks, which in terms of 3GPP refers to LTE. The common standardization makes LTE compatible with legacy networks, which is one of the important arguments of the evolution. As the 3GPP work is a continuous task, there will be further milestones, including Release11. Its focus can be expected to be on the further maturation of LTE and LTE-Advanced, including Systems Architecture Evolution (SAE). The work also continues due to LTE deployments which will reveal further development items.

2.2.11 3GPP2

The Third Generation Partnership Project 2 (3GPP2) is a third generation (3G) telecommunications specifications creating project meant for the North American and Asian markets. It is functioning in a cooperation

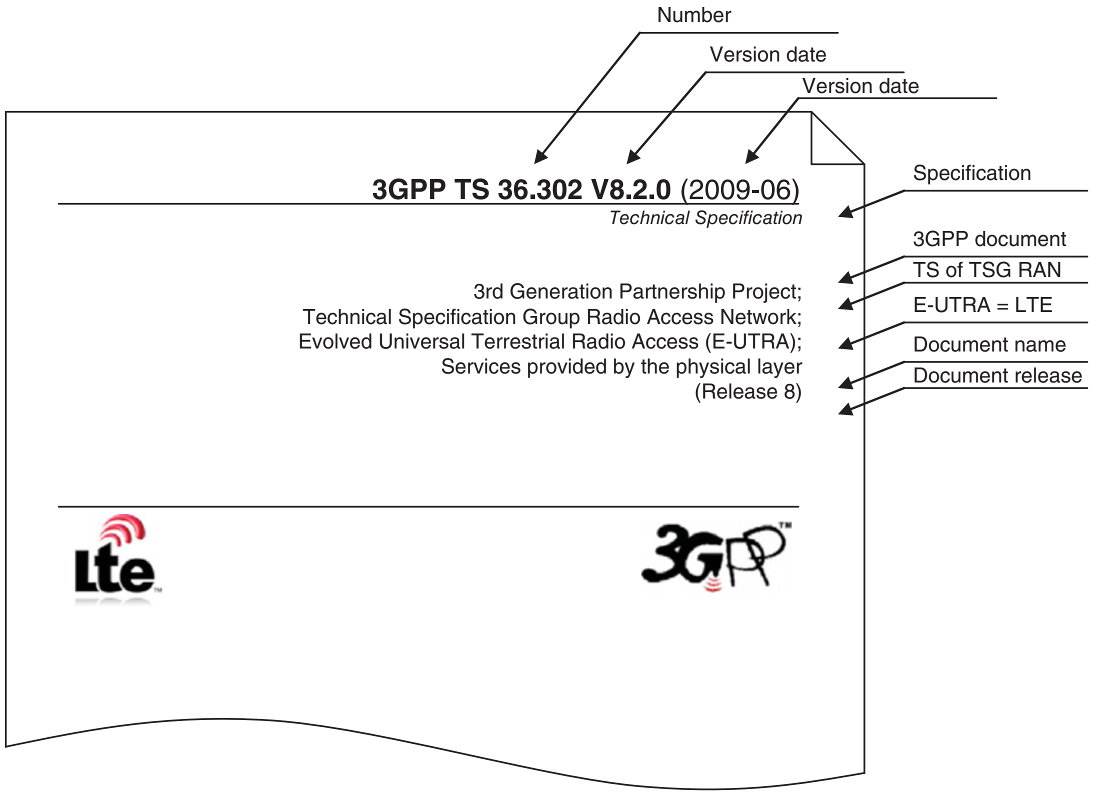

Figure 2.3 *An example of the 3GPP standard's cover page. The LTE and 3GPP logos printed by courtesy of ETSI.*

with 3GPP, and its aim is to develop global specifications for ANSI/TIA/EIA-41 Cellular Radiotelecommunication Intersystem Operations network evolution to 3G and global specifications for the radio transmission technologies (RTTs) supported by ANSI/TIA/EIA-41.

The initiation of 3GPP2 was triggered via the IMT-2000 initiative of ITU. The focus areas of 3GPP are the high speed, broadband, and Internet Protocol (IP)-based mobile systems that feature network-to-network interconnection, feature/service transparency, global roaming and seamless services. In the beginning of 3GPP, there were discussions about embedding both European (ETSI) and American (ANSI-41) development under the same 3GPP concept. After all, it was noted to be most straightforward to create two parallel streams of organizations, the American organization being the 3GPP2. The idea of the work is in any case the same as in 3GPP, taking the advantage of the benefits of collaboration.

3GPP2 is a collaborative effort between five officially recognized standardization development organizations: Japanese ARIB (Association of Radio Industries and Businesses) and TTC (Telecommunications Technology Committee), Chinese CCSA (China Communications Standards Association), North American TIA (Telecommunications Industry Association), and Korean TTA (Telecommunications Technology Association). These organizations are known as Organizational Partners (OP). In practice, the individual member company wishing to participate in the 3GPP2 activities should be affiliated with at least one of the OPs. In addition, there is also a group of Market Representation Partners (MRP) which has the possibility to offer market advices about the services, features and functionality relevant to the points of view of 3GPP2. These partners are currently CDMA Development Group (CDG), IPv6 Forum, and Femto Forum.

There are four Technical Specification Groups (TSG) in 3GPP2, each consisting of representatives from the member companies. These groups are: TSG-A (Access Network Interfaces), TSG-C (cdma2000®), TSG-S (Services and Systems Aspects), and TSG-X (Core Networks). These TSGs meet about ten times a year, and their principal aim is to produce technical specifications and reports. The output of the work of TSGs is reported to the Project's Steering Committee.

For further information, each TSG has a section on the website of 3GPP2. The technical Specifications and Reports of 3GPP2 can be accessed by the public for free.

2.2.12 Broadcast Standardization

2.2.12.1 EBU

EBU (European Broadcast Union) informs its activities and studies in Ref. [25]. EBU is active in international TV and radio broadcasting, like coordination of communications related to the Eurovision Song Contest. The contents of the Eurovision Song Contest are delivered via satellite and fiber network providing transmission services to broadcasters in a global scale. One task of EBU is thus to provide transmission services with wanted quality, from standard definition up to a variety of high definition audio and video in broadcast, file transport or streaming formats.

Eurovision offers four main transmission types called One Stop Shop, Special Events, Unilateral and Space Segment. One Stop Shop combines facilities with transmissions, meaning that the broadcaster needs to make a single call or order to Eurovision, EBU taking care of the rest of the technical arrangements. Unilateral services are point to point transmissions on satellite or fiber, while space segments are segments of satellite time.

The EBU Partnership Program (EPP) aims to strengthen the cause of public service media (PSM) in Europe through tailored interventions to selected Members who need strategic consultancy, lobbying, training and political intervention.

The EBU Partnership was launched in 2009 as the Special Assistance Project. In June 2012, prior to the final signature of a Partnership agreement between the EBU and the European Commission, the project was renamed.

2.2.12.2 Satellite Systems Standardization

The standardization of satellite technologies and solutions is performed in various standardization bodies. There are also highly closed and proprietary solutions, as in the areas of military and space exploration. ESA and NASA are the most widely known representatives of applied space technology and investigation, and produce both highly proprietary solutions as well as globally standardized definitions.

One of the telecommunications related satellite activities is the standardization of the satellite component of IMT-Advanced systems which will play an important role, especially in providing multimedia broadcasting and multicasting services due to its inherent characteristics. Recently, new satellite radio interfaces are being developed which have high commonality between terrestrial radio interfaces.

2.2.12.3 Other Standardization Organizations

In addition to the actual standardization bodies which have the most important impact on the development of solutions that are interoperable internationally, there is also a wide variety of supporting work groups that influence the standardization and/or the practical deployment of the standards. These groups are often called forums. The following sections identify some of the most relevant forums in the telecommunications area.

2.3 Industry Forums

2.3.1 GSM Association

GSMA (GSM Association) represents the interests of mobile operators worldwide, consisting of nearly 800 full members (mobile operators) and more than 200 associate members (network and handset vendors, GRX carriers, roaming brokers etc.). The main difference between these membership categories is that only full members are allowed to vote. GSMA traces its history back to the original EU GSM declaration of 1982, but was formally created as the "GSM MoU Association" in 1995.

As a global industry trade association, the GSMA heavily consists of meetings and discussions between competitors. Therefore a strict antitrust policy is in place to prevent any kind of cartel behavior such as market sharing or price fixing, for example any type of discussion on pricing is prohibited in GSMA.

As is typical in the mobile industry, the requirements, architecture and technical details of roaming are handled in the "real standardization organizations" such as ETSI or 3GPP. Specifications coming out of those SDOs are then reused by groups such as GSMA to fill in items that are out of the scope of the purely technical standardization work, such as real-world practices, commercial models, common agreements, fraud issues, and so on. In the mobile area these and other items relevant for the commercial deployment of roaming are handled within GSMA. Though GSMA is not SDO, it has also been used as a vehicle to develop some completely new services. Recent examples include VoLTE, RCS and oneAPI which could have been standardized in 3GPP or perhaps in OMA.

Work performed in GSMA sometimes also includes profiling of 3GPP specifications – for example when, due to typical political standardization compromise, there are two (or more) options in the specification for handling some item, it is possible for GSMA to have another round of discussions and if possible agree on only one alternative which obviously makes it easier for feasible deployment in the typically rather challenging multioperator environment. That is the reason why it is advisable for operators also to take into account GSMA documentation, for example in the area of 3G or LTE roaming, in addition to the technical specifications originating from 3GPP.

The main work of GSMA is done in various projects and permanent working groups. For example IREG (Interworking & Roaming Expert Group) is responsible for the technical issues related to all inter-operator aspects, SG (Security Group) deals with security related matters of mobile world while BARG (Billing, Accounting and Roaming Group) has overall responsibility for supporting the interoperator business/wholesale charging framework. Working groups produce various PRDs (Permanent Reference Documents), such as AA.60 which is the template agreement for setting up interconnection, SG.20 documents voicemail security guidelines and IR.90 gives guidance for RCS specific interconnection topics.

A major role for GSMA is the public policy front; in practice this means lobbying such as influencing the EU commission not to regulate the inter-European roaming tariffs too much. There's also a specific development fund within GSMA, which means supporting the developing countries with activities such as mobile money, mLearning and green power. GSMA also produces major events such as the Mobile World Congress and Mobile Asia Expo.

Out of scope for GSMA are the non-3GPP technologies, so all the fixed networks and a number of mobile networks such as WiMAX or CDMA are not handled there. Exception to this rule are interoperator networks such as GRX and IPX which have been developed within GSMA to serve the purposes of mobile networks exchanging traffic between each other in roaming and interconnection scenarios.

The impact of GSMA has been clear, especially in the area of roaming, where the common guidelines of GSMA have been very beneficial in achieving true global interoperability on how to really implement roaming technically and commercially. Otherwise there's a concrete danger that each and every one of those nearly 800 operators have their own slightly different way how to implement roaming, both in technical and commercial sense. For example the common roaming agreement templates provided by GSMA have ensured that negotiations with a new roaming partner can be handled more or less the same way every time. The technical roaming database of GSMA called IR.21 allows easy way for the absolutely vital exchange of technical details, such as the IP address of MMSC required to implement MMS interworking, between the operators. These and a number of similar functions provided by GSMA have clearly proven the need for a forum such as GSMA helping the daily life of an operator.

Regardless of the name, GSM Association also is involved with other technologies in addition to GSM and its evolution like GPRS and EDGE. Further technologies are more specifically related to 3G/WCDMA, HSPA and LTE, and their evolution.

The main focus of GSMA is to represent the interests of mobile operators worldwide. GSMA unites nearly 800 of the world's mobile operators operating in about 200 countries. There is also a group of over 200 other companies in the complete mobile ecosystem of GSM Association, including handset makers, software companies, equipment providers, Internet companies, and media and entertainment organizations. One of the most known activities globally of GSM Association is to organize the Mobile World Congress (MWC) and Mobile Asia Expo.

The organization of GSM Association includes GSMA Board which provides a high-level direction and decision making. The board consists of chairman, Director General, and members of the board. Under the GSM Association, there are committees and groups of specialists. The members of GSMA steer and participate in the work of the GSMA through these specialist committees and groups.

The Strategy Committee, the Products and Services Management Committee and the Public Policy Committee of GSMA initiate projects. These initiations are approved via a process called toll gating. Once approved, the projects are reviewed on a regular basis in order to assure they are kept in line with the strategic objectives of GSMA.

The Regional Interest Groups of GSMA facilitate forums that are meant for discussions of the mobile industry, and to address issues specific to certain regions. There are Regional Internet Groups in Europe, Latin America, Asia, the Arab World, Africa and North America.

As for the European Union, GSMA is a representative of over 100 mobile network operators. The main focus of the secretariat of GSMA Europe is to coordinate the efforts of the GSMA to inform the EU-level policy and regulatory developments that impact the mobile industry. This work is done in practice via the Operator Expert Groups. It consists of public policy and technical experts with interests in Europe.

The Chief Regulatory Officers Group for Europe has an important role in the governance structure of GSMA, forming a link between the Operator Expert Groups and the Public Policy Committee of the GSMA Board [26].

In addition to the GSMA Europe, the roles of the other GSMA entities are the following:

- GSMA Latin America (GSMA LA) has four Operator Expert Working Groups for Billing and Roaming (BARG), Regulatory (REGU), Technical and Terminals (TECT), and Security and Fraud (SEGF) issues in the region. The main task of GSMA LA is to host two key plenary sessions per ear. The Chief Regulatory Officers Group for Latin America (CROG Latin America) guides the GSMAs public policy activities in the region and inputs the interests of the Latin American operators into the agendas of the GSMA Board, the Public Policy Committee (PPC) and the Global Chief Regulatory Officers Group (CROG).
- GSMA North America is steered by CTO Advisory Group, which coordinates activities of various technical working groups. These working groups include the Services Working Group, the Smart Card Group, the Terminal Working Group, the Fraud and Security group, the Interworking, Roaming Expert Group (IREG), the Billing, Accounting & Roaming Group (BARG) and the Standards & Wireless Alerts Task Force.
- GSMA Asia consists of several regional working groups aligned with the global GSMA working groups. GSMA Asia also runs several successful conferences and events covering roaming and regulatory issues.

- GSMA Africa is focused on regulatory, environmental, and roaming issues. Its secretariat oversees research on taxation, spectrum and other issues of particular relevance to the African mobile industry.
- GSMA Arab World promotes and facilitates the development of GSM-based services, seeks to enhance investments in network infrastructure and aims to drives innovation, growth and expand consumer choice. GSMA Arab World runs a website.

2.3.2 UMTS Forum

UMTS Forum was established in 1996, with the original aim facilitating discussions related to UMTS systems. There are members from operators, equipment manufacturers and regulators. Also the European Commission has taken an active part in the activities of UMTS Forum. In general, the UMTS Forum helps all key participants of the field in a highly dynamic environment to understand and profit from the opportunities of 3G/UMTS networks and their Long Term Evolution (LTE) [27].

The UMTS Forum is a member of, for example, ITU and participates actively in the work of all three ITU Sector Groups. UMTS Forum also contributes to the works of EC and CEPT, and discusses actively with other governmental, administrative, industry and technical bodies.

The UMTS Forum also contributes to the standardization work of mobile broadband of ETSI (as an observer) and the Third Generation Partnership Project (as a Market Representation Partner). Furthermore, UMTS Forum has dialogue with regulators related to the licensing and deployment of mobile broadband.

The UMTS Forum executes studies, reports and other publications of the telecommunication field. The main interest areas include markets trends, mobile broadband services and applications, key growth markets, spectrum and regulation, as well as technology and implementation. UMTS Forum is active at conferences, seminars and workshops at a global level, and it also informs media and analysts about the telecommunications area, as presented in Figure 2.4.

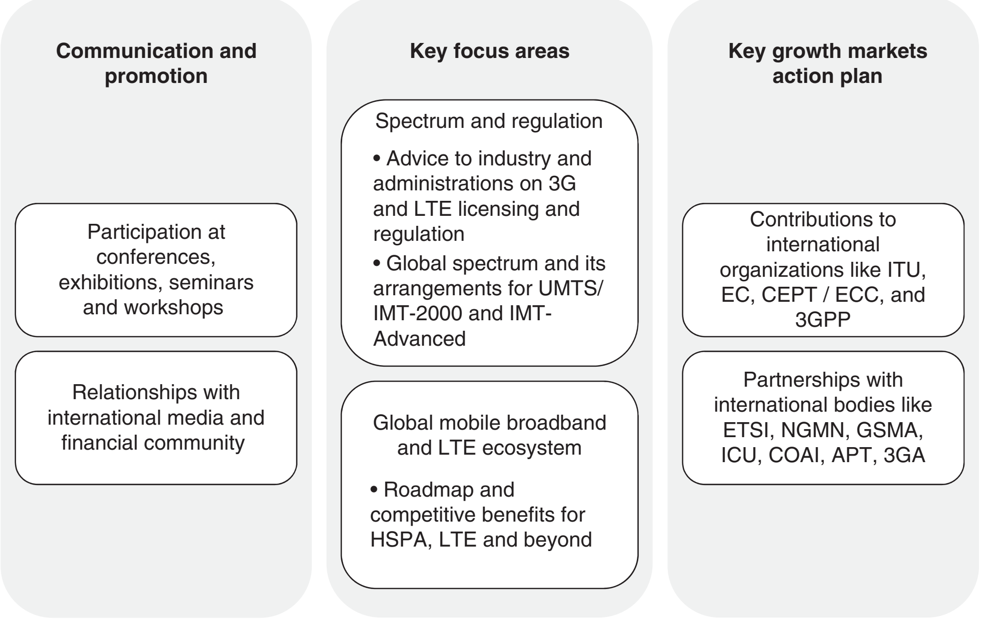

Figure 2.4 *The key areas of work of UMTS Forum as interpreted from source [27]. Data published by UMTS Forum.*

UMTS Forum also keeps track of the statistics of the telecommunications, although the information presented in its web pages is for information only. As an example, UMTS Forum has presented that the global WCDMA subscriptions, which refers to the whole set of WCDMA and HSPA Subscribers was 1159 347 217 in 2012, from which the global HSPA Subscribers was 892 117 434. Furthermore, according to the information, the number of WCDMA Networks Launched was 384, HSPA Networks 417, HSPA+ Networks 182, UMTS 900 Networks 51, and the number of LTE Networks Launched was 80.

2.3.3 WiMAX Forum

Information about the WiMAX Forum can be obtained from Ref. [28]. According to the presentation of WiMAX Forum, it is an industry-led, not-for-profit organization that certifies and promotes the compatibility and interoperability of broadband wireless products based upon IEEE Standard 802.16.

The main goal of WiMAX Forum is to accelerate the adoption, deployment and expansion of WiMAX technologies globally, and to facilitate roaming agreements. The aim of the WiMAX Forum is to promote and accelerate the introduction of cost-effective broadband wireless access services into the telecommunications market. Via the concept of WiMAX Forum Certified products, the equipment is assured to be interoperable with the support of broadband fixed, nomadic, portable and mobile services. As an important way of work, the WiMAX Forum interfaces actively with service providers and regulators to ensure that WiMAX Forum Certified systems meet customer and government requirements.

Currently, the WiMAX Forum has hundreds of members. The members represent operators, component vendors and equipment vendors.

The organization of the WiMAX Forum consists of officers and Board of Directors that are responsible for leading the broadband wireless access (BWA) market adoption of IEEE 802.16-based BWA systems. Means to execute this task include promotional activity, certification and interoperability testing. In addition, the Board and its officers also take care of agency oversight, goal setting, policy review and fundraising.

For the execution of the actual work, there are various working groups within WiMAX Forum. They identify critical focus areas in order to facilitate the introduction of WiMAX Forum Certified products to the telecommunications markets. The working groups are organized under the Technical Steering Committee. The concrete aim of these activities is to develop technical specifications for the WiMAX Forum Certified products.

The overlaying Technical Steering Committee (TSC) guides the non-Advisory Working Groups. It takes care of the high-level coordination of the development of WiMAX Forum technical specifications and certification procedures. Some of the more concrete tasks of TSC are to ensure the consistency of Working Group activities and results, to make sure that the technical planning, specification, and certification takes into account sufficient broadly the representatives of the Principal Membership, and to promote wide acceptance of roadmaps and decisions by the WiMAX Forum Membership.

2.3.4 BMCO Forum and Open IPTV Forum

OIPF (Open IPTV Forum) and bmcoforum (Broadcast Mobile Convergence Forum) have joined their forces and merged the previously separate activities of bmcoforum into the OIPF [29]. This arrangement is planned to form a stronger entity with broader representation and influence across the fixed and mobile broadband industries. The merged activities optimize the resources of these previously separate, leading organizations, and enhance the weight in the standardization of mass market services and devices.

This new setup and the resulting multiscreen approaches that include TVs, PCs, mobile devices and other screens is a logical step for TV and media consumption. One of the aims of this concept is to provide increasing level of interactivity and flexibility as the broadcast is moving towards all-IP era. Different types of content services will thus be delivered into multiscreen environments. Characteristic for this transition is to provide flexibility both in the home and in the moving environment. It should be noted that mobile devices are not limited to the traditional TV contents delivery but also as service control points, allowing selection and control of personal content services. This extension of IPTV to mobile usage and hybrid networks facilitates the development of the broadcast markets towards multiscreen media consumption, and the joining of resources of OIPF and bmcoforum is a logical step.

The bmcoforum is an international nonprofit organization designed to foster the mass market for mobile media consumption. Therefore bmcoforum optimizes the technology mix and business models and brings together players from all parts of the media delivery value chain [30].

The OIPF, on the other hand, is open to diverse participants from the communications technology areas and entertainment industries. It thus brings together network operators, content providers, service providers, consumer electronics manufacturers and home and network infrastructure providers. The members of the Open IPTV Forum are working together on the development of open specifications with the goal of combining the expertise of all involved in helping streamline and accelerate deployments of IPTV technologies. Their aim is to make the next generation of IPTV a mass market service and to maximize the benefits of IPTV for consumers as well as the industry [31].

The resulting Open IPTV Forum enables and accelerates creation of a mass market for IPTV by defining and publishing free-of-charge, standards-based specifications for end–end IPTV services of the future. End-to-end specifications are essential to an effective ecosystem delivering an easy "plug and play" interoperability experience for the end consumer. The Open IPTV Forum is an essential organization independent from the technology behind the industry. The forum is open to participation from the communications and entertainment industries. As a concrete result in the activity, device certification and Interoperability Testing program will trigger the OIPF logo appearing on numerous services and devices. This logo indicates the usability of the services across multiple mobile screens.

2.3.5 Global Mobile Suppliers Association

The Global mobile Suppliers Association (GSA) represents mobile suppliers in international level, including technology areas such as infrastructure, semiconductors, devices, services and applications development, and support services [32]. The website presented in Ref. [32] is meant as a working area for the industry professionals and organizations that represent about 200 countries. It is thus a centralized source of information for the industry, and it includes useful information about surveys, market, technology and subscriptions updates, technology descriptions, information about GSM, EDGE, WCDMA, HSPA, and HSPA+ network deployments, devices availability, applications and services, operator case studies, success stories, and relevant developments in the evolution towards LTE/SAE with detailed information about LTE network operator commitments.

One of the important ways of work of GSA is to provide advice to governments, administrations and policy-makers about conditions for market development. GSA also gives active briefings to media and analysts. There is both publicly available information in Ref. [32], as well as restricted information for the participants.

GSA is a Market Representation Partner in 3GPP and cooperates with other key organizations including COAI, ETSI, GSM Association, ICU and ITU. More information can be obtained from Internet sources like the RSS news feed, Twitter [33], and dotMobi.

2.3.6 CDMA Development Group

The aim of CDMA Development Group (CDG) is to lead the rapid evolution and deployment of CDMA2000 and complementary 4G systems, based on open standards and encompassing all core architectures, to meet the needs of markets around the world [34]. The CDG and its members work together in order to accelerate the definition of requirements for new features, services and applications, to promote industry and public awareness of CDMA2000 and complementary 4G capabilities and developments through marketing and public

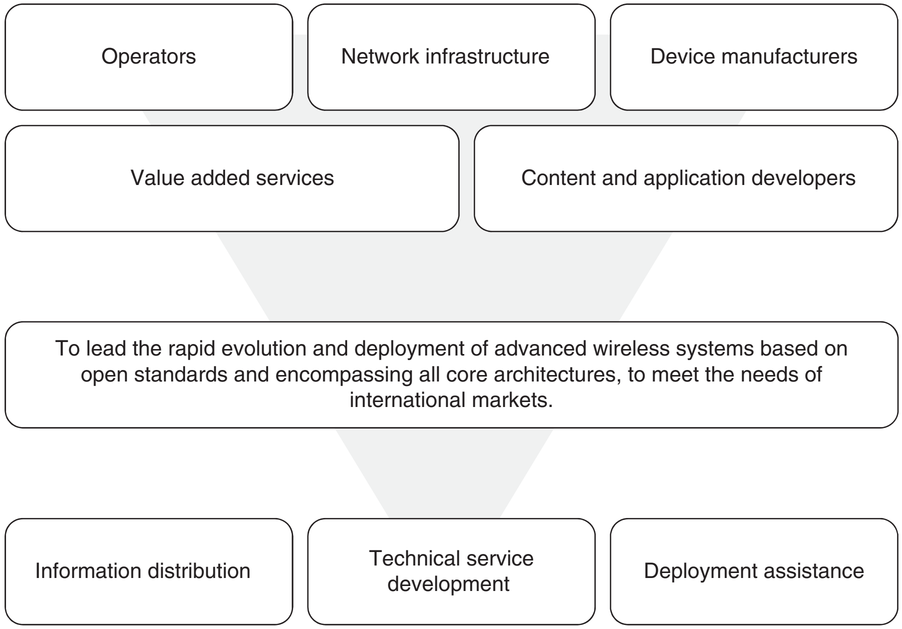

Figure 2.5 *The operational environment of CDG with the high-level tasks [36]. Data published by CDG.*

relations activities, to foster collaboration and the development of consensus among carriers on critical issues to provide direction and leadership for the industry, to define the evolution path for current and next-generation systems, to expand the selection and availability of affordable devices, and to enable and expand interstandard global roaming. CDG also establishes strategic relationships with government ministries, regulatory bodies, and worldwide standards and industry organizations to promote cooperation and consensus on issues facing the wireless community, and it serves as the worldwide resource for CDMA2000 and complementary 4G information. As a result of these activities, CDG aims to minimize the time-to-market of new CDMA2000 and complementary 4G products and services, to enable global compatibility and interoperability among CDMA and OFDMA systems worldwide, and to promote the business objectives of its members.

The CDG was founded in 1993. It is an international consortium of companies who work together to lead the growth and evolution of advanced wireless telecommunication systems. The CDG is comprised of service providers, infrastructure manufacturers, device suppliers, test equipment vendors, application developers and content providers. Its members jointly define the technical requirements for the evolution of CDMA2000 and complementary 4G systems and interoperability with other emerging wireless technologies to expand the availability of wireless products and services to consumers and businesses worldwide.

The primary activities of the CDG include strategic, technical and advocacy efforts that promote the growth and advancement of wireless technologies based on globally accepted open standards. Currently, there are more than 500 individuals working within various CDG subcommittees to lead these efforts. Figure 2.5 summarizes the high-level tasks and work environment of CDG.

2.3.7 Other Standardization Bodies

In addition to the key organizations described in previous sections of this chapter, there is a variety of regional and national standardization bodies and supportive expert groups in the telecommunication area. Some examples of these bodies are:

IEEE (Institute of Electrical and Electronics Engineers) is an entity which aims to promote the education and knowledge of electronics, electrical and information technology. IEEE also participates in the creation of standards, one of the best known set being the IEEE 802 definitions of fixed and wireless local area network variants.
ISO (International Standardization Organization) is a standardization organization that has produced, among others, the OSI model (Open System Interconnection).
IEC (International Electrotechnical Commission) is a standardization body for the electricity area.

2.4 Other Entities

2.4.1 UNDP

In addition to the actual standardization bodies and supporting entities, there is also an important set of other organizations that either directly or indirectly may influence in the development of the telecommunications. One example of these organizations is the United Nations Development Program (UNDP) [35]. This is the UN's global development network, advocating for change and connecting countries to knowledge, experience and resources to help people build a better life. UNDP is active in 177 countries and territories, working with governments and people on their own solutions to global and national development challenges. As they develop local capacity, they draw on the people of UNDP and our wide range of partners to bring about results.

As for the telecommunications area, UNDP may have a coordination role, for example, for the selection of new telecommunications infrastructure via bid processes and evaluation. As an example of various related tasks, there have been renovations of public telecommunications infrastructure in Honduras, 2003, where several telecommunications areas were reviewed and new equipment and transmission was purchased. The evaluation work was done in cooperation with local operators, regulator and independent consultants that were selected via a separate bid process in order to offer additional resources for the evaluation, and to offer special technical knowledge in their areas. Figure 2.6 shows the equipment under renewal.

Figure 2.6 An example of the work of UNDP in the modernization of the public telephony infrastructure and equipment in Honduras. Old cabinets are investigated in Tegucigalpa, Honduras.

2.4.2 IADB

The Inter-American Development Bank (IADB) works also in the evaluation of the needs of the telecommunications. As an example, there have been in-depth investigations of the renovation of the telecommunications infrastructure in Latin America region, in international cooperation coordinated by IADB [36]. To present this example, Development Bank and the local government organizations carried out a feasibility study in Nicaragua, which was related to the technological needs in order to create the local telecenters and Internet cafés to several rural and marginal areas of Nicaragua. Also Finland is participating in this project via World Bank representatives. In this example, users are concentrated in the biggest cities of the county, whereas rural and marginal areas lack telecommunications infrastructure and even basic infrastructure of electrical lines and water delivery systems. There is, though, a great need to enter the Internet world in many areas outside high populated centers, as field studies have shown.

The public use of Internet is more and more popular in the Managua area. This can be seen by fast revision of the city area. There are several Internet cafés in the shopping centers and in other public places. There are also increasing number of hotels adapting to Internet technology, with Internet access provided for the clients in separate communications area and with the possibility of printing documents and send faxes. The overall situation in Managua thus gives an idea of a relatively developed area as telecommunications services are considered. The scene changes, though, drastically when visiting other parts of Nicaragua. If there were more means to use the Internet services, the tourism and local business would most probably develop. For tourists, there are excellent places for spending the vacations in Nicaragua, but in many cases, there is not yet adequate infrastructure of telecommunications, power lines or water tubes constructed.

Depending on the area, the technological alternatives for the connectivity of the Internet cafés might be very limited. There are places in Nicaragua without any telecommunications infrastructure, which gives challenges in offering Internet services. In these cases, the only option might be satellite connection, which results in higher operating costs and thus more expensive prices for customers compared with normal solutions, for example, via cable modem. Nevertheless, data rates via satellite can normally reach up to 128–256 kb/s, which is technologically very interesting aspect. Also mobile communications systems have been deployed in a fast schedule in Nicaragua, which can be one of the bases for the data services. Another alternative might be a combined solution, consisting of the central satellite station and VHF repeater station, and which delivers the traffic within about 70 km radius from the central. This option could be valid in areas that consist of several users spread around the centralized station. As an example, there are areas without telecommunications lines in Nicaragua that are quite difficult to access by roads. In these areas, local businesses, as well as private use, might benefit considerably from the connectivity to Internet. Figure 2.7 shows one of the area types under such investigation.

This example shows that it is important also to take into account the poorest countries when developing the information society. It is also essential that countries without good telecommunications infrastructure would have a chance to access the information services among others. We can clearly see that the provision of Internet services in rural and marginal areas will help to develop local business life, and increases the general quality of life of local people.

2.5 Frequency Regulation

2.5.1 WRC

ITU regularly organizes the World Radiocommunication Conference (WRC). The aim of the conference is to review and revise international radio regulations. WRC also handles the international use of the radio frequency spectrum as well as of satellite orbits. Typically, the conference is organized every three to four

Figure 2.7 *An example of telecommunication consultancy of IADB in Nicaragua. This specific project was carried out in rural and marginal areas of Nicaragua 2003, in order to evaluate the telecommunication infrastructure and to propose development plan for increasing the number of telecenters in the country.*

years. It should be noted that earlier, before 1993, the conference was called the World Administrative Radio Conference (WARC).

The latest WRC was held in 2012 in Geneva along with the Radio Assembly and the first Conference Preparatory Meeting for WRC 2015. There were over 3000 regulators present, as well as ministers and members of the wireless industry. The WRC is thus a major event for designing global radio frequency principles, the main focus being in the key decisions that are impacting the satellite, defence, avionics, shipping, broadcasting and mobile industry.

Some of the key decisions of the WRC 2012 have been:

- Approval of studies in 2012–15 by a joint task group in order to find a new spectrum for mobile broadband and mobile telephony. This decision allows the next WRC 2015 to plan the allocation of additional bands for mobile broadband and mobile telephony.
- Approval of studies in 2012–15 by Study Group 5 (WP5A) related to broadband public safety (PPDR). This decision allows the next WRC 2015 to review the needs of broadband public safety.
- Allocation of DD-2 spectrum in 694-790 MHz to mobile service in Europe, Middle East and Africa (ITU Region 1).

2.6 National Regulators

In order to operate a radio communications network, most countries require the respective operator to obtain a license via the national telecommunications regulators [37]. In general, as the radio frequency spectrum is a limited resource, it needs to be used in an optimal way, so the regulation is badly needed for the selection of frequency bands per operator as well as for the selection of technologies utilized in those frequencies. In practice, the operation of a commercial mobile radio network always therefore requires a respective operating license.

In case of private radio networks like Professional Mobile Radio (PMR) that is meant, for example, for TETRA, the licensing requirements depends on the policies of each country. Typically, for these types of networks that are related to the national security, the private radio system may only require a frequency license.

The usage of radio frequencies is normally thus controlled by the national radiocommunications regulator. Another solution is to control it via a separate frequency management entity to which the regulator delegates responsibility for frequency band management.

The utilization of frequency bands thus requires an application. If the requirements are complied with, the national regulator grants the right to use the respective radio frequencies. This is done for a specified purpose and within a defined geographical area.

In many cases, the radio frequency license may have a cost impact. The expense and duration of the license are country specific, so even in the case of the same technology, some country may apply the rules of so-called "beauty contents," that is, the technically best companies may be granted to operate the technology, whereas some countries are in favor of bid process that may result in very high costs.

The role of regulators contains many aspects, from the national regulation of radio frequencies (coordination of telecommunications system licenses, setting the national limits for the interferences) as well as active participation in international planning of global frequency utilization. One of the main events for the joint planning of the RF utilization principles is the World Radio Frequency conference organized by ITU-R. The event takes place about once per two or three years and dictates the frequency division in the ITU regions between neighboring countries and at a global level. As previous mobile communications systems are eventually ramped down in the commercial markets as well as in the closed environment, and new solutions are appearing constantly, the work is to align the allocations of the frequencies between old and future systems.

2.7 Guideline for Finding and Interpreting Standards

The basic principle when seeking for the source of information via standards is to investigate to which technology area the topic is related to. As an example, for mobile communications in Europe, the most logical place to initiate the study is 3GPP which standardizes the GSM, UMTS and their evolution, LTE and the developed version of it. For American variants, 3GPP2 is the main body for standardization. Both 3GPP and 3GPP2 present the standards in their Internet pages.

Another example of the highest level requirements for mobile communications is ITU, which contains three areas: ITU-R for radio, ITU-T for fixed networks, and ITU-D for development. The Internet pages of ITU can be used for the study of certain standards whilst other standards must be obtained by purchasing them.

References

1. ITU: Overview. http://www.itu.int/en/about/Pages/overview.aspx (last accessed 19 July 2014).
2. ITU Radiocommunication Sector. http://www.itu.int/ITU-R/index.asp?category=information&rlink=imt-advanced-rec&lang=en (last accessed 19 July 2014).
3. ITU-T Recommendations. http://www.itu.int/rec/T-REC/e (last accessed 19 July 2014).
4. ITU-R Recommendations. http://www.itu.int/pub/R-REC (last accessed 19 July 2014).
5. ETSI. http://www.etsi.org (last accessed 23 August 2014).
6. IEEE. Publications & Standards. http://www.ieee.org/publications_standards/index.html (last accessed 19 July 2014).
7. IEEE: About the IEEE Get Program. http://standards.ieee.org/about/get/index.html (last accessed 19 July 2014).

8. Official Internet Protocol Standards. http://www.rfc-editor.org/rfcxx00.html (last accessed 19 July 2014).
9. RFC Search. http://www.rfc-editor.org/rfcsearch.html (last accessed 19 July 2014).
10. IETF RFC 1677. *The Tao of IETF – A Novice's Guide to the Internet Engineering Task Force*.
11. IETF: Request for Comments (RFC). http://www.ietf.org/rfc.html (last accessed 19 July 2014).
12. IETF RFC 2026. *The Internet Standards Process – Revision 3*.
13. IETF RFC 5378. *Rights Contributors Provide to the IETF Trust*.
14. IETF RFC 3979. *Intellectual Property Rights in IETF Technology*.
15. European Conference of Postal and Telecommunications Administrations. http://www.cept.org (last accessed 19 July 2014).
16. Electronic Communications Committee. http://www.cept.org/ecc (last accessed 19 July 2014).
17. European Communications Office. http://www.cept.org/eco (last accessed 19 July 2014).
18. Committee for ITU Policy. http://www.cept.org/com-itu (last accessed 19 July 2014).
19. European Committee for Postal Regulation. http://www.cept.org/cerp (last accessed 19 July 2014).
20. ANSI: Overview. http://www.ansi.org/about_ansi/overview/overview.aspx?menuid=1 (last accessed 19 July 2014).
21. *ANSI Essential Requirements: Due Process Requirements for American National Standards*. The American National Standards Institute (ANSI). Edition: January 2012. 26 pp.
22. ANSI Online Document Library. http://www.ansi.org/library/overview.aspx?menuid=11 (last accessed 19 July 2014).
23. TTC Standard Summary. http://www.ttc.or.jp/cgi/summarydb/index.html (last accessed 19 July 2014).
24. 3GPP. www.3gpp.org (last accessed 19 July 2014).
25. EBU. http://www3.ebu.ch/cms/en/sites/ebu/home.html (last accessed 19 July 2014).
26. GSMA: Committees and Groups. http://www.gsma.com/aboutus/leadership/committees-and-groups/ (last accessed 19 July 2014).
27. UMTS Forum. http://www.umts-forum.org/content/view/2002/125/ (last accessed 19 July 2014).
28. WiMAX Forum. http://www.wimaxforum.org/ (last accessed 19 July 2014).
29. OIPV. http://www.oipf.tv (last accessed 19 July 2014).
30. BMCO. http://www.oipf.tv/bmco (last accessed 23 August 2014).
31. IPTV. http://www.oipf.tv (last accessed 19 July 2014).
32. GSA. http://www.gsacom.com/ (last accessed 19 July 2014).
33. GSA on Twitter. www.twitter.com/gsacom (last accessed 19 July 2014).
34. CDG. http://www.cdg.org/ (last accessed 19 July 2014).
35. UDP. http://www.undp.org/content/undp/en/home.html (last accessed 19 July 2014).
36. eFinland: Information Technology in Nicaragua – Finland Offers a Helping Hand. http://e.finland.fi/netcomm/news/showarticled99c.html?intNWSAID=15848 (last accessed 19 July 2014).
37. TETRA. http://www.tetramou.com/about/page/11999 (last accessed 19 July 2014).

3

Telecommunications Principles

Jyrki T. J. Penttinen

3.1 Introduction

The telecommunications area is significantly full of terminology. This chapter aims to set up the scene by clarifying the essential contents of the terms, and gives ideas on how the terminology differs based on the standards, operator and equipment manufacturer environments.

This chapter also brings views on the evolution of the systems and applications, and identifies the needs that drive the evolution of telecommunications further. In addition, this chapter gives essential guidelines about the frequency spectrum allocations by identifying the ITU regions and principles, and showing also the regional aspects. Frequency regulation in general is at the end the task of each country, so this chapter clarifies the principles of the global and local levels. As one important part of this environment, the interference aspects and shielding are explained.

To complete this chapter, near-future and longer term systems are identified at a high level. For modern systems, the spectral efficiency and optimal utilization of frequencies is one of the most important tasks of operators of wireless communications, so this aspect is covered. Finally, physical aspects of the systems are explained, including the principles of radio interface and radio links, electrical wires and optical fibers. This chapter thus creates the technical, high-level basis for the forthcoming chapters.

3.2 Terminology and Planning Principles

In the telecommunications technology area, it is essential to understand and to be understood as uniformly as possible. The need arises from the substantially international character of the field, as basically all modern telecommunications systems need some sort of interoperability in domestic and international operations. One good example of this is the cellular network user visiting different foreign countries by utilizing the same handheld device.

The Telecommunications Handbook: Engineering Guidelines for Fixed, Mobile and Satellite Systems, First Edition.
Edited by Jyrki T. J. Penttinen.

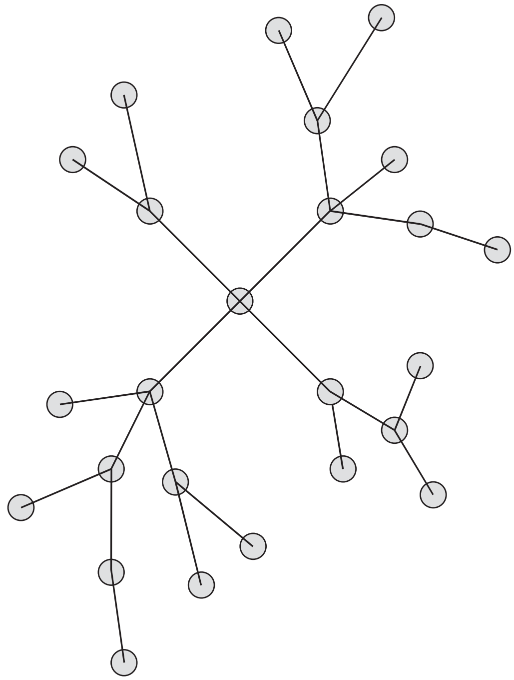

Figure 3.1 *An example of the interconnection of the nodes in telecommunication networks.*

The terminology is challenging to unify, though. There are international standards for the systems that would be quite compatible for the terminology, but then there are different companies like telecom operators, device and network equipment manufacturers, service providers and so on, and each one typically has partially created their own terminology to describe the same topics that are called by other names elsewhere.

The importance of a uniform terminology arises, for example, from the need to commonly agree quality indicators of networks. These key performance indicators may be called by tens if not hundreds of names, and they typically greatly overlap, causing misunderstandings.

The functioning of telecommunications networks is based on physical connections and data flow transferred over various interfaces by utilizing protocols. The physical connections can be based on fixed and wireless methods, each one having special characteristics that have an effect on the reliability and performance of the end-to-end connection – as well as on the terminology. Figure 3.1 shows a high-level example of the connection hierarchy.

3.2.1 Decibel

One of the most important base units in telecommunications is decibel (dB). It is suitable for measuring the relative performance of countless phenomena, for example, for the strength of the audio. A noise level that is hardly possible to hear, like the movement of the leaves of trees, may cause 10−20 dB level, whereas typical office work may generate around 50 dB level. 85 dB already requires protectors for the human ears, and close to the jet the level already exceeds 120 dB, which is the level for potentially harming the human ears.

The decibel is a commonly utilized standard to measure the basic performance of the electronics, like electrical current, voltage, and power levels, which are essential building blocks in the telecommunications. The voltage utilizes 10-base logarithms:

$$A_U(dB) = 20\log_{10}\left(\frac{U}{U_{ref}}\right). \tag{3.1}$$

In this equation, U is the voltage level (volts), and U_{ref} is the reference voltage. As can be seen from the equation, the dB only informs the relative difference of two values. Also the currency in ampers (I and I_{ref}) can be calculated as follows:

$$A_I\,(dB) = 20\log_{10}\left(\frac{I}{I_{ref}}\right). \tag{3.2}$$

It should be noted that both voltage and currency utilizes value 20 in the formula. Instead, the electrical power level utilizes the value of 10 as can be seen in the following formula:

$$P\,(dB) = 10\log_{10}\left(\frac{P}{P_{ref}}\right). \tag{3.3}$$

This difference between the formulas can be seen from the dependency of P, U and I as they behave in the following way:

$$P = UI = RI^2 = \frac{U^2}{R}. \tag{3.4}$$

The squared I and U result in the additional term 2 to the voltage and currency formulas, and thus the coefficient of these formulas is $2 \cdot 10 = 20$.

The decibel offers a practical way to calculate the signal levels in telecommunications simply by adding and extracting the values without complicated conversions. This applies, for example, to the attenuations and gains of different affecting elements in the propagation path. As an example, the link budget calculations are straightforward as the transmitted or received power level is doubled when the dB value is grown by 3, and the power level is put to half when 3 dB is extracted from the original value.

The reference value can be selected freely. It is very typical to utilize, for example, a reference level of 1 mW in the cellular network measurements and dimensioning. When comparing the power level with one mW, the unit for the result is dBm. This can be utilized for both transmitting and received power levels in following way, when the measured power level is informed in watts:

$$P\,(dBm) = 10\log_{10}\left(\frac{P\,(W)}{10^{-3}}\right). \tag{3.5}$$

By applying simple mathematical rules for logarithms, the reversed way produces the Watt value for power, when the measured power value is obtained in dBm:

$$P\,(W) = 10^{\frac{P(dBm)}{10}}. \tag{3.6}$$

The dBm value is widely utilized in the field measurement equipment, so it is possible to compare easily different performance results of the same type of networks from different parts of the world. Table 3.1 presents an example of the GSM 900 system, with power levels expressed in dBm and W.

An equally important and widely utilized unit is dBμV (decibels compared to microvolt) for field strength. Furthermore, it can be informed in dBμV/m (decibels compared to one microvolt in one meter area). This is a common way to express the useful and interfering field strength, for example, in wireless broadcast networks like TV and radio systems. The base for this value is the received voltage level dBμV, which depends on

Table 3.1 Examples of the power levels in W and dBm

Power level (mW)	Power level (dBm)	Power level (mW)	Power level (dBm)
1995	33	31.6	15
1000	30	15.8	12
501	27	7.94	9
251	24	3.98	6
126	21	2.00	3
63.1	18	1.00	0

impedance value. As an example, 0 dBm corresponds to a value of 108.7 dBμV when 75 ohm load is utilized. The same example would result in 107.0 dBμV with the load of 50 ohm.

3.2.2 Erlang

In telecommunications networks, the capacity measurements and dimensioning utilizes typically units in Erlang (Erl). It indicates the average offered capacity of the network, when the network is dimensioned by accepting a certain average blocking probability. In practice, even the basic form of the Erlang can be utilized as a base for dimensioning the voice service in both fixed and wireless circuit switched networks.

In practice, in circuit switched telephony, Erlang refers to the continuous use of single voice connection. In practice, it is used to describe the total traffic volume of one hour. As an example, let us assume a group of voice call users that establishes 50 circuit switched calls during a time period of one complete hour. Further assuming that on average, the length of each call is 5 minutes, the Erlang value for this case can be obtained in the following way:

- Total minutes of traffic during one hour period is the number of calls multiplied by the average duration of each call, that is, 50 calls per 5 minutes, yielding a total of 250 minutes.
- The amount of traffic hours during the one hour period is total minutes per 60, yielding 250/60 = 4.17.
- This result indicates the traffic figure in Erlangs, that is, 4.17 Erl.

The reason for using Erlang is to evaluate the traffic load in the telecommunications networks. The planning of the offered capacity is done based on the expected utilization of the network.

The principle of Elrang is suitable for circuit switched telephony, which refers to the fixed amount of available resources. Nevertheless, for packet switched networks and services, the Erlang as such is not suitable. The offered capacity for packet data can be estimated, for example, by mapping the circuit switched capacity for the time slots or channels to the data flow in bits per second. There are various ways to do this, but the basic idea is to investigate the average data utilization of users by observing the throughput (indented and successful, both being affected under loaded network conditions). The challenge in this approach is that the traffic of packet data networks is bursty, with variable data rates, and thus the exact mapping is not straightforward.

3.2.2.1 Erlang B

For circuit switched systems, the Erlang formula can be presented as follows:

$$A\,(Erl) = \frac{tn}{60}. \tag{3.7}$$

In this formula, t is the length of the average call in minutes, and n is the number of the calls over one hour time period. In this way, assuming that there is only one channel available in the system, and a single user occupies the channel for the whole hour, the traffic on this single channel is in fact 1 Erl. Equally, if there are two separate, consecutive calls of half an hour each, the result is again 1 Erl.

In the real life network, this is a highly improbable situation on average, except during momentary utilization peaks. Otherwise, the average blocking probability would be 100%, meaning that the network would be totally congested at all times. A much more typical and realistic value could be, for example, 150 mErl (milli Erlangs), which effectively means an average occupation of the line during $0{,}150 \cdot 60 = 9$ minutes over a period of one hour.

The Erlang can be utilized for informing the average utilization of a single line or channel, a group of lines or channels, or even a total network whilst the total number of lines or channels is known. It is also possible to create a ratio for the Erlang, that is, by comparing the value over the number of users of the system. This indicates the traffic in mErl.

The assumption for the basic Erlang system is a group of channels that the new user can access, or cannot access in case all of them are already occupied. In other words, there is no queuing assumed in this form of Erlang system, which is called Erlang B. This equation indicates the blocking probability for the new calls as well as for the time, that is, the probability that all the channels are occupied at given moment:

$$B(N) = \frac{\dfrac{A^N}{N!}}{\displaystyle\sum_{n=0}^{N} \frac{A^n}{n!}}. \tag{3.8}$$

This equation has N, which indicates the amount of total channels of the system that are possible to be accessed, and A is the product of the average rate of the calls and the average time the calls occupy the channel. In other words, this is the offered load. This equation can be expressed also in recursive form which is more suitable for computational calculations:

$$\begin{aligned} B(0) &= 1 \\ B(N) &= \frac{AB(N-1)}{N + AB(N-1)}. \end{aligned} \tag{3.9}$$

It is common that the cellular telecommunications networks are dimensioned according to the blocking probability of 2% during the heaviest utilization of the network, that is, during the peak-hour.

As the above shown equation gives the offered load, the effect of blocking should be taken still into account in order to obtain the actual delivered traffic, that is, the traffic that can be served after the rest is blocked. The delivered traffic is obtained in the following way:

$$\bar{x} = A(1 - B). \tag{3.10}$$

Equally, the blocked traffic can be obtained by the equation:

$$m = A - \bar{x} = AB. \tag{3.11}$$

It should be noted that there are also other variants for Erlang formula. The one that takes into account the queuing during a certain maximum wait period until the call attempt is failing is called Erlang C.

Table 3.2 shows Erlang values for a typical dimension criteria for the offered load, and Table 3.3 shows the respective delivered traffic.

As an example of the utilization of the tables above, the offered load for a single TRX (Transceiver Unit) of GSM can be calculated. Assuming that a Full Rate codec is utilized in the TRX, we have a total of 8 timeslots (TSL) in use. For the single TRX case, one timeslot is used for the signaling, which leaves a total

Table 3.2 *Offered load A in Erlangs*

	Blocking probability B (%)							Blocking probability B (%)					
N	0.5	1.0	1.5	2.0	2.5	3.0	N	0.5	1.0	1.5	2.0	2.5	3.0
1	0.005	0.010	0.015	0.020	0.026	0.031	26	15.79	16.96	17.75	18.38	18.92	19.39
2	0.105	0.156	0.190	0.224	0.254	0.282	27	16.60	17.80	18.62	19.26	19.82	20.31
3	0.349	0.456	0.535	0.602	0.661	0.715	28	17.41	18.65	19.48	20.15	20.72	21.22
4	0.701	0.869	0.992	1.09	1.18	1.26	29	18.22	19.49	20.35	21.04	21.62	22.14
5	1.13	1.36	1.52	1.66	1.77	1.88	30	19.04	20.34	21.23	21.93	22.53	23.06
6	1.62	1.91	2.12	2.28	2.42	2.54	31	19.85	21.19	22.10	22.83	23.44	23.99
7	2.16	2.50	2.74	2.94	3.10	3.25	32	20.68	22.05	22.98	23.72	24.36	24.91
8	2.73	3.13	3.40	3.63	3.82	3.99	33	21.51	22.91	23.87	24.63	25.27	25.84
9	3.34	3.78	4.09	4.34	4.56	4.75	34	22.33	23.77	24.75	25.53	26.19	26.78
10	3.96	4.46	4.81	5.08	5.32	5.53	35	23.17	24.64	25.64	26.44	27.11	27.71
11	4.61	5.16	5.54	5.84	6.10	6.33	36	24.01	25.51	26.53	27.34	28.03	28.65
12	5.28	5.88	6.29	6.62	6.89	7.14	37	24.85	26.38	27.42	28.25	28.96	29.59
13	5.96	6.61	7.05	7.40	7.70	7.97	38	25.69	27.25	28.32	29.17	29.89	30.53
14	6.66	7.35	7.82	8.20	8.52	8.80	39	26.53	28.13	29.22	30.08	30.82	31.47
15	7.38	8.11	8.61	9.01	9.35	9.65	40	27.38	29.01	30.12	31.00	31.75	32.41
16	8.10	8.88	9.41	9.83	10.19	10.51	41	28.23	29.89	31.02	31.92	32.68	33.36
17	8.83	9.65	10.21	10.66	11.03	11.37	42	29.09	30.77	31.92	32.84	33.61	34.30
18	9.58	10.43	11.02	11.49	11.89	12.24	43	29.94	31.66	32.83	33.76	34.55	35.25
19	10.33	11.23	11.84	12.33	12.75	13.12	44	30.80	32.54	33.73	34.68	35.49	36.20
20	11.09	12.03	12.67	13.18	13.61	14.00	45	31.66	33.43	34.66	35.61	36.43	37.15
21	11.86	12.84	13.51	14.04	14.49	14.89	46	32.52	34.32	35.55	36.53	37.37	38.11
22	12.63	13.65	14.35	14.90	15.36	15.78	47	33.38	35.21	36.47	37.46	38.31	38.28
23	13.42	14.47	15.19	15.76	16.25	16.68	48	34.25	36.12	37.38	38.39	39.25	40.02
24	14.20	15.29	16.04	16.63	17.13	17.58	49	35.11	37.00	38.30	39.32	40.20	40.97
25	15.00	16.12	16.89	17.50	18.03	18.48	50	35.98	37.90	39.21	40.26	41.14	41.93

Table 3.3 *Delivered traffic in Erlangs*

	Blocking probability B (%)							Blocking probability B (%)					
N	0.5	1.0	1.5	2.0	2.5	3.0	N	0.5	1.0	1.5	2.0	2.5	3.0
1	0.005	0.010	0.015	0.020	0.025	0.030	26	15.72	16.79	17.49	18.02	18.45	18.81
2	0.105	0.154	0.188	0.219	0.247	0.273	27	16.52	17.62	18.34	18.88	19.32	19.70
3	0.347	0.451	0.527	0.590	0.645	0.694	28	17.32	18.46	19.19	19.75	20.20	20.58
4	0.698	0.861	0.977	1.07	1.15	1.22	29	18.13	19.29	20.05	20.62	21.08	21.48
5	1.13	1.35	1.50	1.62	1.73	1.82	30	18.94	20.13	20.91	21.49	21.97	22.37
6	1.61	1.89	2.09	2.23	2.36	2.47	31	19.75	20.98	21.77	22.37	22.86	23.27
7	2.15	2.48	2.70	2.88	3.02	3.15	32	20.57	21.83	22.64	23.25	23.75	24.17
8	2.72	3.10	3.35	3.56	3.72	3.87	33	21.40	22.68	23.51	24.13	24.64	25.07
9	3.32	3.74	4.03	4.26	4.44	4.61	34	22.22	23.53	24.38	25.02	25.54	25.97
10	3.94	4.42	4.74	4.98	5.19	5.36	35	23.05	24.39	25.25	25.91	26.43	26.88
11	4.59	5.11	5.46	5.72	5.95	6.14	36	23.89	25.25	26.13	26.80	27.33	27.79
12	5.25	5.82	6.19	6.48	6.72	6.93	37	24.72	26.11	27.01	27.69	28.24	28.70
13	5.93	6.54	6.94	7.25	7.51	7.73	38	25.56	26.98	27.89	28.58	29.14	29.61
14	6.63	7.28	7.71	8.04	8.31	8.54	39	26.40	27.85	28.78	29.48	30.05	30.52
15	7.34	8.03	8.48	8.83	9.12	9.36	40	27.24	28.72	29.67	30.38	30.95	31.44
16	8.06	8.79	9.26	9.63	9.93	10.19	41	28.09	29.59	30.56	31.28	31.86	32.36
17	8.79	9.56	10.06	10.44	10.76	11.03	42	28.94	30.46	31.44	32.18	32.77	33.28
18	9.53	10.33	10.86	11.26	11.59	11.87	43	29.79	31.34	32.33	33.08	33.69	34.20
19	10.28	11.12	11.67	12.09	12.43	12.72	44	30.64	32.22	33.23	33.99	34.60	35.12
20	11.04	11.91	12.48	12.92	13.27	13.58	45	31.50	33.10	34.14	34.89	35.52	36.04
21	11.80	12.71	13.30	13.76	14.13	14.44	46	32.36	33.98	35.02	35.80	36.43	36.97
22	12.57	13.51	14.13	14.60	14.98	15.30	47	33.21	34.86	35.92	36.71	37.35	37.13
23	13.35	14.33	14.96	15.45	15.84	16.17	48	34.07	35.76	36.82	37.62	38.27	38.82
24	14.13	15.14	15.80	16.30	16.70	17.05	49	34.94	36.63	37.72	38.54	39.19	39.75
25	14.92	15.96	16.64	17.15	17.58	17.93	50	35.80	37.52	38.62	39.45	40.12	40.68

of 7 TSLs for the traffic. Now, assuming our network must be dimensioned in such a way that the busy hour blocking probability cannot exceed 2%, we can note that the offered load must be 2.94 Erl. The 2% blocking probability refers to the failed call attempts as well as to the average blocking periods as a function of time during the complete one hour observation period.

Now, as we know the offered load, the realistic delivered traffic, taken into account that 2% of the call attempts are not successful, is 2.88 Erl, meaning that out of the seven available timeslots, 2.88 timeslots are in use on average.

The Erlang B formula further suggests that if there are more resources available, the efficiency of the delivered traffic increases. This means that as users enter and exit the system, the probability of available resource increases as a function of the number of the resources. This phenomenon is called Erlan gain.

3.2.2.2 Erlang B Extended

Extended variant of Erlang B is similar to the basic Erlang B model, but it takes into account retry efforts. The model has the retry percentage as a variable.

3.2.2.3 Erlang C

Erlang C model is based on the behavior of the calls that makes it possible for all blocked calls to stay in the system until they are handled. This model is useful, for example, for the customer care call center that places the incoming calls in queue whilst the personnel is serving previous customers.

In practical telecommunications networks the Erlang B is giving sufficiently accurate indication for the network load, so it is widely utilized even in cases when the network offers certain queuing functionalities.

3.2.3 Noise and Interferences

The capacity of telecommunications networks is dictated, in addition to the offered load and delivered load, by the noise power N that is added to the useful signal power S. There are various methods to estimate power levels, for example, via Gaussian distribution, depending on circumstances. An important base for all signal transmission is the theoretical limit that can possibly be transferred over the channel. This is called the theorem of Shannon. The maximum capacity is thus:

$$C = W \log_2 \left(\frac{S+N}{N} \right) \tag{3.12}$$

In this equation, C represents the maximum capacity for the information (b/s) when it is transferred within a bandwidth of W (Hz). S represents the signal power (W) and N noise power (W).

This equation can also be presented in a more practical format:

$$C = W \left(\frac{10}{3} \right) \log_{10} \left(1 + \frac{S}{N} \right) = \left(\frac{W}{3} \right) \log_{10} \left(\frac{S}{N} \right) \tag{3.13}$$

This equation is valid for cases when $S/N \gg 1$, and at the same time, the term $(1 + S/N)$ can be simplified to the form S/N. In practice, C can be presented with sufficient accuracy by the following equation:

$$C = \frac{W}{3} \frac{S}{N} \tag{3.14}$$

In practical systems, the performance is often informed by the number of bits per second that can be transferred within band of 1 Hz. In theory, the Shannon sets the limits in the way shown in Figure 3.2 when signal-to-noise ratio and information capacity are the investigated units.

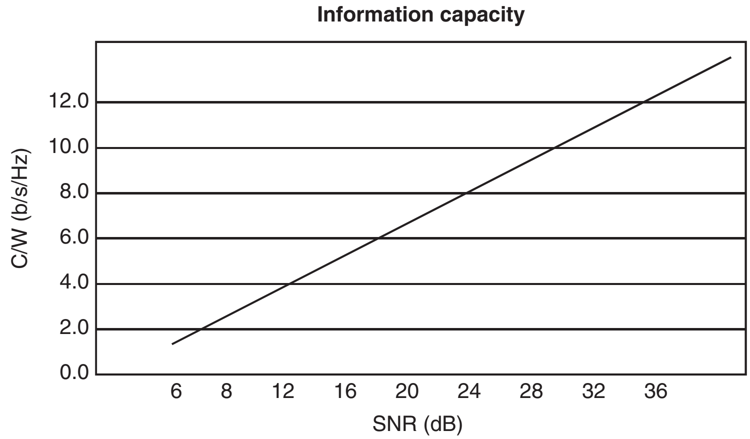

Figure 3.2 *Capacity of information according to Shannon limit.*

In practice, there are various components other than pure theoretical noise level affecting the performance. This is the case especially in the wireless environment. Thus, practical systems result in performance that is always lower than the theoretical Shannon limit indicates. Radio transmission includes components like multipropagated signals that have various different phases and amplitudes, as well as intermodulation and the same and adjacent channel interferences, which lower the performance of cellular networks.

The coding of data transmission finally dictates capacity limits. When heavier signal protection is utilized, less useful capacity can be obtained. On the other hand, with lighter signal protection capacity can be increased, but increased retransmission lowers useful data throughput. The balancing of error coding is thus one of the important optimization tasks in telecommunications. Typically, error coding is adaptive, so depending on the environment the system can adjust the level of correction automatically to a certain limit, which produces a certain amount of error rate which is still considered useful for the fluent user experience. The coding of the line dictates the final value for efficiency which can be typically 1–5 bits/Hz for the radio interface.

The relationship between spectral efficiency, which dictates the data throughput for the end-user, and distance of connection can be observed, for example, by measuring the modern mobile communications networks. The further the user is utilizing the service from the base station, the weaker the C/I ratio goes, which results in heavier coding and thus lower data throughput.

3.2.4 Other Typical Telecommunications Units

The quality of telecommunications can be described via countless units. Some of the essential ones include BER (bit error rate), FER (frame error rate) and PER (packet error rate). They all indicate the proportion of erroneous parts compared to total transmitted traffic. The errors can be measured with several different equipment, including network analyzers, field measurement equipment, and so on, or by utilizing the own methods of the network and its elements. Some examples of the latter are the network measurements for the Key Performance Indicators (KPI) that can be obtained from the operations and management centers of the networks.

Other indicators for the network quality investigations are, for example, dropped call rate for the core network and C/I (carrier per interference) for the radio network. Along with the introduction of the packet data networks both in core and radio systems, many new measurements have been adopted. Typical measurements for IP traffic are unsynchronized packet reception, delays and jitter of data transmission.

The quality of telecommunications networks can also be investigated via subjective measurements. One example of related measurements is MOS (mean opinion score), which consists of a scale of 1–5 that has been designed based on opinions of users for received voice connection quality, 5 referring to excellent, 4 good, 3 average, 2 weak and 1 referring to bad quality.

3.3 Evolution

3.3.1 Mobile Networks

The popularity of mobile communications has grown in an exponential way since the early deployments of the first analog networks. The idea of publicly available radio coverage for wireless communications regardless of location and time has lead to the growth figures of mobile penetration, resulting in more mobile subscriptions than landline subscriptions.

The first generation networks showed the way towards the totally new era where users are not any tighter to the location and time dictated by fixed telephony lines. Nevertheless, there were already some earlier systems deployed in the 1950s through to the 1970s [1]. As an example, there was a system called MTA in Sweden, deployed already in 1956, and operative in Stockholm and Göteborg. With the technology of that era, the equipment was big and heavy, and suitable only in a car mounted environment. Dialing was done via dialing plate, and calls could be delivered in an automatic way. In any case, there were only some hundreds of customers in the network, and as the technology was somewhat limited to support the idea in larger scale, initiation was merely an early trial of the concept without commercial success. The same types of trials in the car environment were also performed, for example, in Finland as early as 1952, again without commercial continuum in that era. The technology was finally mature enough in Finland 1971 for the commercial opening of ARP (Auto Radio Phone), which functioned until 2000 in 160 MHz frequency band. In other Nordic countries, the parallel systems were adopted in 450 MHz band. All these initiations may still be referred to as generation 0, with relatively small user bases.

The evolution started to take off along with the first fully automatic mobile communications system, and also the most international variant of the first generation mobile communications network deployed by that time in Nordic countries, which was NMT 450 (Nordic Mobile Telephone, the number indicating the frequency band in MHz). It was commercialized at the beginning of the 1980s. It was still meant only for analog voice services in a vehicular environment. An advanced version, NMT 900, was deployed in the Nordic countries later in 1980s, with handheld terminals from the beginning of the launch. The system was also adopted later in Switzerland, Russia and some other parts. Among NMT, there were various similar types of mobile networks in Europe, Americas and Japan.

The capacity was relatively limited in second generation systems, and regardless of international functionality especially via NMT, the roaming concept was still fairly limited. These were some of the reasons for triggering second generation mobile communication systems. GSM (Global System for Mobile Communications) has been up until then the most popular variant of the second generation. GSM was designed by ETSI (European Telecommunications Standards Institute). The technology was now completely digital. Amongst the benefits of digital service in quality, the separation of the terminal and subscriber module (SIM, Subscriber Identity Module) offered for the first time independence of subscriber number and device. The Circuit Switched (CS) data of GSM was introduced some years after the first commercial launches. CS data was included in phase 2 GSM standards that functioned with a maximum of 9.6 kb/s. At the same time, Short Message Service (SMS) became available as the networks and terminals started to support the functionality.

Short Message Service (SMS) has been very successful platform for various services. After the initial circuit switched data service evolution, also packet switched data came into the picture. GPRS (General

Packet Radio Service) and its extension, EDGE (Enhanced Data Rates for Global Evolution) showed the way for global wireless data development. Also other second generation systems appeared in the market, like CDMA based IS-95.

GSM is standardized in 3GPP (3rd Generation Partnership Project) since 1999. Its evolution includes DHR (Dual Half Rate) mode that provides four times more voice capacity compared to the original full rate codec, and DLDC (Downlink Dual Carrier) that utilizes two separate frequencies for the downlink data channels, providing around 500–600 kb/s data rates with 5 + 5 time slot configuration in Uplink/Downlink.

In ETSI, and then in the later phase when standardization activities were moved to 3GPP, the standardization of the 2G was executed under the term GERAN (GSM / GPRS / EDGE). GERAN is represented in the standardization body as one complete Technical Specification Group (TSG).

The SMGs (Special Mobile Group) of GSM standardization soon identified the potential limitations of the basic GSM platform. For this reason, the developing of third generation mobile communication system was initiated. The standardization of the UMTS (Universal Mobile Telecommunications System) was initiated in ETSI, first under the name FPLMTS (Future Public Land Mobile Telecommunications System). The standardization was moved to 3GPP along with GSM evolution in 1999. Later, the 3G term UMTS was also developed, the packet data services being the primary indicators for the enhanced 3G: HSDPA (High Speed Downlink Packet Access), HSUPA (High Speed Uplink Packet Access), and the combined evolution version of these, HSPA (High Speed Packet Access). Nowadays, further enhancements of HSPA are indicated by the term HSPA+.

The first theoretical UMTS data rate was 2 Mb/s, from which the practical top speed was somewhat decent 384 kb/s comparing today's maximum data rates. The evolving applications and customer habits required more data rates, which have been tackled by the introduction of HSPA.

Figure 3.3 shows the idea of mobile systems generations, interpreted from the ITU-R web pages. Please note that the terms "3.5 G" and "3.9 G" are practical interpretations of the industry and are not defined as such by ITU.

Up to 3G, the terminology has been mostly logical, and the definition of 3G is aligned with the technical abilities of the networks compared to previous 2G and 1G networks according to commonly understood and accepted principles.

As a conclusion, the term for 4G seems to be somewhat confusing. There are several interpretations, the strict versions coming from the regulation that admits only the evolved versions of LTE and WiMAX to the 4G category, and the loosest ones arising from the market interpretations that also accepts evolved 3G systems like HSPA+ to be represented as 4G technologies due to their considerably higher data rates than have been achieved previously in 3G.

Table 3.4 shows the definitions of different generations as indicated from ITU documentation.

Strictly speaking, as defined in ITU, the LTE defined by 3GPP in the Release 10, that is, LTE-Advanced, would be the first system complying the 4G definition. Nevertheless, the mobile telecommunications industry has mostly interpreted that the LTE is already within the performance of 4G. The practical explanation for these phenomena could be that LTE, in fact, is much closer to 4G than to 3G, and thus 4G has already been adopted in the marketing of LTE. Another explanation for this is also the debate about the final decision making body of the term 4G.

In any case, investigating the history behind the definition of 4G, ITUs Radiocommunication Sector ITU-R had completed the assessment of six candidate submissions for the global 4G mobile wireless broadband technology by 21 October 2010. According to the ITU-R terminology, the fourth generation refers to IMT-Advanced that contains various technical requirements, for example, for the data rates. The proposals resulted in two 4G technologies, "LTE-Advanced" and "WirelessMAN-Advanced." Both of these solutions have thus been recognized officially by ITU-R as 4G technologies, although the interpretation of the mobile telecommunications industry is somewhat looser in case of the first, Release 8 LTE.

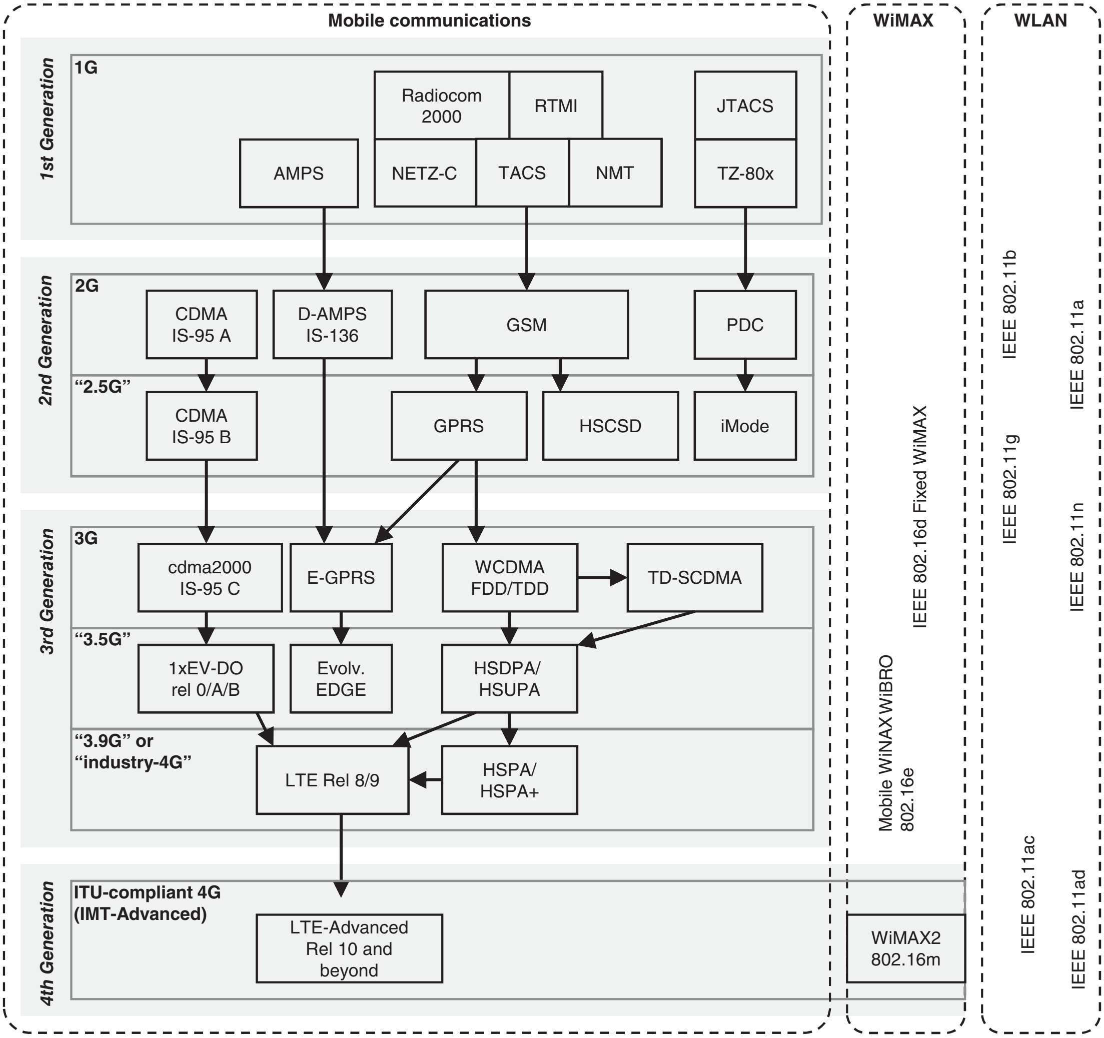

Figure 3.3 *The contents of the 2G, 3G and 4G of mobile communications. According to the original 4G performance definitions of ITU, LTE-Advanced (as of Release 10) and WirelessMAN-Advanced (WiMAX2) would comply with the requirements.*

Regardless the classification of LTE as a last step of 3G, or as a first or pre-step in 4G era, it drives the path towards the ITU-R-defined 4G and latest by the deployment of LTE-Advanced that LTE will present a full-ranked 4G system.

3.3.2 Mobile Data

3.3.2.1 Development of Mobile Data up to 4G

The first generation systems were principally meant for only voice calls, although it was possible to utilize data at some extent via a data modem and data adapter. In this way, the first generation provided peak data

Table 3.4 The definition of the generations of the mobile systems

Generation	Definition
1G	Analog, first completely or almost completely automatic mobile networks, that were principally meant for only voice calls, although special solutions were possible to adapt for data usage. The initial systems were based on the vehicle mounted equipment which could also be carried by the user. The weight was typically several kilograms, and there was a separate auricular for utilizing the equipment. Some examples of this phase of 1G are: NMT-450, Netz-C, AMPS. In the further development of 1G, there were also handheld devices. The devices were typically large in size and relatively heavy, although typically less than 1 liter and 1 kg. Examples of this phase are: NMT-900.
2G	The generally understood differentiator of 2G was the digital functionality of the systems. This has provided a fluent integration of also data services into the system in such a way that the data and messaging can be done by default via the user device itself. Examples of this generation are: GSM, IS-95.
3G	The further development of multimedia-capable systems led into the third generation. The main differentiator of this generation is the possibility to use considerably higher data rates. According to the original set of performance requirements, LTE belonged to the 3G phase.
4G	ITU-R has defined a set of principles and performance requirements for the fourth generation systems. In the initial phase of the compliance review by ITU, there were 2 systems that could comply, which have been the enhanced versions of LTE (as of Release 10) and WiMax (as of WirelessMAN-Advanced). As the mobile telecommunications markets have been growing heavily, and the competition is tougher than ever, there have been also parallel interpretations of the 4G capabilities. Often, the Release 8 LTE is interpreted to belong to 4G, and also HSPA+ is considered by various operators to be a 4G system [5].
5G	Ideas beyond LTE-A, with focus of the deployment in 2020.

rates of even 9.6 or 14.4 kb/s in good radio conditions, for example, overt the NMT-900 system with a standard modem connected to the device. Nevertheless, the first generation was never utilized for data services at a larger scale than in special telemetry or the most active users [2].

In the second generation, it was logical to include data services from the early phase. The GSM specifications provided the basic data bearers up to 9.6 kb/s after the first phase specifications had been released. Ever since, GSM specifications provided more advanced data via the modified coding schemes, making first the 14.4 kb/s possible, and soon up to nearly 60 kb/s via the circuit switched data. Table 3.5 summarizes the data evolution of GSM up to date.

It was only after the creation of packet data service concept of GSM (GPRS, General Packet Radio Service) when data usage started to fly due to the much more cost efficient resource utilization that the permanently reserved circuit switched data could offer. The multislot concept combined with the adaptive channel coding schemes via the EDGE (Enhanced Data Rates for GSM/Global Evolution) provides already as such theoretical data rates of 384 kb/s (DL), which also in practice are comparable with the first data rates of UMTS in downlink. At present, the most evolved version of the GSM data services is called DLDC (Downlink Dual Carrier), which combines timeslots from two separate frequencies in the downlink direction. Combined with the practical multislot concept with adaptive channel coding scheme, the data rate can be around 500 kb/s with the $5+5$ timeslot configuration, and close to 1 Mb/s in the theoretical case of $8+8$ timeslots. The latest variant, E-GPRS2 already includes 32-QAM and new modulation and coding schemes which provide up to about 1 Mb/s data rate in DL by using $5+5$ DLDC, and 0.5 Mb/s in UL via 4 TSL.

Table 3.5 The evolution of GSM data services

GSM data service	Data rate	Notes
Original CS data	9.6 kb/s	The original GSM data was based on circuit switched modem pool in MSC. Later, also ISDN rate adapters were included into the pool for direct ISDN data calls.
Enhanced CS data	14.4 kb/s	Enhancement was achieved by loosening of the code rate and by puncturing of the original data.
HSCSD	56 kb/s	Introduction of multislot technique provided more capacity, yet maintaining the same offered capacity per site.
GPRS	In theory, 21.2 kb/s per timeslot, multiplied by the maximum number of simultaneous timeslots.	First phase of the packet switched data over GSM.
E-HSCSD	2×56 kb/s.	The enhanced version of the HSCSD. In practice, the use is discontinued.
E-GPRS (EDGE)	In theory, close to 64 kb/s per timeslot multiplied by the number of simultaneous timeslots.	EDGE makes it possible to use more variable coding schemes and modulations.
DLDC	Twice as much data rate as the GPRS / EDGE offers in downlink.	By utilizing two carriers per user in the downlink direction, the data rate enhances accordingly.
E-GPRS2	With 5 TSL multislot and DLDC, up to 1 Mb/s in DL, and with 4 TSL, up to 0.5 Mb/s in UL.	Modulation schemes dynamically up to 32-QAM, and additional modulation and coding schemes.

The third generation mobile communications system was designed as multimedia capable from day one. The first basic data rate of 384 kb/s (DL) has increased recently considerably via the introduction of HSDPA (High Speed Downlink Packet Access), HSUPA (High Speed Uplink Packet Access), and nowadays via HSPA (High Speed Packet Access) and its evolved stage, HSPA+.

The current versions of third generation data services are already clearly taking off compared to the first UMTS data service data speeds.

The 3G data evolution is a result of 3GPP standardization. As a rule of thumb, commercial solutions (vendor releases) appear in the markets within 2 or 3 years from the freezing of respective releases of the standards.

3.3.3 Demand for Multimedia

The packet data service of GSM and UMTS has opened the way for the real multimedia era. As mobile data usage grows faster than ever at the global level, the spectral efficiency has been noted to be one of the most critical items for the operators. Among the spectral efficiency, that is, the optimized way to transfer a number of bits over a certain frequency bandwidth, also the bandwidth itself is one of the main drivers for the further evolution of mobile networks [3].

After the increased data rates provided via the HSPA evolution, 3GPP started to evaluate its successor candidates at the end of 2004. The goal for the new radio performance was set clearly higher that in any of the previous WCDMA solutions of 3GPP. The peak data rate requirement was decided to be at least 100 Mb/s in DL, and over 50 Mb/s in UL. In addition, the latency had to be improved considerably. The work name for this new idea was called Long Term Evolution (LTE), which, after this initial study phase, also became the public name for the respective radio interface. Nevertheless, 3GPP standardization calls the radio interface Evolved UMTS Radio Access Network (E-UTRAN).

The evolved radio interface with higher data rates also required significant enhancements in the 2G/3G packet data network side, that is, GPRS core, of the mobile networks. This 3GPP study item was called as System Architecture Evolution (SAE), and it is currently utilized as a practical name of the evolved packet core. In fact, LTE/SAE is a term that is utilized in practice to describe the Evolved Packet Core (EPC). LTE/SAE and EPC are thus parallel terms to describe the same item. EPC is documented in the Release 8 Technical Report GSM TR 23.882 and Technical Specifications GSM TS 23.401 and 23.402. The Release 8 was completed at the beginning of 2009.

EPC is capable of connecting GERAN, UTRAN, LTE, Femto Access Point and other non-3GPP access networks as CDMA, WiMAX and WLAN. The definitions for the handover procedures allow a rapid deployment of LTE in various scenarios. The handover between LTE and CDMA2000 eHRPD (Evolved High Rate Packet Data) handover is a special case that has been optimized.

EPC and the utilized access network are called Evolved Packet System (EPS). The main difference between EPS and previous solutions is that EPS does not contain any more definitions for the circuit switched domain connections, indicating clearly the evolution path towards the all IP environment. In this environment, IMS has an important role. In the all IP architectures, the voice and SMS is handled in alternative ways such as via session continuity that combines 2G/3G and LTE functionalities via system handovers or via the Voice over IP (VoIP) solution.

LTE/SAE provides modern means for utilizing Point-to-Point (PTP) IP based multimedia services that require more bandwidth and lower latency, including mobile TV/audio, online gaming and other applications that work with high data rates, constant connectivity and need service continuity on the move [4]. This is actually an iterative evolution, as the provision of more capacity, increased speed and reliability of the packet delivery also increases the utilization of data transmission of new solutions, including machine-to-machine (M2M, or MTM) communications.

Today, IP packet traffic dominates the traffic. Thanks to the bursty nature of the packet switched data, it is the most practical way to send and receive data in modern networks. The Internet is the most typical example of the functionality and benefits of packet data. It is thus logical that mobile communications have also gone towards the all IP concept. It can be claimed that in the initial phase of mobile multimedia, one of the driving forces has been the UMTS development which has introduced novelty architectural solutions in order to lower the round trip delays and thus to increase throughput, which after all is the concrete technical criteria for a fluent end-user experience.

It can be noted that as the architecture of Release 7 indicates, the Internet-HSPA (I-HSPA) has been the first step towards evolution which moves the functions of the Radio Network Controller (RNC) closer to the user, to the base station. The benefit of this solution has been that the packet switched data connection contains fewer elements than in previous phases of UMTS. This simplification results in shorter signaling delays and thus in lower round trip delays which, in turn, increase the throughput values.

One of the new phenomena along with evolved telecommunications networks for both fixed and wireless environments are new applications. They are also becoming increasingly important for the actual selection of the mobile devices by the end-users. For this reason, mobile device manufacturers are increasingly dependent on the ecosystem that includes devices, operation systems and respective applications.

3.4 Spectrum Allocations

3.4.1 ITU Regions and Principles

ITU (International Telecommunications Union) is the highest institute for the assignment of frequency bands in global level. The principles for frequency utilization are planned at the World Radio Conference (WRC) that is organized currently by ITU-R about every three or four years [5].

The primary task of WRC is to review the Radio Regulations. If noted necessary, WRC also revises the regulations. This can happen, for example, if older telecommunications systems are disappearing from the markets and new frequency divisions are justified. The international treaty governs the use of the radio frequency spectrum as well as the geostationary satellite and nongeostationary satellite orbits. The agenda of these revisions is prepared by ITU Council. As a basis for the new agenda, the Council takes into account the recommendations made by previous world radiocommunication conferences.

The work for the general agenda is already initiated four to six years before the actual event. The final agenda is also prepared well beforehand, as it is set by the ITU Council two years before the conference, via the democratic rules of the majority of Member States.

More specifically, WRC can revise the Radio Regulations and any associated frequency assignment and allotment plans; address any radiocommunication matter of worldwide character; instruct the Radio Regulations Board and the Radiocommunication Bureau, and review their activities; determine questions for study by the Radiocommunication Assembly and its Study Groups in preparation for future Radiocommunication Conferences [5].

The WRC handles the contributions from administrations, the Special Committee, the Radiocommunication Study Groups, and other sources related to the regulatory, technical, operational and procedural matters that should be considered by World and Regional Radiocommunication Conferences. After this, the Conference Preparatory Meeting (CPM) prepares a consolidated report to be used in support of the work of such conferences.

3.4.2 Regional Aspects

ITU defines the highest level rules for the frequency utilization, and the adjustments are done in national level. Regulators of each country can thus design the utilization of the frequencies that are assigned to those countries with certain limits for the interferences and so on. The utilization of the same frequencies in the border areas of different countries requires cooperation between the national regulators and telecommunications companies. This is often a matter of negotiations and coordination of the same channels of the frequency bands in order to avoid possible interferences. This is typical in mobile communications, radio link planning, and broadcast networks.

3.5 Physical Aspects

3.5.1 Radio Interface and Radio Links

The transmission of data and voice connections can be provided with radio links, for example, to the remote areas with difficult access or challenges in the construction of copper wires or fiber optics. In addition, radio link is a feasible solution in more dense areas with capacity limitations of cable network.

The radio link is typically done by utilizing microwaves with a highly directional antenna beam. The physical antenna can be, for example, a disc or horn model. The useful distance of radio links can be, depending on the antenna characteristics, frequency band and power levels, from some kilometers up to several tens of kilometers. For the lowest frequencies, the distance can even be several hundreds of kilometers. The main point in the designing of the radio link is to make sure there are no major obstacles in the first Fresnel zone as it considerably attenuates the received power level. In practice, the link should comply with LOS (line of sight).

One of the main benefits of radio links is the possibility to deploy them in a fast time schedule, given that the frequency band is available for the link. The investment level is also low compared to the physical cables located in poles or below ground. The utilization costs are low, and furthermore, the equipment can be transported easily to alternative location, and links can be redirected by adjusting the antenna directions. Drawbacks include additional attenuation due to rain and occasional interferences caused by other radio transmissions nearby. In some cases, when the transmission antennas are not installed steadily enough, there can be variations in the received power level due to the moving of antennas in the wind. There are also some rare examples of extremely sensible variations due to the seasonal affects of the surrounding environment. One example is a metallic roof near the transmission antenna. The roof may get heated due to the sun rays which twist the roof sufficiently to make modifications in the radio link within the first Fresnel zone. In general, the planning and deployment of radio links require accurate work to function correctly.

Typical radio links function in few GHz frequency bands, for the link distance of some tens of kilometers. Today, for short distance links especially in the city environment, the typical frequency bands are 15, 18, 23, 26, 38, 53 and 58 GHz. In general, the lower the frequency band is, the longer the obtained useful distance also is. It should be noted that the chaining of the links is also possible to extend the final distance.

Depending on the carrier bandwidth, several hundreds of PCM basic channels can be deployed. The basic European variant offers 30 PCM channels for the communications whilst the North American system provides 24 channels.

In addition to terrestrial radio links, satellites are also suitable for radio transmission. They are thus utilized especially for long-haul connections, for example, in intercontinental connections, as one option for telecommunication systems. The satellites are either passive (i.e., merely repeating the received signals further), or active (e.g., they also amplify the received signal prior to repeating it).

The signals of the satellite systems must be able to penetrate the ionosphere in all circumstances, which means that the satellite frequency bands are typically higher than 3 GHz.

3.5.2 Electrical Wires

The foundation for telecommunications systems are physical wires or cables. This applies basically to the core part of all systems regardless of whether they use wireless radio access technologies. The typical core solutions are based on the copper lines or fiber optics. Both are useful in varying conditions, installed in pole supports or underground, under water or indoors.

3.5.3 Copper Lines

Copper lines have been utilized traditionally for telecommunications systems practically since the birth of telephone networks. The benefit of copper is its robustness and easy installation when wires need to be reconnected. The challenge of long copper lines is the gradual attenuation of the signal which has negative effects, especially in analog systems. Long-haul connections thus require various repeaters to keep the signal level sufficiently high. The drawback of repeaters is that – in addition to signal level – they also increase noise level. In analog systems, the resulting signal-to-noise ratio can thus be low enough to cause disturbing effects to end-users.

Figure 3.4 *The basic idea of twisted pair.*

3.5.3.1 Symmetrical Cable

Copper cables are still useful and are widely used in today's networks. The simplest variant is the twisted pair. The basic solution can include two wires wrapped around each other in order to avoid induction interferences. It is useful only for local connections, for example, as a last mile solution for basic telephone lines because of the level of interferences that grows as a function of the distance, for example, in the form of overlapping calls. This problem can be attenuated to some extent by utilizing 2 or 4 pairs within the same line.

Another problem is the relatively high attenuation characteristics of copper cable compared to more modern solutions like fiber optics. The benefits of copper are that they are easy to connect cables together in the field, and easy installation.

The impedance value for the typical symmetric twisted pair copper cable is 120 ohms. In case the resistance value of individual parts of the same line differ from each other, reflections in the connectors will occur, resulting in echoes and other types of interference.

The importance of the twisted pair as shown in Figure 3.4 has reduced along with the increasing popularity of fiber optics. If the distance between the user and the nearest switching center is small enough, a twisted pair can be utilized for ADLS modem located at the end-user premises. Regardless of drawbacks, the symmetrical cable is also useful for basic data rates of PCM transmission. Nevertheless, copper cable is not suitable for high speed transmission of larger areas.

The quality of the twisted pair depends on the symmetry of the individual cables. Figure 3.5 shows typical imperfections of cables. In case the twisting is nonuniform, it may result in EMC (electro magnetic compatibility) between the cables and surrounding environment. On the other hand, there may be interferences caused by induction. In practice, perfect symmetry is challenging to obtain due to the material of the cables and the methods of production.

Also the role of the connectors may be significant for avoiding interferences and connector losses in end-to-end performance. The careful installation of the connectors, as well as good grounding of the elements and cables, as shown in Figure 3.6, is essential for high-quality networks. In the case of symmetric cables, the importance of the high-quality connectors is notable, for example, for short-distance cables (<100 m), with high bit rates (>1 Gb/s).

The performance of cables depends on the frequency of the signal. The electromagnetic interferences are not a major issue in low frequencies, but the higher the frequency is, the weaker the immunity will also be.

One solution for enhancing the immunity is a correct electromagnetic shield which can be done by an aluminum folio that is wrapped around the connector. If there is an extra layer of wired material on top of the aluminum shield, the cable can be grounded relatively easily.

3.5.3.2 Coaxial Cable

The importance of twisted pair is diminishing all the time. A more developed version is a coaxial cable, as presented in Figure 3.7, as its interference shielding is higher. In addition, coaxial cables support much higher power levels that paired cables, depending on the thickness of the cable.

The nominal impedance of the coaxial cable depends on the relative dimensions of the inner part (the transmission wire) and outer part (protecting shield) of the cable. Typical impedance values of coaxial cable are 50 and 75 ohm.

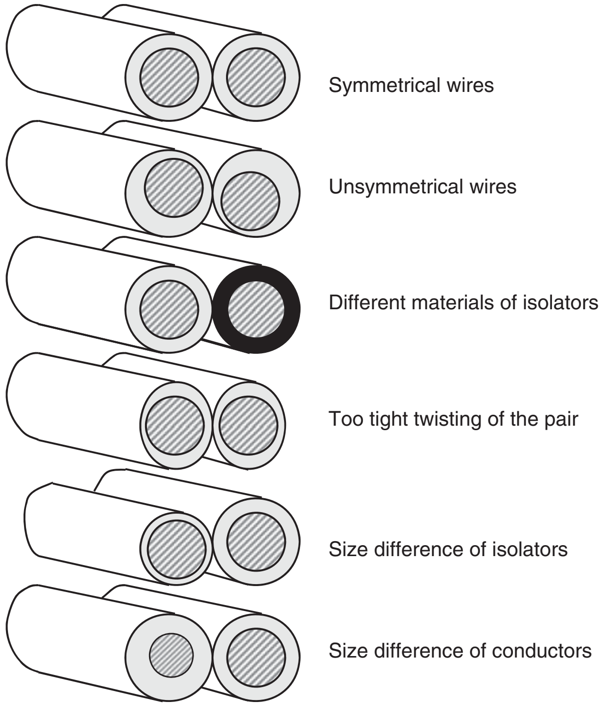

Figure 3.5 *The possible root causes of the nonuniform cables.*

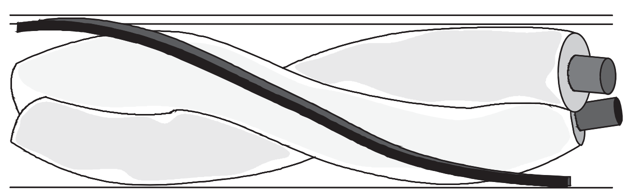

Figure 3.6 *The immunity for the interferences can be enhanced in twisted cables by including a separate, twisted copper line around the actual cables.*

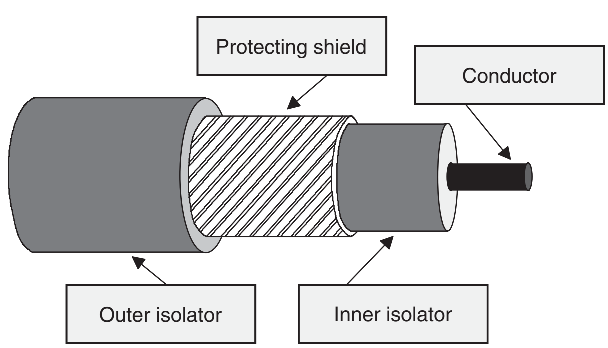

Figure 3.7 *The principle of coaxial cable.*

Table 3.6 *Some attenuation values of typical coaxial cables in temperature of 20 Celsius*

f/Mhz	AJS 75-5 (75 ohm)	RG 58 C/U (50 ohm)
0.1	0.3	0.46
0.5	0.6	0.86
1	0.8	1.2
5	1.9	3.2
10	2.7	4.7
50	6.3	11.0
100	9.1	16.0
500	21.9	39.7
1000	33.0	60.0

The coaxial cable is useful for the signal transmission for core and radio systems, including antenna cabling of cellular and TV systems.

The attenuation value of coaxial cable depends on the diameter of the cable, the frequency of the signal, on the isolating material and temperature of the cable. The isolating of the inner and outer parts of the coaxial cable can be done by air (which means that the inner cable is supported by isolating rings inside of the cable to maintain uniform distance) or by isolating material. Table 3.6 shows examples of the behavior of the attenuation value for typical cables.

The selection of coaxial cable can be optimized as for the power level, attenuation and other characters like flexibility in the installation environment. A cable that supports more power is basically thicker, which also has a cost effect and is more challenging to install due to flexibility issues. For a low-power system like signal transmission a thin cable of, for example, half an inch can be applied, whilst coaxial cable used as antenna feeder for high-power TV tower might need to be several inches in diameter. In a mobile communications network, the antenna system can typically utilize 7/8 inch cables which often offer an optimal balance for the attenuation, flexibility, supported frequency, costs and maximum power levels. As an example, a thin coaxial cable would be easy to install to the mobile communications system's elements between synchronization modules located inside of the building and rooftop mounted GPS antenna, but the attenuation of this solution can be several tens of decibels. If the system does not support high attenuation, the only solution would be to install thicker cable instead, with some more challenges in the flexibility of the cable when it is installed. The attenuation values can be investigated from the technical data information of the respective cable. Figure 3.8 shows some typical examples of attenuation values.

One special solution for coaxial cable is to use it as radiating cable, that is, leaking cable that is meant to function as a radiating antenna element along a longer distance. This solution is especially suitable in tunnels to create radio coverage of a mobile network via a long distance in a difficult environment where radio signals cannot otherwise normally penetrate. Another use case is an elevator tunnel. The radiating cable consists of openings in the protecting shield. The openings are dimensioned in such a way that the impedance is expected, and the radio signal can be directed on the way the openings are. It should be noted that when the radiating cable is installed, for example, to the wall of the tunnel, the openings must be pointing outside of the wall. As the isolating outer material prevents seeing the actual opening physically, the markings of the cable must be noted carefully.

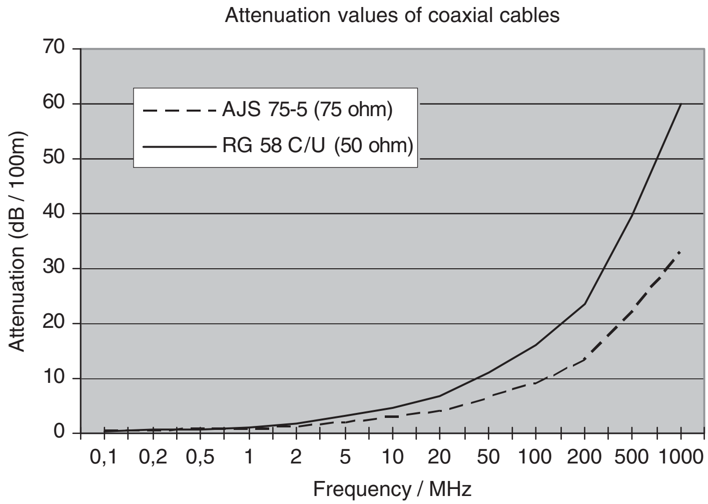

Figure 3.8 *Examples of the attenuation values of typical coaxial cables.*

3.5.4 Fiber Optics

Fiber optics is suitable solution for long-distance data transmission. The basic functioning consists of light emitting diode (LED) or laser that utilizes OOM (on-off modulation). The conversion of electrical to optical, and vice versa in the receiving end, is straightforward. This conversion is called E/O (electrical to optical) and O/E (optical to electrical).

In modern telecommunications networks, fiber optics is a dominating solution for transmission due to its superior abilities to deliver capacity. As a rough estimate about the benefits, a capacity of 2 Mb/s of coaxial cable equals 300 Mb/s of the fiber optics, whereas the short-distance symmetrical cable is limited to 10 Mb/s. Additional benefit of fiber optics is the lighter weight which is beneficial in the installation work. As the attenuation value of fiber optics is very much lower compared to any metallic cables, the need for repeaters is lower, and fiber optics are thus suitable for both short and long-distance core solutions. Also the optical fiber attenuates the signal, which requires digital regenerators. Figure 3.9 shows the high-level physical structure of fiber optics.

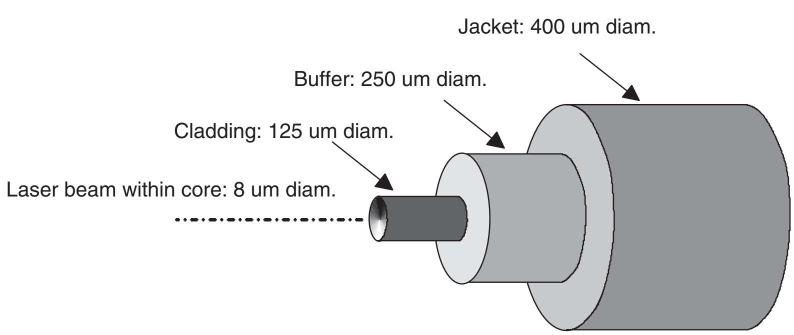

Figure 3.9 *The main components of fiber optics.*

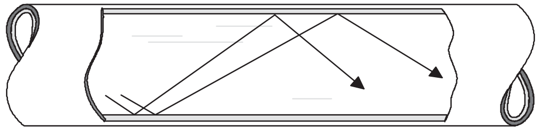

Figure 3.10 *Complete reflected rays in multimode fiber.*

The outer layer of fiber optics is the protecting shield that makes it possible to bend the cable with relatively small diameter. The minimum size of the loop depends on the cable type, and the instructions of manufacturer should be followed in order to avoid the breaking of the interior fiber.

The inner part of the cable consists of the glass fiber and the core part of the cable that is utilized for inputting and outputting the optical transmission. There are three classes of fibers: Single-mode fiber and two variants of multimode fiber. The latter consists of two variants depending on the ways the ray reflects.

3.5.4.1 Single-Mode Fiber

The optical fiber can be understood as optical waveguide in such a way that the fiber supports one or various confined transverse modes. The energy of the light is not totally confined in the core of the fiber. Instead, a considerable amount of the energy travels in the cladding as an evanescent wave. This provides means for the light to travel within the fiber. Fiber that supports only mode is monomode fiber, more commonly known as single-mode fiber. The core of the single-mode fiber is typically about 8–10 micro meters.

The most common type of single-mode fiber has a core diameter of 8–10 micrometers and is planned for use in the near infrared.

3.5.4.2 Multimode Fiber

The multimode fiber supports more than one mode of propagation, which results in the name of this type. The behavior of the light bouncing within the core of the fiber can be modeled assuming geometric optics if the core is large enough to support more than a few modes. Multimode fiber has diameter of a core about 50 micrometers to some hundreds of micrometers.

Rays that meet the core-cladding boundary at a high angle, that is, exceeding the critical angle, are completely reflected. The critical angle which is the minimum angle for complete reflection is the difference in index of refraction between the core and cladding materials.

In graded-index fiber, on the other hand, the index of core refraction will decrease continuously between the axis and the cladding. As a result the light rays will bend smoothly as they approach the cladding instead of total reflection from the core-cladding boundary. These curved paths reduce multipath dispersion. The ideal index profile is approximately comparable to a parabolic relationship between the index and the distance from the axis. Figures 3.10, 3.11 and 3.12 summarize the modes of fiber optics.

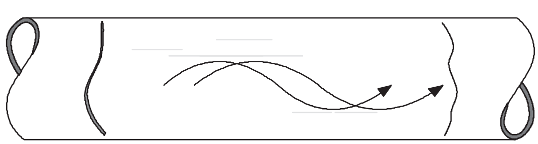

Figure 3.11 *The reflection is smoother in the graded-index fiber.*

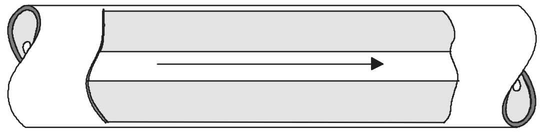

Figure 3.12 *The principle of the single-mode fiber.*

There are both directly propagating rays of light, as well as reflected rays in the fiber optics. The different angles of various angles of the rays result in dispersion. The diameter of the inner part of multimode fiber is typically in range of 50−62.5 μm, and the diameter of the outer glass layer is 125 μm. The reflections are caused because of the relatively big diameter of the inner layer, and in addition, the propagating ray of light is spread. In the reception side, this thicker ray causes uncertainty and thus increases the bit error rate.

The inner part of the single-mode fiber is much thinner, typically in range of 1−10 μm, but the outer layer's diameter is the same as in the case of multimode fiber, that is, 125 μm. This results in lower dispersion, which means that the ray of light propagates straighter in single-mode fiber, and is thus especially suitable for the high bit rate transmissions and for long distances.

The three basic wavelength areas of fiber optics, so-called propagation windows of the optics, are 850 nm, 1300 nm and 1550 nm. The single-mode fiber wavelength areas are 1300 and 1550 nm (meaning the second and third windows of the fiber optics), whereas the multimode fiber has the values of 850 and 1300 nm (the first and second window). The multimode fiber's attenuation value is about 3 dB/km, whereas the single-mode fiber has attenuation of roughly 0.5−1 dB/km.

A single fiber optics cable is capable of transferring multiple channels via FDM (frequency division multiplex). In addition to the lower attenuation, another benefit of fiber optics compared to the copper wire is the immunity to the external interferences due to the lack if induction of electromagnetic fields. The transmission of the fiber optics is not affected by the weather conditions or temperature changes. According to the principles defined by CCITT, the typical time of use for fiber optics is at least 40 years.

Typical use cases for fiber optics are video conferencing, the broadcast transmission of TV and radio, and broadband data transmission, including the fast Internet solutions and B-ISDN. Also cellular networks typically utilize fiber optics as a base of the core network infrastructure.

The evolution path of fiber optics include WDM (wavelength division multiplexing), which is one possibility to increase the capacity considerably. It can be a base for tens of Gb/s through hundreds of kilometers of distances, by utilizing the single-mode fiber.

References

1. Telecom history, 1st generation: http://gkos.com/viestin/nmt.html (last accessed 21 July 2014).
2. Penttinen, J. (2006) *Tietoliikennetekniikka (Telecommunications Technology), Parts 1 and 2*, WSOY, Helsinki, Finland.
3. Penttinen, J. (2011) *The LTE/SAE Deployment Handbook*, Wiley.
4. Discussion about 4G of ITU: http://www.networkworld.com/news/2010/121710-itu-softens-on-the-definition.html (last accessed: 21 July 2014).
5. WRC. http://www.itu.int/ITU-R/index.asp?category=conferences&rlink=wrc&lang=en (last accessed 21 July 2014).

4

Protocols

Jyrki T. J. Penttinen

4.1 Introduction

The protocols are the basis for practically all the data flows of the modern communications systems. The benefits of protocols include the fluent methods of functionality, and in an ideal world, protocols at their levels are capable of understanding the protocols of any other equipment of the same system regardless of manufacturers. In practice, this is rarely the case in wider environments, but in any case, good work has been done in international standardization in order to create as compatible system blocks as possible.

The physical aspects of the connection include the selection of signal types, voltage levels, connector types, cabling and network topology. For a direct, dedicated point-to-point communication only between two users it would be straightforward to send and receive signal levels according to a certain rule such as modulation and demodulation of the signal, for example, in the walkie-talkie connection. Nevertheless, for wider communications even for point-to-point cases, there are hardly ever any possibilities of applying this method. Furthermore, various network elements are involved in the communications route. This means that there needs to be a commonly agreed way in which the communication is handled between network nodes to deal with, for example, the collision of multiple access attempts of users, error recovery, resending, and so on. Protocol basically refers to the standardized method of transmission to a network.

This chapter describes theories of protocols, including the OSI model of ISO, and typical protocols that have been adapted to modern communications systems. There are also many other protocols developed, but this chapter concentrates on the most essential ones of modern telecommunications systems. The chapter also describes the realities of live networks, and provides reasons why the theoretical models are not necessarily deployed as such everywhere.

All telecommunication networks are made up of five basic components that are present in each network environment regardless of type or use. These basic components include terminals, telecommunications processors, telecommunications channels, computers, and telecommunications control software.

The protocols thus take care of the delivery of the message within a communications network. The protocols participate in this process in all the elements of the network, that is, *terminals* (which are for initiating and

The Telecommunications Handbook: Engineering Guidelines for Fixed, Mobile and Satellite Systems, First Edition.
Edited by Jyrki T. J. Penttinen.

terminating the message delivery, acting as input and output devices), *telecommunications processors* (supporting transmission and reception of data between terminals and computers), *telecommunications channels* (provides the actual transmitting and receiving, for example, via media such as copper wires, coaxial cables fiber optics and radio links), *computers* (executing their communication assignments), and *telecommunications control software* (controlling network activities and functionality).

In the fixed telecommunications network, the traditional protocol that has been utilized for decades is SS7. The signaling of, for example, GSM, UMTS and LTE networks is substantially more complicated than in the traditional fixed telephony network. The difference is a result of the mobility of cellular network terminals, which creates needs for additional protocols, mobility and connection management, security and many other parts of the protocol layers. The common signaling systems of the fixed telephony networks have thus been extended for the mobile networks, partially overlapping with the other networks, and partially being network-specific. The common phenomenon is that each time when there are new mobile communications systems there are also additions to the signaling.

In this chapter, the theory of the protocols is presented, with the theoretical OSI model that is adapted to the commercial networks only partially. Then, the protocol solutions of different fixed and mobile networks are discussed.

4.2 OSI

4.2.1 General

The Open Systems Interconnection (OSI) model is a result of the work done in International Organization for Standardization (ISO). The aim of OSI is to present standardized abstraction layers for communications systems. The model consists of logical layers, each having a set of communication functions that communicate with lower and upper layers. In addition, the functionalities at certain layer are connected with each others by a horizontal connection.

As a result of the standardization work, ISO developed the OSI framework architecture. OSI contains two areas: (1) Basic Reference Model, that is, seven-layer model that acts as an abstract model for the networking, and (2) a set of protocols functioning within different layers and providing communications between the layers.

The seven-layer concept has been adopted at least partially in many telecommunications systems, including mobile communications networks, due to its logical functionality. In the model, the protocols of certain protocol layer provide the means for an entity of certain host to interact with another entity at the same layer within another host. The clear benefit of the protocol layer structure is that the service definitions abstractly describe the functionality that layer N-1 receives from the layer N, the term N referring to the number of the originally defined seven layers that operate in the local host.

The OSI definitions can be found, for example, at ITU-T, which has published the OSI layers in the recommendations of X.200-series [1].

The recommendation X.200 defines the seven layers in such a way that the lowest level is in the bottom, and is numbered as 1. The basic functionality can be summarized in the following procedure: An *N+1 entity*, which resides at layer N+1, requests services from an *N entity*, which resides at layer N. At each level, two entities called *N-entity peers* interact with each other as defined by the functionality of the *N protocol*. The interaction happens by sending and receiving PDUs (Protocol Data Units) according to the respective protocol definitions.

In the OSI data flows, an SDU (Service Data Unit) is passed from a certain OSI layer to a lower layer. At this stage, the lower layer has not yet encapsulated it into a PDU. In other words, an SDU is a data flow that is transmitted unchanged.

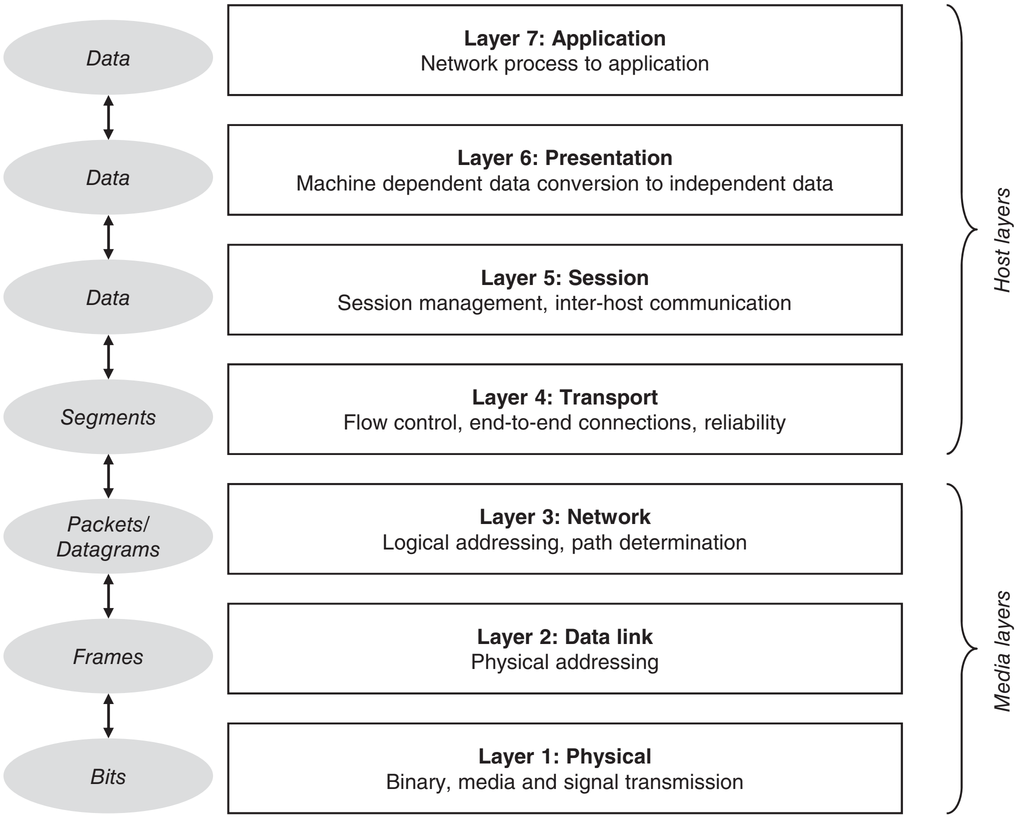

Figure 4.1 *The OSI layer definitions.*

The relationship between PDU and SDU is the following: A certain PDU found at layer N is the SDU of layer N-1. In the transition of SDU to a PDU, an encapsulation occurs, which is executed by the lower layer. The result is that all the data that belongs to the SDU is encapsulated within the PDU. Furthermore, the underlying layer N-1 adds essential information for the fragmentation and defragmentation, that is, headers and footers to the SDU, prior to making it as the PDU of the very same layer N-1. This is essential for making the transmitting of data possible between the origin and destination. Figure 4.1 summarizes the seven OSI layers as defined by ISO.

In the OSI reference model, the units are divided in the following way for each protocol layer: The highest level Application Data Unit (APDU), followed by Session Data Unit (SPDU), Transport Data Unit (TDU), Packet, Frame, and as the most elemental unit, Bit.

Each of the OSI layer communicates with underlying and overlying protocol layers in such a way that underlying protocol adds a certain amount of information to the information flow, like headers and bits required by the protection mechanisms. Also the possible fragmentation of data requires mechanisms for the defragmentation again in the upper layers of the receiving end. Figure 4.2 clarifies the overall idea of the layered structure and functionality [2].

The method of organizing network software, firmware and hardware in a stack of layers has various benefits, basically because layer N provides a service to the next layer (N+1), though keeping the details of its internal state and algorithms hidden [2]. This results in a hierarchical approach, with modularization,

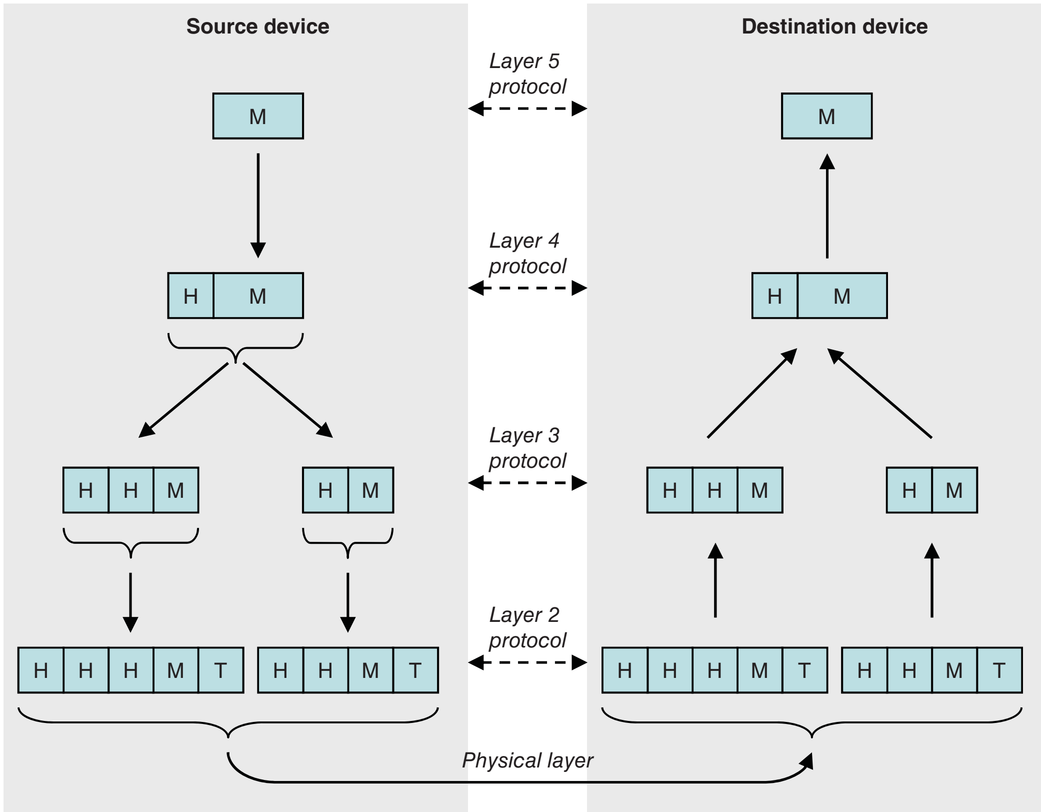

Figure 4.2 The concept of the layered protocols. The lower the protocol layer is, the more overhead there is due to header information (H) which is utilized, for example, for routing and fragmenting/defragmenting of separate packets, sending separate packets, collecting and reconstructing them even if they are received in reversed order. In this figure, M = Message block and T = tail block.

information hiding, and encapsulation of data. Furthermore, it is straightforward to create abstract data types which benefit, for example, object oriented programming [2].

The role of protocol can be a dedicated task in that layer in question, or a multitude of tasks in order to make sure the information flows between the transmitting and receiving ends correctly. Some tasks of protocols are addressing, error and flow control, and routing of data between the origin and destiny. Figure 4.3 further clarifies the role of protocol in general.

It should be noted that the OSI model is ideal and somewhat theoretical, with the aim of modeling as much as possible in the communication chain. In practice, though, the mapping of layers is not always straightforward. Sometimes the functionality of certain solution belongs in more than one original OSI layer's role, whereas sometimes there is no need to present everything in an OSI structure. One such example is the TCP/IP reference model which is lacking in OSI layers 5 and 6, whilst OSI layers 1 and 2 are represented in TCP/IP as one. Figure 4.4 clarifies this.

The key message from the layered protocol structure is that communication is allowed solely between adjacent layers, meaning that there should be a corresponding protocol layer both in the transmitting and receiving ends. Furthermore, there should be communication only via procedure calls and responses. In

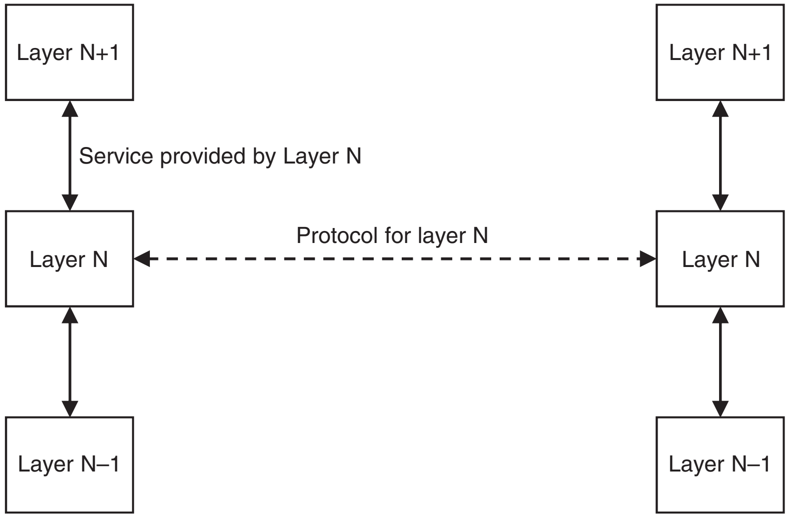

Figure 4.3 *The role of protocol in layer N. The protocol layer receives services from upper layer.*

	OSI model	**TCP/IP**
7	Application layer	Application layer
6	Presentation layer	N/A
5	Session layer	N/A
4	Transport layer	Transport layer
3	Network layer	Internet
2	Data link layer	Host-to-network
1	Physical layer	

Figure 4.4 *An example of TCP/IP protocol layer mapping with OSI model.*

general, the services at different layers are designed via protocols at these layers. For more detailed information about adaptation layers, please refer to [3–8].

4.2.2 Physical Layer (1)

The first OSI layer, physical layer, describes physical and electrical characteristics of the devices that are used in a network. More specifically, the physical layer defines the interactions between devices and the medium for transmission. These physical means for transmission may be, for example, copper wires, coaxial cables, and fiber optics. The physical layer also refers to radio links. As the layer defines the physical characteristics of the media, there are topics like voltage levels, connector pins, cables, repeaters, hubs and other equipment between the user terminals and network elements.

The primary tasks of the physical layer are:

- Management of the establishment and termination of communication link with communications medium.
- Modulation and transmission of the signals over physical communication channels.
- Physical layer also takes part in flow control, and in general, in activities that are meant for the sharing of physical resources.

As some examples, Ethernet cables, as well as cellular systems and Bluetooth radio interface are defined in physical layer.

It can be generalized that the physical OSI layer is related to hardware and streams of raw data bits. The task of the physical layer is to place data onto network media, and after the transmission to another element, to remove the data from media to higher protocol layers.

In computer networking, the physical layer transfers frames from the origin (memory) of computer to the network transmission media. The data is handled as such, so there is no addition of headers or compression or other control overhead, as the higher layers take care of these functions. The tasks of the physical layer are thus related to the mechanical, electrical, functional and procedural sides of the data transmission. The Packet-Protocol Data Unit (PDU) in this layer is the smallest unit of the data transfer, that is, a bit. Simply said, when the cable breaks down, it is a layer 1 problem.

4.2.3 Data Link Layer (2)

The second OSI layer, data link layer, provides means to transfer data between network entities like physical network elements or functional blocks. The data link layer is also able to detect and correct errors that are generated due to the somewhat nonoptimal characteristics of the physical layer.

The data link layer is thus responsible for physically passing data from one node to another. Frame represents the PDU at this protocol level. Some of the tasks of the data link layer are flow control, error detection and error control. In addition to the data link layer itself, it also has two sublayers: Media Access Control (MAC) and Logical Link Control (LLC).

As the physical layer does not take care of the control of the data flows, the data link layer must assure that the data packet sequence numbering, packet addressing and primary error control data are taken care of prior to letting the data be delivered via a physical layer.

4.2.3.1 Connection Oriented System

Connection oriented data link protocols make the framing of the data, take care of error detection and possibly correction, and are able to control transmission data rate. There are many practical means to make these functions happen, like sliding window flow control and acknowledgement mechanisms. Some examples of this are SDLC and HDLC, as well as LAPB and LAPD.

4.2.3.2 Connectionless System

An example of the connectionless system is Ethernet, via the media access control (MAC) sublayer that takes care of the access of multiple users of the network resources and detects errors. The next level up is the logical link control (LLC) layer which is also within the data link layer like MAC. LLC is media-independent solution to take care of the addressing and multiplexing of various connections (data streams of different users). MAC and LLC layers are thus performing the tasks as a pair within the data link layer of OSI.

4.2.4 Network Layer (3)

The third OSI layer is network layer. Its tasks include the needed functionalities for the delivery of variable lengths of data sequences, from the destiny to the receiving entity of another network. As a difference with network layer, the LLC takes care of the connections within the same network. Network layer also takes care of the quality of service (QoS) based on the level requested by transport layer. Other functions of the network layer include routing, fragmentation and reassembly of the data packet units, as well as reporting of faults in the delivery of the data. One example of the equipment at this level is routers, which are designed based on addressing scheme.

In the OSI model, the network layer may have three sublayers:

1. Subnetwork-dependent convergence, which takes care of data transmission between the edges of the network, not only until the next hop.
2. Subnetwork-independent convergence, for example, via IPv6 or CLNP. This method is responsible for the reliable data delivery in such a way that it takes care of the delivery only until the next hop, and in case of errors, the method is able to detect it but not to correct.
3. Subnetwork access, which includes necessary protocols for interfacing with data networks like X.25.

There are numerous protocols defined for the network layer, including protocols for routing, network layer address assignment, multicast group handling, and so on. Nevertheless, in general, the network layer performs addressing and routing. Network layer thus has the necessary functions for making it possible to communicate between indirectly connected entities. The task of network layer is to deliver the messages until the next layer 3 entity via hops, until the final destiny is reached. In practice, this is done by routing data packets between nodes. The data packet thus forms the PDU in this layer.

4.2.5 Transport Layer (4)

The fourth layer, that is, the transport layer of the OSI model, transfers the data in a transparent way. This level takes care of the data transfer between end-users in a reliable way. The transport layer thus has a control mechanism for assuring the reliability of data link. This happens via flow control. This layer also performs segmentation, desegmentation, and error control. In addition, there is acknowledgment mechanism built in the transport layer for successfully received packets.

There are five classes defined in OSI model, from class 0 (TP0) to class 4 (TP4). In general, TP0 provides least features and, for example, no error recovery at all. It is thus suitable for a network that is reliable, whilst TP4 supports most of the functionalities and is thus optimal for unreliable networks, like the Internet.

The following list summarizes the functionalities of each TP class:

- TP0: Connection-oriented network, segmentation / reassembly.
- TP1: Connection-oriented network, concatenation / separation, segmentation / reassembly, error recovery, re-initiation of connection, reliable transport.
- TP2: Connection-oriented network, concatenation / separation, segmentation / reassembly, error recovery, multiplexing and demultiplexing over single virtual circuit, explicit flow control.
- TP3: Connection-oriented network, concatenation / separation, segmentation / reassembly, error recovery, re-initiation of connection, multiplexing and demultiplexing over single virtual circuit, explicit flow control, reliable transport.
- TP4: Connection-oriented network, connectionless network, concatenation / separation, segmentation / reassembly, error recovery, multiplexing and demultiplexing over single virtual circuit, explicit flow control, retransmission on timeout, reliable transport.

In general, transport layer includes tunneling protocols which are able to carry, for example, non-IP protocols. As an example, L2TP carries PPP frames inside transport packet. It is not always explicit what protocol belongs to which OSI model, so depending on the complete functionality of the protocol it can be imagined in different layers. From the typical IP protocols, the transmission control protocol (TCP) and user datagram protocol (UDP) normally are placed on transport layer of OSI.

The aim of transport layer is to get the layer 4 messages delivered to the destination (to the layer 4 of another entity) in a reliable way. In this sense layer 4 communications is end-to-end, within the network, unlike the layer 3 that utilizes the hop-method to reach only the next layer 3 entities. If the message gets too large when delivered to underlying protocol level, the transport layer also takes care of fragmentation. Also the general integrity of the data is concern of transport layer. In this layer, segments form the PDUs.

4.2.6 Session Layer (5)

The fifth OSI layer, session layer, takes care of the controlling of communication between terminals. It thus manages the establishment, maintaining and releasing of connections. The ways of communication are full-duplex, half-duplex or simplex. In practice, the session layer may be implemented in application environments with remote procedure calls.

In general, session layer controls communication between applications and hosts. The PDU in this layer is data. As an example, login passwords may be taken care of in session layer, as well as the exchange of user IDs.

4.2.7 Presentation Layer (6)

The sixth OSI layer, presentation layer, is meant to establish context between higher layer (application) entities. The presentation layer provides transformation of data into the acceptable format for applications. The presentation layer formats and encrypts data that is sent over the network.

In general, the presentation layer executes needed data transformations and formatting that the applications require. Also data compression is performed in the presentation layer. As in the session layer, the PDUs of the presentation layer are formed by data.

4.2.8 Application Layer (7)

The highest, seventh OSI layer is the application layer. It is thus nearest to the end-user. This means that the OSI application layer interacts with the actual application in one side, and the user in the other side of the application whenever the application includes functionalities for the communication.

Some examples of the application layer are X.400 mail protocol and CMIP (Common Management Information Protocol), these being in OSI stack. On the TCP/IP stack, examples include HTTP (Hypertext Transfer Protocol), SMTP (Simple Mail Transfer Protocol), File Transfer Protocol (FTP) and SNMP (Simple Network Management Protocol).

In general, the application layer provides network services that support applications of the host. The PDU of the application layer is formed by data, as in the case of session and presentation layers.

4.2.9 Practice

Despite the good intentions of modeling the communications systems via the protocol layers of OSI, hardly any of the commercial networks have adapted the complete OSI layer structure. Nevertheless, OSI has indicated a functional and logical way to construct communications systems, so practically all the commonly available networks use at least partially the idea of OSI layers. Figure 4.5 shows an example of the diverse practices of supporting protocols in OSI layers in commercial telecommunications systems.

As an example, the 3GPP standardization has taken the OSI layers as a basis for the GSM and UMTS networks as for the lower layers, which are OSI layer 1 (physical), 2 (data link) and 3 (network). The rest of the layers are a kind of fusion of the original and practical solutions, the application layer being on top for the end-user communications. One example of the fusion is the leaving out of originally defined layer 6 (presentation) that theoretically includes also the encryption. In GSM and UMTS, the encryption has been included as a part of the functionality of lower layers.

Table 4.1 summarizes some of the typical solutions that are located in different OSI protocol layers.

	OSI model	OSI protocol suite
7	Application layer	ACSE, ROSE, RTSE, CCRSE
6	Presentation layer	Presentation service/protocol
5	Session layer	Session service/protocol
4	Transport layer	TP0, TP1, TP2, TP3, TP4
3	Network layer	IS-IS, CONP/CMNS, ES-IS, CLNP/CLNS
2	Data link layer	IEEE 802.2, 802.3, 803.5 (Token Ring), FDDI, X.25
1	Physical layer	HW for IEEE 802.3, Token Ring, FDDI, X.25, cellular systems, other packet data networks

Figure 4.5 *A summary of some typical protocols in different OSI layers [3].*

Table 4.1 *Typical examples of the solutions on different OSI layers*

OSI layer	Solutions
1: Physical	Copper wire, coaxial, fiber optics cable, radio link
2: Data link	Bits packed into frames
3: Network	Routers
4: Transport	TCP, UDP, SPX
5: Session	NFS, SQL, RPC, NetBIOS
6: Presentation	MIDI, JPEG, MPEG, ASCII, GIF, TIFF
7: Application	FTP, TFTP, Telnet, SNMP, BOOTP, SMTP, MIME

4.3 Fixed Networks

The fixed telecommunications networks consist typically of three parts, or planes:

- The data or user plane, that is, bearer plane, which is meant to carry the traffic of the terminals of the end-users. The data can be of any information, including voice, files, audio, video and other type of contents.
- The control plane, which is meant to carry control information, that is, signaling. This signaling is meant to control data transmission between network elements and user terminals. Depending on the protocol layer, the signaling has different functions. It can, for example, observe that the communications link is valid, and triggers a supervision alarm if something goes wrong. Other functions of the control plane are flow control, admission control, coordination of the security signaling, and compression.
- The management plane, which is meant to carry the operations and administration traffic utilized for network management. The management layer typically contains functionalities to monitor the network functionality and performance, and communicates between the network elements in the case of, for example, parameter changes or software updates remotely. For fault management, the management system can identify the root cause of the alarms, make further analysis and try to repair the problem remotely.

Figure 4.6 presents the overall idea of these three communication links. Network elements route the signaling and data between terminals according to the user and signaling plane protocol layers, and the management system may be connected to a part or all of the network elements.

4.3.1 SS7

The “traditional” way of doing international telephony network signaling is via the SS7 (Common Channel signaling system no. 7). It can also be called CCS7, SS #7 or C7.

SS7 is defined by CCITT for international signaling of telephony networks. One of the benefits of SS7 has been that the signaling route is independent of the related call, that is, the user and signaling planes can be routed via completely different paths (cables) within the network.

The original SS7 contains two parts: MTP (message transfer part) and TUP (telephone user part). The latter has been defined for the analog telephone network, and also national variants have been defined from that (NUP, national user part). As the role of the analog telephone networks has significantly reduced along with the modernization of the networks via digital systems, there is an extension of the signaling, ISUP (ISDN user part).

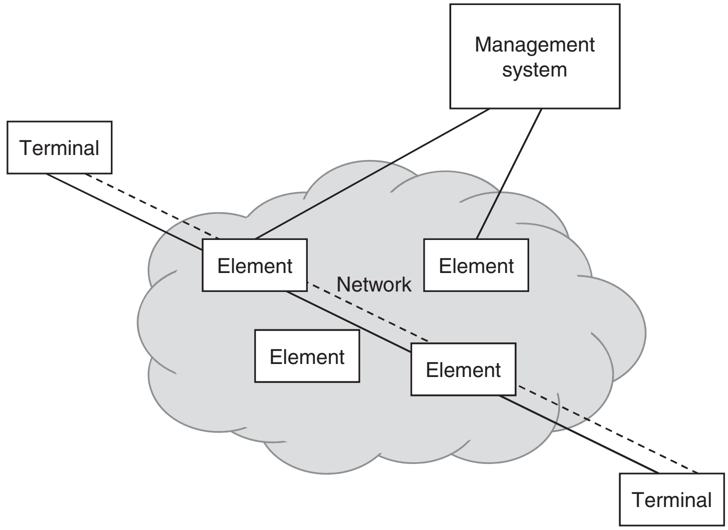

Figure 4.6 The basic principle of the user, signaling and management planes.

4.3.1.1 MTP

The physical layer takes care of the signaling between network elements when the call is being established. MTP can be divided into three layers according to the OSI principles, that is, into layers 1–3. The first layer defines physical and electric characteristics.

The physical layer defines the physical and electronic characteristics. The signaling messages are transferred, for example, in Europe via the 64 kb/s PCM blocks. The second layer is meant for the data link control, which takes care of the errorless transfer of signaling messages between the network elements. In other words, the tasks of this layer include error recovery and the separation of messages. The third protocol stage above the data link layer is the network layer which is meant for the delivery of the signaling messages between the elements. As an example, routing is one part of this protocol layer.

4.3.1.2 TUP/NUP/ISUP

TUP, NUP and ISUP are protocols designed for user part delivery. These can be used for the establishment, maintaining and ending of the calls. The protocol layer structure of the signaling is presented in Figure 4.7.

In fixed network signaling, the TUP/NUP/ISUP protocols are utilized as well as the underlying MTP protocol stack, the same principle has been adopted by mobile communications.

4.3.2 SIGTRAN

The transfer of information consists of the message and related signaling, or control part. The practical means of how these two parts are treated within the network depends on the system. In the earlier days of telecommunications, the information and control part were tight together and were sent to the receiving end physically via the same resources. Some examples of this principle are the first fixed telephone networks, and first, analog mobile telecommunications networks. A concrete example of the latter is NMT, which had a frequency modulated voice channel that was put on hold whilst the related signaling, for example, due to the handover procedure, was performed.

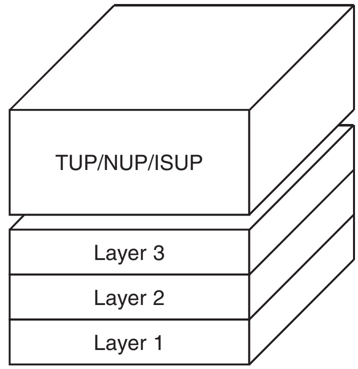

Figure 4.7 *The fixed telephony network's SS7 consists of MTP and TUP/NUP/ISUP layers.*

Nowadays, modern telecommunications systems are based on independent physical resources for the delivery of information and related control data. This separate signaling can, logically, be delivered also within the same physical level, for example, via radio link or optical fiber, but there is no dependency on these two which provides, for example, a fluent continuum of the voice call even during the handover procedure.

SS7 has been useful for a long time for circuit switched telephony networks. Nevertheless, as the development has taken a major leap towards all IP concepts for both fixed and wireless networks, the basic form of SS7 is not optimal in the packet-based environment.

Various intentions were thus made, and many proposals for the solution have been presented, including TIPHON (Telecommunications and IP Harmonization on Networks). One of the major solutions for updating the signaling in the packet switched networks has been developed by IETF under the name SIGTRAN (Signaling Transport). It has been created in the working group with the same name, and it defines a set of protocols that are able to make the SS7 of the public telephone networks transported also over the IP networks. SIGTRAN was published in IETF RFC 2719 with the name " Architectural Framework for Signaling Transport." Figure 4.8 shows the idea of SIGTRAN protocol family as a replacement of the OSI model's lower layers.

The set of protocols thus forms a new transport layer, SCTP (Stream Control Transmission Protocol) with a set of user adaptation layers that are capable of functioning like the services of the lower layers of SS7 and ISDN.

SIGTRAN can thus be considered as an evolved version of the original SS7. It defines adaptors and transport capacity by mixing protocols of SS7 and IP in such a way that it offers the essential from both technologies. The SIGTRAN applications include Dial-Up via Internet, IP telephony connected between IP and PSTN, among other services.

The components of SIGTRAN are presented in Figure 4.9. They are:

- Signaling Gateway (SG). This takes care of the interconnection of SS7 networks, and transmission of signaling messages of the IP nodes.
- Media Gateway (MG). This takes care of the encapsulation of the voice traffic as well as the transmission of traffic towards the destination.
- Media Gateway Controller (MGC). This takes care of the call control between SG and MG. It also controls the access of IP towards PSTN.
- IP Service Control Point (IP SCP). This is an integral part of IP networks, but this makes it possible to map it with SS7 networks.
- IP phone. This is the actual IP device, or IP terminal.

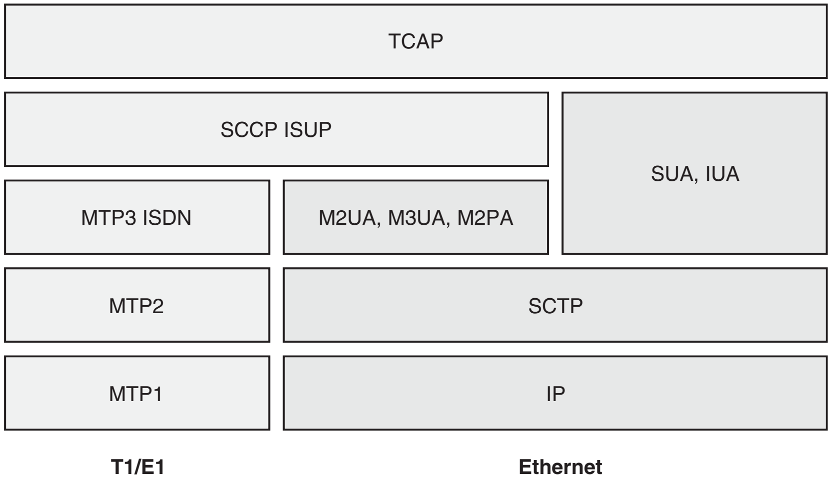

Figure 4.8 *The mapping of the SIGTRAN and T1/E1 protocol layers. SIGTRAN protocol family is designed as a replacement of lower-level OSI network type.*

The set of protocols related to SIGTRAN includes the following:

- SCTP (Stream Control Transmission Protocol). This is the main protocol of SIGTRAN, and is essentially for the transport. This protocol has been first published in RFC 2960, and was later updated in RFC 3309. SCTP is used by one of the following protocols for the layer adaptation: SUA, IUA, M3UA, M2UA, M2PA, V5UA or DPNSS/DASS2. The explication of these is in the following bullet points.
- SUA (Signaling Connection Control Part User Adaptation Layer). An example of this is TCAP. It provides SCCP services in the case of peer-to-peer, like SG-to-IP SCP. Its user is, for example, TCAP.

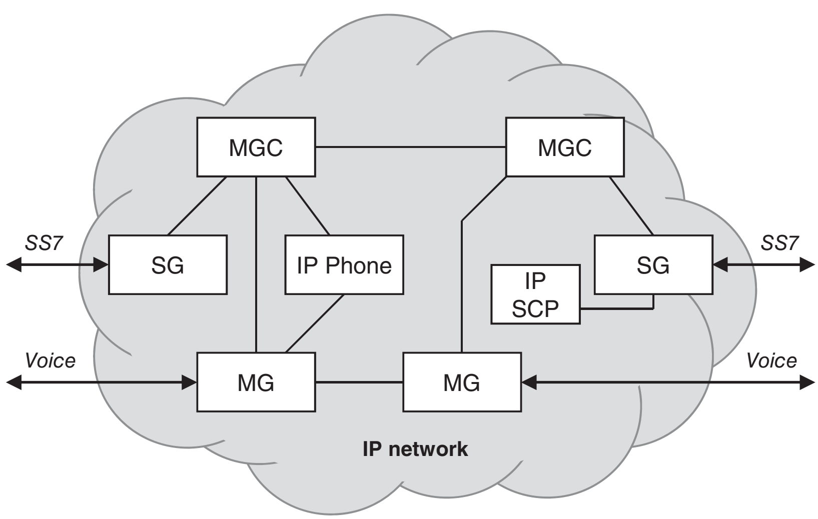

Figure 4.9 *The high-level presentation of the interfaces between IP networks and PSTN for the provision of voice service.*

- IUA (ISDN Q.921-User Adaptation Layer). This has been defined in RFC 4233 and RFC 5133. It provides LAPD (ISDN Data Link Layer) services. Its user is ISDN layer 3 entity (Q.931).
- M3UA (SS7 Message Transfer Part 3 (MTP3) User Adaptation layer). An example is ISUP and SCCP. This has been defined in RFC 4666. It provides MTP3 services in the cases of client-server and peer-to-peer. Its users are SCCP and ISUP.
- M2UA (SS7 Message Transfer Part 2 (MTP2) User Adaptation layer). This has been defined in RFC 3231. It provides MTP2 services in a setup consisting client and server, including SG (server) to MGC (client). The user of this layer is MTP3. M2UA transfers MTP2 user data SG (MTP2 instance) and MGC (MTP3 instance), and M2UA makes it possible for MGC to provide MTP2 service.
- M2PA (MTP2 Peer-to-peer user Adaptation layer). This has been defined in RFC 4165. It provides MTP2 services in a case of peer-to-peer, including SG-to-SG. Its user is MTP3.
- V5UA (V5.2-User Adaptation Layer). This has been defined in RFC 3807. It provides V.5.2 protocol services.
- DPNSS/DASS2 User Adaptation (DUA). This has been defined in RFC 4129.

These protocols are basically utilized in order to replace the original SS7 stacks described in Section 4.3.1 for the transport of the contents in the IP environment. This provides flexible routing of signaling in IP networks. The typical use cases of SIGTRAN are related to VoIP, and in general, multimedia over IP, including video and music.

Figure 4.10 shows the protocol layer structure of SIGTRAN. In the figure, the red elements are new, along with the definition of SIGTRAN, whilst the other elements have already been defined earlier. In addition to the elements shown in the figure, there is also ISUP for the ISDN user part. It would typically be located over MTP3 or M3UA.

The adaptation layers of SIGTRAN serve, for example, the following situations:

- They carry upper layer signaling protocols over IP transport, which is assumed to be reliable.
- They are transparent for the users and services, and thus provide the same services as is the case in PSTN.

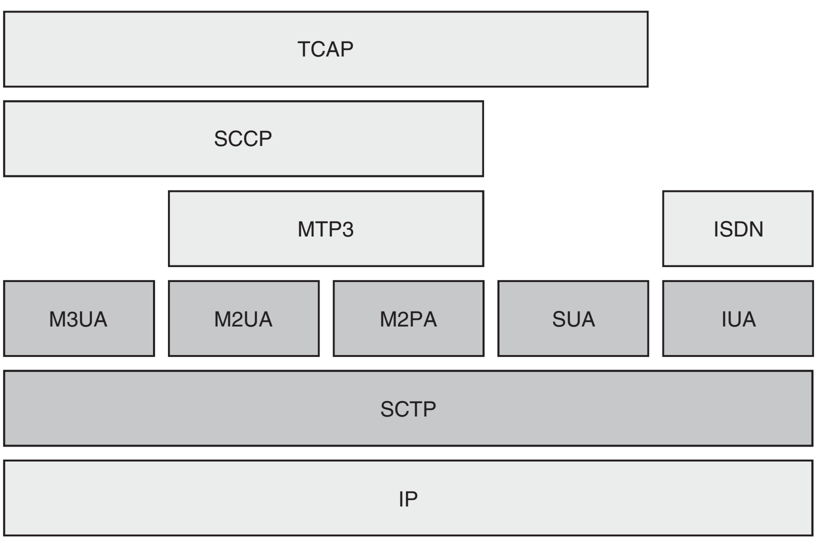

Figure 4.10 *The protocol suite structure of SIGTRAN.*

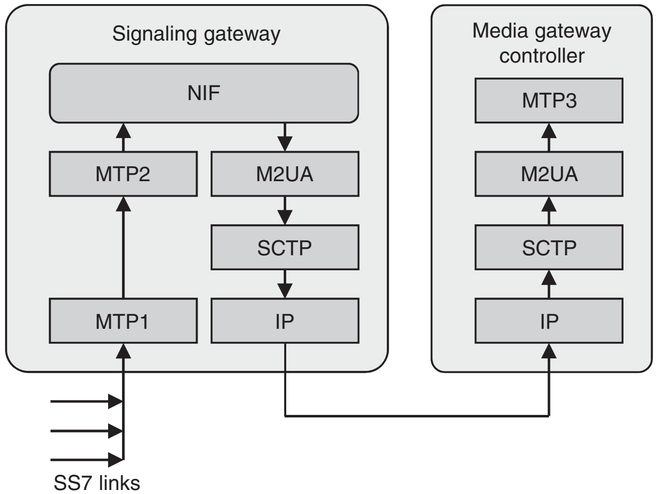

Figure 4.11 *The architecture of M2UA.*

4.3.2.1 M2UA

M2UA provides a link between remotely located MTP2 and MTP3 instances. In the case of M2UA, one user is MTP3 and the other is an SG IWF. M2UA functions according to a client-server model between SG (having MTP2 instance) and MGC (having MTP3 instance). In this model, MGC represents the client and SG the server. Figure 4.11 presents the architectural model of M2UA.

The functionality of M2UA is applicable, for example, in the case where there is a low density of SS7 links at a certain point of network, or if there are a large number of SG functions that are separate.

This SIGTRAN configuration is common in European networks as the SS7 links and voice circuits are typically shared physically.

4.3.2.2 M2PA

M2PA is similar to M2UA, but is meant for the peer-to-peer cases, and replaces the MTP2 link under MTP3. In the case of M2PA, the user is MTP3 at both ends of the connection. M2PA makes it possible for MTP3 layers in SGs to communicate directly.

M2PA allows communications between SS7 systems over IP instead of the use of TDM links based on T1 or E1. The practical benefit of M2PA is that it can be used instead of the MTP2 link in such a way that there is no need to invest in a dedicated SS7 HW.

The architectural model of M2PA is presented in Figure 4.12. This solution is applicable in SG-SG connection, and is functioning in such a way that it bridges two isolated SS7 parts of the network.

4.3.2.3 M3UA

The functioning of M3UA is similar to M2UA, that is, it functions in the same way as a client-server between SG and MGC, in order to provide the upper layer SS7 with protocol remote access to lower layers.

M3UA signaling gateways run lower levels of SS7 protocol stack, that is, in MTP3 and MTP2. On the other hand, M3UA functions as application servers at higher levels, that is, in ISUP and SCCP. M3UA provides reliable and flexible system architecture.

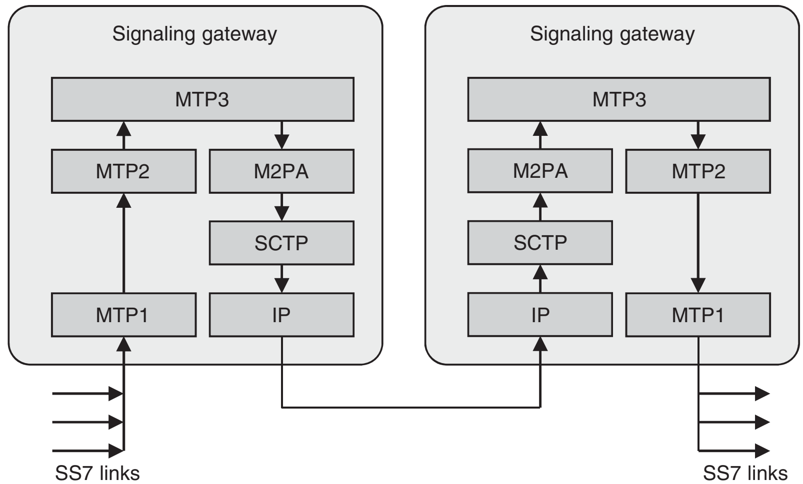

Figure 4.12 *The architectural model of M2PA.*

The architectural model of M3UA is presented in Figure 4.13.

The M3UA is suitable when there is a high density of SS7, making it possible to use a standalone SG. This architecture is applicable when SS7 links are physically accessible as a single point.

The model is thus suitable in North America as the SS7 links are physically separated from the voice circuits.

4.3.2.4 SUA

Via SUA, it is possible for Application part, like TCAP, of IP SCP to be reached via an SG. The architecture of SUA is presented in Figure 4.14.

SUA makes it possible to make the mapping between SCCP addresses and IP addresses at the SG.

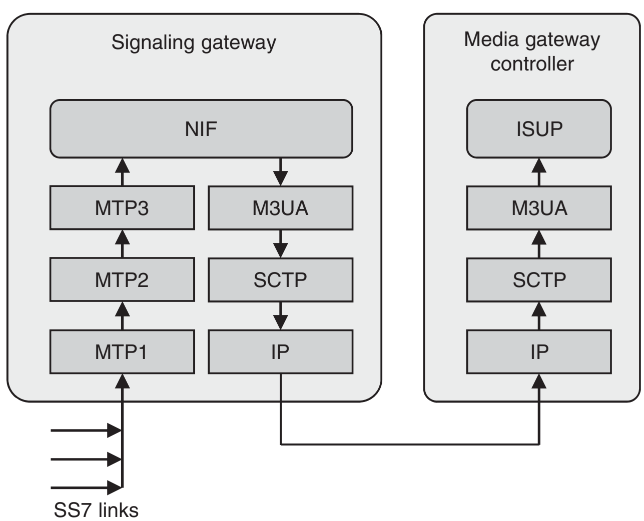

Figure 4.13 *The architectural SG/MGC model of M3UA.*

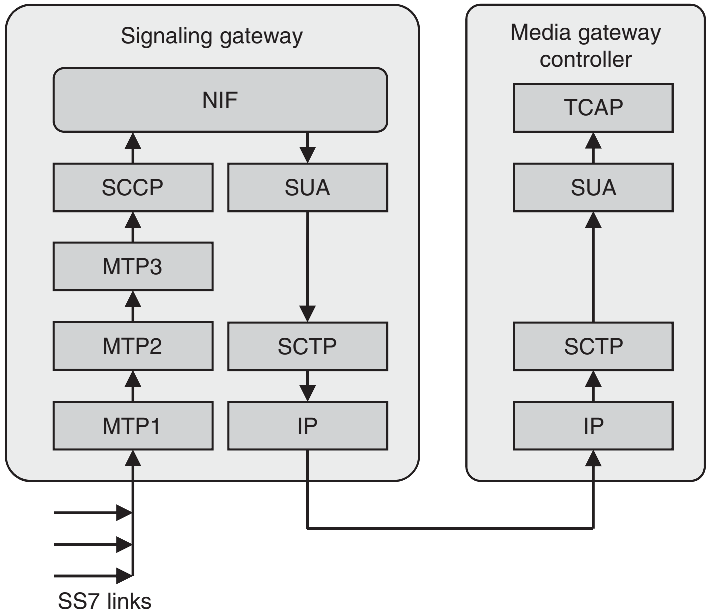

Figure 4.14 *The SUA architecture.*

4.4 Mobile Networks

4.4.1 SCCP

As the cellular networks typically contain many functionalities that are not utilized or known within fixed telephony networks, the SS7 as such is not suitable for the signaling method for the cellular networks. The basic problem arises from the fact that SS7 is not capable of handling the virtual connections, which are essential, for example, in the location update procedures of the cellular systems. On the other hand, for example, MTP is able to route the messages only within a single network, which would cause immediate problems in the national and international roaming of the cellular network users.

SCCP (signaling connection and control part) manages virtual connections, as well as connection oriented and connectionless signaling. SCCP is thus utilized in the establishment of signaling connections, as well as for the supervision of the validity of connection links in a parallel way with the basic Telephone User Part (TUP) and ISDN User Part (ISUP) of the fixed telephony network. At the moment, the fixed telephony network thus contains SS7, which is constructed by MTP, TUP/ISUP and SCCP.

4.4.2 BSSAP

For internal communications between the MSC and BSC/MS, there is a BSSAP protocol (base station subsystem application part). Because there is BSC between the complete chain of MSC and MS, there must also be a virtual connection, which requires SCCP. The SCCP is the underlying protocol, and BSSAP is located on top of that. The role of BSSAP is, among other tasks, to transport the signaling that relates to the authentication of users within the communication line of MS and MSC, as well as the signaling related to the temporal mobile subscriber identity (TMSI). In addition, there is also internal signaling between BSC and MSC elements which is handled by BSSAP.

BSSAP is divided into DTAP (direct transfer application part) and BSSMAP (BSS management application part) messages. The former is meant to the direct signaling between MSC and MS whereas the latter is used in the management of BSS.

4.4.3 MAP

MAP (mobile application part) is utilized in, for example, GSM and UMTS networks between MSC and registers, which are HLR, VLR, AuC and EIR. The communication between two MSC elements utilizes MAP only when there is other than user traffic. In case of routing of the call from one MSC to another MSC, the fixed network principles of TUP, NUP and ISUP are applied.

MAP is used, for example, for location area updates, MT call routing via the MSRN number, and delivery of short messages. The more specific type of MAP protocol is indicated by an additional letter that refers to the interface, so, for example, the protocol utilized between MSC and HLR is called MAP-C.

4.4.4 TCAP

MAP is based on a set of messages between the participating elements, containing requests, responses and related info fields. These events are managed by TCAP (transaction capabilities application part), which resides under the MAP layer.

4.4.5 LAPD/LAPDm

When information is transferred over the radio interface, GSM utilizes LAPDm protocol (link access protocol on the modified D channel) which is an enhanced version of the one utilized in ISDN networks. LAPDm includes an address field, which delivers SAPI (service access point identifier) of the user who is utilizing a certain protocol. When command or directive frames are delivered, SAPI identifies the user related to those specific frames.

LAPDm frame also includes LPD (link protocol discriminator), which is utilized for defining the use of LAPDm. Furthermore, LAPDm contains C/R (command/response), which indicates whether the LAPD frame between base station and base station controller is command or response frame. The last field of LAPDm includes EA (extended address), which is utilized for extending the LAPDm address field to contain more than one octet, if needed.

4.5 Data Networks

4.5.1 TCP/IP

TCP/IP (transfer control protocol / Internet protocol) and UDP (unstructured datagram protocol) have been de facto standards of the IP networks for long time. They are utilized for creating communications connections between different network types and both are based on the flexible delivering of the data packets. Other essential protocols, among countless sets, are the following:

- Telnet is meant for providing remote connections between the host and client machines (terminal and application). Due to the limitations of the original version of Telnet as for the security aspects, another widely utilized protocol is SSH.
- FTP (file transfer protocol) is meant for transferring files over data networks.

- SMTP (simple mail transfer protocol) is meant for delivering emails.
- SNMP (simple network management protocol) is meant for the network management.

TCP/IP is already a senior representative in the data network environment. It was developed in the beginning of 1970s as a protocol for the American ARPANET to serve the needs of academic environment and defense forces. The aim of the development was to create a method that was able to create and maintain communications links regardless of the partial destruction of the networks. All the phases of the development resulted in Internet by 1983.

It should be noted that TCP/IP cannot be compared directly with the OSI model of ISO, although the connectionless TCP is relatively close to the idea of the network layer. Thus, TCP can be related either to the OSI model's third layer, or based on its functionality, also to the fourth layer.

The functionality of TCP is based on the fragmenting of the messages, the retransmissions of the messages, and defragmenting of the messages in the receiving end, as well as prioritization and management of connections (initiation, maintaining and ending). TCP utilizes SAPs (service access point), which correctly handle the dedication of network resources between different users in the routing of the connections. TCP is a connection-oriented method, and thus provides reliable data transfer.

As a summary, TCP provides a reliable connection between the emitting and receiving side, in order to deliver the contents correctly, for example, in the application level. TCP is suitable for transferring a large amount of data in the environment which cannot guarantee the quality of service as for packet losses, or which is not able to keep the order of packets, take care of the duplicated packets and so on. The drawback of TCP is that the addition of the required information for reliable delivery of the packets increases the TCP packet size accordingly.

4.5.2 UDP

User Datagram Protocol (UDP) is a transport layer protocol, and is thus working in the OSI model 4. It is based on the exchange of datagrams. UDP allows the transfer of datagrams over telecommunications networks without prior establishment of connection. This is thanks to the contents of the datagram which already contains sufficient amount of information within its heading fields. Furthermore, UDP does not have flow control, which means that the order of the transmitted and received packets can change. In addition, the sender does not have any means to ensure that the UDP packet has arrived to the receiving end correctly because UDP does not include mechanisms for the confirmation.

One of the typical use cases for UDP is video and audio streaming in real time. Although there would be errors at the receiving end, in order to maintain the real time data flow, it would not be feasible to have resending of the contents in any case. UDP is thus assuming that the network itself provides sufficient capacity and quality for the streams, and for the detection and error correction, a higher level mechanism can be adopted that could include, for example FEC.

UDP is thus a protocol that contains minimum amount of complexity between the network and application protocol levels. UDP protocol is especially useful for the situations when the bit rate is more important than the guarantee for quality of service, for example, in the case of audio and video. UDP has been defined in IETF RFC 768.

The header of UDP is structured in such a way that it has a total of 4 fields, from which 2 are left optional. This high-level principle of UDP datagram is presented in Figure 4.15.

The fields for the origin port and destiny port consist of 16 bits as shown in Figure 4.16. It should be noted that, as UDP is not based on a system containing a separate server that has the status of sending, nor is the origin asking any acknowledgments; the port for the source is left optional (in which case the port should be marked as zero). The range of the UDP ports can be 0-65,535, according to the 16 bit presentation of

+	Bits 0…15	Bits 16…31
0	Originating port	Destiny port
32	Length of the message	Check sum
64	Data	

Figure 4.15 The fields of UDP datagram.

these fields. The port 0 has been reserved although it has not been restricted either if no acknowledgments are expected from the receiving end. The typical port number is from the range of 1-1023 (well-known ports) whereas the ports 1,024-49,151 are registered. The rest of the ports are for temporal use.

After the destiny port, there is an obligatory field for indicating the size of the data part of the UDP datagram, marked in bytes. The minimum value for this field is 8 bytes. Then, there is a field for checksum, with the size of 16 bits. This field contains the result of the checksum for the origin and destiny IP address, protocol, and size of the UDP packet (forming a pseudo-IP heading), the UDP heading, the data, and possible filling of 0-bits until the check field is an integer multiple of 16. The checksum is left optional for IPv4, but for IPv6 it is mandatory. In practice, the checksum is also typically utilized in the case of IPv4.

As a summary, UDP is suitable for the network environment that provides sufficiently high quality for data transmission. As UDP does contain much lower overhead compared to, for example, TCP (8 bytes vs. 20

+	Bits 0…7	Bits 8…15	Bits 16…23	Bits 24…31
0	Originating address			
32	Destiny address			
64	Zeros	Protocol	UDP length	
96	Originating port		Destiny port	
128	Message length		Check sum	
160	Data			

Figure 4.16 The more detailed presentation of UDP datagram.

bytes), it also provides higher data transmission speed. Another benefit of UDP is that it does not add delay in the initiation of the data connection. As a result, a server of a certain application can serve more active clients via UDP compared to TCP.

4.6 Error Recovery

When talking about transparent data transmission, the data is pushed to the system without additional or special error recovery mechanism on behalf of the network, except the ones that are involved in the basic internal data delivery. The transparent data transmission tends to guarantee the bit rate, even if the remaining bit error rate grows. The occurred errors should thus be corrected, for example, in the application layer.

The nontransparent data transmission, on the other hand, tries to recover the occurred bit errors already within the system. When the bit error rate grows, the data throughput lowers accordingly. This method eases the task of the application layer in error recovery.

In any case, one of the essential functions of telecommunications is to assure that the occurred errors can be detected and corrected as efficiently as possible, regardless of error recovery within the underlying protocol layers, or in the application layer.

The errors can occur for multiple reasons. As a result, the received bit stream might contain erroneous bits, or bits can be missing from the messages. The bits may also be duplicated during transmission, or the order of the bits can vary.

4.6.1 Message

A message is constructed by the single bits that forms data blocks, as shown in Figure 4.17. The terminology varies greatly depending on the environment. In case of radio path, radio blocks or radio data blocks are formed, whereas in the fixed environment, data blocks are formed. When this block of bits is received in such a way that the remaining errors exceed the capabilities of the error correction mechanism, we are talking about Block Error Rate (BLER). The error correction can be typically two-folded, meaning that the first phase detects the errors, that is, the existence of the BLER occurrences. The possible second phase tries to further correct the errors, still leaving a certain amount of residual Block Errors in the received bit stream.

The basic principle of the message is that it is first created by a sending terminal. The delivery of the message is done via a network of links and nodes until it reaches the destination terminal. During this procedure, the intermediate nodes take care of the message by routing it via adequate links that lead it to the destination.

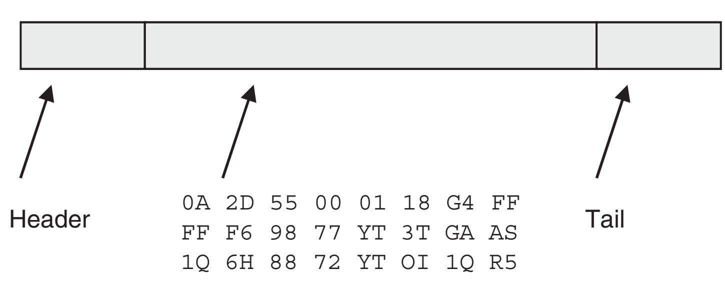

Figure 4.17 An example of a message.

The message is divided into two parts: control and bearer. It should be noted that these parts can be delivered either separately or at the same message. The content of the message resides in the bearer part. This content can be basically anything that the user wants to send, including, for example, audio, video, files, email messages, multimedia, and voice. The control part is meant for guiding the participating nodes for routing the message, by indicating the destination and possible means of delivery (instructions about, e.g., prioritization). The protocols are utilized to handle the delivery of the message.

4.6.2 Error Correction Methods

In order to recover the messages, the erroneous bits must be detected. The simplest way to do this is via parity sum check methodology. The parity check has been utilized since the earliest modem solutions, and is still useful due to its noncomplicated nature. In practice, parity sum check methodology has developed since the early solutions but the basic principle remains the same.

In the parity sum check, additional bits are located to the bit stream prior to transmission to describe the parity, or oddity, of a certain amount of consecutive bits. The indication of parity or oddity should be agreed in advance by the transmitting and receiving end. The parity or oddity can be calculated, for example, for each single bytes, or for a complete matrix that takes into account a set of bytes at the same time.

Neither one of these variants can result in total confidence level for the detection of errors, for example, in the case that two bits are erroneous within the observed set of bits. The matrix form enhances reliability because the sum is checked in both vertical and horizontal orientations (rows and columns), but, if the error happens to occur at the edges of the square shaped area within the matrix, the result is again the same as in the correct form. This means that these fairly basic methods are useful only in very simple transmission cases, and the more demanding transmission requires more advanced methods.

The following example presents the idea of the calculation for the parity bits in the matrix form as interpreted from Ref. [9]. Let us select an original bit stream that is to be transmitted, presented by the following symbols: 10110011, 10101010, 00110011, and 01000111 in a matrix of 4 symbols. In the transmission end, we can now calculate the parity bit for each symbol separately. It can be done by observing the number of 1-bits of the symbols for the rows of the matrix in both vertical direction (columns) and horizontal direction (rows). We can decide that if the number is even, the parity bit will be 1, and if it is odd, the parity bit is 0. Table 4.2 shows the idea of this method.

It can be observed from Table 4.2 that the original contents of the stream of 4 symbols are 4×8 bits = 32 bits. The number of additional bits due to the parity check is $2 \times 8 - 1 = 15$, increasing the number of total bits to 47. This means that the overhead caused by the protection is 15/47 = 31.9%. The positive effect of the protection is that the errors that have occurred can be detected, and in this case, retransmission can be requested if an erroneous packet is received. The negative impact is that the overhead reduces the throughput of the data transmission by 31.9%, that is, the data transmission rate of the bits that the user is sending. If the

Table 4.2 *An example of parity bit creation in a matrix of 4 rows*

	Bit 1	Bit 2	Bit 3	Bit 4	Bit 5	Bit 6	Bit 7	Bit 8	Parity
Bit stream 1:	1	0	1	1	0	0	1	1	**0**
Bit stream 2:	1	0	1	0	1	0	1	0	**1**
Bit stream 3:	0	0	1	1	0	0	1	1	**1**
Bit stream 4:	0	1	0	0	0	1	1	1	**1**
Parity:	**1**	**0**	**0**	**1**	**0**	**0**	**1**	**0**	**0**

offered speed of the network is, for example, 1 Mb/s, the actual data rate that the user sees (assuming only this parity check is applied) would thus be about 2/3 of the total bit stream of the channel.

A more advanced method, still widely utilized in the telecommunications environment, is CRC (cyclic redundancy check). It has various versions, although the basic idea is the same. The base for this method is XOR operation, that is, modulo 2. The rule for this is the following, when two bits are summed: 0 + 0 = 1, 0 + 1 = 1, 1 + 0 = 1, and 1 + 1 = 0. Constant division is applied in this method, with a length of n + 1 bits, where n refers to the length of the check bits. This means that the divider should be at least 2 bits long, with the last bit set to 1.

The divider can be expressed in the form of $x^n + x^{n-1} + x^{n-2} + \ldots + x^2 + x + 1$, where x^n is the most significant bit and each term of x is either 1 or 0. As an example, $x^3 + x + 1$ refers to the stream of 1011 presented in the bit format.

Let us create a CRC protection to the message that consists of letters "OK" when the divider is $x^5 + x^3 + 1$, referring to the bit stream of 101001. Assuming we are using ASCII, the letter "O" can be presented in 8-bit format as 01001111, and "K" can be presented as 01001011. In the CRC method, the bits of the message are combined, and a stream of pure zeros with length of divider are added to the end of the stream. In this example, the additional part is thus "00000", resulting in the following stream:

010011110100101100000

Next, let us identify the first 1-bit of the stream, so that we can start dividing as of that position with the constant divider in the following way:

```
010011110100101100000 (original message with additional part)
 101001               (divider)
-----------------------------------
 00111010100101100000 (sum of the bits via XOR resulting in first round result)
```

Continuing with this principle, selecting again the next 1-bit position, we can obtain the following:

```
00111010100101100000   (previous result)
  101001               (divider)
---------------------------------
  010011100101100000   (second round result)
   101001              (divider)
---------------------------------
   00111000101100000   (3rd round result)
     101001            (divider)
-------------------------
     010001101100000
      101001
------------------------
      00101001100000
        101001
----------------------
        000000100000
              101001
          ------------
              001001 (last result)
                1001 (remaining part, i.e., FCS)
```

The remaining part of this method of the CRC method is FCS (frame check sequence). It is transmitted, during the calculation, instead of the 0-bits that were used only during the calculation, now placed at the end of the bit stream. The complete bit stream is transmitted over the network according to the protocol functionality (which adds part of the headers and tails during the internal transmission process, and removes those in the

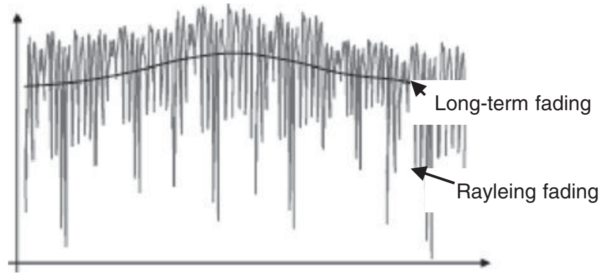

Figure 4.18 *The principle of fast fading, that is, Rayleigh distributed fading [5]. Figure by Jyrki Penttinen.*

respective elements in the receiving end). When received at the other side of the network, the receiving end performs a reversed calculation of the stream by applying the same XOR logics, but now utilizing the FCS bit stream instead of the 0-sequence. If any errors have occurred during the transmission, the remaining part is now something other than 5 zeros.

TCP/IP utilizes an error check method that is based on the 1 complement procedure, because it is easier and faster to perform than that presented in the example above. Although the processing load would not be an issue any more, it was optimized in the early days of TCP/IP. The drawback of that variant is that it does not detect the errors as efficiently as CRC does. In case of TCP/IP, the error check calculation is performed in blocks of 16 bits.

There are countless reasons for errors, for which the radio interface is one of the most probable root-causes for the misbehavior of the bits. The radio path may contain multipath-propagated components which result in a set of received radio transmissions with varying amplitudes and phases. These cause variations in the received power levels of the signals. The more reflections there are, closer to the fast fading, or Rayleigh fading profile the resulting signal is. The Rayleigh fading basically means that the received signal includes significantly attenuated peaks that tend to repeat every half-a-wavelength in the space domain as shown in Figure 4.18. The Rayleigh fading is by nature most common in environments where many reflections happen due to heavy obstacles, such as dense city centers.

The other variants of the fading include, for example, Rice-distributed fading that is basically a combination of slow and fast fading components. The Rice-distribution is most typical in suburban types of environments.

When radio propagation happens more clearly in rural and open areas, the distribution approaches the additional white Gaussian noise (AWGN). This is basically a flat fading model which does not contain strong attenuation peaks as the previous ones do.

The radio interface is not the only cause for bit errors, but is by far more significant in local distances. Fixed lines cause problems in delivery and reception of bits, although the rate tends to be considerably lower compared to radio propagated signals. Nevertheless, the longer the distance, the more noise level increases as well as bit error rate of the cable. For coaxial and other metallic cables, the root cause is based on electrical characteristics and distortions in voltage, whereas optical fiber starts containing problems due to "optical distortions."

4.7 LAP Protocol Family

For the transport layer of the OSI model, many methods have been developed. A commonly utilized method is developed by CCITT, and is named LAP (link access procedure). It also has many further variants. LAP is based on HDLC protocol (high-level data link control) of ISO. The following sections list some of the most common variants of LAP.

4.7.1 LAPB

LAPB (LAP balanced) is defined as a basis for the X.25 packet networks, and is a subset of HDLC. The functioning of LAPB is based on a master–slave type of transmission. In this model, the master gives commands for the slave, which in turn gives answers to the master. When the transmitting and receiving ends of the network contains both master and slave functionalities, the system is balanced. There are only 9 commands in total that LAPB utilizes, of which 3 are for supervision, and the rest are not numbered.

The LAPB definitions contain supervision, information and unnumbered messages. The role of supervision messages is to take care of flow control and error recovery. The method includes response procedures for the erroneously received message (reject), acknowledgment of the correctly received message (receiver ready) and information about overloading (receiver not ready). The not numbered commands are: (1) Set Asynchronous Balanced Mode (SABM); (2) Set Extended Asynchronous Balanced Mode (SABME), Disconnect (DISC); (3) Unnumbered Acknowledgement (UA); (4) Disconnect Mode (DM); and (5) Frame Reject (FRMR).

The information messages are utilized along the actual messages in order to inform about the order number of the next messages. Because this information is embedded into the actual message, no need for a separate capacity is needed for delivering this information in the network.

The unnumbered messages of LAPB protocol are related to the methods of transferring the values of parameters, procedures for ending the transfer of the data, and testing.

4.7.2 LAPD

LAPD (link access protocol in D channel) protocol has been developed for the needs of ISDN network. It is based on the definitions of LAPB, and is defined in ITU Q.921. The main differences between LAPB and LAPD protocols are related to the address fields and delivery of the identification of the control information.

LAPD is capable of multiplexing several connections into a single physical channel between users. The main difference with the LAPB is that LAPD does frame segmentation and defragmentation.

LAPD works in the link layer, that is, in the second layer of OSI model and thus serves the network layer. The frame delivery can be done in a confirmed or nonconfirmed way.

The nonconfirmed way utilizes the transfer of information in the network layer in nonnumbered frames. This results in a lack of error control as such, except the detection of errors which can reject erroneous frames. The benefits of this way include the following: (1) it is suitable for point-to-point transmission as well as for broadcast; (2) the transmission is fast; (3) it is especially suitable for network management functions.

Some of the disadvantages are the following: (1) there is really no guarantee for receiving the frame; (2) the transferring end cannot detect potential errors in the transmission of the information; and (3) there is neither flow control nor error control.

The confirmed mode requires establishment of logical connection between users. This is done for the whole lifetime of the connection, from establishing the initial connection, transferring of data, and terminating the connection. During this connection, the information is located to numbered frames and the reception of the frame correctly or in a faulty way can thus be detected accordingly.

The benefits of the confirmed mode include the guarantee of the reception of the data, and the inclusion of error control and flow control.

4.7.3 LAPF

LAPF (link access procedure for frame mode bearer services) is a protocol developed for Frame Relay networks. It is a variant of LAPB in such a way that LAPF is in fact simplified further from LAPB. The reason for this is that Frame Relay networks have been assumed to cause fewer errors in data transfer, and thus LAPF protocol does not need the error recovery procedures that are applied in X.25 networks.

There are two variants of LAPF: LAPF core and LAPF control. LAPF core provides a minimum amount of functions for controlling of the data link. It is included by default for all Frame Relay implementations. LAPF control, on the other hand, includes error control and flow control and can be selected by the user to provide more reliable, but dedicated data links.

4.7.4 LAPM

LAPM (link access procedure for modems) is a part of ITU-T V.42 error recovery specification. LAPM is largely similar with LAPD protocol, but there are some minor additions. These additions provide a more efficient way to execute the retransmissions because the LAPD retransmission happens as of the first frame that contains erroneous bits, whereas LAPM handles the retransmissions more selectively.

4.8 Cross-Layer Protocol Principles

Source Ref. [4] describes a set of cross-layer designs. The cross-layered approach means that there is a direct but controlled violation to the principles of layered protocol architectures like OSI. This violation can be done, for example, by allowing communications between nonadjacent layers, or by sharing variables between layers. There is also a possibility of deploying protocols that span more that one layer, that is, they do not communicate necessarily only with adjacent protocol layers as has been defined in the OSI model.

Some needs for the cross-layer adaptation would be, for example, to provide enhancements to the performance which is most logical in a wireless environment due to the sometimes highly unpredictable characteristics of the radio interface. If there are, for example, typical short-term attenuations in the received signal level due to Rayleigh fading in urban environment, the original TCP protocol assumes that the resulting packet loss rate indicates, instead, network congestion, and the protocol has been designed in such a way that it will automatically slow down the transmitting data rate. As the radio interface often experiences these high attenuation peaks that are very short in the time domain, the slowing down of the data rate is in fact sometimes a wrong measure that does not necessarily enhance the performance but actually makes it worse, if there is no blocking due to simultaneous data transmission of various users. The slowing down has logically been the most adequate means to act as the quality of the transmission wire can be assumed to be constant, and the basic reason for packet losses is basically a congestion.

As radio conditions vary in time and space from the wires environment, including slow and fast fading for slow and fast moving terminals, the radio interface would benefit from enhanced protocol functionalities designed especially for mobile communications.

There are different possibilities for designing protocol architecture for the cross-layered concept, as presented in Ref. [4]. One is to allow upward information flow by jumping over part of the protocol layers between the destination protocol layers. This idea provides notifications from lower layers to TCP in order to better and faster inform, for example, the congestion and high bit error rate directly to the adequate layer. Another solution is a backwards information flow in a similar manner but from notifications from higher layers to lower layers bypassing part of the layers in between. In this way, for example, the application can inform link layer about the more specific requirements for prioritizing purposes. These two approaches can further be combined so that there is collaboration, for example, between link layer and physical layer, again bypassing the layer in between, for making the collision resolution more realistically.

Another solution would be to add new capabilities of the layers, and so-called vertical calibration, which makes it possible to adjust parameter values jointly across the layers in such a way that the communication goes directly from various layers to other layers.

The practical implementation of the cross-layered concept may be such that there is direct communication between certain layers that are not necessarily neighboring layers. Some examples of the solutions are the utilization of protocol headers and internal packets. Another approach may be a common database to which each participating protocol layer can connect directly. In practice, this solution that requires an interface between each layer and the database is rather challenging because the database structure needs to be planned carefully.

There is also a possibility to quit the structured approach completely and instead make communications between protocols cross all the layers rather via a protocol heap than via neighboring protocols, that is, all the protocols would be capable of communicating with all the others. This approach requires new abstractions which may be still more challenging than in the previous one due to potential side-effects that are not known when parameter values are tuned between layers – there might be totally opposite effects elsewhere that depend directly indirectly on the tuned parameter values. Nevertheless, the flexibility of this solution is better than in any previous approaches. Also the maintenance of this mesh approach is more challenging than in a clearly structured OSI model. For more detailed information about the commonly used protocols in telecommunications systems, please refer to [10–18]. Specifically for DVB-H system, please refer to [19].

References

1. The X.200 recommendations of ITU. http://www.itu.int/rec/T-REC-X/en (last accessed 23 July 2014).
2. Layered protocol architectures. http://users.encs.concordia.ca/~glitho/F09_Chap1_Elec6861.pdf (last accessed 23 July 2014).
3. Penttinen, Jyrki (2011) The planning and optimisation of DVB-H radio networks. Doctoral dissertation, Aalto University. Espoo, Finland.
4. Cross-layered protocol architectures. http://users.encs.concordia.ca/~glitho/F09_Chap2_Elec6861.pdf (last accessed 23 July 2014).
5. IETF RFC 2960. Stream Control Transmission Protocol.
6. IETF RFC 2719. Framework Architecture for Signaling Transport.
7. SS7 MTP3. User Adaptation Layer.
8. SS7 MTP2. User Adaptation Layer.
9. Penttinen, Jyrki. Tietoliikennetekniikka (Telecommunications), part 1. Published in Finnish.
10. SS7 MTP2. User Peer-to-Peer Adaptation Layer.
11. SS7 SCCP. User Adaptation Layer.
12. IETF RFC 3057. ISDN Q.921 User Adaptation Layer.
13. V.5.2. User Adaptation Layer.
14. IETF: http://www.ietf.org (last accessed 23 July 2014).
15. Other protocols. http://users.encs.concordia.ca/~glitho/F09_Chapter6_Elec6861.pdf (last accessed 23 July 2014).
16. TCP/UDP Basics. http://users.encs.concordia.ca/~glitho/F10_Chapter4_Elec6861.pdf (last accessed 23 July 2014).
17. Wireless TCP. http://users.encs.concordia.ca/~glitho/F09_Chap5_Elec6861.pdf (last accessed 23 July 2014).
18. SIP session signaling. http://users.encs.concordia.ca/~glitho/F10_Chapter4_INSE7110.pdf (last accessed 23 July 2014).
19. Examples of ISO OSI model applications. http://www.gocertify.com/quizzes/osi/osi-1/Question-6.html (last accessed 7 April 2013).

5
Connectivity and Payment

Jyrki T. J. Penttinen

5.1 Connectivity

Connectivity in telecommunications refers basically to all the means and methods that can be used for transferring signals and/or contents between two points, including wired and wireless environments. The widest definition thus includes, for example, radio and fixed interfaces of cellular, transport and core networks for end-to-end connectivity, whilst more limited environments include near-distance connectivity of, for example, Bluetooth transmission.

This chapter concentrates on the latter environment as the relevant telecommunications systems are described elsewhere in this book. More specifically, near-distance connectivity of the following technologies is explained:

- Fixed connectivity, including USB (base solution and enhanced variants) and serial port (significance disappearing in current solutions).
- Wireless connectivity, including Bluetooth and NFC.

Furthermore, as one essential part of the NFC, mobile payment principles are also discussed.

5.2 Definitions

The usability and quality of the near-distance connectivity depends on the maximum data rates with a still acceptably low bit error rate. As transmission technologies are evolving, ever higher data rates can be offered over more developed solutions.

Data rates are typically informed in bits per second, which are marked in this book as b/s. The velocities that are relevant now and could be relevant in the near future are shown in Table 5.1.

The Telecommunications Handbook: Engineering Guidelines for Fixed, Mobile and Satellite Systems, First Edition.
Edited by Jyrki T. J. Penttinen.

Table 5.1 The relationship between main data rate categories

Unit	Symbol	Relation	Bytes per second
Kilobit per second	kb/s	1024 b/s	125 B/s
Megabit per second	Mb/s	1024 kb/s	125 kB/s
Gigabit per second	Gb/s	1024 Mb/s	125 MB/s
Terabit per second	Tb/s	1024 Gb/s	125 GB/s
Petabit per second	Pb/s	1024 Tb/s	125 TB/s

For the data rates, it should be noted that the relationship between bits per second (b/s) and bytes per second (B/s) depends on the system, that is, how many bits are used to describe bits. In a typical case, ASCII 8-bit format (American Standard Code for Information Interchange) is used which means that there is a relationship as shown in Table 5.1.

It should also be noted that the data rate depends on the ability of the system to include certain amount of bits per symbol. In telecommunications and in electronics in general, baud (Bd) is utilized for informing for symbol rate, that is, the value of symbols per second. It is used, for example, for the modulation rate, each modulation method having abilities to carry certain amount of bits per symbol. As an example, when the decision point of a certain modulation scheme is changed, it triggers a signaling event. If each symbol represents 3 bits, the actual data rate in bits per second is actually threefold compared to the symbol rate making it respectively more efficient than lower level modulation schemes. As a general rule of thumb, as the efficiency of the modulation increases, it is also more sensitive to bit errors due to tighter decision areas of the modulation scheme, which means that in practice these modulation schemes require better signal-to-noise ratio (SNR or carrier per noise, C/N) and signal-to-interference (SIR or carrier per interference, C/I) levels.

It should also be noted that although the baud rate is related to the gross bit rate in b/s, it should not be confused with it. The modem manufacturers typically generalize this, or make a distinction by applying an alternative term of characters per second (CPS). The latter can be obtained by dividing the baud rate by the number of bits per character which results in the number of characters per second. As a summary, in order to avoid any confusion with the real value of Baud, the respective case should be investigated individually.

5.3 IP Connectivity

Connectivity in telecommunications refers increasingly to the Internet or in general to IP connectivity. Connectivity in this sense thus means the system which is utilized in the subnetwork in order to connect to the packet data networks.

We can distinguish connectivity of the local area networks (LAN) to the "old-fashioned" dial-up access via modems, and broadband access via IP-based methods [1].

Dial-up networking is losing its relevance heavily as broadband IP access methods are increasing. The following chapters describe typical broadband access methods.

5.3.1 Multilink Dial-Up

An already somewhat historical and only relatively rare method was multilink dial-up. This refers to channel bonding. It increases the bandwidth respectively as the amount of links increase. In practice, this means that the user experiences the multilink dial-up in such a way that it combines two or more dial-up connections and handles them as a single data channel. The drawback of this method is that it requires two or more modems, phone lines and dial-up accounts, with respective increased costs. In addition, the Internet Service Provider (ISP) needs to support multilinking.

This method can be called inverse multiplexing. Its popularity was limited to only a few enthusiasts in the era prior to the introduction of ISDN, DSL and other similar technologies to the markets. In any case, there was a niche-market for some modem vendors at that moment.

5.3.2 ISDN

Integrated Services Digital Network (ISDN) has been the first concrete step to increase data transfer capabilities of the public switched telephone network in such a way that data functionality and capacity are integrated to the network itself, instead of users having dial-up modem connections. In addition to the data, ISDN also integrates the voice service over the same subscription line. In practice, ISDN has been used for voice service as well as for lightweight video conferencing and other data applications. During the most active era, ISDN was especially popular in Europe.

The basic ISDN-BRI, or ISDN-B, offers two 64 kb/s channels for user data and voice. They can be used separately or combined to provide a total of 128 kb/s channel. For enterprise usage, there was also a higher capacity variant available, ISDN-PRI or primary rate ISDN, which is able to combine 30 bearer channels, providing a total of 30 E1 channels with 2 Mb/s in European variant. For the American variant, 23 T1 channels can be combined to provide a total of 1.5 Mb/s per subscription line.

5.3.3 Leased Lines

Dedicated lines, or leased lines, are typically used by Internet Service Providers. It is also a typical solution for any cellular carrier operator in areas or cases when an operator needs to advance quickly in deployment but there is a lack of own capacity or lines.

Dedicated lines can be used to connect Local Area Networks to public Internet, or via a separate IP network infrastructure. The physical layer can thus contain public telephone network infrastructure, dedicated E1/T1 slots, or any other available transport technologies available in the area. Furthermore, the solution can contain fixed lines and radio links. The latter ones are suitable for reaching end-users in remote areas or places where deployment of wired connections is not straightforward.

In a typical case, the cost of leased lines is relatively high, but lower than building own infrastructure during various years. The final decision about whether to choose leased or own lines depends on the business models and expected return of investments. There are examples of technoeconomic optimization presented in, for example, Chapter 22 of this volume (Planning of Mobile TV Networks) for identifying the relevant parameters, and to show the methodology for the analysis.

5.3.4 E1/T1

T-carrier technology provides data rates from 56/64 kb/s (DS0) to 1.5 Mb/s (DS1 or T1), or to 45 Mb/s (DS3 or T3). A T1 line is designed to deliver a maximum of 24 voice or data channels (DS0). This means that end-users are able to use part of the capacity for data and another part for voice traffic. Logically, the whole 24 channel capacity may also be used for pure data.

A single DS3 (T3) line may deliver a maximum of 28 DS1 (T1) channels in such a way that also partial T1 lines can be used in multiples of a DS0 resulting in 56–1500 kb/s data rates.

The corresponding standard is J1/J3 in Japan, whilst in Europe, E-carrier is in use. The latter provides a total of 32 channels, each with 64 kb/s data rate, via E1 (2.0 Mb/s). The next step of the European standard is E3 which provides 512 channels (16 x E1), resulting in a total data rate value of 34.4 Mb/s.

SONET (Synchronous Optical Networking) in the USA and Canada, and SDH (Synchronous Digital Hierarchy) in the rest of the world are multiplexing protocols for digital, high data rate transfer over optical fiber. For the lower transmission rates, an electrical interface can be applied, such as copper wires.

Framing can be done via OC-3c (optical) or STS-3c (electrical). The letter c refers to "concatenated," which means a single data stream instead of various multiplexed data streams. The OC-3c category is able to deliver data rates of 155.520 Mb/s. A single OC-3c contains three OC-1, each capable of delivering 51.84 Mb/s. This means that a full DS3 can be included in this category. For higher data rates, OC-3c multiples of four are utilized, which results in OC-12c with 622.080 Mb/s, OC-48c with 2.488 Gb/s, OC-192c with 9.953 Gb/s, and OC-768c with 39.813 Gb/s.

The 1, 10, 40, and 100 Gigabit Ethernet of IEEE 802.3, that is, GbE, 10GbE, 40GbE, and 100GbE, respectively, provide digital data to be transferred via copper lines within 100 m distance, and via fiber optics within 40 km.

5.3.5 Cable Modem

A cable modem provides Internet access. It is based on hybrid fiber coaxial cable, which is familiar from TV antenna cabling. The last mile from the cable Internet provider's head end up to the customer's location can be fiber optics or coaxial cable. The broadband cable access provides a continuous connection with an Internet Service Provider.

Depending on infrastructure maturity, the maximum downlink bit rates may be approximately 400 Mb/s for business environment, and up to 100 Mb/s for residential environment. The uplink may be up to 20 Mb/s.

5.3.6 DSL

Digital Subscriber Line (DSL) provides connectivity to the Internet via telephone network. DSL operates via single telephone line. As a result of suitable filtering, DSL works without disturbances with voice phone calls, unlike in the case of earlier dial-up connectivity. DSL operates in high frequencies for the data connections, whereas the audio frequencies of the same subscription line are utilized by telephone communication.

5.3.7 Power-Line Connectivity

Power-line Internet, that is, BPL (Broadband over Power Lines) shares the communication capacity and electric power delivery via the same physical cabling. The benefit of this solution is the already established power line infrastructure. As a result, BPL is able to provide Internet access in rural and marginal areas with low cost for additional transmission equipment and related accessories.

The data rate of BPL is asymmetric, ranging typically between 256 kb/s and 2.7 Mb/s. One of the limiting factors of BPL is that data signals do not pass through power transformers, which require an additional repeater at the location of each transformer. As a general development path, BPL is advancing faster in Europe than in the USA due to differences in the power system design in this sense, because the transformer typically serves only a small amount of households in the USE whereas the served number of households is greater in Europe, being typically 10–100.

5.3.8 ATM

ATM (Asynchronous Transfer Mode) is a wide area networking standard. It can be used for the offering of Internet access either directly or as a base for other access technologies. As an example, DSL implementations may rely on ATM layer for offering various different technologies over the same link.

Customer LANs are typically connected to an ATM switch or Frame Relay node in such a way that they are using leased lines and variable data rates. As the other technologies are rapidly deployed for the delivery of packet data, including Ethernet over optical fiber, MPLS, VPN, cable modem and DSL, it has resulted in the lowering of the importance of both ATM and Frame Relay.

5.4 Wired Connectivity

Modern telecommunication equipment requires connectivity for charging and information transfer. In the initial phase of mobile communications back in 1990s, connectivity of equipment was based typically on adapters and RS-232C connectors. It provided connectivity of the device with, for example, external modem. The adaptation of the mobile device with laptops was also made via PCMCIA cards still in the beginning of 2000s.

A real confusion was related to the connectors of chargers. Even the same equipment vendors provided several models which were not directly compatible with each other, although typically the change to other equipment of the same or different vendors could be done via adapter. In addition to the form of the connector, the polarity of the charger needed to be reviewed. In any case, for consumers this was not by any means an ideal situation.

Nowadays, connectivity has been harmonized further, and many vendors have adopted USB as a common platform for both charging as well as for wired information transfer, for example, for synchronizing of the contents of the device.

5.4.1 USB

Universal Serial Bus (USB) refers to the industry standard developed in the mid-1990s. USB standards define the respective communication path including cables, connectors and communications protocols. The complete definition set also includes the means to use USB for power supply purposes.

USB is especially useful for – and the original aim was to include these – connectivity of peripherals to the computers, including printers, pointing devices, external mass storages, network adapters, keyboards, and variety of other typical accessories that are utilized with computers. Ever since the original USB standard was created, there has been an increasing amount of new type of devices that can be used via USB, so it typically is the base solution for the smart phones, and PDAs. As the USB offers both an efficient way to transfer data between computer and connected devices, and as it is also used for charging the batteries of this equipment, it has rapidly replaced the earlier connectivity solutions like serial and parallel communications like RS-232 and printer ports. As a result, the new portable devices typically do not include any more parallel nor serial ports. The goal of the original development group of the participating companies was thus effectively reached as the end-users rapidly got used to easy connectivity according to the plug-and-play principle.

The original USB solution has evolved further, and the steps have been standards 2.0, 3.0, and the latest one 4.0, each containing enhanced data rates and increased functionalities as described in the next chapters.

5.4.2 USB Development

The first USB 1.0 specification was ready in the beginning of 1996. It provided data rates of both 1.5 Mb/s for low speed transfer, for example, for game control peripherals, and 12 Mb/s for full speed transfer, for example, for external hard disk drives. Nevertheless, it was the USB 1.1 in 1998 that started to be a popular base for wider utilization.

As a continuum of the development, the USB 2.0 was ratified by the USB Implementers Forum (USB-IF) in 2001. As a result of this phase, USB 2.0 now provided data rates up to 480 Mb/s which was a significant step compared to the original USB 1.0 / 1.1 data rate.

Furthermore, USB 3.0 was finalized 2008. It provides increased data rates up to 5 Gb/s, which again is a major step compared to the previous USB 2.0 definitions. In order to achieve this speed, the bus solution of USB 3.0 is changed to a more efficient one called SuperSpeed, which is also used as a marketing term for USB 3.0. Nevertheless, the USB 2.0 bus solution is possible to use in a parallel fashion to guarantee backwards compatibility.

The future of the USB beyond the USB 3 standard is still under work, but currently the most probable evolution path leads to the Type-C connector which would replace gradually the previous variants for forming a uniform, symmetric connectivity that fulfils the technical performance of at least USB 3.1. The first specifications for the Type-C connector are expected to be finalized by the first half of 2014. The size of the Type-C connector would be comparable with the Micro USB plug. Furthermore, it is planned to be reversible, that is, it functions regardless of the orientation, and it also would replace the UBS-on-the-go fulfilling the ideas of this concept.

5.4.3 General Principles of USB

USB is based on asymmetrical architecture. It contains host, various downstream USB ports, and a set of peripherals. The topology of USB is a combination of tiered and star.

USB hub can be extended, and thus additional USB hubs can be connected to the setup. A maximum of 5 tier levels is possible to use in the USB tree structure. Furthermore, USB host is allowed to use various host controllers, and each host controller can contain one or more USB ports. USB architecture is thus flexible and the inclusion and removal of multiple devices via the same USB host is possible in a dynamic way. The specifications allow a total number of 127 devices to be used with a single host controller so that the devices are connected in series. It should be noted that albeit USB allows various devices to be connected to the same equipment, all devices share the common bandwidth. In practice, some computers may use more than one USB root hubs, allowing the user to spread devices across more than one root hub as presented in Figure 5.1. Furthermore, for those systems that are able to have separate root hubs, one is USB 1.1 only whereas the others are USB 1.1/2.0 compatible. This is a practical solution because some older 1.1 devices do not provide a good compatibility between USB 1.1 and 2.0.

USB device can consist of various subdevices connected in a logical way. These devices are called device functions, and each USB device can provide one or more functions. A practical example of this principle is a video camera providing the video device function of the USB device. In addition, this device can include an audio device function via the built-in microphone. The whole device in this case can be called a compound USB device. If these logical devices are connected separately to the USB hub, each device also uses different addresses and separate, physical connections and connectors.

The actual communications of the USB devices via the logical connections are called pipes. A pipe is effectively a connection between the host controller and logical entity of the device that is referred to as endpoint. The terms "pipe" and "endpoint" are in practice used similarly, although the endpoint is inbuilt permanently whereas the pipe can be opened and closed by the user. The maximum number of USB pipes or endpoints is 32.

Furthermore, the pipe can be used for streams or messages. The *message pipes* are bidirectional and are used for control purposes of actual data transfer via relatively short commands and status responses. On the other hand, the stream pipe is unidirectional and is used for data transfer. The type of data transfer in this case can be isochronous, interrupt or bulk transfer. "Isochronous transfer" refers to data transfer with partially guaranteed data rate, but there is a possibility of data loss. "Interrupt transfer" refers to the data transfer

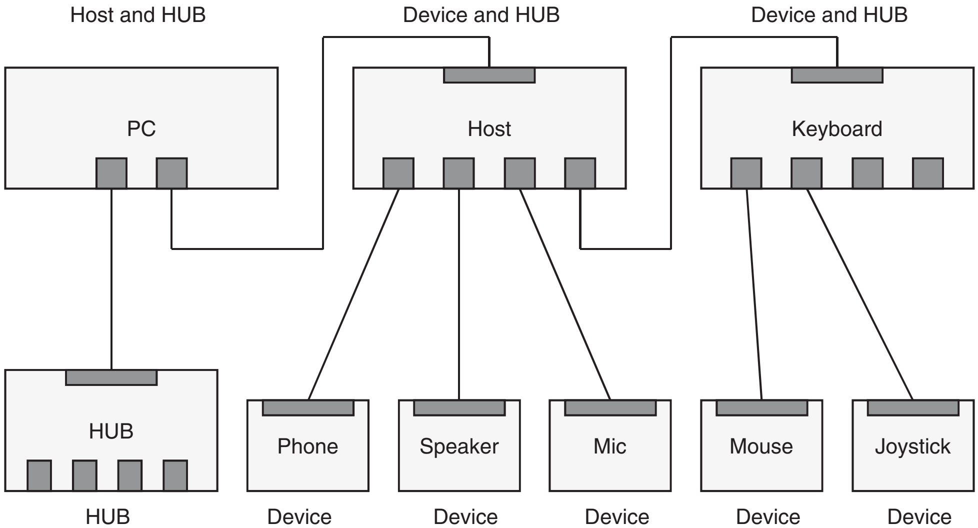

Figure 5.1 *An example of a desktop USB environment. All USB peripherals are controlled by the USB host controller located at the PC motherboard or PCI (Peripheral Component Interconnect) add-in card. Assistance is granted by subsidiary hub controllers. Data shown in web page of Electronic Design. Nevertheless, this is commonly known info.*

for devices with guaranteed quick response. Examples of these devises are keyboards and pointing devices. Lastly, "bulk transfer" refers to large occasional data transfers that may utilize the complete bandwidth, yet without guaranteed bandwidth or latency. File transfer represents one example of this type.

As soon as USB device is connected the first time to the USB host, it triggers a USB device enumeration process. This is initiated by reset signaling that is sent out to the USB device in question. The outcome of the reset signaling is the possible data rate of the USB device, which is now known by the host. Furthermore, the device will have a unique address of 7 bits assigned by the host. In addition, the device drivers are loaded for providing actual communications between the host and USB device. Finally, the device state is set to "configured," and is ready to be utilized. It should be noted that restarting the host triggers a new enumeration procedure for all devices connected to the hub.

As of now, the host controller starts to request upon need the transfers between the devices and host.

Furthermore, USB defines connections to USB mass storage devices. Some examples of typical USB storages are flash drives and external hard disk drives. USB definitions allow transparent utilization of the mass storage devices, so booting of the machine from USB devices is also possible.

5.4.4 Physical Aspect of USB

Figure 5.2 and Figure 5.3 show the physical aspects of USB connectors, while Table 5.2 presents the connectivity of different USB cables.

The data cables of USB 1 and USB 2 are based on a twisted pair in order to reduce noise and crosstalk phenomena. For these type of cables, the twisted pair contains the "Data +" and "Data −" conductors. USB 3 cabling, on the other hand, contains double wiring compared to USB 2 in order to provide SuperSpeed data transmission. This has resulted in a larger diameter of USB 3 cable.

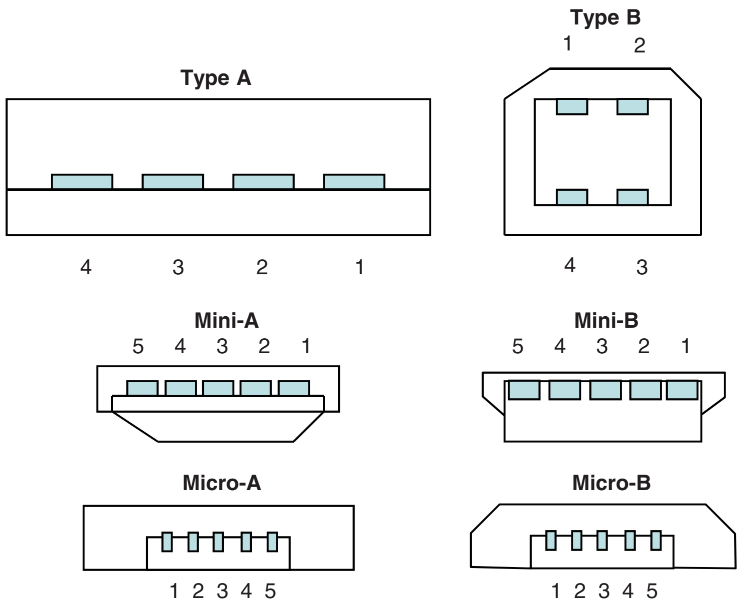

Figure 5.2 *USB types.*

Figure 5.3 *Physical aspects of the currently typically utilized USB connectors: Micro-B plug and standard-A plug.*

Table 5.2 *The matrix for the USB cable connectivity (receptacle and plug)*

	Plug				
Receptable	Type A	Type B	Mini-B	Micro-A	Micro-B
Type A	yes	—	—	—	—
Type B	—	yes	—	—	
Mini-B	—	—	yes	—	—
Micro-AB	—	—	—	yes	yes
Micro-B	—	—	—	—	yes

5.4.5 Cable Length

The maximum length of USB 1 type cable is 3 meters for low speed operation (1.5 Mb/s), and USB 2.0 allows a maximum cable length of 5 meters with high-speed data rate devices (480 Mbit/s). The most important limitation for cable length is the maximum tolerated round-trip delay which is approximately 1.5 μs for the reception of answer from the USB device to the USB host command. As time is needed for USB device response, and furthermore, there are delays due to the maximum number of hubs, the maximum acceptable delay per cable is 26 ns. In practice, the USB 2.0 specification dictates that cable delay should be less than 5.2 ns per meter.

The USB 3.0 standard does not directly dictate maximum cable length. Nevertheless, it states that all cables shall meet electrical specification so, for example, for copper cabling that has AWG 26 wires, the maximum practical length is 3 meters.

5.4.6 Power

The USB 1-series and 2.0 specifications have a 5 V supply via a single wire in order to provide power to the USB devices. The specification dictates limits in such a way that the maximum voltage is 5 V $\pm$ 5% measured between the positive and negative bus power lines. The USB 3.0 specifies the voltage range for low-powered hub ports as 4.45–5.25 V.

USB 2.0 has a unit load of 100 mA, USB 3.0 has a unit load of 150 mA. In the case of USB 2.0 port, the device is allowed to consume a maximum of 5 unit loads, that is, 500 mA, whereas the number of units is 6 in USB 3.0 ports, corresponding to 900 mA.

USB specifications define low-power and high-power devices. The USB 2.0 low-power device may consume a maximum of 1 unit load with the minimum operating voltage of 4.4 V. The value is 4 V in USB 3.0. On the other hand, the high-power device may consume the maximum number of unit loads the standard allows. It should be noted that all USB devices initiate the operation in low-power mode, and may request high-power mode later. The system provides the high-power mode, depending on availability on the bus.

There is a solution for devices that possibly require more current than the specifications allow via a single USB port. In order to avoid any damages or malfunctioning, these devices may have an external power source, or a USB cable with two USB connectors for the data connectivity and power via one connector, and for additional power via the other connector.

5.4.6.1 USB 1.0

USB 1 (full speed) data rates of 1.5 Mb/s (Low-Bandwidth) and 12 Mb/s (Full-Bandwidth) were specified in 1996. Nevertheless, the definitions did not contain extension cables or pass-through monitors. Only a low amount of USB devices were introduced in commercial market until the readiness of USB 1.1 was achieved in 1998. That solved the previous issues of 1.0, from which a major part was related to the hub. The USB 1.1 was basically the first revision that was widely adopted. Tables 5.3 and 5.4 present the pinout of USB 1 and 2.

5.4.6.2 USB 2.0

USB 2.0, that is, High Speed USB, was released in 2000, and accepted commercially 2001. It improved transfer data rate by offering speeds up to 480 Mb/s, although the practical effective throughput can be estimated to be up to about 280 Mb/s.

Table 5.3 USB 1.x/2.0 pinout according to the standard

PIN	Term	Colour of cable	Info
1	Vbus	Red or orange	Voltage +5V
2	D−	White or gold	Data −
3	D+	Green	Data +
4	GND	Black or blue	Ground

Different USB 2.0 variants are as informed in Ref. [2]:

- USB 2.0 Standard-A plug and receptacle. A Standard-A USB plug inserts into a USB host or a hub and carries both power and data.
- USB 2.0 Standard-B plug and receptacle. A Standard-B plug typically plugs into a large device, such as a printer.
- Micro-USB 2.0 (Micro-A, Micro-B and Micro-AB) plug and receptacle. Micro-USB connectors carry both power and data, and support USB On-The-Go. They are used in small portable devices, such as smartphones, digital cameras, GPS devices and more.

5.4.6.3 USB 3.0

The USB 3.0 is an evolved version of USB 2.0. The improvements are the following:

- Theoretical nominal data transfer rate of 5 Gb/s. The transfer type of USB 3.0 is referred as Super Speed (SS). The characteristics of the physical layer of USB 3.0 are similar to Generation 2 of PCIe (Peripheral Component Interconnect Express) and SATA (Serial AT Attachment). The USB 3.0 bus provides theoretical raw data throughput of 4 Gb/s, and the practical data rate is about 3.2 Gb/s.
- USB 3.0 uses two-way data transfer for reception and transmission.
- Link power management states of U0 to U3.
- Enhanced data bus utilization via asynchronous notifications of the host of its readiness.
- Stream Protocol allows a large number of logical streams within an endpoint.

USB 3.0 transmission data rate is about 10 times faster than in the case of USB 2.0 (480 Mbit/s). As USB 3.0 has full duplex communications method compared to the half duplex of USB 2.0 the effective band of USB 3.0 is 20 times that of USB 2.0.

Table 5.4 USB 1.x/2.0 Mini and Micro pinout

PIN	Term	Colour of cable	Info
1	Vbus	Red	Voltage +5V
2	D−	White	Data −
3	D+	Green	Data +
4	ID	N/A	Provides information about host and slave connections: host connected to ground and slave not connected.
5	GND	Black	Ground

Figure 5.4 *USB 3.0 Micro-B.*

There are two options for the physical connector for USB 3.0, based on the standard-A. It allows USB 3.0 Standard-A plug or a USB 2.0 Standard-A plug. It should be noted that it is allowed to plug USB 3.0 Standard-A connector into USB 2.0 Standard-A receptacle. The USB 3.0 connector, as shown in Figure 5.4 is based on the same physical principle as its predecessor, with additional five pins.

The VBUS, D−, D+, and GND pins are required for USB 2.0 communication. The additional USB 3.0 pins define two differential pairs and one ground (GND_DRAIN). The two additional differential pairs are meant for the SuperSpeed data transfer purposes via dual simplex SuperSpeed signalling whilst the GND_DRAIN pin is meant for drain wire termination and to maintain signal integrity.

It should be noted that USB 2 and USB 3 ports can coexist on the same device and they look similar. For this reason, USB 3 connector is distinguished by a blue insert. Table 5.5 summarizes the USB A and B type pins, and Table 5.6 shows the pins for USB 3 powered connector.

The USB 3.0 variants are the following, as informed in Ref. [2]:

- SuperSpeed USB cables and connectors contain 5 additional wires compared to USB 2.0. A SuperSpeed USB plug and a Hi-Speed USB receptacle results in the device working at Hi-Speed USB rates. If Hi-Speed USB receptacle and a SuperSpeed USB plug are used, the device will only work at Hi-Speed USB rates. In order to achieve the data throughput of SuperSpeed USB, a user must have a SuperSpeed USB host, a SuperSpeed USB device and a SuperSpeed USB cable.

Table 5.5 *Standard A and Standard B connector*

PIN	Colour	Term of A-connector	Term of B-connector
1	Red	Vbus	Vbus
2	White	D−	D−
3	Green	D+	D+
4	Black	GND	GND
5	Blue	StdA_SSRX-	StdB_SSTX-
6	Yellow	StdA_SSRX+	StdB_SSTX+
7	Shield	GND_Drain	GND_Drain
8	Purple	StdA_SSTX-	StdB_SSRX-
9	Orange	StdA_SSTX+	StdB_SSRX+
Shell	Shell	Shield	Shield

Table 5.6 USB 3.0 Powered-B connector

PIN	Term	Info
1	Vbus	Voltage
2	D−	USB 2.0 differential pair with D+
3	D+	USB 2.0 differential pair with D−
4	GND	Ground for power return
5	StdB_SSTX-	Superspeed TX differential pair with StdB_SSTX+
6	StdB_SSTX+	Superspeed TX differential pair with StdB_SSTX-
7	GND_Drain	Ground for signal return
8	StdB_SSRX-	Superspeed RX differential pair with StdB_SSRX+
9	StdB_SSRX+	Superspeed RX differential pair with StdB_SSRX-
10	DPWR	Power originated from the device
11	GND	Ground return to power (DPWR)
Shell	Shield	Metal shield of the connector

- USB 3.0 Standard-A plug and receptacle. A Standard-A USB plug inserts into a USB host, or a hub, and carries both power and data. The USB 3.0 Standard-A plug and receptacle are backward compatible with the USB 2.0 Standard-A plug and receptacle.
- USB 3.0 Standard-B plug and receptacle. A Standard-B plug typically plugs into a large device, such as a printer. The USB 3.0 Standard-B receptacle is backward compatible with the USB 2.0 Standard-B plug.
- Micro-USB 3.0 (Micro-B) plug and receptacle. A Micro-USB 3.0 plug is for small, portable devices, such as smart phones, digital cameras, GPS devices and more. The Micro-USB 3.0 receptacle is backward compatible with the Micro-USB 2.0 plug.

More information about SuperSpeed can be found in Ref. [3]. The USB 3.0 specification can be found in Ref. [2].

5.4.6.4 USB Low Energy

USB Low Energy (LE) is an energy optimized addition to the USB connectivity family. Figure 5.5 presents example of LE device, other practical solutions including many types of small objects and smart wearables.

Figure 5.5 *Typical example of USB Bluetooth Low Energy device.*

5.4.6.5 USB OTG

The USB was originally designed as an interface between PCs and peripherals. It has in fact become a successful, all-purpose PC interface in the whole history of computers, and along with USB 3 standardization, there is now increased capabilities for the SuperSpeed communications over USB [4].

As defined from the beginning of USB, USB communication is defined between a host and a peripheral. This is due to the fact that the original focus of standardization was to place the workload on the host that has been PC, whilst USB peripherals can be left as relatively simple devices. This means that the PC side needs to provide power to peripherals, with 0.5A at about 5V as defined in USB 2, or 0.9A at about 5V as defined in USB 3. The PC side is also assumed to support all possible data rates, that is, SuperSpeed, Low Speed, Full Speed and High Speed. Furthermore, PC side is supposed to support all defined data flow types, that is, control, bulk, interrupt and isochronous.

Along with the general development of processors, there is no longer a clear division between PC and peripherals in regard to computing performance. Furthermore, as the chipsets have become more economical, the previous role division of PC and other devices has become less important. It is clear that currently there are many devices – which are not exactly PCs as they were understood to be some years ago – that are needed to connect directly to peripherals. Some examples of this need are printers that can be connected directly with cameras, or smart devices that can be connected to USB headsets [4]. On the other hand, not all possible host functionalities are needed for these non-PC devices.

For this reason, there is a need to partially support the USB host functionalities in non-PC devices, and there is now a specification which defines these non-PC hosts as Targeted Hosts. A Targeted Host is a USB host that supports a specific, targeted set of peripherals [4].

Furthermore, the developer of each Targeted Host product defines the set of supported peripherals on a Targeted Peripheral List (TPL). A Targeted Host needs to provide only the power, bus speeds, data flow types, etc., that the peripherals on its TPL require. This concept is called as USB On-The-Go (USB OTG).

USB OTG device is based on a single USB connector which is a Micro-AB receptacle. It can accept Micro-A as well as Micro-B plugs, and can be attached to any of the defined USB cables and adapters as defined in Micro-USB 1.01 standard.

The USB OTG concept works in such a way that USB OTG device which uses A-plug insertion is called A-device. This device takes care of the powering of the USB interface upon request, and takes the role of host. On the other hand, USB OTG device with the B-plug insertion is called a B-device. As is logical, this type of device automatically takes the role of a peripheral.

Furthermore, the USB OTG device that has no plug inserted acts by default as a B-device. The flexibility of the OTG is shown in cases when the application that uses a B-device requires the role of host. Via the Host Negotiation Protocol (HNP), it is now possible to temporarily grant the role of host for the t B-device.

5.4.6.6 Proprietary Connectors

In addition to USB connectors defined by the standards, there are also proprietary USB connectors in the markets. The reason for these variants may be technical or marketing. In these proprietary cases, the complete functionality of the ports or cables connected to standard USB ports may not be guaranteed because the set of supported functionality may be limited to, for example, the use for only charging of the battery without data transfer capabilities. Furthermore, the supported charging current of these proprietary devices may differ from the official, standard solutions which might make them incompatible with other devices even when utilized only for charging.

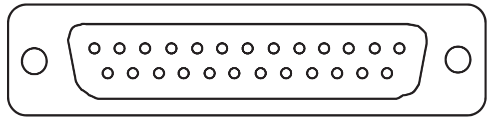

Figure 5.6 *DB-25 connector based on RS-232 standard.*

5.4.7 Serial Port

In telecommunications, RS-232 refers to a series of standards that provide connectivity for serial binary data and control signals between DTE (Data Terminal Equipment) and DCE (Data Communication Equipment). The standard describes electrical characteristics and timing of signals, among other characteristics like the description of signals as well as physical size and order of connector pins. RS-232 has commonly been used in computer serial ports, but it has had several special use cases for managing and controlling devices. The latest standard is TIA-232-F, named "Interface Between Data Terminal Equipment and Data Circuit-Terminating Equipment Employing Serial Binary Data Interchange," from 1997.

At the moment, RS-232 has lost its importance practically completely in the consumer markets of computers, and it is hard to find a laptop or PC with RS-232 connector. Before USB took over RS-232, it was used as a multipurpose connector for connectivity to modems, printers, mice, data storage and other peripheral devices. The benefit of RS-232 was the common base for typical peripherals, but the practical drawback was the low transmission speed and wide voltage margin. Furthermore, RS-232 had problems due to nonstandard pin assignment of circuits on connectors, and incorrect control signals which was one of the reasons why a more stable standard was needed. RS-232 devices are currently found merely in special industrial and scientific instruments whilst USB has completely taken over the consumer markets. Figure 5.6 presents the DB-25 connector that has practically disappeared from the consumer equipment as the USB connector has replaced it.

Nevertheless, for the devices requiring physical RS-232 connectivity, there is a wide range of adapters in the market that takes care of the conversion between USB and RS-232.

5.5 Radio Connectivity in the Near Field

The mobile communications for the wider areas contains cellular technologies of 2G, 3G and evolved versions like LTE and LTE-Advanced. These actual cellular technologies can be utilized, for example, for synchronizing the contents of the devices via data services, SMS, and cloud services. Typical use case would be the sending of visit card from the contact list of A-subscriber to B-subscriber via SMS. Also, the whole contents of the device can be backed up to the cloud service.

For the near field, connectivity can be utilized, for example, for local information sharing, to Internet connection, etc. There are various methods in the commercial market, and each solution provides optimum use cases depending on data rates and distance. The currently available methods are:

- Infrared (IR). This can be used to create a link between two devices (e.g., mobile to mobile, or mobile to laptop) for information sharing, including photos, contacts and other contents. The importance of this method has been dramatically decreasing in the recent years, and many devices do not include IR any more, nor recent laptops.
- Bluetooth (BT). The popularity of this method has been increasing along with technical advances since the first versions.

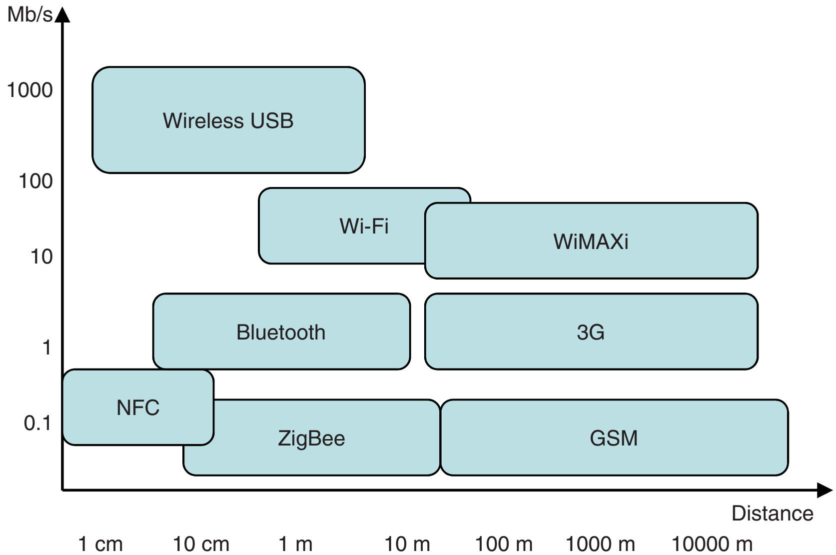

Figure 5.7 *Examples of wireless connectivity solutions with respective coverage and data rate.*

- Wi-Fi. The importance of Wi-Fi has been increasing gradually as there are more hotspots available in public places. Wi-Fi is typically a method to connect to Internet services as an affordable alternative to cellular connectivity. Typically, SW updates of smart devices are done increasingly via Wi-Fi as an alternative to SW downloading via laptop/PC connected to a mobile device. Operators are also increasingly interested in offering Wi-Fi Offloading solutions to consumers to balance the cellular network load.
- Near Field Communications (NFC). This is the closest range technology that is gradually appearing into new markets as of 2012. It can be used in many solutions via tap gesture, including information sharing (similar to RFID), establishing connections for audio/video, and for performing secure payment.
- Radio Frequency ID (RFID) is based on readable and optionally writeable tags. This is slightly out of the scope of mobile devices, although it can be integrated into the device / SIM.
- Wireless USB
- Ultra wideband (UWB)
- ZigBee.

Figure 5.7 summarizes and compares typical wireless connectivity technologies. Figure 5.8 presents the difference between 2.4 and 5 GHz variants of Wi-Fi.

5.6 NFC and Secure Payment

5.6.1 General

Near Field Communication (NFC), is a short-range wireless communication technology that enables the exchange of data between different devices, including, for example, handheld mobile devices. The technology is based on the high-frequency radio interface that provides functional connections within a distance of about 10 centimeters between NFC enabled devices. It should be noted that maximum practical distance depends

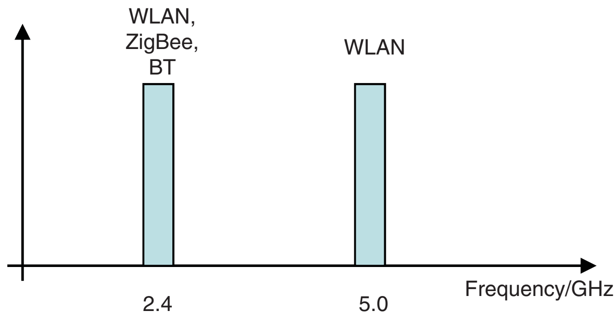

Figure 5.8 *The high-level division of the 2.4 GHz and 5 GHz frequency band utilization.*

on the specifications as well as additional requirements, for example, the distance requirements can be tighter for operators and credit card companies.

NFC is based on the extension of the ISO/IEC 14443 proximity-card standard referred as RFID (Radio Frequency ID). Nevertheless, it should be noted that NFC is not RFID. Even if NFC and RFID do include common functionalities, RFID is about small and economical tags that are readable within certain distance wirelessly. One of the examples of RFID is a warehouse inventory tagging, which provides instant information about, for example, the number of items of different types and their characteristics found within that specific spot.

Instead, NFC is about point-to-point communication between two NFC devices which can be physically, for example, computers, cellphones, laptops and PDAs. Unlike in the case of one-way information transfer of RFID, NFC allows bidirectional communication between two devices.

NFC combines the interfaces of a smartcard and a reader into a single device. An NFC device can communicate with both existing ISO/IEC 14443 smartcards and readers, as well as with other NFC devices, and is thereby compatible with existing contactless infrastructure already in use for public transportation and payment. NFC is primarily aimed to be used in mobile phones.

5.6.2 Readers and Tags

Readers, or devices, are basically always powered when they function. For the power, readers use a battery or an external power source. Readers need to be able to create electromagnetic field that is utilized for the radio transmission. The reader is active as soon as it produces an RF field.

Tag normally has no power. It gets sufficient amount of energy via the nearby RF field, and can thus respond via a carrier modulation which is one form of a passive communication. Since tags are not able to activate radio channels, two tags cannot communicate with each other like NFC devices do.

In general, NFC can be divided into the following features:

- *Tag reading* (telephone numbers, URLs, visit cards). As an example, there have been tag readers in the real environment that can be used for calling taxi, to order the taxi via SMS, and to read bus stop time schedules, since 2011.
- Easy *setup*, including, for example, BlueTooth headset or other accessory pairing by tapping the tag.
- *Sharing* of, for example, photo or other supported contents from a handheld device to another phone, or, for example, to a digital photo frame by touching it. As soon as the initiation has been done, the actual transfer of the contents happens via BlueTooth.

- *Payment and Ticketing* (P&T) can be applied, for example, in a similar manner as in the case of credit cards and digital bus tickets. The possibilities for the practical solutions are actually endless. It should be noted that P&T is relatively complicated ecosystem, which includes middleware software (including APIs) and adaptation (such as SWP), as well as hardware platform (NFC chip, antenna), Secure SIM cards, OTA (Over the Air), Trusted Service Manager (TSM), purchase readers, different wallet and payment applications and midlets (from operators and/or credit card companies). In addition, credit companies and operators require different certificates such as EMVCo, MasterCard Certification and Visa Certification. Also NFC Forum has its own certificate for these purposes.

5.6.3 Architecture

The Near Field Communication Forum (NFC Forum) has defined the architecture of NFC. The NFC Forum's policy regarding the use of the trademarks NFC Forum include, for example, rules that define the permission to use the NFC Forum logos, which is granted to designated members only. It should be noted that technically, the NFC Forum differentiates between NFC Forum Devices, which are actually within the scope of the Forum, and NFC Forum Tags, which are not.

An NFC Forum Device is a device that complies with the High Level Conformance Requirements and implements at least the mandatory parts of the NFC Forum Protocol Stack and at least the mandatory NFC Forum Operating Modes. The mandatory NFC Forum Operating Modes are the NFC Forum Peer Mode and the NFC Forum Reader/Writer Mode. The optional support means that NFC Forum Devices can optionally support NFC Forum Card Emulation Mode. Furthermore, an NFC Forum Device can additionally support optional parts of the stack and also additional protocols and applications that are not defined by the forum.

An NFC Forum Tag can be any contactless component that an NFC Forum Device is capable of accessing, as defined by one of the Type X Tag operation specifications. NFC Forum Tags are not required to support the complete specification for the NFC Forum Protocol Stack.

The *NFC Forum Protocol Stack* includes protocols for communication between NFC Forum Devices, between NFC Forum Devices and NFC Forum Tags, between NFC Forum Devices and technology-compatible contactless smart cards, and optionally between NFC Forum Devices and existing reader/writer terminals. It does not make any assumptions about the implementation or the overall architecture of NFC Forum Devices.

The NFC Forum Protocol Stack supports the following operating modes:

- The *NFC Forum Reader/Writer Mode*. This mode is capable of reading from and writing to NFC Forum Tags. In addition, this mode allows communication with compatible smart cards.
- The *NFC Forum Peer Mode*. This mode is meant to communicate with other NFC Forum Devices.
- The *NFC Forum Card Emulation Mode*. This mode is optional and emulates the behavior of a smart card or tag. Communication with existing technology compatible reader/writer terminals is possible in this mode.

In *NFC Forum Reader/Writer Mode*, an NFC Forum Device has capability to at least communicate with NFC Forum Tags. The device can possibly exchange data with NFC Forum Tags by NFC Forum's or third-party message formats. The device may also communicate with a variety of components like smart cards, memory cards and tags, provided they are compliant with some of the contactless technology types. An NFC Forum Device supports the RF interfaces NFC-A, NFC-B and NFC-F.

In *NFC Forum Peer Mode*, an NFC Forum Device has the capability to communicate with another NFC Forum Device. The service discovery protocol is the mechanism used to identify common services supported by both NFC Forum Devices.

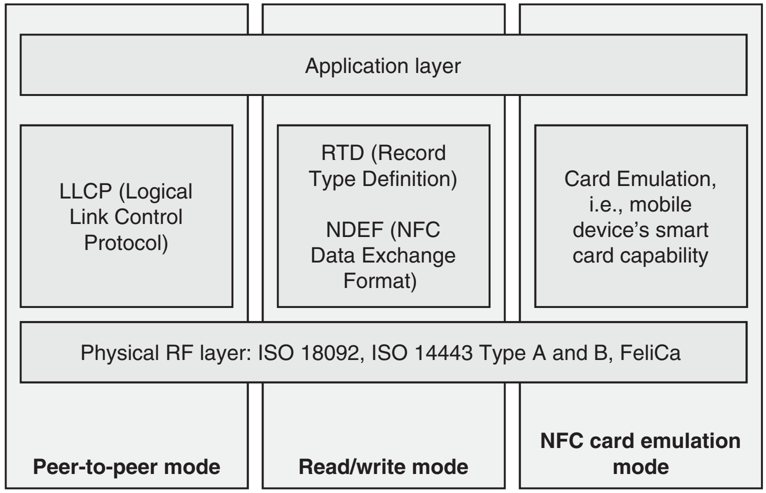

Figure 5.9 *NFC architecture as defined by NFC Forum.*

The *NFC Forum Card Emulation Mode* allows an NFC Forum Device to behave like a smart card or tag in front of a conventional technology-compatible reader/writer. This mode includes the emulation of memory cards and tags, and the emulation of smart cards is intended mainly for portable devices that can be conveniently presented to reader/writers. Using this mode, existing technology-compatible terminal infrastructures (e.g., for payment and ticketing) can communicate with NFC Forum Devices supporting NFC Forum Card Emulation Mode.

The NFC-defined architecture is shown in Figure 5.9 [5]. The technical architecture contains an initial set of mandatory tag formats based on ISO 14443 Type A and 14443 Type B standards, and Sony's FeliCa. These include the following items:

- NFC Data Exchange Format (NDEF) which specifies common data format for NFC Forum devices and NFC Forum tags.
- NFC Record Type Definition (RTD) which specifies standard record types used in messages between NFC Forum devices and between NFC Forum devices and tags.
- Text RTD which is meant for records containing plain text.
- Uniform Resource Identifier (URI) RTD which is meant for records referring to an Internet resource.
- Smart Poster RTD which is meant for posters incorporating tags containing text, audio or other data.

5.6.4 Standardization

The development is done in NFC Forum. It also indicates the implementation of the standards documented in the forum. The NFC forum was established 2004, and ever since, the number of participating members has grown significantly.

The mission of the NFC Forum is to advance the use of NFC technology by developing standards-based specifications that ensure interoperability among devices and services. The means can be to encourage the development of products using NFC Forum specifications, to educate the market globally about NFC technology, and to ensure that products claiming NFC capabilities comply with NFC Forum specifications.

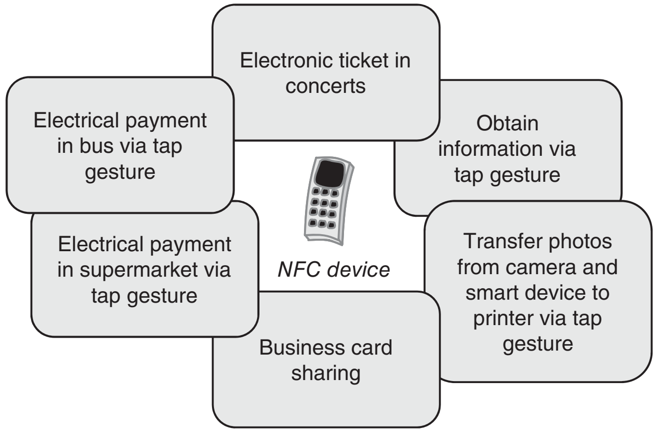

Figure 5.10 *Typical use cases for NFC. Data published by NFC Forum. Nevertheless, the book example is showing very general info, publicly known.*

5.6.5 Products and Use Cases

5.6.5.1 Commercial NFC Products

There are various NFC enabled devices in the markets. Some of the typical use cases of NFC are:

- Connect Electronic Devices according to Peer-to-Peer Data Exchange concept.
- Access Digital Content according to Reader/Writer concept.
- Make Contactless Transactions according to Card Emulation concept.

Figure 5.10 presents more use cases for NFC.

The use cases can be related, amongst a vast amount of situations, to the following: enhancements to loyalty programs (by tapping loyalty card, e.g., in the airport), electrical format of offer coupons, content gathering and transferring, access card to physically closed locations, assets management, reporting, and for making connections.

The following sections present some use cases that are expected to appear and develop in real life.

5.6.5.2 NFC in Transport

Transport environment is one of the most logical for NFC usage. The payment of train tickets, bus tickets and taxi rides can be done, for example, with a mobile device with embedded NFC functionality on the go. The air flight environment also provides interesting cases that can be handled via NFC, like flight reservation and loyalty program management, and entering VIP lounges and boarding areas by utilizing the NFC enabled device. Also the luggage tracking can be combined easily to the same NFC concept. It is logically possible to make the payment also beforehand in retail shops.

The NFC-enabled mobile device is suitable for ticketing as the tickets can be stored to the device beforehand, as well as for access to transit areas and vehicles. At the same time, the user can investigate time schedules and maps of public transportation by utilizing smart posters with the same device. As an extension to the traditional functions of the passenger, the user can also download special offers from the smart poster to the device, in order to get discounts from the related travels.

There have been functional solutions about the ordering of taxi via the tag. As an extension to this idea, the address of the user can be informed to the taxi driver's NFC-enabled device.

5.6.5.3 NFC in Retail

The payment can be done with NFC mobile devices at contactless points of service. The loyalty programs and utilization of coupons are straightforward. Furthermore, the downloading of coupons and other special offers can be done directly from smart poster to NFC phone, and the transfer of coupon can be done fluently to friends with a note of recommendation of products. In another direction, the user can collect information about purchases by reading the product history. Also the touch tags can be used to collect shopping lists, with additional offers of the retailers. The NFC environment can combine various functions, like the collection of deposits from bottle recycling machines.

5.6.5.4 NFC in Public Sector

The NFC can be used to pay community services like parking lots, with a record of the parking stall in NFC phone. The NFC enabled device can be used to access parking areas, buildings and offices by using NFC mobile device and contactless reader. In general, the NFC device can serve as ID card, visa and passport.

5.6.5.5 NFC in Health Care

The NFC device can be used to make health care payments, to identify the patient, and to show health care insurance information. The device can also contain the health care history of the patient, including access to graphical contents like previous X-ray pictures, and important data as the health data that shows the normal values of patient, possible illnesses that should be taken into account if the patient is found, for example, during severe health conditions and the patient is not able to communicate. NFC can also be utilized by the patient to access restricted areas in hospitals, and to show prescriptions in pharmacies that the doctor had ordered previously in paperless format. The doctor can see the history of, for example, purchased medicines and, if the device is connected to a larger mobile health management system, also the health values of the patient.

5.7 Secure Payment

5.7.1 Softcard

Softcard (formerly known as ISIS) is a mobile commerce joint venture between AT&T Mobility, T-Mobile USA and Verizon Wireless, with the aim of providing consumers and merchants with an open and secure mobile commerce platform that will provide a mobile alternative for actions how consumers shop, pay and save. The Softcard mobile commerce platform will be available to merchants, banks and mobile carriers. Softcard and ISIS are trademarks of JVL Ventures, LLC in the US and/or other countries.

Especially for the US market area, three of the largest US wireless service providers, that is, AT&T Mobility, T-Mobile USA, and Verizon Wireless, have united to build a nationwide mobile commerce network utilizing smartphone and NFC technology. This cooperation includes the bringing together of merchants and consumers, Softcard mobile commerce network providing the actual shopping experience for the end-users.

GEMALTO, the digital security related company, has been selected by Softcard to secure its mobile commerce platform through GEMALTOs Allynis Trusted Service Manager (TSM) solution. Another example of activities of Softcard is the cooperation with Giesecke & Devrient (G&D), which also is security solutions and smart card provider.

Softcard aims to enable consumers to experience the speed, security and convenience of mobile contactless payment using NFC technology at retail outlets such as restaurants, movie theaters and drug stores. Consumers will be able to securely pay, present loyalty cards, and redeem coupons with a tap of their phones.

Giesecke & Devrient and GEMALTO are companies providing electronic passports and identity cards, contactless payment cards, as well as subscriber identification modules (SIM) and universal integrated circuit cards (UICC) in mobile phones. Also, in the emerging machine-to-machine applications these companies are suppliers of wireless modules and machine identification modules (MIM). To operate these solutions and remotely manage the software and confidential data contained in the secure devices these companies also provide server platforms, consulting, training, and managed services to assist customers achieve their goals.

According to Ref. [6], Softcard acting as its own TSM as well as enabling other TSMs, intends to push secure information and apps between banks and others to the handset. As an commercial example, if Verizon, AT&T and T-Mobile end up equipping each of their phones with the Softcard wallet, Softcard could potentially facilitate other mobile wallets to work through Softcard in order to function on the three MNO networks. The combined size of the market served by Softcard that it has considerable clout over, remains the single most threat to, e.g., Google Wallet. Another example about the high market dynamics is the introduction of Apple Payment in 2014.

Related to the key North American markets, Ref. [6] speculates that Verizon, despite being a traditional CDMA network, foresees SIM Card based phones on its 4G LTE network that will use the Softcard wallet. Sprint on the other hand is waiting to see if Softcard plans on having a presence beyond the SIM based secure element approach before initiating activities.

5.7.2 Background to Secure Payment Standardization

NFC is a subset of the RFID domain, and is based on a proximity range frequency of 13.56 MHz. This frequency range is dominated by the ISO-14443A, ISO-14443B, FeliCa and ISO-15693 tag standards.

Nevertheless, the 14443A, B and 15693 do not define security architecture. The Ecma-340 standard is meant for information exchange between devices that have more capabilities than merely simple memory storage. It is based on the stack utilized in 14443A standard, but it allows more functionality in addition to the reading and writing memory. Nevertheless, it does not contain security architecture either.

The focus of NFC Forum is the standardizing of the application domain. As an example, a data storage standard (NDEF, NFC Data Exchange Format), and mapping for tag types are defined by NFC Forum.

As a result of the focus, the security topic has been decided to be handled in the application layer. NFC Forum has a separate working group for security that focuses on identifying potential threats and attacks against NFC.

5.7.3 Functionality of Secure Payment

The secure payment via NFC can be based on a Secure Element which is a tamper resistant device with an embedded microprocessor chip. Secure Element stores applications to perform secure execution and keys to perform cryptography such as ciphering, authentication, or signing for NFC services. Secure Element can store multiple applications to support NFC services such as payment, offers, and loyalty. These Secure Element applications are accessible by mobile applications through the baseband and accessible by contactless readers through the contactless interface.

The supported Secure Element form factors can be, for example, UICC, Embedded Secure Element (i.e., within the handheld terminal's HW), NFC Enables such as MicroSD with integrated antenna and SIM-Wrapper.

Furthermore, The Secure Element needs to be based on a tamper-resistant chip, which may be, for example, soldered on the device (Embedded SE), or included in a removable form factor. The Secure Element must accept commands coming from the mobile device (in contact mode through NFC controller for Embedded SE or through ISO7816 hardware interface for UICC) or from the external antenna (in contactless mode).

Its features must be sufficient to cover the most common applications, including those that require fast cryptographic computations (secret key and public key). The memory requirements may depend on the possibility to provision new applications via OTA.

In addition, The Secure Element should support Java Card specifications and implement Java Card API.

5.7.3.1 Embedded Secure Element

In the embedded Secure Element solution (eSE), the secure element of NFC is integrated into the hardware of the device. An example of this solution is Google Nexus S, which uses NFC chip of NXP. It can be speculated that RIM and Google, and possibly Microsoft and Apple, might prefer eSE solution [6]. Apple has, in fact, launched the iPhone 6 model with embedded SE 2014. The most important benefits of eSE are:

- It provides a common architecture for content providers. This does not depend on the mobile system.
- All the data is encrypted when it is stored. It also remains encrypted during the processing along the complete route the data is present.

Some of the drawbacks of eSE are:

- It might be difficult to transfer applications to a new handset.
- There is not yet too large base for phones supporting an integrated NFC Chip.
- For all the new device models, the payment applications must be retested which may lead to delays in the device development.
- If the device is delivered to the maintenance, the secure element is exposed to others. Even if this is a highly hypothetical case, and there is encryption, there could be fraudulent intentions.

5.7.3.2 SIM Based Secure Element

It could be reasoned that SIM, which has already been designed for a reliable means to authenticate and authorize subscribers, and is thus an important base for the billing, would also be a logical storage for the Secure Element that supporting mobile payments. It is considered secure, and Over The Air (OTA) provisioning is an additional benefit of it. The other side of the coin is the amount of control a single entity would get, that is, operator. For that reason, the solutions that are based more strongly on the ecosystem via external SE approach and Trusted Service Managers (TSM) are still important alternatives. One of the clear examples of this approach is the operator-lead Softcard that started the development as an independent payments processor, and morphed later into a TSM.

The benefits of SIM based secure element are:

- This solution is generally preferred by Mobile Network Operators (MNO), and is controlled by the issuing party.
- As SIM is already an established technology, the solution complies easier with the security standards of financial Institutions.
- The solution is independent of the handsets, which provides faster development and deployment as this method is independent of handsets.
- The Over-the-Air (OTA) provisioning is possible to apply with this solution, meaning that new secure payment-related applications can be downloaded remotely.
- If the device is stolen or lost, the applications located into SIM can be blocked by the operator.
- The SIM can be relocated to other devices.
- SIM also supports various security compartments and thus a number of different cards.

Some of the cons of the solution are:

- The solution is based on the functionality of the operator's processes and thus requires cooperation with operator network.
- In the case of various payment applications within a single SIM, it might not be straightforward to divide the responsibilities of the control and visibility of credit cards from separate banks.
- The sharing of the costs between the operator and other parties, for example, when operator applies fees for transactions (revenue sharing vs. flat fee).

5.7.3.3 Secure Digital Card Based Solution

The third solution is based on a combination of SD card and NFC antenna that allows the handset to communicate with contactless readers. This solution, with the Secure Element stored on SD card, is agnostic of both network operators as well as device manufacturers. In general, the Secure Element can thus be located to the microSD card whilst the handset takes care of the physical NFC functionality.

As an example, DeviceFidelity provides a microSD card based Secure Element. The company has partnered with VISA on its In2Pay microSD solution to offer NFC payment capabilities across VISAs payWave platform. DeviceFidelity allows its microSD cards to be issued and personalized like traditional smart cards. It has partnered with Vivotech to add OTA provisioning capabilities to its In2Pay microSD product [6].

Some of the advantages of this solution are:

- It facilitates rapid application deployment.
- It functions with existing hardware.
- It does not depend on the mobile network operators or device manufacturers, and may thus be preferred by financial institutions.
- It allows the bank institute that issues the card, to own the secure element.

5.7.4 EMV

Related to the secure payment via NFC, the EMV Integrated Circuit Card Specifications for Payment Systems are global payment industry specifications that describe the requirements for interoperability between chip-based consumer payment applications and acceptance terminals to enable payment. The specifications are managed by the organization called EMVCo.

Named after the original organizations that created the specification, Europay, MasterCard and Visa, the EMV specifications were first published in 1996. According to the statistics of EMVCo. In the year 2010, there were one billion active EMV chip cards used for credit and debit payment, at 15.4 million EMV acceptance terminals deployed around the world.

5.7.5 Practical Solutions

The first operator adapting the NFC secure payment was Orange providing the service since 2012. It can be expected that various operators will follow the trend as of 2014.

5.7.6 Other Payment Solutions

The markets of secure payment via mobile devices are expected to be grown. It can be stated that from the alternatives, Google Wallet has been informed quite actively in the public media. There are also other alternative solutions appearing in the markets.

5.7.6.1 Google Wallet

The Google Wallet mobile application has been designed to securely storing the credit cards and offers of the users on the phone. When checking out at brick-and-mortar stores that accept Google Wallet, users are able to pay and redeem offers with the same device by tapping the phone at the point of sale [7].

Google has coordinated partnerships, for example, with MasterCard, Citi, Sprint and First Data. Google Wallet works with its Nexus S smartphone equipped with an NFC embedded chip, combined with MasterCard PayPass terminals used by over 140 000 merchants. Google plans to subsidize NFC equipped POS Terminals by VeriFone to select retailers during its trial. The solution encrypts data on the NFC embedded chip, and requires that the customer use a PIN to authorize every transaction. In addition, Google has also announced that it will use fingerprint sensors as an added security measure. According to Ref. [6], Google Wallet is expected to be ported also to other platforms and devices like iOS and Blackberry.

Google Wallet has also received some negative publicity as can be interpreted, for example, from Ref. [8]. According to the source, the PIN for the payment can be revealed relatively easily by subject matter experts. When this vulnerability was revealed, Google was receptive to the findings, and attempts at a fix have to be balanced with the coordination of the banks, since changing the way the PIN is stored would have impacts on responsibilities for security. Meantime, users were advised to take precautions by refraining from rooting the phone, enabling the lock screen, disabling the USB debugging, enabling Full Disk Encryption and keeping the handset up-to-date. Later, Google has clarified that the potential problem is only for the case of users of rooted devices. The advice was thus updated to encouraging people not to install Google Wallet on rooted devices, and to always set up a screen lock as an additional layer of security for the phones.

Furthermore, according to Ref. [6], Google has shrugged off the coup by Softcard in signing up MasterCard, VISA and American Express to handle payments generated through Softcard, by pointing out that it remains at least a year ahead of Softcard, and that it supports an open platform. Google Wallet is designed to be interoperable but so far these efforts have not been met with reciprocal affection from other players in the payment ecosystem. If Softcard MNOs (Verizon, T-Mobile & AT&T) mandate that mobile wallets must work through Softcard on their networks, Google may be forced to modify Google Wallet so as to work with a SIM based secure element.

5.7.6.2 Visa

During February 2012, Visa has announced a mobile payment solution that is competing with Google Wallet and the carrier-backed Softcard payment system. The solution is based on the "Visa-certified" NFC-equipped smartphone, with what the consumer can contact the company and activate the handset for mobile payments. In the solution, the device is linked securely with a user's bank account. This provides mobile payments in those locations where Visa's payWave system is accepted. This means that as is the case in the secure provisioning of payment cards traditionally, Visa has now extended the idea for mobile technology to securely provision mobile payment accounts over the air [9].

The company has announced that Intel Atom-powered smartphones and tablets will be the first Visa-certified devices to allow mobile subscribers to securely make NFC purchases.

VISA plans to have a multipronged strategy including bets on NFC (via Softcard and DeviceFidelity) while launching a P2P payment service separately in 2012 [6].

5.7.6.3 American Express

According to Ref. [6], American Express has a strategy in extending its proprietary payment network in to online, mobile and NFC based proximity payments space. AmEx through its recently launched Serve platform, while being similar to PayPal, aspires to be something much broader that integrates mobile payments, loyalty programs and other social and connected services. It has signed up Sprint and Verizon, while its partnership with Payfone will allow millions of customers from either Mobile Network Operator (MNO) to use AmEx to pay using their mobile phone number. The Serve digital wallet service is accepted by the millions of merchants who accept AmEx.

5.7.6.4 Square

According to Ref. [6], Square allows credit cards to be transacted via a mobile phone equipped with a square reader. As a potential disruptor in the POS market, Square started off at the low end, creating its own market and moving up market to eventually dethrone traditional POS terminals vendors.

Nevertheless, Square has no NFC presence so far. For that reason, Square would be disruptive in an environment that lacks a traditional POS or Card Infrastructure, for example, in Africa and Asia. Source [6] speculates that if Square keeps plastic alive in those nations, then NFC will face some uphill battles for adoption outside the USA, in case Square is able to do what it does best: capture the low-end market and steadily move upmarket to ultimately push out incumbents.

5.7.6.5 Others

Some other initiatives are listed below, summarized from Ref. [6]:

- Bank of America, Wells Fargo and Chase have formed a venture to enable P2P payments for their customers. ClearXChange allows customers to send money to each other without needing to open a separate ClearXChange account. It is attempting to make P2P payments fluently for banking customers. Bank of America, Wells Fargo & Chase partnered with DeviceFidelity and VISA to use its In2Pay microSD solution to run NFC payment trials across VISAs payWave platform. These trials lend evidence to the fact that financial institutions are testing the waters with their own mobile payment applications to test market adoption while waiting it out for Google, Softcard and the broader industry to offer a standard and a clear way ahead.
- The iPhone users are ready to use their iPhone as of September 2014 to make mobile payments upon the support of the system by merchants. According to publicly available statements, it can be speculated if more customers trust Apple and Google to run their mobile wallets, compared to AMEX, VISA and MasterCard. If Apple enters banking by acquiring one, and becomes a credit source for its iTunes accounts, then the 200 million credit portfolio it has already built up in iTunes will make it a strong competitor to banks. One speculative possibility may be that Apple decides to circumvent the POS backbone and use the over 100 million IP enabled devices (iPhones) to further disrupt the payment industry.
- Other than Apple, Amazon, Microsoft & Facebook could also launch a mobile payment solution and enter local commerce or partner with one of the above wallet solutions.

5.8 Bluetooth

5.8.1 General

Bluetooth (BT) is a standardized short range wireless radio technology making it possible to connect electronic devices with each other [10]. BT can be generalized as technology replacing cables that have been utilized in the initial phase of the mobile devices. BT is essentially a global standard for short-range wireless connectivity, allowing a broad range of electronic devices to connect and communicate with each other. In addition to mobile devices, some examples of the devices are: headsets, MP3 players, PCs and peripherals.

Table 5.7 The approximate useful distances of Bluetooth devices as a function of the transmitter power level. Practical value may vary greatly depending on obstacles and radio conditions

Class	Max power (mW)	Max power (dBm)	Distance (m)
1	100	20	100
2	2.5	4	10
3	1	0	5

5.8.2 Bluetooth RF

Bluetooth is basically meant for replacing wires between equipment like computers, their peripherals and mobile devices. Characteristic to Bluetooth is low power consumption and short functional range. The latter depends on the power classes of Bluetooth. The aim of Bluetooth is to offer low-cost transceiver technology. As the interface is based on RF, one of the benefits over older technologies like infra red is that no line-of sight is requires as long as the received power level is high enough. Table 5.7 presents the power classes of Bluetooth.

5.8.3 Bluetooth Profiles

Bluetooth profiles can be defined as general behaviors through which Bluetooth enabled devices communicate with other devices. This means that in order to be able to connect the devices together via Bluetooth technology, both must support and understand the common Bluetooth profile in use. The Bluetooth profile describes the possible applications that can be used in the connection, and how Bluetooth is used. As an example of the profile, a File Transfer profile defines how devices should use Bluetooth in order to transfer files between devices, which can be physically, for example, mobile device and PDA.

In order for the Bluetooth device to connect to another, both devices must share at least one of the same Bluetooth profiles. Another example would be a Bluetooth headset that is utilized via a Bluetooth enabled cellphone. Both the headset and mobile device should have and use the Headset (HS) profile, which basically defines the way to initiate, maintain and release the connection between, for example, headsets and mobile devices.

There are various Bluetooth profiles developed. The manufacturer of the Bluetooth device assigns a set of Bluetooth profiles for the device to a certain set of applications that work with other Bluetooth devices.

According to the Bluetooth standards, all Bluetooth profiles should include as a minimum the following set of information:

- Dependencies on other profiles.
- Recommended user interface formats.
- Particular parts of the Bluetooth protocol stack used by the profile. To perform its functions, each profile uses particular options and parameters at each layer of the stack. This may include an outline of the required service record, if applicable.

Most Bluetooth devices are given just a few profiles. For example, a Bluetooth headset will use the Headset Profile, but not the LAN Access Profile which defines how devices use Bluetooth technology to connect to local area networks.

Profile	Name	Use cases
Advanced Audio Distribution Profile	A2DP	Defines the level of the quality of the audio (stereo or mono) that is streamed between devices over Bluetooth. Is capable of transferring unidirectional stereophonic audio stream, for example, music between MP3 player and headset. Relies on AVDTP and GAVDP. Includes mandatory support for the low-complexity SBC codec and supports optionally MPEG-1, MPEG-2, MPEG-4, AAC, and ATRAC.
Attribute Profile	ATT	Wire application protocol for Bluetooth Low Energy specification.
Audio/Video Remote Control Profile	AVRCP	Provides a standard interface to control, for example, TV and Hi-fi equipment. Allows a single device (remote control) to manage a set of different A/V equipment. Has several versions for the support of functionality, state information and images.
Basic Imaging Profile	BIP	Can send images between devices. Has ability to resize and convert images.
Basic Printing Profile	BPP	Allows devices to send, for example, text, emails and vCards to printers. Is based on print jobs. Note that HCRP does not need printer-specific drivers, so BPP is especially suitable for embedded devices like mobile devices and digital cameras which are challenging to update with drivers.
Common ISDN Access Profile	CIP	Provides unrestricted access to the ISDN services, data and signaling.
Cordless Telephony Profile	CTP	Defined for cordless phones so that they function with Bluetooth.
Device ID Profile	DIP	Provides device to be identified more detailed compared to the more limited Device Class that is available already in Bluetooth. Can show the identification of the manufacturer, product ID and version, as well as the version of the Device ID specification being met. Can be used, for example, for downloading needed drivers to the connecting device.
Dial-up Networking Profile	DUN	Provides access to Internet and other dial-up services over Bluetooth, for example, via laptop by utilizing mobile device for the dial up. Based on Serial Port Profile (SPP) and AT command set specified found in 3GPP 27.007.
Fax Profile	FAX	Provides standard FAX functionality between mobile device or fixed-line phone and a computer equipped with fax software, according to the AT commands defined in ITU-T.31 or ITU-T.32.
File Transfer Profile	FTP	Provides browsing, manipulation and transferring of files and folders.
Generic Audio/Video Distribution Profile	GAVDP	Creates the foundation for A2DP and VDP.
Generic Access Profile	GAP	Is used as a base for all other BT profiles. Describes the discovery and connection establishment of Bluetooth units.
Generic Attribute Profile	GATT	Functions for profile discovery and description services for the low energy protocol of BT. Defines the grouping of ATT attributes.

(continued)

Profile	Name	Use cases
Generic Object Exchange Profile	GOEP	Functions as a base for other data profiles. Based on OBEX.
Hard Copy Cable Replacement Profile	HCRP	Functions as an alternative to a cable connection of devices and printers. Please note that printer-specific drivers are required with this profile.
Health Device Profile	HDP	Provides transmission and reception of Medical Device data.
Hands-Free Profile	HFP	Provides communication between car hands-free kits and mobile devices. Uses Synchronous Connection Oriented link (SCO) for the monaural audio channel.
Human Interface Device Profile	HID	Supports devices like mouse and keyboard. Provides a low latency and low-power link.
Headset Profile	HSP	The most typical profile for the connection of mobile device and headset. Based on SCO audio encoding of 64 kbit/s CVSD or PCM. Includes functionality for call indication, answering and termination of call and volume tuning.
Intercom Profile	ICP	Telephone control protocol based profile which is based on SCO for audio. Allows voice call between Bluetooth capable handsets.
LAN Access Profile	LAP	Provides mobile device to access LAN, WAN or Internet via other device that has a physical connection to the network. In addition, allows the device to join an ad-hoc Bluetooth network.
Message Access Profile	MAP	Provides means to exchange messages between devices. Typically used in handsfree environment of cars.
OBject EXchange	OBEX	OBEX is related to OPP. OPP uses the Access Point Interfaces of OBEX profile. The OBEX operations which are used in OPP are connect, disconnect, put, get and abort.
Object Push Profile	OPP	This is a basic profile for sending objects. Examples of these are pictures, business cards and calendar markings. In this profile, the communication is initiated by the sender (client), not the receiver (server).
Personal Area Networking Profile	PAN	Personal Area Networking Profile (PAN) profile allows the use of Layer 3 Bluetooth Network Encapsulation Protocol for the transport of data over Bluetooth link.
Phone Book Access Profile	PBAP / PBA	Phone Book Access Profile (PBAP), or Phone Book Access (PBA). This is a profile providing the exchange of Phone Book Objects between devices. A typical use case is the transfer of data between a car kit and a mobile phone so that the car kit can display the name of the A-number, facilitate the car kit to download the phone book in such a way that the user may initiate a call from the car display instead of the phone UI. This profile contains 2 functionalities: PSE (Phone Book Server Equipment) for delivering phonebook data, and PCE (Phone Book Client Equipment) for receiving data.

Profile	Name	Use cases
Serial Port Profile	SPP	Serial Port Profile (SPP) emulates serial cable in order to replace RS-232. It provides, that is, the control signaling of RS-232. Please note that this profile acts as a basis for DUN, FAX, HSP and AVRCP.
Service Discovery Application Profile	SDAP	Service Discovery Application Profile (SDAP) instructs how applications can use SDP to discover services of remote devices.
SIM Access Profile	SAP, SIM, rSAP	SIM Access Profile (SAP, SIM), or remote SIM Access profile (rSAP) allows devices, for example, built-in GSM car phones to connect to a SIM card of mobile device via Bluetooth. This means that the car phone does not need a dedicated SIM card of its own.
Synchronization Profile	SYNCH	Synchronization Profile (SYNCH) allows synchronization of Personal Information Manager (PIM) items. The roots of this profile are from the infrared specifications (IRDA), and they have been adopted by the Bluetooth SIG to form part of the main Bluetooth specification. For this reason, in practice, the profile is also referred to as IrMC Synchronization.
Video Distribution Profile	VDP	Video Distribution Profile (VDP) makes it possible to transport video stream. The practical use cases include streaming of stored video of PC media center to portable player, and live video from digital video camera to TV set. It should be noted that the support of H.263 is mandatory of this profile, and optional cases include MPEG-4 Visual Simple Profile, as well as H.263 profiles "3" and "8."
Wireless Application Protocol Bearer	WAPB	Wireless Application Protocol Bearer (WAPB) is used for carrying Wireless Application Protocol (WAP) over Point-to-Point Protocol over Bluetooth.

5.9 Hearing Aid Compatibility

5.9.1 T and M Rating

HAC (Hearing Aid Compatibility) for wireless devices, like mobile phones and their accessories refers to the reduction of the interference noise in hearing aids. Furthermore, a mobile phone that has been equipped with HAC is capable of working with a hearing aid's possible telecoil for use with the telephone and with assistive listening devices [6].

In order to use the telecoil, the hearing aid is switched to the "T" position. Alternatively, a telecoil program can be selected form the hearing aid typically via a push-button. Once activated, the telecoil receives the audio generated by telephones via induction. The benefit of telecoils with telephones is that they provide with the hearing aid's volume control by minimizing feedback loop that otherwise disturbs the hearing. Another benefit of the telecoil is that it reduces background noise.

There are two hearing aid ratings for telecoils, that is, M and T. Furthermore, the ratings consist of a no. 3 or 4, that is, the categories can be M3, M4, T3 and T4. Thos markings indicate that the mobile phone in question has been designed to be compatible with hearing aid. The letter M refers to microphone rating and T is designed for telecoil rating.

Mobile device manufacturers are required to provide several hearing aid compatible models in the case when the device is used with hearing aids set to standard microphone setting M3 or M4 since mid-2000s, and slightly later, also the hearing aid compatibility for telecoil mode T3 or T4 has been required in the markets. It can be generalized that the higher the number in question is, the more immune the device is for interferences, that is, the higher the rating, the less likely the practical experience for the interference is.

If the device has either T3 or T4 rating, it must also have either M3 or M4 rating to offer the complete set of interference reduction functionality. Furthermore, the hearing aids also have to been typically rated in M and T scales 1–4 although it is not a mandatory requirement in the markets. There are, for example, various hearing aids with M2/T2 rating available. The new hearing aids might not use, though, RF-immunization.

Especially for the American market, the Hearing Aid Compatibility Act of 1988 (HAC Act) of the FCC requires telephones manufactured or imported for use in the United States after August 1989 to be HAC-compatible. Some years after that decision, it was also decided that mobile phones are part of the rules.

5.9.2 HAC Compatibility Aspects

Hearing aids can operate in two modes: acoustic coupling or telecoil, that is, inductive coupling. The acoustic coupling mode variants work in such a way that they receive and amplify all sounds. The receiving includes both wanted signals like the audio of telephone, but also unwanted signals like background noise. On the other hand, the hearing aids that operate via telecoil coupling mode, are able to reduce the background noise level because they can turn off the microphone and instead receive solely the signals from magnetic fields generated by telecoil-compatible telephones. As an example, in the USA major part of the hearing aids are based on the latter principle.

In practice, it is possible to integrate telecoil with “In-The-Ear” and “Behind-The-Ear” -types of aids, but typically smaller hearing aids do not have room for the telecoil.

The performance of the hearing aids for lowering the unwanted signals also depends on the technology in question. As an example, it is easier to cope with signals of CDMA or WCDMA than GSM due to the nature of the signals, as CDMA is based on a wider spectrum and signals that can be below noise level, whilst the TDMA signals of GSM may cause interference signals in the audible spectrum. With proper design, it is possible to cope with the interference level in any case, regardless of the technology.

As an example of the FCC requirements for hearing aid compatibility, it has been noted that analog wireless telephones did not generate as much interference as digital systems, due to the nature of the signal shapes (analog signal being “rounded,” as in the case of FDMA wave form compared to the impulse type of digital signaling) usually do not cause interference with hearing aids. Thus, FCC has adopted specific hearing aid compatibility rules for digital wireless telephones. According to FCC information, the standard for compatibility of digital wireless phones with hearing aids is set forth in American National Standard Institute (ANSI) standard C63.19. A digital wireless handset is considered hearing aid-compatible for inductive coupling if it meets a “T3” (or “U3T”) rating under the ANSI standard [11]. In addition to rating wireless phones, the ANSI standard also provides a methodology for rating hearing aids from M1 to M4, with M1 being the least immune to RF interference and M4 the most immune. To determine whether a particular digital wireless telephone is likely to interfere with a particular hearing aid, the immunity rating of the hearing aid is added to the rating of the telephone. A sum of four would indicate that the telephone is usable; a sum of five would indicate that the telephone would provide normal use; and a sum of six or greater would indicate that the telephone would provide excellent performance with that hearing aid [11].

The default FCC rules for the mobile devices are the following for acoustic coupling:

- Handset manufacturers must meet at least an M3 rating for one-third of the handset models that it offers to service providers per digital air interface.

- Each nationwide carrier operator must meet at least an M3 rating for 50 percent or eight of the handset models it offers to consumers, whichever is less, per digital air interface.
- Each non-nationwide wireless service provider must meet at least an M3 rating for 50 percent or eight of the handset models it offers to consumers, whichever is less, per digital air interface.

For Inductive Coupling, the default rules are the following:

- Each handset manufacturer must offer at least two T3-rated handset models per digital air interface. In addition, manufacturers have to ensure that one-third of their handset models per air interface meet at least a T3 rating.
- Each wireless service provider must meet at least a T3 rating for one-third or ten of the handset models it offers to consumers, whichever is less, per digital air interface.

The more specific rules and exceptions can be found in the FCC web pages [11]. As an example, manufacturers and service providers have been required to post information about their hearing aid-compatible handset offerings on their websites since 2009.

5.9.3 TTY/TDD Compatibility

TTY refers to a TeleType device, and TDD means Telecommunication Device for deaf people. A TTY/TDD solution helps deaf people or the hard of hearing, as well as the speech-impaired, to utilize telephone devices for the communication. This is done via text messages. In order to function, the solution requires a TTY/TDD functionality for both sender and receiver ends [6].

The solution is applicable for fixed line as well as for mobile communications systems as long as the devices are compatible with the solution. The difference with text message service (SMS) is that TTY/TDD provides synchronous textual conversation, comparable to a phone call that is performed via writing and reading text.

Text telephone devices (TTY and TDD) are used by people with hearing or speech disabilities in order to send and receive text messages over telephone networks. Wireline telephone and analog cellular networks were typically compatible with TTYs, but digital wireless networks caused new challenges. The rules have been updated accordingly in order to provide TTY users with a complete emergency 911 calling. Consumers are thus typically able to use TTYs with digital wireless phones, including 911 calls, if the phone is TTY-compatible. The requirements for being able to send SMS message via the 911 emergency number are also being discussed, and, for example, Canada will require compatibility from mobile equipment manufacturers and operators to initiate 911 calls via SMS based on a predefined list of users who are registered to use this type of service.

More information about hearing aid equipment, compatible mobile phones and other relevant information can be found in Ref. [12], and the following links: [13, 14].

5.10 Other Connectivity Technologies

5.10.1 G.V2A

G.V2A is an ITU-T Recommendation on an automotive interface for applications external to the vehicle gateway. The G.V2A defines communications interface between external applications and a Vehicle Gateway Platform (VGP). It also enables applications running on nomadic devices like mobile phones and portable music players, and in the cloud (e.g., navigation servers) to interact with drivers in a safe way. Furthermore, it allows external applications to leverage the user interface of the vehicle platform, and it enables vehicle platforms to control the timing and format of all application messages to the driver.

5.10.2 MirrorLink

5.10.2.1 General

MirrorLink is based on a set of nonproprietary technologies [15]. It uses standard Internet technologies such as Internet Protocol for compatibility with various devices. It also uses technologies already common in the car environment such as Bluetooth and USB and Wi-Fi. MirrorLink is based on the Universal Plug and Play (UPnP) for controlled access to applications. Furthermore, Virtual Network Computing (VNC) replicates the handheld device's display on the navigation screen of the car and communicates the user inputs back to the phone. In addition to Bluetooth, the audio is also possible to deliver via the Real-Time Protocol (RTP).

Furthermore, MirrorLink provides a mechanism ensuring that solely those applications that are officially approved are accessible while driving. The applications will be approved via a standardized test process.

5.10.2.2 Certification Program

MirrorLink certification is based on MirrorLink Specification Conformance, client/server interoperability and a review of a series of administrative requirements.

Certification Program Processes works in such a way that the certification is granted to a product of a specific manufacturer, product name, model number and revision number. Any change to the identifying markers of a product renders the product not certified. Only the original device manufacturer (ODM) may apply for certification. If a product is to be relabeled or remarketed for use by another member company, the processes contained in the Program Management Document must be followed.

For more detailed information about mobile payment technologies, please refer to [16–23]. For hearing aid details, please refer to [24–30] and for USB connectivity, please refer to [31–37].

References

1. IP connectivity, basics. http://en.wikipedia.org/wiki/Internet_connectivity (last accessed 25 July 2014).
2. Universal Serial Bus Revision 3.0 Specification: http://www.usb.org (last accessed 25 July 2014).
3. Super Speed USB: http://www.usb.org/press/USB-IF_Press_Releases/SuperSpeed_10Gbps_USBIF_Final.pdf (last accessed 25 July 2014).
4. On-The-Go and Embedded Host Supplement to the USB Revision 3.0 Specification. Revision 1.1. 10 May 2012. http://www.usb.org/developers/docs/ (last accessed 25 July 2014).
5. NFC Forum specifications: http://www.nfc-forum.org/specs/spec_list/ (last accessed 25 July 2014).
6. Hearing Aid Compatibility. http://www.pagepluscellular.com/ (last accessed 25 July 2014).
7. http://www.nfc-forum.org (last accessed 25 July 2014), NFC Forum's official web portal.
8. Cherian Abraham, Principal Analyst, Mobile Commerce & Payments Practice / Co-Founder, DROP Labs. Mobile Payments. A study of the emerging payments ecosystem and its inhabitants while building a business case. 36 pages.
9. http://www.paywithisis.com (last accessed 25 July 2014), ISIS web portal.
10. Bluetooth: http://www.bluetooth.com (last accessed 25 July 2014).
11. FCC information for hearing aids http://www.fcc.gov/guides/hearing-aid-compatibility-wireless-telephones (last accessed 25 July 2014).
12. CTIA – The Wireless Association. http://www.accesswireless.org/Home.aspx (last accessed 25 July 2014).

13. HAC information, FCC: http://www.fcc.gov/encyclopedia/hearing-aid-compatibility-and-volume-control (last accessed 25 July 2014).
14. HAC information, HearingLoss: www.hearingloss.org (last accessed 25 July 2014).
15. Mirrorlink. http://www.mirrorlink.com/ (last accessed 25 July 2014).
16. ISO/IEC 14443 proximity-card standard.
17. NFC Forum Architecture. NFC ForumTM. NFCForumArchitecture_1.2, 2009-SEP-02. 13 p.
18. High Level Conformance Requirements (HLCR) of NFC Forum.
19. Standard ISO/IEC 18092:2004.
20. Web page of NFC Forum: http://www.nfc-forum.org/home/ (last accessed 25 July 2014).
21. Wb page of NFC Forum specifications: http://www.nfc-forum.org/specs/ (last accessed 25 July 2014).
22. Press release of Isis, Austin, TX, Dec 12, 2011 (http://news.paywithisis.com/) (last accessed 25 July 2014).
23. EMVCo certificate: https://www.emvco.com/ (last accessed 25 July 2014).
24. Figure of Jim Stapleton, Isis, presented at RAMP Advanced Commerce & Mobile Retail Services Summit: http://news.paywithisis.com/wp-content/uploads/2011/12/Isis-Travel-made-easy.png (last accessed 25 July 2014).
25. Isis web page: http://news.paywithisis.com/ (last accessed 25 July 2014)
26. JVL Secure Element Requirements. Version 1.01. External distribution. Issued: 01/10/2011. 49 p.
27. Update on NFC Forum Activities. Gerhard Romen, Nokia, chairman of NFC Forum Marketing Committee. Presentation material, 30 November 2006. 17 p.
28. USB HUB design principles. http://electronicdesign.com/boards/use-these-design-and-test-methods-create-robust-usb-devices (last accessed 25 July 2014).
29. USB standards. http://www.usb.org/developers/ (last accessed 25 July 2014).
30. HAC information, AccessWireless: http://www.accesswireless.org (last accessed 25 July 2014).
31. EMVCo - A Guide to EMV. Version 1.0. May 2011. 35 p.
32. Google Wallet vulnerable to "brute-force" PIN attacks (update: affects rooted devices): http://www.engadget.com/2012/02/09/google-wallet-open-to-pin-attacks/ (last accessed 25 July 2014).
33. Description of Google Wallet: http://www.google.com/wallet/what-is-google-wallet.html (last accessed 25 July 2014).
34. Announcement of Visa's mobile payment: http://www.bgr.com/2012/02/27/visa-announces-new-mobile-payment-solution/ (last accessed 25 July 2014).
35. Public overview of WP and ISIS solutions: http://www.geek.com/articles/mobile/windows-phone-8-incoming-heres-what-you-need-to-know-20120620/ (last accessed 25 July 2014).
36. http://www.emvco.com (last accessed 25 July 2014), Global standard for credit and debit payment cards.
37. FCC, TTY. http://www.fcc.gov/guides/use-tty-devices-digital-wireless-phones (last accessed 25 July 2014).

6

Fixed Telecommunications Networks

Jyrki T. J. Penttinen

6.1 Introduction

The fixed telecommunications networks have a long history. Since the beginning of the networks – which at the first phase were only dedicated point-to-point lines – they were utilized for telegrams via Morse codes. The voice was adapted to the wired environment after the invention of telephone devices. There have been various persons contributing to the initial innovation and enhancements of telephone equipment and systems in 1800 century. The first steps towards the communication lines were taken in 1870s. Since then, the wired lines and networks have been utilized for voice calls along with the deployment of the commercial systems. The voice is still an essential service, although the end-user's connection point has been moved from the fixed lines towards wireless access in a very fast transition period during the 1990s, and remarkably towards IP solutions as of 2000s.

6.2 Network Topologies

The elements of the fixed telephone network can be connected via various principles. The most important ones are bus, ring and star topologies. There are many other variants, too. The extreme representatives of these are the straight line connecting directly two devices, and at the other extreme, the open Internet with all the possible systems connected via IP.

The quality of service (QoS) has been one of the most important aspects of voice telephony networks. For that reason, it has been typical to assure functioning of the most critical nodes and elements by duplicating the hardware modules, complete elements, interfaces and other parts of the system. The assurance can be done, for example, via 2n principle (doubling the elements or functional blocks).

The simplest format of telecommunication network topology is point-to-point, that is, direct connection of two devices, or nodes. As there are more devices, they can be connected in bus topology, as is done in the LAN solutions. The devices can also be connected in star and ring topologies.

The Telecommunications Handbook: Engineering Guidelines for Fixed, Mobile and Satellite Systems, First Edition.
Edited by Jyrki T. J. Penttinen.

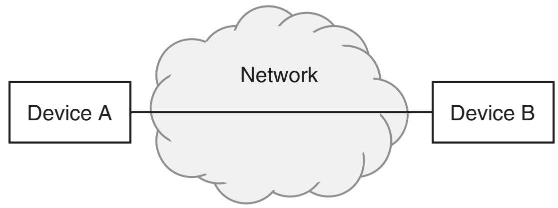

Figure 6.1 *Point-to-point topology.*

The network standards typically dictate the topology, or set of topologies that are possible to adapt in the respective deployments. As an example, LAN, that is, Ethernet is typically deployed according to the bus topology although it can also be implemented in a star or ring topology by applying interconnecting equipment for these solutions. Some networks, like Token Ring, are logically only applicable in the ring topology due to the fundamental idea of the system. In practice, the combination of different topologies is also possible, as is the addition of several same topologies in a chain.

The ITU-T Blue Book [1] is a reference for the appropriate guidance on telecommunications reforms and restructuring. Relevant to this chapter, references [2, 3] summarize background information for the previous development of exchange technologies, and current cloud based technology description can be found in [4]. Reference [5] summarizes the LTE related environment.

6.2.1 Point-to-Point

This network type was basically the very first solution for the telephony system. The limitations of it mean that it is used merely in special solutions as in temporal military base connections. Figure 6.1 shows the principle of point-to-point topology.

6.2.2 Bus

The bus topology is straightforward to install, and its maintenance is simple. The devices can be easily added and removed in this environment. The drawback of bus topology is that all the data and signaling is routed via all connection points, that is, the network handles the whole traffic generated by the users. For this reason, bus topology represents broadcast networking. This creates capacity limitations, and increases potential security risks such as eavesdropping as well as problems to the service when a break in the physical wiring happens. The latter can result in isolation of a complete bus area. Figure 6.2 shows the principle of bus topology.

The initial phase of bus topologies in LAN environment was based on coaxial cables. In LAN environment, the data rate is bi-directional. The data rate has been growing since the initial values of 10/100 Mb/s, nowadays 1 Gb/s being typical value.

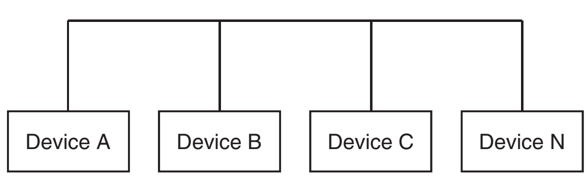

Figure 6.2 *Bus topology.*

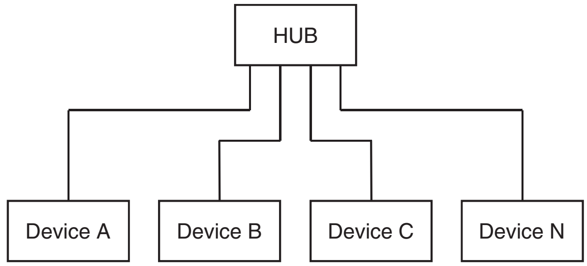

Figure 6.3 *Star topology.*

The method of the resource reservation typically is based on medium access control (MAC) that utilizes CSMA (Carrier Sense Multiple Access). In the worst case, this method leads to collisions of the data packets and retransmissions which lower the useful capacity of the system. In addition, the limitation of the original LAN bus networks has been 2.5 km, or 1.5 miles, with additional limitation of a single segment having been a maximum of 500 m. This is due to the relatively high attenuation of high-frequency signals.

Networks that use bus topology are defined in IEEE 802.3. A single segment of the LAN bus supports $2^{10} = 1024$ network addresses.

6.2.3 Star

Star topology as shown in Figure 6.3 allows communication between specific groups of devices/nodes in such a way that other devices/nodes are not aware of the communication. This means that the security of star topology is higher than in bus topology. If one leg of the network is damaged, it would not affect the functioning of the rest of the network.

There is a central node in the star topology to which all participating nodes/devices are connected. In practice, the central node is a hub, switch or router. The data rates of star topology vary between 10 and 100 Mb/s, based on 10Base-T and 100Base-T. The physical connection of the participating devices/nodes can be unshielded twisted pair (UTP) or Shielded Twisted Pair (STP), which is at the same time one of the most significant benefits of the star topology. Unlike in the bus solution, multiple users can take full advantage of the capacity at the same time in the star topology. The drawback is logically the event of the central hub failing, which breaks down the whole star network.

6.2.4 Ring

As with bus topology, ring topology as shown in Figure 6.4 also represents the broadcast type as the traffic is routed via all the nodes/devices. Nevertheless, ring topology provides high reliability in contrast to bus topology, yet with a low amount of wiring. In this topology, each one of the devices/nodes has two connection points to the infrastructure. If there were breaks in any part of the ring, it would not affect the functionality of

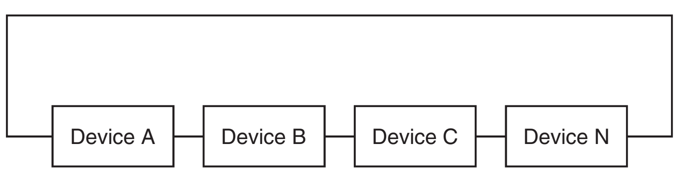

Figure 6.4 *Ring topology.*

the rest of the network. The drawback of this topology is that the devices must be designed especially for the ring topology in order to function. Another drawback is that when the nodes/devices are added or removed, the whole network must be interrupted.

The transmission happens only in one direction in the ring topology, around the closed loop of the network. Each node/device receives the packets and relays them forward, until the correct receiving end is found amongst all the nodes/devices.

The physical solution is typically based on coaxial cable or fiber optics. The most widely known system utilizing the ring topology is Token Ring network to be used as a variant of Local Area Network (LAN).

Token ring LAN is in practice a protocol which is located to the data link layer (DLL) of the OSI model. Token Ring LAN is based on a three-byte frame, or token. The token "travels" within the ring until the destiny station is found. Token-possession grants the possessor permission to transmit on the medium. The Token Ring has been standardized in IEEE 802.5.

When no information is sent at the moment, empty information frames circulate on the ring. As soon as some device/node has a message to be sent out, it inserts a token in an empty frame. This token is a field in the frame that indicates the existence of the message. Also the actual message is inserted in the same frame, together with the destination identifier.

As a next step, the frame travels to the following devices/nodes in order. Each device examines the frame. Finally, the device to which the message is sent copies the contents of the frame, and changes the token back to empty, to indicate that the frame is not active any more. At this point, the contents are not deleted until the frame again reaches the original sender, who makes a note about the copied and received frame by the receiving end. At this point, the original sender removes the message from the frame.

As of this point, the empty frame continues its circulation, and is thus ready to be filled in by the device that has something to be sent out within the network. It should be noted that the principle of the token can also be used with bus topology LANs.

IEEE 802.5 working group standardized Token ring LAN speeds of 4 Mbit/s and 16 Mbit/s. Also data rates up to 100 Mbit/s was standardized and marketed. Even 1 Gb/s was finally standardized, but equipment of this rate was never marketed. Finally, the Toke Ring disappeared from the markets.

6.3 Redundancy

There have been many examples in public about major problems if some of the essential elements or interfaces have not been assured. The concentration of the capacity to fiber optics has sometimes caused a situation where whole cities are served via single core cabling. If this is damaged, the consequences can be dramatic for the immediate functioning of telecommunications, and nondirectly even for the economy of the country. As is the case for any other technical equipment, the functioning of telecommunications connections and elements are not possible to guarantee 100%. The outage times are typically regulated by telecommunications authorities at the national level. Depending on the areas, the acceptable limit of nonfunctional services can be, for example, in the range of 99.9999%.

In order to measure service outages, the commonly adopted term is the mean time between failures (MTBF) as shown in Figure 6.5. This represents the estimated elapsed time between inherent failures of a system during operation. MTBF is calculated based on the average time between failures of a system. The assumption of MTBF is that once it has occurred the failure is repaired according to the mean time to repair (MTTR). As a corollary, the mean time to failure (MTTF) indicates the average time to failures assuming that the failure is in fact not repaired.

The Internet has been changing the traditional way of dimensioning and deployment of telecommunications networks. The VoIP service is more popular on the Internet between clients, and also telecommunications

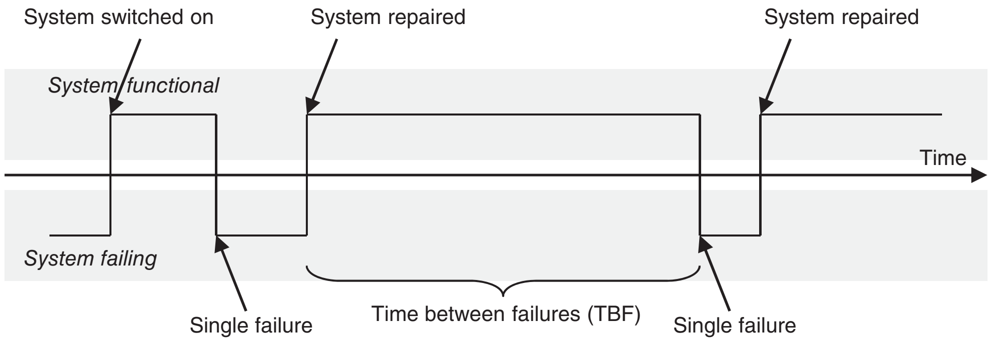

Figure 6.5 *The principle of TBF (Time Between Failures).*

operators have changed the way of work in this previously very tightly closed business. Despite its superior flexibility over traditional circuit switched systems, VoIP also has some fundamental drawbacks like limitations in the quality of service guarantees.

The benefits of IP are related to network management, and more specifically, it's automatization. The IP core is basically a set of countless routers and bridges, which are capable of automatically and dynamically rerouting the connections when there are load issues or failures in the elements or interfaces. The drawback of this extremely flexible functionality is that some failed elements can be left to the networks causing, for example, occasional high duplicated and failed packet transmissions because the rest of the network is capable of omitting the wrong packets. These elements may occupy capacity unnecessarily and consume energy in the ever-growing IP network infrastructure.

The bus topology is commonly utilized in local area networks. Ethernet is the most commonly known example of bus topology. Bus topology is easy to deploy, and its possible drawbacks are related to limitations of challenging principles as the transmitting stations are trying to initiate data sending, which can result in overloading of the network. Another limitation of bus topology is the limitation of security and confidentiality because access of external parties into the bus might be challenging to observe. Eavesdropping is thus relatively easy to do in this type of network because all traffic goes through the point where external person connects the equipment. The solution for increasing security in these cases is the utilization of some higher level scrambling system, like VPN (Virtual Private Network).

In ring topology, each user is connected to the neighboring party. Some examples of ring topology are Token ring and FDDI. Star topology contains concentrator and terminals connected to it. Star topology is utilized, for example, in the Ethernet variant of 100 Mb/s.

In practice, the telephone network can be done by deploying many topologies, and a mix of principles, for example, as a combination of star and ring topologies. In all cases, the assuring of the correct functioning of the network is an important task of the telecommunications operator.

6.4 Telephone Network

A typical way to deploy the "old-fashioned" circuit switched telephone network is shown in Figure 6.6. National calls are handled by the interworking exchange elements, and international traffic is directed via the specially selected exchanges that are interfacing with neighboring countries and international telecommunications networks. Also cellular networks are connected to the national and international infrastructure via the exchange layer.

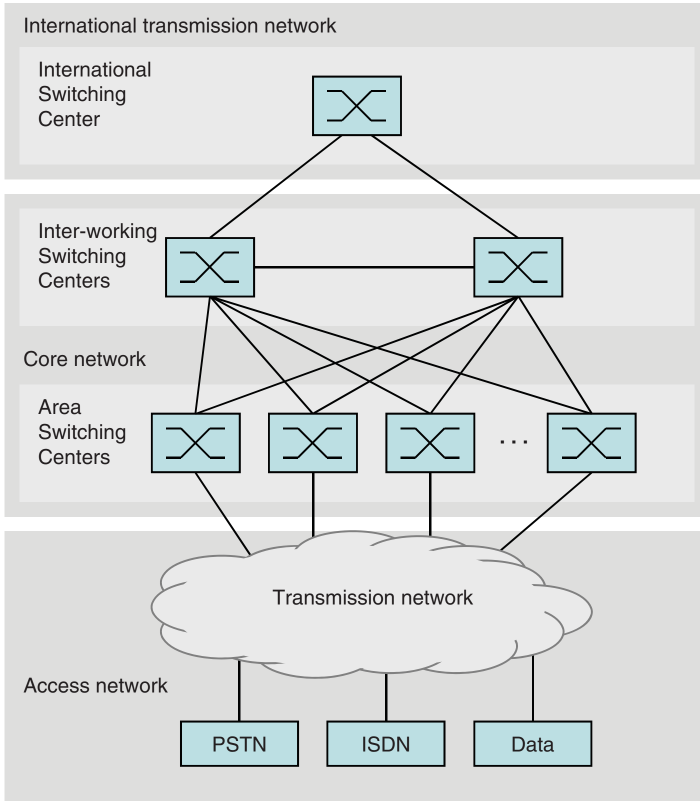

Figure 6.6 *A typical example of the telecommunications network solution at the national level.*

Under the interworking exchanges, there is a set of regionally arranged local exchanges. Under those, there is still the actual access network layer, which is physically constructed until the final connector of end-users.

6.5 User Devices

6.5.1 Telephones

6.5.1.1 Analog Dialling Plate

In the era of manual exchanges, voice calls were initiated by the end-user by the movement of an induction system of the device which made the bell ring in the exchange site. The personnel activated the line with the originator of the call in order to communicate via voice with whoever the subscriber would like to contact.

Later, the automatic telephone network's terminal equipment was based on the dialing plate. When first selected, the number is returned to the initial position in such a way that it generated pulses by cutting the line in intervals corresponding to the number dialed (from 1 to 9, number 0 being the last and generating 10 pulses). The exchange interpreted the pulses as a series that indicated the complete telephone number of the B-subscriber, and commuted the connection to the B-subscriber via the relays. This extremely simple

Figure 6.7 An example of wall-mounted analog telephone equipment, the inner parts exposed.

principle was utilized for decades until the digital systems have been replaced the analog relay exchanges in the commercial networks.

The current terminals (TE, terminal equipment) of the digital telecommunications systems are typically based on DTMF (dual tone multi frequency) dialing keyboard. It functions as indicated by the term by generating two parallel tones when the keys are pressed. The combination of these two different frequencies of tones is interpreted as numbers in the exchange side. As the voice calls have been originally designed to be delivered in the frequency range of 300−3400 kHz, these tones fit into the system's frequency band.

The telephone equipment is principally meant for voice calls, but there is a wide range of other equipment that can be utilized via the telephone network, as answering machines, modems, television sets, facsimile equipment, and short range wireless terminals. Telecommunications equipment is possible to utilize as a part or connected to a computer or PDA (personal digital assistant). The equipment that is meant to be used especially for data transfer is called DTE (data terminal equipment). Figure 6.7 presents the structure of analog telephone devices.

Although their importance is decreasing all the time, there are still places where public telephones can be found. As an example, in Latin America the infrastructure of public telephone equipment was modernized at the beginning of the 2000s.

6.5.1.2 DTMF

When the user selects the B-number, the delivery of these must be done in such a way that the network understands the numbers in a globally uniform format. The established way to do this is the utilization of DTMF (Dual Tone Multi-Frequency). In selective DTMF dialing, the numbers and special characters are transmitted to the network via a string of digits. This method can be utilized in both fixed telephony networks

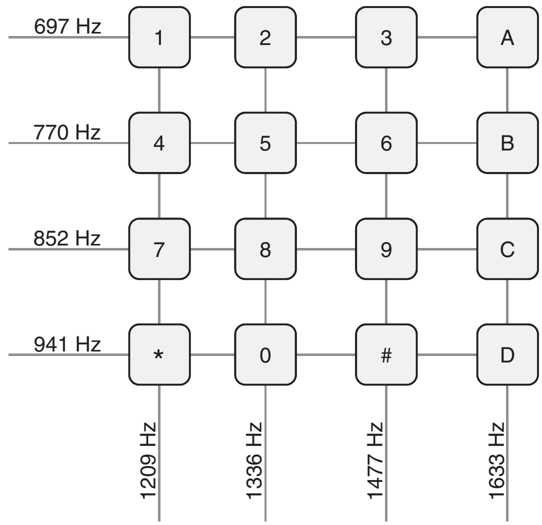

Figure 6.8 *The symbols and respective frequency pairs of DTMF.*

as well as in mobile communications networks. As the term indicates, DTMF produces, when the user presses the dialing key, a tone that consists of two separate sinusoidal frequencies, which the network commutation element can map into the numbers and characters of ∗ and #. Figure 6.8 shows the combination of frequencies for each dial key.

The actual physical method and user interface can vary in practice, as long as the result is the audio frequency formed by the DTMF pairs of tones. This means that dialing is possible whilst the user is able to produce frequencies, for example, via voice recognition methods which makes the use of the physical dialing pad unnecessary.

6.5.2 Data Equipment

6.5.2.1 Modems

Modem (modulator/demodulator) is still useful equipment for the data transfer in the home environment as well as in industry. The form of modem has changed from its first commercial variants to more advanced ones, and they support much faster data rates and access methods. The idea has remained the same, though, and the modem thus adapts the bit stream of the user to the transmission line. Before, the modulation was done for the analog format of telephone lines in such a way that the voice call bandwidth of 300–3400 Hz was utilized. Nowadays, along with complete digitalization of the telephone network infrastructure, adaptation is done to the digital systems.

In addition to adaptation of user data to the network, modems also take care of error correction between the transmitting and receiving end. The base band modems, in turn, are meant for fast data transmission in relatively short distances, and are used, for example, in connecting local area networks with each others.

The modems can be divided into 2- and 4-wire versions. The 4-wire modems are duplex variants in such a way that both transmission directions have their own 2-wired connection. The system for 2-wired variant is, in turn, simplex, half duplex or full duplex. The full-duplex modems can be further divided into symmetric and asymmetric variants.

The latest version of the basic modem variant that is utilized via the circuit switched data connections over the telephone network is defined in ITU-T V.90 specifications. It provides a data rate of 56 kb/s in reception and 33.6 kb/s for transmission. Along with the modernization of the telecommunications networks, these basic modems have gradually disappeared for digitalized markets, but are still used in countries where analog telecommunications networks are found.

The AT commands (attention) of modems are used in the initiation of connections, maintaining and releasing them. The basic principle of modem connections via the circuit switched networks is to create the data connection between two modems. The A-subscriber can connect either directly to the modem of the B-subscriber, or to the modem of a network service.

During the time the AT command set has been extended for supporting new functionalities of the services. In addition to fixed telecommunications networks, the data services of mobile networks like GSM and UMTS, along with the packet switched data (GPRS, General Packet Radio Service) can also be utilized via AT commands. As an example of GPRS data connection, the initiation of the data call can be made by typing "ATD" (AT command for dial) with GPRS-specific "number." It is a parameter value of "∗99#" which the packet data network is able to interpret as a command for the packet data connection, resulting in the routing of the call to the packet data network. It should be noted that in the GPRS connection, there are actually no modems situated in the GSM core network in the traditional meaning of the terminology, like in the circuit switched connections were the case with the modem pool, but instead, the Packet Control Unit (PCU) separates the GPRS data connections in the BSC, and directs connections to the GPRS core network. In this sense, the user's device is "fooled" to believe the AT-commends are for dialing in the traditional way although the actual target of the command is the GPRS network's Access Point (AP) that leads into the public Internet.

6.5.2.2 xDSL

DSL (digital subscriber line) equipment is designed for environments where the last mile of the connection would otherwise experience problems via traditional modem and cobber line. There are various different versions of DSL, so xDSL is utilized as a general term for describing the overall concept. In addition to traditional DSL modems, there are also solutions for cable television network and optical fiber. The popularity of the latter version is growing fast, but whilst waiting for advances in the optical infrastructure for last mile solutions, the traditional xDSL is also useful. The xDSL variants are the following:

- ADSL (asymmetric digital subscriber line)
- HDSL (high bit rate digital subscriber line)
- SDSL (symmetric digital subscriber line)
- VDSL (very high speed digital subscriber line)
- IDSL (ISDN digital subscriber line).

ADSL is, as the name suggests, a system that offers different bit rates depending on the transmitting direction of the connection. The usability or maximum length of this last mile solution depends on the thickness of the cable. As an example, 0.5 mm cable is capable of transferring 2 Mb/s data rates over 5 kilometers, whilst 0.4 mm cable variant and 6 Mb/s rate the length is only half of the other one. When ordering the ADSL service, it is thus recommended to review the maximum length to the nearest exchange, in order to estimate the achievable data rates.

There is an ADSL modem on both sides of the connection. The first ADSL channel is meant for voice traffic, and its bandwidth is fixed to 0–4 kHz. The middle band is of 4–100 kHz and is meant for bidirectional transfer. The high band is of 100 kHz–1.1 MHz and is designed for reception of a high data rate bit stream. Figure 6.9 shows the band division of the ASDL system.

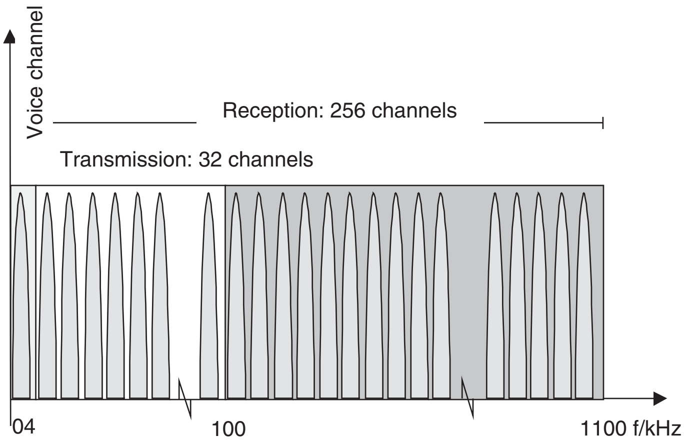

Figure 6.9 *ADSL consists of three bands. This figure presents the DMT version of ADSL.*

The voice channel of the line thus remains usable, regardless of the presence of ADSL. Nevertheless, the voice channel must be separated from ADSL frequencies by applying a separate filter, which in practice is connected to the telephone connector in the user premises. This makes it possible to utilize both voice and data services in a parallel way.

In the DMT variant of ADSL (discrete multi tone line coding), all the channels of all the ADSL bands are fixed to the blocks of 4 kHz each. In the middle band, there is a total of 32 channels, and in the high data rate, high frequency band the number of channels is 256. The functioning of the channels is, in practice, dynamic. Furthermore, the channels are independent from each other in such a way that the maximum offered capacity of each channel is 15 bits per symbol.

Another typically utilized ADSL variant is CAP (carrierless amplitude and phase modulation). It is similar, but simplified version of QAM. As the final modulation method is not explicitly standardized, there are different equipment versions of CAP.

HDSL is a system which can replace PCM transmission with the user lines. When unbalanced cobber cable is utilized, the practical achievable data rates are in the range of 1–2 Mb/s, with typical distances of 3–5 km.

SDSL provides up to 3 Mb/s data rates. SDLS supports equal data rates in both transmission directions. Nevertheless, the system has only gained popularity in the USA.

VDSL, that is, VHDSL offers much higher data rates compared to the above mentioned ones. The system functions either in a single connection, or via various cobble lines, and offers up to 55 Mb/s data rates in a typical environment. In indoor networks, achievable data rates can be in the range of around 155–622 Mb/s. Nevertheless, VDSL is still in an early phase.

Lastly, ISDL is a hybrid solution, that is, a combination of ISDN and DSL. The difference of this system, compared to the ISDN, is that ISDL bypasses the central tone-based network and redirects data streams to a separate network infrastructure.

6.5.2.3 Cable Modems

Along with xDSL solutions, also cable modems are rapidly growing solution for the offering of high-speed data to the households. Cable modems provide several Mb/s data rates in reception, and are relatively simple to connect into the CATV (cable television) via coaxial cable.

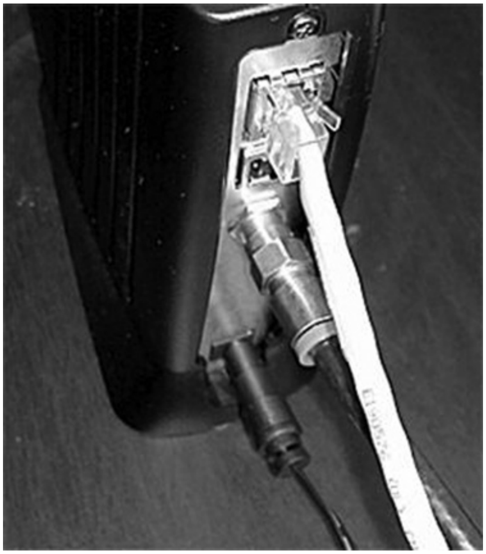

Figure 6.10 The cable modem is simple to connect and set up. The connection to the PC or laptop is done typically by utilizing LAN or USB cable, and the connection up to the server of the service provider happens via the television cables.

The typical solutions of cable modems can be done separately for the reception, meaning that the return-channel must be constructed, for example, via the modem connection of telephone network. Nevertheless, modern solutions are based on bidirectional transfer via the same physical cable.

In addition to CATV systems, data connections can also be done via the public electrical network. This solution provides data transmission in a very wide area. Figure 6.10 shows connectors of typical cable modem equipment.

6.6 Plain Old Public Telephone System (POTS)

6.6.1 General

The Plain Old Telephone System (POTS) or Public Switching Telephone Network (PSTN) has served fixed telephony customers for decades. Its principle was useful for circuit switched voice calls, as well as for data transmission in an analog way. For data transmission, a modem (modulator / demodulator) was needed in order to transfer the bits over analog lines.

The role of the POTS/PSTN in its analog format has practically disappeared along with the fast development of digital systems. The digitalization of the telephone networks has reached practically 100% level globally. In addition, the role of mobile telecommunication has taken over the voice service of fixed networks. It is logical, as the communication is not any tighter to a certain location. Furthermore, the ever growing set of additional services makes the mobile communication incomparable with old systems. It can be estimated that the role of fixed telephone systems changes dramatically, offering principally data services from where voice would be only one data service amongst the whole set of other type of services like multimedia communication. Furthermore, the general trend has been the decrease of subscription and communication costs.

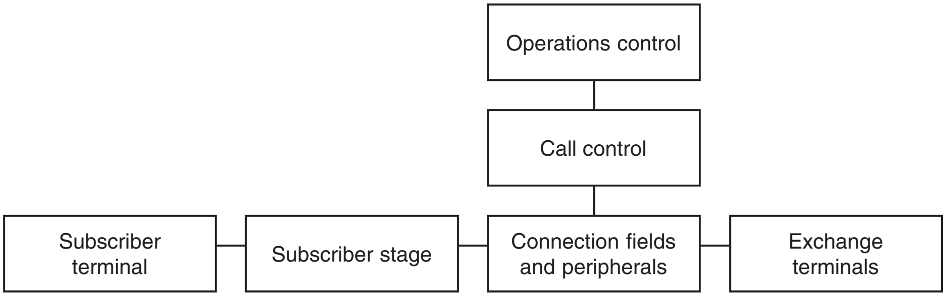

Figure 6.11 *The block diagram of telephone exchange.*

6.6.2 Numbering of Telephony Networks

The international functioning of the telephone system is based on standardization of subscriber numbers and communications principles. The uniformity of the systems has been taken care of by the ITU (International Telecommunications Union). This guarantees that the public telephone networks are interoperable with each others. The numbering plan for the global method for designing the numbers are presented in the Blue book of ITU-T.

6.6.3 The Principle of Telephone Networks

The fixed telephone network contains telephone exchanges and switching systems, which form network nodes.

The common principle of the systems is that they consist of a combination of hardware and software). As the role of software is increasing considerably, one of the most important maintenance aspects is related to SW upgrades and corrections.

The functioning of modern digital switches is based on the frame structure of time division Pulse Code Modulation (PCM). Typically, the signaling of the telecommunications is made via ISDN user part (ISUP).

The other typical aspect in modern communications is that interfaces are standardized heavily, but actual functioning and architecture of each element is manufacturer-dependent. This includes also the principles of redundancy in order to cover the faulty situations due to, for example, HW or SW breakdown. Typical solutions are doubling or creating N parallel functional blocks.

The general block diagram of the telephone networks can be seen in Figure 6.11. The most important functionality of the network is logically to deliver the calls via teleservices, as well as to provide additional services. The basic building blocks for the connection are the user part and signaling part. The combination of these provides the possibility to deliver voice and data traffic, as well as signaling. If the users are connected directly to the exchange, its term is Local Switching Center (LSC). An exchange that delivers the traffic of LCS elements is called transit switching center (TSC) or gateway switching center (GWSC). Furthermore, the International Switching Center (ISC) handles the traffic between the national and international level.

6.6.4 Billing Principles

The billing can be done in various manners via the telephone network. The basic principle that has been utilized traditionally is based on the circuit switched mode of operation, which makes it straightforward to charge based on the elapsed time of the call, with the resolution of pulses that may occur, for example, every second or minute. The actual price level of the pulse can be varied depending on the hour and week day, season and so on. Some operators have provided local calls free of charge (based on, e.g., a flat monthly fee).

Pulse based charging has many variations. One of the possibilities is to charge the establishment of the connection, or alternatively already when the call was tried. Then, the consecutive time is charged based on the pulses that have occurred. The first pulse can occur either randomly or after a certain constant elapsed time period.

Along with the heavy transition of telephone networks towards IP based solutions the charging principles have also changed. In fact, VoIP calls were already possible a long time ago via public Internet service providers. The possibilities for charging have amplified at the same time as packet data services have grown, and it is thus possible to charge based on, for example, the transferred bits and the quality of service level offered. Again, at the same time, the trend is that the cost of calls for end-users is getting into considerably lower level.

Some commercial examples of Internet voice call providers are Skype and MSN messaging. The services use Internet voice call via the computer or a separate device that contains a voice codec, or set of voice codecs that are adopted according to the situation, lower bit stream providing less quality but saving at the same time most of the bandwidth.

The highest quality IP voice call system requires a data rate comparable with the basic ADSL bit stream. As an example, the early phase basic GPRS would not have been able to always offer sufficient capacity per user for high-quality VoIP connections. On the other hand, as an example of the other extreme, the PoC or PTT (push-to-talk over cellular) is based on the very low bit stream. This corresponds to approximately 8–9 kb/s for the delivery of voice traffic.

Nowadays, there are various applications on top of the actual networks that can be utilized for VoIP calls – whilst both ends have the same application and codecs in use. These can be utilized, if network does not support, for example, IMS or other network-based VoIP services.

6.6.5 Comparison of Current VoIP Solutions

6.6.5.1 VoIP in Fixed Network

VoIP typically refers to communication protocols and related techniques for providing transmission of voice communications as well as multimedia contents over IP-based packet data networks. In this context, VoIP can be understood as a synonym for Internet telephony and IP telephony. Other commonly utilized term is Voice over Broadband (VoBB).

Voice is only one subset of the possible data that is transferred over the Internet. Other services that are typically utilized in fixed and mobile telecommunications networks include facsimile, short message service (SMS), and other messaging services like MMS (multimedia messaging service), can also be transferred over the Internet instead of PSTN (Public Switched Telephone Network).

The actual techniques and platforms differ greatly in IP and PSTN, although the basic principle is comparable. In both cases, to initiate, for example, voice calls requires initial signaling and building up the actual data channel where the contents are then transferred. For making the transfer of the original contents in digital network, the analog presentation of the information must first be digitalized, and the voice should be encoded. In IP networks, the data stream is then encapsulated into packets that are possible to deliver via routers between the origin and destination. Once received on the other side, the reverse procedure takes place in order to present the original analog format to the end-user.

As for the terminology, IP Telephony and VoIP, it can be generalized that IP telephony is a solution for digital telephony systems that are based on IP protocols for voice contents delivery, whereas VoIP is a special case of IP Telephony for general IP networks.

As a general note, the initial solutions of VoIP were based on digital telephone networks. After that, the first solutions started to appear with optimized methods for still relatively closed IP networks, which in any case, at the same time, impacted largely on traditional technoeconomical models of traditional voice service provision. One of the early examples of these services is Skype that started to offer free calls between Skype direct users in such a way that there were still no possibilities to utilize access to other networks.

The next step has been the introduction of a concept of Federated VoIP. This is a clearer intention to change the architecture of legacy networks. These new solutions may allow interconnections freely between two domains of the Internet independently of the circumstances, that is, users are freer to initiate calls.

The common procedures of VoIP solutions include session control, and the setup and termination of calls via a set of protocols. The environment for VoIP also requires one or a set of audio codecs in such a way that the common codec is selected in the initial negotiation of system parameters before the actual call takes place. The codecs basically convert the original voice to an audio stream that can be de-encoded at the other end. There are various codecs available in VoIP environment, and each one has been optimized for a certain connection like narrow-band, broadband, highly erroneous or error-free connections and so on. The achieved voice quality depends on the bandwidth as well as compression and error correction capabilities. The selection of the suitable codecs is one of the countless tasks of network operators as too high quality calls require excess of capacity. The historical aspects are also taken into account, as part of the codecs also support PCM codecs that are based on μ-law and a-law variants, that is, G.711 and G.722 definitions of ITU-T.

6.6.5.2 Mobile VoIP

Along with the increasing penetration of the mobile phones, VoIP is increasingly popular in cellular environment. Mobile VoIP, or mVoIP can be understood as an extension to a Voice over IP network that increases the usability in mobile environment with the ability of, for example, performing handovers between cellular base stations with hardly noticeable breaks in the connection. Mobile VoIP can be categorized into short range and wide area solutions. The former is limited typically in a physical area that could be, for example, under a single LAN that serves few base stations. The latter refers to solutions that are available for users over a whole system, like 3G and 4G networks. The protocols are selected respectively in such a way that the short range solution could include, for example, Wi-Fi, whereas the latter relies on protocols that are included in the mobile communications networks.

In the cellular environment, VoIP is increasingly available by default on various modern smartphones, tablets and other devices capable of transferring data in wireless environment.

Technically, smartphones and other wireless devices can be prepared for VoIP connections in several ways. One way is to integrate all the needed protocol stacks onto the device by default in the level of HW and operations system, in order to take full advantage of, for example, IMS connections. In this solution, the handset acts as a standard SIP client capable of initiating the SIP connections and transferring data via RTP for the voice channel. This solution provides the possibility of using standard VoIP protocols like SIP over any broadband IP network, like Wi-Fi, WiMAX, HSDP/HSDP+, and American EVDO rev A.

There are also other solutions, like soft-switch-based gateway for providing SIP and RTP with the SS7 of mobile networks. The benefit of this solution is that the device continues its original functions like CDMA, GSM and UMTS. The solution adds a SIP application server that provides SIP-based services.

For the end-user, hotspots like Wi-Fi access points offer an economical – typically free-of-charge – way to utilize VoIP. The drawbacks of Wi-Fi hotspots are limited coverage area even in dense cities, and lack of mobility as the handovers are typically not supported. Also the most popular hotspots may experience blocking of access service when various users share the same access point. For larger coverage and higher mobility, mobile communications networks nowadays provide sufficiently good throughput, but the user should be aware of the charging plans for non-flat rate subscriptions to avoid excess payments. In addition, it is essential to make sure that charging is understood in roaming environments as the bill is typically much higher in these cases, both in international as well as in national roaming scenarios.

6.7 Integrated Services Digital Network (ISDN)

ISDN (integrated services digital network) has been expanded heavily after the first practical deployments were completed. Nevertheless, the relatively long development and standardization of ISDN resulted in regional subsystems, and thus, for example, in Europe, there was a local system deployed in 1989, which was the so-called Euro-ISDN. It was compatible with the parties that had signed the MoU (memorandum of understanding), and it was compatible with the basic ISDN services defined by ETSI.

ISDN was a logical step in the digitalization of the previous analog telephony network. Via ISDN, all services, including voice service, have been completely digitalized. In addition to the digital transmission of services, ISDN offers wider services along with data transmission and calls. Furthermore, it is possible to connect several user equipments in the same subscription line – including analog variants.

ISDN as such did not change the earlier principles of analog networks, so as an example the international numbering of subscription lines is equal with the principles utilized earlier in the analog variant.

6.7.1 Standardization of ISDN

International standardization has resulted in the adaptation of ISDN globally. The basic definitions of ISDN are described in the blue book of CCITT, that is, the I-series, which was published in 1988, after several years of standardization work. The I.420 and I.421definitions include the basic and broadband versions of ISDN in such a way that the European variant is based on 30B+D channels whereas the US and Japanese variants have 23B+D. This means that the data rate of broadband ISDN in the USA and Japan is 1.5 Mb/s, and in Europe it is 2 Mb/s.

6.7.2 Principles of ISDN

There are two types of channels in the ISDN system. These are B for the bearer and D for the signaling data. B channels are used for data, which may include anything from files and voice to multimedia. The D channels are intended for signaling and control, although they can also be used for data to some extent.

The ISDN network provides customers with basic and broadband data rates. The European variant is based on a pair of B channels, each having 64 kb/s for user data, combined with an additional 16 kb/s signaling D channel. When these narrow-band channels are utilized together, the resulting data rate is thus 144 kb/s. In the broadband version, it is possible to use 30 channels, each consisting of 64 kb/s, resulting in a total data rate 2 Mb/s. In addition, the signaling capacity for the broadband version is a total of 64 kb/s.

ISDN defines the access types in the following way:

- Basic and broadband data classes.
- Network services, which are the signaling connections between user and network (UNI, User to Network Interface).
- Additional services, which provide extra functionality of calls.
- Tele services, which provide connection between end-users.

ISDN was designed from the beginning to be compatible with earlier analog systems. Nevertheless, the digitalization of telephone networks was so fast in the 1990s that there are now few analog networks in the world.

6.7.2.1 Narrowband ISDN

N-ISDN (narrowband ISDN, or basic ISDN) which has 2B + D channels, supports the connection of up to 8 parallel ISDN devices, and 2 of them can be active at the same time. The narrow-band ISDN utilizes 4-wires connection, from which 2 wires are meant for reception and the other 2 for transmitting. The narrowband ISDN uses Basic Rate Interface (BRI), also called basic rate access (BRA), which thus refers to the $2 \times B +$ D channels of the narrowband ISDN.

The total data rate of narrow-band ISDN is 192 kb/s, which is the sum of 2 x 64 kb/s of the B channels, and one D channel. In addition, this data rate contains supervision, synchronization, and maintenance signaling. It should be noted that D channel can also be utilized for user data in addition to signaling.

6.7.2.2 Broadband ISDN

Broadband ISDN (B-ISDN) provides with the use of a full capacity of a single E1 or T1 line. As for the terminology, Primary Rate Interface (PRI), also called primary rate access (PRA) in Europe, contains a higher number of B channels and a D channel with a bandwidth of 64 kb/s. The number of B channels for PRI depends on the country. As an example, in North America and Japan the number is 23B+1D, with a total bit rate of 1.544 Mb/s according to the T1 capacity. In Europe, India and Australia the number is 30B+1D, with a total bit rate of 2.048 Mb/s according to the E1 capacity.

For broadband ISDN, there is a set of possible signaling channels:

- H0: 384 kb/s.
- H10: 1472 kb/s.
- H11: 1536 kb/s.
- H12: 1920 kb/s.

Broadband ISDN can be defined in its basic format as containing 30 B + D channels. It is also possible to define variants of these channels, for example, H12 + D or 5 H0 + D. These variants were originally meant for telephone exchanges and other environments that require a high capacity.

6.7.3 ISDN Reference Model

The reference model of ISDN defines the different ways to connect to the infrastructure, as is shown in Figure 6.12. It should be noted that, for example, the reference models of GSM and UMTS are heavily based on the ISDN model.

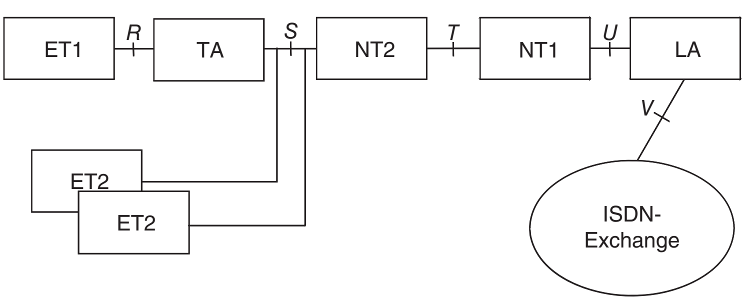

Figure 6.12 *The ISDN reference model with elements and interfaces.*

ISDN defines the following functional elements and interfaces in different reference points:

- TE1 terminal equipment can connect to the 4-wired ISDN as such via S interface. In other words, TE1 is thus ISDN terminal equipment. Its practical form can be, for example, telephone, facsimile equipment of group 4, and teletext equipment. The S interface supports data rates of narrowband ISDN.
- TE2 terminal equipment consists of machines that have not been designed originally for ISDN network. Some examples of these are telephone equipment of the previous analog network and old-fashioned data terminals. This type of equipment can be connected to ISDN via R interface which makes terminal adaptation via TA.
- TA (terminal adapter) makes it possible to connect TE2 type of terminals to ISDN network. TA makes the required interconnection between R and S interfaces.
- NT2 (network terminator) takes care of signaling between user equipment and network. It also manages the multiplexing of traffic and signaling of terminals, as well as local delivery of transmissions and internal calls. T interface supports the data rates of broadband ISDN. This interface is between user and network.
- NT1 takes care of the physical connection to the ISDN network. Its functions include quality monitoring of connection, synchronization of user terminals to the network, multiplexing of single connections and power supply.
- LA (line adapter) functions between 2-wired U interface and 4-wired V interface of the ISDN exchange.

6.7.4 ISDN Signaling

ISDN signaling can be divided into two parts: (1) signaling between user and network via the D channel, and (2) signaling between the network elements. The latter is defined according to SS7 standards.

The method applied between user and network is DSS (digital subscriber signaling), which is based on LAPD methodology. The signaling of D channel is threefold:

- The first level defines the physical connection between TE and NT2. The definitions of electrical characteristics and physical connectors as well as line coding belong to this physical level. The B and D channels are multiplexed in this level into the same physical line.
- The second level, that is, the transport layer, defines the connection between user and network. This level also takes care of multiplexing of terminals into the same line.
- The third level, that is, network layer defines the signaling method between users and network.

The signaling of network elements, in turn, is based on the SS7 within the PCM channel of 64 kb/s. An important part of the set of signaling messages is MTP (message transfer part), which is common to all users, as well as ISUP (ISDN user part), which is in use between ISDN and other networks. The older version that was used in the analog networks was TUP (telephone user part), which also contained country-dependent variants.

6.7.5 ISDN Services

ISDN services can be divided into three categories: network, tele and additional services. The network services provide with different means for data transmission, because tele services actually function on top of them for tele services. The added services enrich the user experience of normal phone calls, and make it possible to use different control functions related to the actual call.

6.7.5.1 Network Services

Network services are defined between terminals and offer the possibility to transfer teleservices between users. This means that the network must first be able to offer a sufficiently good capacity for supporting the data rate the user wishes to utilize, up to the maximum data rate that has been defined for the user.

The main classes of the circuit switched network services have been defined in the following way:

- 3.1 kHz audio service corresponds to the general voice service of the telephone networks in the bandwidth of 3.1 kHz (containing the audio frequencies of 300 Hz–3.4 kHz). This network service provides voice calls, or alternatively a narrowband modem data service or group 2 or 3 facsimile transmission.
- Transparent (constant bit rate) 64 kb/s data service provides a full B channel transmission data rate in such a way that the network guarantees error correction.
- Conversation class provides transfer of ISDN basic calls.

ISDN also defines network services for packet data in the following way:

- Virtual call provides packet data transfer between users and ISDN packet control unit, based on definitions of virtual connections.
- Permanent virtual connection establishes the reservation of the network resources between the users in such a way that there is no need to separately establish or end calls.

6.7.5.2 Teleservices

Teleservice refers to communication between end-users. These services are thus active between actual terminals from one terminal to another. The teleservices of ISDN are defined in the following way:

- Voice service is similar to the previous voice service of the telecommunication networks, based on the same principles, but offering now voice call via ISDN terminals.
- A voice call of 7 kb/s data rate offers higher quality voice call thanks to the wider bandwidth compared to the original 3.1 kHz band of tradition voice call.
- Facsimile group 2 and 3 offers the service via the analog adapter.
- Facsimile group 4 offers an ISDN-specific service which is faster than the previous facsimile service, with a wider variety of related functionality like multiple colors.
- Video text offers a rich method to present text, audio, pictures, moving pictures and video. In practice, the importance of this service has been low.

6.7.5.3 Additional Services

ISDN defines a large variety of additional services related to controlling and enriching the call. Some examples of the additional services are:

- Call forwarding.
- Conference call.
- Call hold, parallel initiation of another call, and connection of participants to conference call.
- Display of A-subscribers number for B-subscriber, as well as prohibition of display.
- Closed user group, which gives possibility to define ISDN subscriber numbers that can be called to.

- Direct calling, which gives possibility to call B-subscriber without dialing the complete number.
- Call continuity provides the possibility to transfer the original call to another terminal in such a way that the call does not drop.
- Charging display in real time.
- Definition of up to 8 individual telephone numbers for a single ISDN subscription.
- Direct call to ISDN subscriber by bypassing the exchange.

6.7.5.4 Use Cases

One of the typical environments for ISDN services is related to business exchanges, either narrowband or wideband variants. Currently, for private consumers, ISDN may be a method that assures basic voice calls and data services, the main use environment being dominantly in higher speed packet data side, for example, via xDSL or cable modem systems. The latter provides a more logical technological platform for much higher data rates.

6.8 Intelligent Network (IN)

6.8.1 IN Principles

Intelligent Network (IN) is an extension of the already existing network infrastructure for the addition of services in a centralized way. The benefit of IN is thus the fluent introduction of services as not all the traditional telephone network exchanges need to be updated to support the achieved services. IN also provides additional services that utilize open interfaces. This means that in principle the IN networks can be deployed and extended independently of network element providers.

The history of IN is already relatively old. In fact, the idea was deployed already in 1960s when the telephone network was connected to SPCs (stored program control). Along with the innovations, there was finally a solution called AIN (Advanced IN) in the USA in 1989. The equivalent solution in Europe was standardized in ITU-T, and as of 1989 in ETSI. These standards belong to the series of Q.12*xx*. These standards define capability sets (CS), which, in turn contain basic IN services.

The first packet, CS1, was released in 1993, which in practice initiated the IN deployments in Europe.

6.8.2 IN Elements

An essential element of IN is SCP (Service Control Point) which makes it possible to offer the IN services in a centralized way. The calls that uses IN services are thus directed to this point via SSP elements (service switching point).

The SCP redirects the calls with the help of core INAP, which is an intelligent network application protocol. Examples of this routing is the call from the fixed telephone network's exchange or GSM network's mobile switching services exchange between SSP and SCP.

In SCP, there is service management system (SMS) and databases that contain information about the services of the users. SMS takes care of the management and maintenance of these databases. Figure 6.13 clarifies the architecture of IN and the main elements.

For the communications between IN and other networks, there is STP element (Service Transfer Point) for the routing of the signaling. STP is packet switched, IN-specific network element which specializes in relaying signaling via the principles of SS7.

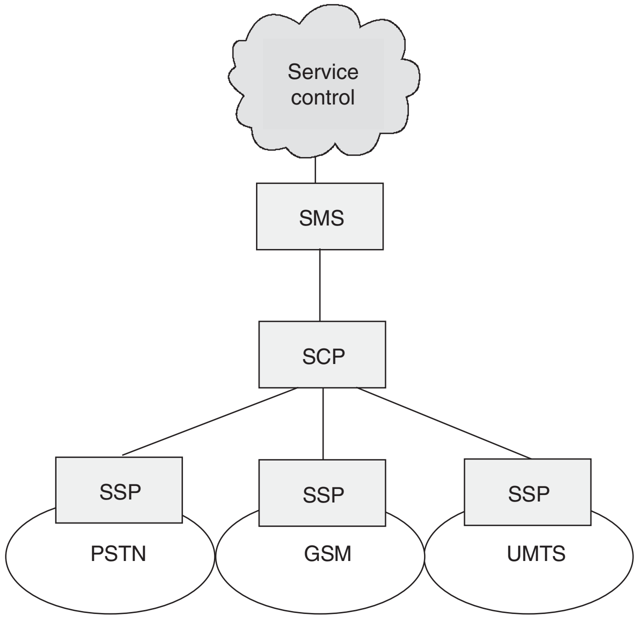

Figure 6.13 *The architecture and main elements of IN.*

6.8.3 IN Applications

The applications of IN can vary very much. Some examples are prepaid service numbers and personalized, PIN-protected services that can be accessed from any subscription line via the dial tone and that can be paid, for example, by utilizing credit card. Other examples are reversed billing (B subscriber pays the connection that is initiated by A-subscriber), division of the bill between A and B subscribers, voice menu and virtual private network (VPN). The latter is a practical way to deploy, for example, the exchange functions of small and medium-sized companies in such a way that the company does not need to invest in separate hardware and transmission lines.

IN is also suitable for deploying short dialing numbers, call announcements and automatic redialing of the number. In fact, IN can provide many similar types of services from the network side that it has been possible to handle via the terminal, and also many functionalities that have not been available before via network or terminal.

In practice, there is a range of subscription numbers reserved for national IN services, indicated by an operator-specific prefix of the number. These numbers, as well as others that have been defined as IN numbers but are not necessarily distinguished by their prefix, are redirected to SSP and finally to SCP which handles the call according to rules applied to that specific number. SCP can, for example, perform a number transform, that allows the special number to be redirected to the terminal of the final B-subscriber. The database dictates the charging principle of these calls, including the cost of the call which can be, for example, a single cost or based on active call time, in addition to normal call expense, including the variations of hour and day. As soon as the final number of the B-subscriber has been resolved from the data base, the call is forwarded to the SSP of the receiving end, and finally to the subscriber.

One of the special benefits of the IN concept is related to the transfer of subscriber numbers. The technology of IN makes it possible to comply with the typical regulatory requirements to keep the original subscription number in case, for example, the user wants to terminate the contract with one mobile operator and initiate

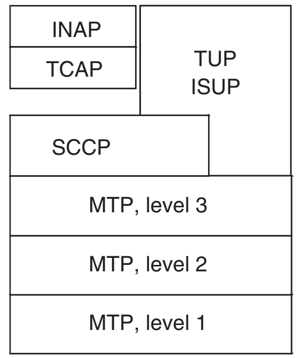

Figure 6.14 *The protocol stacks of IN network.*

the contract with a new operator, still keeping the original subscriber number, including the original prefix of the number that previously might have indicated the operator's network. The liberalization of the number portability has thus been possible via this IN functionality.

The protocol stack of IN is based on the principles of SS7 as can be seen in Figure 6.14. Whilst the analog network's user part has been TUP, and nowadays ISUP, the corresponding user part in IN environment is INAP.

As a summary, IN provides the possibility to deploy a whole set of totally new services, and facilitates the flexible management of subscriptions. It should be noted that especially mobile communication networks have been heavily integrated to IN to optimize the service offering and management of subscriptions.

6.9 SIP

6.9.1 Background

The Session Initiated Protocol (SIP) was born in 1995. It was originally merely a component for Multicast Backbone (MBone) network which was basically an experimental network on top of the Internet at that time. After that, SIP was developed from basic elements of this experiment. The actual SIP was defined first time in IETF MMUSIC working group (Multi-Party Multimedia Session Control). The first version of SIP was published via IETF in 1997, and it appeared in IETF document RFC 2543 in 1999. The work continued and the result was then IETF RFC 3261 which replaced the older definitions.

SIP in general is a set of protocols that is meant to be used in advanced telephony services and multimedia over the Internet. It should be noted that VoIP (Voice over IP) has further been developed from the idea of SIP in order to offer lower quality and more economical voice calls over the Internet.

SIP is utilized, for example, in IMS environment. Please refer to Chapter 9 for a more detailed description of the IMS.

6.9.2 Functionality of SIP

SIP is a signaling protocol with multiple functions. The most important functionality of SIP is related to the initiation, maintaining and releasing of voice calls between two users connected to IP network. SIP is capable of handling the user's location, and SIP is thus suitable also for mobile environment.

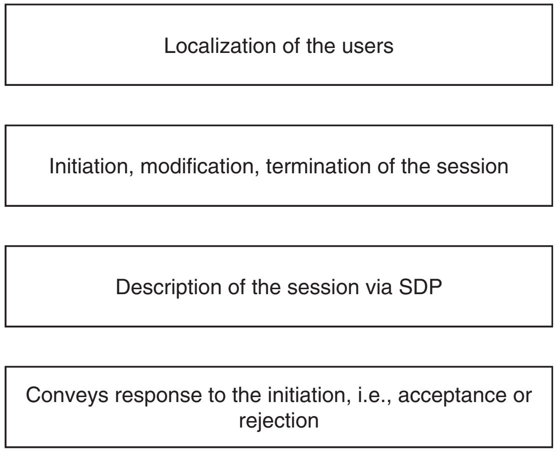

Figure 6.15 *The main tasks of SIP.*

It should be noted that SIP is a signaling protocol. For that reason, SIP is not aware of the actual details of communication. For that, there is a separate SDP (Session Description Protocol). Figure 6.15 summarizes the main functionalities of SIP.

Figure 6.16 shows the principle of routing the SIP call.

In Figure 6.16, the following procedure is executed when the call is initiated:

1. The A-subscriber sends request to the network in order to initiate a new session. This request is called INVITE which is routed to the B-subscriber via its outbound SIP proxy server. The address

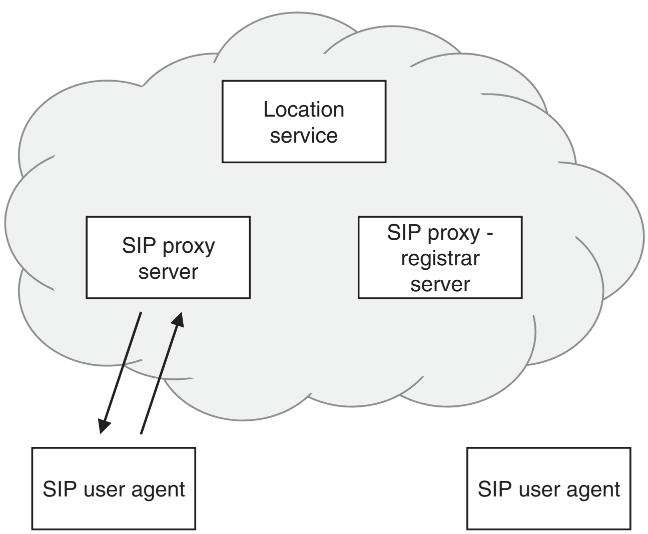

Figure 6.16 *The principle of the signaling flow in SIP connection.*

format in this communication is SIP URL (Uniform Resource Locator), as an example: "sip:jyrki.penttinen@finesstel.com". The outbound SIP proxy functions for the B-subscriber's domain in the initiation of the call. In the actual communications, as SIP is only meant for the initialization of the call, the proxy would not be used any more, but the call is routed directly to the B-subscriber without proxy.

2. Outbound SIP proxy then seeks the IP address of the SIP proxy of the B-subscriber via DNS. After getting the information, the outbound SIP proxy then forwards the request to this SIP proxy.
3. The next step is to locate the B-subscriber. For this, the proxy server makes a query of the SIP location service. The location service answers the last known name, for example, "Alpha." The location is known by the location service because the user (in this case, B-subscriber) has to register the location(s) with SIP registrar server before the calls can be set up. Please note that the functionalities of the SIP proxy and the SIP Registrar Server can be integrated physically into single equipment.
4. The INVITE signaling is then sent to the SIP user agent of the B-subscriber, which has the address of jyrki.penttinen@alpha.finesstel.com.
5. When the call request is successful, the SIP user agent returns a "200 OK" message to SIP Proxy-SIP Registrar Server, Outbound SIP Proxy Server, and finally to the original SIP User Agent of the calling party. After the reception of the confirmation, the data transfer can begin between A- and B-subscriber via SDP (Session Description Protocol), which manages, for example, the list of available codecs for the selection of one for the communication, as well as the port numbers.

Figure 6.17 shows the signaling flow of the SIP call.

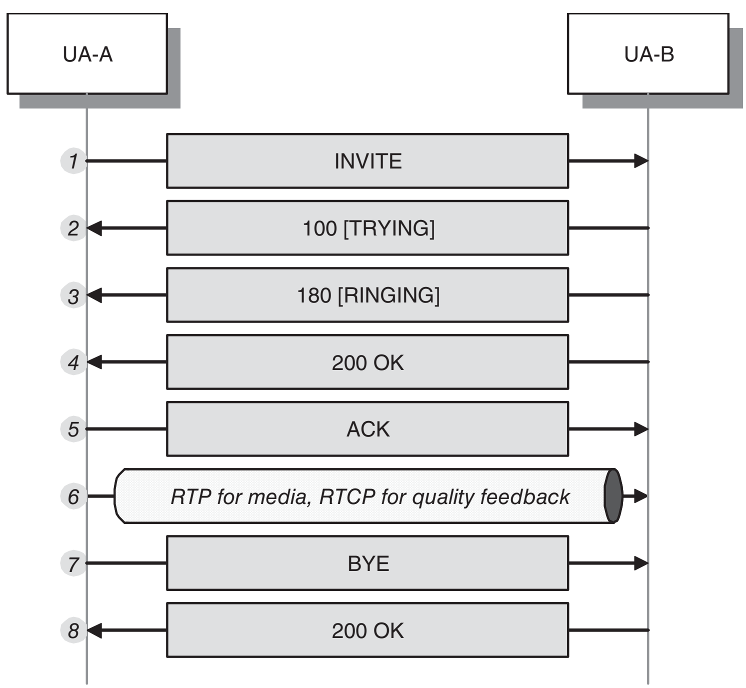

Figure 6.17 *The principle of the SIP call.*

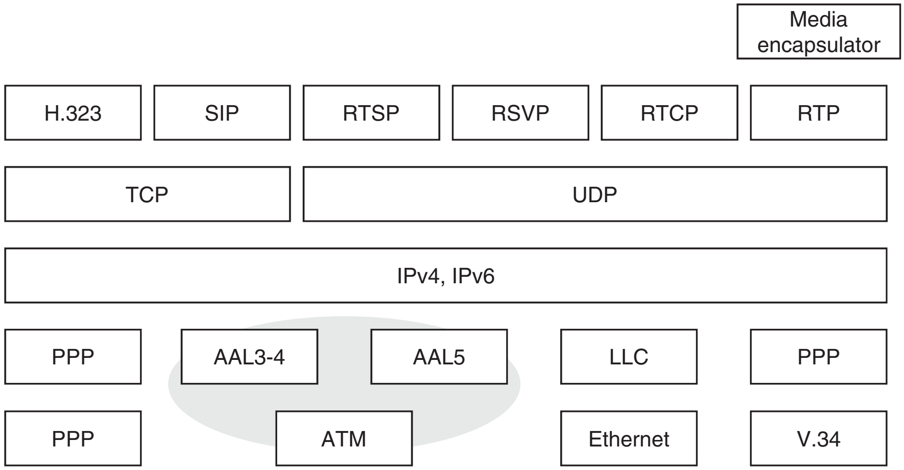

Figure 6.18 *The Internet Multimedia protocol stack.*

6.9.3 SIP Elements

SIP has elements from HTTP (HyperText Transfer Protocol) and SMTP (Simple Mail Transport Protocol). The former is basically meant for web browsing whereas the latter is used for the delivery of e-mails.

SIP URLs represent Internet addresses for the users. The form is http://alpha.beta.com. This means that there is a protocol scheme defined first, and then the domain name. Also the port number can be included to URL, separated by ":" after the domain name. There may also be other attributes, like file name after the port number. Other protocol schemes may be, for example, FTP (for accessing file that is accessed via anonymous FTP) and SIP (SIP address), MAILTO (for indicating email sending to respective address).

6.9.4 Protocol Stack for Internet Multimedia

Data streaming over IP, like VoIP, is basically a method for streaming data between users. There are various protocols that belong to the Internet Mutimedia Protocol Stack which is defined by IETF, like SIP, H323, RSVP (Resource Reservation Protocol), RTCP (Real-time Transport Control Protocol), RTP (Real-time Transport Protocol) and RTSP (Real Time Streaming Protocol). Figure 6.18 presents the stack.

6.9.5 Initiation of Call

6.9.5.1 Peer-to-Peer Call

The SIP operation has peer-to-peer operation. This mode means that two SIP user agents (UA) can communicate directly. The UAs may be any of the solutions that are able to communicate via IP, like LAN telephones. In peer-to-peer mode, the A-subscriber, that is, the calling UE, initiates the call by sending INVITE signaling to the B-subscriber, that is, the called UE. After the INVITE message, a "180 Ringing" message is sent back to A-subscribers UA, and whenever the B-subscriber answers, a "200 OK" message is still sent back to A-subscriber. Prior to opening the actual communication, the A-subscriber still sends an "ACK" message to the B-subscriber, and then the audio/video/data connection can begin. When the communication is over, the terminating UA sends a "bye" message. This is acknowledged by the other via a "200 OK" message.

Table 6.1 The fields of SDP

SDP field	Information	Optional / mandatory
a	Attributes of the media session. The format:a=rtpmap	O
b	Information of the bandwidth	O
c	Information of the connection	M
e	Email	O
i	Information of the session	O
k	Encryption key	O
m	Information of the media	M
o	Owner / creator, session identifier	M
p	Phone number	O
r	Repeat times	O
s	Name of the session	O
t	Initial and ending time of the session	M
u	Uniform Resource Identifier	O
v	Version of the protocol	M
z	Time zone corrections	O

6.9.5.2 Session Description Protocol

SDP (Session Description Protocol) describes multimedia sessions. SIP utilizes SDP for describing needed attributes for SIP sessions. The SDP parameters and their values are encapsulated within the SIP request. SDP contains, for example, the following media session related information, via a series of lines (fields) with abbreviation of each item in lower-case letters:

- IP address or host name.
- Port number by UDP or TCP.
- Media type.
- Media encoding scheme, for example, G.723.1, PCM a/myy law, and so on.
- Session name.
- Initiation and ending times.
- Connection information about the session.
- Bandwidth (kb/s).

There are both mandatory and optional fields as shown in Table 6.1.

6.10 Telephony Solutions for Companies

6.10.1 Centrex

6.10.1.1 Principle

Centrex (Centralized Exchange) refers to a telephone exchange-type of service that is manageable by a customer (company). This solution provides functionalities that telecommunications operators offer in general, but for a limited set of users, like the employees of a certain company. The switching of calls is performed physically in a centralized way instead of locating the equipment at the company premises. It can be generalized that Centrex solution outsources the exchange functionality of companies in such a way that the operator takes care of the operation and maintenance of services, including the SW and HW of the respective equipment. The customer would not have thus physically the equipment, which provides benefits especially for optimizing fault management and maintenance.

Centrex is thus a service offered by telecommunications operator in such a way that the operator allocates a part of the capacity of PBX (Private Branch Exchange) for the exclusive use of on organization that will be the customer of the operator. Centrex provides functionality of a private telephone switch without need for own PBX equipment or work force with knowledge of the system. Centrex is especially useful for small companies, but is beneficial for organizations of all sizes. For the latter case, the economical benefit is lower, though.

In practice, the solution is based on centralized equipment at exchange sites of operator. Between the centralized site and customer's physical location, there is a set of transmission lines. Physically these lines can be "old-fashioned" paired copper cables, or nowadays by default fiber optics. Centrex functions as a kind of SW-emulated PBX which mimics the functionalities of the original HW of the exchange. The benefit of the solution is that the special functionalities that the customer requires can be offered without complicated HW. Typical solutions offer PBX switching functionalities of 3–5 digits of subnumbering scheme, and the functionalities like call forwarding, conference calls and so on.

The initial Centrex solution for the fixed environment provides Centrex lines which are available for the Centrex users from the local exchange. The extensions are possible to provide different physical locations easier than in case of the company-owned and operated PBX in such a way that the functionality feels like the lines are within the same building.

Some examples of environments that benefit the Centrex solutions are:

- Small startup companies that are growing yet wanting to optimize the capital expenditure.
- Companies that have country-wide operations and offices spread over large geographical areas. Some examples of these are hotel chains, universities, banks and financial institutions, as well as government offices. The benefits of centralized PBX functions include the easiness to setup the services, assurance of the reliability of the connections even in faulty situations, cost savings, and high quality level of customer service.
- Temporal locations where PBX services are needed for only limited time.

6.10.1.2 Functionality

In practice, when organization becomes a customer of the operator and is provided with Centrex solution, the organization is allocated with a set of extensions of PBX for closed group utilization. Physically, the Centrex service can be within a single building, or it can cover various geographical sites. For the latter, a VPN solution (Virtual Private Network) is applied. As the physical equipment resides at the operators' premises, no additional equipment is needed at the customer site. In practice, the setting up of the service is comparable with the activation of single telephone lines.

Typically, at least the following set is possible to use per extension:

- DDI (Direct Dialling In). This functionality includes customer-controlled call barring, call transfer and hunt groups.
- Voicemail, with Call Forwarding and Call holding functionalities.
- Camp on Busy, with Ring Back and Call Waiting functionalities.

For the selection of Centrex service, it is advisable to make a cost analysis for the Capital Expenditure (CAPEX) and Operation Expenditure (OPEX) by taking into account the difference between old solution and

Centrex as a function of at least couple of years. The points to be considered for own solution may include the initial cost of equipment, installation costs, rental of lines, maintenance costs, system management costs, and costs for expansions. These points can be reflected against the benefits of Centrex, which are, for example, lack of equipment that needs otherwise maintenance, possibility to obtain PBX functionality, no charges for internal calls within the same user group, possibility to extend the service geographically.

A practical example of the Centrex concept is Central Office Exchange or Centrex. It is a standardized service, and is used by various companies. It serves as a replacement for the old PBX systems that companies still have. There are various provides and several commercial names for this solution.

The functionality of Centrex is based on phone lines created by using NARS (Network Access Registers). After the initial setup, the hosting provider manages forthcoming calls to these lines via a central office. There are no differences in call management of Centrex compared to normal business and residential line management.

6.11 Transport

The traditional transport of telecommunications networks has been based on dedicated, circuit switched (CS) lines. The physical means have been typically based on fiber optics. The CS lines have been done typically via TDM, more specifically via E1 and T1.

As the telecommunications world is going strongly towards all-IP concept including the delivery of voice traffic, these traditional solutions are reaching practical limits especially as for the nature of the IP traffic. Along with the modernization of radio access technologies, also the core network is converting towards the IP.

As the transition is not rapid due to the large proportion of the legacy systems in the networks, there are techniques that can be utilized for supporting the old principles and new IP concept. One of these is Carrier Ethernet Transport that encapsulates the contents and old format into IP packets that can be delivered over any IP network as such. More advanced technologies include the cloud concept which can also be used, among other tasks, for load balancing of transport networks.

The Carrier Ethernet Transport (CET) technology can be utilized for deployment of new backhaul networks as shown in Figure 6.19. For connectivity, so-called pseudo wire solutions can be applied for the emulation of the TDM and ATM, if native solutions are not available.

The CET concept is a cost-effective solution that can replace the traditional time division multiplex (TDM) transport solutions such as SDH/PDH. It is possible to deploy CET for both access and aggregation networks. Basically all LTE, 3G and 2G traffic types can be delivered over the packet-based backhaul infrastructure. The main benefits of CET solution are: support of standardized services of variety of physical infrastructures, wide scalability of the bandwidth (from 1 Mb/s to over 10 Gb/s), high reliability, support of Quality of Service options. It also offers the possibility to monitor, diagnose and manage the network in a centralized way. CET has been standardized by the Metro Ethernet Forum, so it provides vendor-independent implementations.

6.12 Cloud Computing

6.12.1 General

Cloud computing is based on the sharing of resources in order to achieve capacity gain and economies of scale. The drawback of dedicated data delivery networks is that their typical delivered load is only a fraction of the dimensioned maximum capacity, due to the fact that there are occasional load peaks that

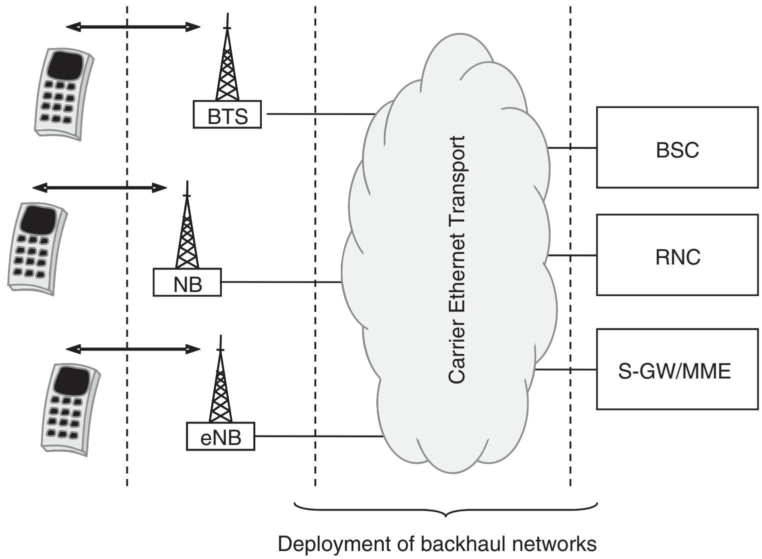

Figure 6.19 *The principle of CET.*

demand considerably more capacity. Instead of all the capacity operators managing this type of highly over-dimensioned networks for the sake of the serving of these capacity peaks, the whole delivery infrastructure can instead be concentrated into a single, large IP cloud. As the capacity peaks of different operators or areas within a single operator typically won't occur exactly at the same time, total offered capacity of the cloud can be considerably lower than the sum of the individual networks would be. This concentration of the capacity into the cloud is one of the benefits of cloud computing in telecommunications.

As the concept is still relatively new, there might be concerns about connectivity control, security and privacy as the functionality and contents is distributed outside the previously very local environment. The important task for network service providers is thus to address these concerns and make the respective measures to minimize the potential security risks.

Cloud computing is thus a model that enables on-demand network access to a set of shared computing resources that are configurable.

6.12.2 Principles

It starts to be clear that the future of corporate IT is in private clouds. Cloud computing provides a totally new way to process services via flexible computing networks. Some important influencing parties have been public providers like Google and Amazon.

An important need for cloud computing is the possibility to outsource IT tasks to cloud services available over the Internet. It seems that this need will increase, which creates a natural evolution for more advanced cloud services that can be offered for large companies to build private cloud networks. The benefit of this evolution is that increasing amount of resources can be managed via a single point and utilized by applications and services based on the need. Naturally, the development may take years for the large-scale support of the cloud services. The server virtualization is one of the keys for the creation of internal and external cloud networks and services. There are also other areas that need to be created and developed like a meta operating system that is needed to manage distributed resources as one computing pool. A meta operating system is a virtualization layer between applications and distributed computing resources. It takes the advantage of distributed computing resources in order to manage and control applications and related tasks like error control. In addition, there is also a need for a service governor for decision making about the final allocation and prioritizing of the computing resources for applications.

The initial evolution of cloud networks and services is driven by the needs of large companies, but eventually cloud computing can also become an important daily tool for smaller companies.

6.12.3 Benefits

Cloud computing offers the possibility for dynamic network structuring. An example of this functionality is software-based networking as a part of cloud computing access via low cost and flexible adjustments by the cloud customers. Cloud networks also provides safety features for essential network functions like IP routing, address management, NAT, authentication, QoS.

Via software-based networking of cloud computing, the network architecture of customers may be replicated regardless of location. The replication includes changes to topologies and policies, making cloud networking a logical platform for the disaster recovery in a dynamic and cost effective way.

References

1. International Telecommunication Union. *ITU-T Blue book (Telecommunication Policies for the Americas)*. 2nd edition, 2000.
2. Telecom Advice, Virtual Exchange. http://www.telecomsadvice.org.uk/infosheets/centrex_virtual_exchange_service.htm (last accessed 26 July 2014).
3. Phone System Chattery, Virtual Centrex. http://phonesystemchattery.typepad.com/blog/2010/07/get-to-know-the-centrex-virtual-pbx-system.html (last accessed 26 July 2014).
4. Network World, Cloud Development. http://www.networkworld.com/news/2008/111208-private-cloud-networks.html (last accessed 26 July 2014).
5. Penttinen, J. (2011) *The LTE/SAE Deployment Handbook*, John Wiley & Sons, Ltd.

7

Data Networks

Jyrki T. J. Penttinen, Tero Jalkanen and Ilkka Keisala

7.1 Introduction

Data networks are the basic building blocks for a major part of the telecommunication services today. After the long history of circuit switched, analog telephony systems, the change towards digital and packet-based systems has been radical. The fast transition has provided totally new user experiences that support the high demands of the modern multimedia environment. Both fixed and wireless networks develop in a parallel way, in order to support ever growing demand. The need for speed is due to advanced applications and higher technical quality of contents, including transition to high definition video and audio, and advanced multimedia services like video conferencing. As the number of users is growing extremely quickly, the constant signaling caused by messaging types of services – even if the service itself would not require high bandwidth – creates challenges for the dimensioning of transport networks.

As the capacity is a potential limitation for the transport network itself, as well as for the addressing of the equipment, the evolution is going rapidly towards IPv6. It is important to handle the transition time correctly in order to have fluent functionality between networks that support the old technology, and those that contain the latest solutions.

7.2 IPv4

IP (Internet protocol) has been defined for the efficient delivery of the packet data. The base for IP are the routers that deliver the packets from the sender to the receiver dynamically via different routes. The original idea of this adaptation for varying conditions of the network is thus still valid, the first solutions developed for the military environment that take into account the potential danger of transmission lines breaking down, and the traffic routed on fly via alternative routes. After the military solutions, the original IP was defined in 1981 by IETF (Internet Engineering Task Force), which is still the building block for the global Internet of today.

IP is suitable for the delivery of basically all types of packet data, from the smallest local networks up to the largest ones at the international level. The length of the IP packets is not fixed in size, which provides

The Telecommunications Handbook: Engineering Guidelines for Fixed, Mobile and Satellite Systems, First Edition.
Edited by Jyrki T. J. Penttinen.

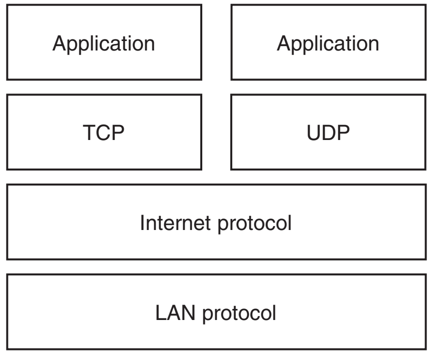

Figure 7.1 *The relationship between IP and other protocols on top of IP.*

the possibility to utilize many different networks between the sensor and receiver, varying the capacity and performance. This is possible by the fragmentation of the user packets of the sender, and reassembly in the receiving end.

IP as such does not guarantee the delivery of packets by any means as for the quality, flow control nor order of different packets delivered via the packet networks. This means that there should be additional solutions on top of the IP networks, which takes care of flow control. The most common ones for this purpose are TCP (transmission control protocol) for connections that require additional means for error recovery, and UDP (user datagram protocol) for applications that take care of their own error detection and recovery.

Still by far the most common Internet protocol is IPv4, which traditionally has been abbreviated as IP. It has been the basis for practically all modern Internet solutions for fixed and mobile communications, including, for example, GPRS. The high-level protocol stack of IPv4 is shown in Figure 7.1.

The heading of the IPv4 packet has 32 bits, which has been divided into 3 different variants depending on network type. The type can be A, B or C class as presented in Figure 7.2. The A class has the largest network numbering field compared to sub network addressing, and is thus suitable for large organizations. The B class is suitable for small and medium size organizations, whereas C class addressing is suitable for private users.

As the utilization of the Internet is growing extremely fast, there have been discussions of potential ending of available addresses of IPv4. The next step has been IPv6 which aims to guarantee availability of addresses long into the future. Instead of the 32 bit structure of IPv4, IPv6 is based on the 128 bit address field. Other terms that have been utilized for the IPv6 are IP Next Generation or IPng.

The transition towards IPv6 has already been planned for a relatively long time. As an example, the very first GPRS specifications of 3GPP (ETSI) release 97 took this into account, providing a smooth transition to

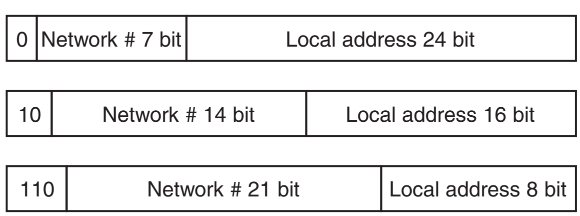

Figure 7.2 *The addressing of IPv4 classes.*

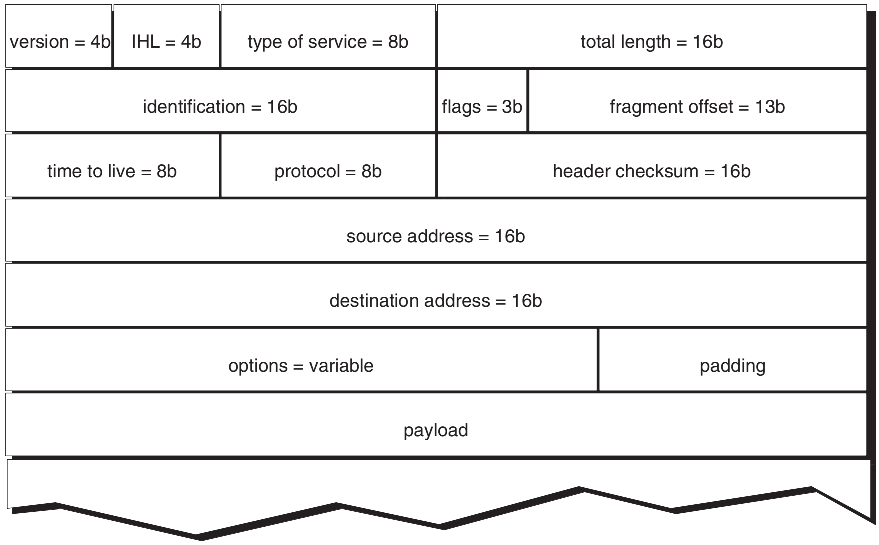

Figure 7.3 *IPv4 data frame.*

IPv6 when the moment was logical. Because IPv4 has been so widely adopted, the transition has taken time though. The mechanisms have been developed to cope with this transition time.

As IPv4 has been utilized basically everywhere, the transition has taken a relatively long time. In order to cope with this phenomenon, mechanisms have been developed so that IPv4 address field can be embedded into that of IPv6. The addressing has been defined in IETF RFC 1884.

IP datagram is a frame that contains both header and user data fields. Figure 7.3 presents the IPv4 datagram.

IETF is a forum that makes decisions based on democratic voting principles. The IP definitions are called RFC (Request for Comments). IETF consists of various working groups that focus on certain details of IP development. Email distribution lists are widely utilized in the planning, which makes the work of IEFT very flexible and dynamic compared to many other organizations with more traditional meeting protocols.

7.2.1 General

Although data networks are becoming more common in vast steps, and the data services have a rapidly growing role in modern society, the voice service is still one of the most important services. The packet switched data lead by Internet protocol has been noted suitable for all type of data, including text, pictures, video and binary code, so the logical solution for the era towards all IP concept is to transfer also voice over IP networks.

The transfer of voice service was already possible as such in the 1990s especially as the Internet took off as a platform for users outside its previous relatively closed societies and military environment. The basic idea has not changed, the participating users sharing the common voice codec in one or another way in the terminals. As the number of codecs is growing, basically all imaginable environments can be covered via IP networks, including the interconnection of IP, fixed and mobile telephone networks.

The amount of operators offering these interconnection points is growing steadily. The common fact for all these service providers is that they tend to offer IP calls with more economical prices both in the national as well as international environment compared to traditional fixed or mobile operators. The drawback has been the variations of the quality of these IP calls which sometimes make the connection break down – which has been relatively rare in the previous generation fixed telephony networks. The balance of the achieved quality of service and the expense of the call is thus one of the tasks that end-users evaluate in order to decide in which circumstances the combination is acceptable.

The same principle of IP calls can also be offered via packet switched data services of mobile networks by integrating a related SW in the device, or by installing a separate application to capable devices. The data stream produced by the voice service is then treated as any other data service over the cellular network. If the data usage plan contains sufficient data transfer per month, this option can provide either free-of-charge calls between users sharing the same codec and Internet connection, or considerably lower expenses by calling the normal subscription of fixed or mobile user.

7.2.2 IPv4 Addresses

In the initial phase of IPv4 designing and deployment, the Internet addresses were divided into 3 categories. These are class A, B and C addresses. These classes provide the designing of subnetworks accordingly. The class is indicated in the address field in such a way that A refers to bit "0," B refers to bits "10," and C refers to bits "110." Based on this information, it is explicitly clear how many bits the network number has been granted, and from which position the host number begins with. The IPv4 address classes are presented in Figure 7.4.

7.2.2.1 Class A

Each one of the representatives of class A has an 8-bit prefix which is typically marked as "/8" (slash eight). In this prefix, there is the first "0"-bit indicating the class A. Followed by that, there is a 24 bit address space indicating the actual network station. The total amount of the class A subnetworks can be 126 (2^{27}-2).

The class A addresses take half of the capacity of the complete IPv4 address space, which provides a total of 2^{32} addresses. That's the most important reason for the extremely low availability of new class A addresses. The addresses 0.0.0.0 and 127.0.0.0 have been reserved for other purposes, the first one for the default route, and the second for the loopback functionality, that is, the testing of the station's own TCP/IP functionality.

Each one of the /8 sub networks provides a total of 2^{24}-2 hosts. It should be noted that the addresses ending "0.0.0" and "1.1.1" cannot be utilized in normal IP traffic.

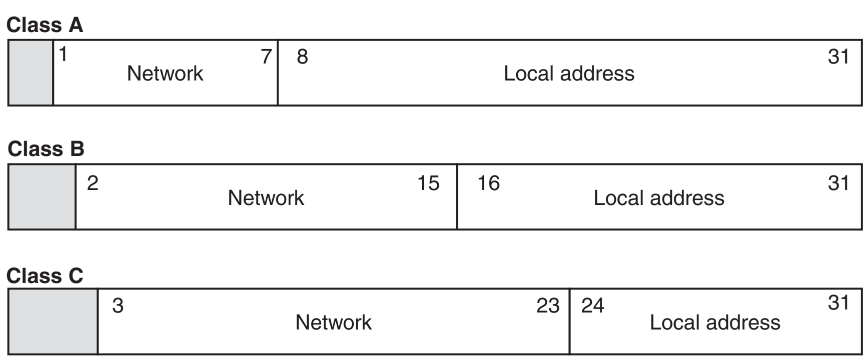

Figure 7.4 *The IPv4 address classes and respective numbering of the address.*

7.2.2.2 Class B

Class B addresses have a prefix of 16 bits followed by 14 bits for the addresses of stations. The class B addresses are often marked as /16. The maximum number of class B subnetworks can be 16 384 (2^{16}-2), and the class provides a maximum of 2^{30} addresses per subnetwork. The class B address space reserves thus 25% from the total address space of IPv4. This class has been meant for the use of large companies and organizations.

7.2.2.3 Class C

Class C addresses consist of a prefix of 24 bits and an 8-bit address space for the stations. Following the same principle with class A and B addresses, the class C addresses are typically marked as /24.

Class C provides 2^{21} subnetworks and 254 (2^8-2) addresses per network. This class represents 12.5 percent of all IPv4 addresses.

7.2.2.4 Other Classes

In addition to previously mentioned classes, IPv4 contains definition for IP Multicasting via class D, and its prefix is "1110." Class E is marked with prefix "1111," and is reserved for testing purposes.

7.2.2.5 Development of the Addresses

The explosive growth of the Internet has surprised developers and has been causing challenges in the assignment of addresses and management of routing tables. As an example, class B addresses have been handed out even to relatively small user groups which have wasted the capacity.

For this reason, the addressing was developed further and network identifiers can be delivered to subnetwork identifiers and their network station identifiers. The other area of development is related to the use of the prefix. The routers of the network investigate only the prefix when they are making the routing, in order to deliver the packets to the subnetwork of the destination. The routers of these subnetworks, on the other hand, utilize extended prefix when routing the packets between subnetworks. The extended prefix is identified by the subnetwork mask. If a company has, e.g., a class B address a.b.c.d. and the company wishes to utilize all three octets of the address from the address, the subnetwork mask must be defined as 255.255.255.0.

In addition to the subnetwork mask, modern routers can also utilize more compact notation for indicating – with the accuracy of bit – the length of the mask. An example of this could be the notation of a.b.c.d/24, which means the same as previously, that is, the mask of 255.255.255.0.

7.2.3 Notation of the Address

The IP addresses are presented in the practice in a number format in such a way that each of the 8-bit field of the address field corresponds to a decimal number separated by a full stop. As there is a fixed prefix with each class, the notation is as presented in Table 7.1.

7.3 IPv6

Internet Protocol Version 6, IPv6, represents the new generation of the Internet. It provides various benefits compared to the previous version of Internet Protocol, IPv4. The common idea of both of these protocols is

Table 7.1 The address space per IPv4 class

Class	Address, initial	Address, final
A	1.0.0.1	126.255.255.255
B	128.0.0.1	191.255.255.255
C	192.0.0.1	223.255.255.255

to define network layer protocol. It dictates how data is transferred between sending and receiving entity over packet switched data networks (PDN). Internet itself is one of PDNs, but IPv4 and IPv6 works basically over any packet data network. The IPv6 is also sometimes called the Next Generation Internet Protocol, IPng.

As is the case for IPv4, also IPv6 is defined in a set of RFC documents, Request for Comments. IPv6 is thus the next step in Internet. Nevertheless, as the transition takes a relatively long time due to the existence of services based on IPv4 addressing, both variants are used in a parallel fashion for years to come. The basic idea is, in any case, that the older variant will gradually disappear [1].

7.3.1 Principles

Internet Protocol version 6 has been defined in a set of IETF recommendations that begins with the introduction of RFC 2460 in 1996. The ultimate aim of it is to replace the older IPv4 (defined in RFC 791) due to limitations of old definitions, including the lack of address space. These limitations have been noted as important obstacles for increased utilization of Internet, and that is why the next generation IP, or IPv6 was developed. As a detail about the version numbering – the next generation would have been logical to number as IPv5, but that term was already utilized in other experimental network which obligated IETF to choose the term IPv6 for the next generation system.

According to the historical data, there was only about 10% IPv4 addresses left by the year 2010. The total number of IP addresses that the IPv4 allows is 2^{32}, that is, slightly over 4 billions. This is inadequate for the needs, since there is a vast amount of IP-based machines that, in addition, might utilize solely various IP addresses.

The amount of unique addresses in IPv6 is 2^{128} in order to guarantee availability for the long-term future. With increasing IPv6 support, there have been major manufacturers driving the technologies. In addition, important entities like the USA government are promoting the new technology by obligating federal agencies to deploy IPv6.

The major difference between IPv6 and IPv4 is header addressing. As these differ significantly from each others, the versions are not compatible. In many details, IPv6 is actually an extension to IPv4. For that, many transport protocols need only little or no changes in order to operate within IPv6. The protocols that need modifications are the application protocols that integrate network addresses for the functioning, including, for example, FTP, NTPv3 and NTPv4.

7.3.2 IPv6 Address

7.3.2.1 Address Field

The address field of IPv6 is more complete than in IPv4. The most significant difference is the length of the address field, which has grown fourfold being now 2^{128} corresponding to $3.4 \cdot 10^{38}$. Figure 7.5 shows the principle of the IPv6 address field.

The vast amount of addresses can be visualized by assuming the number of inhabitants of the world as 7 billions. On average, there would be $4.8 \cdot 10^{28}$ IPv6 addresses available for each person.

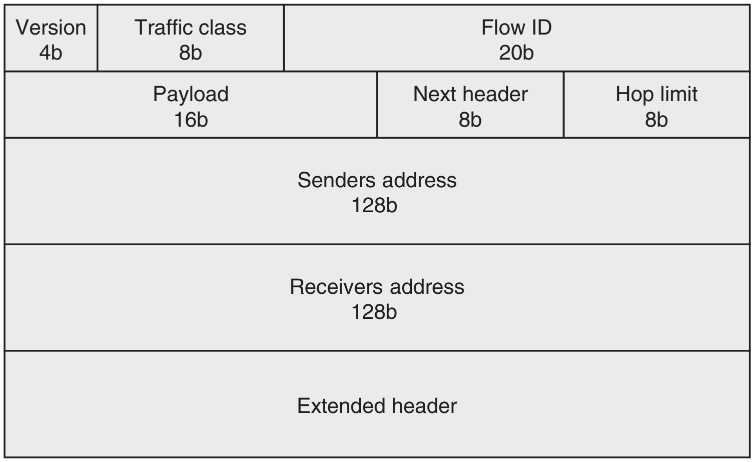

Figure 7.5 *The address field of IPv6 can be utilized for 2^{128} addresses.*

7.3.2.2 Presentation of IPv6 Address

Via the IEFT mechanisms, the IPv6 addresses are compatible with IPv4 addresses, which facilitate the deployment of IPv6 systems, first as isolated parts of the networks, and finally growing to provide a full IPv6 addressing in the complete transmission chain.

The IPv6 addresses can be presented in 3 ways: Preferred, Compressed and Mixed. Furthermore, there are 3 different types of addresses, for unicast, anycast and multicast.

The preferred presentation form contains a series of hexadecimal integer numbers, each containing 4 numbers in 16-bit presentation format. Different numbers are separated from each others by colon. There are 8 number fields in total in this presentation of the address, and each number is presented in hexadecimal format as shown in Figure 7.6.

As can be observed from Figure 7.6, the addresses of IPv6 are separated from each other by colon. If some of the field contains only zero-bits, the number can be left without marking in the compact presentation format of IPv6, resulting in two colons in series.

It should be noted that an IPv6 address can also contain an IPv4 address. In this case, the address consists of six IPv6 address fields followed by colon, and IPv4 address marked in 4 fields as is the case in its original presentation format.

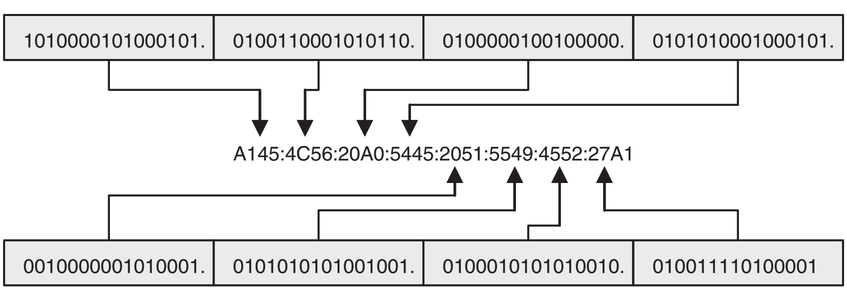

Figure 7.6 *The presentation of IPv6 address.*

7.4 Routing

The functionality of the Internet is based on the dynamic routing of the packets between the sender and receiving ends. The logical transport of the packets is handled by TCP/IP protocol set (transfer control protocol / Internet protocol), and the physical base for the delivery of the packets as a large set of routers.

The routers form a complex structure. There are 12 master routers in the highest level of the hierarchy of Internet. If some of these master routers are damaged, the remaining ones assure the correct functioning of the routing, still. When an IP packet arrives at a router, it checks if the receiving end, or the routers that lead to the receiving end, is/are marked into the routing table of this investigating router. If this is not the case, it routes the packet further, and so on, until the correct subnetwork is found in some of the routing tables. The rest of the routing is then straightforward according to the routing table information, until the packet has been delivered to the final destination address.

The main idea of the routers is to work as a redirecting element in the third OSI layer. There are also repeaters, which function in the first protocol layer, and bridge that functions in the second protocol layer.

Each IP network and subnetwork contains its own IP network address. The delivery of the packets can be done in a controlled way with the help of routers because the routing tables of the routers that connect networks with each other contain the information about the addresses of the other networks. The table consists of the addresses of networks and other routers as well as the number of hops leading to these addresses. Figure 7.7 shows an example of the routing information, and Figure 7.8 shows the general principle of the routing.

The addresses can be seen from the headers of the messages. Each protocol layer adds its own header together with other control information. The header of MAC layer contains the address of the device, the header of the IP layer contains the header according to IP protocol, and TCP contains the needed information for the transportation, as can be seen in Figure 7.9. The total amount of this overhead can be thus considerable.

- C:\>tracert www.finesstel.com
- 1 192.168.194.67
- 2 192.168.197.1
- 3 148.233.151.1
- 4 inet-mex-nextengo-8-s4-1-0.uninet.net.mx [148.233.148.190]
- 5 200.38.196.33
- 6 inet-cal-onewilshire-1-pos7-0.uninet.net.mx [200.38.209.10]
- 7 69.224.111.181
- 8 bb2-g8-1-0.irvnca.sbcglobal.net [151.164.42.42]
- 9 core2-p5-0.crrvca.sbcglobal.net [151.164.41.14]
- 10 core2-p10-0.crhstx.sbcglobal.net [151.164.42.4]
- 11 151.164.191.192
- 12 core2-p8-0.cratga.sbcglobal.net [151.164.241.86]
- 13 core2-p6-0.crhnva.sbcglobal.net [151.164.41.206]
- 14 core1-p8-0.crhnva.sbcglobal.net [151.164.188.21]
- 15 bb1-p4-0.hrndva.sbcglobal.net [151.164.191.98]
- 16 ex1-p11-0.eqabva.sbcglobal.net [151.164.40.49]
- 17 nb18b11-ge0-0.nb.telenor.net [206.223.115.133]
- 18 nb10b11-pos4-1.nb.telenor.net [217.70.227.37]
- 19 nb03b11-pos3-0.nb.telenor.net [217.70.227.45]
- 20 217.70.229.114
- 21 hel-ficix1-fe0-0.utfors.net [212.105.101.130]
- 22 r1-ge2.hki.nbl.fi [80.81.160.235]
- 23 virtual8.nebula.fi [217.30.180.25]

Figure 7.7 *An example of the IPv4 routing between the device in Mexico and server located in Finland via an GPRS operator of Mexico.*

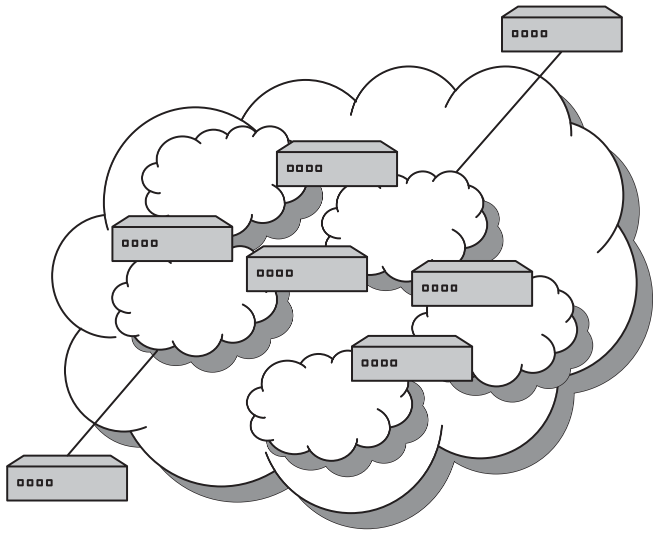

Figure 7.8 The routing of packets between the subnetworks contains typically several routers in the interface between different networks.

When the packet arrives to the network router, it investigates the contents of the MAC packet in the IP layer in order to resolve the address of the destination. The, the router builds renewed MAC packet and transfers it to the next router which is identified by the new MAC address.

TCP takes care of the errorless transfer of the packets. There is no restriction to utilize also, for example, UDP, but in that case, some higher layer like application should take care of the error detection and correction.

Whenever new router is attached to the network, it sends a message that indicates to which networks it is connected. The functionality of IP networks is thus highly dynamic, and the base for them is the ability to recover from faulty situations as fast as possible. If, for example, a router is damaged, the connection is rerouted immediately via other, still functional routers, which, in turn, correct their routing tables at the same

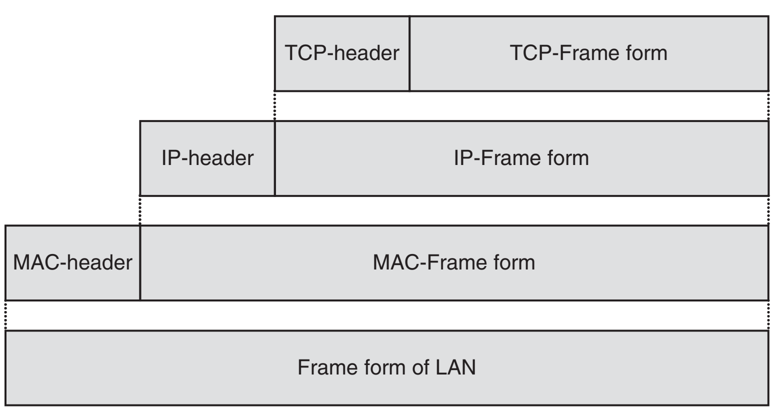

Figure 7.9 The principle of the forming of data packet. Each protocol layer adds information to the message, for example, in the form of headers. The lower the protocol layer is, the longer the packet also is.

time. The packets will thus reach the destination as long as there are alternate elements and interfaces in the communications route.

7.5 ATM

Prior to the full-scale IP concept, ATM (Asynchronous Transfer Mode) has been utilized for packet transport. It is a transport technique that utilizes the network resources dynamically. Unlike in STM (Synchronous Transfer Mode), which has been designed for circuit switched connections, ATM has been made especially for the delivery of data packets with varying length and delays. ATM is thus suitable for many types of high-speed systems like broadband ISDN network as well as UMTS. Among other solutions, ATM can work as a basis for PDH and SDH networks.

ATM is based on data packets that are called cells with a fixed size of 53 octets. The cells are delivered via ATM switches, which are comparable with the circuit switched exchanges of fixed telephony networks.

The ATM cell consists of a header field of 5 octets and data field of 48 octets as presented in Figure 7.10, each octet consisting of 8 bits. The notification of Figure 7.10 is the following:

- CGF = general flow control
- VPI = virtual path identifier
- VCI = virtual channel identifier
- CLP = cell loss priority

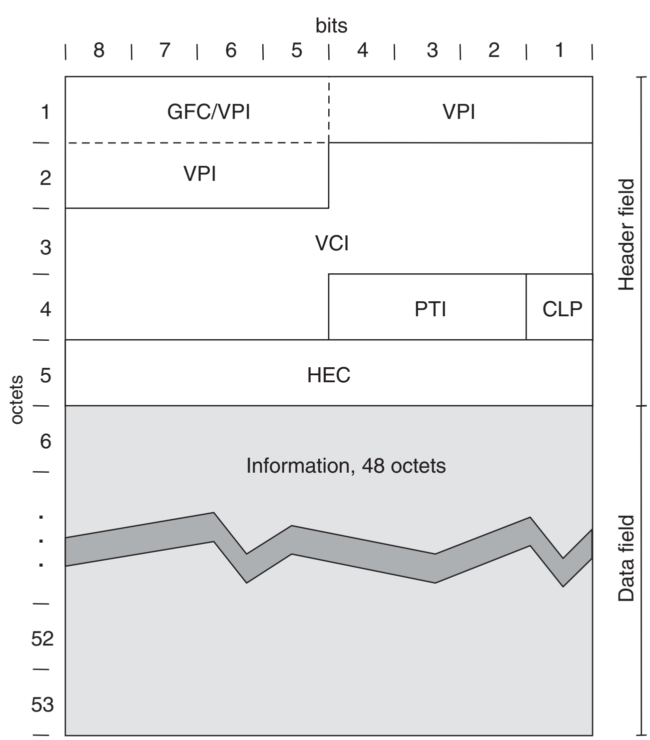

Figure 7.10 *The fields of ATM cell.*

Table 7.2 The class division of ATM

AAL type	AAL-1	AAL-2	AAL-3/4 or AAL-5	AAL-3/4 or AAL-5
Size	Constant	variable	variable	variable
Time dependency	Yes	yes	no	no
Connection oriented	Yes	yes	yes	no
Applications	Voice calls, circuit switched data	Audio and video	X.25 and FR	IP

- HEC = header error control
- PTI = payload type identifier, referring to XYZ infotype, where X = 1 means user data, Y = 1 means blocking, and Z is a bit between users.

There is a so-called double layer concept in use with ATM, with a term of AAL (ATM adaptation layer). The lowest level of ATM protocol structure is the physical layer. The second layer is the actual ATM layer, and AAL resides on top of that. With the use of AAL, it is possible to utilize different types of transfers in the ATM cells.

There are 5 sublevels in AAL, which are referred to AAL-1 … AAL-5. Each AAL sublevel is meant to a corresponding functional environment. AAL-1 is suitable for CBR (constant bit rate) type of circuit switched data transfer. AAL-2 is designed to, for example, low bit rate video transfer, and is suitable for connection-oriented VBR (variable bit rate) data transfer. AAL-3 and AAL-4 (commonly referred to AAL-3/4) are suitable, for example, for the connectionless or connection-oriented Frame Relay traffic. The most powerful variant is AAL-5 which is suitable, for example, for fast bit-rate local area networks. Table 7.2 shows the division of AAL levels into classes of A–D. In addition, there are examples about the networks that are suitable for corresponding AAL levels.

The underlying ATM layer of AAL takes care of, for example, the flow control and error correction according to the specifications of CCITT I.150 and I.361. The ATM layer handles the establishment of the connections via VP (virtual path) and VC (virtual channel) as presented in Figure 7.11. It should be noted that in a single ATM channel, there can be one or more virtual path, which in turn, may consist of one or various virtual channel.

Depending on the connection point, the ATM interface is called either UNI (user node interface) or NNI (network node interface). ATM Forum has defined various frame structures and thus different bit rates according to the transmission system used by ATM.

As an example, E1 or E3 are suitable for the physical transport for ATM, providing the data rates of 2,048 kb/s or 34,368/ kb/s. The STM-1 of SDH, in turn, provides with data rate of 155.52 Mb/s and STM-4 a data rate of 622.08 Mb/s.

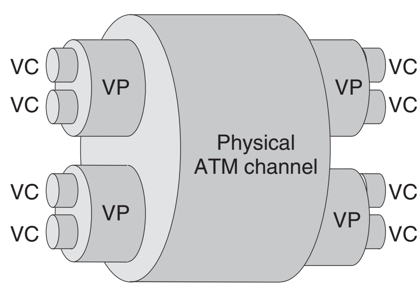

Figure 7.11 *The virtual paths (VP) of ATM can consist of one or more virtual channels (VC).*

7.6 Frame Relay

7.6.1 Definitions

Frame Relay (FR) is network technology specified by CCITT (currently ITU-T) and ANSI. Frame Relay is actually not a single network as such, but it may consist of various elements. FR network includes end-points, which can be, for example, computers and servers. There is also access equipment, which can be, for example, bridges and routers, as well as network equipment which can be, for example, centers and network routers. As FR is not a single network, it is often presented as a cloud in the technical presentations. Instead of physical connections, there are virtual route definitions in use.

Frame Relay is already relatively old technology. It was utilized still, for example, in the *Gb*- interface of the initial GPRS solutions between the GSM BSS and SGSN.

The frames of FR are based on the LAP-D protocol (link access protocol on D-channel) of CCITT Q.922 recommendation. The corresponding frame definition of ANSI is defined in T1.618 specification.

FR is suitable for data transfer of bursty nature because the length of the frame can vary from only a few up to over thousand characters. As an example, frames of IP and X.25 fit directly into the frame structure of FR, which means that there is no need to segment those separately into the FR frames.

FR utilizes PVC (permanent virtual circuits). They take care of the correct order of the packets in the receiving end. From one access point, there can be a connection to a single or various PVC tunnel which means that the parallel connections into several addresses in possible.

The data rate of FR can be 64 kb/s, 128 kb/s, 256 kb/s, 384 kb/s, 512 kb/s or 1984 kb/s. In the physical interface, there is V.35 applied to the connections below 2 Mb/s, and the fixed telephony G.703 or G.704 for the connections of 2 Mb/s.

7.6.2 Functionality

The Frame Relay network consists of DTE (data terminal) and nodes that can be, for example, router and server. DTE contains data field and a header with a total of 16 bits from which 10 bits are reserved for DLCI (data link connection identifier). There is also a bit called C/R (command / response) which separates the transfer and reception. There is also EA (extended address bits), which indicates if there are 3 or 4 bytes present. Figure 7.12 presents the Frame Relay frame.

The flow control of the header is called FECN/BECN (forward/backward explicit congestion notification). In addition, the header contains DE (discard eligibility indicator) and FCS (frame check sequence). Figure 7.12 clarifies the principle of the FR frame.

The connections of FR network are type VC (virtual circuit). These VCs are software defined routes between ports of FR network. There are 2 kind of virtual connections in FR: SVC (switched VC) connections and PVC (permanent virtual circuit) connections.

The idea of PVC is to maintain the connection between two connection points as fixed. This means that the communications are point-to-point. Although the connection points are physically the same, their data transfer can utilize variable routes depending on the situation. The utilization of PVC requires preplanning because the respective capacity can not be increased during the connection.

SVC is defined separately for each connection. It is thus more complicated to define SVC compared to PVC. On the other hand, the PVC connection is transparent for the end-user, meaning that there is no need to define connection by the user. In SVC connection, the network provides the required capacity for the user depending on the need. At the same time, it takes into account the resource utilization of various users.

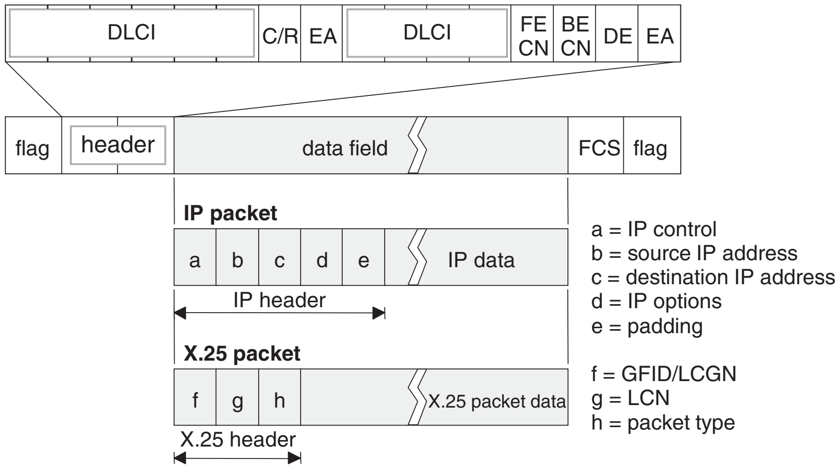

Figure 7.12 *The Frame Relay frame.*

7.7 LAN and MAN

Table 7.3 shows the list of solutions developed under IEEE 802 standardization. Some of the standards are related to access methodology whereas other part complements those, for example, by describing additional security measures.

7.7.1 IEEE 802.1 (Bridging)

IEEE 802.1 refers to the bridging (networking) and network management. The respective working group takes care of the definition of 802 LAN/MAN architecture, internetworking between 802 Local Area Networks (LAN), Metropolitan Area Networks (MAN) and Wide Area Networks (WAN). Furthermore, the working group has responsibilities related to 802 Link Security, 802 overall network management, and protocol layers above the MAC & LLC layers.

The complete list of the current activities of 802.1 working groups can be found in Ref. [1].

7.7.2 IEEE 802.2 (LLC)

IEEE 802.2 refers to a software component of a computer network, defining the Logical Link Control (LLC) of the OSI Model. More detailed description of the IEEE 802.2 can be found in Ref. [2].

7.7.3 IEEE 802.3 (Ethernet)

7.7.3.1 Principle

The currently most typical way to deploy local area data network is based on Ethernet. It is based on the carrier sense multiple access with collision detection (CSMA/CD) which means that the user who is sending data reserves the resource whilst the others wait for the liberation of the line.

The history of Ethernet begins from 1970, with the ALOHA radio network that was built in order to develop the principles of ARPANET. As a result of the joint development work of Xerox, and later DEC and Intel

Table 7.3 IEEE 802 versions for wired and wireless local area networks

IEEE recommendation	Description
IEEE 802.1	Connectivity, bridging and network management, LAN/MAN architecture, link security
IEEE 802.2	LLC layer (inactive)
IEEE 802.3	LAN CSMA/CD (Ethernet)
IEEE 802.4	LAN Token bus (discontinued)
IEEE 802.5	LAN Token Ring, MAC layer (discontinued)
IEEE 802.6	Metropolitan Area Network, MAN (discontinued)
IEEE 802.7	Broadband LAN using Coaxial Cable, Definitions for the broad band use (discontinued)
IEEE 802.8	Fiber Optic TAG, Definitions for the optical fiber use (discontinued)
IEEE 802.9	Integrated Services LAN (ISLAN or isoEthernet), Definitions for the connection of voice and data networks (discontinued)
IEEE 802.10	Security aspects of LAN (discontinued)
IEEE 802.11	Wireless LAN (WLAN)
IEEE 802.12	100BaseVG, Priority according to the demand (100VG-AnyLAN) (discontinued)
IEEE 802.13	Not assigned
IEEE 802.14	Cable modems (discontinued)
IEEE 802.15	Wireless PAN
IEEE 802.15.1	Bluetooth certification
IEEE 802.15.2	Coexistense of IEEE 802.15 and IEEE 802.11
IEEE 802.15.3	High-Rate wireless PAN
IEEE 802.15.4	Low-Rate wireless PAN (e.g., ZigBee, WirelessHART, MiWi, etc.)
IEEE 802.15.5	WPAN Mesh networking
IEEE 802.15.6	Body area network, BAN
IEEE 802.16	Broadband Wireless Access (WiMAX certification)
IEEE 802.16.1	Local Multipoint Distribution Service
IEEE 802.17	Resilient packet ring
IEEE 802.18	Radio Regulatory TAG (Technical Advisory Group)
IEEE 802.19	TAG Coexistence
IEEE 802.20	Mobile Broadband Wireless Access
IEEE 802.21	Media Independent Handoff
IEEE 802.22	Wireless Regional Area Network
IEEE 802.23	Emergency Services Working Group
IEEE 802.24	Smart Grid TAG
IEEE 802.25	Omni-Range Area Network

(DIX), the first principles of the LAN were born in 1980. At the same time, IBM developed parallel principles of Token Ring.

The first official version of Ethernet was defined in IEEE 802.3 by 1983. It provided a basic bit rate of 10 Mb/s, and there were four different commercial variants. The MAC layer of all these variants uses CSMA/CD (carrier sense multiple access/collision detection). The channel reservation is done based on the existence of the other users on the line. In the case of collision, that is, when more than one user sends data at the same time, each user waits a time period that is defined randomly.

The total length of the original 802.3 network, including the individual segments and routers, is 2.5 km. The first Ethernet system was based on the thick Ethernet cable and is called 10Base-5. The no. 5 refers to the maximum length of a single segment which is about 500 meters. The other variant is called thin Ethernet, that is, 10Base-2. The no. 2 in this case refers to the maximum length of a single segment which is about 200 meters.

The Ethernet network is based on the terminal equipment that is connected in a parallel manner. The end-points of the Ethernet network need terminal resistors of 50 ohms, and the protecting shield of the cable is one of these that must be grounded. The resistors attenuate the reflections and resulting interferences.

Evolved variant is 10Base-T. It is based on concentrator from which a cable is connected to each terminal that is located a maximum of 100 meters to the concentrator. The benefit of this system is that there is a possibility of using the network also for the higher speed transfer than 10Base solutions.

The capacity of 10 Mb/s Ethernet is especially limited. This can be seen by the saturation as a function of the load. This is a result of the collisions of data packets which trigger a random delay for each user, the aim being that each user eventually could access the network. The optimal throughput can be achieved when the load of Ethernet is about 40 percent.

The demand of data transfer has caused the further development of the Ethernet concept. The next step in this evolution path was Fast Ethernet that provided 100 Mb/s data rate. A parallel intention of FDDI would have provided the same data rate but it was noted to be too complicated and expensive in practice. Compared with the previous variant, the benefits of 100 Mb/s Ethernet was a 10-fold data rate as well as better tolerance against the errors. The actual data transfer can be done with 10BaseT cables, and also the frame structure is identical with the corresponding network definitions.

The 100 Mb/s Ethernet network is called 100Base-T. It has 5 different variants: MAC layer, independent information transfer (MII) and 3 physical variants that are 100BaseTX, 100BaseT4 and 100BaseFX.

The shorter the segment is in the MAC layers CSMA/CD system, the faster the data transmission rate is. The data rate can also be increased by shortening the space between frames. The evolution of the data rates of Ethernet networks is based on these variants.

The definitions of Fast Ethernet networks include the interfaces of MAC layer and physical 100 Mb/s Ethernet systems. This optimizes the data rate.

From the variants of Fast Ethernet, 100BaseTX is based on the 10Base-T network topology with cobber cables, and it provides with a data rate of 100 Mb/s and signaling rate of 25 Mb/s. The network topology is star, which means that the concentrator manages the user traffic in both directions. Each direction occupies its own pair of cables which eliminates the collisions of the data packets.

Figure 7.13 shows an example of PCMCIA LAN card that was utilized typically with laptops prior to USB standards [3]. Nowadays this type of equipment is not supported any more.

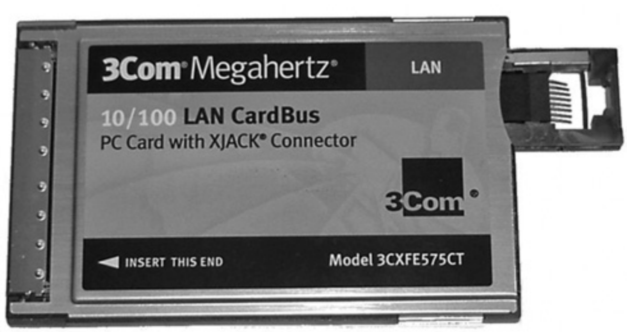

__Figure 7.13__ An example of older, 10/100 Mb/s LAN card that has been used via PCMCIA. This model had connector that was possible to pull out to connect the LAN cable. Currently, PCMCIA LAN cards are replaced by either integrated LAN equipment or by external equipment based on USB connectivity [3]. Photo by Jyrki Penttinen.

The variant that is based on fiber optics is called 100Base-FX. Thanks to the optical transfer, a single segment is up to 2 km if no other changes in the network are applied.

The next step in the evolution has been Gigabit Ethernet. This is an extension of the previous solutions in such a way that it provides a data rate of 1 Gb/s in either half duplex or full duplex mode. The definitions are backwards compatible with basic and fast Ethernet networks. The Gigabit half duplex version still includes the CSMA/CD method, although with minor modifications. The maximum length of the segments are thus backwards compatible, that is, 100 meters from concentrator, which means 200 meters between two terminals. When maximum data rate is used.

7.7.3.2 Deployment of Ethernet

The basic elements of Ethernet type of local area networks can be seen in Figure 7.14, and they are:

- User equipment (terminal) is typically portable or desktop PC. These are also called clients.
- Terminals are connected to LAN via adaptors.
- The peripherals are connected to the LAN in the same manner as terminals, via LAN adapters. In practice, the adapter is typically integrated into the peripheral. Commonly used peripherals include network printers and mass storages.
- The operation system of the network must be compatible with both client and server equipment in order to transfer data. Some examples of the OS are Microsoft Windows NT, Novell NetWare and Apple AppleShare.
- LAN server is in practice a computer taking care of the data communications between terminals. Based on the task that the server performs, it can be called, for example, application server, printer server, data transfer server and so on.

7.7.3.3 Current Activities

The IEEE 802.3 work groups have produced actively standards since the 1970s, along with the first experimental Ethernet definitions produced in 1973. The first specification under the name IEEE 802.3 was produced in 1983, and it defined the original 10BASE5 10 Mbit/s data rate over a thick coaxial cable according to the CSMA/CD process. 802.3a specification followed in 1985 which defined 10BASE2 10 Mbit/s over thin coaxial cable. Ever since, there have been tens of 802.3 specifications defining, for example, repeaters, 10BASE-T over twisted pair, 10BASE-F over Fiber-Optic, 100BASE-T2 over low quality twisted pair, 1000BASE-X

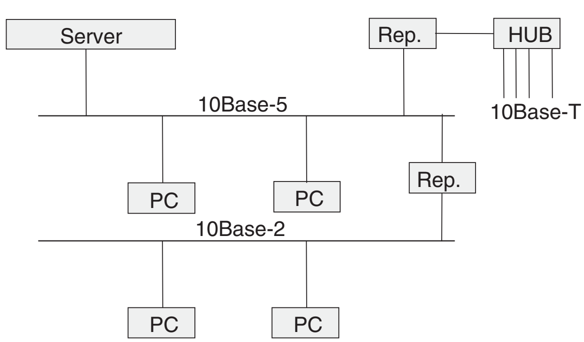

Figure 7.14 *An example of the Ethernet connections.*

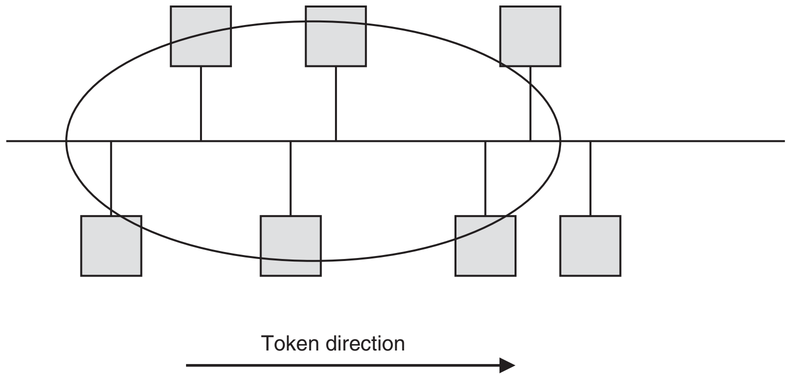

Figure 7.15 The principle of the token bus with token passing.

Gbit/s Ethernet over Fiber-Optic, and 10GBASE-LRM 10 Gbit/s Ethernet over multimode fiber. One of the most relevant recent additions is 802.3ba from 2010, which defined 40 Gbit/s and 100 Gbit/s Ethernet.

Some examples of the ongoing activities are IEEE P802.3bj that refers to the 100 Gb/s Backplane and Copper Cable Task Force, and IEEE P802.3bm which refers 40 Gb/s and 100 Gb/s Fiber Optic Task Force.

The complete list of 802.3 activities can be found in Refs. [4] and [5].

7.7.4 IEEE 802.4 (Token Bus)

IEEE 802.4 defines Token Bus principle as presented in Figure 7.15. It is a network solution that is based on a token ring protocol, the physical layer being a coaxial cable. The basic functionality is such that the token travels through the network nodes. The node that possesses the token is admitted to transmit. In case the node does not have contents to transmit, the token travels to the next node according to a virtual ring principle. The virtual ring is possible as all the nodes must know the address of their neighboring node. The task of the token bus protocol is thus to notify the other nodes about the connecting and disconnecting of nodes.

Token bus has never been a solution for large audience, and its use has been limited to the industrial environment. In fact, experiences from practical maintenance and operation were not promising at the time it was deployed. There were intentions to enhance functionality, but currently the working group is disbanded and no related standardization efforts are active.

7.7.5 IEEE 802.5 (Token Ring)

IEEE 802.5 defines Token Ring as presented in Figure 7.16. It is a parallel solution with the Ethernet concept. The market share of Token Ring has been considerably lower than LAN solutions already in the initial phase of the deployments. Token Ring has been used mainly in the network environment of IBM.

The basic version of Token Ring provides data rate of 4–16 Mb/s, and the evolved variants 100 Mb/s and 1 Gb/s. In a single loop, it is possible to connect 70–260 terminals, depending on the cable type. The physical solution can be straight cobber cable pair, or fiber optics.

The Token Ring concept is based on a logical ring in such a way that both server and terminals are connected to the same ring according to the IEEE 802.5 standard. The actual ring is formed by a set of MAU (multi station access unit). It is possible to connect up to 8 terminals to a single MAU which forms a star topology. The connection is done via adapter cables, and when MAUs are connected, it forms a full star. The order of

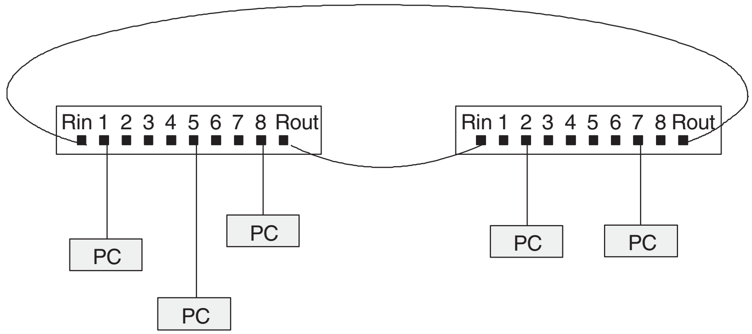

Figure 7.16 *The principle of Token Ring network.*

transfer is dictated by a token. When the same token returns to the original destination after a full round, the respective message is removed by the sender.

More information about IEEE 802.5 can be found in Ref. [6].

7.7.6 IEEE 802.6 (MAN)

IEEE 802.6 defines Metropolitan Area Networks. The aim of the standard is to improve the previous ANSI-defined FDDI network (Fiber Distributed Data Interface). The original FDDI standard was not successful as it was too expensive to implement. The IEEE 802.6 standard is based on DQDB (Distributed Queue Dual Bus), which provides data rates up to 150 Mb/s, and the network can handle distances up to 160 km via optical fiber of 1310 nm wave length. The idea of IEEE 802.6 network is based on two unconnected unidirectional buses.

Also this standard failed, for the same reasons as the FDDI. Instead of FDDI, the current Metropolitan Area Networks are deployed typically via SONET (Synchronous Optical Network) or ATM (Asynchronous Transfer Mode) concepts. The most actual deployments use also native Ethernet or MPLS.

7.7.7 IEEE 802.7 (Broadband LAN via Coaxial)

IEEE 802.7 covers broadband local area networks, and more specifically it defines Broadband LAN using Coaxial Cable.

Since the working group issued its recommendation in 1989 it has not been involved in any activities.

7.7.8 IEEE 802.8 (Fiber Optic TAG)

IEEE 802.8 refers to a Fiber Optic Technical Advisory Group (TAG) which had the task of developing a LAN standard for fiber optic media used in token pass computer networks from which FDDI and token bus were examples.

As the token pass networks did not take off, related activities have not been pursued in this group.

7.7.9 IEEE 802.9 (ISLAN)

IEEE 802.9 refers to a Working Group that developed standards for integrated voice and data access over category 3 twisted-pair network cable installations. The most significant standard has been known as isoEthernet. The topic of this group can be called Integrated Services LAN (ISLAN or isoEthernet).

The basic idea of isoEthernet is to combine an Ethernet of 10 Mb/s and 96 times ISDN-B channels with 64 kb/s data rate of each.

Regardless of light vendor support, isoEthernet did not turn out to be successful because Fast Ethernet advanced on the other side commercially. This working group was thus discontinued.

7.7.10 IEEE 802.10 (Interoperable LAN Security)

IEEE 802.10 is a former standard that is not valid any more. The idea of this standard was to define security functions usable in local area networks as well as in metropolitan area networks, based on IEEE 802 protocols.

802.10 standards defined security association management and key management, access control, data confidentiality and data integrity. Nevertheless, the IEEE 802.10 standards were withdrawn during 2004 and the respective working group is no longer active. In any case, wireless networks security was standardized in 802.11i.

7.7.11 IEEE 802.11 (WLAN)

The IEEE 802.11 definition is in fact a set of standards in order to implement wireless local area network (WLAN) communications, the radio frequency bands being 2.4 GHz, 3.6 GHz and 5 GHz. This set is developed and maintained by the IEEE LAN/MAN Standards Committee (IEEE 802).

The IEEE 802.11-2012 standard has been complemented by subsequent amendments. This complete set of standards acts as a base for wireless network products using the Wi-Fi brand. The set of standards of IEEE 802.11 are named with letters, the original Wi-Fi variant being without this extension. The variants are shown in Table 7.4. In addition, there are several other variants like TV White Space amendment IEEE 802.11af. The updated list of variants can be found in http://standards.ieee.org.

Table 7.4 *The IEEE 802.11 variants*

802.11 variant	Release date	Frequency / GHz	Bandwidth / MHz	Data rate Mbit/s per stream	Modulation method
802.11	June 97	2.4	20	1 / 2	DSSS / FHSS
802.11a	September 99	5 / 3.7	20	6 / 9 / 12 / 18 / 24 / 36 / 48 / 54	OFDM
802.11b	September 99	2.4	20	1 / 2 / 5.45 / 11	DSSS
802.11g	June 2003	2.4	20	6 / 9 / 12 / 18 / 24 / 36 / 48 / 54	OFDM / DSSS
802.11n Narrower band (1)	October 2009	2.4 / 5	20	7.2 / 14.4 / 21.7 / 28.9 / 43.3 / 57.8 / 65 / 72.2	OFDM
802.11n Wider band (1)	October 2009	2.4 / 5	40	15 / 30 / 45 / 60 / 90 / 120 / 135 / 150	OFDM
802.11ac	November 2011	5	20	Max. 87.6	OFDM
	November 2011	5	40	Max. 200	OFDM
	November 2011	5	80	Max. 433.3	OFDM
	November 2011	5	160	Max. 866.7	OFDM
802.11ad (WiGig)	December 2012	2.4 / 5 / 60	2160	Max. 7000	OFDM

Note:
(1) This variant allows a maximum of 4 streams.

7.7.12 IEEE 802.12 (100BaseVG)

100BaseVG (Voice Grade) is a 100 Mbit/s Ethernet standard which is meant to work over four pairs of category 3 UTP wires. This variant is also known as 100VG-AnyLAN as it is able to handle Ethernet as well as Token Ring frames.

100BaseVG has already disappeared from the markets. It was ratified by the ISO during 1995 but did not manage to succeed to the following decade.

7.7.13 IEEE 802.13 (Unused)

This variant is not defined in IEEE.

7.7.14 IEEE 802.14 (Cable Modems)

IEEE 802.14 defines the cable modem. It is in practice a network bridge providing bidirectional communication on a hybrid fiber-coaxial (HFC) and RFoG. It is based on radio frequency. As the HFC and RFoG provide high bandwidth, it is a suitable base for broadband Internet access for households.

7.7.15 IEEE 802.15 (Wireless PAN)

Wireless PAN refers to the personal area network in very short distances. It is a computer network which provides communications between terminals that can be, for example, computers, PDAs and telephones. The PAN communications can be created between the terminals as such which is referred to as intrapersonal communications. The communications can also be done towards higher level network in uplink.

A Wireless PAN (WPAN) refers to a PAN that can utilize any physical wireless network solution as a base, including Bluetooth, infrared (IrDA), Wireless USB, ZigBee or BAN. There are no limitations as such for the physical solution as it merely functions as a bit pipe for WPAN. For this reason, the underlying network can also be PAN or other wired solution like USB and FireWire.

The useful coverage area of PAN can vary typically between only a few centimeters up to a few meters.

The Wireless PAN is suitable, for example, for medical purposes for transferring telemetry data from the human body. It would also work, for example, for intelligent clothing.

The task groups of Wireless PAN are listed in Table 7.5.

7.7.15.1 IEEE 802.15.1 (Bluetooth)

Bluetooth technology is defined in the first task group of 802.15. This group has defined the physical layer of Bluetooth, as well as Media Access Control (MAC). The specification is valid for wireless connectivity for a variety of devices, including fixed, portable and moving devices. The first standards are from 2002.

7.7.15.2 IEEE 802.15.2 (Coexistence of 802.11/15)

IEEE 802.15.2 defines the coexistence of two WPANs, IEEE 802.15 and IEEE 802.11.

7.7.15.3 IEEE 802.15.3 (High-Rate WPAN)

IEEE 802.15.3 defines the MAC and physical layer of the High-Rate Wireless PANs, with the data rate varying between 11 and 55 Mb/s. The 802.15.3a was meant to include the physical layer definition of Ultra

Table 7.5 The IEEE 802.15 WPAN task groups

Task group	Description
1	WPAN and Bluetooth
2	Coexistense
3	High Rate WPAN
4	Low Rate WPAN
4.1 (4a)	WPAN Low Rate Alternative PHY
4.2 (4b)	Revision and Enhancement
4.3 (4c)	PHY Amendment for China
4.4 (4d)	PHY and MAC Amendment for Japan
4.5 (4e)	MAC Amendment for Industrial Applications
4.6 (4f)	PHY and MAC Amendment for Active RFID
4.7 (4g)	PHY Amendment for Smart Utility Network
5	Mesh networking
6	Body Area Networks
7	Visible light communication
8	Wireless Next Generation Standing Committee

Wide Band (UWB) for enhanced multimedia purposes, but due to the disagreements of the more detailed access technology, the work was discontinued.

Instead, IEEE 802.15.3b-2005 was created in order to enhance the functioning, implementation and interoperability of MAC, yet being backwards compatible with the original IEEE 802.15.3. Furthermore, there has been an IEEE 802.15.3c which provides an addition of a millimeter wave option to the physical layer for 57–64 GHz band. This solution provides about 2 Gb/s data rate which is sufficient, for example, for streaming of high definition television contents. Furthermore, there is an option for 3 Gb/s data rate.

7.7.15.4 IEEE 802.15.4 (Low-Rate WPAN)

In contrast for the high data rate WPANs, the IEEE 802.15.4 defines a low-rate wireless PAN. Some practical forms of this definition are ZigBee, WirelessHART and MiWi.

The idea of this low bit rate solution is to increase the battery durability whilst the technical complexity is maintained low. The battery life could be considerably longer than in typical WPAN variants, up to months or years.

IEEE 802.15.4-2003 defines the base for the physical and data link layers. It is possible to use various networks on top of IEEE 802.15.4, including the Mesh NWs like IEEE 802.15.5, WirelessHART and ZigBee.

7.7.15.5 IEEE 802.15.5 (Mesh NW for WPAN)

Mesh networking for WPAN is defined in IEEE 802.15.5. Mesh networking refers to the architectural model that provides interoperable, scalable and stable wireless mesh networks.

IEEE 802.15.5 defines low-rate WPAN mesh network (based on IEEE 802.15.4-2006 MAC), and high-rate WPAN mesh network (based on 802.15.3 or 802.15.3b).

7.7.15.6 IEEE 802.15.6 (Body Area NW)

Body area network (BAN) is meant for low-power and short-range wireless networking. It provides means for optimal use of devices in and around human body, yet not limiting to the human body environment. Some of the focus applications are related to the medical solutions, as well as on the personal entertainment and in general consumer electronics.

7.7.15.7 IEEE 802.15.7 (Visible Light)

IEEE 802.15.7 defines Visible Light Communication (VLC) for physical and MAC layers. The idea of this technology is to provide means to transfer data, for example, within a room by emitting light from the roof to the terminals on the table. The technology requires basically a line of sight (LOS) connection between transmitter and receiver. Li-Fi is a subset of VLC.

IEEE 802.15.7 defines the short-range optical wireless communications using visible light in a spectrum of 380–780 nm wavelength. The aim of the standard is to provide possibility to transfer such data rates that they are sufficient to support audio and video multimedia services. In addition, the standard considers mobility of the visible link at the reasonable extend. The challenges that the standard is tackling are the compatibility with visible-light infrastructures and effects of the noise and interference from, for example, ambient light. Important aspect in this standard is the compatibility to the health and other environmental regulations, meaning that the standard will adhere to any applicable eye safety regulations [7].

The motivation for this standard arises from the fact that the visible light is noted to serve as a useful communication medium due to the further developments in scientific fields. In fact, solid-state light sources are replacing conventional sources in signaling, illumination and display infrastructures [7]. This leads to the possibility of using light sources as physical bearers for the signals. Furthermore, the visible light frequency band is not regulated unlike the radio frequency spectrum, and does not have correlation with RF interferences which makes it an interesting option in the highly restricted RF environment.

As communications may happen within a relatively short distance and in LOS, for example, within the same room, the technique may provide good physical layer security. Furthermore, it can be assumed that the standard provides an economical way to communicate due to the low complexity level of the systems.

Some of the possible applications of this technique include secure point-to-point communication, indoor Location Based Service (LBS), secure point-to-multipoint communication in closed environments such as offices, hospitals and inside of airplane, as well as Intelligent Transportation System (ITS) and in short-distance broadcast [7].

7.7.16 IEEE 802.16 (WiMAX)

Broadband Wireless Access, that is, WiMAX is defined in IEEE 802.16. It is in fact a set of wireless broadband standards of IEEE. The root for WiMAX proceeds from 1999 when IEEE initiated the work on drafting Wireless Metropolitan Area Networks (WMAN).

The standard terminology for IEEE 802.16 is Wireless MAN. Nevertheless, in practice, the solution based on the standard is called WiMAX (Worldwide Interoperability for Microwave Access). This name is a result of the work of WiMAX Forum industry alliance.

The task of WiMAX Forum is to promote and certify compatibility and interoperability of products based on IEEE 802.16 standards.

The IEEE 802.16 evolves, and an amendment 802.16e-2005 was released for the next wave of WiMAX solutions.

IEEE 802.16.1 refers to the Local Multipoint Distribution Service (LMDS). This is a broadband wireless access technology which was initially planned for digital television environment (DTV). The idea of LMDS is to be used in fixed wireless point-to-multipoint solutions, to replace the so-called last mile for the delivery of content to end-users.

LMDS typically functions in microwave bands around 26–29 GHz frequency, or in the case of the USA, also within 31.0–31.3 GHz.

The modulation method of LMDS can be phase-shift keying (PSK) or amplitude modulation (AM), which dictates the achieved data rate. The coverage area of LMDS is typically around 2–2.5 km. Nevertheless, it is possible to optimize the radio link by utilizing highly directional antennas in order to achieve longer distances.

7.7.17 IEEE 802.17 (Resilient Packet Ring)

IEEE 802.17 defines Resilient Packet Ring (RPR). It has been designed for optimized data transmission over optical fiber ring networks.

RPR provides resilience (50 ms protection) that is comparable with SONET/SDH networks. Nevertheless, instead of circuit switched communications, RPR is based on packet data transmission.

RPR is based on dual counter rotating rings. They are also called ringlets, which are formed by creating PRP stations at nodes where traffic is expected to be dropped per flow. PR is based on MAC protocol messages for the traffic direction.

7.7.18 IEEE 802.18 (Radio Regulatory TAG)

The IEEE 802.18 refers to the Radio Regulatory Technical Advisory Group, that is, RR-TAG. It is a working group taking care of LAN/MAN Standards. The current status of the working group and the active standards it is taking care of can be found in Ref. [8].

The task of the RR-TAG is to monitor the interests of the active projects in national and international level. It also prepares comments and policy recommendations about the subject matters to regulators which, in turn, takes care of the equilibrium of the interests of the respective projects.

7.7.19 IEEE 802.19 (Coexistence TAG)

IEEE 802.19 refers to the Wireless Coexistence Technical Advisory Group (TAG) within the IEEE 802 LAN/MAN Standards Committee. This Technical Advisory Group takes care of the coexistence of the unlicensed wireless networks.

In practice, several of the IEEE 802 wireless standards operate in unlicensed spectrum which means that there may be several systems sharing the same frequency band, and there is thus a need to handle the coexistence in a controlled way in order to minimize the respective interferences.

7.7.20 IEEE 802.20 (Mobile Broadband Wireless Access)

IEEE 802.20 refers to the Mobile Broadband Wireless Access (MBWA). It has been a standard association specification for mobile Internet access networks.

7.7.21 IEEE 802.21 (Media Independent Handoff)

IEEE 802.21 refers to the Media Independent Handoff. This standard supports algorithms which are capable of providing seamless handover between the same type of networks. Furthermore, the standard provides possibility to use handover between different network types, which is referred as Media Independent Handover (MIH), or vertical handover.

This standard allows handovers to and from IEEE 802.3, 802.11, 802.15 and 802.16 networks as well as 3GPP and 3GPP2 networks.

7.7.22 IEEE 802.22 (Wireless Regional Area Network)

IEEE 802.22 refers to the Wireless Regional Area Network (WRAN) which utilizes the white space of the TV frequency spectrum.

IEEE 802.22 WRAN focuses on the use of Cognitive Radio (CR) techniques in order to provide sharing of unused spectrum allocated in different regions of the Broadcast Television Service. The idea is to reutilize the capacity in such a way that the use would not cause interferences. The benefit of this idea is to provide broadband access in geographical areas that are challenging to cover otherwise due to the difficult topology or for being remote areas with low density of population.

WRAN is the first global initiative to provide a standard radio interface which would be based on the principles of the Cognitive Radio and opportunistic use of TV bands in such a way that the result is interference-free.

The main focus of the technology is to reutilize the unused TV bands of the broadcast networks yet assuring that the systems operating at the same band are not disturbed, that is, digital and analog TV, as well as the licensed low power devices like wireless microphones.

The sub-publications of this standard are IEEE P802.22.1 to enhance harmful interference protection for low power licensed devices operating in TV Broadcast Bands, and IEEE P802.22.2 to act as a recommended practice for the installation and deployment of IEEE 802.22 Systems.

7.7.23 IEEE 802.23 (Emergency Services Working Group)

IEEE 802.23 refers to Emergency Services Working Group.

7.7.24 IEEE 802.24 (Smart Grid TAG)

IEEE 802.24 refers to the Smart Grid Technical Advisory Group (TAG). The tasks of this group are to act as a liaison and point of contact with, for example, regulatory agencies, industry organizations, government agencies and IEEE societies, in order to discuss the use of 802 standards in Smart Grid applications. The group also has many other high level tasks, such as facilitation of coordination and collaboration between the IEEE 802 groups and obtaining presenters upon need for discussing IEEE 802 standards in Smart Grid applications. In addition, the group produces information packages about the subject matter.

7.7.25 IEEE 802.25 (Omni-Range Area Network)

The IEEE 802.25 refers to the Omni-Range Area Network.

7.8 Wi-Fi

WLAN (Wireless Local Area Network), which also is known as a brand Wi-Fi, as we know it today is based on the IEEE 802.11 standards. IEEE developed standards in 1990s to provide a Wireless LAN solution. The solution uses unregulated 2.4 and 5 GHz frequencies and so it was possible to use this solution basically everywhere without regulator permission. It was the first time when an affordable wireless solution was available basically for everyone. You could create a Wi-Fi network of your own, to your home, office, everywhere, and this does not require specific skills.

In early 2000 the cellular networks were using GPRS technology and data speeds were typically 30–40 kb/s so it was no surprise that users were really interested to use Wi-Fi. Also the price of the cellular data was very high, especially when roaming. So, it was no wonder that high data speed with low cost also lead to interest in WLAN technology among MNOs (Mobile Network Operators) as well as creating new wireless operators: Wireless Service Providers, WISP. All these were providing Wi-Fi services, first in their own market and later implementing also a possibility for their customers to roam in other operators' Wi-Fi networks.

The first Wi-Fi implementations were usually Wi-Fi PCMCIA cards or USB cards targeted to be used in the laptops. Nowadays almost all devices have an integrated Wi-Fi card, either on motherboard or a separate module on motherboard. That has also lead to smaller power consumption. We could say that the Wi-Fi is the most supported wireless technology in the portable consumer devices today.

There are 4 generations of the WLAN, described in the Table 7.6. First one that was implemented was 802.11b. It provided theoretical throughput of 11 Mb/s on 2.4 GHz frequency band. At the same time IEEE published 802.11a standard that provided 54 Mb/s throughput on 5 GHz band. It has been used more in the corporate environment. The next standard was 802.11g on 2.4 GHz band providing theoretical throughput of 54 Mb/s. This technology might be the most used technology at the moment among the Wi-Fi users. The newest standard is 802.11ac and having as high as 600 Mb/s theoretical throughput. Naturally theoretical throughputs are not achieved in real environments, maybe two-thirds or half of the theoretic speed.

These standards are backward compatible, so usually Access Points can negotiate the speed according to the technology the device supports.

7.8.1 Standardization

7.8.1.1 Institute of Electrical and Electronics Engineers (IEEE)

IEEE 802.11 is an IEEE specification for the WLAN. It is the family of the specifications, the most common versions implemented are 802.11b (11 Mb/s) and 802.11g (54 Mb/s) at the moment. IEEE is developing

Table 7.6 *WLAN generations*

Wi-Fi variant	Frequency Band	Maximum data rate (Mb/s)
802.11a	5 GHz	54
802.11b	2.4 GHz	11
802.11g	2.4 GHz	54
802.11n	2.4 GHz, 5 GHz; selectable between 2.4 and 5 GHz; Concurrent 2.4 and 5 GHz	600

WLAN continuously, one example being 802.11n specification which provides theoretically speed of 600 Mb/s, but the real speeds are something like 100–200 Mb/s, depending the environment. This version can be seen to provide the same speed as 100 Mb/s Ethernet. 802.11n is also backwards compatible, providing compatibility mode speeds according to for example, 802.11g or 802.11a.

802.11n also supports MIMO (multiple-input, multiple-output) technology, like LTE does. This technology takes advantage of using multiple antennas as well as multiple channels to achieve higher throughput. 802.11 includes also several amendments, like 802.11u.

IEEE has also developed 802.1X, Port Based Authentication protocol. 802.1X protocol runs between the Mobile Terminal and the Access Point for the purposes of authentication. The Port Access Entity (PAE) implemented on the Access Point is responsible for blocking all user traffic until the user is successfully authenticated.

7.8.1.2 The Internet Engineering Task Force (IETF)

IETF develops and promotes mainly protocols and leaves the system architecture work to the other organizations, like to the 3GPP. It has developed protocols like TCP/IP, RADIUS, Diameter and Extensible Authentication Protocol (EAP). Initially IETF was US government activity, but early 1990s it was changed to be an independent, international activity associated with the Internet Society.

7.8.1.3 3G Partnership Program (3GPP)

3GPP scope is the interworking of WLAN with other cellular technologies. As 3GPP-WLAN interworking concentrates on interfaces between 3GPP elements and the interface between the 3GPP system and the WLAN, the internal operation of the WLAN is only considered in order to assess the impact of architecture options/requirements on the WLAN. 3GPP-WLAN interworking shall be independent of the underlying WLAN Radio Technology and defines two interworking scenarios, WLAN Direct IP Access and WLAN 3GPP IP Access.

7.8.1.4 GSM Association (GSMA)

GSMA is an association of Mobile Network Operators (MNO) and related companies devoted to supporting the standardizing, deployment and promotion of the GSM mobile telephone system. It does not create specifications itself, but makes recommendations and best practices based on existing specifications and standards.

GSMA is handling Wi-Fi from the interconnection and roaming point of view. It recommends in the IR.61 document a common technical solution for roaming between Wi-Fi Service Providers. Authentication methods recommended are layer 3 WEB username/password authentication, EAP-SIM or EAP-AKA authentication.

The recent GSMA activity is a joint project with the WBA about to speed up the EAP-based authentication method implementations in roaming environment, called the Wi-Fi Roaming Task Force. This activity also includes implementation of the new HS2.0 specification and how two roaming entities, GSMA IPX and WRIX can cooperate in the future.

7.8.1.5 Wi-Fi Alliance (WFA)

WFA is a nonprofit industry association to work for adoption of Wi-Fi worldwide. It works for the technology development, market building, and regulatory programs around the Wi-Fi. WFA works through the certification programs. It provides a widely recognized designation of interoperability and quality and helps to ensure that Wi-Fi-enabled products deliver the best user experience. The Wi-Fi Alliance has completed more than 15 000 product certifications, encouraging the expanded use of Wi-Fi products and services in new and established markets. WFA provides a "Wi-Fi certified" logo to those products that fulfill their requirements, that is like a guarantee for the product to work in different Wi-Fi networks.

The recent WFA activity is to enhance Wi-Fi network discovery, to follow more cellular network automatic discovery implementation. The idea is to ease up how to connect to different Wi-Fi networks. Specification is called HotSpot 2.0 (HS2.0) and it includes also usage of 802.1x and EAP-based authentication methods.

7.8.1.6 Wireless Broadband Alliance (WBA)

Wireless Broadband Alliance is an alliance that works to create specifications and guidelines to make Wi-Fi interworking and roaming more efficient and reliable for operators and service providers. Alliance includes Wi-Fi operators, transmission providers, Wi-Fi equipment vendors and others around the Wi-Fi services. Wi-Fi operators, MNOs, separate Wi-Fi service providers, are using WBA WRIX (Wireless Roaming Intermediary eXchange) infrastructure to implement Wi-Fi roaming services. WRIX provides for the operators WRIX-i (Interconnect), WRIX-d (Data Clearing) and WRIX-f (Financial Settlement) functionalities.

WRIX roaming is based on layer 3 WEB authentication, but the latest release of WRIX includes support for 802.1X and EAP authentication providing the transport and indicating which radius attributes to use. This specification recommends Network Operators to support EAP-SIM, EAP-AKA, EAP-TLS, and EAP-TTLS.

The WBA activity areas are for the following points of view:

- Customers and devices: To make the utilization of Wi-Fi as easy as possible for the end-users regardless of the device type.
- Operators and Hotspots: To create specifications and guidelines in order to make the deployment of Wi-Fi interworking and roaming fluent.
- Wireless technologies (WiMAX, 3G, LTE): To facilitate the interoperability of different technologies.

WBA scope is also to facilitate interoperability with other technologies, such as 3G, LTE and WiMax. For that purpose guidelines and specifications have been published.

7.8.2 Wi-Fi Authentication and Accounting

Before the user can authenticate and have access to the Internet, the device must be associated to the Wi-Fi network. This happens by scanning the available Wi-Fi networks, and selecting the one that is desired. Another way is to configure the desired Service Set Identifier (SSID) to the device and then select that to be connected. Certain devices just scan available Wi-Fi networks, and if they find open Wi-Fi network they just connect to it.

Wi-Fi networks can be configured to be open, available to all who just like to connect, or to require certain authentication to connect to. In the home environment users often let the network be totally open and SSID broadcasted, so that anyone can connect and use the connection. This is not recommended since it allows the misuse of the connection, allows anyone to see the data that is transferred and also to hijack the connection.

It is recommendable to configure the Wi-Fi access point of Wi-Fi networks to use a preshared key to restrict the possibility to connect to the network and to encrypt the data. To get connected to the Wi-Fi access point preshared key must also be configured correctly to the Wi-Fi device. The preshared key can be inserted manually to the Wi-Fi device, or it can be automatically populated in certain devices using for example, WPS (Wi-Fi Protected Setup) functionality. However, WPS is not recommended to be used, because of a major security flaw. When using a preshared key radio interface is also automatically encrypted so enhancing the security and privacy.

First generation WI-Fi encryption was WEP (Wired Equivalent Privacy), but it was realized to be weak, so the enhanced solution WPA (Wi-Fi Protected Access) was developed. It uses TKIP (Temporal Key Integrity Protocol) and RC4-based ciphering. TKIP employs a per-packet key, meaning that it dynamically generates a new 128-bit key for each packet and thus prevents the types of attacks that compromised WEP. WPA supports also optionally AES ciphering.

In the new devices WPA2 (Wi-Fi Protected Access 2) security protocol, based on IEEE11i, has replaced WPA. As a new feature it introduces CCMP, an AES based strong encryption. There are no known successful attacks towards AES.

It is also possible to configure the SSID to be hidden, so that the network name is not broadcasted and so not so easy to find. This might add security on certain cases, but monitoring of the data traffic might reveal hidden SSID for the person who knows what to do. Activating the MAC filtering also adds security, but requires maybe more than basic knowledge of Wi-Fi technology on home environment.

RADIUS protocol transmits also accounting data. It includes for example, connection time, data usage, location where data is used. The Wi-Fi Service Providers use this information for end-user billing. Wi-Fi Service Providers (SP) have different data models for their customers, like flat rate, usage-based, time-based, roaming and so on.

On early days it was common to have a billing by megabyte, since MNOs also used that kind of model in cellular networks. Hotels provided often time based like 1 hour or 24 hours vouchers or paid by the credit card. When the cellular data subscriptions got cheaper, Wi-Fi flat rate subscriptions became more popular. The most advanced MNOs also include in their mobile data package the Wi-Fi usage, so the customer does not need to care what kind of technology is used for the connection.

7.8.2.1 Username-Password WEB Authentication

The public Wi-Fi service operators are typically using layer 3 browser based username/password login. The device must be attached to Wi-Fi and it has received an IP-address. When the browser is opened the user is redirected the to web login page, where to insert correct credentials, username and password. In the roaming case the home operator must be typically selected from the drop down menu. Usually the login procedure uses an https connection but the data after secured login is open to be seen by everyone because of the lack of encryption on radio interface. Another problem is that once a Wi-Fi device is attached to the network, it gets the IP-address. This is using lots of IP-address resources when you must provide addresses to the devices that are not your customers and might never authenticate to your network.

Also automation of the authentication is not easy to implement even WFA has specified a WISPr attribute to ease authentication procedure. Many mobile operators have developed an operator branded connection manager to simplify the Wi-Fi authentication process but also to include 3G connection functionalities to the same software. WEB authentication is widely used at the moment and works pretty well in laptops and tablets. Using it in smartphones is quite challenging because it requires first opening the browser and then inserting username and password. Something better is needed for Wi-Fi usage in smartphones.

7.8.2.2 802.1X

Enhanced Wi-Fi authentication solution is based on IEEE standardized 802.1X, a port-based access control protocol, that can be used either in wireless or wired environment. It is running between the Mobile Terminal and the Wi-Fi Access Point to authenticate the user. The AP has also a RADIUS client function which is

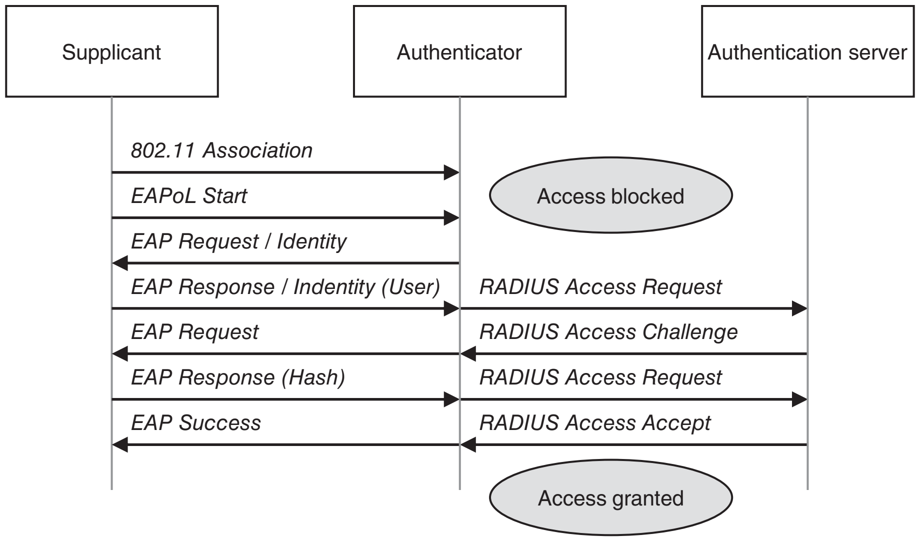

Figure 7.17 Flow chart of successful EAP authentication.

responsible for initiating the RADIUS protocol which is finally terminated on the Home Wi-Fi Network RADIUS Server. 802.1X brings some of the cellular network features, like encryption, to the Wi-Fi networks.

802.1X provides an authorization framework that allows or disallows traffic to pass through a port and thereby access network resources. This protocol enhances the authentication so that before a successful authentication there is only a port open for the EAP packets to authenticate the user and all other traffic is blocked until the authentication is completed. Only after successful authentication is the IP-address granted for the user and connection to Wi-Fi network allowed. This mechanism really reduces usage of the IP-addresses. After successful authentication AP sends RADIUS Accounting Start message to the RADIUS Server.

802.1X framework consists of three main components as shown in Figure 7.17:

1. **Supplicant** is a host with software, which requests authentication and access to the network. Usually it is a client in the terminal.
2. **Authenticator** is a device, which blocks or allows the traffic to pass the port. There are two ports: an uncontrolled port and a controlled port. The uncontrolled port allows EAP authentication traffic to pass through, while the controlled port blocks all other traffic until the supplicant has been authenticated. Typically this is a WLAN access point or an access controller, depending on the architecture of the network.
3. **Authentication Server (AS)** validates the credentials of the supplicant and notifies the Authenticator that the supplicant has been authorized. The Authentication Server can contain a database or it will proxy the authentication request to the appropriate database, for example, in EAP-SIM case to the HLR. Typically AS is a RADIUS server but Diameter servers are also available on the market.

In the radio interface between supplicant and authenticator EAPoL protocol (EAP over LAN) is used. It includes the EAP protocol to be transferred over the RADIUS protocol. Authenticator rips the EAP part of the EAPoL, and sets it over the RADIUS to communicate to the Authentication Server. The new protocol Diameter shall be used in this interface, too.

RADIUS and Diameter protocols support EAP-framework. EAP is an authentication framework, not a specific authentication mechanism. It provides some common functions and negotiation of authentication methods called EAP methods. Commonly used modern methods capable of operating in wireless networks include EAP-SIM, EAP-AKA, EAP-TLS, EAP-TTLS and EAP-LEAP. Requirements for EAP methods used in Wireless LAN authentication are described in RFC 4017.

802.1X introduces also the key exchange for the data encryption. As stated earlier, WEP encryption was realized to be weak. Based on 802.1X feature to deliver dynamically the encryption keys it is possible to use enhanced security of WPA and WPA2.

The new Wi-Fi networks on the MNO environment are usually supporting 802.1X. It provides better security, makes possible to use enhanced authentication methods and new Wi-Fi network discovery features.

7.8.2.3 EAP-SIM, EAP-AKA and EAP-AKA′

The success of smartphones has also created a need for the data Offloading. Maybe the most essential success factor in Public Wi-Fi services from the user point of view is the transparency of the authentication and how to join to the Wi-Fi network. In MNO environment, where most smartphones are used, the EAP-SIM (RFC 3579), EAP-AKA (RFC 3580) and EAP-AKA′ (RFC 4372) authentication methods, are highly useful. These methods use SIM or USIM card for the user authentication. On the network side it uses the Mobile Network subscriber information that resides in the HLR or HSS. It uses the same data that is used for the 2G/3G/4G network authentication but in the enhanced way.

The benefit of EAP-SIM in the smartphone is that it already includes the SIM card and the user database (HLR/HSS) already exists. So the MNO can reuse the existing infrastructure to provide a transparent authentication to the Wi-Fi network. Also the user does not need to set username & password in the client. Also Wi-Fi charging can be integrated to the cellular data and the end-user gets one bill for the data usage.

EAP-SIM works with the older SIM cards and is based on GSM authentication. EAP-AKA and EAP-AKA′ require USIM, a 3G SIM card, to work and they use 3G network AKA authentication on enhanced way to authenticate the Wi-Fi user. First authentication is using mobile identity IMSI, and in later authentications is used Temporary IMSI for the security reasons.

7.8.2.4 EAP-TLS

EAP-Transport Layer Security (EAP-TLS) is defined in IETF RFC 5216. It provides strong security and uses certificate and password for the user authentication. It is supported natively in several operation systems for example, in Windows, Windows Phone, Mac OS and iOS. Most EAP-TLS implementations require a unique client certificate, which has reduced the adoption of the EAP-TLS on the mass market. It is mostly used in the enterprise environments for laptops. Typically the certificate is delivered by the smart card that is inserted to the laptop smart card reader or by installing certificate to the laptop before the laptop delivery to the employee.

As an example enterprises are building Wi-Fi networks using 802.1X and EAP-TLS. Once the laptop has an individual certificate (smart card or preinserted), and it connects to the office Wi-Fi network, it can be automatically authenticated towards the company authentication server. Password can be given separately or using operation system single-sign-on features. As a result the user is authenticated, data in the air interface is encrypted and the user gets access to the Intranet. This is really feasible from the employee point of view.

7.8.2.5 EAP-TTLS

EAP-Tunneled Transport Layer Security (EAP-TTLS) is an EAP protocol that extends TLS as defined in the IETF RFC 5281. The difference to the EAP-TLS is that it does not require the individual client certificate to authenticate to the server. In the authentication process the server is first authenticated to the client securely (optionally client is authenticated to the server) and after this the server can establish a secure connection, tunnel, to authenticate a client, as presented in Figure 7.17. In that secure tunnel credentials (e.g. username/password) are transferred to the authentication database. Tunnel is providing protection from eavesdropping and man in the middle attack.

There are two distinct versions of EAP-TTLS: original EAP-TTLS (EAP-TTLS v0) described in RFC 5281 and EAP-TTLSv1 that is available as an Internet draft.

EAP-SIM, EAP-AKA and EAP-TTLS are feasible authentication methods for the Wi-Fi Service Providers to use, since they do not require unique certificate to the device. In managed corporate environment EAP-TLS is providing a feasible, user friendly authentication method to the Wi-Fi networks.

7.8.2.6 3GPP WLAN Interworking

3GPP scope is the interworking of WLAN with other cellular technologies. As 3GPP-WLAN interworking concentrates on the interfaces between 3GPP elements and the interface between the 3GPP system and the WLAN, the internal operation of the WLAN is only considered in order to assess the impact of architecture options/requirements on the WLAN. 3GPP-WLAN interworking shall be independent of the underlying WLAN Radio Technology and it defines two interworking scenarios, WLAN Direct IP Access and WLAN 3GPP IP Access.

3GPP is working for the WLAN interworking with other cellular technologies. It concentrates on the interfaces between the 3GPP system elements and the WLAN. It does not consider WLAN internal operation as such, only if there is an impact of 3GPP architecture on the WLAN. 3GPP interworking is independent of the used WLAN Radio Technology and it defines two different scenarios for the interworking.

The simplest scenario is to use EAP-SIM/AKA/AKA′ authentication and authorizing through 3GPP system to access direct to Internet, which is called WLAN Direct IP Access. This scenario is basically in use today; authentication is done towards the HLR/HSS and there is no service level integration to the 3GPP networks.

More enhanced scenario is to integrate cellular network's services to be available through WLAN access. That is called WLAN 3GPP IP Access, which allows Wi-Fi devices to establish connectivity with External IP networks. For that purpose 3GPP has specified use of IKEv2 IPSec tunnel to the Packet Data Gateway (PDG) that resides on the edge of 3GPP Home Network. Tunnel setup is typically done by using EAP-SIM, EAP-AKA or EAP-AKA′ authentication, and at the same time the user is authenticated by the SIM/USIM based authentication mechanisms. If a user chooses to access the Internet directly using the local IP network, no service selection information is passed to the PLMN.

In all other cases, where WLAN 3GPP IP Access is desired, the service selection information shall contain the name of the W-APN (Wireless APN, APN like in the cellular networks, but for the WLAN access) to which access is requested. The 3GPP AAA Server in the Home network shall verify the users subscription to the indicated W-APN against the subscriber profile retrieved from HSS. The 3GPP AAA Server selects a W-APN based on the requested W-APN and on the user's subscription/local policy.

In Figure 7.18, the shaded area refers to WLAN 3GPP IP Access functionality and use of the PDG. Wa/Wd interfaces refer to WLAN Direct IP Access that is already used today.

7.8.3 Wi-Fi Offloading

The background today for MNO Wi-Fi usage is usually a data Offloading. Offloading in this context means that data is transferred from the more expensive technology mobile networks to the cheaper technology Wi-Fi networks.

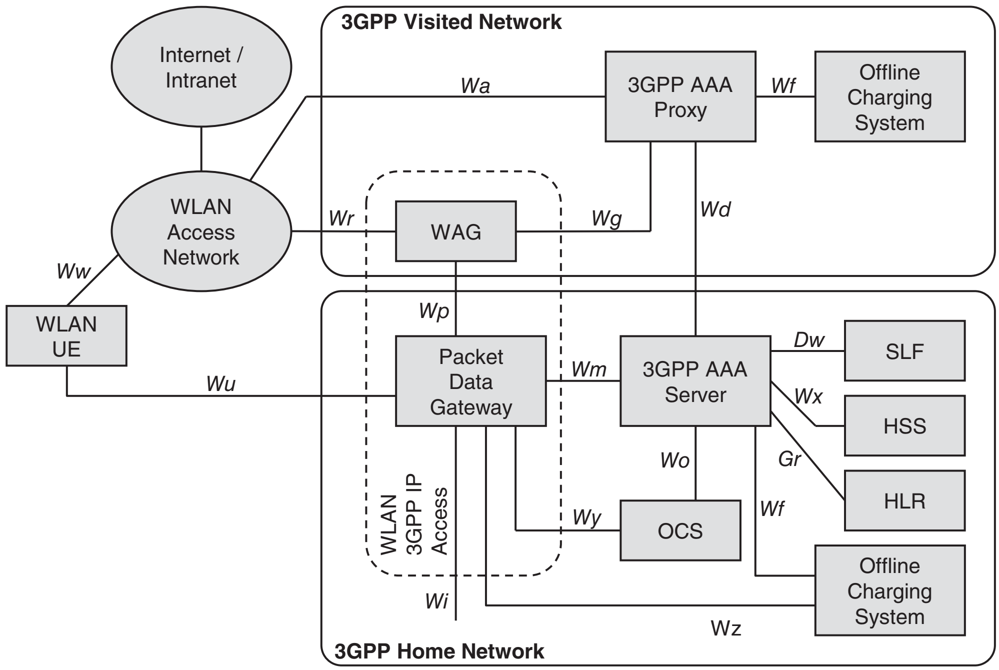

Figure 7.18 *This figure illustrates WLAN networks from the roaming and 3GPP interworking point of view.*

Offloading has become more and more important among the MNOs. Earlier Wi-Fi was seen as a high speed connection providing a better user experience to the mobile users. GPRS, EDGE and even the basic 3G did not provide a reasonable user experience. Slowly the mobile network speeds have been enhanced and at the same time amount of the data users and terminals have expanded, especially the smartphone and tablet penetration. From the MNO perspective this has led to the data explosion in the mobile networks.

Today Wi-Fi is still providing a high data connection, but that is not anymore the key element of the Wi-Fi for MNOs. Today MNOs see that Wi-Fi provides an affordable way in operator environment to offload data from more expensive mobile networks to the cheaper, unregulated Wi-Fi networks. MNOs have all the assets that can make Wi-Fi Offloading successful in operator environment, like devices with SIM and USIM, EAP-SIM and EAP-AKA authentication methods, HLR and HSS are already in place and Operator Channel is selling these devices as well 3G/LTE tablets. Another reason why Wi-Fi is so interesting is that almost all wireless consumer devices support Wi-Fi today. This expands the potential MNO customer base. If MNO has a Wi-Fi network, it can also provide Wi-Fi subscriptions to the Wi-Fi only devices such as non-3G tablets and on that way have a better service offering compared to the MNO that has just 3G/LTE.

On certain areas a good reason for the Offloading might also be a poor indoor coverage, like on urban and sparsely populated areas. Customers are used for good data speeds today, and suddenly indoor speeds are worse than expected. Wi-Fi Offloading can solve or at least help this problem.

Laptops that are really data consuming devices would really benefit from the Offloading, but the problem is often the authentication. Typically OS like Windows supports EAP-TTLS and EAP-TLS, but not EAP-SIM nor EAP-AKA. Laptops are always having Wi-Fi today but inbuilt 3G modem in laptop is not so common. External USB modems usually do not expose the interface to the SIM/USIM making the usage of the SIM card impossible for Wi-Fi authentication purposes. Recently operation systems have been enhanced to support natively EAP-SIM/AKA/AKA′. iOS supports today EAP-SIM, and Windows Phone 8 supports natively EAP-SIM and EAP-AKA. The fact is that the more there is native support, the better user experience is achieved. This can be seen for example, in case of iPad connecting to 802.1X Wi-Fi networks and automatically using EAP-SIM authentication. The users usually do not realize that they are using Wi-Fi instead of 3G, they are just happy for the connection speed.

The basic Offloading scenario happens when a user has manually configured the home Wi-Fi credentials (earlier mentioned preshared key and SSID) to the device. Most users are doing this even the device supports mobile networks. The earlier reason for this has been the price of the mobile data subscription, poor data speed in mobile network or just the small size of the data package on mobile networks. From the MNO point of view this kind of Offloading is preferable, but it happens usually only in home environment. Elsewhere the user is usually using a mobile data.

The more sophisticated approach to the offload is to provide MNO Wi-Fi network profile automatically to the device. The profile is using 802.1X standard and for the authentication EAP-SIM or EAP-AKA is used. This happens already today, for example, Apple is working together with the MNOs, and they provide a mechanism how to transfer the correct profile when users for the first time use iPhone or iPad equipped with a SIM/USIM card. Nokia has had this feature for years (already in Symbian), and the new Windows Phone 8 supports this feature as well Android 4.0.4. MNOs might also use OTA (Over the Air) to update the Wi-Fi profile of the mobile device. There are different ways how to insert profiles to devices, but the most important thing is that users do not need to know about it. The correct profile must just be there and work transparently providing a good, transparent user experience.

The most advanced approach to Offloading includes a client in a terminal as well as network support for Offloading. There are different kinds of solutions, but maybe the most advanced solutions are integrated to the cellular networks. Solution analyzes the mobile network data traffic and load, and when it realizes a high data usage in a cell, it can transfer the user from cellular data network to the Wi-Fi network. Naturally this requires awareness of the available Wi-Fi networks and possible roaming agreements if the networks are not owned by the MNO. The question is how to get the user to install separate client. If the terminal is subsidized, mobile phone can include this as a factory setting when the user buys the device. Client can also include other advanced features, such as mobile advertisement, channel to the MNO services and so on. Laptop environment might be somehow easier; MNO can provide an enhanced connection manager that combines 3G/4G and Wi-Fi management as well as Offloading features. Laptops are usually heavy data users, so the benefit of the Offloading is obvious too.

Swisscom is a forerunner in this area with its Unlimited product that was launched in 2004. The product was targeted to laptop and business users, and included support for 2G, 3G and Wi-Fi (802.1X and EAP-SIM authentication) also supporting Mobile IP. It always selected the best/fastest network automatically, and Mobile IP took care of session continuity. At the same time GSMA was trialing EAP-SIM authentication in the Seamless Access WLAN (SA-WLAN) project and proved that Wi-Fi roaming is possible using EAP-SIM instead of the legacy WEB-login and also using RADIUS protocol in the inter operator interface.

The conclusion was that RADIUS with small enhancements can support EAP authentication in the roaming environment and the automated, for user transparent authentication is an essential asset for the success of the Wi-Fi in the operator environment. SA-WLAN also created the first prototype of EAP based roaming hub functionality, Wireless Roaming Proxy (WRP). The driving forces behind this GSMA activity were TeliaSonera, Swisscom and T-Mobile. Unfortunately the time was not right; it took 5 more years to see the rocketing smartphone penetration and MNOs to realize that Wi-Fi can be a good add-on to their wireless offering with an automated EAP authentication.

An interesting take on Wi-Fi Offloading is to turn the situation on its head and deploy Wi-Fi Onloading instead. In practice this means that the mobile operator could deploy Wi-Fi only network in some cases, that is, use the Wi-Fi radio network not to load the traffic from the mobile network situated in the same area but rather to rely on Wi-Fi to provide all the operator services such as voice and SMS for the customers of the mobile network. Obviously this requires these customers having Wi-Fi enabled devices and the operator to have some sort of enhanced QoS & security mechanisms in place in the Wi-Fi environment.

7.8.4 Wi-Fi Roaming

7.8.4.1 General

Roaming in general is the possibility to use another operator's network with your home operator credentials. This possibility is widely implemented in cellular networks, where it is automatic when you travel from country to country. Cellular networks are well standardized so it has been easy to implement roaming and that has been one key factor for the success of GSM technology. On Wi-Fi there have not been such standards.

Wi-Fi was originally designed to be a local wireless solution at homes and offices using the unregulated frequency bands. When it provided better data speeds it was obvious that there were operators to provide Wi-Fi services to the customers and even the technology was unregulated. There were many kinds of active players: MNOs, Wi-Fi Service Providers and so on. The service was usually targeted to business users. When customers got used to Wi-Fi on their own network, there was a need for roaming too. In this way they could also avoid high mobile data roaming costs since Wi-Fi operators typically provided cheaper data than MNOs.

The fact that Wi-Fi had no standardized roaming infrastructure as such led to the situation where different parties created their own roaming environments. The roaming interconnection was usually based on the RADIUS protocol since that was already used in the home Wi-Fi network. In earlier days there were certain problems because RADIUS was not originally designed for the roaming purposes. Also the parameters that were sent using RADIUS were partly different based on the agreements between the roaming partners. That was a starting point to enhance RADIUS to work better in the roaming environment.

Diameter was developed to answer better to the roaming challenges. It includes more features than RADIUS and is specified to be used in LTE Enhanced Packet Core. Still the usage of Diameter in Wi-Fi environment is quite restricted, for example, it is not supported widely in Wi-Fi access points.

7.8.4.2 Activities around Roaming

GSMA has been active in Wi-Fi roaming since 2005 and created IR.61 document, including recommendations on how to implement username/password based WEB-login Wi-Fi roaming. The roaming agreements were bilateral and the parties were MNOs. Later IR.61 was enhanced to cover EAP-SIM and EAP-AKA based roaming, too.

WBA created an own roaming alliance. That includes MNOs and WISPs as well as other players around the Wi-Fi. The alliance introduced WRIX (Wireless Roaming Intermediary eXchange), a roaming hub. It provides a modularized set of standard service specifications to facilitate commercial Wi-Fi roaming between operators including WRIX-i (Interconnect), WRIX-d (Data Clearing) and WRIX-f (Financial Settlement) functionalities. Connections are based on the VPNs over Internet. WBA is using RADIUS protocol for roaming connections.

One of the first to react on enterprises demands on Wi-Fi roaming was iPass. It is a company that buys Wi-Fi capacity as wholesale from Wi-Fi operators and sells it to iPass customers. Typically their customers are international enterprises. From the user point of view it is not a question of roaming, they just use iPass service. The usage of service requires a client who has the intelligence to find the correct Wi-Fi networks and also it takes care of authentication of different Wi-Fi networks. Also data control of user is possible. Originally this service was targeted only to laptops, but it is also available for smartphones today. iPass also uses RADIUS protocol and VPN over Internet for the interconnection.

Roaming always requires interconnection. The original GSMA approach was to use IPX (IP Packet Exchange) for the WLAN AAA traffic (authentication, authorization and accounting) transmission as they use IPX for 2G/3G/4G data transmission. The IPX is a global, private, IP based network which supports end-to-end quality of service and the principle of cascading interconnect payments and it consists of a number of IPX carriers. The usage of IPX has been minor so far for Wi-Fi purposes, MNOs have usually used bilateral VPNs. MNOs are also agreeing data clearing and financial settlement bilaterally.

WBA and iPass are using VPN over Internet for the interconnection. WBA provides WRIX-i and in the best case the WBA operator member has only one connection to WRIX hub providing all required interconnection services. WBA favors a nonbilateral approach, so when Wi-Fi operator joins WBA, its customers can roam in the Wi-Fi networks of other WBA operator members. So, one agreement makes all WBA members' Wi-Fi networks available for your customers. This kind of approach is needed to make Wi-Fi roaming successful.

7.8.4.3 Other Players

Roaming is not just Wi-Fi operators, it is much more. iPass is a good example for the innovative approach to provide a good enterprise service. Buy capacity wholesale and sell it to the correct target group, enterprises. Roaming ecosystem also includes carriers and hub providers, like MACH and Aicent, who provide IPX services for the MNOs, but also hub services for the Wi-Fi service providers. Roaming brokers provide charging services for the operators. Sometimes access point owners just sell possibility to use the radio capacity, providing for the Wi-Fi operator a possibility to add an operator SSID to the access point owner's AP.

Roaming is usually understood to happen in another country, but it can also be domestic roaming, when you use other Wi-Fi network with your credentials in your home market. This does not happen very often; reason for this is the competition on the home market. If the roles and the business models of the Wi-Fi operators are different, domestic roaming may happen.

In certain dense areas free Wi-Fi is provided, for example, by the city administration or any other party. Transportation companies are sometimes providing Wi-Fi in the buses. Railroad companies have built Wi-Fi to the trains. However the experience in trains might often be poor, because of many users and restrictions on backhaul transmission.

It is clear that when Wi-Fi roaming becomes more common, there is a need for the enhanced interconnection architecture. Direct roaming connections are too complicated to manage in the long run. Complexity of connections rises potentially when there are more roaming partners. One connection point, like WRIX-i, is in that case really valuable, greatly simplifying network topology.

7.8.4.4 HotSpot 2.0

The nature of Wi-Fi was originally "wireless technology on unregulated frequencies, available for all." The natural outcome from this was that it was never designed to have such advanced features as the mobile networks have, like roaming, network discovery, billing and so on. Roaming was developed because there were Wi-Fi operators and they needed this feature to support their business model to serve their customers better. IEEE never standardized roaming architecture, it was just developed using the existing protocols (like RADIUS), based on a best practices approach.

There has always been a dilemma for Wi-Fi roaming; How do I know what is the correct SSID for which I have a roaming agreement? How do I insert a correct SSID to my terminal and what kind of authentication is used? When using a legacy WEB-login you must know the roaming partner and then you manually insert the username/password to authenticate yourself. You must also select your home Wi-Fi operator domain, hopefully from the dropdown box. Quite complicated!

For this purpose WFA has developed HotSpot 2.0 (HS2.0) specification. HotSpot 2.0 specification provides an automatic network discovery for Wi-Fi networks, a functionality that is presented and familiar from mobile networks. HS2.0 is based on IEEE 802.11u standard. This functionality automates correct SSID discovery and simplifies authentication, thus increasing Wi-Fi usage and data Offloading. It also uses EAP methods for authentication and WPA2 for encryption of the data.

The functionality is based on the AP advertising/broadcasting roaming partners NAI Realm list for which the AP/Service Provider has a roaming agreement. The functionality is as for mobile networks. When an HS2.0 capable device tries to get connected to a Wi-Fi network, it reads the AP advertised NAI Realm list and looks for the Home Wi-Fi SP NAI Realm. If it finds its Home SP NAI realm, the Mobile device matches the NAI Realm list to its Home SP NAI Realm and starts the authentication process with the correct EAP-method.

HS2.0 also requires changes to the device. The main functionalities are to scan different SSIDs advertised NAI Realms, match those to the device including NAI Realm and then to begin the authentication. HS2.0 also requires WPA2-Enterprise encryption to the AP and to the device instead of TKIP or WEP. In fact HS2.0 does not allow use of WPA or WEP.

Even HS2.0 supports all kinds of NAI Realms; it is recommended that MNOs use 3GPP defined NAI Realm for Wi-Fi usage (FQDN: wlan.mnc.mcc.3gppnetwork.org), since generation of this NAI Realm can be automatically created from the (U)SIM card in the mobile device. This helps the usage a lot, since there is no need for user interaction to complete this task.

HS2.0 specification includes support for EAP-SIM and EAP-AKA, mentioned earlier in this chapter, as well as EAP-TTLS and EAP-TLS. Specification mandates a Hotspot 2.0 SP (MNO) having SIM/USIM infrastructure to support SIM/USIM credentials and their associated EAP methods and shall support at least one of the following: username/password or certificate credentials and their associated EAP method.

The most important thing with HS2.0 is that it brings Wi-Fi to the same level as mobile networks are concerning security and ease of use. This will greatly increase usage of Wi-Fi.

7.8.4.5 ANDSF

Access network discovery and selection function (ANDSF) is part of LTE evolved packet core (EPC) of the system architecture evolution (SAE) as defined in 3GPP. The focus of ANDSF is to provide assistance to user equipment (UE) in order to discover non-3GPP access networks, for example, Wi-Fi and WiMAX that are suitable for data communications at the location in question in addition to 3GPP access networks. Furthermore, ANDSF is designed to provide the UE with rules policing the connection to these networks.

Operators can list the preferred networks and provide automatically respective policies via ANDSF. ANDSF thus offers the possibility for carriers to enable Wi-Fi hotspots with secure connectivity and seamless experience in locations where roaming between cellular and Wi-Fi networks is controlled by the operator.

There is a need for ANDSF due to the rapid increasing of mobile subscriptions and respective demand for mobile broadband data transfer. Mobile data utilization as well as the number of users can be expected to grow considerably in the forthcoming years. The combination of ANDSF and Hotspot 2.0 is an efficient enabler for enhanced and fluent user experience across Wi-Fi and cellular networks. In order to maintain an adequate quality of service level, Hotspot 2.0 provides the first step solution for roaming. Table 7.7 compares HotSpot 2.0 with ANDSF [9].

After the launching of Release 1, the next logical step would be network-directed roaming based on the Hotspot 2.0 Release 2 combined with the ANDSF enhancements according to 3GPP Rel 12. Hotspot 2.0 is a Wi-Fi discovery and management mechanism suitable for mobility between Wi-Fi hotspots, and it is claimed to be a complete, carrier-enterprise-user solution. The challenge of HotSpot 2.0 is, though, that does not provide guidance for users about identified hotspots, nor does it distribute traffic intelligently between Wi-Fi and cellular networks. The solution would be that the Hotspot 2.0 can be an enabler but further integration

Table 7.7 Comparison of HotSpot 2.0 and ANDSF [9].

	ANDSF	Hotspot 2.0 (Passpoint)
Standardization	3GPP	IEEE 802.11 and Wi-Fi Alliance
Information available	Via cellular and Wi-Fi. UE can reach ANDSF server via any access	Via Wi-Fi
Prioritization between 3GPP and Wi-Fi	Yes, information when and where to select 3GPP or Wi-Fi	Only Wi-Fi selection
Prioritization of Wi-Fi networks	Yes, different Wi-Fi networks can be prioritized	Own Wi-Fi can be prioritized but other Wi-Fi networks have the same priority
UE location information	Yes	No
Roaming support	Yes	Yes
Type of access network	No	Yes: public, private, with guest access, chargeable public, free public, etc.
Venue	No	Yes, group (residential, business), type, name etc.
Wi-Fi access point performance information	No	Yes. UE can avoid congested access points
Network authentication type	No	Yes

Source: Data by courtesy of Dr. Harri Holma.

is needed for cellular-related data transfer schemes. Companies can manage the user access to their Wi-Fi networks by applying ANDSF, which provides a broader-ranging policy scheme compared to Hotspot 2.0.

7.8.4.6 Relationship with LTE

Expanding the network capacity after certain limit is expensive. New technologies like Long Term Evolution (LTE) help the existing situation. Although LTE provides higher speeds to consumers, it is predicted that this capacity is not sufficient in future, when LTE devices come more popular. Traditionally Wi-Fi and cellular networks have been seen as competitive solutions, but it's clear the latest development especially in the area of Wi-Fi Offloading have changed the situation considerably and nowadays it's widely understood that technologies can well complement each other.

One possible scenario is to use base stations that also include Wi-Fi radio beside cellular radios to address these ever-increasing capacity demands. Wi-Fi could use the same transmission than 3G/LTE and could be connected to the 3GPP Enhanced Packet Core (EPC). Since Wi-Fi is unregulated MNO could basically increase "free of charge" capacity. Another way of Wi-Fi & LTE coexistence is using Wi-Fi access as a wireless backhaul link for the cellular traffic between the base station and core network. Naturally this should be considered only in special cases where the normal wired backhaul options are not available, but anyway could be used at least as a temporary solution.

This is part of the so-called Heterogeneous Networks, HetNets concept. 3GPP defines HetNets to be cellular deployments where small cells with lower transmission power are positioned within the coverage area of a macro cell, thereby enhancing both capacity and coverage especially in more challenging situations such as indoors which can often suffer from poor cellular coverage. HetNets can use the same or a different RAN technology to the carrier's macro cell deployment, and can even encompass Wi-Fi hotspots that dynamically offload capacity from the carrier's cellular network thus creating in essence a seamless extension of the mobile network. Here, as always when talking about Offloading, the seamless handover from technology to another is really essential. Basic IP level handovers should work relatively well across different IP access networks without any specific additional network assistance but customers using more delicate applications and services such as voice and video calls will likely see/hear a break when the handover takes place. This break might be short enough not to cause real interference but depending on the scenario it might even cause the session to end, for example due to IP level timer expiring. This is an important area where a lot of work is done, for example in IETF and 3GPP related to topics such as network-based flow mobility.

An operator selecting to do the tight integration of Wi-Fi access into the mobile core system means that Wi-Fi is seen more or less as the same kind of network as LTE by core nodes such as GGSN/PGW, once some kind of Access Gateway is deployed between the Wi-Fi cells and the core network. This allows the operator to use functions such as charging, DPI, policy control and QoS handling both for Wi-Fi and LTE network using the same core architecture such as PCC and IMS. This model provides network controlled handovers and full session continuity also in the case of device moving between Wi-Fi and LTE coverage, requiring the device to support functionality such as ANDSF which would allow the operator to install and modify operator policies in the devices. Providing operator services such as Voice over LTE coupled with Voice over Wi-Fi would be rather straightforward as they would use the same model of SIP + RTP traffic without any protocol interworking or transcoding required. Including also 3G voice would be easy there in case of using Voice over HSPA which again uses exactly the same voice architecture.

It's now rather widely seen that a seamless combination of LTE (big and small cells), Wi-Fi and wired wideband IP networks will form the practical basis for the "5G" service in the near future, simply because they all complement each other quite nicely. With this kind of combination you get massive bandwidth at home or office via solutions such as Ethernet of FTTH, great bandwidth when in a Wi-Fi hotspot, while still enjoying good bandwidth and great coverage when moving around. All this would happen automatically with the device hopping between the available access networks according to policies set by the operator and/or user.

7.9 Inter-Operator Networks

7.9.1 Introduction

A special case of general data network is the network between service providers / operators, such as GRX which has been used to carry all IP based 2G/3G roaming traffic between GSM mobile operators since the year 2000. The inter-operator network can be an extension of "normal" data network by reusing for example Internet but it can also be a specific network created just for this purpose, either logically or physically separated from other data networks [10].

In the inter-operator environment data networks are being utilized for both roaming and interconnection purposes, that is, traffic needs to find its way from the visited operator to the home operator or there's a need for exchanging traffic between two home operators.

7.9.2 Overview

The background for creating a specific inter-operator network instead of simply reusing already existing data networks is closely related to the reason why there are other closed networks such as Intranets in place: people want to avoid others from seeing certain traffic, in order to avoid any potential malicious attacks and also to keep things simple by ensuring all the parties involved in this closed network play by the same rules. This is rather difficult to achieve in a completely open environment where you have a potentially huge number of different parties connect to the same data network. Inter-operator networks typically carry both signaling and media traffic, for example in the case of voice roaming using LTE/SAE the traffic consists of Diameter, SIP, RTP and RTCP protocols.

A number of services can be run over these networks, perhaps the most typical example where the inter-operator network is being used is the voice interconnection. Typically the circuit switched TDM connections are used today for voice interconnection. This is fine for a "traditional" CS voice but in case of, for example, two VoLTE operators it would mean that the originating operator converts IP-based voice to CS-based voice, which is used over the TDM interconnection interface, and then the terminating operator converts the voice back to IP-based voice so that the end-user using VoLTE can understand it. Naturally this PS/CS conversion, occurring twice, just adds delay, potentially degrades voice quality and, generally speaking, adds unnecessary costs to the voice interconnection due to both operators having to use their CS/PS conversion nodes (such as MGW/MGCF). In a nutshell it can be said that exchanging traffic between IP-based core systems using a TDM-based interconnection network is not really reasonable either from the technical or commercial point of view.

7.9.3 Different Solutions

A technically and commercially preferable approach would probably be to go for the IP based interconnection – that is, both signaling and media would be run over a data network instead of using the traditional circuit switched solution. From the deployment perspective there are a number of options for this IP interconnection solution:

1. Leased line
2. Connection point
3. Internet
4. GRX/IPX.

Within a limited area (such as single country) a dedicated leased line, whether a physical or logical connection, is quite often used between the operators. The downside of this approach is that it is not a valid option for international connections because arranging bilateral connections to tens or even hundreds of interconnection partners simply does not work.

The connection point is a slightly evolved version of (1), This basically requires a single point where all the interconnected parties bring their own leased lines. This helps to avoid the main problematic issue of solution (1), which is arranging and operating multiple separate connections, as all the interconnected operators would have just a single line to the connection point regardless of the number of interconnection partners connected there. This kind of approach is used today, for example, by some VoIP providers to arrange the peering of traffic within a "club" of providers. However, this still does not provide a feasible option to produce the global connectivity required for an operator.

The Internet, obviously, is a wonderful thing but, when it comes to the carrier-grade connectivity that is typically seen as core issue of an operator-offered service, it lacks certain important features such as predictable latency and availability. From a security point of view, the Internet might not be the first choice of a person responsible for the security aspects of an operator service such as voice. Another aspect is the desire of many operators to use the multilateral model instead of bilaterally arranging connections to each interconnection partner, which requires a lot of manual work. In practice there is a need for some service provider(s) to take care of "brokering" the technical and commercial requirements between the operators so that an operator can have only a single connection to this "broker," which then forwards the traffic to other destinations accordingly.

For the reasons listed above, the GSM Association developed GRX (GPRS Roaming eXchange), which has been commercially used since the year 2000 for arranging the IP connections needed in PS roaming.

Basically GRX is a dedicated IP network completely separated from Internet, open only for the members of GSMA. GRX consists of around 20 GRX carriers connected to form an "IP cloud," which is necessary for the global connectivity required by hundreds of 2G/3G mobile operators around the world that are using GRX to provide, for example, routing of traffic between visited and home operator in 3G data roaming.

The concept of GRX has been further developed to include full QoS support, hubbing model (the multilateral connectivity option) and inclusion of non-GSMA operators, mainly FNO (Fixed Network Operators). This new concept is called IPX (IP eXchange) and it is a potential candidate for the interconnection needs of the LTE/SAE operator, in addition to supporting the roaming scenarios. GRX and IPX are described in a number of GSMA documents, the main paper being IR.34. IPX is a managed private IP backbone controlled by the commercial Service Level Agreements (SLA) that define the level of service such as throughput, jitter, availability, and Mean Time Between Failures offered by the IPX Provider to the operator. This means that deviations from the level of service formally defined in the SLA are subject to possible penalties. IR.34 defines a set of QoS parameters – for example the availability of IPX to an operator that is connected via a single connection is 99.7%. Upgrading this connection to a dual connection increases availability to 99.9%. The average monthly packet loss ratio of IPX when using the traffic class AF1 is defined as less than 0.1%. Another example of QoS values defined in IR.34 is the round-trip time – for example, the delay value for conversational streaming traffic class between Northern Europe and Southern Europe over the IPX network is listed at 75 ms.

It is entirely possible to use a combination of these alternatives – for example an operator could choose to handle national interconnection via leased lines while deploying IPX for all the international connections. This kind of mix does place an additional burden on the design and operation of an operator's internal and external IP core network; for example, outgoing traffic towards Operator C needs to be routed differently from the traffic going towards Operator D if those are reached via different inter-operator IP networks. Therefore it would probably make sense to try to minimize the number of different solutions used simultaneously, if possible.

It is important to notice that, if the intention is to deploy new IMS-based services such as RCS (using Presence, Video Share, Image Share, IM, etc.) one anyway needs to deploy an IP-based interconnection in order to support these new services. Running for example, Presence traffic over TDM is not a feasible option.

Note that it is also possible to run CS voice over an IP interconnection, using, for example, a 3GPP Release 4 core architecture of MSC-S/MGW "soft-switch" nodes. Signaling between MSC-S nodes would use SIP-I while the media runs via RTP/RTCP. Both control and user planes are natively IP based. This kind of architecture is widely deployed by the operators around the world – for example Nokia Siemens Networks has provided their MSC-S product to more than 260 customers. This would, for example, allow the voice interconnection traffic to run over IP-based NNI, even though the UNI would still be traditional CS based. This kind of general TDM replacement by IP-based networks is an ongoing trend in the operator community at the moment, even for pure 2G/3G operators, simply because the IP-based solution is often considerably cheaper than comparable TDM solution.

Finally it is important to notice that in multioperator environments other operators have an impact on the solution deployed individual operators. The reason is simply that there is no point in selecting the Internet as the interconnection mechanism, for example, if one's major interconnection partners are going for IPX.

References

1. Introduction to IPv6. http://ipv6.com (last accessed 26 July 2014).
2. IEEE 802.2 (Logical Link Control) description. http://standards.ieee.org/about/get/802/802.2.html (last accessed 26 July 2014).

3. Penttinen, J. (2006) *Tietoliikennetekniikka (Telecommunications Technology)*. WSOY, Finland.
4. IEEE 802.3 (Ethernet). http://www.ieee802.org/3/ (last accessed 26 July 2014).
5. IEEE 802 working groups. http://www.ieee802.org/ (last accessed 26 July 2014).
6. IEEE 802.5 (Token Ring). http://www.ieee802.org/5/www8025org/ (last accessed 26 July 2014).
7. IEEE 802.15.7 (Visible Light Communications) http://www.ieee802.org/15/pub/TG7.html (last accessed 26 July 2014).
8. Active IEEE working groups. http://www.ieee802.org/18/ (last accessed 26 July 2014).
9. Interview with Dr. Harri Holma. January 2013.
10. Penttinen, J. (2011) *The LTE/SAE Deployment Handbook*, John Wiley & Sons, Ltd.

8

Telecommunications Network Services and Applications

Jyrki T. J. Penttinen

8.1 Introduction

The advances in telecommunications systems have facilitated numerous solutions that enrich the communications between users. Telecommunications solutions have provided ever enhanced ways also for the communications between systems (M2M, Machine to Machine).

As the actual access and transmission technologies advance, they provide a platform for ever developing applications.

8.2 Voice

Voice call continues being the most important service in the telecommunications networks. Voice call is one of the teleservices of telecommunications networks.

ITU-T defines "telephony service" in such a way that it provides users with the ability for real-time two-way speech conversation via the network [1].

The recommended overall support of teleservices by, for example, ISDN has the following standards: telephony (I.241.1), teletex (I.241.2), telefax (I.241.3), mixed mode (I.241.4), videotex (I.241.5), and telex (I.241.6). During the lifetime of ISDN, it can thus be noted that telephony is practically the only one from the list that has been used on a larger scale. It is worth noting, though, that ISDN can offer telephony as a basic telecommunication service, teleservice or bearer service.

According to Ref. [2], bearer services are fully described by prose definitions and descriptions, by attributes and by dynamic descriptions, which altogether define the service characteristics as they apply at a given reference point where the customer accesses the service. In practice, the bearer service provides means to transfer the contents within and between the network as shown in Figure 8.1.

The Telecommunications Handbook: Engineering Guidelines for Fixed, Mobile and Satellite Systems, First Edition.
Edited by Jyrki T. J. Penttinen.

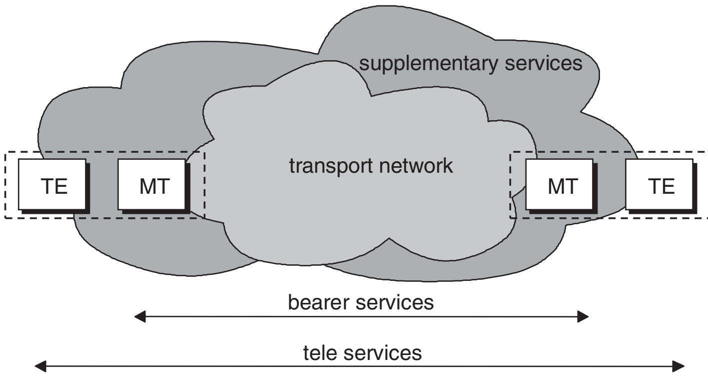

Figure 8.1 *The principle of tele services and bearer services.*

In the analog and ISDN type of telecommunications networks, voice calls have been traditionally delivered in circuit switched format. Along with the popularity of packet data services both in the fixed and mobile environment also the voice service is heavily converting to packet switched. VoIP (Voice over IP) is one of the results of this evolution. Along with voice delivery via packet switched networks, the communication also has much more possibilities with enriched contents. One of the solutions for enhanced communications is RCS, Rich Communications Suite that combines possibilities like instant messaging, video conferencing, and so on.

8.3 Messaging

The importance of instant messaging and awareness of the presence of the users have increased heavily since the social networking got popular.

Instant messaging (IM) is a communication method over packet data networks offering fast delivery of text messages between users. If push-mode is used, IM provides possibility for real-time online chat. In addition to point-to-point messaging, IM can also be used for multicast communications. Advanced variants of IM provide voice and video calling, video chat and accompanying hyperlinks. Figure 8.2 presents the basic principle of IM architecture.

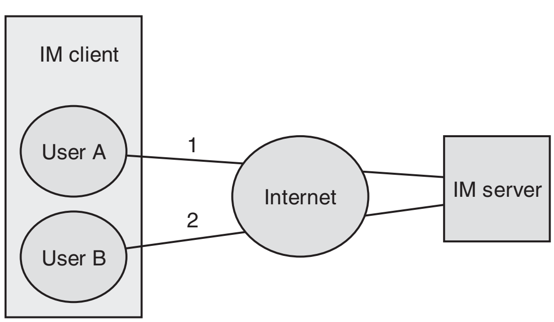

Figure 8.2 *The principle of IM system architecture. In this generic example, the message is first sent to IM server (1). Then, IM server sends the message to the receiving party (2).*

These applications are typically based on a common server that manages user data, the presence information as well as location information. This facilitates the routing of the calls.

IM can be based on the functionality of the fixed network such as via the application servers of IMS (IP Multimedia Subsystem), via the services of mobile communications networks like VoLTE (Voice over LTE) and RCS (Rich Communications Suite) or by the upper protocol layer applications (e.g., via Skype and other IM-capable SW). For the latter, the IM services generally provide the respective client, which can be a separately installed SW or client that is based on the Internet browser. Typically these solutions only work between the same provider's SW, that is, there is only limited or not at all interoperability between users that are not utilizing the same client. For the most common IM services, third party client software applications are available that are able to interconnect the services.

Typical instant messaging applications have functions like file transfer, support of contact lists, and support of simultaneous chat. In a limited environment, for example, between users that are utilizing the same provider's SW, this is feasible, but amongst larger groups or within and between companies, this may be too limited. For office use, the IM service is typically integrated with other enterprise applications. These belong to the category of enterprise applications, or enterprise application integration (EAI).

Telecommunications markets have had several solutions for unifying IM. Some examples are SIP (Session Initiation Protocol) and SIP for Instant Messaging and Presence Leveraging Extensions (SIMPLE), Instant Messaging and Presence Protocol (IMPP), and Instant Messaging and Presence Service of OMA (Open Mobile Alliance). Nevertheless, the wider unification of major IM providers (AOL, Yahoo! and Microsoft) has not been successful so each one still typically uses proprietary protocols. Figure 8.3 presents the overall development of IM systems.

It is possible to unify the systems by combining separate protocols within the IM client application, or within server application.

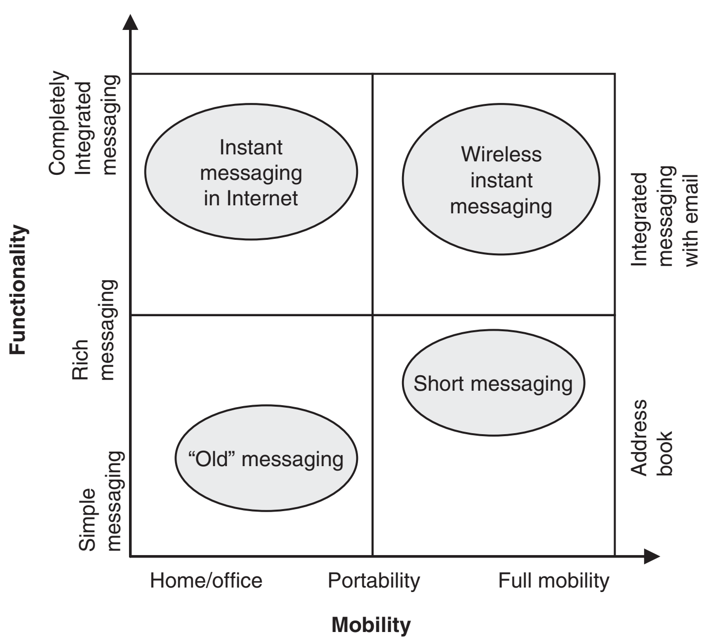

Figure 8.3 *The general steps of IM development.*

In addition to fixed IM services, there is also mobile instant messaging (MIM). It provides wireless access for instant messaging services via mobile phones, smartphones and other cellular devices capable of data transfer.

IM applications can be divided into two categories: Enterprise Instant Messaging (EIM) and Consumer Instant Messaging (CIM) [3]. Enterprise solutions typically use an internal IM server. The CIM is a suitable option for smaller scale companies as it is more economical, and does not heavily require new hardware or SW for the server.

8.4 Audio and Video

8.4.1 Streaming

8.4.1.1 Base for Streaming

Streaming media, like audio or video and a combination of these, is a type of multimedia that is transferred constantly between the sender and receiver. In a typical case, a service provider delivers the contents to the end-user either via point-to-point or point-to-multipoint mechanisms. The streaming refers thus to the process of delivering media. In practice, when the streaming is initiated, it can be played back rather soon even if the transmission of the whole content has not been ended yet. One example of the streaming systems is DVB, Digital Video Broadcasting, which is an umbrella for various methods to deliver video broadcasting. Another typical modern streaming solution is Internet television for the delivery of media for the end-users.

The term “streaming media” refers to media other than video and audio such as live closed captioning, stock ticker, and real-time text. This type of media is considered as streaming text. The term “streaming” was first used at the beginning of 1990s to distinguish video on demand on IP networks. At that time, the delivery of video was typically referred to as store and forward of video, which can be misleading [4].

For live streaming over the Internet, essential elements are video camera and microphone for capturing the varying visual and audio contents, encoder for digitalizing the contents, a media publisher, and a transmission network for the delivery and distribution.

8.4.1.2 Protocols for Streaming

For the streaming, there are various protocols available and under further development. Each one of the protocols have an optimal functioning for certain environments and content types, so at the moment there is no universal protocol available for delivering ideally all the content types.

Datagram protocols, including the typically utilized User Datagram Protocol (UDP), deliver the media stream in a format of fragmented, small packets. The benefit of this method is simplicity. Nevertheless, datagram delivery lacks of mechanism for assuring of content delivery. The detection of lost packets is thus left to higher protocol layers, and basically the application layer is the one that takes care of retransmissions upon the need, based on error detection and correction techniques.

More advanced methods like Real-time Streaming Protocol (RTSP), Real-time Transport Protocol (RTP) and Real-time Transport Control Protocol (RTCP) have been optimized for streaming of media over packet data networks. The RTSP functions over various transport protocols, whereas RTP and RTCP function on top of UDP.

An alternative solution suitable for streaming in web environment with varying contents is adaptive bitrate streaming. As an example, functionality of HTTP adaptive bitrate streaming is based on HTTP progressive download with very small files that are comparable with data packets as are used in the streaming of RTSP and RTP.

On the other hand, reliable protocols like Transmission Control Protocol (TCP) are able to guarantee delivery of media stream. Their functionality is based on utilization of timeout and retry of data. Compared to the systems that lack guaranteed content delivery, the drawback of reliable protocols is increased complexity. For cases where the data is inevitably lost and retransmission is needed, buffering of data lowers the detectable effects for the user's point of view. Logically, the buffering delay is acceptable for systems that do not require fast response time. Video on demand is one of the suitable services that this type of protocol can handle, but for the real-time two-way video conferencing the delay may be unacceptable. It is commonly understood that delays greater than 200 ms are already interpreted as annoying.

The idea of unicast protocol is to send an individual copy of the media stream to single recipients from the server. This is a suitable method for video on demand types of services over Internet, but the scalability of the method is limited in case there is a higher demand for a certain contents, such as a live football match that is observed by a multitude of individuals.

The idea of multicast protocol is that it broadcasts the same multimedia contents over a larger area of the network, either within a limited part or within the entire network, to a group of clients that are subscribed to receive the stream. The basic reason for the multicast protocols is the possibility to reduce the network load between server and clients. As the stream in the multicast environment is based on real-time delivery without possibility to acknowledge the reception of the contents, there may be errors in the reception that prevent the receiving of the complete contents. The quality of reception can be enhanced by adopting forward error correction methods, combined with other supporting methods like buffering in the player of the stream.

IP Multicast sends single media streams to a group of recipients in packet data network. One example of multicast protocol is Internet Group Management Protocol, which is suitable to manage the delivery of multicast streams to LAN groups. It should be noted that IP multicast requires that inter-LAN routers and firewalls must be configured to allow the delivery of packets directed to multicast groups. The task is easier if the same organization is managing the contents and has control of the network. This results in the possibility of using routing protocols like Protocol Independent Multicast, that are capable of delivering data streaming to multiple LAN segments.

Furthermore, peer-to-peer (P2P) protocols are also suitable for streaming multimedia contents over packet data networks. They are based on management of prerecorded streams that can be delivered in a controlled way between computers.

There are endless ideas for broadcast delivery that can be enriched compared with typical TV and radio transmissions. Computer networks and application allow means for including interactivity to content in such a way that the basic content is delivered via broadcast network, and, for example, by selecting a related hyperlink, the connection can be directed via P2P network that delivers more specific details or additional value services to individual users.

8.4.1.3 Streaming in Practice

Let's look at an example of streaming in the home environment via WLAN. In this example, a WLAN network is set up covering the house. A separate room for media server is done by setting up a PC that contains WLAN adapter, audio contents in MPEG3 format and media player for the streaming of the audio files. The server is connected to the WLAN network. In this example, Windows is the platform and Windows Media Player is used.

For the actual room where the audio is consumed, equipment for receiving the stream is set up in such a way that the media player allows shared utilization of contents.

The basic principle is equally simple in a commercial environment as in the home environment, although the setup in the transmitting end may consist of more complicated systems.

8.5 Health Care

According to Ref. [5], Telecommunications will enable capacity expansion in healthcare. In fact, along with the evolved data networks, both for fixed and mobile environment, current and future data rates and low latency values function as a base for highly useful content delivery and added value services for health care purposes.

The benefit of modern data networks combined with advanced applications is that the healthcare resources can be utilized much more efficiently because of the possibility of consulting patients and using information sources remotely in a global level. The benefit can be seen at every sociological level and culture. For modern information societies, diagnostics can be done faster and more reliably, and for very remote areas, local resources can be advised remotely for helping with fast conclusions in the case of urgent health care when no physical travel of the patient or specialists is possible.

Also the longer lasting follow-up is possible to execute more efficiently and reliably, thanks to remote connections and surveillance systems that automatically transfers vital signals to the centralized systems, and upon need, triggers alarms if, for example, the insulin level of the patient is dropping to too low a level.

The key for advanced e-healthcare is the convergence of telecommunications and health care systems. The resulting integrated systems can be used to increase the well-being of the population. As the collected data may increase exponentially, it is of importance to increase the capacity for storing the data, and making more advanced methods for analysis and postprocessing of the data. Along with increased data rates and quality of the networks, it is also possible to deliver higher quality descriptions of the patients, like high-definition X-ray photos for further analysis for the specialized entities. It is also vital that cross-checking of information can be done in an efficient and automated way, yet respecting the individual's rights to privacy. The concept of cloud computing and storage is one of the most logical bases for increasing the efficiency of the provision of the complete picture of the patient's health profile, which helps, for example, in the assurance of correct treatment in case of emergency and when the patient is not able to communicate on time about potentially life-threatening conditions like allergies.

As one of the positive results in the increased possibilities to assure information flow between health care entities, interconnected health records (EHR) also increase the collaboration of hospitals, health personnel and medicine and health care equipment providers, combined with health records and forecasts of case-by-case analysis. This provision of the complete picture assures that the treatment can be optimized as the physical location of each entity with specialized expertise is no longer a limiting factor.

Table 8.1 shows some typical examples of healthcare network services.

8.6 Education

Along with more advanced telecommunication networks and solutions, the e-learning is one of the areas that benefits greatly.

Table 8.1 *Examples of healthcare network services*

Entity	Description	Technology
Network	Wired and wireless services that are base for the interconnection of providers, patients and payers via healthcare infrastructure.	Fibre optics, broadband access, Ethernet, Wireless LANs, WiMAX, Cellular networks.
Equipment	Devices for the healthcare at hospitals and home.	RFID, Imaging, Storage and disaster recovery.

E-learning refers to the use of electronic media and suitable data delivery methods for education. It can be generalized that e-learning may use as a basis all types of technological methods. Some examples of typically utilized terms are multimedia learning, technology-enhanced learning (TEL), computer-based instruction (CBI), computer-based training (CBT), computer-assisted instruction (CAI), Internet-based training (IBT), web-based training (WBT), online education, virtual education and virtual learning environments (VLE).

The common factor of all types of e-learning includes media that deliver text, audio, images, animation and streaming video. E-learning can also include various delivery channels like audio and video storages, terrestrial and satellite TV and radio, Internet and advanced multimedia broadcast services.

The benefit of e-learning is that it can be utilized as a supporting method for physical classes in a so-called blended way, as well as for the remote learning. There are thus no limits in the type of methods, whether based on individual learning or synchronized learning in groups, with or without an instructor. One of the clearest benefits of e-learning is that it provides means to apply distance learning.

E-learning can be divided into synchronous and asynchronous types. The idea of synchronous learning is that it happens in real-time, and there are more than one participant interacting with the contents. On the other hand, asynchronous learning is based on self- learning without dependency on others.

The benefit of synchronous learning is that it provides the possibility to exchange ideas and information with other participants. Online, real-time live teacher interacting with the pupil or a group of pupils is one example of synchronous learning instruction and feedback. The practical means for creating the synchronous e-learning include combined instant messaging and VoIP platforms like Skype, Internet chat rooms and online virtual classrooms.

Asynchronous learning can be based on the utilization of various technological means, which characteristically are based on non-real-time communications or consumption of information. Some examples of asynchronous learning methods are email, Internet blogs, wiki pages, discussion forums and hypertext documents. Also audio books and audio/video courses with any current format are examples of asynchronous e-learning. The benefit of asynchronous learning is that the pupil can select the most suitable times to learn, and it is possible to repeat the contents as many times as needed. The disadvantage is the slower feedback rhythm, and lack of real-time conversations.

It can be generalized that both the asynchronous and synchronous learning methods are very much based on self-motivation and self-discipline.

8.7 CSTA

Computer-Supported Telecommunications Applications (CSTA) refers to an abstraction layer that is meant for telecommunications applications. Thanks to the abstraction, it is independent of the underlying protocols. As it includes a telephone device model, CTI applications function with it in a wide range of telephone devices.

Currently, most CTI applications are built on it, and it is maintained by ECMA International.

CSTA is based on normalized Call Control model. There is also a set of call associated and physical device features. The complete set of standard is not needed in the implementation as long as profiles are provided. As an example, the Basic Telephony profile provides sufficiently functionality to make a call, answer and clear connection.

As an example of the Microsoft environment, an important element is the Telephony Application Programming Interface (TAPI), which provides computer telephony integration and enables computers with Microsoft Windows operating system to use telephone services. It should be noted that TAPI versions vary depending on the Windows version.

The idea of TAPI is that it allows applications to control telephony functions for the use of data, fax and voice calls. TAPI includes elemental functions like dialing, answering and terminating a call. Also a set of supplementary functions are available like hold, transfer, conference call and call park according to the typical telephone systems.

TAPI is used to control modems or devises under Private Branch Exchange (PBX). In addition, TAPI can provide control for voice-enabled telephony devices like voice modems.

8.8 Advanced Telecommunications Functionalities

Along with the development of the telecommunications and data networks, the more advanced applications have the possibility to enter the markets. Some potential applications can be based on interactive TV, video games and virtual reality. The high-speed data networks with higher quality as for the latency and reliability provide a base for the next level of the user experiences. Also already established applications and solutions benefit from network development making the contents delivery richer.

Some of the new media markets in telecom that have been growing considerably within recent years are commercial online services, Internet in general, interactive TV, home shopping, videogames and virtual reality.

8.8.1 Email

8.8.1.1 General Advances

Email has established itself as the most utilized information delivery method since the initial, pregraphical Internet back in the end of 1980s. The cultural change has been dramatic as the speed of contents delivery and editing can be done close to real-time compared to previous methods that were typically based on physical paper delivery via postal service or fax, or voice communications via fixed line telephone. As email provides a near-real-time way to exchange messages and also the ability to deliver editable attachments, it is close to the most optimal communications method currently utilized in the telecom markets.

Compared to the original email utilization via computers connected to LAN, the advances of telecommunications networks have provided a remarkable increase in mobility. Table 8.2 shows general statistics of email utilization in fixed and mobile environments, as of 2011 [6].

It is nowadays common practice that different entities like Internet service providers offer email accounts free of charge to end-users, for example, by displaying advertisements to the end-users. Along with the interest of governments, it is also notable that state and municipal governments along with libraries have

Table 8.2 *Typical service utilization in smartphone and fixed network environments [6]*

Service	Smartphone use (%)	Fixed network use (%)
Email	61	77
Facebook	16	9
Twitter	2	N/A
Text messages	9	3
News site	4	4
Search / portal site	2	N/A
Other	N/A	3

taken initiatives to provide terminals and kiosks to access public records and selected information resources, and it is highly likely that the development leads to more general Internet access provision with email accounts for citizens. In fact, there have been early concepts already, for example, in France prior to the graphical Internet era, as the government provided terminals for the citizens to access the information sources via Minitel, that is, a terminal utilized in 1980–2012.

It also is notable that, after the fixed data network access and smartphone mobile access to ever advanced email utilization, there is an increasing variety of other consumer devices such as TV set-top boxes, telephones and other equipment that can have email access enabled. It is also possible that pay terminals for email and Internet access may increasingly develop depending on the demand of general pay phone markets.

8.8.1.2 Functionality of Email

Typically, email is based on Simple Network Management Protocol (SNMP). It is a protocol that manages IP devices in the delivery of messages. The complete chain of SNMP includes routers, switches, servers, workstations and printers, among other equipment. SNMP is used in network management systems in order to monitor network devices. SNMP is one of the components of the IP Suite defined by the IETF (Internet Engineering Task Force).

The management data of SNMP is seen as variables of managed systems. The parameters describe system configuration. The basic functionality of SNMP is based on these variables which can be queried and set by managing applications.

Figure 8.4 presents the idea of the functionality of SNMP. There are manager and master agent elements. The agents are connected to these Master Agent and Manager elements, and the message retrieval happens via trap function.

8.8.1.3 Management Information Base (MIB)

SNMP uses an extensible principle, and available information is defined by management information base (MIB). MIB describe the structure of the management data of a device subsystem. MIB is a virtual database for the management of entities in a communications network.

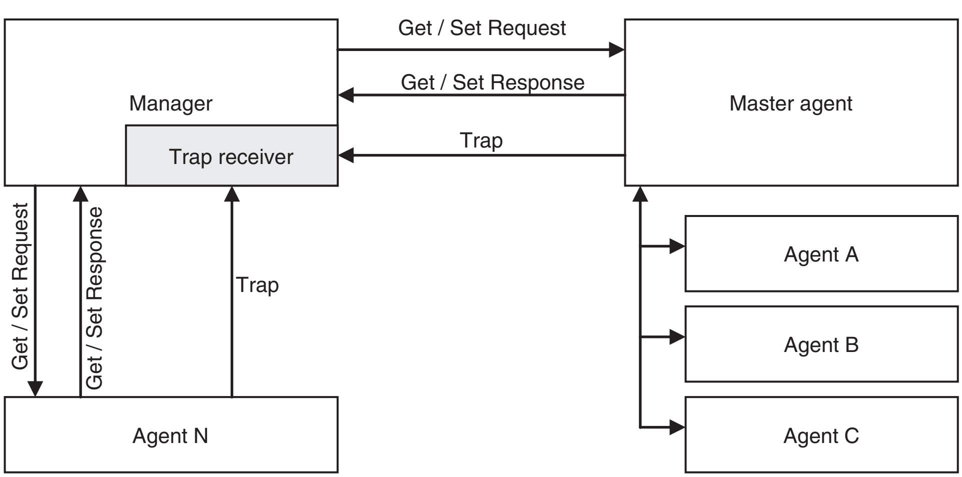

Figure 8.4 *The functionality of SNMP.*

Table 8.3 *The structure of the SNMP*

IP Header	UDP Header	Version	Community	PDU Type	Request ID	Error Status	Error Index	Variable Bindings

MIB is often associated with SNMP, although it is used generally in OSI/ISO Network management model. The objects in the MIB are defined using a subset of Abstract Syntax Notation One (ASN.1) which is called SMIv2 (Structure of Management Information Version 2) defined in RFC 2578.

SNMP operates in OSI Layer 7, that is, in application layer. For the requests, the SNMP agent receives them via UDP port no. 161. The agent response is sent back to the source port on the manager. The manager receives notifications via Traps and InformRequests on port 162.

SNMPv1 defines five protocol data units (PDU), and SNMPv3 includes additional two PDUs as shown in Table 8.3.

The SNMP PDUs are the following:

- GetRequest is a request from manager to agent to get the variable value(s).
- SetRequest is a request from manager to agent to change the variable value(s).
- GetNextRequest is a request from manager to agent to discover available variables and their values.
- GetBulkRequest is a request from manager to agent for various iterations of GetNextRequest. This request was added in SNMPv2.
- Response is an answer from agent to manager. It returns bindings of variables and acknowledgements for GetRequest, SetRequest, GetNextRequest, GetBulkRequest and InformRequest.
- Trap (or as of SNMPv2, "SNMPv2 Trap") is an asynchronous notification from agent to manager. It contains sysUpTime value, the type of trap and optional variable bindings.
- InformRequest is acknowledged asynchronous notification, as of SNMPv2. It is currently used to manager to manager and agent to manager communications.

8.8.2 Videoconferencing

During the history of telecommunications, there have been various intentions to create videoconferencing systems. The transfer of video was rather limited in the fixed analog telephony networks, and due to limited quality and data rates, videoconferencing did not take off in the analog era. Some examples of intentions were Picturephone of AT&T, and Vistacom. For Picturephone, The first test system was already built as early as 1956, with picture frame rate of 1–2 per second. By 1964 a complete experimental system was set up and tested. Equally as happened to Vistacom's initial solutions via ISDN, the results of Picturephone showed that was not yet the time for the solution. People, it turned out, didn't like Picturephone [7].

The popularity of teleconferencing between conference rooms has ever since increased as a result of faster data rates and higher quality networks, advanced compression techniques, as well as due to the availability of sufficiently efficient equipment. The benefit of teleconferencing is clear as it saves time and money, making physical traveling less needed. The global standardized protocols for videoconferencing play an important role in the increasing popularity of systems in order to lower the issues in interworking, including personal solutions between desktops. As videoconferencing technology evolves, there are also other added value services available to enrich the conferencing experience, like parallel visual presentation, dashboards, and accompanying file and multimedia transmissions.

In the lighter way, the end-users have been able to do videoconferencing already as long as the Internet has existed, by setting up video camera and microphone, with a commonly used application like Skype for point-to-point videoconferencing. Furthermore, the development of smart devices makes it possible to initiate videoconferencing regardless of location, as long as there is sufficient bandwidth available in the coverage area via cellular technologies, Wi-Fi or other wireless access technology.

As an example of instant messaging, Microsoft Lync is an instant messaging client used with Microsoft Lync Server or Lync Online, representing enterprise solution. The basic feature set includes instant messaging, Voice Over IP and video conferencing integrated to the software. Other examples of IM protocols are AIM and Yahoo!

8.8.3 Telecommuting

Telecommuting refers to working remotely, or in general, at home office by using telecommunications technologies and networks. This method is utilized by both self-employed personnel as well as regular employees. This has a positive impact on the reduction of transportation which, in turn, aligns with green values. In the regular working environment, telecommuting where applied happens typically only a smaller part of the time due to the need for physical interfacing at office. Nevertheless, when possible to apply, the benefits of telecommuting are related to lower traffic and saving of time as physical transportation can be eliminated. Many tasks can be done remotely, and especially along with development of telecommunications networks and systems, basically all the typical ways of communication are possible, including videoconferencing, file handling, conference calls. Furthermore, together with ever increasing options of broadband wireless access, concepts like Cloud Computing and Cloud Storage make the mobile office as seamless as possible as the documentation can be done regardless of physical location in such a way that the cloud can handle heavy processing, and the stored document can be synchronized nearly in real time to all the physical devices the user connects to the cloud.

8.8.4 Advanced Applications

8.8.4.1 Development

The development of smartphones and tablets has created a totally new way of consuming information, to utilize the personal and company services, and for entertainment purposes.

8.8.4.2 Application Search

The development of applications is fast. The increase of applications for different operating systems is exponential. Consumer markets have also changed as have economic aspects. Whereas the typical situation at the beginning of 2000 was still a relatively low number of programs utilized in the longer term, smartphones and tablets are now providing a huge number of programs that provide to the customers a large variety of competing platforms and software solutions. The software is also typically nowadays much more economical, and new variants of software arrive in markets faster than ever, both for “serious” use as well as for entertainment.

Along with the increase of available applications, there is one fundamental problem increasing. The challenge of current consumers is how to understand what applications are actually available, and how to find the most suitable applications.

Equally as for the Internet information search, there have been intentions to develop more efficient search engines for applications.

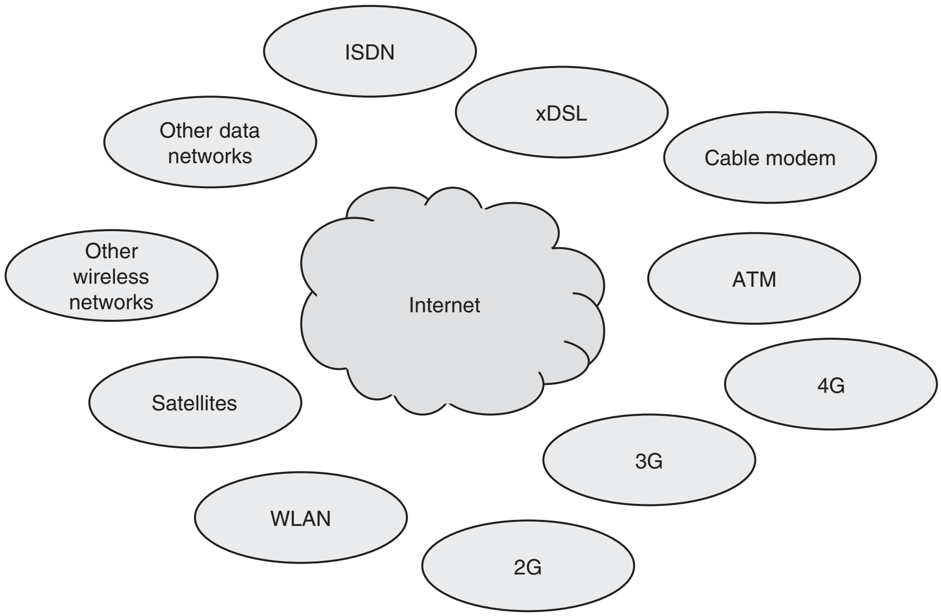

Figure 8.5 *Internet is a set of various subnetworks capable of utilizing different access technologies both via mobile as well as fixed networks. The business exchanges can be based on basically any of the known technologies although the final selection is dictated by the availability, capacity, quality and cost of the service.*

8.9 Business Exchange

In addition to private consumer use, the public IP network is also increasingly suitable for business use. The VoIP service is one of the most concrete examples of adaptation of new methods to the business environment [8].

The Internet was initially designed for military and academic use via the ideas of DARPA (Defence Advanced Research Projects Agency) and NFC (National Science Foundation), back in the 1960s and 1970s. The resulting definitions of NSFNET are still forming the basis for the modern Internet. Finally, graphical Internet browsers expanded utilization for private and enterprise use in the beginning of 1990s. Nowadays, services from email up to complex video conferencing via Internet are essential in business use.

The basic idea of the Internet provides the possibility to interconnect various network types, so by adapting data networks, the mix can contain almost all type of networks as shown in Figure 8.5.

Along with the increase in Internet service provision, an increasing number of solutions are also available for business use.

Voice and video services are thus logical solutions both for private and business use. From alternative solutions, for example, H.323 and SIP (Session Initiation Protocol) offer the possibility to deliver multimedia over data networks, and are also suitable for modern cellular networks.

8.10 Public IP Network Develops to NGN

NGN (Next Generation Network) is the developed phase of IP networks. As defined by ITU-T, the NGN refers to the packet switched network which is capable of offering various types of services, including tele services. It can take advantage of broadband and quality of service-aware transport networks, and whose higher protocol layer service functionality is totally independent from the underlying transport technologies. NGN offers in practice an unlimited handovers between networks.

The NGN networks are based on the principles of Internet like IP and MPLS (Multiprotocol Label Switching). The application layer may be based on the H.323 protocol, although the general trend is favoring SIP (Session Initiation Protocol) as the multimedia service base. Before, H.323 has been the reference protocol of the multimedia area, but its market share has been decreasing due to the fact that it did not perform well in the network address translation (NAT) or along with the firewalls.

Along with the increased popularity of VoIP services, the importance of SIP has been growing at the global level. Nevertheless, also H.323 is still globally supported as it does include enhanced functionalities. SIP protocol is, on the other hand, forming a part of IMS concept (IP Multimedia Subsystem) and VoIP both for the fixed as well as mobile networks – and the deployments of IMS are increasing currently heavily.

Related to the voice services, one of the important NGN elements is the software based exchange, that is, softswitch which controls the VoIP calls. It also facilitates protocol integration within NGN. The most important task of softswitch element is to act as interface between PSTN (Public Switched Telephone Network) and NGN. The connection is done via signaling Gateway and Media Gateway elements. There is also MGC (Media Gateway Controller) in NGN. The Call Agent refers to the equipment and systems of NGN that control the calls.

IP Multimedia Subsystem (IMS) has been standardized architecture for offering NGN with the functionalities of Internet media services in a way they have been defined by ETSI (European Telecommunications Standards Institute) and 3GPP (3rd Generation Partnership Project).

8.10.1 IMS Architecture

In IMS architecture, all services are executed via the home network. In addition, the architecture of IMS supports roaming in the same manner as it has been defined for the current visited and home networks. The difference is that the visited network provides with local access point for the SIP connection, via Proxy Call State Control Function element (P-CSCF). The local control of the connection is executed by Policy and Charging Rules Function element (PCRF) as the functionality has been defined in 3GPP specifications. In the home network, the user services are offered via single application servers (AS) which in turn have been defined as logical service functionalities.

An example of AS is TAS (Telephony Application Server) that is meant for additional services of the network, including call forwarding. A subscriber number display is given and other services are familiar from ISDN.

Other examples of AS are PoC AS (Push to Talk Application Server) and PS (Presence Server). Depending on the equipment vendor the AS can be either integrated or stand alone solution. For the integrated solution, the AS functionality is included into the same physical element of some other network functionality.

The core of the IMS home network architecture contains Call State Control Function (I-CSCF) element and Serving Call State Control Function (S-CSCF) element. The task of I-CSCF is to seek for a suitable S-CSCF element for the served IMS-registered user. S-CSCF is, on the other hand, in charge of the execution of the service by selecting suitable application servers for the session. This element also takes care of the authentication of the user and executes registering to IMS.

Furthermore, 3GPP has defined Home Subscriber Server (HSS) which contains subscription data for the use of the network, including the IMS, packet and circuit switched network user profiles. Please note, though, that in practice HSS forms part of the physical HLR element so that it can serve both IMS as well as cellular subscriber data. Also, the equipment vendors may integrate HSS functionality into the HLR, Home Location

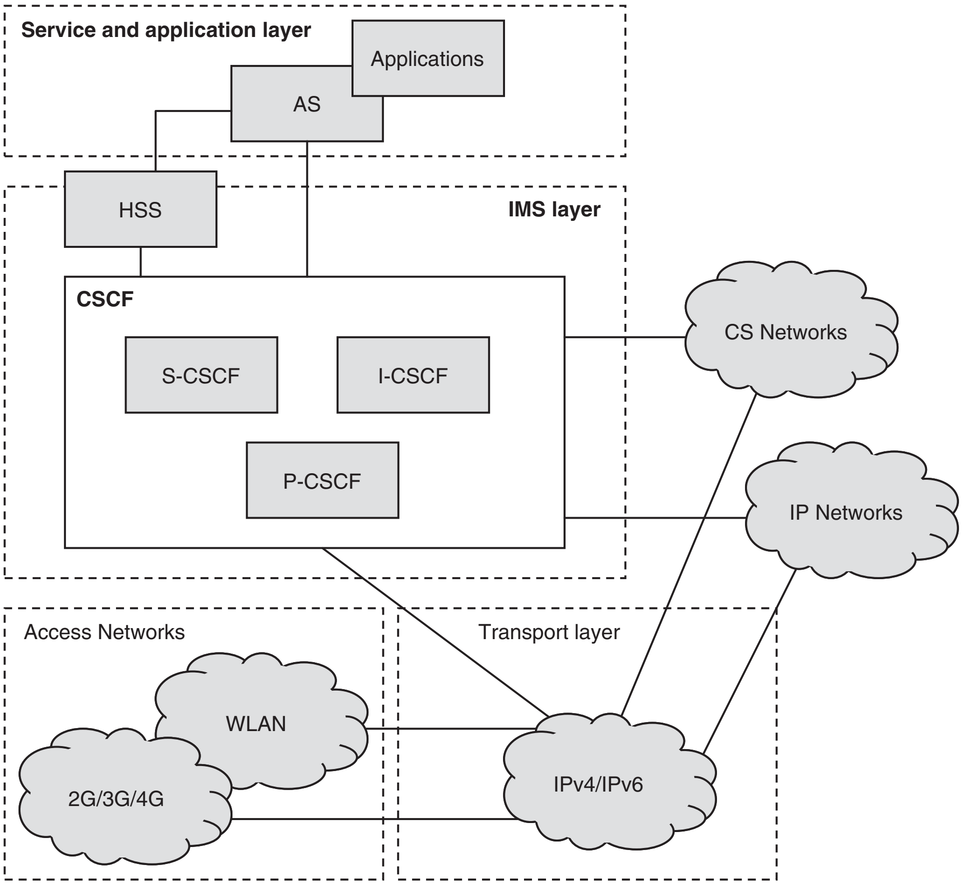

Figure 8.6 *The architecture of IMS.*

Register of cellular network instead of having it as a separate element. In case two IMS networks communicate with each other, NNI (Network-Network Interface) is utilized.

Although IMS has been defined principally for IP communications, IMS can also communicate with circuit switched networks. The latter is in fact still a basic requirement because part of the subscribers uses services of circuit switched network. For this reason, 3GPP has standardized IMS-CS inter-working function, Media Gateway Control Function, and IMS Media Gateway Function.

Figure 8.6 presents the principle of IMS architecture both for home and visited networks.

The IMS definitions include voice and video calls, as well as text message transfer over IP networks. The following sections describe the functionalities and elements of IMS.

8.10.1.1 P-CSCF

Proxy Call State Control Function (P-CSCF), that is, SIP proxy, is the first contact point both in home and visited networks for the users terminal in order to initiate the IMS calls. P-CSCF initiates signaling by selecting a suitable Interrogating Call State Control Function (I-CSCF) from the home network which can be located independently from the physical location of the user.

8.10.1.2 I-CSCF

Interrogating Call State Control Function (I-CSCF), that is SIP proxy, is typically the first contact point within the home network. In some cases also the visited network may contain the I-CSCF in order to hide the network topology behind it.

P-CSCF contacts I-CSCF during the IMS registering in order to get access to IMS services. The interface between P-CSCF and I-CSCF elements is based on the standard SIP-protocol of 3GPP (Session Initiation Protocol). The actual SIP message routing is based on the principles of Domain Name System (DNS).

8.10.1.3 S-CSCF

Serving Call State Control Function (S-CSCF) element functions as registering SIP-element for IMS subscriber in such a way that it is the contact point of IMS network also for IMS authentication (based on AKA). S-CSCF also coordinates, which of the IMS services are offered and in which order for the IMS subscriber. S-CSCF performs authentication and informs HSS about the IMS subscriber's registering status via Diameter-based Cx and Dx interfaces.

HSS must be aware of the identity of S-CSCF, for example, in case it needs to take care of routing when SIP connection is terminated requested by I-CSCF. This routing of the terminating request is similar to the functionality of HLR of 2G/3G cellular networks, as well as the functionality of gateway MSC of the traditional circuit switched cellular networks. The interface between I-CSCF-, P-CSCF- and S-CSCF elements is based on the SIP standardized by 3GPP.

8.10.1.4 Other IMS Elements

The actual services of IMS network can be utilized by end-users via AS (Application Server).

Emergency Call State Control Function (E-CSCF) is functionality required for emergency calls via IMS network, either via home or visited network. The E-CSCF functionality is initiated by P-CSCF.

Home Subscriber Server (HSS) of IMS is used as the principal data base for the profiles of IMS subscribers. HSS includes the authentication information as well as data needed for user services so that each service can be taken into use when user initiates the call. In case network has more than one HSS elements, the Subscriber Locator Function (SLF) has the information about the correct HSS element where the user data of question is stored.

Other elements are Media Gateway Control Function (MGCF) and IMS-Media Gateway (IMS-MGW) which are used when SIP session is routed between IMS subscriber and circuit switched points. In this case, MGCF takes care of the mapping of the signaling between the networks. It also controls the subscriber resources in communications between networks. Media Resource Function Controller (MRFC) and Media Resource Function Processor (MRFP) provide supporting functions which are, for example, audio notifications and alarms. They also take care of DTMF (Dual Tone Myulti Frequency) signaling.

8.10.2 SIP

The signaling method of IMS network, Session Initiated Protocol (SIP), was already developed in 1995. After initial testing, the first, SIP method was defined by IETF MMUSIC working group (Multi-Party Multimedia Session Control). As a result, the actual SIP specifications were published by IETF in 1997.

SIP is, in fact, a set of protocols designed for the advanced voice call and multimedia transfer via Internet. In addition to SIP, there has also been active development of alternative solutions for the Internet that may offer higher quality and lower user expenses.

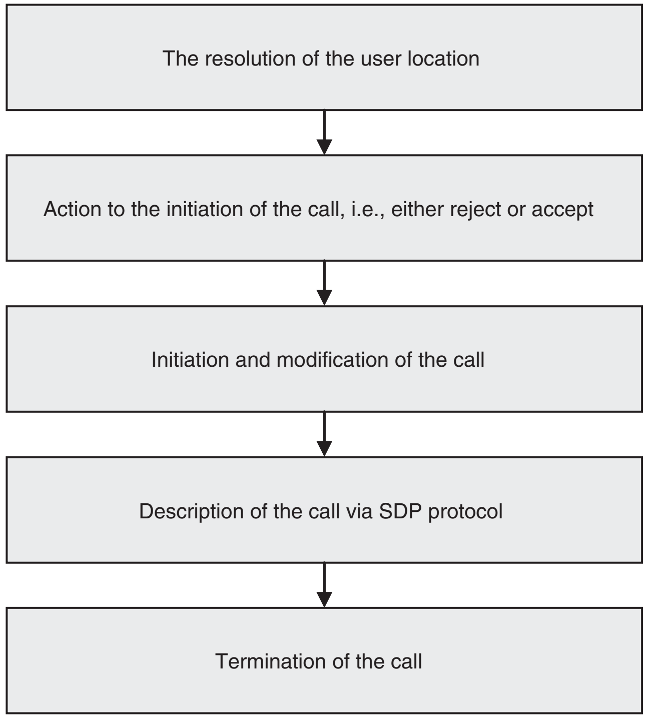

Figure 8.7 *SIP functionalities.*

8.10.2.1 SIP Functionality

SIP is a signaling method which has various tasks. The most important functionalities are the initiation, maintaining and termination of calls between two users who are connected to the Internet. SIP is capable of handling tasks also in the case of users moving from one access point to another which makes SIP especially suitable for mobile communications. Figure 8.7 presents the most important functionalities of SIP.

It should be noted, though, that SIP is specifically a signaling protocol which means that it is not aware of the details of communications in the user plane. For this, another method is applied apart from the SIP, that is, SDP (Session Description Protocol).

As for enterprises and housing communities, the IMS maintained by telecommunications operator provides logical means for VoIP service provision. This is because IMS supports services like enterprise exchange via both fixed and mobile networks in such a way that seamless handovers are possible during the connection.

8.11 Voice Service Access Points

The business exchange systems are based increasingly on VoIP (Voice over IP). The benefit of the packet switched systems in the scalability of the service capacity in a flexible manner. The packet transmission is bursty in nature, and users share the common resources unlike in previous circuit switched systems which had fixed reservation of channels. The latter ones thus needed to reserve at the beginning of the connection, and could be liberated only when the actual data call was terminated regardless of transmitted or received data. The charging of circuit switched data was thus tight to the time of the resource reservation whereas the packet data connections can be charged in a realistic way based on actual transferred data.

8.11.1 VoIP as a Solution for Fixed Communications Networks

When defined in the broadest way, VoIP refers to communications protocols and to related techniques which provide the transfer of voice and multimedia connections over IP based networks. VoIP can thus be interpreted as a synonym for Internet calls, or IP calls. Another term used to describe VoIP technique is Voice over Broadband (VoBB).

Voice call represents only a part of the data services over Internet. In practice, all the known services that are familiar from the circuit switched telephony networks can be utilized via IP packets over Internet, including, for example, fax service.

Even modern techniques and platforms are considerably different from previous circuit switched principles, there are common tasks. These include the initiation, maintaining and termination of the calls via suitable signaling techniques. In the case of voice service, the audio information needs to be first digitized and encoded, meaning that the contents is compressed and protected via common principles that both A- and B-subscriber can utilize per connection. In the case of IP networks, the bit stream is transferred over the router networks to the receiving party in such a way that the receiving equipment processes the arriving packets, assuring that the packets are harvested in correct order and with acceptable bit error rate, in order to be converted back to the analog audio format.

In practice, the IP call can be interpreted as a solution that transfers the voice via the IP protocols. Instead, VoIP represents a subset of IP call which refers to the IP calls in the public IP networks.

The original VoIP solutions were based on digital public telephony networks. After the first phase, more advanced solutions started to appear on the markets, but still within relatively closed IP networks. Nevertheless, these were the first steps towards major changes in the business models of telephony networks. One of these early stage solutions has been Skype which provided direct, free-of-charge phone calls between users. At this stage, it was not yet possible to utilize Skype calls between other telecommunications networks.

The next step was the developed phase of VoIP services which started to change significantly the models and architectures of old networks. These new solutions provided the possibility to utilize IP voice service between any Internet access points regardless of circumstances, including customers utilizing data services of mobile communication networks.

The common idea of VoIP functionalities is to include the initial signaling to the session, as well as the creation of the actual session, the maintaining of the connection, and controlled termination of the call based on jointly utilized protocols. In addition, VoIP requires one or more audio codecs in such a way that both sender and receiver may select the best suitable variant for communication per session.

For VoIP solution, there are various codecs developed and adopted. Each one of the codecs has been optimized to a certain environment, depending on the bandwidth and error levels of the connections. Although in general, the new codecs are more developed than previous ones, and can optimize better the quality as a function of available bandwidth, the earlier variants also need to be taken into account in the set of supported codecs due to wider utilization level. Part of the codecs thus support still old PCM technique based on both A and u law, that is, G.711 and G.722 definitions of ITU-T.

8.11.2 Residential Areas

For the transfer of VoIP at the residential buildings, the local area networks are still typically dedicated to single households or users. In practice, the networks are typically of small scale, and are based on wireless LAN solutions that are connected to the infrastructure of the Internet Service Providers (ISP). Typical communication platforms are ADSL (Asymmetric Digital Subscriber Line) and other DSL variants, as well as cable modems. As a physical base, fiber optics are getting increasingly popular, providing increased data rates for residential buildings, whilst the importance of old-fashioned copper lines is considerably reduced.

In addition, data connectivity via cellular network infrastructure is getting more popular along with advanced data services due to lowering of expenses. In this solution, the data connectivity of residential buildings is not tied to a fixed location but services can be utilized in a dynamic way also outside the building in wide areas, provided that the offered data rates, coverage area and quality of service level are adequate, for example, for VoIP services.

8.11.3 Business Environment

For business use, within a company building, the contract between operator and business defines the technical values for, for example, maximum available capacity (number of voice call channels) and quality of service in terms of, for example, data rate and delay.

There are several solutions available in the market with common characteristics. In a typical business VoIP solution, business entities are connected together via the telecom center of the business building to the public telephony network via IP connectivity and SIP signaling protocol. For operations, a separate business subscription and corresponding telephone number are utilized with respective QoS (Quality of Service). Characteristics of the typical solution are the infrastructure agnosticism that provides connections for both fixed line telephones as well as to the mobile communications subscriptions.

In addition, the VoIP service provider may offer wider solutions which connects the physically separated local area networks of the business locations to the IP services of the operator, and at the same time offer remote connectivity for personnel.

8.12 Mobile Services

Mobile VoIP is one type of VoIP service which is suitable for traveling personnel. One natural part of the development of mobile communications, also integrated VoIP solutions are taking place more often in the business environment. The Mobile VoIP can be defined as an evolved variant of the fixed network VoIP service which is capable of managing the continuum of the voice call by utilizing seamless handovers in the radio interface.

Mobile IP can be divided into short range and wide area solutions. The short range solution is limited to a certain location, for example, within a single LAN area that contains only few radio access points. On the other hand, the wide area solutions provide the possibility of utilizing the VoIP services within the whole service area of the complete mobile communications network, including the foreign countries via roaming.

The Mobile VoIP protocol are selected in such a way that the short range solution is based on Wi-Fi whereas the wide range solutions support the protocols of mobile communications networks, like HSPA and LTE. In the latter case, VoIP is becoming an increasingly typical integrated solution for smart devices and tablets, and other portable devices that are capable of cellular network connectivity. In these cases, the VoIP solution can be either integrated as a native code for the operation system, or a separate application that can be preinstalled by the operator, or installed later by the user.

One principle is to integrate all needed protocol layers in the chipset level for optimization of IMS connectivity in the fastest way. In this case, the mobile device acts as SIP client equipment capable of initiating SIP connections and utilizing data as a bearer for voice calls via the radio interface. This solution is based on standard VoIP protocols, including SIP over any wireless broadband IP network including Wi-Fi, WiMAX, third generation HSPD/HSPD+ (High Speed Packet Data) and LTE (Long Term Evolution), as

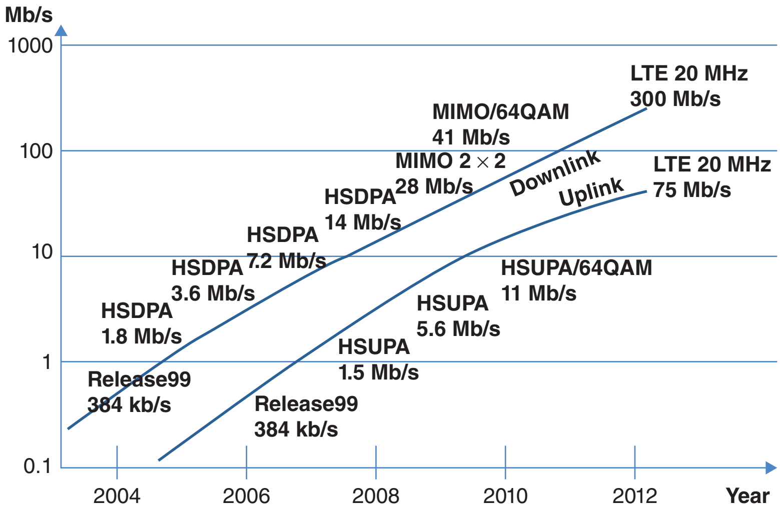

Figure 8.8 *The evolution of 3G networks towards 4G era. Basically all of the current 3G network solutions are sufficiently capable of supporting VoIP services over packet data.*

well as fourth generation LTE-Advanced and evolved version of WiMAX with the performance as defined by ITU-T. In addition, the developed data services of even GSM may be adequate for VoIP solutions even if GSM has not been designed originally for voice calls over packet data.

Figure 8.8 presents the 3G network evolution towards the fourth generation.

In addition to the previously mentioned techniques, there are also other enablers for VoIP services, such as soft-switch gateway which is capable of utilizing SIP and RTP protocols with the SS7 signaling system within mobile networks. The benefit of this specific solution is that the device supports as such original radio functionality like the HSPA interface. The solution adds SIP application server which provides the actual SIP services.

For the end-user, the local Wi-Fi access points offer a cost-effective (typically free-of-charge) way to utilize VoIP services. The drawback of these Wi-Fi hotspots are the limitations of the coverage area even in dense urban city centers, and they typically lack mobility management due to the immaturity of handover definitions to other communications networks. Nevertheless, solutions are expected in the near future via, for example, the Wi-Fi Offloading concept which provides a seamless handover between Wi-Fi hotspots and cellular networks. Even so, the hotspots are sometimes relatively loaded in the peak hour which may cause congestion or dropped VoIP calls. For this reason, the most modern, high data rate cellular networks are especially suitable for VoIP service. The data rate is in adequate level already in typical, advanced HSPA networks, and LTE with its evolution path offers the most logical base for VoIP with high-quality connections.

As the technical barriers are being removed gradually along with the advanced cellular network deployments, the remaining part for the businesses is to analyze the feasibility for mobile VoIP in terms of expense. There may be flat-rate principles used with possibly a maximum data reserve defined, which is the most adequate for business use. The other charging principle is based on charging of transferred bits, which requires in practice tools for keeping track on consumed data for avoiding surprises in the final charging. These tools include apps that keep track of data consumption. In both cases, it should be noted that the roaming cases may cause surprises in the charging, so the study of the roaming prices is highly recommended prior to trips abroad.

8.12.1 Mobile Exchange

The Mobile Exchange may be based on circuit switched voice channels, as has been the case especially for second generation mobile networks (Centrex). Along with the benefits of IP data via mobile networks, circuit switched exchange services are disappearing fast.

The VoIP solution is the modern variant and default for business usage in case of Mobile Exchange deployment.

8.12.1.1 The Suitability of Mobile Networks for VoIP Services

The base for Mobile VoIP is the cellular network with sufficient available capacity. The mobile network consists of two parts: radio network and core network. Neither one of these should create bottlenecks for the VoIP service provision. If such an event is likely, the task of the operator is to assure the extension of the capacity and coverage for offering the minimum agreed quality of service level for the customers.

Although second generation networks could be utilized for VoIP, their data rate and availability are typically limited for larger service offering. Thus, 3rd generation networks are required as a minimum for packet data based voice services.

8.12.1.2 Transport Networks

The transport networks of mobile communication networks have traditionally been based on PCM (Pulse Code Modulation) as for the circuit switched services. PCM is, on turn, based on 64 kb/s timeslots per customer for voice calls and circuit switched data connections. Along with standardization of GPRS (General Packet Radio Service), the transport networks started to take the form of packet networks.

Along with the development, the circuit switched transport networks can also be adapted to the IP era. One of the transition phase solution for carrying the circuit switched voice calls is Carrier Ethernet which is a solution that encapsulates circuit switched connections to IP packets within the cellular networks as presented in Figure 8.9.

Prior to the purchase of mobile exchange, the business area must be assessed thoroughly in order to understand the feasibility for mobile communications, that is, the current level of coverage and quality of service needs to be reviewed. This avoids any unpleasant surprises with lack of service level, in case coverage is missing in part of the business location, including lower floors indoors. In case coverage area is limited, the operator may be able to offer repeater solutions within the business location, or even build additional

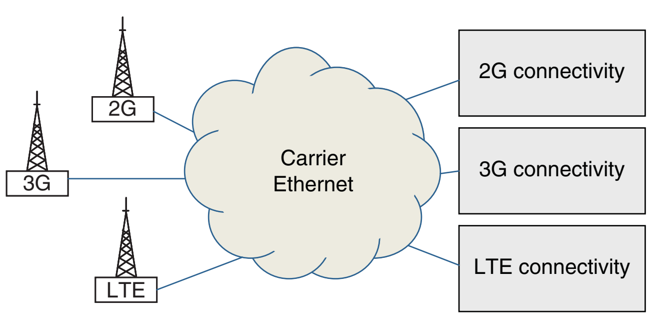

Figure 8.9 *Carrier Ethernet solution. This solution connects the radio network via IP packets.*

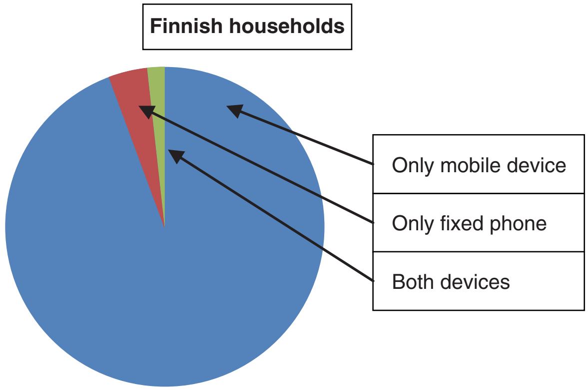

Figure 8.10 Snapshot of telecommunication equipment utilization of households in Finland, 2012 [11]. Data published by Statistics Finland.

coverage and capacity by providing additional base stations. Operator coordination is essential in order to assure that repeaters do not cause interferences in the rest of the radio network.

8.12.1.3 Needs for Mobile Exchange

The utilization of mobile communications networks has exceeded that of fixed networks for years ago in many market areas. Figure 8.10 presents a snapshot of a typical situation of the share of mobile equipment vs. fixed telecommunications equipment in developed market [9].

Even a snapshot, Figure 8.10 indicates the general trend of telecommunication equipment utilization in consumer and business segments. As a summary, the personnel of businesses are having more and more mobile handheld equipment in use, for both company and personal communications. At the same time, the role of fixed telephone equipment is reducing drastically.

Thus, the deployment of Mobile Exchange in the business environment is a relevant option. The most important reasons for the deployment are:

- Reduction of total communication costs by combining technologies.
- Support of remote work.
- Enhanced ways to reach personnel.
- The optimization of the maintenance of different systems and equipment by outsourcing the services.

Due to the clear benefits of the mobility and cost-effectiveness of the outsourced mobile services, businesses are moving towards Mobile Exchange solutions.

8.12.1.4 Characteristics and Deployment Options

Compared to old business solutions for telecommunications, the Mobile Exchange provides a set of novelty solutions both for mobile device users as well as for switch personnel of the company. There are also new solutions available for customer service which eases troubleshooting and shortens customer care time. Nevertheless, typically also all the older solutions are still available in Mobile Exchange.

The general trend is that the personnel of companies have increasingly mobile device for the business communications. In addition to high-quality voice calls, the data rates are increasing which is beneficial for utilizing business solutions via remote connections.

The functionalities that enrich voice communications in the Mobile Exchange environment are, for example, the following:

- Conference call between 3 or more parties.
- The delivery of A-subscriber's number to the B-subscriber through the exchange whilst the personnel connect the call.
- The further forwarding of already forwarded call by the exchange personnel, to another mobile subscriber or to a subscriber of a fixed network.
- When calling from the exchange area, the conversion of the A-subscriber's number to a company number instead of the original A-subscriber number. A practical case for this modified A-subscriber number could be the direct fixed telephone number in order to hide the mobile subscription.
- Management of work hour information via mobile device. Examples of the modified info are dates and hours of absence.
- The creation of different answering profiles. An example can be a different message in voice mail during vacations with a functional option to either leave a message (in order not to disturb the B-subscriber during the vacation), or to connect directly to the B-subscriber.

Figure 8.11 presents the architecture of the integrated Mobile Exchange as a part of the company's communications solution. In addition to the functional set of services of the Exchange, the device itself may also be

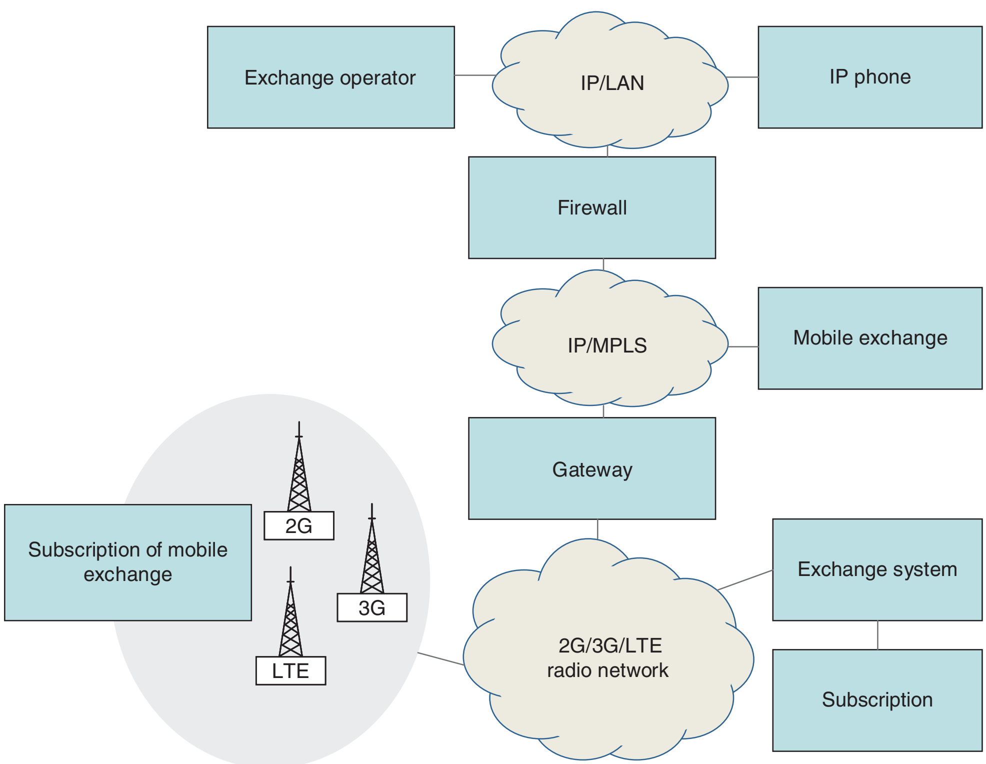

Figure 8.11 *The Mobile Exchange can be integrated as a logical part of the business communication service solution.*

capable of utilizing different business-related apps and integrated services. Some typical examples are email, calendar and VPN connections to the company's IT services in such a way that the information and contents are synchronized in all the devices of the user when connected.

An important enabler for wide and flexible synchronization in a seamless way is cloud network. An example of the benefits of the cloud service is that when once updated via one device, the calendar info, contents of documents with texts and figures, and so on are synchronized practically in real time in wireless and fixed company equipment as well as in home devises. It is thus straightforward to continue with the same document at a different location. As a concrete example of the integration, Windows Phone 8 and Windows 8 devices are compatible and thus able to utilize the same Office suite. As a result, the editing of the same documents in a seamless way is possible.

Although the role of facsimile is lowering drastically, it is still possible to utilize it via mobile device, separate app, or network service.

Also the Exchange personnel are provided with new services which have not been typically available in previous solutions. These services include the following:

- It is possible to observe the status information of the C-subscriber without need to wait for the answer. This makes it faster to realize if the number is, for example, occupied or the subscription is cancelled, so that the personnel would not even try to connect the call. As a result, the customer service time is faster.
- Queuing functionality to the occupied number.
- The returning of the already connected call back to the Exchange personnel in case the B-subscriber won't answer, or the line is occupied.
- Barring service to call forwarding, preventing the call to be transferred to, for example, voice mail.
- Communication between Exchange personnel and B-subscriber via short message service, for example, for leaving messages electronically from the A-subscriber.
- Direct forwarding of the call to the voice mail of the B-subscriber.
- Observation of the status information of the workers at the Exchange.
- The management of the priority order of the Exchange personnel based on the calling parties.

One of the benefits of the Mobile Exchange is that it provides the possibility to connect calls regardless of the physical location of the Exchange personnel which eases the outsourcing of the human resources. Furthermore, via the mobile communication networks, it is possible to take advantage of the Intelligent Network (IN) services.

The Mobile Exchange can be deployed as an additional functionality of the already existing fixed exchange. Another option is to obtain the service as a turnkey solution from the operator as a network service which removes the need for physical equipment for the company.

The first option is heavier because it requires transport connections via the own exchange equipment of the company, and the equipment requires regular maintenance. The latter solution is based on complete outsourcing of the service which provides the full set of Exchange functionalities yet without need to maintain own equipment.

In the planning phase of the Mobile Exchange deployment it is essential to investigate the locations where the company workers are typically utilizing the services. Even if the mobile network coverage areas are widening, radio reception may be low especially in remote areas as well as inside the company buildings. Especially buildings with relatively tight metallic constructions or supporting material, including windows with metallic particles, may be problematic as they attenuate the radio signals due to the Faraday cage effect. In case the company requires Mobile Exchange services in such locations, one option is to order a separate repeater solution to the indoors of the company in cooperation with the carrier operator. Another solution is that the operator deploys an additional base station near to the coverage-limited location in order to enhance the service area.

The Mobile Exchange solution is seamless to the company workers as the personnel experiences the services exactly as they are utilized in the fixed network cases. The difference is the enhanced answering time and more complete chain of the reach. In the company point of view, the solution is straightforward and fast to activate because operators' already existing infrastructure provides the network services as such without additional equipment. This removes the need for equipment integrations and interoperability testing as well as maintenance work completely. The operator also offers the subscription numbers and manages the functionality of the system. In the case of failure, operator guarantees the recovery based on the service level agreement (SLA). Furthermore, the network service and quality reporting can be included as one part of the complete solution for the customer.

Connectivity to the mobile network services via the own equipment of the enterprises provides the same set of services as the direct connections of the operator. The difference is that the enterprise will have to perform the equipment maintenance and error recovery in the former solution. Furthermore, the routing of connections via the own infrastructure increases the delays and points of failures. In that sense the most logical solution is the latter one, with direct connectivity to the operator infrastructure [10].

The main components of the voice service are:

- Mobile exchange and information system.
- Connectivity to Intelligent Network (IN).
- Connectivity to cellular network.
- Fixed line networks and VoIP systems.
- Terminals.

The Mobile Exchange provides the exchange personnel with a platform which can be utilized in a versatile manner for connecting the calls. The service contains the terminals for cellular and fixed networks. Fixed network calls can be established by circuit switched or packet switched VoIP connectivity.

A typical example of the functionality of mobile exchange is the status display of the subscription via the user interface of the exchange personnel. The status information of the B-subscriber can be obtained via the interface to the cellular network's visitor and home location registers. By utilizing this functionality, the exchange personnel can speed up the work as there is no need to try to establish the call with users that are not available. Also, the erroneous call establishment attempts are lowered.

The information system with respective database as presented in Figure 8.12 has an important role in the mobile exchange. The centralized information system has interfaces to various systems which offer optimal reachability of workers, as well as related location based services. The information system can be connected, for example, to the time management system and calendar. The added value of this solution is that workers are able to inform about absence via mobile device and web browser. The reachability information in this solution is thus easy and dynamic.

8.12.1.5 The Role of IN

The Intelligent Network (IN) has SCP (Service Control Point) which acts as a centralized contact point for services of the network. The calls that are using IN services are directed to SCP via SSP (Service Switching Point). Between the connection points, the calls are routed via core INAP (Intelligent Network Application Protocol). An example of this is the connectivity between SSP and SCP of fixed telephony network or cellular network. In SCP, there is an SMS (Service Management System) as well as data bases for user services. SMS is meant for management and maintenance of the data bases. Figure 8.13 presents the principle of IN.

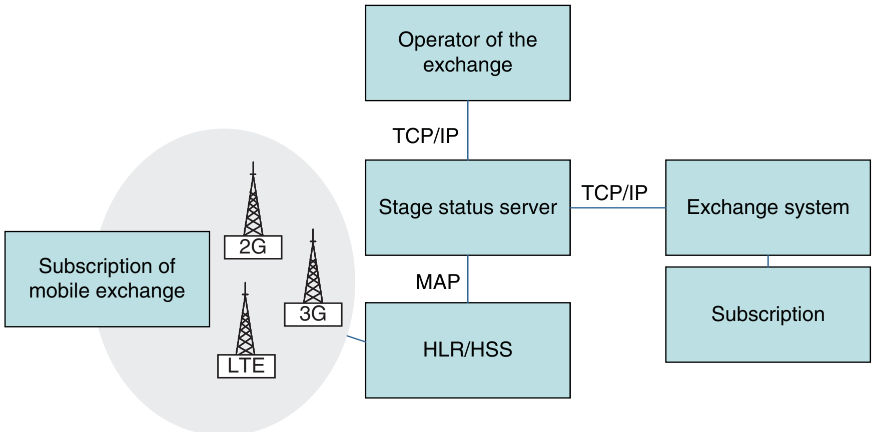

Figure 8.12 *The Mobile Exchange can be integrated via the operator systems to the status registers of the mobile exchange as well as to the enterprise exchange.*

The applications of IN are, for example, prepaid service numbers, personalized and PIN-protected services, reversed charging, splitting of charging between A- and B-subscribers, voice notifications menu as well as VPN (Virtual Private Network). The latter is a useful base, for example, for the exchange of a small scale company in such a way that the company would not need to invest in exchange equipment and transmission lines.

Signaling between IN and other networks is based on STP which is a packet switched IN element offering signaling between network elements via common channel signaling.

IN provides a logical base also for various added value services. As IN services of the operator reside in centralized points, the utilization, management and development of services is much easier compared to distributed models.

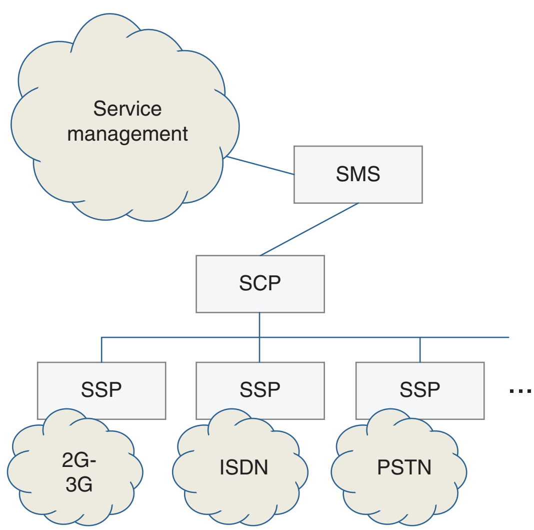

Figure 8.13 *The principle of Intelligent Network (IN). IN offers services in a centralized way via SCP (Service Control Point) and by routing the calls to SSP (Service Switching Point).*

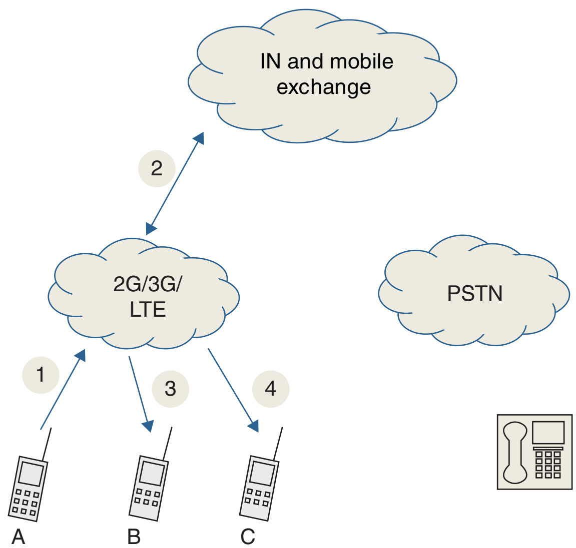

Figure 8.14 *Example of call routing from A-subscriber (A) to the subscriber of the company (B) via direct call.*

The numbering of company telephone subscriptions is typically done in such a way that flexible use of IN services is also possible. In the case of mobile exchange, IN routes the calls entering the company subscriptions. It also provides enhanced functionalities for customer service as well as reachability functions like call forwarding according to the predefined criteria. Thus, IN is an essential part of the mobile exchange of companies.

Figures 8.14 and 8.15 present two typical cases about call forwarding via mobile exchange.

In the example of Figure 8.14, the A-subscriber calls to the B-subscriber so that the call is first routed to the mobile network (1). The network recognizes that it is an IN call, which triggers a signaling to SCP and SSP (2). They answer the signaling by giving the routing information for B-subscriber (2), and the call can now be connected directly to the physical number of the B-subscriber (3).

The combination of IN and mobile exchange the direct call to the company subscriber is flexible. In case the B-subscriber line is busy or the user terminal is switched off, mobile exchange can thus forward the call to another representative according to the predefined reachability chain, either to mobile equipment (4) or to a fixed line telephone.

The example of Figure 8.15 shows how the fixed line subscriber calls to the company exchange (1). The telephony network recognizes the connection to represent an IN call and initiates thus respectively signaling with IN (2). IN, in turn, recognizes that the aimed B-subscriber belongs to users under the mobile exchange, which triggers further signaling to the mobile exchange (3) resulting in the signaling being routed to the mobile exchange via the mobile network.

Now, based on the information of the IN and mobile exchange the call attempt is routed to the exchange personnel (4) who answers to the arriving call via the mobile network, or alternatively via fixed telephone network if the company uses fixed exchange. At the same time when the call arrives to the terminal of the exchange personnel, the User Interface (UI) of the mobile exchange shows information about the arriving call. Furthermore, the exchange personnel can now also observe the status of the B-subscriber. If the latter is available according to the UI, the exchange personnel can thus connect the call (5, 6).

When the B-subscriber answers, the mobile exchange disconnects the call leg with the A-subscriber and mobile exchange (4), and the personnel is again ready for receiving further calls. At this point, when the

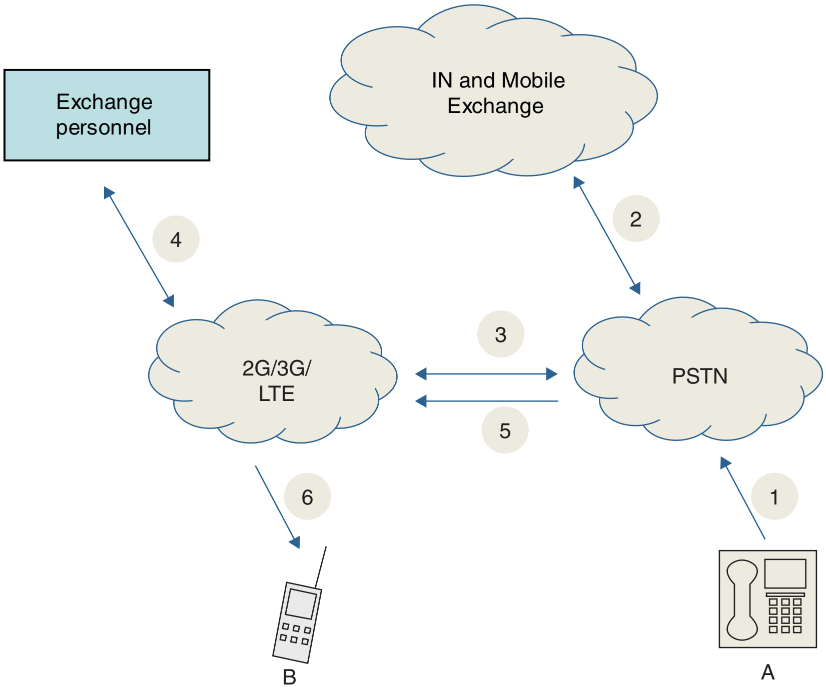

Figure 8.15 *An example of the A-subscriber calling via a fixed line telephone to the company exchange because the direct B-subscriber's number is not known.*

call is connected between A- and B-subscribers, the original A-subscriber number is now shown for the B-subscriber. As soon as the B-subscriber answers, the call is now a direct connection (1, 5, 6) so that the exchange is not any more taking part of the routing.

The examples of Figures 8.14 and 8.15 indicate that IN has an essential role in the call routing to B-subscriber. The mobile exchange, in turn, provides the needed functionalities for the call coupling. For the actual bearer for the voice calls, either fixed telephone network or mobile communications networks can be utilized upon the situation. It should be noted that as for the mobile exchange, it is irrelevant which type of network is used for the calls. It is also not important for overall functionality which type of handheld mobile devices the company has whilst they are generally accepted by regulation. Thus, the mobile exchange and fixed network exchange function seamlessly together.

Major part of the practical solutions for mobile exchange is such that they still cooperate in a parallel way with the previously purchased fixed network exchange. After the deployment of the mobile exchange, it is logical that subscriptions for employees function in the same manner as for the provided set of services and the user experience in general. The difference is that the mobile exchange provides better reachability, which in turn enhances the general quality of customer service.

In the longer run, mobile exchange may replace partially or completely fixed network exchanges in companies. For customer service, the benefit is that the physical location is no longer important as customer service activities can be performed remotely by utilizing web-based software, by default via secured HTTPS connections.

In mobile exchanges, the centralized information system of mobile communication networks is in essential role. It also eases integration of information systems to other needed interfaces, as is the case for connectivity between the information system to the mobile switching centers and its registers. For the fixed network exchange, the information system with status and calendar information can be deployed via TCP/IP protocol.

8.12.2 The CAPEX and OPEX of Mobile Exchange

The mobile exchange is a product of carrier operator and resides thus within the cellular network. This is easier for companies as there is no need for equipment installations, maintenance and fault management as those are part of the contract with operator that provides the mobile exchange service. In practice, the deployment of the mobile exchange concept is done via turnkey principles, the providing party coordinating all the practical steps for integration, testing, and getting the service on with planned service level agreement (SLA) and charging of the use. The charging may be based, for example, on the number of used lines and/or monthly fee, the number of available functionalities and minimum quality of service.

For the analysis of complete cost impact, it is recommendable to select use cases that best represent typical traffic profiles of the company. The analysis should take into account the initial costs (Capital Expenditure, CAPEX) as well as the longer term costs of operation (Operation Expenditure, OPEX). The most realistic estimate of benefit can be achieved by analyzing the longer term effect for some years ahead in such a way that costs generated by older own equipment are taken into account, including maintenance, increased probability for mean time for failures, need for SW and HW updates with respective costs of labor work and material. In the most detailed analysis the nondirect cost impacts are also taken into account, including the difference between breakdown times, and respective impact on customer satisfaction, which in turn may be affecting negatively the whole image and thus business opportunities of the company if the critical limit is exceeded. It is recommendable to look on the relevant aspects during 3 years time period which already indicates the trend of the cost behavior. Table 8.4 summarizes the main points that can be selected as a basis for the analysis.

Based on Table 8.4, the more concrete cases can now be analyzed in the following way:

- **Case 1**: The original solution which has only a fixed network exchange. The possible mobility can be offered by deploying wireless network access points of, for example, DECT system for respective handheld devices (which is now relatively uncommon) or Wi-Fi hotspots which could be used for Voice over IP devices. The wireless network would be covering the indoors by default but otherwise it is limited as for the wider coverage.
- **Case 2**: The original solution of fixed network exchange combined with mobile exchange with a certain percentage of the users out from the total number of employees. As an example, 30% can be assumed to use both fixed telephones as well as mobile phones (having both terminals), whilst the rest of the employers are relying only to the fixed phones.
- **Case 3**: Limited utilization of the fixed network exchange combined with the mobile device utilization. This model is a so-called single terminal case, that is, each employer has either fixed or mobile terminal.
- **Case 4**: Solution based on purely mobile terminals. In this model, the company only has mobile exchange.

As a general observation, the savings of the company typically increase along with approaching the single device model. It is highly probable that along with the mobile exchange contract, the total savings compared to the other cases are greater, although the estimate needs to be done for each case individually by selecting the cost items as realistically as possible, also for the near future if only possible (including the most probable scenarios for extending the business and locations, and thus the needed capacity). One of the additional benefits of the mobile exchange model is, though, that reachability of the personnel increases clearly. On another note, the total costs of the call are lower the more the internal calls are used compared to calls to external subscriptions. Furthermore, as previous fixed exchange capacity can be reduced and thus the amount of the transmission lines lowered, the costs for fixed exchange lowers respectively.

Table 8.4 *Aspects that can be taken into account in the comparison of the total costs of fixed exchange owned by company, and mobile exchange service provided by operator*

Cost item	Fixed exchange	Mobile exchange
Deployment of the service	The previous investment would probably not cause additional impacts for the original installation effort.	There is an initial one-time cost impact because of the interoperability adjustments, interface and service setup.
Monthly fee of the equipment	Own exchange of the company does not cause additional direct operational costs.	There is no separate equipment installed in the customer premises, and thus there is no direct monthly fee.
Monthly fee of the software	Typically, the SW is owned by the company. Nevertheless, there may be license fees involved, for example, due to database use.	The monthly fee charged by the operator covers the utilization of the service as well as the mobile exchange user interface. There is a monthly fee which depends on the amount of provided functionalities, number of subscriptions, and possibly on the capacity, quality of service and other items that are included into the service level agreement.
Expenses of the equipment	Own equipment of the company requires maintenance which triggers a maintenance fee. The potential failures must be taken into account by optimizing the use time before the probability of the failures increases, that is, the purchase and deployment of modernized equipment on time is important.	The monthly service fee of the operator covers typically all the expenses for normal and abnormal cases.
Recovery from the failures	As the equipment is aging, the probability for increased mean time between failures increases. As a result, there may be complete service breakdowns. Also the warranty of the equipment may be expired when the worst type of failures are increasing.	Operator takes care of the recovery from the failures according to the agreed service level.
Costs of telecommunications	Own equipment requires operational connections which have a separate cost impact.	The provider of the service takes care of all telecommunications transmission and signaling costs. Furthermore, the provider may give regular reports about the load, failures and maintenance.
Other operation costs	There are indirect costs from the use of own equipment, for example, due to the increased electricity bill, costs of the equipment room, its cooling and ventilation, and so on.	No additional operation expenditure as it is included into the monthly fee of operator.

8.12.3 Deployment of the Mobile Exchange

In the selection of the mobile exchange solutions, it is recommendable to investigate the offered services and their real need and compliance for the quality for the normal operations of the company. The main point of the mobile exchange concept is the high mobility and cost-effective communications. Thus, the more personnel is moving outside of the company premises, the more cost-effective the mobile exchange concept generally becomes.

The deployment of the mobile exchange requires detailed analysis of the current and future situation of the company. The plan is recommendable to make jointly with the company and providing operator. If the mobile exchange is meant to be an addition to the already existing fixed exchange, the SW or HW of the latter probably does not require updates as such. Nevertheless, the completely mobile exchange based solution is the most straightforward as it totally removes the need for fixed exchange maintenance. In this case, the most important tasks of the complete mobile exchange solution is to assure that the already existing customer data is collected properly and transferred into the new systems.

As an additional benefit for the mobile exchange, it does not require a separate cabling in the office premises. Nevertheless, the potential drawback is that the construction material of the office premises might lower the useful radio reception and transmission. Also, the growth of the user base is one of the aspects to be considered, which might cause potential capacity problems under the mobile exchange. Prior to the service, it is thus essential to investigate the potential coverage and capacity limited locations now and in the future. The network measurements are recommended as a basis for the analysis, especially during the peak hour which represents the most heavily loaded case for both radio interface as well as the core transmission capacity utilization. In case potential capacity or coverage problems may be expected, it is highly recommended to discuss the possibilities for installation of additional base stations on the area, or by adopting repeaters within the company premises.

As a final point, it is also of high importance to organize adequate training for personnel in order to put mobile exchange services into use as efficiently as possible.

References

1. ITU-T. Integrated services digital network (ISDN) service capabilities. Definition of teleservices. ITU-T Recommendation I.240. 1993. 5 p.
2. ITU-T. Integrated services digital network (ISDN) service capabilities. ITU-T Recommendation I.230. 8 p.
3. IM in Enterprises. http://im.about.com/od/imforbusiness/a/topbizims.htm (last viewed 27 July 2014).
4. International Research Center. http://www.researchedge.com/uss/adv_tele.html (last accessed 27 July 2014).
5. Telecom, IT and healthcare: Wireless networks, digital healthcare, and the transformation of US healthcare 2012-2017. May 2012. The Insight Research Corporation. 17 p.
6. Statistics of e-mail utilization. http://www.emailmonday.com/mobile-email-usage-statistics (last accessed 27 July 2014).
7. AT&T Picturephone. http://www.corp.att.com/attlabs/reputation/timeline/70picture.html (last accessed 27 July 2014).
8. Puhelinjärjestelmät (Telephone exchange systems). Published in Finnish by Sähkätieto, Finland, 2013.
9. Wireless Village. System Architecture Model. Version 1.1. 2001-2002 Ericsson, Motorola and Nokia.
10. Vistacom video solutions. http://ins-news.com/en_UK/100/683/1235/?ls-art0=60 (last accessed 27 July 2014).
11. The official statistics of Finland, 2012: www.stat.fi (last accessed 27 July 2014).

9

Transmission Networks

Jyrki T. J. Penttinen and Juha Kallio

9.1 Introduction

The transmission lines form the base for practically all types of networks, transporting the contents of the user to the destination in the core part of the networks. This applies to voice service as well as data services and signaling. As the data rates tend to increase and the applications and services are constantly becoming more demanding, as for data consumption, transport networks need to support very high data rates and quality of service requirements. Among the fixed data network evolution, especially the wireless data networks take major leaps towards very high data rates, and the transmission of the core should be at least as efficient so that it does not create bottlenecks.

9.2 Physical Transmission Systems

Transmission can be built up, for example, by utilizing good old copper wires, or more advanced lines such as fiber optics, radio links, or via satellite connections. For each one of these, there is an optimal environment that can be balanced between the cost of the deployment and operating expenditures, bit rates, length of transmission line, easiness to install the equipment, availability of power, presence of interferences, and other technical and nontechnical parameters. Although in general, the role of copper wires is getting less significant along with the deployment of fiber optics, to mention one example, the copper is still useful in some areas.

The transmission line capacity can be increased by utilizing the same physical channels for different users, and even for different systems. This is called multiplexing. It can be done in practice via FDM (frequency division multiplexing) and TDM (time division multiplexing).

FDM is an older method of these two, and is utilized, for example, for radio and television broadcast transmission in coaxial lines. The method is thus traditionally based on analog technology. TDM, instead, converts the signals from analog format to digital (if they are originally analog ones), by using A/D- and D/A-converters. Once digitalized, the signal can be transmitted between channeling elements via, for example, PDH or SDH techniques.

The Telecommunications Handbook: Engineering Guidelines for Fixed, Mobile and Satellite Systems, First Edition.
Edited by Jyrki T. J. Penttinen.

Transmission can be done either via cables or via wireless connections. For the latter solutions, there are various solutions for constructing radio links.

FDMA (frequency division multiplexing access) means that the same physical transmission path is divided between several users or channels in such a way they are served in different frequencies. Purely FDMA technology was applied widely in the radio interface of analog, 1st generation mobile systems, such as NMT (Nordic Mobile Telephone), AMPS, TACS and Netz-C. In those systems, each user reserved a whole frequency channel during the complete circuit switched call, until the connection war terminated.

TDMA (time division multiplexing access) means that the same transmission resource is divided between different users in such a way that each one reserves the resources only during a part of the complete time. One of the examples of this method is the radio interface of GSM and TETRA. The key for making this happen is the digital system, which provides the possibility to squeeze into bursts in such a way that the original piece is digitalized by sampling, and the resulting data block is transferred with higher data rate than the actual playback of the stream would require. As an example, the original, full-speed voice codec of GSM requires about 12.2 kb/s bit rate in order to be streamed in the same rate that it was digitalized. In addition, there are certain overhead bits required for controlling the call, so a single full-rate GSM channel requires exactly 16 kb/s time slot. If this is transferred now with a bit rate of 64 kb/s, the contents can be delivered to the receiver in a quarter of the actual time needed to stream the contents in audio format for the user. This therefore means that in the same 64 kb/s transmission stream, there can be three other calls delivered accordingly to their own destinations, and a total of four voice calls can be served via the TDMA time slot structure of the transmission. In GSM, the voice codecs have been developed considerably since the first phase of GSM, so the half-rate voice codec can serve a total of 8 users in a single 64 kb/s time slot of the core network, or up to 16 users when recent technology called OSC (Orthogonal Sub Channel) is applied. The voice codecs and OSC are described in detail in Chapter 11 of this volume.

A more developed technique compared to the FDMA and TDMA is CDMA (code division multiplexing access), which separates different users via the bit patterns dedicated for each user. As the technique spreads the bits in the radio interface over a relatively wide band, users can pick up their own contents by correlating their respective user codes, whilst the traffic of other users is interpreted merely as a background noise. When correlating the contents of a certain user by utilizing this known pattern, the own contents bits can be amplified from the signals that are even below the thermal noise level. As an example, IS-95 is a 2nd generation mobile system based on the CDMA, and UMTS/HSPA+ represents a 3rd generation CDMA system.

9.3 Coding Techniques

There are clear benefits in digital transmission techniques over old analog methods. The general characteristic of analog technology provides with signals that are represented directly by the variations of frequency, amplitude and/or phase. This means that the original contents (like the original voice of the user) are in theory transmitted exactly like it is originally captured via the electrical signals of a microphone. Nevertheless, there is certain noise level added to the original contents, and it increases along the transmission route. In addition to the noise, there is also attenuation of the original signal. The signal can be amplified by the repeaters that are installed along the transmission route, but the drawback of the purely analog equipment is that in addition to the original signal, it also amplifies the inevitable noise. There is thus a certain critical level in the analog systems when the noise level is simply too strong to be received in a decent manner. This is problematic for both audio and data transmissions.

Digital transmission utilizes voltage levels representing the zeros and ones of bits. The original contents, if it is in analog format like human speech, are first digitalized via the A/D conversion (Analog to Digital). The quality of the received contents depends thus greatly on the quality which was applied in digitalization.

This depends mainly on the sampling rate of the original analog contents, as well as on the resolution, that is, on the maximum scale of the bits.

The sampling rate is important to dimension correctly in order not to create (too much) loss of information. The sufficiently high sampling rate of the original analog signal can be reasoned by applying Nyquist theorem. According to the theorem, a band-limited analog signal can be completely reconstructed from an infinite sequence of samples on condition the sampling rate exceeds a certain amount of samples per second, which is 2 times the highest frequency occurring in the original signal.

The scale of representation of the contents in a bit format depends on the system. Generally utilized scale is based on the 8-bit format as this is relatively easy to apply in many parts of the digital systems, including the transmission itself as well as typical file systems. When the signal is represented via the changes of only bit levels, the interpretation of the signal in the receiver end is typically very close to the original bit pattern in the transmitter end, up to a certain signal-to-noise level that starts to destroy the interpretation of the signal in a quite dramatic way. This is one of the key characteristics of the digital transmission compared to the smoother analog transmission, as the digital signal functions relatively well up to the critical level, whilst analog signal gradually gets noisier.

The receiver end can thus interpret the bits within practical noise level limits, and the original contents, if they were in analog format originally, can be revealed by performing D/A conversion (Digital to Analog).

Even digital transmission experiences increased noise level during transmission, which makes voltage levels (that represent the bits) deform, and in addition, the levels attenuate similarly as in the case of analog transmission. Repeaters are thus needed in digital transmission lines. There are basically two main types of repeaters: straight and degenerating variants. The first one simply receives the bits, amplifies them, and resends them further to the next repeater, or to the final destination. The latter contains more functionalities as they actually interpret the signals in bit format (by the original numeric values), and create again the format utilized in the transmission line by amplifying the contents, "refreshing" thus the bit presentation. This way obviously provides more reliable transmission during the complete transmission line.

9.3.1 Unipolar Format

The simplest method to present the bits of the signal is the straight binary format, that is, a voltage level that represents either zero or one as a function of time. This is called a unipolar signal. This format can be either RZ (return to zero) or NRZ (nonreturn to zero).

It should be noted that this format is not utilized in commercial telecommunications transmission. Instead, the line encoding must be done in a more reliable way. Figure 9.1 shows the idea of RZ line code.

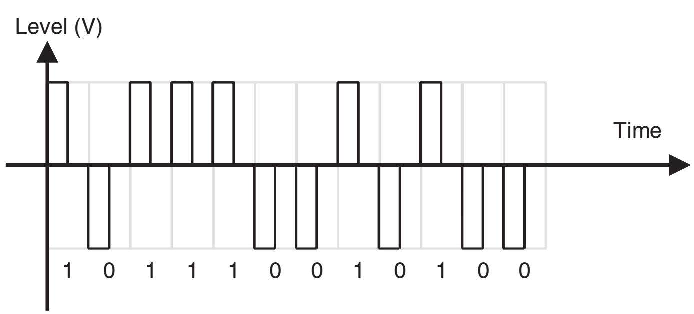

Figure 9.1 *An example of RZ line code.*

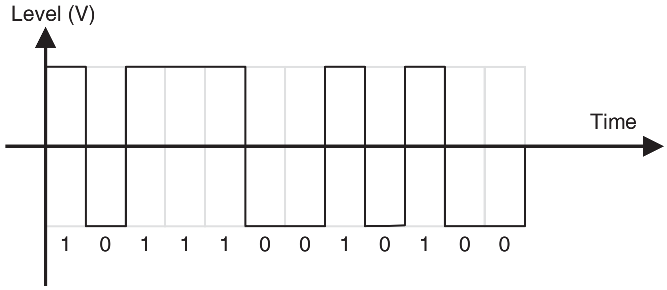

Figure 9.2 *The idea of NRZ line code.*

The RZ line code means that the signal returns to the zero-level after each pulse has occurred. This also happens in the case where various consecutive 0- or 1-bits are present in the signal, which means that the RZ signal is capable of clocking itself. This results in the freedom of the separate clock signal together with the signal. The drawback of RZ line code is the need of twice the bandwidth compared to a nonclocking signal in order to provide the same data-rate compared to the NRZ line code.

In RZ line code, the level represented as zero between the actual bits is the rest condition. That stage is typically selected in the halfway between the 1-bit and 0-bit of the actual signal. Figure 9.2 shows the idea of NRZ signal.

The NRZ line code is such a binary code that represents the higher bits (1) by one significant condition which is usually a positive voltage whereas the lower bits (0) are represented by another significant condition which is typically a negative voltage. It should be noted that there is no other neutral or rest condition in the NRZ line code, unlike in the case of RZ line code. This leads to the fact that NRZ is not inherently a self-synchronizing code, and thus additional synchronization techniques are required in order to keep the bits in order. The solution could be, for example, parallel synchronization signals. Especially in the asynchronous communications, the lack of the neutral state requires separate bit synchronization as the clock signal is not available.

In general, NRZ as such can be utilized in both synchronous and asynchronous transmission, and transitions simply occur as a function of time representing an integral clock cycle.

9.3.2 Bipolar Format

If the original RZ-type of signal is modified in such a way that the level of every other 1-bit is actually converted to a reversed level (-1), we can obtain a bipolar representation of the signal called AMI (alternate mark inversion). This means that the bit levels can be presented via voltage levels of $+1$, 0 and -1. Regardless of its benefits in increasing reliability due to clearer synchronization of 1-bits, this format also has drawbacks due to the occasional, longer periods of 0-bits.

9.3.3 Modified AMI Codes

9.3.3.1 Zero Code Suppression

The early technique for ensuring minimum density of marks has been “zero code suppression.” This line coding sets the least significant bit of the transmitted stream of each 8-bit byte to a value of 1. This has avoided the need to modify the AMI code, but on the other hand, only limited data rates to 56 000 bits per second could be achieved per single DS0 voice channel. The additional problem resulted from low minimum density of levels of 1 (12.5%), which occasionally caused increased clock slippage on the span. Due to these limitations, zero code suppression was not ideal for the ever-increasing bandwidth demand, for example, based on the ISDN basic speed of 64 kb/s. It was thus replaced by B8ZS.

9.3.3.2 HDB3

The transmission line coding can further be developed in such a way that the pair bits of the original voice channel are inverted prior to line coding. Furthermore, the accuracy of synchronization can be enhanced if the system can allow only 3 consecutive 0-bits with the same level. The possible fourth 0-bit is coded thus as a "broken bit." The existence of the possible broken bit can be interpreted via the rules of alternating the sign of the 1-bit, that is, the presence of either bit +1 or −1 reveals the existence of the broken bit. This line coding is called HDB3 (high density bipolar 3), and has been used in all levels of the European E-carrier system.

9.3.3.3 B8ZS

The North American systems utilize bipolar with eight-zero substitution (B8ZS). It is used in the North American T1 (digital signal 1) with 1.544 Mbit/s line code. The idea of B8ZS is that it replaces each string of 8 consecutive zeros with the special pattern "000VB0VB." Depending on the polarity of the preceding mark, that could be 000+−0−+ OR 000−+0+−.

9.3.3.4 B6ZS

The other North American variant of line code is bipolar with six-zero substitution (B6ZS), and is used with T2 rate of 6.312 Mbit/s. This line coding replaces 6 consecutive zeros with the pattern "0VB0VB." Depending on the polarity of the preceding mark, that could be 0+−0−+ OR 0−+0+−.

9.3.4 Delta Modulation

A simple but highly useful method for converting the original, analog presentation of the signal to digital format is delta modulation. Instead of presenting the complete voltage value of the signal, it only codes changes in the original signal levels. This means that increasing the signal level is represented by +1, and decreasing the signal is represented by −1.

The benefit of this method is that the signal can be coded in a simple yet very reliable way. Nevertheless, the simplicity is also a drawback in these situations when the signal level changes rapidly, as coding can happen after the actual signal level has already changed. This problem can be observed in practice, for example, via digital television systems when highly motional contents are presented – part of the visual contents (pixels) might get blurred, that is, the resolution of certain details may get momentarily lower.

9.4 PCM

9.4.1 Principles

The commercial telecommunications transmission networks have been constructed by utilizing PCM (pulse code modulation). PCM provides digitalization of all the contents of telecommunications networks, including voice calls, audio and video, text and multimedia.

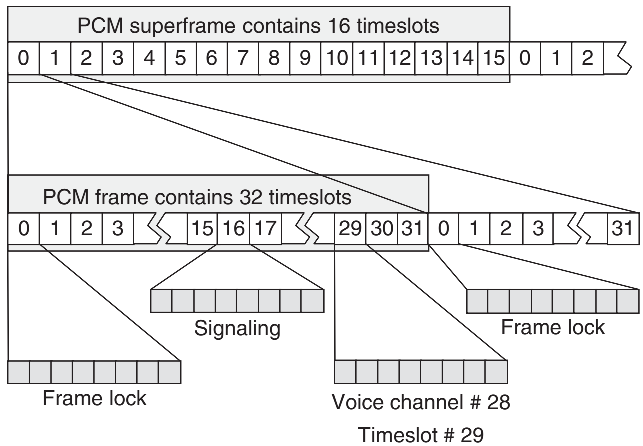

Figure 9.3 *The principle of PCM timeslots, frames and superframes. The PCM timeslots no. 0 and 16 are reserved for signaling and frame lock, and the timeslot no. 1–15 and 17–31 corresponds to the voice channel nos. 1–15 and 16–30. The superframe of PCM is repeated every 16 frame intervals.*

The physical layer of PCM is typically based on fiber optics, but also various other solutions are feasible, from coaxial cable to satellite connections. Especially in remote areas, like deserts and archipelagos, transmission between telecommunications network elements is easiest to deploy via radio links. The network layer of PCM can be done, among other possibilities, via ATM (asynchronous transfer mode).

The PCM technology is based on TDD (time division duplex). In a single PCM timeslot, it is possible to deliver 64 kb/s (European A-law weighted system). The European system is based on the single PCM line with a total transmission rate of 2 Mb/s, which is divided into 32 PCM timeslots. Two of the timeslots are reserved for signaling purposes, alarms and synchronization, leaving a total of 30 channels for end-users.

The principle of the PCM timeslots is presented in Figures 9.3 and 9.4. The figures show the European variant. In the American version, the PCM is based on μ-law weighted system, which provides with a total of 1.5 Mb/s that is divided into 24 PCM timeslots.

The duration of a single PCM timeslot is 3.9 micro seconds, and the duration of a single bit within the timeslot is of 488 ns. A single PCM frame lasts 125 micro seconds, and the duration of a superframe is of 2 ms.

The data rate that PCM timeslot provides is 8 x 8000 kb/s = 64 kb/s, and the signaling timeslot (no. 16 in the PCM timeslot structure) per single voice channel corresponds to 2 kb/s in average.

The complete data rate of a single PCM line is thus 32 x 64 kb/s = 2048 kb/s. The next levels of PCM system provide total data rates of 8448 kb/s, 34 368 kb/s, 139 264 kb/s and 565 148 kb/s. These classes can be utilized in the transmission of both for fixed telecommunications as well as for mobile systems.

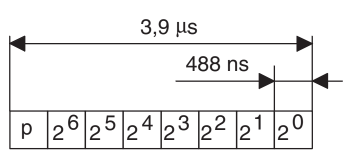

Figure 9.4 *A single PCM timeslot for the voice channel consists of 8 bits, from which the first one indicates the polarity of the signal.*

9.5 Coding Techniques

The analog signal is converted to the PCM transmission path by sampling and digitalizing the original analog presentation of the signal as shown in Figure 9.5. According to the theorem of Nyquist, the sampling rate should be at least twice as high as the utilized bandwidth in order to present the signal in reliable manner. That means that for the telephony networks, the basic audio signal of 300–3400 Hz (corresponding to 3.1 kHz band) can be presented by digitalizing it via PCM methodology.

The sampling rate of PCM is 8000 Hz, which is sufficiently high for presenting the analog voice contents with 8 bits. As the PCM has 256 quantization levels, the result is a bit stream of 64 kb/s.

The quantization of PCM is done in a nonlinear way, that is, the spacing between quantization points grows as a function of amplitude as presented in Figure 9.6. The term for this technique is "compress and expand" and is designed to function optimally in a wide dynamic range.

In North America, the compounding is done based on the μ-law weighting. The idea is exactly the same as in the European variant, but the US variant utilizes 15 segments instead of 13 of the A-law weighted version. In practice, when the A and B subscribers of the telecommunications networks (including mobile communications) are in Europe and North America, there is a need of mapping the A and μ-law versions at some point of the connection.

An inevitable side effect of the sampling is quantization distortion, which is a result of the uncertainty of the correct value of the original signal in the form of bits. Quantization distortion is not constant over the

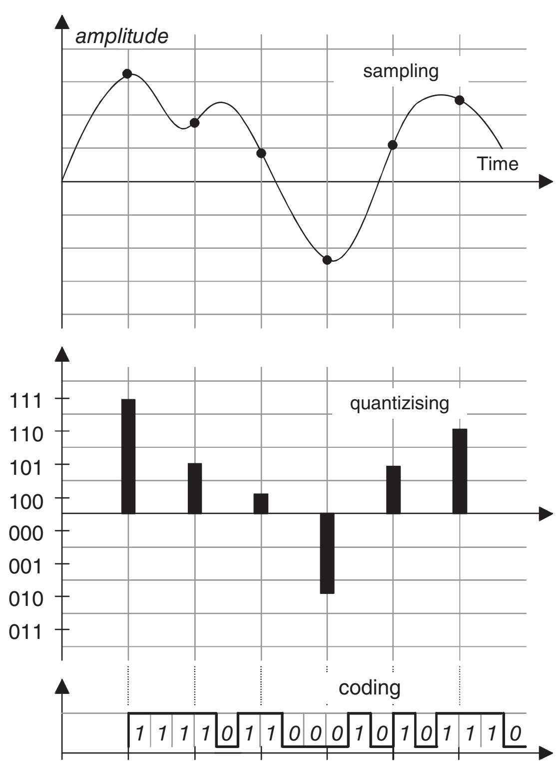

Figure 9.5 The principle of the sampling of the signal with 3 bit coding. In PCM, the same principle is applied, but by utilizing 8 bits.

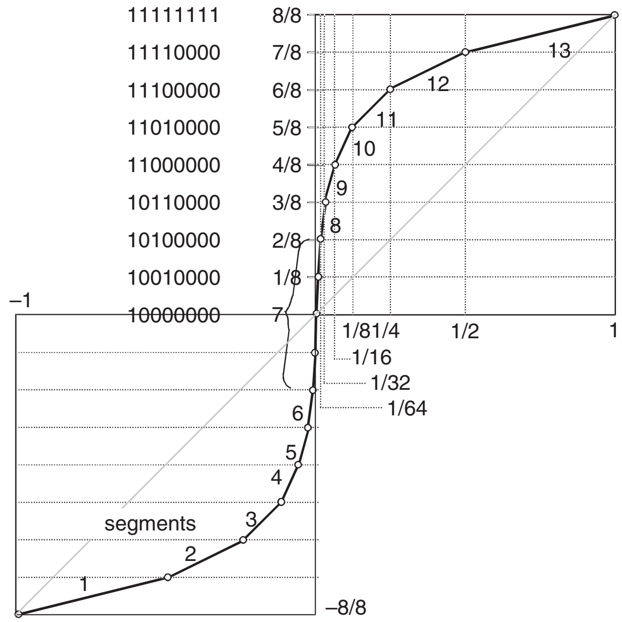

Figure 9.6 The idea of the A-law weighted PCM compounding technique. The voltage level for this solution varies between the values of –8/8 and +8/8 compared to the reference voltage.

whole dynamic area of the band, which is the reason for applying compounding to the PCM sampling. The quantization values that are presented in Figure 9.6 have been selected in an empirical way. The amount of sampling points could logically be higher in order to provide higher quality of the voice calls, but at the same time it would consume more capacity from the transmission. The selected level is sufficiently good for voice call purposes, exceeding clearly the quality that had been obtained in previous, analog versions.

In the receiving end, when the signal is decoded back to the analog format, the code values of the bit stream are interpreted in the same manner as is done in the regenerating transmission line repeaters. The bits are then presented in voltage levels corresponding to bit values. In this phase, when there are suddenly changing levels in the signal, it causes harmonic components for the resulting spectrum. For this reason, low-pass filters are applied in the receiver, which at the same time limits the utilized spectrum to 3.4 kHz upper frequency level.

The mapping of the PCM voltage levels and corresponding bits is the following, as shown in Figure 9.7. The PCM code word’s first bit indicates the polarity. The three following bits indicate the segment, and the four last bits indicate the discrete level within that segment. When the code word is transmitted, the first bit to be transmitted is the polarity bit (B1), corresponding to 2^7. This is at the same time the most significant bit (MSB) of the code word. The last bit to be transmitted is the B7, which is the LSB, least significant

B1	B2	B3	B4	B5	B6	B7	B8
2^7	2^6	2^5	2^4	2^3	2^2	2^1	2^0

Figure 9.7 The mapping of PCM bits.

bit. In practice, the utilized voltage range of the element is dimensioned according to the saturation point of the coder.

Next, the compression that was utilized in the PCM coder is removed by performing a nonlinear amplification, meaning that the signal level is amplified less for the levels that are lower. This technique is called expansion. The combination of compression and expansion is called compounding. It provides with linear transmission of the signal, and in addition, the quantization distortion is relatively constant over the whole dynamic range of the system.

9.6 PDH

The old technique that was utilized in circuit switched telecommunications networks was PDH (plesiochronous digital hierarchy). It was based on the 30 channel PCM transmission system. The PDH technique allows the combination of the basic velocities of 2 Mb/s to higher level transmission systems which provide accordingly higher total data rate. Figure 9.8 shows the levels of the PDH.

As can be seen in Figure 9.8, the first level of transmission is directly derived from the basic PCM, and the higher levels are the multiples, providing digital channeling. As an example, the second level PDH contains 120 basic PCM channels. When the additional synchronization is taken into account, the final bit rate for this level is 8445 kb/s. It should be noted that the highest, V-level of PDH was not standardized completely so there were proprietary solutions.

9.7 SDH

As the PDH systems have gradually moved to history, the enhanced version SDH (synchronous digital hierarchy) has been utilized for providing higher bit rates in a flexible way. In addition, SDH can be more

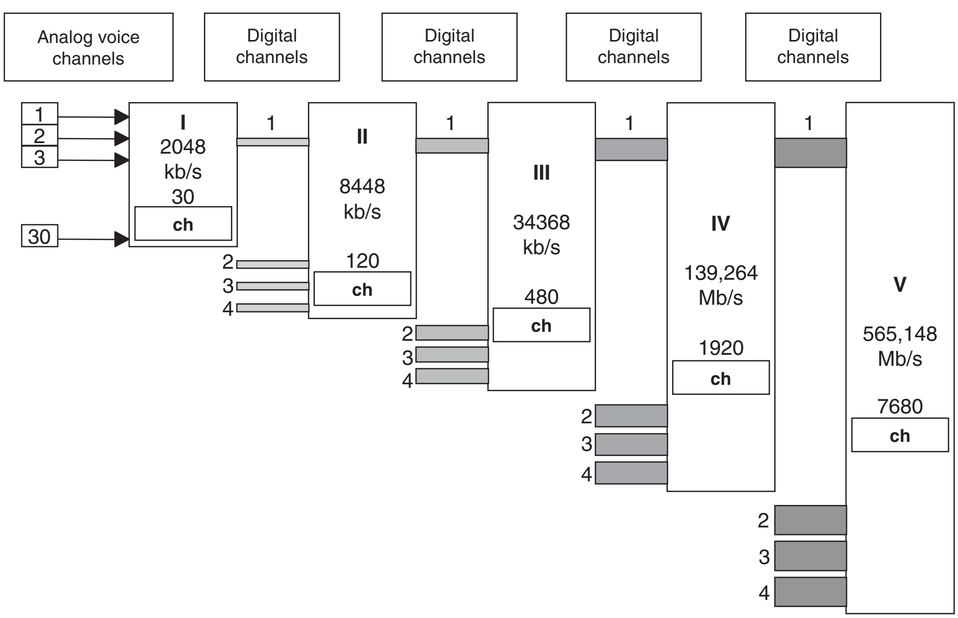

Figure 9.8 *The levels of PDH data rates.*

Table 9.1 *The hierarchy levels of SDM*

Level	Data rate
STM-1	155 520 kb/s
STM-4	622 080 kb/s
STM-16	2 488 320 kb/s
STM-64	9 953 280 kb/s

easily applied to management systems. This is due to the fact that signaling use-cases, for example, for system management were taken into account already in the standardization phase of the system, which guarantees sufficient capacity for signaling.

The benefit of signaling for management systems is that the operations and management center can receive an accurate description of the occurred fault, and also detailed level information can be delivered about the functioning of the network. In addition, SDH is suitable for configuring network settings remotely. This provides means for executing, for example, changes in the switching matrix remotely.

As was the case for PDH, also SDH is based on PCM coding of analog signals. The base line for both PDH and SDH is the basic 30-channel PCM line with 2 Mb/s. The most important difference between PDH and SDH is in canalization hierarchy. PDH is not flexible in this sense, because the output of an individual PCM channel must be done by deconstructing the higher level channels to the basic channels. Instead, the SDH provides the separation of the PCM channels directly from the hierarchy.

Table 9.1 summarizes the hierarchy levels of SDH and the respective data rates. The levels correspond to STM transport frames (synchronous transport module), from which the basic level, STM-1, has a data rate of 155.520 Mb/s.

All the SDH transport modules utilize fiber optics as a physical means. The lowest bit-rate connection, STM-1, can be done in a limited way also via radio links or coaxial cables, and the next one, STM-4, still functions in a limited way via radio link.

The SDH network can be deployed via star, ring, and loop topologies. The failure of the node equipment does not thus break down the network, but redundant connections can guarantee the continuum of the service.

As an example, in typical telecommunications networks, the long-haul network may be deployed via STM-16 connections, the area networks via STM-4, and the user connectivity networks via STM-1. It should be noted that the network management system is connected in any case to each of the network types.

There are three main node types in SDH: Terminal Multiplexer (TM), Add/Drop Multiplexer (ADM) and Digital Cross-Connect (DXC). TM is used to connect the PDH transmission to SDH-type optical signals. ADM adds or separates subchannel signals to/from the main signals, up to the frame structure of 64 kb/s. This can be done thanks to the synchronous transmission method of SDH. The main signals are treaded further via the DXC cross connection elements.

9.8 WDM

SDH provides a maximum data rate of 2.5 Gb/s, or in certain cases, up to 10 Gb/s. The connections via the latter data rate may already experience the dispersion effect of pulses in fiber optics. When part of the pulses arrives with different timing to the receiving end the transmission is distorted. For a typical service level of voice calls, this is not a significant problem though.

The general trend has been that the proportion of fixed network voice calls has already been declining for some years, compared to the popularity of public Internet as an enabler for VoIP calls – among all the other services that are most fluent to transfer via packet switched network infrastructure. The pricing of services has favored this transition greatly, and the importance of the Internet has exploded. This means that even parallel connections of SDH-16 are far away from the optimal solution for responding to the ever increasing capacity demands of Internet users. In addition to the inevitable capacity limitations, they are also relatively expensive to deploy and operate – one example being underwater fiber optics between the Americas and Europe.

WDM (wavelength division multiplexing) is a suitable solution for delivering high-capacity SDH connections. The idea of this technique is to combine various parallel 2.5 GHz transmissions, which are called single WDM channels. These channels differ from each other by the wavelength which makes it possible to distinguish between different WDM channels from each others. This technique is relatively easy to deploy, as it can utilize the same physical fiber optics, and it would be a matter of simply changing the transmission and receiving equipments in both sides of the cable.

One example of the ITU-T – defined 40-channel WDM system is wavelength of 1528 – 1560 nm. This variant provides with a data rate of 100 Gb/s.

9.9 Carrier Ethernet Transport

The Carrier Ethernet Transport (CET) technology is a modern concept to connect the radio access and core parts of different systems in a flexible way as presented in Figure 9.9. CET is especially useful in the deployment of new backhaul networks. In case the native solutions are not available, the connectivity is done via CAT by pseudo wires which emulate the TDM and ATM.

In the current enhancements of the mobile communications networks, the CET concept offers a cost-effective solution that is able to replace the time division multiplex (TDM) transport solutions, including SDH and PDH. It is possible to deploy CET for access networks as well as for aggregation networks. This

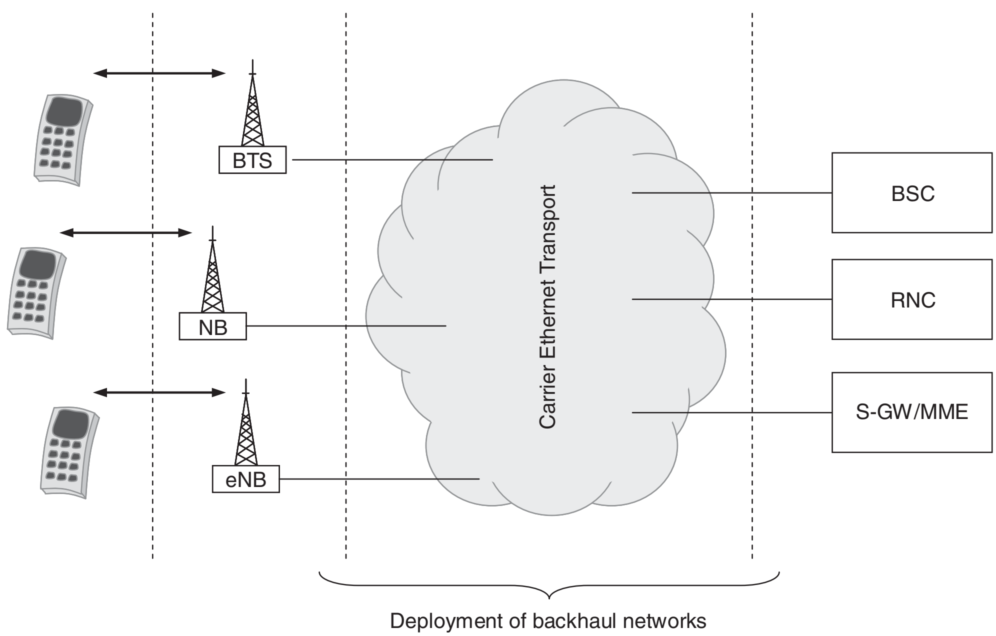

Figure 9.9 *The main idea of the CET concept.*

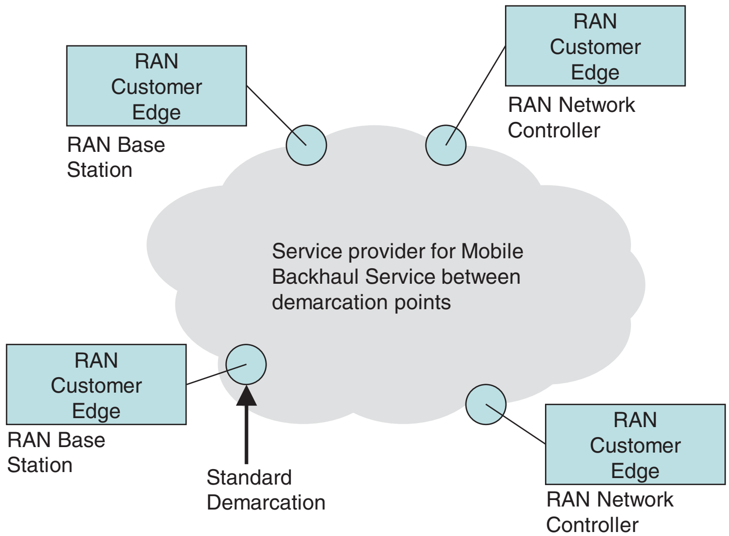

Figure 9.10 *The principle of Carrier Ethernet (CET) according to Metro Ethernet Forum; functional elements as defined by MEF 22.1. The CET Mobile Backhaul Service refers to standard demarcation, standard and scalable services with QoS, and service management and reliability.*

means that via CET, all LTE, 3G and 2G traffic profiles can be delivered over the packet-based backhaul infrastructure.

The main benefits of CET solution are: support of standardized services of variety of physical infrastructures, wide scalability of the bandwidth (from 1 Mb/s to over 10 Gb/s), high reliability, and support of the Quality of Service options. It also offers the possibility to monitor, diagnose and manage the network in a centralized way.

CET has been standardized by the Metro Ethernet Forum (MEF), which provides vendor-independent implementations [1]. Figure 9.10 shows the general principle of the CET solution as interpreted for the MEF documentation.

The transport network must meet the synchronization requirements for all the services it delivers, both in the time-slot structure of TDM as well as in the packet based structures. The standards and recommendations related to the synchronization are the following:

- ITU-T G.781 describes a clock hierarchy deployment model as well as a clock selection process for TDM networks.
- ITU-T G.81x and ITU-T G.82x series recommendations include the performance specifications for TDM networks and synchronization. They also include relatively tight definitions as Maximum Time Interval Error (MTIE) and Time Deviation (TDEV) resulting in compliance for the delivery of ppb accuracy.
- ITU-T G.82xx recommendations include the definitions of the ITU-T Telecom profile for Time of Day transfer. In general, this set includes broad requirements of synchronization over packet networks.
- ITU-T G.8261 defines the limits for SyncE networks and includes aspects for deploying synchronization in a packet network. It also includes test cases for the phase prior to actual deployment.
- ITU-T G.8262 defines performance requirements for accuracy of synchronous EEC. It should be noted that these definitions can also be found in the ITU-T G.81x and ITU-T G.82x series.

- ITU-T G.8264 defines the Ethernet Synchronization Messaging Channel (ESMC) protocol, which is used for managing the hierarchy and deployment of SyncE. ITU-T G.8264 recommendations include refreshed definitions of ITU-T G.781, covering the same aspects for an Ethernet synchronization network.
- ITU-T G.8265 defines the ITU-T Telecom Profile for frequency transfer.
- IEEE-1588 defines Precision Timing Protocol (PTP). IEEE-1588v2 standardizes a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems. It can be used, for example, as defined in ITU-T G.8265 Telecom Profile for frequency transfer.
- MEF 22.1 is divided into Phase 1 and Phase 2. In Phase 1, synchronization architecture performance assumes only TDM interfaces towards base stations. Phase 2 is under development and will add enhancements for interoperable deployments and increases performance of synchronization. MEF-22.1 is based on ITU-T and IEEE standards.

It should be noted that SyncE cannot be deployed on legacy Ethernet networks unless the hardware and interfaces are renewed, which can be a limiting factor in network deployments between operators. Nevertheless, SyncE is not affected by network impairments such as frame delay range (FDR). It should also be noted that IEEE 1588v2 can be deployed with legacy Ethernet network elements which do not contain the protocol, if the performance of the slave clock recovery algorithms are at an adequate level in case load-dependent network impairments occur.

Also various dedicated synchronization interfaces have to be supported, including primary reference clock (PRC), building integrated time source (BITS) for T1, E1, 2 MHz, and 6 MHz, one pulse per second (1PPS), and time of day (ToD). Table 9.2 summarizes the interfaces.

The technology interfaces are, amongst others, seen in Table 9.2.

E1/T1 (PDH), SDH/SONET, GPS and 1PPS+ToD are the systems typically used in the synchronization in TDM networks. As an example, 2G base stations can be connected to the backhaul network by using E1 or T1. For those operators that deploy TDM networks would also maintain some T1 or E1 lines for making sure that the synchronization is still maintained. In the solution, the operator is thus able to transport data services via parallel way, by utilizing asynchronous packet network. Furthermore, Ethernet can connect the 3G and 4G base stations to the packet based backhaul. Via this solution, the operator can thus deploy services in a hybrid manner, keeping voice traffic on the TDM network and data services on the packet network. Due to relatively large investments during the past years (decades), operators might not want to swap the existing T1 and E1 transmission lines in a fast time schedule. It can be thus estimated that that T1/E1 line synchronization will remain popular with operators at least through 2013.

Figure 9.11 presents an example of the way to deploy new IP or Ethernet base stations in the intermediate transition phase when the TDM is evolved to packet transport variant or to hybrid-type of networks. The benefit of this solution is that the already existing networks can still manage the seamless synchronization of SONET/SDH and Ethernet in the case of these combined technologies. SDH/SONET can be utilized as a base solution during the transition phase in order to providing a transport solution and end-to-end synchronization over certain parts of the hybrid network. In this solution, SyncE provides similar functionality for the rest of

Table 9.2 *Technology interface comparison*

Interface	Physical layer	Packet layer
MEF-compliant UNI	SyncE	IEEE 1588v2
TDM	SDH / SONET, PDH	ACR / NTP (MEF8)
Nontraffic interface	2 MHz, 6 MHz, 1 PPS, BITS	NTP, ToD

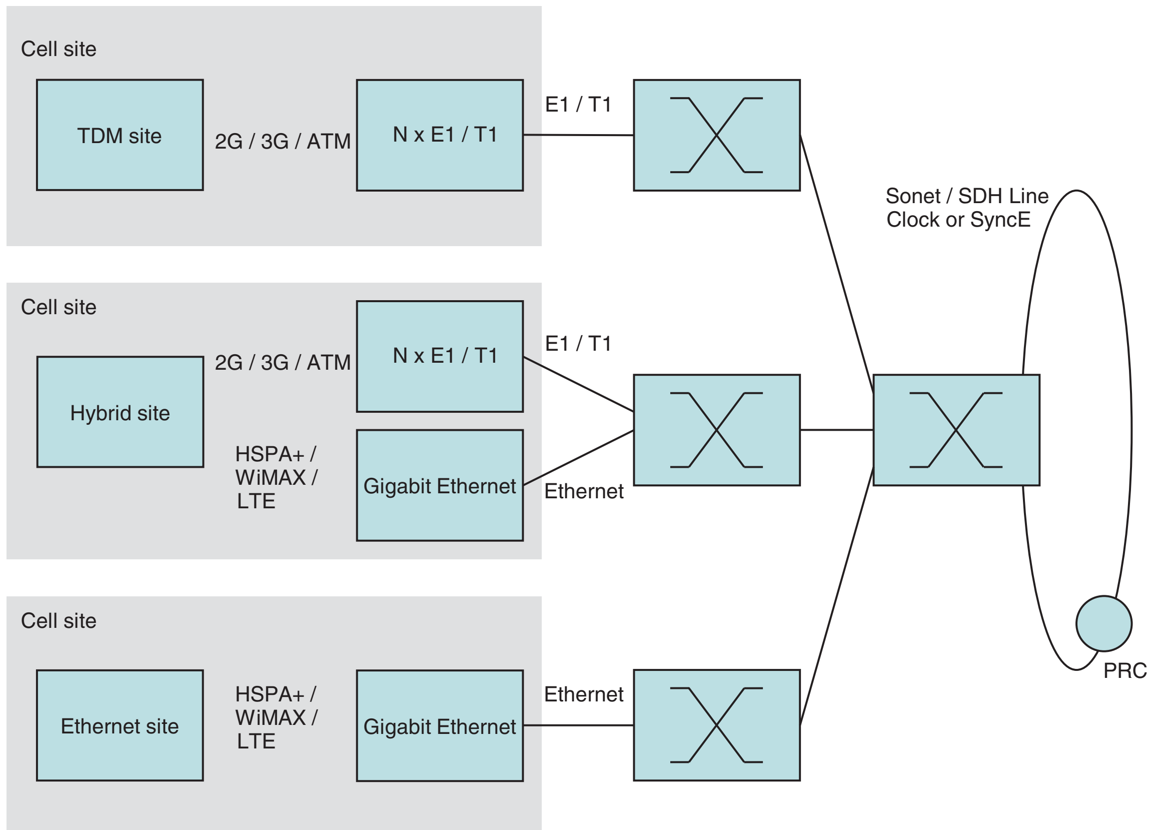

Figure 9.11 *The architecture for synchronization distribution of TDM and Ethernet base stations. In this example, the primary reference clock (PRC) is originated from SONET/SDH.*

the network, including end-to-end synchronization. Furthermore, the solution provides seamless interworking regardless of the selected network synchronization.

9.10 IP Multimedia Subsystem

9.10.1 IMS Architecture

IMS architecture has been defined according to the following basic principles. First, of all the IMS architecture is home network centric in the sense that all services are executed by the home network. IMS architecture does support roaming model with familiar concepts of visited and home network, but the visited network mainly provides a local access point for SIP connectivity as Proxy Call State Control Function (P-CSCF) as well as local policy control functionalities in the form of Policy and Charging Rules Function (PCRF) as defined by 3GPP. In the home network actual services are provided by individual Application Servers (AS), which are typically defined as logical functionalities based on their nature of service such as Telephony Application Server (TAS) for telephony supplementary services, Push to Talk Application Server (PoC AS) for Push-to-Talk services as well as Presence Server (PS). Application servers can be manifested by different physical implementations that can either reside in standalone hardware or integrated as part of some other functionality, depending on vendor.

The Core IMS architecture in the home network is built from Interrogating Call State Control Function (I-CSCF) as well as Serving Call State Control Function (S-CSCF). I-CSCF is responsible for resolving suitable S-CSCF for served IMS registered subscribers whereas S-CSCF is responsible for orchestration of service execution by selecting proper application servers for session as well as authenticating and performing IMS registration procedures jointly with other CSCF roles and naturally terminal. The 3GPP defined Home Subscriber Server (HSS) function contains subscription data related to use of network services, including IMS and also Circuit Switched and Packet Switched subscription profiles. However in practical deployments HSS does not contain both CS/PS and IMS related subscription data but instead HSS products are introduced beside standalone HLR network elements in order to provide support for IMS as well as optionally for LTE related subscription data. The issue that HSS products may not support legacy Packet Switched (GERAN/UTRAN) or Circuit Switched data means that in some cases communication service providers are keen to deploy LTE subscription data into HLR instead of new HSS products. The end result is likely in any case the same.

CSCF and S-CSCF are always located in the home network of given subscribers. In case two IMS networks are involved in a communication session between end-users then standardized Network–Network Interface (NNI) will be used to interconnect these IMS networks together.

Interworking with Circuit Switched networks is required, since most end-users are still using Circuit Switched services, and is the basic requirement for any IMS deployment today. Therefore 3GPP has also standardized IMS-CS interworking via Media Gateway Control Function and IMS Media Gateway functions. Figure 9.12 presents the high level IMS architecture in both visited and home networks [2].

The following paragraphs describe in more detail required parts of the IMS architecture and technologies needed for native voice and video telephony over IP and Short Message Service (SMS) over IP.

9.10.1.1 P-CSCF

Proxy Call State Control Function (P-CSCF) acts as the SIP proxy and as the first point in either the home or visited network where the end-user terminal takes contact in order to obtain access to IMS services. P-CSCF will select suitable Interrogating Call State Control Function (I-CSCF) within the home IMS network (which can be in a different country) in addition to the following tasks.

P-CSCF is responsible for providing sufficient security measures in order to keep integrity and security of SIP signaling between terminal and itself as well as asserting the subscriber's identity towards other IMS network elements such as S-CSCF. P-CSCF will not be changed in a typical case during active IMS registration and it needs to be able to handle both own and inbound roaming subscribers from other networks. Security and integrity protection is achieved through the use of IPSec but in the past IPSec has not been widely supported by SIP capable endpoints. This is now expected to be changed since adoption of GSMA "IMS profile for voice and SMS."

P-CSCF is responsible for handling resource reservations via Policy Charging Control (PCC) architecture optionally deployed in access network (such as LTE). P-CSCF does this via Diameter based Rx interface to Policy and Charging Rules Function (PCRF). P-CSCF in this respect implements Application Function (AF) as defined in 3GPP standardized PCC architecture [t]. When PCRF is used in network it will communicate via Diameter based Gx interface to Policy and Charging Enforcement Point (PCEF), which in case of LTE resides in Data Network Gateway (PDN GW). P-CSCFs responsibilities include taking part in codec negotiation (media negotiation) between intended SIP endpoints and then based on negotiated result, requesting

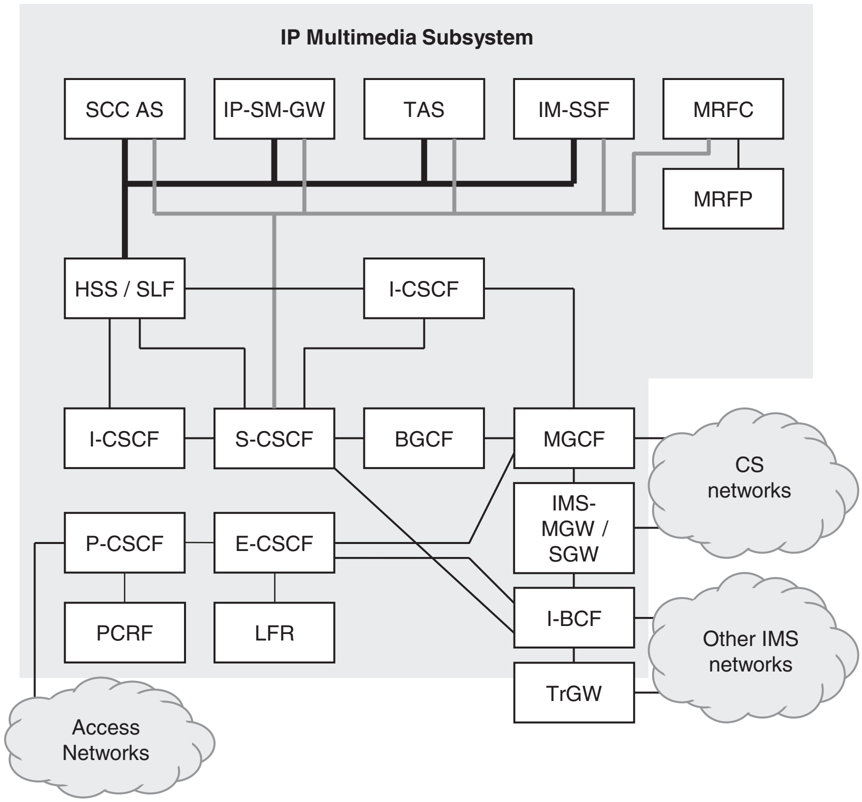

Figure 9.12 *The IMS architecture.*

needed resources from PCRF. In voice or video telephony over LTE this means resources for both voice and/or video codec.

Use of PCC is optional and thus may or may not be deployed by communication service providers. PCC may even be deployed before any voice or video telephony is deployed in order to categorize different users (gold, silver, bronze) or to enable IP flow based differentiated QoS for Internet services. Additionally some access network side products implementing for example, PDN GW functionality may have inbuilt functionalities to provide Quality of Service (QoS) for basic data services without need for Gx interface or PCC architecture. This means that P-CSCF may not need to support Rx interface and if it does then it may be the first time when that interface is used within the network.

From a practical deployment point of view P-CSCF may be colocated as part of products that implement other IMS functions or in some cases within the Session Border Controller product already deployed at the edge of the network. Both scenarios are valid and depend on the existing network architecture. Additionally it is possible that LTE is deployed as a solution for Broadband Wireless Access connecting for instance the entire home and all related IP capable equipments into Internet and communication service provider services. This means that in practice such P-CSCF may have to support simultaneous use of PCRF as well as media anchoring, for instance to overcome problems caused by far-end NAT in similar fashion as those supported by Session Border Controllers do today – despite the product being used for P-CSCF.

9.10.1.2 I-CSCF

Interrogating Call State Control Function (I-CSCF) acts as the SIP proxy and typically acts as the first point of contact in the home IMS network. However in some cases the visited network may also have I-CSCF functionality in order to hide the topology of the network. I-CSCF is contacted by P-CSCF during IMS registration in order to obtain access to IMS service as well as hiding the topology of the home IMS network from external world (Topology Hiding, THIG) Interface between P-CSCF and I-CSCF is based on 3GPP standardized Session Initiation Protocol (SIP) and routing of SIP messages is based on the Domain Name System.

I-CSCF is also the first node that interfaces with HSS of IMS subscriber. I-CSCF uses the Diameter based Cx interface to fetch subscriber information from HSS during IMS registration as well as deciding which Serving Call State Control Function (S-CSCF) will be suitable for a given IMS subscription. In case home IMS network supports multiple HSS instances (network elements) then I-CSCF may use Diameter based Dx interface towards Subscription Locator Function (SLF), which may be either co-located with other IMS functionalities (such as HSS) or deployed in standalone fashion. SLF redirects Diameter requests to the appropriate HSS which contains subscription data of the given IMS subscriber.

From voice and SMS over LTE point of view there are no specific new requirements for I-CSCF and therefore I-CSCF is not described in more detail in this book.

9.10.1.3 S-CSCF

Serving Call State Control Function (S-CSCF) acts as the SIP registrar of IMS subscriber by acting as the end point in the IMS network for IMS authentication (AKA) as well as coordinating which IMS services and in which order will be applied for a given IMS subscriber. S-CSCF will perform authentication and inform HSS about registration status of IMS subscriber via Diameter based Cx and/or Dx interfaces. HSS needs to be aware of the identity of S-CSCF, for instance in order to handle routing of terminating SIP sessions correctly when interrogated by I-CSCF. This routing of terminating requests loosely resembles the behavior of HLR and gateway MSC in traditional Circuit Switched mobile networks. The interface between I-CSCF, P-CSCF and S-CSCF is based on 3GPP standardized SIP.

Beyond functionalities listed above, S-CSCF is also responsible for deciding whether a given IMS subscription is entitled to certain type of communication based on media type (voice, video) as well as translating used identities in the SIP signaling to SIP Uniform Resource Identity (URI) format in case Telephony Uniform Resource Locator (URL) has been used by terminal.

S-CSCF involves required Application Servers (AS) into SIP session in order to provide actual services for IMS subscriber based on user profiles retrieved from HSS during registration or in case the user profile has been changed by the communication service provider. In case of voice, video and SMS over LTE this means that Telephony Application Server (TAS) as well as IP-Short Message-Gateway (IP-SM-GW) functionalities are notified of IMS registration by using 3rd party registration procedure and succeeding SIP messages related to these particular services will be routed via these application server instances in order to invoke service execution.

In practical deployments capabilities of S-CSCF vary between network vendors. In some cases S-CSCF may even have inbuilt application server functionalities in order to achieve higher flexibility to route SIP sessions as well as manipulate SIP headers within SIP messages. Similarly same functionalities may be used in some cases to develop more advanced Service Control Interaction Management (SCIM) in order to achieve higher grained control of service interactions in case more than one service will be applied for single SIP session.

In case of voice, video and SMS over LTE the S-CSCF represents important building block within IMS architecture as described in this chapter but as such these use cases will not pose any significant requirement independent of other IMS use cases.

9.10.1.4 E-CSCF/LRF

Emergency Call State Control Function (E-CSCF) is functionality that is required to complete emergency IMS session either in visited or in home network. E-CSCFs is invoked by P-CSCF after P-CSCF detects that the nature of the session is an emergency for instance based on the value of the received Request URI parameter. After this E-CSCF will resolve required location information through the help of the Location Retrieval Function (LRF) which again is able to either use received signaling level information (P-Access-Network-Info header of SIP message) from terminal or use Location Service framework (LCS) possibly existing in the network. Location information is typically used at least to select Public Safety Answering Point (PSAP) which is responsible for emergency calls from given location. Conversion of location to PSAP address is done at LRF and this PSAPs routable address (for instance SIP URI or TEL URL) is returned to E-CSCF in order to route calls either via MGCF to Circuit Switched networks in case PSAP does not have native SIP connectivity or by using SIP.

9.10.1.5 Home Subscriber Server and Subscriber Locator Function

Home Subscriber Server (HSS) acts as the main subscriber data repository of IMS user profiles. This data contains information related to identities and services of given subscriptions. Subscriber Locator Function (SLF) is required in case the IMS network has multiple HSS entities and the requesting function (e.g., I-CSCF or AS) requires knowledge about which individual IMS user profile is located in which HSS entity.

From voice and video telephony over the LTE point of view, the data stored within HSS may contain in addition to the identity of Telephony Application Server (TAS) entity and also information about provisioned supplementary services. This information may be stored in XML document format as defined originally by 3GPP for XCAP based Ut interface but also it may be stored optionally as binary based format that resembles more to the way how HLR store information today in Circuit Switched network. In case XML document format is used, then HSS may not have any understanding about actual content of this document since it is stored as part of generic-purpose Application Server specific data containers in IMS user profiles.

From the Short Message Service point of view HSS needs to be provisioned with identity of IP-Short Message-Gateway (IP-SM-GW) AS entity that is responsible for handling that particular IMS subscriber. This information is then used by S-CSCF to select IP-SM-GW when the IMS subscriber performs IMS registration.

From the IM-SSF point of view HSS needs to be provisioned with the identity of IP Multimedia – Service Switching Function (IM-SSF) entity that is responsible for handling that particular IMS subscriber. This information is then used by S-CSCF to select IM-SSF when the IMS subscriber performs IMS registration.

9.10.1.6 Application Servers

Application Servers (AS) can be considered to be the workhorses of IMS architecture to provide business critical services for IMS subscribers. The underlying architecture of various Call State Control Functions are just as important but have less significance when considering the service logic itself.

3GPP has standardized logical AS entities which can be productized in a vendor specific manner. However some grouping of functionalities can be found from the market such as that voice and video telephony related services are supported by a single product but then more advanced, programmable services on top of frameworks such as JAIN Service Logic Execution Environments (JSLEE) are productized as part of other products.

From voice and video telephony point of view, one of the most important 3GPP standardized functionality is called Telephony Application Server (TAS), which is responsible for providing 3GPP defined Multimedia Telephony (MMTel) services for IMS subscribers entitled to the service. From the Short Message Service over IP (SMS) point of view the most important 3GPP standardized functionality is called IP-Short Message-Gateway (IP-SM-GW) that provides business logic for handling of the Short Message Service as well as interworking to legacy Circuit Switched networks, when required. Interworking to legacy Intelligent Network (IN) services may be required. In that case IMS architecture has a dedicated AS functionality for IM-SSF, which is able to translate SIP session to appropriate CAMEL or INAP service control protocol towards existing Service Control Point (SCP). Additionally in case there is need to support service continuity and the network supports IMS Centralized Services architecture then the so-called Service Centralization and Continuity Application Server (SCC AS) is involved in the path of IMS session. SCC AS is responsible for important tasks such as anchoring the session for possible forthcoming domain transfers due use of Single Radio Voice Call Continuity (SRVCC) as well as to perform Terminating Access Domain Selection (T-ADS) to select either Circuit Switched or IP based access network for terminating call in case the terminal can be reached via both accesses.

Rich Communication Suite is a separate IMS application suite that uses own specific application server functionalities such as Presence Server, Instant Messaging application server as well as XML Document Management Server (XDMS). In practice these functionalities are not mandatory for implementation of "IMS profile for voice and SMS," with exception of XDMS in case it is used in context of Ut interface for Multimedia Telephony, but these functionalities may also be deployed in parallel, if so wished by the communication service provider.

In case AS instance require access to IMS user profile that is stored within HSS, this access is possible via the Diameter based Sh interface. In case network has multiple HSS instances, then the Diameter based Dx interface needs to be used towards SLF in order to redirect AS to HSS containing the wanted IMS user profile. 3GPP has also defined Diameter based Si interface which could be used by IM-SSF to fetch IN-related subscription data from the IMS user profile in HSS. However, this Si-interface may not be required if IM-SSF is able to use other mechanisms to fetch the required data from the subscriber data repository, which can be the situation if IM-SSF is colocated with some other product such as MSC. All in all it is possible that various different kinds of AS implementations exists in the market and not all interfaces are supported by them if the same end to end functionality can be achieved some other way without visible impact on terminals and other IMS entities.

9.10.1.7 MGCF, IMS-MGW, I-BCF and TrGW

The Media Gateway Control Function (MGCF) and IMS-Media Gateway (IMS-MGW) are functionalities which are typically involved when SIP session is routed between IMS subscriber and Circuit Switched endpoint. In this case MGCF is responsible for signaling related tasks such as conversion between SIP and SDP signaling used in the IMS network as well as signaling protocols used in Circuit Switched networks such as ISDN User Part (ISUP), Bearer Independent Call Control (BICC) as well as even specific variant of SIP protocol which tunnel the ISUP messages, SIP-I. MGCF also controls the user plane resources required for such interworking and located in IMS-MGW via H.248 protocol based 3GPP Mn interface. In a typical case multiple IMS-MGWs can be controlled by single MGCF and vice versa, thus maximizing the flexibility of network planning.

IMS-MGWs at minimum need to be able to handle transport level interworking for instance between TDM and IP based transport but in addition to this also codec level interworking that is usually called transcoding.

Transcoding may be support for both voice and video codecs or only voice codecs depending on capabilities of the IMS-MGW product used.

In practical deployments in mobile networks the MGCF and IMS-MGW are typically colocated in mobile soft switching solution consisting of MSC Server (MGCF) and MGW (IMS-MGW). This way it is possible to optimize media plane routing in such calls that require use of MGCF and IMS-MGW and either originate or terminate to Circuit Switched mobile terminal since no separate transit MGW may be required.

Voice and video telephony over LTE requires that MGCF and MGW are able to support codecs mandated by the 3GPP specifications as well as GSMA "IMS profile for voice and SMS." Support for High Definition voice with Wideband Adaptive Multi Rate (WB-AMR) speech codec requires additional capability from MGCF and IMS-MGW to support interworking of SIP session to Circuit Switched call by using Transcoder Free Operation (TrFO) or Tandem Free Operation (TFO) depending on the call scenario. These two technologies are mandatory in order to support WB-AMR in Circuit Switched networks.

In order to support interworking between SIP based video telephony and 3G-324M as defined by 3GPP [5], depending on capabilities of used products for IMS-CS interworking, either integrated or standalone video gateway should be used. In case standalone video gateway installation is used that is different from MGW used for audio-only calls, then routing of calls need to be done in such a way that voice calls and video calls towards IMS are routed for instance with a different prefix in front of the called party number in order to use different gateways correctly.

In case of interworking via IMS Network Network Interface (IMS-NNI) to other IP based networks it is possible to deploy Interconnection Bearer Control Function (I-BCF) together with Transition Gateway (TrGW) functionality. I-BCF and TrGW may be used in order to provide security functionalities to prevent Denial of Service (DoS) attacks from unsecure IP interconnections but also to perform user plane related functionalities such as transcoding, if required for IMS sessions that break-in or break-out from IMS. Additionally it may be possible, depending on the product that offers I-BCF and TrGW functionalities, to use the same product also for SIP-I interworking between Circuit Switched core networks. In this way it is possible to achieve synergies between these different domains.

9.10.1.8 Media Resource Function Controller and Processor

Media Resource Function Controller (MRFC) and Media Resource Function Processor (MRFP) provide media plane related functionalities in case they are needed from the IMS network. These capabilities typically mean injection of in-band tones and announcements as well as collecting in-band information such as DTMFs. Additionally these functions may provide support for network based conferencing similar to what exists in Circuit Switched mobile networks today as Multiparty supplementary service. Typical commercial MRFC/MRFP products have a great amount of flexibility and support multimedia in various use cases including conferencing.

Despite the fact that 3GPP originally standardized two separate functionalities for Media Resource Function (MRFC/MRFP) these are typically in commercial products sold as standalone entity with possible capability to also have functionalities deployed in separate manner, if so required. In addition to standalone elements it is possible that some vendors may have again colocated the relevant functionality for certain services (such as voice conferencing or capability to deliver in-band voice announcements for voice telephony) in some existing product and thus provide more value for their communication service provider customers having the product.

Voice and video telephony over LTE is considered to require support for similar in-band interaction that exists in today's Circuit Switched mobile telephony. This means that similar announcements given by network as well as tones also need to be available when voice is deployed over the IMS network. Similarly, but less often, consumer ad-hoc conferencing functionality is also required from Circuit Switched networks which also places similar requirements for the IMS network.

9.10.1.9 SRVCC and ICS Enhanced MSC Server

Current modern mobile networks have MSC Server system which enables communication service providers to use packet switched transport for Circuit Switched calls as well as for signaling. Similarly Media Gateway platforms may have additional capabilities to support other use cases beyond Circuit Switched calls.

3GPP Release 8 has defined new functionality for MSC Server to assist in the service continuity procedure termed Single Radio Voice Call Continuity (SRVCC) as part of 3GPP TS 23.216. SRVCC means continuation of voice call when the terminal moves from LTE to Circuit Switched network. SRVCC enhanced MSC Server has specific GTP based Sv interface towards MME function. This Sv interface as defined in 3GPP TS 29.280 is used by MME to request MSC Server to reserve required radio access resources from target Circuit Switched radio access (GERAN/UTRAN) for SRVCC, which may occur. SRVCC enhanced MSC Server will prepare resources either to locally or remotely connected IuCS or A interface. In case target radio access is controlled by another MSC Server (MSC-B) then SRVCC enhanced MSC Server will perform normal Inter-MSC relocation. After target Circuit Switched radio access resources have been committed then SRVCC enhanced MSC Server will establish calls on behalf of the terminal to specific address given by MME via Sv interface. This address is related to current SCC AS of that particular subscription and involved in the original call establishment.

SRVCC procedure has been gradually improved between 3GPP Release 8 and Release 10 releases to support more functionality such as capability to support multiple simultaneous calls (active and held) as well as capability to perform reverse SRVCC from Circuit Switched network to LTE. In order to support functionalities beyond 3GPP Release 8 additional requirements set by 3GPP IMS Centralized Services (ICS) architecture needs to be taken into use. This can occur in a phased manner, in case IMS based voice over LTE has been deployed commercially by using 3GPP Release 8 standardization baseline. 3GPP Release 9 introduces specific "MSC Server assisted mid-call" functionality that is based on ICS enhanced MSC Server functionality. This functionality is required to support SRVCC for multiple ongoing call (active and held) in case the terminal is not ICS enabled, that is, it does not have capability to use Circuit Switched network as bearer for a session established by using SIP. Additionally, in case reverse SRVCC is required, then also Circuit Switched calls originated by the terminal need to be anchored in IMS (SCC AS); which means that IMS Centralized Service architecture in full effect need to be deployed into the network.

9.11 Case Example: LTE Transport

9.11.1 Ethernet Transport

The basic LTE/SAE solution includes electrical and optical Ethernet interfaces which provides the operator with the lowest transport cost with high offered transport capacity. More specifically, the physical solution can be a Gigabit Ethernet 100/1000Base-T with electrical connectivity via the RJ-45 standard and 1000Base-SX/LX/ZX with optical connectivity. Furthermore, the logical functionality includes the automatic negotiation of the mode and data rate. Figure 9.13 shows the Ethernet solution for LTE/SAE transport.

9.12 Cloud Computing and Transport

Cloud computing refers to virtualization of resources, which can be computing capacity and hosting facilities. The benefit of cloud computing is increased efficiency and dynamics of resources, which in turn offers cost savings in network hosting. This is a result of virtualization and uniform resource base, and in general, economies of scale.

For applications that require storage, the limiting factor is latency and packet loss rate. In general, internal bandwidth of the cloud data center requires several times more bandwidth compared to the connection

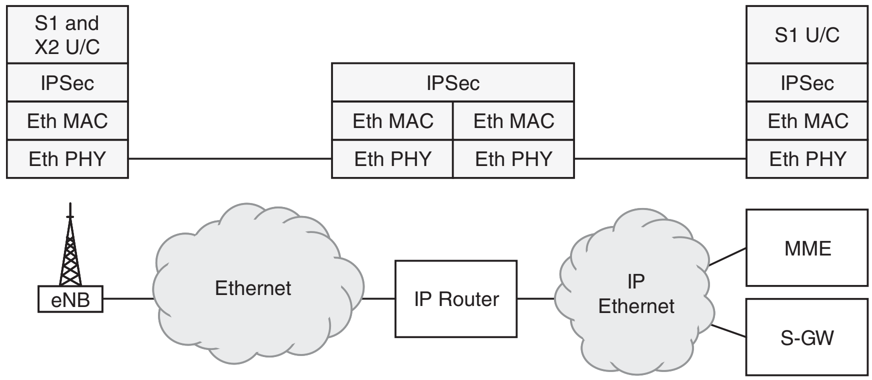

Figure 9.13 *The Ethernet solution for the LTE/SAE transport.*

between users and data centers. For coping with the challenge of the bandwidth, there are techniques that optimize band per user, for example, by splitting the connection into several parallel streams, thus reducing the load per server. This, in turn, requires more accurate synchronization of various locations.

There is seen a transition of local computing and storage resources to cloud based data centers. This requires interconnection arrangements which need to be equipped with additional security level for application access and for interconnection of private and public clouds.

The different traffic flows between cloud and other entities are the following, as defined in Ref. [3] and shown in Figure 9.14:

- Traffic from data center to user is traffic that flows from the data center to end-users through the Internet or IP WAN.
- Traffic from data center to data center is traffic that flows from data center to data center.
- Traffic within data center is traffic that remains within the data center.
- Consumer traffic is a type of traffic that originates with or destined for consumer end-users.
- Business traffic is a type of traffic that originates with or destined for business end-users.
- Cloud data center traffic is a type of traffic associated with cloud consumer and business applications.
- Traditional data center traffic refers to traffic associated with noncloud consumer and business applications.

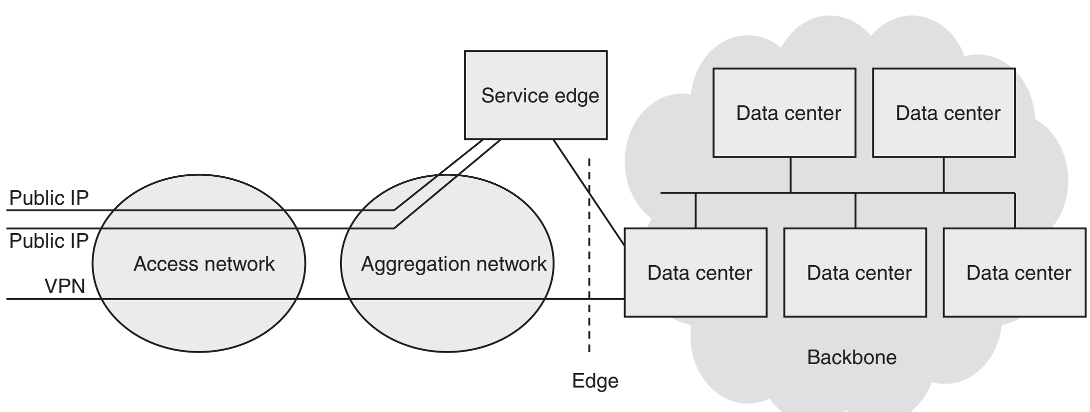

Figure 9.14 *The transport connectivity of cloud service.*

From the security point of view, one practical and efficient method is VLAN Layer 2 isolation technique. It is also a useful solution for resalable network virtualization.

Cloud computing network connectivity is possible to offer by bridging VLANs with an overlay which eases geographic restrictions due to more reasonably sized user groups without the need for complicated virtualized networking component definitions such as switching or routing. Also the architecture is simpler thanks to virtualized control plane.

The near future development of clouds needs to take into account estimated use cases. It can be assumed that in the consumer segment, applications like video and audio streaming are amongst the strongest drivers for increased cloud traffic. On the other hand, more modern services such as personal content lockers also contribute to the growth of capacity demand. Personal content lockers refer to storing and sharing of music, photos, videos and other, for example, multimedia type of data through an intuitive user interface with non-significant cost levels. In addition, the increased markets of tablets, smartphones, and other multimedia capable mobile devices can be used for accessing personal content lockers on the road [4].

Typical requirements for cloud network access and performance are the following, as defined in Ref. [4]:

- Broadband ubiquity which refers to fixed and mobile broadband penetration while considering population demographics to understand the pervasiveness and expected connectivity in various regions.
- Download data speed as a result of increased penetration of mobile and fixed bandwidth-intensive applications. This is a critical item for a high quality of service.
- Upload data speed is also required for the same reasons and in the case of download speeds, for delivering content to the cloud. It can be estimated that the importance of upload data speed will increase further as a result of, for example, storing of large files to virtual file systems, and due to demand for consumer cloud game services and backup storage.
- Network latency refers to delays experienced by end-users in case of, for example, VoIP services, uploading large real-time data streaming files, online banking services via mobile broadband, among many other critical cases that require respond times in order of few milliseconds. For these cases, the reduction of packet delivery delay in both uploading and downloading is essential for ensuring high-quality end-user experiences.

As a summary, as concluded in Ref. [4], data center virtualization is important traffic promoter in the fast development and expansion of cloud computing. This makes it possible to offer flexible, fast deployable and highly efficient services. Furthermore, wide adoption of multiple types of devices that can be utilized independently form the location and independently from the networks, is also increasing utilization of cloud computing. Cloud-based data centers are optimal for supporting a higher amount of virtual machines and workloads than traditional data centers. Source Ref. [4] estimates that by the year 2016, close to two-thirds of workloads will be processed in the cloud.

References

1. Metro Ethernet Forum. http://metroethernetforum.org/ (last accessed 27 July 2014).
2. Penttinen, J. (2011) *The LTE/SAE Deployment Handbook*, John Wiley & Sons, Inc.
3. Ericsson. Ericsson Review 284 23-3174. Transport networks in the cloud age. 1/2012. 6 p.
4. Cisco, White paper, 2011, Cisco Global Cloud Index: Forecast and Methodology, 2010–2015. http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns1175/Cloud_Index_White_Paper.html (last accessed 27 July 2014).
5. ITU-T H.324 specification for multimedia conferencing over circuit switched networks.

10

Modulation and Demodulation

Patrick Marsch and Jyrki Penttinen

10.1 Introduction

Modulation refers to the way in which one or more properties of a carrier signal may be varied in order to embed an information signal into the carrier signal and send this over a physical transmission channel, e.g. over air or cable. One generally distinguishes between analog modulation, where an analog information signal is modulated, and digital modulation, where a sequence of data bits is modulated onto a carrier signal.

This chapter provides a brief overview of analog modulation techniques and describes in detail the digital modulation techniques that are most relevant to today's telecommunication systems.

10.2 General

Let us assume that we want to transmit an audio signal such as the simple sine wave depicted in Figure 10.1, having a frequency within the audible range of, say, 10 kHz. Clearly, if we transmit this as it is, for instance just sending this 10 kHz signal over a wireless channel, this would be rather inefficient because

- propagation conditions at these frequencies are very bad, and
- it would only be possible to transmit one such audio signal in a certain area, as multiple transmissions would otherwise be interfering with each other.

For this reason, information signals such as audio signals are typically modulated onto carrier signals of higher frequency. This way, frequencies can be chosen that are particularly suitable for, e.g., wireless transmission, and also many information signals can be transmitted in parallel by simply modulating these onto carriers of different frequency.

Once a signal is modulated and sent over the physical channel, the receiving end demodulates the signal in order to recover the original message. This process of modulation and demodulation has led to the term "modem."

The Telecommunications Handbook: Engineering Guidelines for Fixed, Mobile and Satellite Systems, First Edition.
Edited by Jyrki T. J. Penttinen.

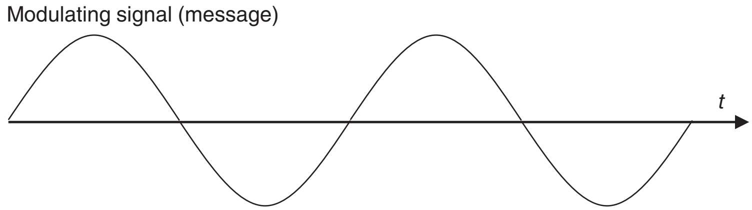

Figure 10.1 *Example of a signal or message to be modulated (sine wave).*

In practice, all the modulation and demodulation methods that telecommunications systems use are based on the variation of amplitude, frequency or phase of the carrier signal, or on a combination of these, as explained in the following section.

10.3 Analog Modulation Methods

10.3.1 Amplitude Modulation

The simplest form of analog modulation is amplitude modulation (AM). It provides the modulation of a signal $m(t)$, for instance an audio signal as in the previous example, to the carrier frequency $c(t)$ by utilizing the original amplitude variations of the signal $m(t)$. Figure 10.2 presents an example of the amplitude modulation of the example signal from the previous figure in the time domain.

In practice of course, the modulation signal hardly consists of a single sine wave. Instead, as Fourier decomposition dictates, $m(t)$ consists of a sum of various sine waves of multiple frequencies, amplitudes and phases. Through modulation of $m(t)$ onto carrier frequency f_c, each frequency component of $m(t)$ with a frequency f_i results in a signal at frequency $f_c + f_i$, and a signal at frequency $f_c - f_i$. The set of frequencies $f_c + f_i$ above the carrier frequency is called upper side band (USB), and the set of frequencies $f_c - f_i$ forms the lower side band (LSB). The modulated signal thus consists of a symmetric, mirror-imaged set of positive and negative frequency components as presented in Figure 10.2. It is possible to transmit only one side band of the modulated signal (i.e. either the LSB or USB), while still being able to completely reconstruct the original modulating signal (given that this is real-valued).

In all the above-mentioned cases, the final quality of the reception depends on the received power level of the signal. At the transmitting end, the grade of the amplitude modulation is defined as $m = (E_{max} - E_{min})/(E_{max} + E_{min})$. If m exceeds the value of 1 (100%), the AM signal is over-modulated, which results in distortion.

The benefit of amplitude modulated transmission is the simplicity of the transmitter and receiver. AM was thus the first modulation method utilized in early radio broadcasting stations in the beginning of 1900s. Regardless of the active digitalization of the broadcast systems, AM is still widely utilized especially in medium wave bands, and partially also in short and high frequency bands, e.g., for delivering news, which does not require much bandwidth.

10.3.2 Frequency Modulation

Frequency modulation (FM) is another analog modulation method. The difference between AM and FM is that in FM, the amplitude of the carrier is kept constant, but the frequency is varied according to the original source of the signal which may be, e.g., audio. The information of the frequency modulated signal is thus contained in the frequency of the carrier.

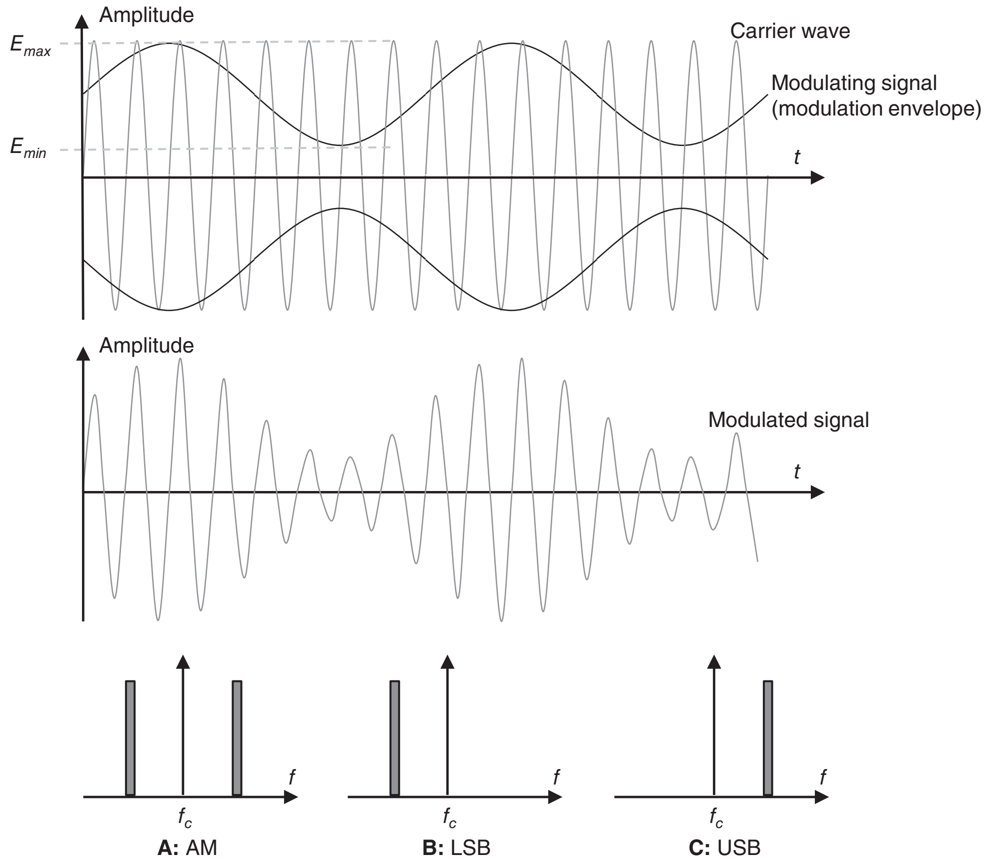

Figure 10.2 *An example of amplitude modulation (AM).*

In FM, the modulating signal thus lets the carrier frequency vary, and the variations are controlled by the frequency as well as the amplitude of the modulating wave, as can be seen in Figure 10.3. The resulting spectrum is also different from AM, as the figure shows. Compared to only one spectral component on both sides of the AM central frequency, or a single component in case of SSB, there is an infinite amount of spectral components in FM in theory; in practice, limited bandwidth can deliver the relevant part of these components.

The digital correspondence of FM is Frequency Shift Keying (FSK), as explained in Section 10.4.5. Analog FM is still utilized widely in the VHF band for radio broadcasting, including high quality music. The term FM band, which is used in commercial markets, indicates the frequency band dedicated to FM broadcasting, which is typically 87.5–108 MHz in most countries or 76–90 MHz in Japan. The digital audio is described in more detail in Chapter 15, Terrestrial Broadcast Networks.

10.3.3 Phase Modulation

Phase modulation (PM) works in such a way that information is encoded through a variation of the phase of the carrier wave.

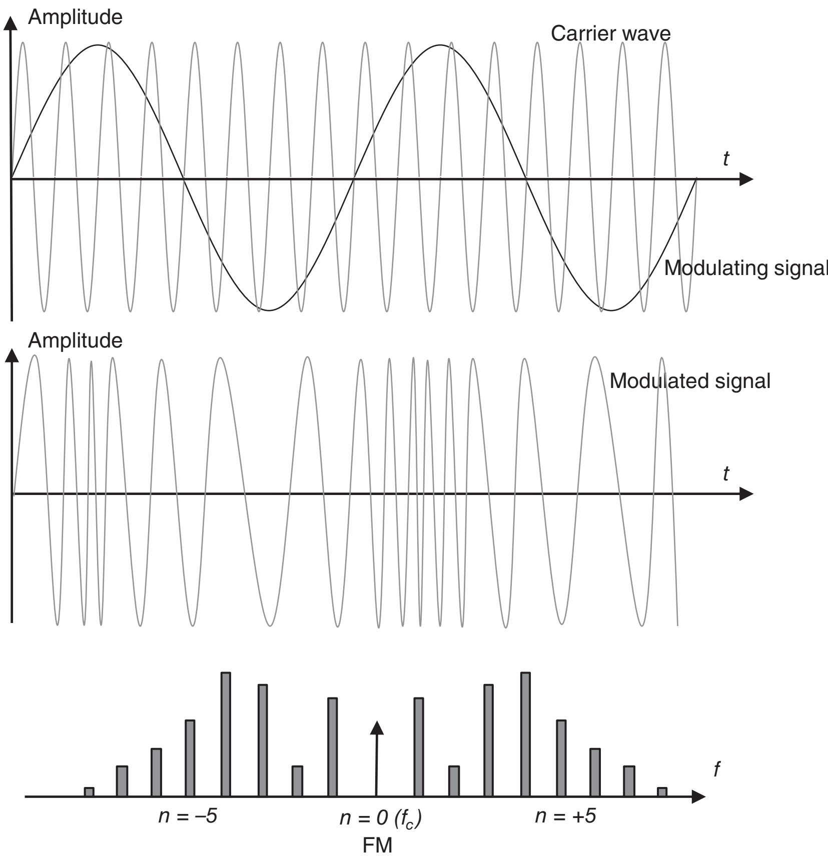

Figure 10.3 *The amplitude is kept constant in Frequency Modulation (FM).*

If the original signal as a function of time t, i.e. the message, is $m(t)$, and the carrier $c(t)$ onto which $m(t)$ is to be modulated is $c(t) = A_c\sin(\omega_c t + \phi_c)$, where A_c is the amplitude of carrier, ω_c is the carrier frequency ($2\pi f_c$), and ϕ_c is the phase shift, the modulated signal will be $y(t) = A_c\sin(\omega_c t+\phi_c + m(t))$. The effect of $m(t)$ modulating the phase can now be seen in this format. It should be noted that in addition to the phase shift, also the frequency of the carrier signal is inherently changing. For this reason, phase modulation can actually be seen as a special case of FM where the carrier frequency modulation is a result of the phase modulation's time derivative.

PM is used in many digital transmission coding schemes, e.g., in GSM and Wi-Fi via its digital correspondence, Phase Shift Keying (PSK), which is explained in Section 10.4.3.

10.4 Digital Modulation and Demodulation

Nowadays, most information is stored and processed digitally, and hence in this section we will now discuss how digital information can be modulated onto an analog carrier signal. In fact, even in cases where the source information is available in analog form (e.g. an audio signal), it is in most cases beneficial to sample and quantize this information into a digital representation and then apply digital modulation for the reason that

- the digitization process can be designed such as to minimize any particular distortion metric (for instance such that it is for the human ear barely hearable that the signal was digitized), and
- the modulation and transmission can be designed sich that at the receiver side the digital representation of the signal can be fully reconstructed – meaning that one can achieve an exactly defined level of quality of reconstructed signal at the receiver side.

In the sequel, the term symbol refers to the time period in which a group of data bits is mapped in any way to an analog pulse shape and modulated onto the carrier. Consequently, the symbol rate refers to the frequency at which groups of data bits are modulated onto the carrier. Assuming that no coding is applied, that is, source data bits are directly modulated, the overall data rate can hence be calculated as the symbol rate times the number of bits connected to each symbol. Note that most practical communication systems use coding which means that redundancy is purposely introduced into the source data sequence before modulation takes place, such that the overall useable data rate is decreased, but the robustness of transmission is increased. By doing so, the transmission can be adjusted to the capacity of the channel, meaning the amount of information that can be transmitted over the channel with an arbitrarily low probability of error.

10.4.1 Amplitude Shift Keying (ASK)

The most straightforward way to embed digital information in an analog carrier signal is to turn the carrier signal on and off according to whether a certain data bit to be represented is 1 or 0. In this case, the period in between two potential switching points would correspond to one symbol carrying exactly one bit of information. This is known as *on-off-keying* (*OOK*), and an example for a bit sequence modulated via OOK is given in Figure 10.4.

An alternative approach is to not map a zero data bit to a carrier signal which is turned off, but rather to a carrier signal with a negative amplitude. This is commonly referred to as *amplitude shift keying* (*ASK*), or in this particular case *2-ASK*, as two possible amplitude values exist per symbol. An exemplary modulated signal is again shown in Figure 10.4.

Under certain amplitude choices, 2-ASK is actually the same as OOK, except that all potential amplitudes are shifted by the same offset. More precisely, OOK corresponds to a 2-ASK modulated signal, but with a constant DC signal in the baseband added to it. This fact already reveals the main downside of OOK, namely

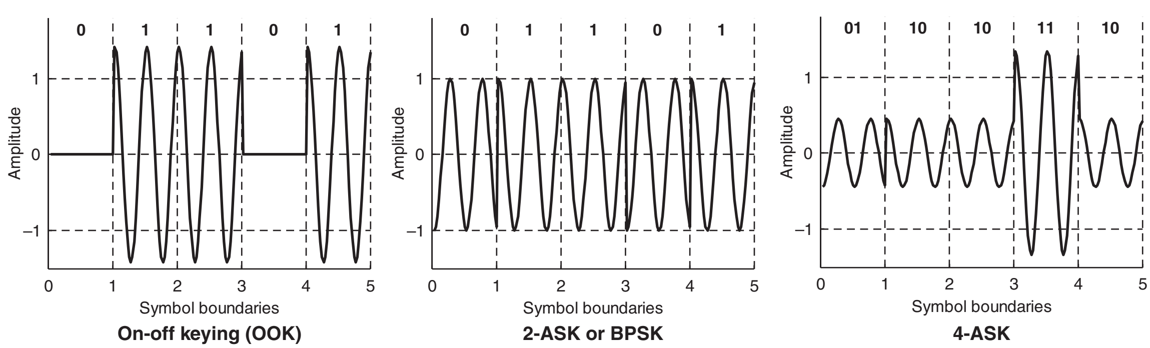

Figure 10.4 *Exemplary signals modulated with OOK, 2-ASK or 4-ASK.*

that part of the transmitted power does not contain any information at all. This part of the power may be used for synchronization purposes, but if we have other means to perform synchronization, this portion of power is basically wasted. Another disadvantage of OOK is that the instantaneous transmit power fluctuates between zero and full transmit power, while it is desirable from a hardware implementation point of view to have a rather constant transmit power. Note that 2-ASK is equivalent to a scheme called *binary phase shift keying* (*BPSK*) which will be explained in the next section.

Of course it is possible to embed more data bits in each symbol by using a higher number of amplitude values. An example of 4-ASK utilizing 4 discrete amplitude values is hence also given in Figure 10.4.

10.4.2 Phase Shift Keying (PSK)

Another way to modulate digital data onto an analog carrier is to apply data-dependent phase shifts to the carrier signal, commonly referred to as *phase shift keying* (*PSK)*. In fact, we have already observed some examples of such a modulation technique in the last section when applying a negative amplitude to the carrier, for instance in the case of 2-ASK, as this is clearly the same as introducing a 180 degrees phase shift to the carrier. For this reason, as mentioned before, 2-ASK is equivalent to the most basic form of PSK, namely *binary phase shift keying* (*BPSK*). In this section, we will now see how finer phase shifts can be applied to embed more information bits per symbol.

A popular PSK scheme is *quaternary phase shift keying* (*QPSK*), where four different possible phase shifts are applied to the carrier signal, hence representing 2 bits per symbol. An example for a modulated signal is given in Figure 10.5. One issue which equally applies to QPSK and the ASK schemes considered before is that from one symbol to the next the phase may be shifted by 180 degrees, which may lead to large amplitude fluctuations connected to the low-pass filtering used in most practical receiver implementations. For this reason, different variants of QPSK have been introduced to ensure that the maximum phase shift between two successive symbols is reduced.

One such variant is *π/4-QPSK*, where different possible phase shifts are used for symbols with even or odd time indices. For instance, if in every odd time instance the possible phase shifts are {0°, 90°, 180° and 270°}, the possible phase shifts in every even time instance are {45°, 135°, 225° and 315°}. This means that

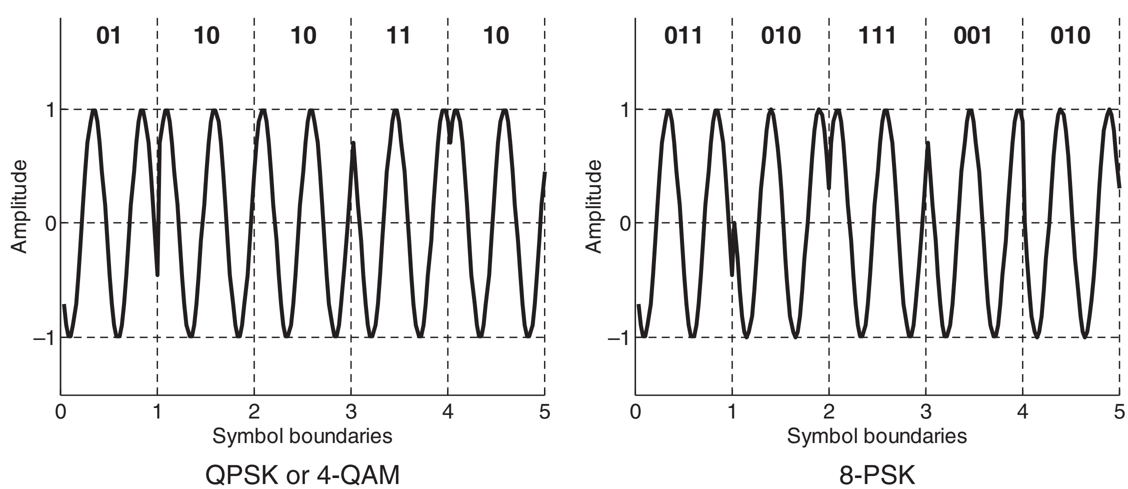

Figure 10.5 *Exemplary signals modulated with QPSK and 8-PSK.*

the maximum phase shift that can occur between two successive symbols is 135 degrees. As this scheme also offers benefits in terms of ease of implementation, it is widely accepted and used in various communications standards, for example in DECT.

Another variant further reducing the potential phase shift between successive symbols is offset QPSK (OQPSK). Here, the 2-bit value to be modulated in each symbol is not applied in the form of a completely new phase shift at the beginning of each symbol, but instead only the first bit is considered for a potential change of phase at the beginning of each symbol, and the second bit is applied in the middle of each symbol. Hence, the phase is effectively changed twice as often as in ordinary QPSK, but through a smart association of bits to phases it can be guaranteed that the maximum phase jump at each point in time is limited to 90 degrees.

Clearly, one can also use more information bits per symbol by introducing finer phase steps. For instance, 8-PSK uses 8 different potential phase values in steps of 45 degrees, meaning that 3 information bits are represented by each symbol. While PSK in general has the benefit that the signal amplitude and power stays constant over time, we will see later that higher order PSK variants such as 8-PSK or 16-PSK and so on are in fact inefficient in terms of the amount of information that can be embedded in one symbol and still be reliably detected at the receiver side.

10.4.3 Combinations of ASK and PSK

As we have seen before, there are some parallels between amplitude and phase shift keying, and it is also possible and beneficial to combine both forms of modulation. A popular form of illustrating any combination of ASK and PSK schemes is to use so-called polar coordinate plots indicating the possible amplitudes and phases that can be applied to each symbol. This is shown in Figure 10.6. For pure ASK schemes, for example, we can see that the information is embedded only in the amplitude of the signal, whereas for pure PSK

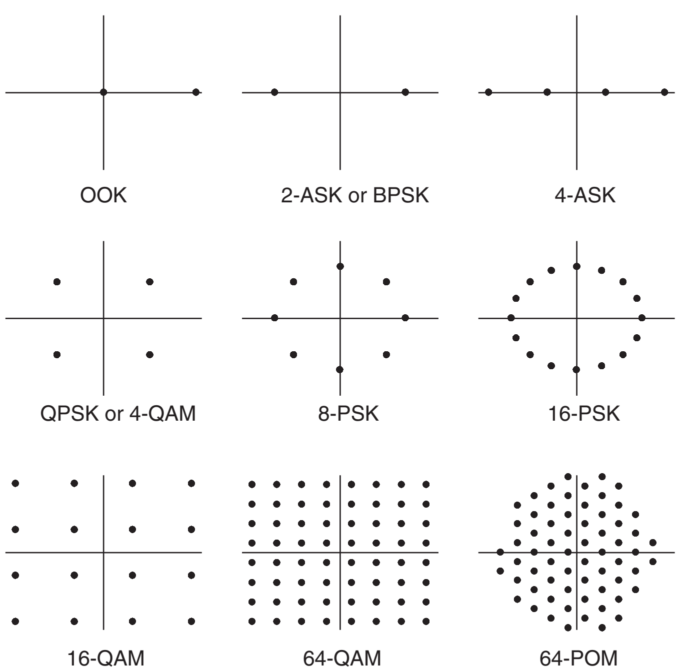

Figure 10.6 *Polar coordinate diagrams for different amplitude and phase modulation schemes.*

schemes the amplitude is constant, and hence all possible constellation points lie on a circle and differ only by phase.

A popular combination of ASK and PSK is so-called quarternary amplitude modulation (QAM), where possible constellation points are arranged in a polar coordinate plot in the form of a block. In 4-QAM, 2 bits are represented in the form of a square of 4 possible constellation points (it is in fact equivalent to QPSK), in 16-QAM, 4 bits are represented in the form of 16 possible constellation points in a 4-by-4 grid. The benefit of these schemes is that the receiver can be implemented rather easily, as linear decision boundaries can basically be drawn through the constellation diagram. Other known variants are *circle optimized modulation* (*COM*) and *power optimized modulation* (*POM*), where in the latter variant the design criterion is to fit all constellation points at equal minimum neighbor distance into the diagram such that the average transmit power per symbol is minimized.

10.4.4 Frequency Shift Keying (FSK)

In this form of digital modulation, the information bits are represented in a change of frequency of the carrier, as illustrated for the case of 2-FSK, also known as *binary frequency shift keying* (*BFSK*), for an exemplary signal in Figure 10.7. For any arbitrary number of different carrier frequencies between which the signal can be switched, one uses the general term multiple frequency shift keying (MFSK).

While in the case of ASK or PSK phase shifts in the modulated signal were not fully avoidable, it is rather straightforward in the context of FSK to let the phase at the beginning of each symbol start with the phase value from the end of the last symbol, a technique which is commonly referred to as *continuous phase frequency shift keying* (*CPFSK*). The effect of this is shown on the right side of Figure 10.7, where we can see that phase jumps are now completely avoided.

An important parameter in the context of FSK is the so-called *modulation index*, which expresses the ratio between the maximum deviation of instantaneous frequency from the carrier frequency and the symbol rate. The minimum modulation index at which the possible output signals resulting from different information bits are still orthogonal to each other (i.e. fully distinguishable by a correlation receiver) is 0.25. A popular scheme

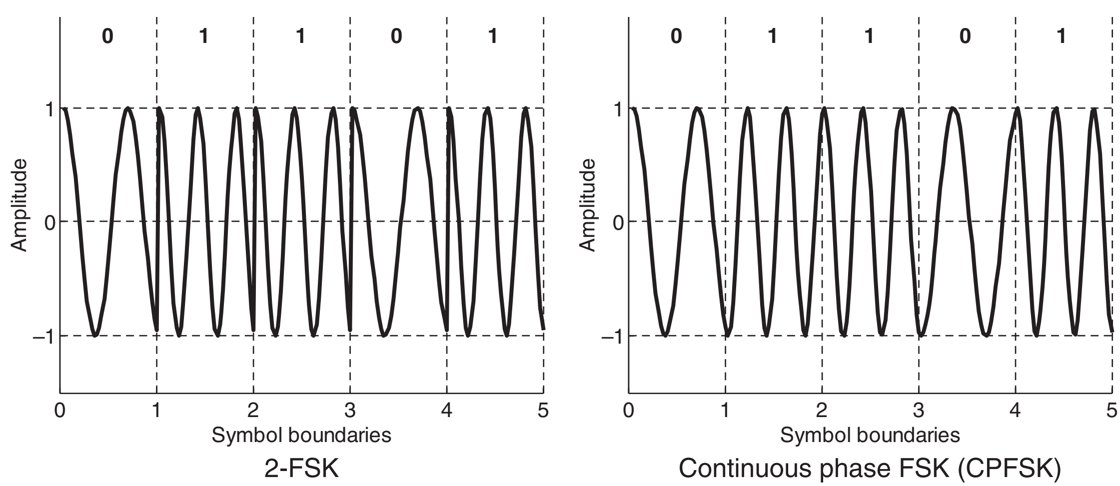

Figure 10.7 *Exemplary signals modulated with FSK or CPFSK.*

which exploits exactly this minimum modulation index is *minimum shift keying* (*MSK*) [1]. In this particular variant of CPFSK, 2 information bits are mapped to each symbol in a similar way as in OQPSK, meaning that one bit is applied at the beginning of each symbol, and the other bit in the middle of the symbol. A variant of MSK called Gaussian Minimum Shift Keying (GMSK) is known for its high bandwidth efficiency and for instance used in GSM, and will be explained later in more detail.

10.4.5 Modulation from a Mathematical Perspective

A different perspective on the modulation schemes presented in the last sections can be obtained by describing any modulated signal generically as

$$s(t) = A(t) \cdot \cos(2\pi f_c t + \phi(t)), \tag{10.1}$$

which basically describes a sinusoidal signal of frequency f_c to which a time-variant phase shift $\phi(t)$ is applied, and which is subject to a time-variant amplitude $A(t)$.

In the case of pure ASK, we apply a different amplitude A_k in each symbol by setting $A(t) = A_k$, but do not apply any phase offset, yielding

$$s(t) = A_k \cdot \cos(2\pi f_c t). \tag{10.2}$$

In the case of pure PSK, we set the amplitude $A(t) = 1$ to a constant value and apply a different phase offset in each symbol k, yielding

$$s(t) = \cos(2\pi f_c t + \phi_k), \tag{10.3}$$

Any combination of ASK and PSK can be expressed by restating the transmitted signal as

$$s(t) = I_k \cos(2\pi f_c t) - Q_k \sin(2\pi f_c t), \tag{10.4}$$

where I_k and Q_k are the real-valued and imaginary components of the term used for the joint phase and amplitude modulation of each symbol.

Finally, FSK can be represented by having an information-symbol dependent and linear increase of the phase offset over time, that is,

$$s(t) = \cos(2\pi f_c t + \omega_k t). \tag{10.5}$$

10.4.6 Pulse Shaping and Power Spectral Density of Modulated Signals

In the previous sections, we have assumed for the purpose of simplicity that a hard switching of amplitudes, phases or frequencies is performed between successively modulated symbols. In practice, this is not feasible, as the transmitted signals would otherwise occupy an infinite bandwidth. In this section, we will discuss how pulse shaping is used in practical systems to ensure that the transmitted signals are bandwidth-efficient while still enabling successful demodulation at the receiver side.

Any of the previously discussed digital modulation schemes can be expressed as a timewise multiplication of a complex-valued baseband signal with the carrier signal. In frequency domain, that is, when looking at which portions of spectrum are occupied by the signals, this mathematical operation corresponds to a so-called convolution, meaning in our case that the spectrum allocation of the complex baseband signal is simply shifted to be centered around the carrier frequency. This means that the actually transmitted signal can only be bandwidth-efficient if the complex baseband signal is also bandwidth-efficient.

There is a fundamental property in signal theory that any signal can only be strictly constrained in either time or frequency, but not in both dimensions simultaneously. More precisely:

- a rectangular pulse shape, where each pulse is strictly constrained in time domain, requires an infinite spectrum in frequency, while on the other hand
- a signal which is fully constrained in frequency is infinitely spread out over time, meaning that signals connected to different symbols overlap and potentially cause demodulation errors.

When designing a communication system, we hence have to find a trade-off such that the transmitted signals are moderately well constrained in frequency (such that we occupy as little spectrum as possible) and at the same time the signals connected to each modulated symbol are only spread out in time to an extent that still enables reliable demodulation. Regarding the latter aspect, one often refers to the two so-called *Nyquist criteria* [2] that should be fulfilled as well as possible to enable reliable demodulation:

- The *first Nyquist criterion* states that if a receiver samples a modulated baseband signal in the middle of each symbol, this sample should not be affected by the transmission of any preceding or succeeding symbols.
- The *second Nyquist criterion* states that if a receiver samples a modulated baseband signal on the edge between two symbols, then the obtained value should be the average of the two symbol values and not otherwise distorted by the transmission of preceding or successive symbols.

Note that the Nyquist criteria refer to an idealistic transmission over a frequency-flat channel, that is, a channel which does not introduce any scattering. Clearly, if the channel does introduce scattering and the receiver obtains multiple copies of the original signal which are time-shifted to an extent larger than the symbol length, both Nyquist criteria could never be met, regardless of any choice of pulse shape.

The extent to which the Nyquist criteria are fulfilled is often illustrated in the form of so-called *eye diagrams*, as shown for an exemplary pulse shape and BPSK modulation in Figure 10.8. The diagram shows all possible baseband signal realizations for the period of one modulated symbol which can be caused by any possible sequence of symbols. We can see that in this example the first Nyquist criterion is met perfectly, as the baseband signal in the middle of the symbol can only take the value 1 or -1 depending on the data bit associated to the symbol of interest, regardless of the values of preceding or succeeding symbols. The second Nyquist criterion, however, is only met moderately well, as we can see that the baseband signal at the symbol edges can take various different values depending on other symbols, and not only the values 1, 0 and -1 that could result as the average of two adjacent symbols. Simply speaking, the robustness of demodulation depends on how wide the "eye" is opened, that is, how large the white area in the middle is.

It is essential that the first Nyquist criterion is strictly fulfilled, as otherwise even the transmission over a perfect channel and under perfect receiver conditions would be subject to intersymbol interference, which is typically not desired. The second Nyquist criterion need not be met perfectly (and it is typically not met perfectly in any practical communication system), but the better it is fulfilled, the more robust the communication system is to time synchronization errors between transmitter and receiver.

A pulse shape used in various communications systems such as UMTS is generated via a so-called *raised cosine filter* [3]. The definition of the exact pulse shape resulting from this filter contains a parameter which allows adjusting the trade-off between improving the bandwidth-efficiency of the transmitted signal and improving the decoding performance. For this so-called *roll-off factor* which can take values between 0 and 1, a low value means that bandwidth is used most efficiently, while a value of 1 means that the signal occupies a wider spectrum, but then both the first and second Nyquist criterion are strictly met. In practice, a rolloff-factor smaller than 0.5 is typically used, such as for example 0.22 in UMTS. Figure 10.9 illustrates different pulse

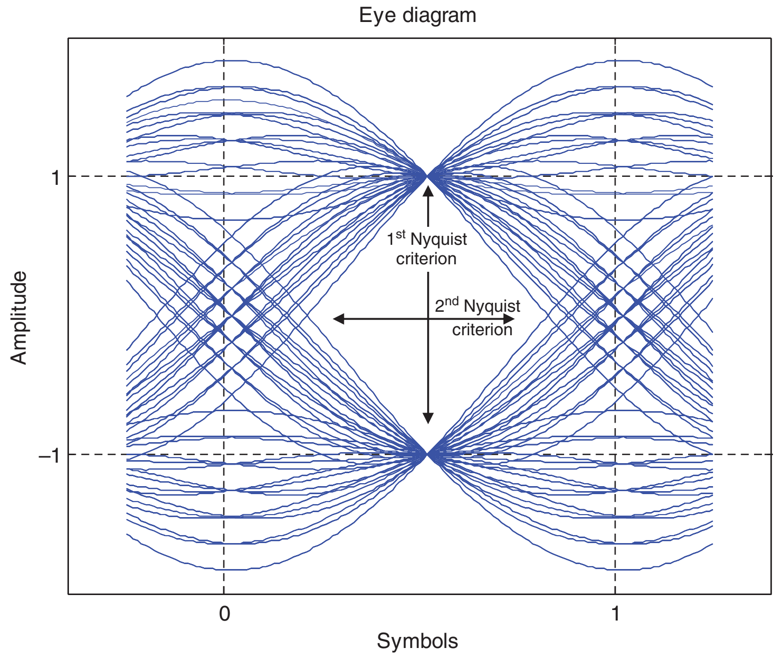

Figure 10.8 *Eye diagram illustrating to which extent the first and second Nyquist criteria are fulfilled by a given pulse shape (in this example based on a raised cosine filter with rolloff factor of 0.22).*

shapes and resulting power spectral densities for a communication system using (for illustration purposes only) a symbol rate of 1 kHz, BPSK modulation, and a carrier frequency of 5 kHz.

While the raised cosine filter is used in most communication systems, also other pulse shapes are used. For instance, GSM and DECT use Gaussian Minimum Shift Keying (GMSK), which is based on FSK, and using a Gaussian pulse shape.

10.4.7 Typical Transmitter- and Receiver-Side Signal Processing

Having understood how to generate a modulated signal in principle, we will now look at how this is typically realized in the form of the signal processing chain at the transmitter side, and how a receiver can be built that is able to successfully demodulate the received signal again. The example depicted in Figure 10.10 would be suitable for any form of digital amplitude or phase modulation or combination thereof (for instance 8-PSK or 16-QAM). This can be expressed as a separate modulation of the in-phase and quadrature phase of the baseband signal.

At the transmitter side, the data bits to be transmitted are typically encoded and then mapped to in-phase and quadrature phase modulation symbols, as we have learnt in the previous sections. These symbols, each representing one or multiple data bits, are still discrete in time. They are made time continuous by applying a pulse-shaping filter, such as the raised cosine filter. The time continuous baseband signals are then upconverted to the desired carrier frequency by being multiplied with the carrier signal and a copy of this which is phase-shifted by 90 degrees, respectively. Finally, the two signals are added, sent through a bandpass to ensure that any aliases possibly introduced through the upconversion are removed, power amplified and transmitted.

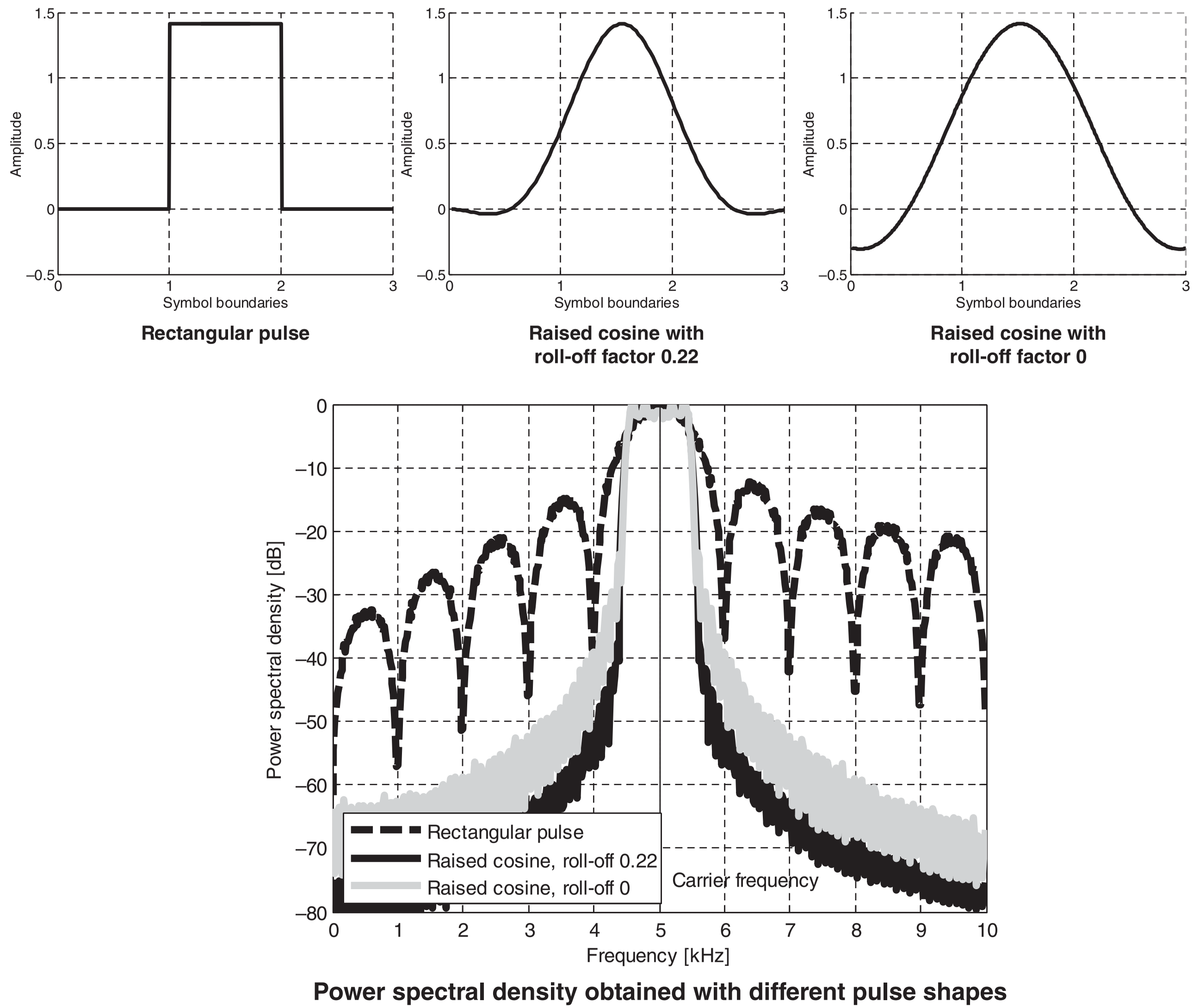

Figure 10.9 Pulse shapes and resulting power spectral densities for rectangular pulses or a raised cosine filter with rolloff-factor 0.22 or 0, for a system with 1 kHz symbol rate, BPSK modulation and 5 kHz carrier frequency.

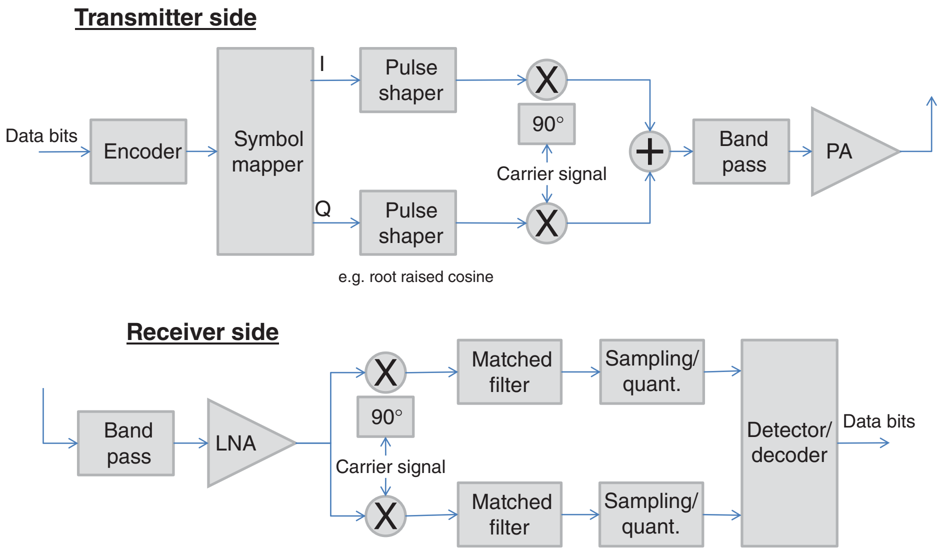

Figure 10.10 *Typical transmitter and receiver side signal processing chains, here shown for the example of any digital phase and amplitude modulation scheme.*

At the receiver side, the antenna signal is typically filtered via a bandbass-filter and amplified with a so-called *low-noise amplifier* (*LNA*). Then, the signal is down-converted from being centered around the carrier frequency to a baseband signal, simply by a multiplication with a reference carrier signal and subsequent low-pass filtering. Analog to the transmitter side, we can obtain the in-phase and quadrature phase components of the baseband signal by separately multiplying the received signal with a reference carrier signal and a 90 degrees phase shifted copy of this. Clearly, the reference carrier signal has to be synchronized precisely in both frequency and phase to the carrier inherent in the received signal, as otherwise the in-phase and quadrature phase components of the baseband signal would experience cross-coupling. This synchronization is realized differently in various communication standards.

After down-conversion, we have hence obtained the imphase and quadrature components of the modulated and pulse-shaped baseband signal, but distorted through the transmission over the channel, and with additional thermal noise and possibly also interference from other transmissions. The question is now how the receiver can best process these signals such that demodulation is as robust as possible, which is equivalent to maximizing the *signal-to-interference and noise ratio* (SINR).

It is easy to derive mathematically that the SINR is maximized if a filter is applied that is matched to the filter used at the transmitter side for pulse shaping, but taking into consideration also the effect the channel had on the signals. More precisely, the channel transfer function of the receive filter should correspond to the (complex conjugate of the) channel transfer function describing both the pulse shaping filter at the transmitter and the distortion introduced by the channel.

In practical communication systems, it is typically desired to have a fixed receive filter, in particular if this is implemented via analog RF circuitry, hence it is not possible to adjust the receive filter to the channel characteristics. In this case, one chooses a receive filter that is at least matched to the pulse shape filter at the transmitter side, assuming that this is standardized and hence exactly known to the receiver. The approach of using a matched filter is in fact pursued in various communication standards, such as UMTS, where a *root raised cosine filter* is used at both transmitter and receiver side, so that the overall filter characteristics from transmitter to receiver then correspond to that of the raised cosine filter considered before, and the receiver SINR is maximized at least in the case where the channel does not introduce frequency-selective distortion.

After filtering, the signal is then sampled at discrete time intervals representing the centers of the modulated symbols, and quantized to discrete amplitude values. Note that sampling always requires that the input signal is previously low-pass-filtered such as to suppress any frequency components larger than half the sampling frequency. If not, so-called *aliasing* would occur, meaning that any signal at frequencies beyond half the sampling frequency would appear as an interference artifact in the frequency range of interest. In most receiver architectures, the receive filter will already remove higher frequency components, so that no additional low-pass filter is needed before sampling.

A common implementation variant where receive filter and sampling/quantization units are integrated into one component is based on a *correlator*. In this case, the baseband signals in the in-phase and quadrature phase branches are multiplied with a signal representing the transmitter pulse shape and integrated over time, for instance via a simple capacitor. At the end of each symbol, the charging value of the capacitor is then quantized, and the capacitor is uncharged again in preparation for the next symbol. Mathematically, this is equivalent to the matched filter approach stated before, and constitutes the best possible receiver given that the thermal noise and interference the transmission is subject to is Gaussian and spectrally white (i.e. having a flat spectral density).

Finally, the time-discrete in-phase and quadrature phase samples are fed into a detector and possibly decoder to determine which bit constellation (or bit sequence in the case of a coded transmission) has been transmitted at highest probability. The exact detection and decoding strategies employed depend on various transmission characteristics, such as the symbol rate, the delay spread of the channel, or the usage of, for example, multiple antennas.

10.4.8 Digital Modulation Schemes Used in Practical Systems

In this section, and in the form of Table 10.1, we briefly summarize the different digital modulation schemes introduced before, state their key advantages and disadvantages, and mention in which communication systems these are used.

Table 10.1 starts with pure amplitude shift keying (ASK) schemes, which have the main disadvantage that one dimension of the IQ baseband signal is completely ignored, hence half the potential capacity of the transmission is simply wasted. One particular variant of ASK is on-off-keying (OOK), which has the additional disadvantage that part of the transmitted power is invested into a carrier frequency signal which bears no information. ASK schemes are for instance used in optical communications or cable modems.

Phase shift keying (PSK) schemes utilize both the in-phase and quadrature phase components of the baseband signal and embed digital information in the phase of the carrier signal. These schemes have the benefit that all modulated symbols have the same amplitude, meaning that the envelope of the modulated signal stays constant.

Combinations of ASK and PSK such as quarternary amplitude modulation (QAM) or power optimized modulation (POW) give up the property of a constant signal envelope, but have the benefit of better spacing potentially transmitted symbols in signal space (see the constellation diagrams for these schemes in Figure 10.6). This means that for the same average transmit power and same number of bits per symbol, QAM and POM schemes provide a lower probability of symbol error than PSK schemes. QAM schemes are nowadays the most popular digital modulation techniques and are employed in a wide range of communications systems such as DVB, HSPA and LTE. While POM schemes provide a further improved performance in terms of robustness for a given transmit power, they are less popular due to their irregular constellation diagrams and consequently more complex receivers.

Table 10.1 Comparison of different digital modulation schemes and their usage in practical communications systems

Modulation scheme	Bits/symbol	Advantages/disadvantages/ comments	Examples of utilization
Amplitude shift keying (ASK)		Disadvantage: Essentially inefficient, as number of bits per symbol can be doubled through QPSK at same bandwidth and same robustness	
OOK	1	Disadvantage: Part of transmit power wasted into basedband DC signal compontent (but may be used for sync)	Optical communication systems (e.g. IrDA).
2-ASK or BPSK	1		Cable modems, space exploring equipment
Phase shift keying (PSK)		Advantage: Constant envelope Disadvantages: Phase jumps between symbols cannot be fully avoided but alleviated through OQPSK or π/4-QPSK. For higher-order modulation, robustness lower than QAM.	
QPSK or 4-PSK, OQPSK, π/4-QPSK	2		Satellites, cable modems, CDMA, TETRA, IS-54, HSPA, LTE, LTE-A
8-PSK	3		E-GPRS, aircraft equipment
Combinations of ASK and PSK		Disadvantage: No constant envelope, potential phase jumps between symbols	
16-QAM	4	Advantage: Regular constellation diagrams facilitating simple receivers	DVB-T, DVB-C, cable modems, HSPA, LTE, LTE-A
32-QAM	5		DVB-T, micro wave links
64-QAM	6		DVB-C, modems, HSDPA, LTE, LTE-A
256-QAM	8		DVB-C, digital video (USA), modems
64-POM	6	Advantages: Highest robustness of all schemes compared here Disadvantage: Irregular constellation diagram requires more complex receiver	
Frequency shift keying (FSK)		Advantages: Constant envelope, phase jumps can be avoided with CPFSK Disadvantages: Less robust than for example, QPSK. Disadvantage: Deteriorating spectral efficiency as M increases.	
FSK, CPFSK, MSK	1		GSM, CDPD, DECT
M-FSK	$\log_2 M$		

Frequency shift keying (FSK) approaches embed digital information in the carrier signal by slightly changing the carrier frequency. FSK approaches are widely used in, for example, older wireless communications standards (e.g. GSM or DECT) for reasons of simple transmitter and receiver implementation and also good spectral efficiency for low modulation orders. FSK, however, becomes very spectrally inefficient and unrobust for higher modulation orders (i.e. a higher number of modulated bits per symbol), and has hence become rather irrelevant in latest communications standards.

10.4.9 Multiplexing, Multiple Access and Duplexing

In the past sections, we have discussed how single streams of digital information can be modulated onto analog carrier signals. In practical communication systems, we typically face the challenge that multiple streams of information have to be *multiplexed* into one single chunk of bandwidth. When these different streams of information originate from multiple devices or are targeted to multiple devices, one typically talks about *multiple access* solutions. As we will see later, multiple access solutions are always strongly related to a particular multiplexing concept, but they typically inherit additional challenges due to the fact that either the transmitter or receiver side consists of spatially distributed devices. Finally, the term *duplexing* refers to the challenge of using a certain chunk of spectrum for bidirectional communication, as for instance for the uplink and downlink in mobile communications systems.

10.4.9.1 Multiplexing and Multiple Access

Regarding *multiplexing* and *multiple access solutions*, we in principle have the following options:

Time division multiplex (TDM). In this most intuitive approach, multiple transmissions are multiplexed in time, that is, they are handled sequentially, such that only one transmission is taking place at a time. The corresponding multiple access scheme, for example in a cellular context, is called **time division multiple access (TDMA)**. In the uplink, this clearly requires that all devices are precisely synchronized, such that their transmitted signals are aligned well in time at the receiver side and inter-user interference is avoided. TDMA is for instance used in the Global Standard for Mobile Communications (GSM), where each available frequency carrier is divided into 8 recurring time slots.

Code division multiplex (CDM). In this approach, the signals connected to multiple transmissions are multiplied with spreading sequences before pulse shaping and modulation. By doing so, multiple transmissions are spread over a larger spectrum and superimposed in time and frequency, but can be despread and separated again at the receiver side if the used spreading sequences have low cross-correlation. The corresponding multiple access scheme, **code division multiple access (CDMA)**, is for instance used in the third generation of cellular systems. While the key challenge in the TDMA uplink is synchronization, the CDMA uplink mainly suffers from the so-called *near-far problem* [4], which arises if multiple devices are received at very different power by a base station. In this case, the residual inter-user interference after despreading will still be large, despite the low cross-correlation properties of the codes. Wideband CDMA (WCDMA) and HSUPA systems alleviate this issue by applying a very fast power control in the uplink that ensures that all devices are received at similar power. More details on this can be found in Section 12.2.

Frequency division multiplex (FDM). A further option is to multiplex transmissions by letting them occupy different subparts of the available spectrum. The corresponding multiple access scheme, **frequency division multiple access (FDMA)** is for instance used in satellite communication, where each user is assigned one or multiple frequency bands or channels. A challenge inherent in both uplink and downlink transmission is the potential crosstalk between transmissions on adjacent frequencies. A special variant of FDM is *orthogonal frequency division multiplex* (*OFDM*) [5], where a particular signal processing at transmitter and receiver side is used to generate a large number of narrowband and orthogonal sub-carriers that can be used for communication connected to individual users. This scheme, which is used in LTE, will be explained in detail in the next section.

Further means to reuse the same spectrum for multiple transmissions are **spatial division multiplex (SDM)**, where multiple streams are multiplexed based on the spatial properties of a transmission channel spanned between multiple antennas at transmitter and receiver side, or **polarization division multiplex (PDM)**, where different polarization angles are exploited over which orthogonal transmissions can take place. These two techniques are typically used in conjunction with any of the multiplexing and multiple access schemes stated above, and will be treated in more detail in various later chapters of this book.

10.4.9.2 Duplexing

For *duplexing*, that is, for transmitting bidirectionally over the same chunk of spectrum, one typically considers the following two variants:

Time division duplex (**TDD**). Here, a certain amount of time is reserved for one link direction, and the remaining time for the other. TDD has the downside that it requires very precise synchronization between all communicating entities in a system and/or a sufficiently dimensioned guard interval between the two link directions. Otherwise, strong interference will occur between the two forms of transmission, if for instance the downlink transmission from a base station to a device is interfered by a badly synchronized uplink transmission from another device in close proximity to the receiver. On the other hand, it has the benefit that the channel that both transmissions see is reciprocal; this can be beneficial, for example, for spatial division multiplexing schemes where precise channel knowledge is needed at the transmitter side. TDD is for example used in TDD-LTE.

Frequency division duplex (**FDD**). In this more popular variant, the two link directions use different subparts of the available spectrum. The most severe issue here is the potential self-interference a device may cause if part of the transmitted power spills over into the receiver chain. To avoid this, communicating devices must fulfill very sharp spectral masks (i.e. strongly suppress any out-of-band radiation), and one typically uses a large spacing between uplink and downlink carriers (in FDD-LTE, for example, pairs of carriers for uplink and downlink are typically spaced by several tens of MHz).

10.4.10 Orthogonal Frequency Division Multiplex

Orthogonal Frequency Division Multiplex (OFDM) and orthogonal frequency division multiple access (OFDMA) are particular multiplexing and multiple access schemes, respectively, including a particular choice of pulse shape and modulation concept. OFDM(A) is for example used in WiMax and in the downlink of LTE systems, and offers the following benefits:

- It splits a transmission into a large number of narrow-band, orthogonal subcarriers, which can be
  - assigned flexibly to multiple communicating users at typically fine granularity and
  - used to exploit diversity in frequency.
- It makes transmission robust to even highly frequency selective channels.
- It enables easy transmitter-side and receiver-side signal processing even in the context of multiple-antenna transmission.

The core principle of OFDM is to modulate digital information onto a high number of subcarriers in parallel, but in exchange use a rather low symbol frequency. An LTE system using an overall downlink bandwidth of 5 MHz, for instance, uses a baseband sampling frequency of 7.68 MHz. Groups of 512 successive and complex-valued samples are then denoted as one so-called OFDM symbol and correspond to the time-domain equivalent of 512 complex-valued and discrete signals modulated onto (in principle) 512 adjacent subcarriers with a spacing of 15 kHz.

In practice, only 300 of these subcarriers are used, in order to ensure that the overall transmitted signal is constrained well within the 5 MHz system bandwidth. Also, the central subcarrier is avoided to ensure that the resulting baseband signal contains no DC component. Consecutive OFDM symbols, each having a length of 512/7.68 MHz = 67 us each, are additionally spaced by roughly 37 samples or 4.6us, leading to an overall OFDM symbol rate of 14 kHz. The space between symbols contains a so-called *cyclic prefix*, which is obtained by taking the last 37 samples of each OFDM symbol and copying this in front of the symbol. The insertion of a cyclic prefix serves two purposes: On one hand, consecutive OFDM symbols are spaced sufficiently to ensure that even in a severe multipath environment any interference between successive OFDM symbols is avoided. On the other hand, it leads to the fact that the convolution of the transmitted sequence of symbols with the channel is circularly symmetric, which is required to ensure that all subcarriers remain fully orthogonal, even after the transmission over a highly frequency-selective channel. Beside the avoidance of intersymbol-interference mentioned before, a core benefit of an OFDM transmission is that each subcarrier can be treated as an individual channel and equalized individually, which strongly simplifies equalization in the context of, for example, multiantenna transmission and reception.

A typical OFDMA transmitter and multiple receiver chains as in the example of an LTE downlink are shown in Figure 10.11. Note that all baseband signals are complex-valued, but are here illustrated in the form of only one signal processing path for brevity, and not in the form of separate in-phase and quadrature phase paths as in Figure 10.11. The following signal processing steps are pursued at the transmitter side for each so-called OFDM symbol:

- **Encoding and modulation**. Groups of information bits connected to multiple users are coded and modulated into multiple discrete, complex-valued signals, for example using 16-QAM.

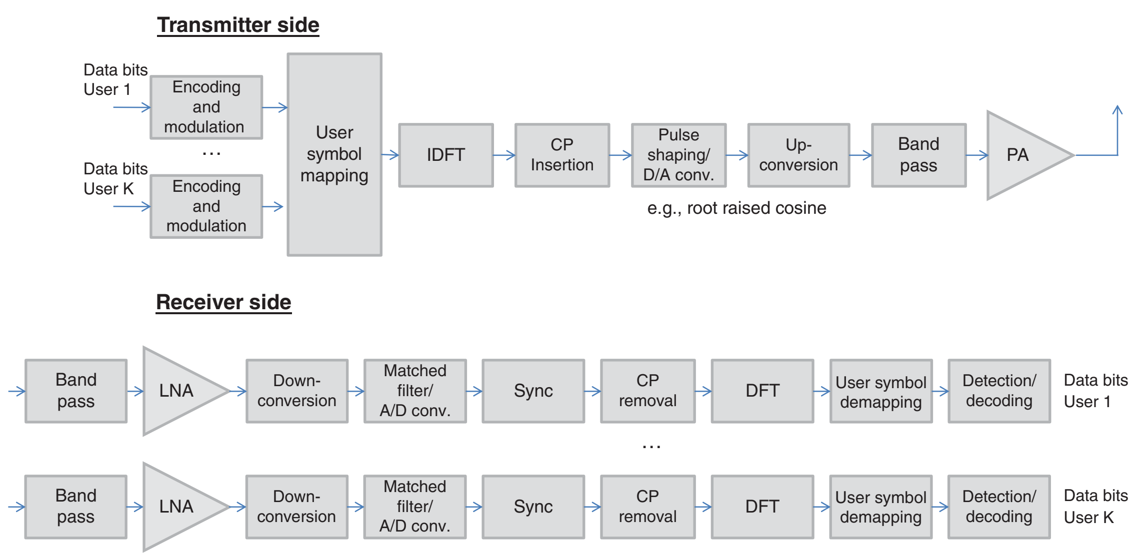

Figure 10.11 OFDMA transmitter and receiver chains, as used in, for example, an LTE downlink.

- **User symbol mapping**. The modulated signals of all users are mapped onto the N available subcarriers such that we have one discrete, complex-valued signal for each subcarrier.
- An **Inverse Discrete Fourier Transform** (IDFT) is applied, mapping N modulated symbols in frequency domain to N consecutive samples in time
- A **cyclic prefix (CP) is inserted**, meaning that a certain number of samples from the end of the series of N samples is copied and inserted before the beginning of the series of samples.
- The series of samples is **D/A-converted** through the application of a pulse shaping filter, **up-converted**, **amplified** and **transmitted**.

At each receiver, the following steps are performed

- The received signal is **filtered by a band-pass**, **amplified**, **down-converted** and **A/D converted** through matched filtering.
- The receiver **synchronizes** itself to the beginning of an OFDM symbol. This can be done easily by looking at the local autocorrelation of the received samples with the same sequence shifted by N samples. A high autocorrelation value indicates that one has found the cyclic prefix before the beginning of an OFDM symbol.
- The cyclic prefix is removed.
- A Discrete Fourier Transform (DFT) is applied to map N consecutive samples in time to N modulated symbols in frequency domain.
- The receiver extracts the modulated symbols from the subcarriers allocated to the user.
- The modulated symbols are equalized, detected and decoded.

Beside the benefits stated before, OFDM bears the following disadvantages:

- As the principle of OFDM relies on the orthogonality of adjacent subcarriers, the scheme is very sensitive to inaccurate synchronization in frequency. More precisely, an OFDM transmission starts being strongly impaired if transmitter and receiver are missynchronized by more than about 3% of the subcarrier spacing, that is, in the case of LTE by more than 450 Hz. Considering that the carrier frequencies used in cellular communication systems are on the order of GHz, this means that the accuracy of oscillators used at transmitter and receiver side must be in the order of less than 0.5 parts per million (ppm).
- The avoidance of intersymbol interference and consequently simple equalization is bought by the insertion of cyclic prefixes, meaning that transmit power and spectral efficiency are wasted in comparison to a system using sophisticated receiver techniques to equalize intersymbol interference instead of avoiding it.
- As an OFDM transmission basically resembles a parallel and uncorrelated transmission of a large number of subchannels, the transmitted signals have a large peak-to-average-power ratio (PAPR), which requires power amplifiers with a large linear range, or a reduction of average transmit power.

As the latter aspect is mainly problematic for the cellular uplink, where expensive RF components at the transmitter side are to be avoided, 3GPP has decided to use so-called **single carrier frequency division multiple access (SC-FDMA)** in the LTE uplink. This is in principle the same as OFDMA, except that modulation takes place in the time domain, and each transmitting device uses a small DFT to map the modulated signals to frequency domain, after which the signals are mapped to the overall number of subcarriers available in the system and the same further procedure is applied as in OFDMA. SC-FDMA offers a significantly improved PAPR as opposed to OFDMA, but has the downside that the effort for signal processing at transmitter and receiver side is slightly increased.

References

1. Pasupathy, S. (1979) *Minimum Shift Keying: A Spectrally Efficient Modulation*, IEEE Communications Magazine.
2. Proakis, J.G. (2005) *Digital Communications*, 3rd edn, McGraw-Hill.
3. Anderson, J.B. (2005) *Digital Transmission Engineering*, 2nd edn, Wiley Interscience.
4. Rappaport, T.S. (2002) *Wireless Communications: Principles and Practice*, 2nd edn, Prentice Hall.
5. Fazel, K. and Kaiser, S. (2008) *Multi-Carrier and Spread Spectrum Systems. From OFDM and MC-CDMA to LTE and WiMAX*, 2nd edn, John Wiley & Sons, Ltd.

11

3GPP Mobile Communications: GSM

Jyrki T. J. Penttinen

11.1 Introduction

This chapter describes the architecture, core and radio systems of GSM. This chapter also includes essential information about the interfaces, elements, modulation, coding, modes, functionality and performance of the system, as well as frequencies, terminals, planning, dimensioning and optimization of radio, transport and core networks. Later, the chapter presents practical use cases. There are also modern functionalities presented to optimize the capacity and interference levels of GSM as one part of the refarming strategies, and information about the advanced data and signaling methods which can be utilized until the end of the useful lifecycle of GSM networks.

11.2 Development of GSM

Mobile communications have become the most important method for exchanging information between persons, with regards to both voice and data. Long before devices utilizing electromagnetic waves were invented people tried to send and receive messages across distances. We have come a long way since the methods of the early days – like drumbeats and smoke signals. However, the actual content of person-to-person communication probably remains largely unchanged.

The communication of data was ignored in the early days of the development of modern mobile communications networks. However, even first generation analog devices such as NMT-phones (of Nordic Mobile Phone system) could be used to transfer data with the help of modems and data adapters. Reflecting the situation based on current solutions, the data rate was not impressive due to technical restrictions, but functioning connections could be set up. Even the transfer of surveillance videos over NMT-networks was trialed in the 1990s [1].

GSM (Global System for Mobile communications) has served the world's mobile markets loyally since the first commercial networks back at the beginning of the 1990s. Even if third generation mobile communications

The Telecommunications Handbook: Engineering Guidelines for Fixed, Mobile and Satellite Systems, First Edition.
Edited by Jyrki T. J. Penttinen.

with its evolution path towards fourth generation are taking over the markets, the number of GSM users still remains high. According to GSMA, the total connections of GSM mobile communications user numbers passed the 3 billion mark globally in 2008. The GSM industry surpassed its first billion connections in 2004, and the second one in 2006 [2]. In 2012, the GSM Association estimated that GSM-based technologies serve 80% of the global mobile market, which means that there have been over 5 billion GSM users in more than 212 countries and territories. It also means that GSM is clearly the most popular cellular system ever used in the globe so far. After the peak, the number is currently declining, but was still at 4.4 billion in March 14 [3].

The latest development of the GSM system provides a considerable capacity extension of voice users. Functionalities like Orthogonal Sub Channel (OSC) and Dynamic Frequency and Channel Allocation (DFCA) make utilization of GSM more attractive than ever before in the long history of the system [4–12]. As the third and fourth generations are in any case more spectral efficient, the utilization of GSM will lower gradually, and users will purchase more advanced devices capable of taking full advantage of the latest technologies and applications. In any case, according to the messages from key operators and the industry, it may be possible that the GSM system will still be up and running at least until 2020, providing basic voice calls and being a growing base for specialized solutions like machine-to-machine (M2M) communications in wide geographical areas.

GSM and other digital second-generation systems were also designed to provide data transfer integrated to the system itself. The utilization of data has been possible via GSM in practice since about 1994, soon after the first phase of GSM had been deployed at the beginning of the decade. In contrast to the principles of earlier analog networks of 1980s, data service functions in a digital form all the way from the terminal to the Mobile services Switching Center (MSC) in second-generation networks. In fact, the definition of second generation mobile systems is the ability to deliver end-to-end digital communications. For example, in the early phase of GSM, circuit switched data connection could be set up by using a mobile handset connected to a PC in much the same way as a connection is set up using a modem and PC in the fixed network. Thus traditional AT commands function in both GSM-phones and fixed line modems as was defined in the ETSI technical GSM specification 07.07. As the GSM system is digital, the user does not need a separate modem for data connections since the modem is located in the interworking functions (IWF) of the MSC element. In case of the traditional circuit switched data call, this modem communicated with a modem in the external network, which made it irrelevant for the user whether the connection was set up using a GSM-handset or fixed line modem. Also fully digital connections from the GSM-network to for example, the ISDN network have been possible since the 1990s.

The original data rate of the GSM system was 9.6 kb/s. This difference in data rates was quite significant at the time compared to ISDN with 64 kb/s or ADSL with data rates of hundreds of kb/s. Nevertheless, also the GSM system has since been developed, resulting in highly improved data rates. The gradual development was first put to the circuit switched data domain, with data compression via a data connection supported by the network. Data compression was also possible on the application level defined by the user. Using compression, data rate is increased "virtually," as the actual transmission does not require faster data rates. GSM specifications have also supported ITU V.42bis compression. Compression could be used both between the terminal and the MSC and the fixed network. Especially text-based files could be compressed to $^1/_4$–$^3/_4$ of the original size if compression was used on both parts of the connection. The actual end-to-end data rate was increased respectively.

The single time slot of the GSM system creates a bottleneck for data rate. The multislot technique has been designed to solve this problem. The multislot technique allows simultaneous use of several time slots by a single user, which increases data rate by the number of time slots in use. HSCSD (High Speed Circuit Switched Data) was the next step in the evolved GSM data services, which utilized both enhanced channel coding and multislot techniques to achieve data rates ranging between 28.8 and 56 kb/s. Timeslots could already be used either symmetrically or asymmetrically in both directions. HSCSD was deployed in practice from 1999 until the packet switched data took over the GSM data service offering.

In theory, HSCSD could have utilized all eight TRX time slots per user. However, networks or terminals did not support the full set due to the complexity of the technique. In addition, a network can normally rarely guarantee simultaneous use of eight timeslots for extended time periods. Thus, the realistic reachable maximum multislot level could be offered via up to 4 + 4 (downlink + uplink) configuration. Despite the practical limitations, HSCSD offered significant improvements to the GSM data rate prior to the transition towards packet switched era.

The circuit switched data had some drawbacks. It typically took anywhere from a few seconds to about 20 seconds including modem handshaking to set up a connection between a GSM network and an external network, whether using a basic or HSCSD modem. Connecting to the ISDN network is faster as the rate adapter units are fully digital. At the time of the development of the data services, the GSM network could support both V.110- and V.120-protocols, of which V.120 was more efficient in transmitting the user's data.

All methods described above represent the circuit switched era, even if a user could have connected directly to an X.25-type of packet data network via a PAD-element (packet assembly/disassembly). In this case the user could set up a circuit switched connection to the MSCs inter working functions, from which point onwards the connection was packet switched. Thus the "traditional" GSM packet connection cannot be called an end-to-end packet switched connection, although the utilization of X.25 was very rare in practice.

GPRS (General Packet Radio Service) is a packet switched data transfer method which has also been specified for GSM. With GPRS, a user no longer dials a separate modem or ISDN-number. GPRS works like the Internet so that terminals and network elements each have their own Internet-addresses. GPRS is an extension of the GSM network, and implementing GPRS does not in principle cause changes to the functioning of the GSM network. GPRS utilizes the same radio interface resources as GSM, but so that circuit switched GSM traffic has a higher priority by default.

The GPRS system can use the multislot technique. Four channel coding classes were specified into the basic GPRS system, according to the Release 97 ETSI specifications. Both the number of adjacent channels and channel coding can change automatically during a connection depending on cell traffic and connection quality according to the same principle as with the HSCSD-technique. As GPRS could, according to the original specifications, use already in theory between 1 and 8 simultaneous timeslots per user and as the lightest channel coding class offers approximately 21.4 kb/s data rate per time slot, the theoretical data rate was approximately 170 kb/s per user in the early days of GPRS. In reality, data rates were typically some tens of kilobits per second with this first phase solution. This is because the equipment did not fully support all optional functionalities included in GPRS specifications. In addition, especially during peak hours, the network may not necessarily be able to continuously provide all resources requested by the user. Even if data rate is not close to the theoretical ones, the connection method itself is more flexible compared to traditional circuit switched GSM data transmission, facilitating connections to bursty Internet-type of networks. The basic form of GPRS was already implemented in many GSM networks from 2000.

GSMs development stages include a new modulation method, 8-PSK, for the radio interface. Better known as EDGE (enhanced data rates for global evolution) the method stands for a new type of modulation, in effect the parallel functioning of the modulation methods of GMSK, the basic GSM modulation method, and the 8-PSK method. The modulation method and channel coding to be used are chosen dynamically depending on the connection quality at the time. EDGE can be used in connection with GPRS, HSCSD, and basic GSM voice services. The combination of GPRS and EDGE is called EGPRS (enhanced GPRS), and the combination of HSCSD and EDGE is known as EHSD (enhanced high speed data). In connection with voice services EDGE is known as EAMR (enhanced adaptive multi rate codec).

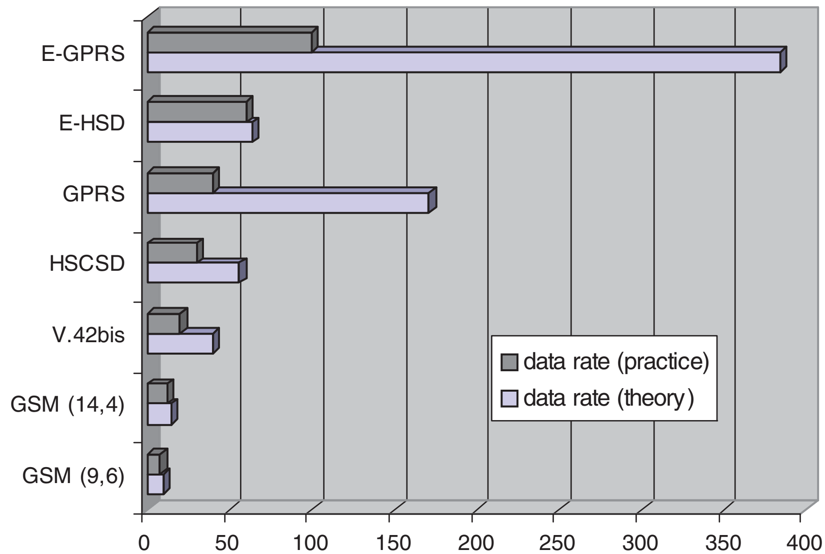

Figure 11.1 *The evolution path of the GSM data transfer methods and the corresponding average downlink data rates in theory and in practice up to E-GPRS. DLDC and EGPRS2 increase the data rate further.*

Even at the beginning of the advanced GSM data services, the data rate of EGPRS would have been able to exceed 400 kb/s using all eight timeslots and the lightest channel coding, but as with basic GPRS, EGPRS also faced the same actual network capacity constraints and equipment limitations. In reality, data rate could be crudely estimated to vary between 50 and 100 kb/s still in 2005. EHDS data could reach approximately 64 kb/s due to the A-interface constraints of the GSM network. However, the 64 kb/s rate could be achieved even with one or two timeslots, in contrast to earlier data services. Using EDGE with GSM voice connections enables for example, stereo and high-quality transmission of music.

Figure 11.1 shows the development path of GSM data transfer methods with the theoretical maximum data rates per user. In addition, the figure includes an estimate of the actual data rates achieved in GSM – or UMTS networks with normal traffic levels.

Even if the specifications include many optional capabilities for example, for channel coding classes and the multislot technique, not all of them could be used in the early phases of the service. The actual data rates achieved varied depending on the planned capacity of the network, traffic peaks, and the terminals in use. If the network and the terminals supported at some later stage all of the eight TDMA timeslots and all the channel coding classes, EGPRS connections could also in practice reach 400 kb/s peak data rate.

The GSM packet data services have then developed further, and the current data rate can be up to about 500 kb/s in the Downlink Dual Carrier (DLDC) – in theory, by utilizing all the 8 timeslots in both carriers, the DL speed can even reach 1 Mb/s. The latest step is EGPRS2 as defined in Release 7 and it provides the possibility to utilize QPSK, 16-QAM, 32-QAM, higher symbol rate and turbo coding in downlink. The data rate is accordingly higher, up to 120 kb/s per timeslot in downlink direction, which can be combined with multislot technique and dual carrier functionality. In this way, the practical downlink data rate may be up to 1 Mb/s, and uplink rate up to 0.5 Mb/s.

The further advances of GSM, as of Release 8, include functional and performance enhancements, but the reasonable and practical data rates seem to saturate into the values described above. Release 8 also includes enhancements for voice calls via MUROS which is basically an enhancement for the OSC/VAMOS concept, doubling the voice capacity compared to the full and half rate codecs whenever the received power level and quality values are high enough. This eases further optimization of GSM in order to pave the way for refarming.

11.3 Specification of GSM

From the mid-1990s ETSIs SMG-groups worked on specifying the GPRS service for the GSM system. Specifications for the first stage GPRS were almost complete by the late 1990s. GPRS affects many GSM specifications, both with regards to the radio interface and the fixed network. Even if GPRS does not in itself affect the functioning of the basic GSM system in any way, many additions have been made to GSM specifications due to GPRS. This is natural as GPRS signaling influences the entire network, for example, the registers in the MSC, the base station controllers, and base stations.

The most important groups which took part in GPRS specification were SMG2 (radio interface specification), SMG3 (fixed GSM network specification), SMG4 (GSM data services), SMG6 (billing aspects), and SMG10 (security issues). Other SMG groups have also participated in the work. For example, SMG1 worked to harmonize the new terms created with GPRS. Figure 11.2 presents an example of ETSI standard format.

Since 2000, the GSM and UMTS specification work was transferred from ETSI to 3GPP (3rd Generation Partnership Project). The basic principle of the specification numbering was changed at the same time yet maintaining the idea for different stages of the specifications. Also, the evolution of the 2G and 3G systems were taken care of in 3GPP, including the HSPA (High Speed Packet Access) and LTE (Long Terms Evolution), as well as the LTE-A (Long Term Evolution Advanced) [13]. Nowadays the best way to find the relevant GSM, UMTS/HSPA and LTE specifications is to enter 3GPP web page www.3gpp.org.

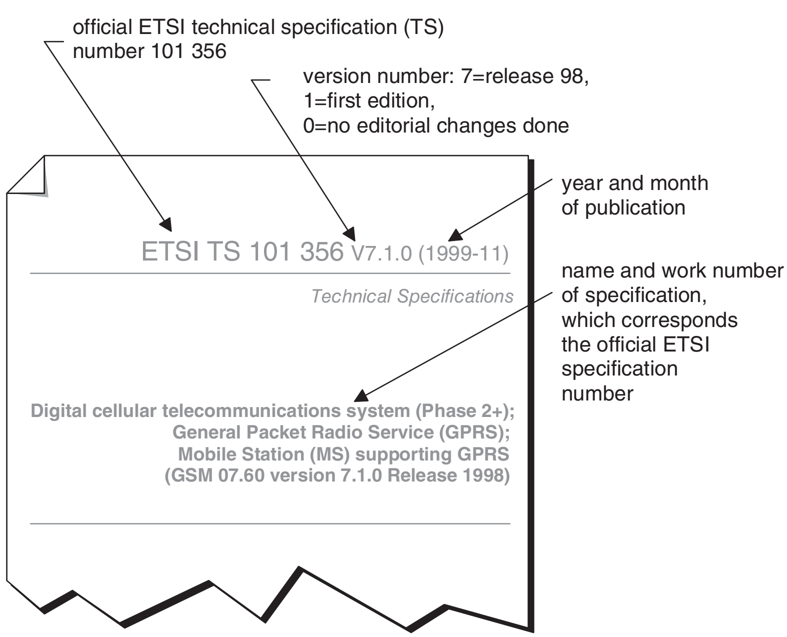

Figure 11.2 *Example of the GPRS specification front page. This is release 98 specification number ETSI GSM TS 07.60, which contains the definitions for the GPRS MS. Layout shown by courtesy of ETSI.*

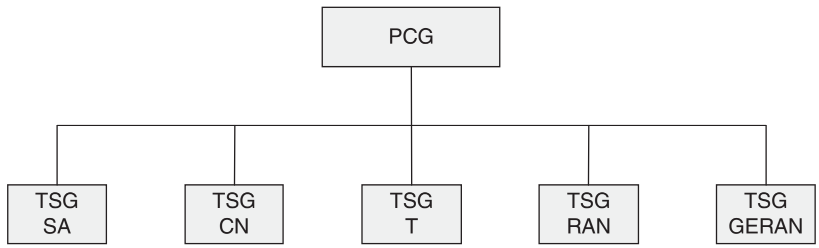

Figure 11.3 *UMTS specification working groups in 3GPP organization. The further development of GPRS is done under GERAN group (GSM EDGE radio access network).*

When searching for GPRS-related information from GSM specifications one needs to consider the different versions of the specifications, and the different phases of the systems. Every GSM specification is given a tripartite version number, where the first part represents the version; the second stands for an altered version after a technical change; and the third part refer to a specification after editorial changes. The version numbers and stages correspond with GSM pre-release 4 specifications so that version 3 stands for stage 1 and version 4 stands for stage 2 specifications. Higher version numbers refer to GSM pre-release 4, stage 2+ specifications. Version 5 refers to release 96, version 6 to release 97, version 7 to release 98, and version 8 to GSM/UMTS-release 99. GSM/UMTS release 4 is version 4, and release 5 is version 5. Versions 0, 1, and 2 refer to incomplete specifications. Figure 11.2 shows an example of a GPRS specification with definitions.

ETSIs SMG groups have since been discontinued, and nearly all GSM .specifications have been transferred to MSG (Mobile Specification Group) under 3GPP, which works on UMTS system specifications. Only section 13 GSM specifications remain under ETSI as ETSI TC MSG, as these specifications relate to European regulators' restrictions. In practice, this change takes effect as of release 99 specifications.

3GPP's organization consists of a project coordination group (PCG) and five technical specification groups (TSG), as can be seen from Figure 11.3. The future development of GPRS takes place mainly in the GSM EDGE radio access network (GERAN) team, the newest of the five teams. The other teams are system architecture (SA), core network (CN), terminals (T), and radio access network (RAN).

The first UMTS specification, known as release 3, complies with GSM release 99. Release 3 is followed by official UMTS releases 4 and 5, earlier called with a common name release 2000. GSM release 99 is superseded by a so-called GSM pre-release 4; the specifications of which have been numbered according to earlier GSM specifications starting from no. 1. The actual UMTS specification more or less follows the same subject headings as the GSM specifications, but numbering begins at 21. Thus, UMTS no. 21 corresponds to GSM no. 1. Specifications after GSM release 4 are numbered from 41 onwards, respectively.

The GSM radio interface keeps evolving via the 3GPP specification work. The latest GSM (GERAN) releases, that is, following the UMTS and LTE specification numbering, keep evolving to provide additional features. The recent evolution contains more GSM frequency bands, support of additional location technologies, increased data throughput, enhanced spectral efficiency and better hardware efficiency, interworking with other radio access technologies, and support of IP-based interfaces.

11.4 Architecture of GSM

11.4.1 General

The GSM network consists of the network and switching subsystem (NSS), the base station subsystem (BSS), and the operations subsystem (OSS), which controls the functioning of the NSS and the BSS. Figure 11.4

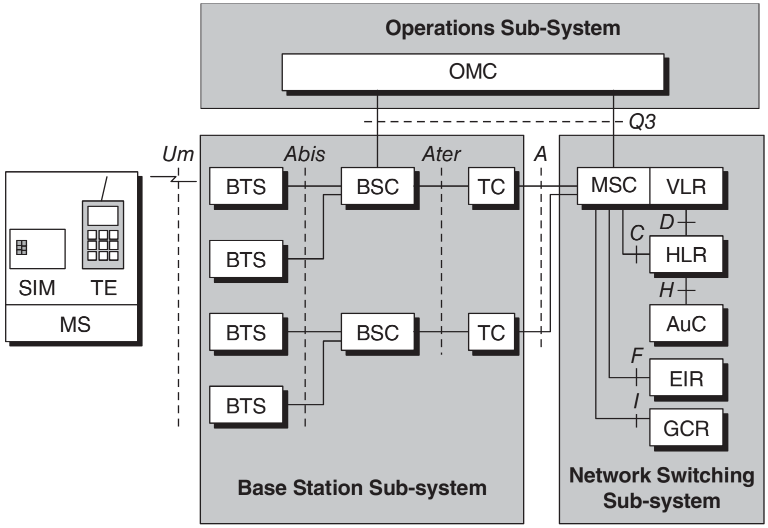

Figure 11.4 The main systems of GSM network.

portrays the GSM system architecture including the most important interfaces. In addition to the elements shown in the figure a GSM network contains elements for example, for the storing of voice mailbox messages, and the collection of billing data. These elements have not necessarily been defined by the specifications, which means that they are often operator-specific and manufactured by external vendors.

The most important function of the mobile services switching center (MSC) is to switch the calls made to the GSM network to the right BSS via the base station controller (BSC) and the base transceiver station (BTS), and to switch calls received from the BSS to other networks. Network registers containing information on the location of the subscriber are used to route the calls. The calls can also be internal to the GSM network which means that calls originate and terminate within the range of one or several MSCs.

The functioning of the MSC is based on the switching centers and protocols of the fixed line network. Thus the switching techniques of the GSM network uses the same principles as the centers of the fixed network, and equipment vendors typically use identical solutions in both cases. The GSM network naturally includes more elements and functions related to the mobility of the subscriber, which need to be taken into account in the functioning of the centers.

The interface used for voice traffic between the MSC and external voice traffic networks is identical to the centers of the fixed line network. Interworking functions (IWF) between the MSCs and external networks enable data connections between the mobile and fixed networks. These interworking functions include for example, modem equipment required for circuit switched data connections.

11.4.2 Area Specifications

The GSM service area refers to the area where GSM services can be used. Thus the maximum service area consists of all functioning GSM networks, which allow roaming either using the same GSM terminal or via so-called SIM-roaming, where the user must switch the SIM card to another terminal supported by the visited network.

The area covered by the radio network of a single MSC is called a center area, which in turn consists of several location areas (LA). The location areas are formed by groups of cells within a certain geographic region within a center area. Incoming calls to the center area are routed to the right location area based on the location data contained in the visitor location register (VLR) of the GSM network. The subscriber's home location register (HLR) contains information about the VLR, where the subscriber has last registered. The VLR contains information on the specific location area, where the subscriber has last connected to the network.

A cell is the coverage area of one or several transceivers (TRX) connected to a certain base station and serving a single geographical area. The typical number of TRXs in one cell ranges from 1–3 in rural regions and could be 6 or more in urban areas. Specifications do not determine how many TRXs a base station needs to support, which means that the number can vary between equipment vendors and different base station models.

Cell size is affected by direction patterns, amplification, and height of the antennas, as well as by the profile of the terrain surrounding the antennas, and the transmission power in use. Cells can be classified by their environment and size into pico, micro, and macro cells, and so-called umbrella cells. Pico cells are small in size and often used in indoor solutions. Microcells utilize antennas, the height of which remains lower than the average height of the buildings in a given area, whereas macro cells reach above building tops. The umbrella cell is a large solution used to increase coverage area by filling in shaded areas. It can also allocate capacity between the cells in its coverage area.

The maximum number of BSCs in the area covered by one center depends on the dimensioning of the operator. Limiting factors include for example, the different solutions of the equipment vendors, and the capacities of the registers. The specifications do not define the maximum amount, and the same principle applies to the number of BTSs and cells within a given BSC.

11.4.3 The Base Station Subsystem (BSS)

The base station subsystem (BSS) is in direct contact with the mobile terminals via the radio interface [14,15]. The BSS is also in contact with the network and switching subsystem (NSS). The function of the BSS is to connect the terminals to the NSS.

The BSS consists of base stations (BTS, base transceiver station) and base station controllers (BSC) controlling the BTSs, and the transcoder / rate adapter unit (TRAU).

11.4.3.1 The Base Transceiver Station (BTS)

The physical equipment center together with the base station equipment is known as the base transceiver station (BTS). The BTS equipment includes TRX elements, power supplies, combiners, power dividers, antenna cabling, the mast, encryption equipment, antennas, and mast amplifiers. In some cases, there might be separate unbreakable power system (UPS) for maintaining power.

One TRX transmits traffic on one frequency, if synthetic frequency hopping has not been specified. Each GSM frequency has been divided into eight time slots (TS or TSL), which means that one frequency can have a maximum of eight users in the basic case. If half rate codecs are used, the maximum number of simultaneous users reaches 16. Some time slots are required for signaling, which means that not all time slots in a cell can be used for the transmission of voice traffic.

A BTS can transmit the traffic of one or several cells. A cell is the coverage area formed by one or several TRXs, where radio frequencies can be utilized with a certain probability. Each cell has its own cell identity (CI), observed by the terminals in the cell's area. Figure 11.5 shows an example of a cell with three frequencies, with three time slots reserved for signaling.

The antennae of the cell can be omnidirectional. Such a cell is useful especially in rural areas, which require extensive coverage areas. Directive antennas can be used to efficiently cover areas such as major highways

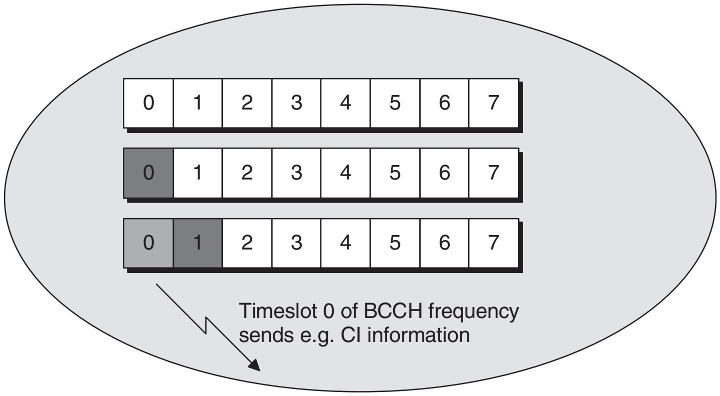

Figure 11.5 *An example of a 3-TRX cell. The cell sends constantly cell specific information on it's time slot 0 of the BCCH frequency. This information consists of for example, the Cell Identity (CI) and other relevant information about the base station identity and radio parameter values. There are also other needed signaling channels marked in Figure as grey colour.*

on both sides of the BTS. They are more popular in modern network planning than omnidirectional antennas, as they offer more efficient use of capacity.

11.4.3.2 The Base Station Controller (BSC)

The basic function of the base station controller (BSC) is to take care of the radio resources within its area. The MSC switches calls via the right BSC to the mobile station (MS), while the BSC handles events in the radio interface at the time of the connection.

The area covered by each BSC typically contains multiple base stations. Location areas are formed by grouping these base stations. According to specifications, a location area can contain entire areas or parts of two or more areas controlled by one BSC. Figure 11.6 illustrates a situation where cells in location areas LA1

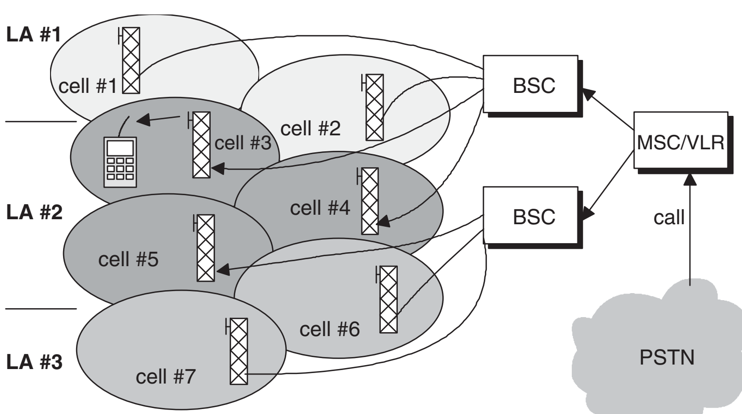

Figure 11.6 *The GSM location areas can consist of part of BSC, or several BSCs can form a single LA. In this example, the location area LA#2 is formed by two BSCs.*

and LA2 belong to the control areas of two separate BSCs. Location area LA2 has been split between these two BSCs.

Should a mobile phone in idle mode move from one location area to another the mobile phone informs the BSC. For this to happen the MS needs to have its own signaling channel, assigned to the MS by the BSC when the MS needs to update its location to the network. The BSC passes on the information regarding the new location area of the MS to the visitor location register.

When there is an incoming call to an MS in a certain location area the BSC pages the MS via all cells in the location area. The BSC defines the channel after the MS has responded via the best cell. In the case of outgoing calls BSC allocates the channel after the MS has asked the BSC for a channel. In Figure 11.6 the BSC pages the MS via cells 3, 4, and 5. The MS responds via cell 3, where the call is set up.

The BSC is constantly aware of all channels available or in use in each cell in its area. In addition, the BSC monitors the quality of the connections based on the measurement results delivered by the MSs in busy mode, that is, when they have active call. Thus the BSC controls the allocation and switching of channels during the set up and duration of a call in all such cases, where the call is made within the area served by the BSC. The BSC alerts the MSC only when channels are switched. Handover can also occur between two BSCs, in which cases also the MSC participates. In Figure 11.6, if the MS switches from cell 3 to cell 5, it also switches BSCs, and the MSC in between these two BSCs takes part in the handover. Handover can also occur between two different MSCs using a special handover number.

Handover can occur due to the low signal level in the coverage area or the variance in service quality. Both the MS and the BTS monitor signal strength and quality and send the results to the BSC. The BSC monitors the utilization of resources by individual connections. It processes and averages out these results reported by MSs and BTSs, and decides when handover should take place. The BSC also considers the channels in use and evaluates the interference levels in the available channels, and sends a channel allocation to the BTS.

In addition, the BSC controls the radio parameters, as traffic channel frequency hopping sequences, power control, discontinuous transmission and receiving, and the location areas. For example, in the case of frequency hopping the BSC records and sends the related parameters to the BTS. Thus the BSC signals directly with the MSs, without the participation of the MSC.

For each connection, the BSC chooses an available PCM (pulse code modulation) time slot or possibly a submultiplexed 16 kb/s time slot from the Abis interface. The MSC chooses the A-interface between the BSC and the MSC, but the MSC can, if necessary, alert the MSC and give up the channel for example, in error situations.

11.4.3.3 Transcoder

The transcoder/rate adapter unit (TRAU) forms the third part of the base station subsystem. The TRAU takes care of speech encoding and decoding, and adapting the data rates to match the transmission types of the internal GSM network and external networks. The corresponding element in the GPRS network is the packet control unit (PCU), which separates GPRS data packets from the GSM traffic.

The transcoder decodes speech which has been encoded by a mobile phone to a form understood by the normal fixed line network, and vice versa. The idea is that from the mobile terminal to the transcoder, speech, data and signaling traffic can be fitted in a 16 kb/s-submultiplexed PCM timeslot, although the use of full 64 kb/s timeslots is also possible. In any case, a full 64 kb/s PCM timeslot is required from the transcoder to the MSC. In the digital fixed network speech is transmitted using the same 64 kb/s timeslots as a so-called A-law form in a European system, and as a μ-law form in an American version. Thus in the GSM network speech can be compressed to $\frac{1}{4}$ of what is required in a fixed line network, which has a more old-fashioned encoding method.

TRAU frames have a fixed size of 320 bits. In addition to speech TRAU frames, there is a data frame which is used for data transfer (except not with a 14.4 kb/s service) and synchronization (only with a 14.4 kb/s service). In respect to GPRS the data frame specifically refers to the PCU frames handled by the PCU element. A so-called extended TRAU frame can be used with the 14.4 kb/s service. Also PCU frames can be called extended frames. Like the TRAU frame the PCU frame also has 320 bits.

In addition to the above-mentioned frames, O&M frames are also transferred between the base station and the TRAU to enable several BSC-internal functions. TRAU frames are generally transferred between the base station and the TRAU once every 20 ms intervals. Whenever the frame does not contain voice or data traffic, an idle frame is used between the base station and the TRAU.

The transcoder can be located in connection with the base station, the BSC, or the MSC. Typically the transcoder is located in the BSC- or MSC site. The proximity of the transcoder to the MSC enables the efficient utilisation of the GSM network's transmission capacity. If the transcoder is set between the BSC and the MSC, an Ater-interface is formed between the BSC and the transcoder. The functions of the transcoder are detailed in specification ETSI GSM TS 08.60. According to specification ETSI GSM TS 03.60, the PCU can be located with the BTS, the BSC, or the SGSN element.

11.4.4 Network Switching Subsystem (NSS)

The network switching subsystem (NSS) consists of the mobile services switching center (MSC) and the related registers such as the home location register (HLR), the visitor location register (VLR), the equipment identity register (EIR), and the authentication center (AuC). The most recent addition to the registers specified by the GSM specifications is the group call register (GCR). The mobile application part (MAP) between the MSC and the registers is done using SS7 (Signaling System #7).

The NSS forms the connections between phones external to the GSM network, and between GSM terminals. The NSS also switches the calls made within the GSM network, which consist of calls made within one MSC or from one MSC to another. Calls between MSCs can be direct or they can be routed via the public telephone network. The connection is always routed via a minimum of one MSC in the region, even if the connection were established between two GSM terminals within the same cell.

Standard interfaces have been defined in the MSCs for connections external to the GSM network. The connections use the standard fixed line numbering system and the established SS7. The MSCs contain interworking functions for data connections to external networks.

11.4.4.1 Mobile Services Switching Center (MSC)

The construction of an MSC resembles that of a fixed network center, and looks much like a typical digital gateway center. In addition to the functions of a normal center, an MSC has several specialized functions related to the mobile network, of which mobility management is the most important. The various registers of the MSC also handle subscriber identification and the protection of the radio interface [16, 17].

In addition to the registers the MSC contains interworking functions (IWF), which in effect refers to typical modems, ISDN rate adapters, and other devices used via data connections. The IWF can also contain the echo reduction devices used with voice traffic, but which may never be used for data or fax connections.

If a call originating in an external network or center is routed to an MSC in whose region the receiving terminal is not located, the call is rerouted to the right MSC by a gateway MSC (GMSC).

The most important function of an MSC is to switch, maintain, and complete calls within its area. In the case of the GMSC, the call would have to be relayed to the next MSC. Calls can be internal to the network and made within one MSC or between different MSCs. The call can also be placed between a GSM network and external networks.

The MSC oversees a call that has been switched. However, the BSC controls the events in the radio interface in such cases where the mobility of the mobile phone is restricted to the BSC area. Mobility management between two BSCs is handled by the MSC in between the BSCs. Mobility can occur also between two MSCs, in which case both centers participate in events during a connection.

Information related to calls can be extracted from the MSC, for example, with the help of the operations and maintenance unit. The loads on the channels used by the MSC can be examined. Depending on the configuration of the channels, issues such as the share of data calls of total traffic and other statistics on calls can be studied. In addition to providing meaningful data, such results also help find potential equipment failures. Billing data in the form of billing tickets are also collected from the MSCs.

One or several base station controllers connect to each MSC on the A-interface. Channel configurations occur using basic 2 Mb/s PCM timeslots. An MSC can also have one or several 2 Mb/s PCM connection lines to external telephone – or ISDN – network centers via the standard G.703/G.704 interface. The center also potentially contains a user interface, for example, via the X.25-packet switched network.

The MSC is generally expected to be capable of switching 64 kb/s timeslots. The MSC must also be capable of signaling with external networks using signaling system no. 7 (SS#7). SS#7 is the only internationally compatible signaling system, which can be used to control connections either between GSM networks, or a GSM network and external networks.

11.4.4.2 Registers

Home Location Register The home location register (HLR) contains subscriber and billing information and other additional services. The information can be permanent or changing in nature. Permanent subscriber data include the mobile subscriber ISDN number (MSISDN), the international mobile subscriber identity (IMSI), protection parameters and subscription type. Changing in nature are the information related to the ability to contact a subscriber such as information on switching on to the network, and routing data, which reveals in which HLRs area a subscriber is currently located, calls switched by the subscriber, and other service requests.

The database of the original HLR needs to be expanded to also include the GPRS register (GR). Thus the HLR of a basic GSM network requires a software update in order to function with GPRS.

Visitor Location Register In practice, a visitor location register (VLR) has been integrated with every MSC. For this reason an MSC is often referred to using the term MSC/VLR.

When a mobile phone registers to a new center area, the VLR of the area contacts the HLR for subscriber information, and alerts the HLR of the mobile phone's new position [18]. The information resides in the VLRs memory as long as the mobile phone is in its area. This means that the subscriber information can always be found in the VLR of the MSC/VLR area in question in addition to the HLR. When the subscriber moves within the network to a new MSC/VLR area, the subscriber information is removed from the previous VLR and transferred to the new VLR. The same principle is in use with roaming.

The VLR contains, among others, the MSISDN and IMSI codes, the temporary mobile subscriber identity (TMSI), the mobile station roaming number (MSRN), and the location area (LA). In addition the VLR contains service configurations and protection parameters (the so-called protection triplets).

The cooperation of the HLR and the VLR is evident as the HLR always contains information on the VLR in whose area the mobile phone currently is located, or was last located. The VLR contains more specific location data based on the location area.

Equipment Identity Register Each mobile phone is equipped with an identity number, which is stored in the equipment identity register (EIR). The identity number has nothing to do with the subscriber number, and identifies only the equipment. The EIR signals with the MSC, so that the network can, if necessary, check the equipment information. Information on faulty or stolen phones or mobile phones under surveillance can be stored in the EIR. It is possible to prevent the use of a mobile phone via the EIR and using the international mobile equipment identity (IMEI) code, which identifies each GSM and GPRS terminal.

The EIR consists of the white, grey and black lists. The white list contains all equipment types that have been type approved to be used in the networks. The grey list can contain equipment, which needs follow-up, such as equipment which has received only temporary approval in the type approval process. The black list can contain illicit equipment, such as stolen phones or equipment, which has not received type approval. Even if a phone were placed on the black list, calls could be switched to the European emergency no. 112.

The international version of the EIR is called the central EIR (CEIR), and is physically located in Ireland. The CEIR compiles the information on illicit equipment using the international X.25 packet switched network. When information on stolen or illicit equipment is added to the black list of one operator, the information is also passed on to the CEIR and to the EIRs of all operators connected to the CEIR.

Authentication Center The authentication center (AuC) is used to store the confidential authentication numbers, which are specified for a subscriber when subscribing to the network. When a call commences the authentication numbers given by the AuC to the VLR are compared with the numbers sent by the mobile terminal. The call is blocked if the caller has no authority to use the mobile network.

The memory of the AuC contains the parameters required to identify the subscriber, to check usage rights, and to protect information. Neither the EIR nor the AuC are mandatory elements for the functioning of the network, but the AuC is in widespread use.

The AuC calculates a so-called protection triplet for each subscriber using the A3-algorithm, the subscriber's protection key "ki," and the random number RAND. The protection triplet is transferred to the VLR in whose area the subscriber has registered, and where a subscriber begins the setting up of a connection.

Group Call Register Specifications ETSI GSM TS 03.68 and TS 03.69 specify GSM group calls. Group calls are related to work phase 2+ and the service known as voice group call service (VGCS). The functionality requires a group call register (GCR), which is in contact with the MSC via the I-interface.

11.4.4.3 Interworking Functions

The MSC elements of GSM network have direct connections to the gateway centers of the fixed line network for voice connections. However, the MSCs must have special equipment to enable GSM data connections. This equipment is located in the interworking functions (IWF) of an MSC.

Depending on the operator, the IWF can contain audio frequency modems, ISDN rate adapter units (UDI or unrestricted digital information with V.110 rate adaptation, or RDI, restricted digital information with V.120- rate adaptation), packet switched data adaptation units, and possibly units, which support direct data connections.

According to the basic GSM solution, every MSC has its own IWF. In addition, specification ETSI GSM TS 03.54 specifies a shared IWF (SIWF), which means that several MSCs can use the resources of single IWF for data connections.

11.4.5 Operations Subsystem

The operations subsystem (OSS) has several functions, which require connections with the BSS and the NSS, as shown in Figure 11.7. The OSS has not been specified in extreme detail, which leaves a lot of room for

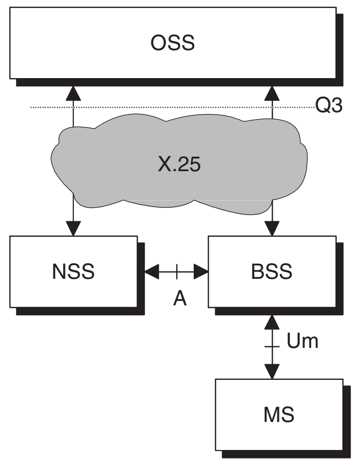

Figure 11.7 *The connectivity to OSS.*

choice for the equipment manufacturers in the design of the system. In contrast, the interfaces between the OSS and other elements have been specified.

The most important functions of the OSS are:

- Network operation and maintenance
- Controlling subscriber information, such as billing data
- Controlling the mobile stations.

The OSS contains one or more OMCs (Operations and Maintenance Centers), with which GSM network element programs can be installed, parameters can be inserted, and the status of the elements can be monitored. The OMC is in contact with the MSCs and the BSCs. The OMC sends maintenance messages to the MSCs via the BSC and the Abis- interface. The OMC can be operated from workstations via the man–machine interface (MMI).

The OMC and the elements connected to it use a Q3-interface, which has been specified in detail. The design of the actual OMC is up to the manufacturer, which means that the user interface can be either text-based or graphical. In practice the network elements of one manufacturer only function with the same manufacturer's OMC.

The OMC is used to transfer files and other data, collect information from network elements, and store and analyse measurement results. The OMCs responsibility is to operate the network under stable conditions, which means that service quality and measurement results are monitored. Also, changes are made to the network via the OMC. Software updates, extensions to the network, and parameter changes can be controlled via the OMC-workstation, and there is no need to physically visit the base station or MSC sites.

11.5 Functionality of GSM

11.5.1 Frequencies

GSM uses time division multiple access (TDMA) technology. In the GSM system, eight time slots (TS) are multiplexed for a given frequency. Thus each frequency can have a maximum of eight traffic time slots while using a full rate speech codec. The offered capacity can be approximately doubled using a half rate codec, as a single time slot can be allocated for two users.

As each GSM operator must have multiple frequencies, GSM is in fact a combination of TDMA and FDMA (frequency division multiple access) access technologies. GSM 900 frequencies are divided into three categories. The standard or primary GSM (P-GSM) operates in the 890–915 MHz frequency range uplink (from the terminal's transmitter to the base station's receiver) and 935–960 MHz downlink (from the base station's transmitter to the terminal's receiver). The P-GSM system contains a total of 124 frequencies, or a 25 MHz frequency band, which is typically shared between competing operators. The NMT 900 system, for example, also operates on the same frequency band.

The extended GSM (E-GSM) has been specified in the 880–915 MHz range (uplink) and 925–960 MHz (downlink). In addition, GSM specifications include the so-called GSM for railways (R-GSM), for which the frequency ranges have been set to 876–915 MHz (uplink) and 921–960 MHz (downlink) as defined in ETSI GSM TS 05.05 The common name for all of the above systems is GSM 900.

In addition to GSM 900, a version of GSM exists known as digital cellular system for 1800 MHz (DCS 1800). DCS 1800 typically operates in a personal communications network (PCN). The DCS 1800 uses frequency ranges 1710–1785 MHz (uplink) and 1805–1880 MHz (downlink), which means that the frequency range is 75 MHz uplink/downlink, and contains 374 frequencies.

The US-based GSM operates in the PCS 1900 network (personal communications system for 1900 MHz), which consists of several frequency ranges in the 1900 MHz area, including both narrow band and broadband systems. The digital mobile systems are CDMA (IS-95), GSM and North American TDMA (IS-136), of which the GSM is based on the European version.

All GSM systems are based on the same specifications. The differences are mainly in the radio interface transmission strengths, the number of channels, and frequency ranges. GSM Association has established standard names – GSM 900, GSM 1800, and GSM 1900 – for the networks. Specification 3GPP UMTS TS 45.005 also identifies the possibility of using frequency ranges 700 MHz, 750 MHz, and 850 MHz in the GSM system.

Currently, there are 14 GSM bands defined in 3GPP TS 45.005 as shown in Table 11.1.

Table 11.1 *The current GSM frequencies. The number in the system name indicates the band*

System name	Uplink (MHz)	Downlink (MHz)	Channel #
T-GSM-380	380.2–389.8	390.2–399.8	Dynamic
T-GSM-410	410.2–419.8	420.2–429.8	Dynamic
GSM-450	450.6–457.6	460.6–467.6	259–293
GSM-480	479.0–486.0	489.0–496.0	306–340
GSM-710	698.2–716.2	728.2–746.2	Dynamic
GSM-750	747.2–762.2	777.2–792.2	438–511
T-GSM-810	806.2–821.2	851.2–866.2	Dynamic
GSM-850	824.2–849.2	869.2–894.2	128–251
P-GSM-900	890.0–915.0	935.0–960.0	1–124
E-GSM-900	880.0–915.0	925.0–960.0	975–1023, 0–124
R-GSM-900	876.0–915.0	921.0–960.0	955–1023, 0–124
T-GSM-900	870.4–876.0	915.4–921.0	Dynamic
DCS-1800	1710.2–1784.8	1805.2–1879.8	512–885
PCS-1900	1850.2–1909.8	1930.2–1989.8	512–810

The channel bandwidth in the GSM system is 200 kHz. Thus a 200 kHz channel space is included in the beginning of each frequency range, so that the first P-GSM frequency for example, is in fact 890.2 MHz.

11.5.2 Channels

Each TDMA time slot in the GSM system is called a physical channel. One physical channel can have traffic channels, control channels, and their combinations, depending on the channel configurations in use. These are called logical channels. For example, one physical channel can be used to transfer one full rate channel or two half rate channels and related control channels.

GSM channels are specified as shown in Figure 11.8. In addition, GPRS uses new packet-specific channels, which have been described in the chapter covering the GPRS radio interface.

Traffic channels are used to transmit the sender's speech and data. The control channels are used when the phone connects to and from the network, to follow the location of the phone, and to set up, maintain, and end calls. The channels are also used to send parameters required by the above-mentioned functions, and to send the measurement results of the MS and to relay short messages. The control channels can be either one-way or two-way, depending on the situation.

The control channels can cause a fair amount of traffic in addition to the traffic channels, as they are used to relay short messages in addition to signaling. Careful network planning requires the allocation of sufficient capacity to the control channels in order to avoid bottlenecks and suboptimal functioning of the network.

As shown in Figure 11.8, the control channels can be classified into broadcast control channels (BCH), common control channels (CCCH), and dedicated control channels (DCCH).

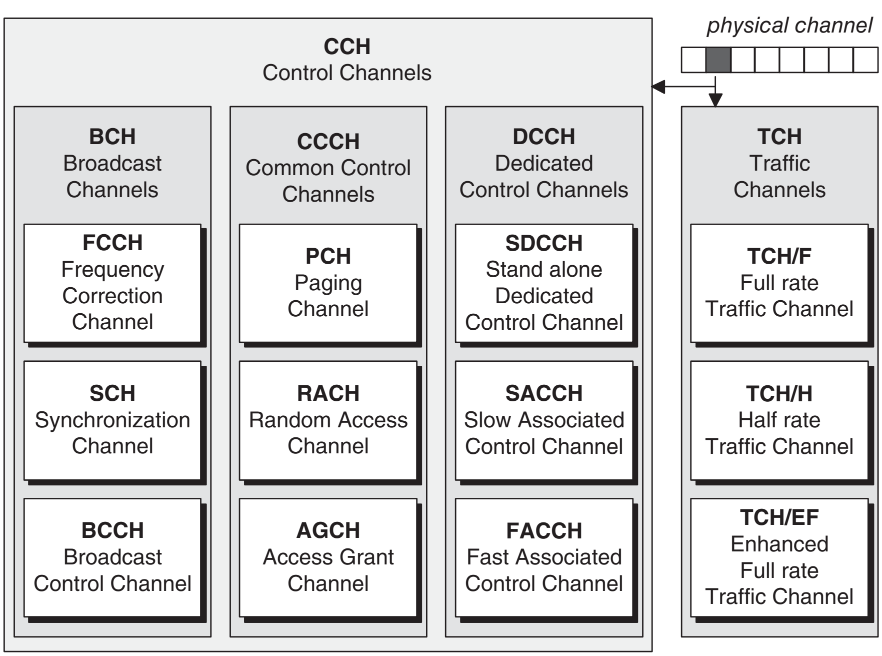

Figure 11.8 The logical channels of GSM. Each physical channel (i.e., time slot) can transfer one or more logical channels.

11.5.3 Traffic Channels

A traffic channel can either be a full rate traffic channel (TCH/F), a half rate traffic channel (TCH/H), or an enhanced version of the two, where speech connections use an enhanced full/half rate codec. Of these, all except the enhanced half rate codec have been specified. It can be considered included in the adaptive multi rate codec (AMR), currently under specification. GSM traffic channels are two-way or full-duplex channels.

A full rate traffic channel can be used to transfer speech information with a raw bit rate of 13 kb/s and user data within a 300–9600 b/s frame, where the rate is either 12.6 or 3.6 kb/s. If a half rate codec is being used, the speech information raw bit rate is halved to 6.5 kb/s. The data rate for user data falls to 300–4800 b/s, transmitted in a 3.6 or 6 kb/s frame, respectively.

The same regular bursts are used with the half rate codec as with the full rate codec, but each user uses a time slot only in every other frame. Thus the number of bits in a single burst remains unchanged, but they are transmitted half as often. This leads to an approximate doubling of capacity, or perhaps even slightly more than double due to the trunking gain effect, in compliance with traffic theory. The codec for speech transmission is so advanced in comparison to the original full rate codec, that the quality of the connection is not affected in any significant way.

11.5.4 Control Channels

Control or signaling channels are used when a mobile phone connects to and from the network, to follow the location of a mobile phone, and to set up, maintain, and end a call. The channels are also used to send the parameters related to the above-mentioned functions, and to relay short messages. Control channels can be one-way or two-way, depending on the situation.

11.5.4.1 Cell Search and Measurement

When connecting to the network an MS searches the frequency range for suitable broadcast channels (BCH), which are either the home operator's channels, or in the case of roaming, either automatically available or user-defined channels of the visited operator. The channels of other operators are automatically rejected from the MSs list. At this point, the primary channels in which the MS is interested, are the frequency correction channels (FCCH), and the synchronization channels (SCH).

When the mobile phone checks a frequency range, it first searches for the frequency correction channels. The information bits of all bursts in these channels has been set to zero, so that when an MS receives and demodulates the frequency correction channel, the result is a sine wave. The MS uses the FCCH to adjust its timing, so that it can also demodulate the bursts of the other channels. The FCCH gives the MS a rough indication of the limits of the time slots in the cell in question, and specifically information on the broadcast channel time slot 0. The broadcast channel time slot 0 is important, for it is ultimately the source of all relevant cell-related information for the MS.

Once the MS has located the sine wave, that is, the FCCH, the MS also finds the synchronization channel (SCH) located on the same frequency. This is due to the structure of the frame where the SCH always follows the FCCH channel on the physical time slot 0. With the help of the SCH, the MS is able to find out the specifics of the time slots in the cell. The MS also receives sufficient information from which to infer the numbers of the frames and time slots in the cell's super frame structure (the numbering of which consists of 8, 26/51 and 2048 timeslots). From this point onward the MS can handle the monitoring of the borders of the time slots, and can increase the numeric value of a new time slot by one.

Having located the FCCH and the SCH, the next job of the MS is to find the cell's broadcast control channel (BCCH). The BCCH is an important channel as it is used to transmit all relevant information on the

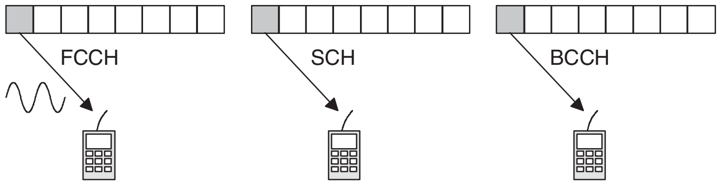

Figure 11.9 MS finds the BCCH frequencies by the FCCH. After finding FCCH channel, MS finds easily the SCH channel. Then, MS can finally start observing the BCCH information.

cell in question. Each cell transmits the BCCH in the broadcast channel time slot 0, like the FCCH and the SCH. Having found the SCH the MS is aware of the borders between the time slots, and quickly demodulates the BCCH burst. The MS can infer the other frequencies used by the cell, the frequency hopping sequence, channel combinations, and information on the neighbouring cells from the information transmitted by the BCCH channel.

The FCCH, SCH, and the BCCH are only transmitted downlink, which means that the phones monitor these channels as shown in Figure 11.9.

An MS in active mode independently maintains a list of the best BCCH channels with the help of cell choice criteria.

Each cell must contain a BCCH channel in the time slot 0 of the BCCH frequency. Specifications allow the use of multiple BCCH channels within the same TRX in special cases. In such a case, these additional BCCH channels use physical time slots 2,4, and so on. In the case of multiple BCCH channels FCCH and SCH channels are only used in the BCCH channel on time slot 0.

11.5.4.2 Establishment of Connection

The paging channel (PCH) is used to set up a call or signaling connection in the case of a mobile terminated call (MTC). The phones constantly monitor the BCCH frequency for occasional signaling channels. When the phone is in active mode, the VLR with the subscriber information is aware of the phone's location by location area (LA). The location area consists of a group of cells, and when there is an incoming call to the phone, the phone is signaled via all cells within the location area.

The phone recognises its own identity from the information on the signaling channel. Typically the temporary mobile subscriber identity (TMSI) is used in the signal. If the network does not support TMSI or fails to send it for some other reason, the international mobile subscriber identity (IMSI) is used. The phone answers the signal sent via the PCH channel using the random access channel (RACH). The phone also begins signaling on the RACH channel in the case of a mobile originated call (MOC). The network answers the RACH-call sent by the phone via the access grant channel (AGCH), which it uses to send information on the SDCCH channel (used to initialize the call) it has allocated to the phone. The combination of the PCH and AGCH channels is often called the paging and access grant channel (PAGCH), even if it has not been separately specified in the specifications.

The PCH, RACH, and the AGCH are all one-way; PCH and AGCH are sent using downlink, and RACH is sent using uplink, as illustrated in Figure 11.10.

11.5.4.3 The Call

Of all connection-specific call functions only the information related to establishing the call are transmitted on the stand alone dedicated control channel (SDCCH). In addition, short messages can be relayed on the

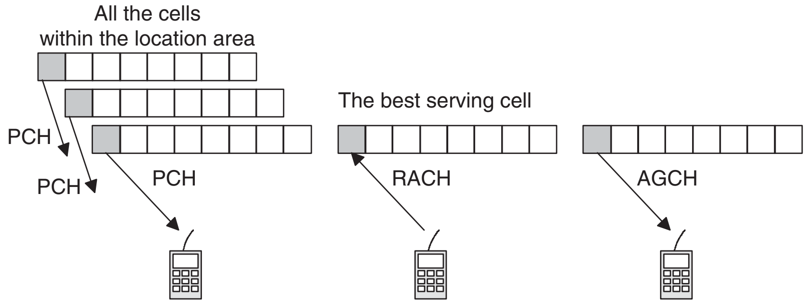

Figure 11.10 The network calls the MS via all the PCHs within the MSs location area. The MS answers to the network via the best serving cell's RACH channel, and gets acknowledgement via AGCH.

SDCCH when a call has not been set up. Prior to all other signaling, the authentication of the user and the checking of the IMEI code from the EIR to establish the legality of the terminal take place on the SDCCH. In addition, signaling related to the protection functions of the radio interface take place on the SDCCH. After the protection functions, all signaling is protected, if the network supports radio interface protection. If the connection is established only to signal information related to for example, location area updates, the connection is terminated after the signaling is complete. But if the connection is to set up a call, the SDCCH allocates a traffic channel (TCH) and a slow associated control channel (SACCH).

The SACCH is used to transfer information related to the maintaining and ending of the call, and short messages between calls. SACCH channel information is sent with intervals of only 26 bursts, together with information related to the traffic channels. The entire SACCH message, which consists of four bursts, is sent in 4×120 ms, or less than half a second. This is sufficient for most functions transmitted via the SACCH, such as the measurement data of neighbouring cells, timing advance (TA), and MS power control.

If the transfer rate of the SACCH is insufficient, the fast associated control channel (FACCH) is used. For example, channel switches require such fast messages. The FACCH operates in steal-mode, which means that it replaces 20 ms slots from the traffic channel. A user cannot detect the interruptions to the voice connection.

All connection-specific control channels, or SDCCH, SACCH, and FACCH, are used both in uplink and downlink, as shown in Figure 11.11.

11.5.4.4 Cell Broadcast

In addition to the above channels, a part of the SDCCH (1/4 or 1/8, depending on the channel configurations) can be permanently allocated as a cell broadcast channel (CBCH).

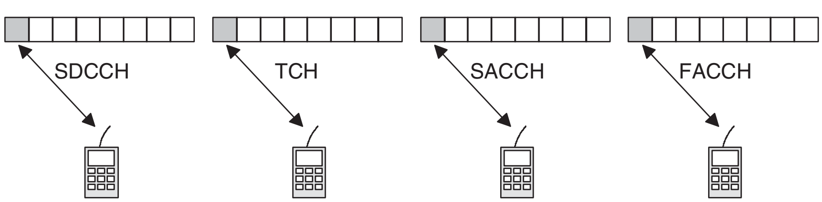

Figure 11.11 In the beginning of the call establishment, SDCCH channel delivers control information between the network and the MS. Finally, the TCH starts to handle the traffic, whereas the call control information is delivered on SACCH and FACCH.

The CBCH is used to send out a cell broadcast message to all mobile phones located in a certain geographical area. The terminals must support the cell broadcast function in order to receive CB messages. The terminal can also filter messages based on their title. This feature allows a user to block the reception of messages containing advertisements. Cell broadcast means a mobile phone does not send information; it only receives messages sent by the network.

11.5.5 Multiframes

When a phone is in call mode (voice, data, or fax call), the sender's data is sent in a 26-multiframe. Frames 0–25, each of which consists of eight time slots, are used to send intermittently the sender's data, control information related to the call, and measurement data.

The channel order is repeated by multiframe as indicated in Figure 11.12. Figure 11.13 portrays the 26-frame structure, which is made up of 24 frame periods (or 24 times 8 time slots) of traffic channels and one period of slow associated control channel (SACCH) related to the 24 traffic channel periods. One period has been left idle during which the MS has the opportunity to monitor neighbouring cells. The fast associated control channel (FACCH), which "steals" time slots intended for traffic channels, must be used for channel switches and other fast signaling. The interruptions caused to the transfer of voice are unrecognisable to the user, as they are hidden by repeating and moderating previously correctly received frames.

The frame structure of the control channels is repeated after every 51 frames. The information of FCCH, SCH, BCCH, PCH, and AGCH channels is sent using a certain sequence during the frame structure, depending on the channel configuration and the direction. Figure 11.14 represents the BCCH or 0- time slot signaling downlink. The other time slots of the BCCH frequency are used by traffic channels (L) in this example.

Figure 11.15 represents one example of the logical channels that can be sent downlink in the BCCH time slot. In its basic form, the BCCH time slot is used to send FCCH, SCH, BCCH channels and CCCH channels, which consist of PCH and AGCH channels. Each PCH channel is divided into several sections, and each MS only listens to the paging signals in its own section. The downlink direction of the BCCH time slot can also be used to send SDCCH and SACCH channel sections, in addition to the above sections.

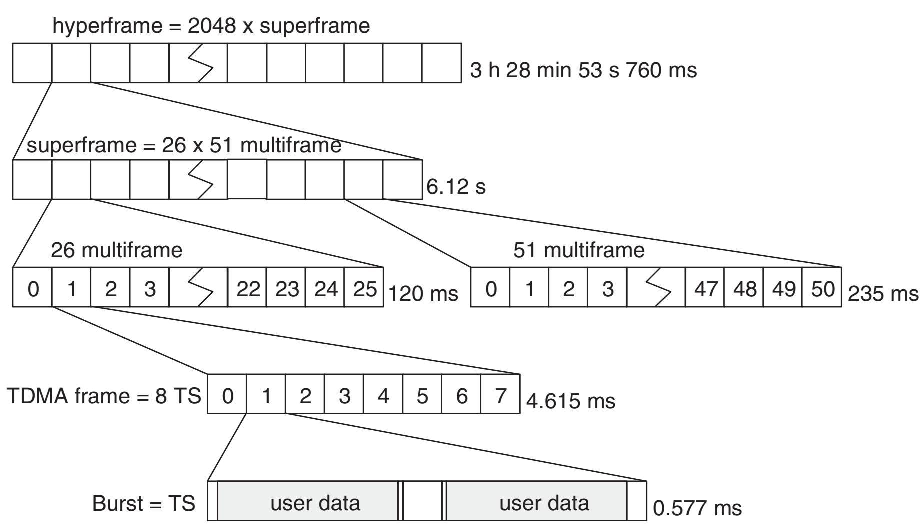

Figure 11.12 *The frame structure of GSM radio interface.*

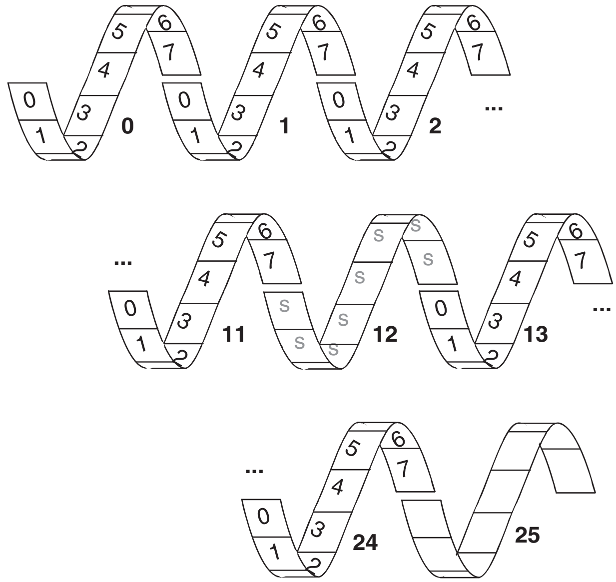

Figure 11.13 The 26-multiframe structure is utilised by TCH. One part of the information flow is reserved to slow associated control channel (S=SACCH) and one part is left empty (idle) for example, for neighbour cell field strength measurement data transfer. Accordingly, the 51 multiframe is utilised by control channels (in BCCH frequency's time slot no. 0), for example, for transferring the BCCH and paging channel information.

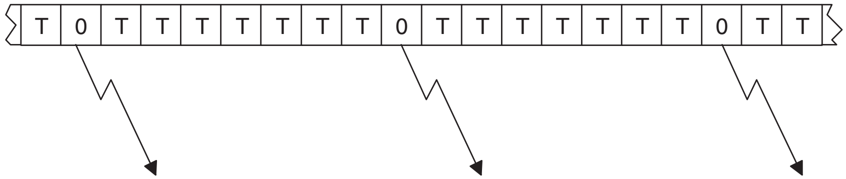

Figure 11.14 The cell sends information on timeslot 0 of BCCH frequency in downlink direction. In this example, the other timeslots are reserved for the traffic channels (T).

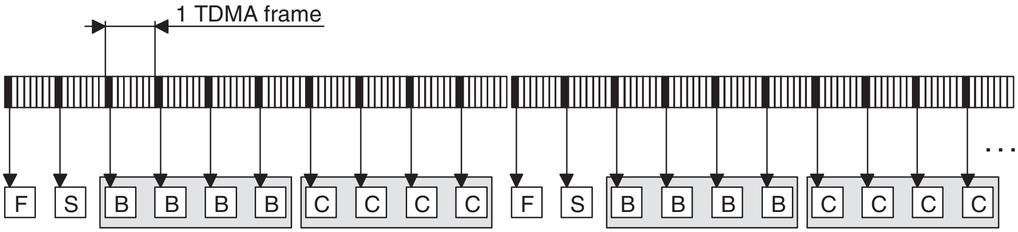

Figure 11.15 Example of the contents of the BCCH timeslot 0 in downlink direction. There are 20 TDMA frames presented; the same sequence is repeated every 51 frames. F=FCCH, S=SCH, B=BCCH, C=CCCH.

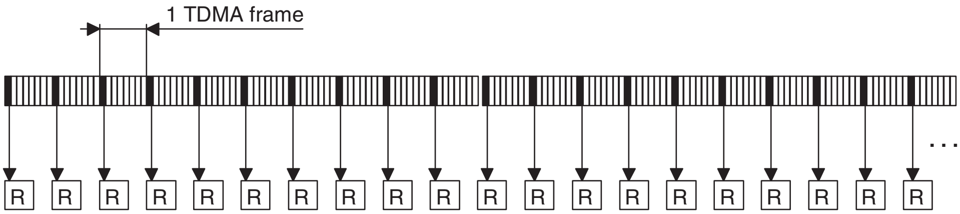

Figure 11.16 *Example of the contents of the BCCH timeslot 0 in uplink direction. Since the MS does not send any other information than RACH bursts (when necessary), all of the logical channels are defined as RACH channels.*

Figure 11.16 shows an example of an uplink BCCH time slot, which consists solely of RACH channels. Part of the SDCCH and SACCH channels can be defined to the uplink BCCH time slot, which means that approximately half of the 51-frame structure consists of RACH channels. Such a solution is suitable for regions that are not overloaded, and where the collisions of RACH bursts are thus rare.

The TDMA frames of 26 traffic channels (TCH) or 51 control channels (CCH) make up one multiframe. Multiframes are used because slow channel types do not always require time slots from consecutive TDMA frames.

51 traffic channel multiframes and 26 control channel multiframes are sent until they are again in the same phase. This structure is known as the superframe. 2048 superframes form a hyperframe, the duration of which is nearly 3.5 hours. Such a frame composition is used for frequency hopping and protection algorithms.

11.5.6 Channel Configurations

Logical channels can be formed using several types of channel configurations. According to specifications ETSI GSM TS 04.03 and TS 5.02 the permitted combinations of logical channels for one physical channel are[1]:

- TCH/F + FACCH/F + SACCH/TF
- TCH/H(0,1) + FACCH/H(0,1) + SACCH/TH(0,1)
- TCH/H(0,0) + FACCH/H(0,1) + SACCH/TH(0,1) + TCH(1,1)
- FCCH + SCH + BCCH + PCH + RACH + AGCH
- FCCH + SCH + BCCH + PCH + RACH + AGCH + SDCCH/4(0–3) + SACCH/C4(0–3)[2]
- BCCH + PCH + RACH + AGCH[3]
- SDCCH/8(0–7) + SACCH/C8(0–7)[4].

For example, the combination on the first row means that information is sent on the traffic channel on a certain time slot, until the SACCH information is sent once during a 26-frame structure. If FACCH is used, it replaces one traffic time slot. In other words, only one of the above-mentioned logical channels is sent at one time over a time slot or a physical channel.

The TCH/F in the list refers to full rate and TCH/H half rate traffic channels. TH and TF indicate that the signaling channels are parts of certain full or half rate traffic channels.

[1]In addition, ETSI GSM specifications specify multislot combinations.
[2]This replaces SDCCH/4(2) when cell broadcast channel CBCH is used.
[3]This combination is already being used, when the same TRX contains another BCCH channel with FCCH and SCH.
[4]This replaces SDCCH/8(2) when cell broadcast channel CBCH is used.

Depending on the load on the network in a given geographical area it may be necessary to change the cells' channel configurations. For example, a cell's SDCCH capacity could be increased in a region with a lot of signaling due to short messages and changes of location area. This could be done by configuring one or several of the cell's time slots with the last alternative in the above list where the SDCCH and SACCH channels have been combined and there is no traffic channel.

11.6 Numbering of GSM

11.6.1 Subscriber Numbering

11.6.1.1 MSISDN

The MSISDN (mobile station ISDN number) is the international subscriber number. MSISDN is based on the CCITT recommendation E.164, which means that it can be utilized as an international GT address (Global Title). The MSISDN contains the following fields:

MSISDN		
CC	NDC	SN

The CC field is the country code based on E.164, containing 1–3 digits. The NDC field is the national destination code, which means the code of the GSM network in the national numbering plan. The SN field is the subscriber number. In the national format, the MSISDN is of the following format:

MSISDN		
P	NDC	SN

The P field is the prefix for the MSISDN.

The GSM specifications define a "+" sign that can be utilized to replace the outgoing international number. The sign is replaced to the original number format in the number analysis of each country. The numbering analysis also decides if the call is national, and removes the international code.

The incoming circuit switched data call requires in practice a separate MSISDN so that the network can direct the call correctly for the user. The same principle applies also to the incoming fax service. In case the subscriber has the circuit switched data and fax services, normally has three different MSISDN numbers.

The MSISDN numbers are stored permanently in the SIM card as well as in the subscriber profile of HLR. The MSISDN are also stored temporally to the latest VLR where the customer has registered.

In case the GSM network consists of more than one HLR, the MSISDN indicates the HLR where the subscriber information is stored. In this way, the system can always make the inquiry directly from the correct HLR. In practice, the HLR can be distinguished, for example, based on the first digit of the subscriber number.

An Example of MSISDN Use Case A user whose home operator is TeliaSonera Finland, is visiting Germany and has been registered to local D1 network, being thus in roaming. The users national MSISDN is 040 0436302. Another user initiates a voice call from Sweden, being the A subscriber. The A subscriber selects the international outgoing number which in Sweden is 009, then the Finnish country code 358, the GSM network code in international format (in this case the first 0 is dropped resulting in "40"), and finally the B subscribers number 0436302. If the A subscriber selects the number via mobile network, the complete number can also be typed as "+358400436302". The call is directed to the Finnish GSM network's MSC which knows, based on the first digit of the B subscriber number, which is the HLR of that specific subscriber. The HLR contains now updated information about the VLR where the B subscriber is registered, that is, in the D1 network, and redirects the call to Germany.

11.6.1.2 IMSI

A single subscriber may have various MSISBN numbers. This is one reason why GSM system utilizes instead primarily an IMSI, international mobile subscriber identity, which identifies uniquely the subscriber. IMSI is utilizes, for example, for the charging of the calls. IMSI is a series of numbers containing a maximum of 15 digits, and that has the following format:

IMSI		
MCC	MNC	MSIN

The combination of MNC and MSIN is called NMSI, national mobile subscriber identity, resulting in the following format:

IMSI	
MCC	NMSI

The MCC field of IMSI is based on the X.121 recommendation, that is, it is a three-digit mobile country code. It should be noted that the MCC differs from the CC, and is not visible for the users. The MCC is defined in the blue book of CCITT, in the recommendation E.212. The MNC is a mobile network code, and is a 2-digit code that defines the GSM network. It should be noted that the MNC differs from the NDC. The MSIN is the mobile subscriber identification number, and is individual for each user within the GSM network. The MNC and MSIN form together the NMSI.

Even if the first parts of IMSI and MSISDN, and possibly also the rest of the digits differ from each others, also the IMSI should indicate the HLR of the subscriber. Like in the case of MSISDN, the digit or combination of the digits that indicates the HLR, depends on the operator's own strategies.

The IMSI is utilized for the subscriber identification within the GSM network, but normally it is protected over the radio interface via a separate, temporal number.

11.6.1.3 TMSI

TMSI is a temporary mobile station identity that is used for protecting the identity of the subscriber when related information is transferred over the radio interface.

The VLR allocates the TMSI when the subscriber registers in the respective VLR area. The network knows the TMSI as well as the terminal as it is delivered for the mobile station. When the MS registers within a new VLR area, the previously dedicated TMSI of the old VLR is liberated for other MSs. This means that the system should have enough number reserve for TMSIs. It should be noted, though, that the use of TMSI is not obligatory in the GSM network.

In fact, the GSM specifications do not define the use of TMSI very deeply. The most important definition is related to the length of the TMSI, which is 4 octets of binary number. In addition, the network cannot dedicate a TMSI where all the 32 bits are ones. This is due to the functionality of the system as the SIM informs with this value that the TMSI can not be obtained from the network. Otherwise, TMSI selection is up to the operators strategies.

The specifications also define a LMSI which is a local MS identity number. This speeds up the fetching of the subscriber information from VLR. The same definitions apply for LMSI and TMSI.

11.6.1.4 MSRN

The MSRN is a mobile station roaming number. It is a temporal number that is utilized in the routing of the incoming GSM call. MSRN is not subscriber specific, and it is not known by the A subscriber nor B subscriber. The VLR delivers the MSRN automatically when the MS registers to VLR. Depending on the operators' strategy, the MSRN can be assigned for the MS for the whole time MS resides in the VLR, or only for each active call in case of arriving calls.

The MSRN is based on the numbering recommendation E.163. In practice, there is a certain number reserve for MSRNs from the fixed telephone network or ISDN network of each MSC/VLR location. The prefix of the MSRN functions as the direction of the subscribers' telephone switching center and the rest of the number specifies the subscriber [19].

An Example of Call Routing via MSRN Let's assume the A subscriber of fixed telephone network resides in Sweden. The B subscriber is a subscriber whose home network is Chinese, and he is roaming in a Finnish GSM network. The B subscriber is registered in a VLR that is located in Tampere, Finland. Figure 11.17 clarifies the example.

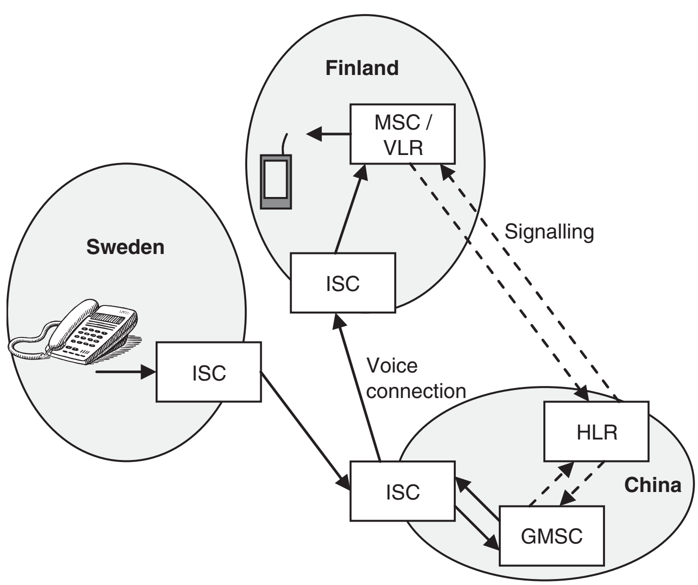

Figure 11.17 *Example of roaming.*

The A subscriber of Sweden selects the outgoing international code, the Chinese country code, the network code and subscriber number of the B subscriber. The dialed number is thus “009 86 ab cdefg”. The call is routed to the international exchanges of the fixed telephone networks of Sweden and China, and further to the first Gateway MSC of Chinese GSM network. The GMSC investigates the subscriber number and resolves the HLR where the B subscribers’ information is stored. The GMSC signals with the HLR in order to resolve what the VLR is where the subscriber is actually registered. At this stage, the HLR signals with the VLR of Tampere via Common Channel Signaling #7 procedure and asks for a MSRN number for this specific B subscriber. In this stage, the VLR selects an available MSRN from the numbering reserve of the local fixed telephone network of Tampere area where the VLR is located. The MSRN is 358 3 98 12345 in this example. Now, the VLR stores this MSRN with information that indicates that this MSRN should be mapped to the B subscribers ISDN number. Then, the VLR sends the MSRN for the Chinese HLR by utilizing GT (global title). As soon as the HLR has received this information, it delivers the GT for the GMSC that made the request. In this stage, the HLR does not take part of the signaling any more so it simply destroys the MSRN information.

The GMSC starts now forming the second leg of the actual call by “dialing” to the MSRN number which leads to the MSC of Tampere where the VLR is located. When the MSC/VLR of Tampere receives the signaling, it reviews what IMSI the MSRN means. Now, the MSC/VLR sends a paging, by utilizing previously mapped TMSI instead of IMSI/MSISDN, over the location area (LA) where the MS has recently registered. The B subscriber receives the paging and answers via the standard call initiation procedure.

This example shows how the voice or circuit switched data call is formed from Sweden via China to Finland. According to the typical charging principles, the A subscriber pays the call costs of the first leg from Sweden to China, and the roaming customer pays the second leg of the connection from China to Finland. As this example clearly indicates, the call channel directly from Sweden to Finland would be more economical for the users. The functionality is called Support of Optimal Routing, SOR, and the intentions have been to define it in ETSI GSM TS 02.79 [20]. In practice, though, the adaptation of the functionality has not been straightforward.

11.6.2 Mobile Numbers

11.6.2.1 IMEI

The IMEI is international mobile equipment identity, which is meant for the identification of the hardware of the mobile equipment. The IMEI does not have a direct connection with the subscriber numbering like MSISDN or IMSI because the removable SIM cards are meant for storing all the relevant subscriber information. If the mobile equipment authentication is activated in the network, the call is allowed only if the IMEI is noted to be permitted [21].

The IMEI code is utilized for the identification of fraudulent or stolen equipment via the equipment identification register (EIR). The register contains white, grey and black lists. The white list contains the permitted equipment types. The grey list is used for the equipment that is followed, and the black list contains the equipment that is not allowed to be used in the network.

If the equipment is found in the black list, no data or fax calls are permitted. Any other signaling, including short messages, is not allowed either. There is one exception, though, which is the call to the emergency center that functions always regardless of the black list definitions.

The IMEI is 15-digit number which consists of the following fields:

IMEI			
TAC	FAC	SNR	SP

The TAC is a type approval code which contains 6 digits. The FAC is a final assembly code with two digits. The SNR is a serial number with six digits, and SP is a spare number with one digit (which is always 0 if MS sends it).

Because the IMEI code explicitly reveals the type and commercial mark of the handset, it can be used as a basis for the statistics collection, for example, for the distribution of the models as a function of network areas. The specifications define the IMEI in such a way that it should be protected against intentions to change it.

The IMEISV is an IMEI software version number, and it contains TAC, FAC and SNR fields as well as 2-digit SVN, software version number.

11.6.2.2 SIM

Each SIM of GSM network has a unique ICC number, which is an integrated circuit card code. It indicates the operator code, manufacturing time, manufacturer code, serial number and revision number [22].

11.6.3 Network Numbering

11.6.3.1 LAI

Each GSM location area (LA) has its own LAI, location area identification, which has the following form:

LAI		
MCC	MNC	LAC

The MCC is a mobile country code, and the MNC is a mobile network code. These two numbers are the same as within the IMSI code. The LAC is a location area code, which has a value between 0 and 65 535, that is, it is of length of two octets. The code can be presented in hexadecimal format, excluding the values of 0000 and FFFF which are reserved for the special situations if the LAI is missing.

Each cell broadcasts the LAI constantly within the BCCH frequency's timeslot 0, and the terminals observe this information. When the MS in idle mode changes the cell order in such a way that the new best one belongs to the new LA, the MS initiates the location area update procedure (LA Update). It contains in practice the same information flow as in the initiation of the call, including the authentication of the subscriber.

11.6.3.2 BSIC

The BSIC is a base station identity code. It is a higher level "color code," which distinguishes the cells from each others, for example, in border areas of different GSM networks. The BSIC is formed by the following fields:

BSIC
NCC
BCC

The BSIC can have values in range of 64 numbers. The 3-bit NCC is a network color code, which distinguishes the GSM networks from each others. The specifications define a NCC code for each country for the separation of the international neighboring networks. The NCC codes have been selected in such a way that no neighboring country has same number. The 3-bit BCC is a base station color code, with values of 0 … 7. This code prevents the interference between the cell information.

11.6.3.3 CI

The CI is a cell identity with values of 0 … 65 535. The cell is identified within the location area based on this code.

The CGI is s cell global identification which consists of the LAI code (i.e., MCC, MNC and LAC fields) and CI field.

11.6.4 Other Numbers

11.6.4.1 HON

The HON is a handover number. It is used in a case when the connection is moved from one MSC are to another. The same principles apply to HON as in case of MSRN.

11.6.4.2 RSZI

The RSZI is a regional subscription zone identity. It can be used for dividing a single GSM network in separate zones that provide roaming. The RSZI consists of CC and NDC fields as well as two-octet zone code (ZC). The latter describes the zones that contain the allowed and prohibited location areas as defined in ETSI GSM TS 03.08.

11.6.4.3 Location Number

The Location Number consists of CC and NDC fields as well as LSP field (locally significant part). The Location Number defines a certain GSM area in such a way that it does not reveal a more specific structure of the network.

11.6.4.4 Element Codes

The MSC elements and the respective VLR elements can be identified with the international PSTN/ISDN numbers or with signaling point codes (SPC). The HLR can be located also via HLR identity that contains the MCC and MNC fields of the IMSI code as well as one or more MSIN numbers.

11.7 GSM Data

11.7.1 Principles

General packet radio service or GPRS is a packet switched extension of the GSM system [23–31]. Prior to GPRS, all transmission of data over the GSM network has been circuit switched in nature. This means that a connection has been established and maintained even if data were not be transmitted continuously. In comparison, GPRS has been designed for the transmission of bursty data based on the Internet protocol (IP). Even if a connection were established logically between a terminal and for example the Internet, the service would not require a continuous physical connection. The physical connection is active only for the duration of the actual data transmission. In addition, several GPRS users can share certain GSM radio interface resources with GSM users, if necessary [32].

In addition to entirely new, GPRS-specific elements, GPRS requires changes to existing GSM network elements. The service also requires an IP-based GPRS backbone network. The backbone network can be constructed using existing infrastructure and for example Asynchronous Transfer Mode (ATM) technology.

GPRS allows an operator to bill only for the data transmitted instead of billing for connection time as is customary in the circuit switched service. Paying only for transmitted data is an attractive option from the user's point of view, especially when the data being transmitted is bursty in nature and there are variable requirements for transmission.

The multislot technique, where several TDMA time slots can be configured for one user, increases transmission speed accordingly. Better quality connections increase data speed, as when the quality is sufficiently high, different channel coding algorithms can be used. The closer the terminal is to the base station, the less error correction is required, and more data can be transmitted.

Even if GPRS requires the operators to commit to investments, the system has been designed to utilize capacity in a very flexible manner. Existing resources, left unused by the circuit switched GSM connections, can be utilized for the transmission of packet switched data traffic as shown in Figure 11.18. In the basic version of GPRS GSM connections have been prioritized over GPRS. Thus GPRS traffic is stalled when there is time congestion in voice traffic.

Connections can be made from the GPRS network to the Internet for example and X.25 protocol-based packet-switched networks. The GPRS network is visible to external data networks as a traditional Internet subnetwork or an X.25 node. GPRS terminals and network elements have their own Internet addresses as shown in Figure 11.19.

Even if GPRS utilizes the same radio interface resources, GPRS terminals are entirely different than GSM phones. After being switched on, a GPRS-MS may connect itself logically directly to the GPRS backbone network, and is on standby to send and receive data. This means that the separate establishment of a separate connection typical with circuit switched connections is not required except for fast signaling. The initial signaling takes approximately as long as the initial signaling to set up a basic GSM call, depending on the configuration of the GSM network. Thus the protection of the radio interface, the checking of the Equipment Identity Register and user authorization add to the time delay before the real connection.

The traffic is based on the IP protocol, which means that subscriber numbers are not used at all. Typically it takes a few milliseconds to complete the initial signaling and begin the transmission of packet switched data, as the signaling complies with the same principles as with the voice service.

The principle of the GPRS radio channel has been configured to be flexible. In theory, the TDMA frame allows the use of up to 8 timeslots per user connection so that the number of uplink and downlink timeslots can vary, and the directions of transmission are completely independent of each other. The radio interface

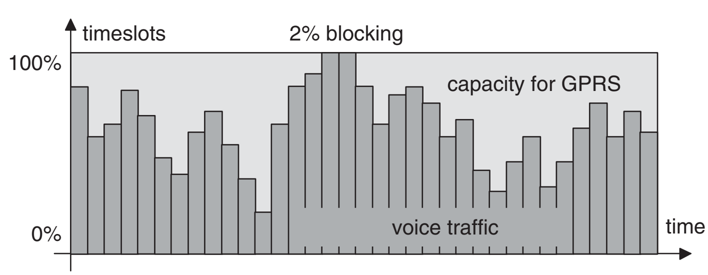

Figure 11.18 *GPRS can utilize the "leftover" timeslots from voice traffic.*

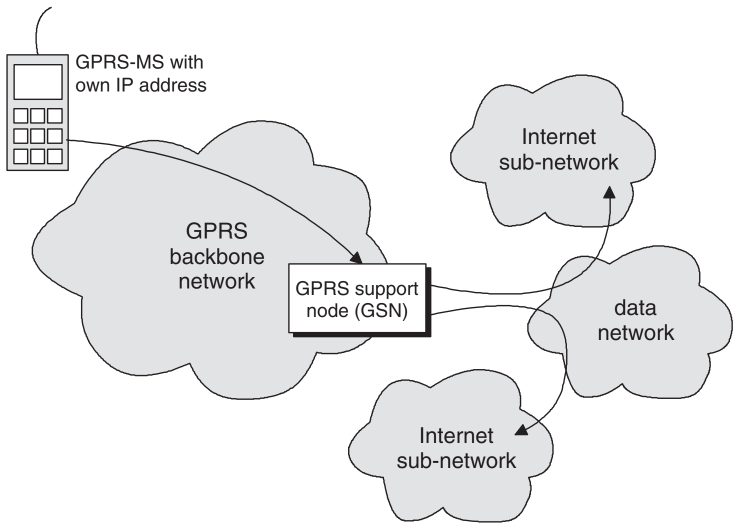

Figure 11.19 *GPRS is seen by the other data networks as Internet or X.25 subnetwork.*

resources can be distributed dynamically between voice and data traffic depending on the load on the network's services and the configuration of the network. In theory, the maximum peak data rate achieved using the maximum number of timeslots and the lightest possible channel coding is 171 kb/s. The realistic speed depends heavily on the capabilities supported by the equipment, the load of the network, and the level of interference. A likely transmission speed during the initial phase of the GPRS network is thus constrained to some tens of kilobits per second per user.

The Point-to-Point (PTP) and Point-to-Multipoint (PTM) services configured for GPRS can in principle be compared to the Short Message Service (SMS) and Cell Broadcast (CB) of the GSM network, but they are more diverse and can include for example, video in addition to text. GSMs SMS has also been configured to function in the GPRS network. However, the PTM service has been excluded from the first, release 97-compliant GPRS specifications.

11.7.2 The Network Architecture

Figure 11.20 portrays the GPRS network architecture, including the most important elements and network connections.

New elements when compared to the traditional GSM network are: the serving GPRS support node (SGSN), the gateway GPRS support node (GGSN), the GPRS register (GR), the packet control unit (PCU) which is attached to the base station controller (BSC), and the separate GPRS backbone network. The BTSs and BSCs may require software and hardware changes or upgrades depending on the equipment and software versions. Also the Home Location Register (HLR) will require a software upgrade. The Border Gateway (BG) element can be used for roaming between operators' networks. In addition to the above, GPRS specifications allow for the possibility of so-called legal interception, which requires specific elements defined in detail in Refs. [33–35].

The term GSN (GPRS support node) can be used to cover SGSN and/or GGSN elements. A GPRS network can contain one or several GSNs.

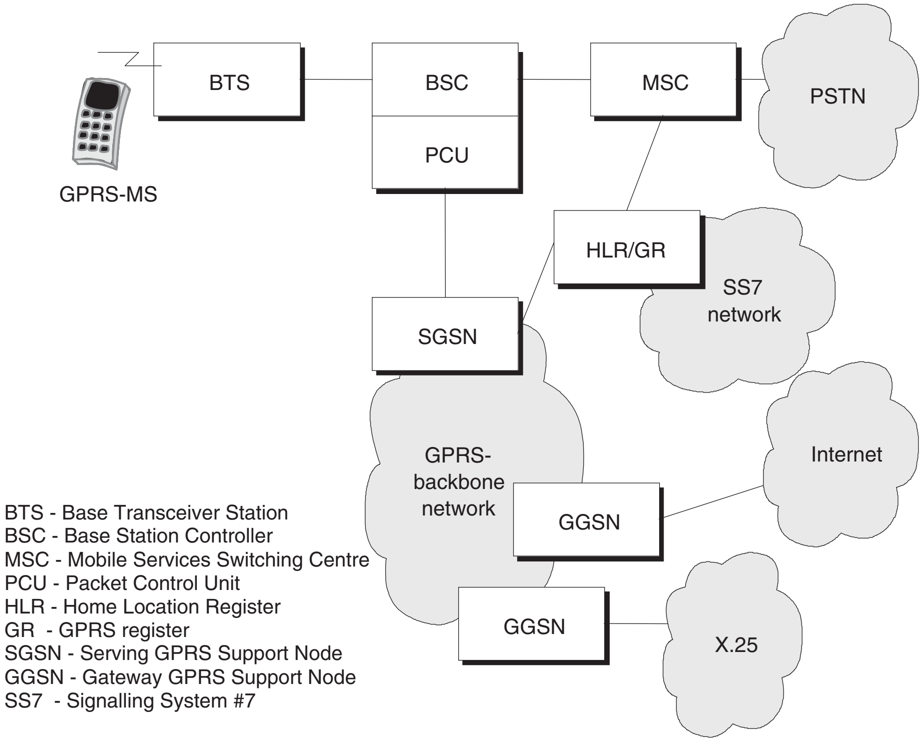

Figure 11.20 *GPRS architecture. The GGSN elements are utilized for connecting GPRS network to the external data networks.*

11.7.2.1 SGSN

The SGSN functions on the same level in the hierarchy as the MSC, following the location of individual GPRS terminals. The SGSN knows the location of the MS by cell or routing area (RA) depending on the terminal's mobility management (MM) state. The SGSN also takes care of user authentication and ciphering, so that the GPRS connection is protected between the MS and the SGSN. The SGSN also handles access to the network.

The SGSN connects to the PCU in the BSC via the Frame Relay network. The PCU can in fact also be integrated with either the BTS or the SGSN. On the other hand, the SGSN communicates with the other GPRS network elements such as the GSN and the BG.

The SGSN can send the user's location related data to the MSC/VLR via an optional Gs-interface. The SGSN can also receive paging from the MSC/VLR via the same Gs-interface. The SGSN also includes an interface to the CAMEL-function's GSM-SCF (service control function) element.

In addition to the above-mentioned functions, the SGSN can be used to collect required charging data on GPRS-connections by forming charging data records (CDR). The most important fact influencing charging is usage of the radio interface resources. In comparison to the GSM network, GPRS charging can, for example, be based only on the amount of sent and received data, even if the connection were logically formed for a longer time period. The circuit-switched connection traditional to the GSM standard is charged based on the length of the connection, even if data were not transferred at all. The charging data is directed via the Ga-interface to the charging data function (CDF), which in turn is connected to the billing system.

Other functions related to the SGSN are for example, compression of the user's packet data and the segmenting of the data packets prior to sending them to the PCU and the MS. The handling of cell selection also belongs to the SGSN. The GPRS service does not use handover, as does the GSM system. Instead, GPRS uses cell selection and reselection functions.

11.7.2.2 GGSN

The GGSN enables data transmission between data networks and the GPRS network. The GGSN is switched to the SGSN via the IP-based GPRS backbone network. The GGSN is visible to external networks as a router of the IP network or node in the X.25 network. The GGSN is in the interface between the GSM network or the UMTS network and an external data network, and its' functions inside the network are the same for both the GSM and UMTS networks.

When the MS and the SGSN are in standby-mode, the GGSN knows the position of an MS according to the SGSN when the subscriber is also roaming in other GPRS networks. If mobility management is set to ready-mode, the SGSN knows the position of the MS by cell. The GGSN is in principle similar to a router in the traditional fixed line network, but there is a significant difference between the mobility management of the GPRS and fixed data networks. Thus a GGSN must be able to route data connections in a mobile environment. In a sense the GGSN "anchors" a connection from the moment it is formed. "Anchoring" means that the connection is established via a certain GGSN, but depending on the GPRS subscriber's location and movements, the GGSN is capable of routing the connection via multiple SGSNs during the same data connection.

One or several SGSNs can be switched to one GGSN element. Data packets moving between the SGSN and the GGSN are tunneled using a specific GPRS tunneling protocol (GTP), which makes the system very flexible and allows the specification of new protocols to the GPRS system in the future.

As with the SGSN, the GGSN is also capable of collecting charging data records. Thus the same Ga-interface connects the GGSN and the CGF-element. The difference is that the GGSN can collect charging data related to the external data networks.

According to specifications, the functions of the SGSN and the GGSN can be combined to one physical element or they can be physically separated. The SGSN and the GGSN can include for example, the routing functions of an IP-, ATM-, or other network, and they can be switched to IP-based routers. When the SGSN and the GGSN are located in different networks, they can be joined via the Gp-interface. The functions of the Gp-interface resemble those of the Gn-interface, but it can be used to handle the protection measures required for connections between networks. The protection between two networks must be agreed separately by the two network operators.

The specifications also define the legal interception element for use by authorities. This means that it must be possible to follow the transmission of data via the GGSN, if necessary. This functionality is part of other regulations concerning authorities. As with the interception of regular voice calls made over fixed and mobile networks, the authorities must show substantial basis for suspicions of potential usage of the networks for criminal purposes to justify following GPRS traffic.

11.7.2.3 PCU

The Packet Control Unit (PCU) handles the connection between the GSM base station system (BSS) or the UMTS network's UMTS Terrestrial Radio Access Network (UTRAN) and the GPRS backbone network. The PCU separates GPRS packets from circuit switched connections and sends them over to the SGSN element, either via the Gb-interface of the GSM network or via the Iu-interface of the UMTS network.

The PCU forms so-called PCU-frames, which resemble TRAU-frames by the number of bits and by duration. Thus a PCU-frame consists of 320 bits, and the transmission length is 20 ms. Due to the similarities of the frames, the BSS and the UTRAN can handle PCU-connections flexibly.

In addition to placing the data in the PCU-frames, the PCU also handles part of the radio functionality in the GPRS network. This means that the PCU takes care of distributing resources to GPRS users, and controls access to the network.

11.7.2.4 HLR/GR

The GPRS register (GR) contains GPRS subscriber and routing information, which have been combined with the subscriber's IMSI-field. The IMSI can have several GPRS profiles. The GR has in practice been joined with the HLR by adding GPRS-specific data fields, meaning that HLRs require a software upgrade in order to function with GPRS.

The Packet Data Protocol Context, address, service level, and GGSN address are static subscriber information. The SS7- and IP- addresses of the SGSN, the ticket showing the accessibility of the GPRS-MS, and the GGSN list are dynamic information. SGSN-addresses are checked during the routing area update, and the GGSN list is updated on the basis of the accessibility of the GPRS subscriber.

Two interfaces have been specified for the HLR/GR: the Gr to join the HLR/GR and the SGSN, and the Gc in between HLR/GR and GGSN. Both interfaces use MAP-signaling. Gr is meant to be used for the updating of user location data, so that the dynamic database of the HLR/GR contains the information of the current SGSN. The Gc is optional, and its functionalities can be realized using the SGSN. The Gc is used to activate the PDP-context when the subscriber is about to receive a packet, but the PDP-context has not yet been established. The GGSN then asks the HLR/GR for the address of the SGSN in service using the IMSI as a primary customer related key.

11.7.2.5 SMS-GMSC

The GPRS network requires alterations to the gateway MSCs, so that current short messages (SMS) could be sent and received also over the GPRS network.

11.7.2.6 MSC/VLR-SGSN

The MSCs can be updated to perform optimally in connection with the GPRS service. For example, paging, combined GPRS-based and non-GPRS-based calls and circuit switched calls can be directed more efficiently via the SGSN.

The Gs-interface in between the MSC/VLR and the SGSN has been specified as optional. The interface is intended for the synchronization of the functions of the MSC/VLR and the SGSN especially in connection with paging and SMS transfers. The Gs-interface is in principle similar to the A-interface.

11.7.2.7 PTM-SC

Point-to-Multipoint (PTM) services and the PTM Service Center (PTM-SC) are part of the second stage of GPRS.

11.7.2.8 BG

GPRS networks can be joined either directly via the GGSN and the SGSN, or by using a separate Border Gateway (BG) element. The specifications do not define any obligatory protocol between BGs, which means that operators signing a GPRS roaming agreement must also agree on using a common protocol. The BG elements can be connected using a separate data network, which can be the internal data network of the operators. A logical solution for the GPRS roaming is the GRX (GPRS Roaming Exchange) which connects the parties via a single point for each operator.

11.7.3 GPRS Interfaces

Every GPRS Public Land Mobile Network (GPRS-PLMN) contains two types of access points. The Um or radio interface is the interface between the terminal and the network, and R (receive) and S (send) are reference points in the sending and receiving of messages (Figure 11.21).

In between GPRS networks is the Gp-interface, via which two separate and independent GPRS networks send messages to each other. The Gi-interface is in between the GPRS network and the fixed packet data network (PDN). The Gi-interface has not been defined in the specifications [36].

The GPRS service connects logically to the GSM system via two network entities: the GPRS support node (GSN) and the GPRS register. Figure 11.22 portrays the principles for the logical interface architecture. It can be seen from the figure that GPRS brings many new, partly optional, interfaces to the GSM system. The short message service has been specified to function with the GPRS service by defining a connection from the GPRS network to the short message service center (SM-SC). Short messages are transported in the GPRS network using the same MAP protocol as is used in the basic GSM network.

GPRS supports services based on standardized data protocols, and GPRS specifications define the connections to X.25 and IP-networks. In addition to common protocols, some GPRS-specific protocols are utilized within the GPRS network. These specifications do not determine which protocols are used for connections external to the GPRS network. Thus operators need to agree on the use of shared protocols in the traffic between the operators' networks.

Figure 11.22 also shows, that GPRS has been specified to function with GSM and UMTS networks. The functionalities of the networks do not differ in principle, and the GPRS backbone network can in fact function as a kind of a combining element between 2G and 3G networks, securing packet-switched services over GSM and UMTS networks.

One example of such cooperation could be the packet switched connection in an operator's network, which supports both GSM and UMTS systems. When UMTS-service is unavailable, but the GSM coverage is sufficient to set up a connection, the user can be offered a restricted data service. It may not be sufficient for real-time video transmission, but it makes sense to offer a restricted service rather than no service at all. GSM and UMTS networks support handover between systems during a connection. Thus once a user returns

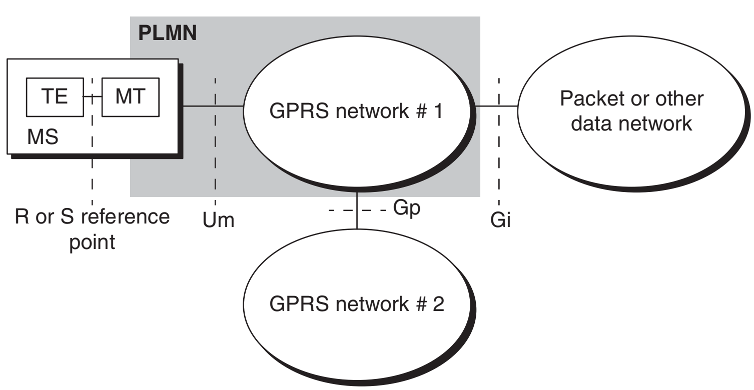

Figure 11.21 *The access points and reference points of GPRS.*

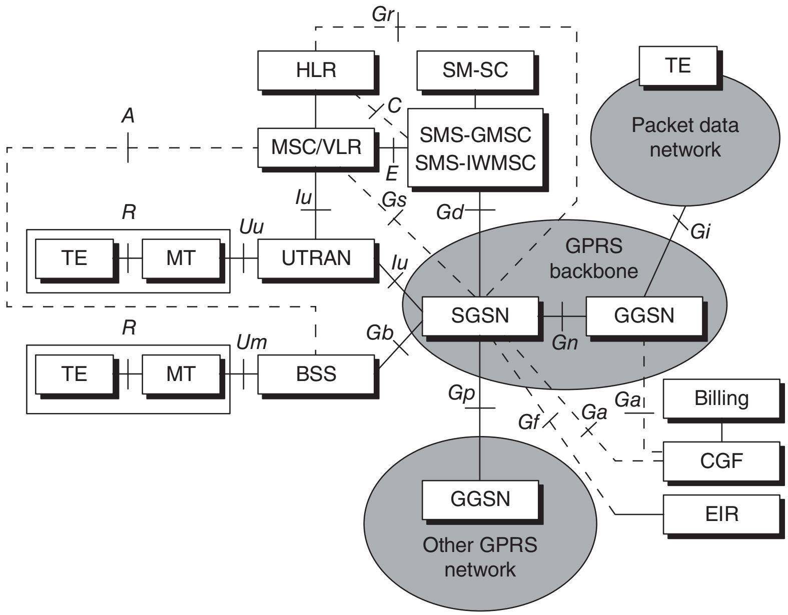

Figure 11.22 *The principle of the logical architecture of GPRS. The dotted line means signaling connection, whereas uniform line means combined signaling and data connection.*

to an area with actual UMTS coverage and enough resources are available, the connection is handed over to the higher bit rate UMTS- network.

Table 11.2 lists the most important interfaces in the GPRS system.

11.7.4 Special GSM Solutions

11.7.4.1 GSM-R

In the beginning of phase 2+ of GSM, it was noted that also railways would benefit in the internal GSM service. This idea resulted in a separate 4 MHz share below E-GSM frequency band (extended band) corresponding an additional 876–880 MHz frequency band in uplink and 921–925 MHz in downlink. This band is meant for the special needs of railways, including the delivery of the control data of the trains, added value services, automatic charging of passenger's train tickets, transport load and passenger statistics, among other information. In addition to ETSI, also International Union of Railways (Union Internationale des Chemin de Fers) was involved in the specification of this special variant of GSM.

GSM-R is a multinational mobile communications system, which is compatible with the control system of the railways. It includes, for example, voice service and data transfer services like HSCSD and GPRS. Logically, R-GSM is compatible also with the public variant of GSM.

As is the case in the public variant of GSM, also R-GSM is regulated by the national frequency authorities by granting the licenses and collecting possible frequency utilization fees.

11.7.5 Machine-to-Machine Communications

The importance of the machine-to-machine communications (M2M) via GSM is growing. In fact, as the more advanced networks are taking over gradually the voice and data traffic of GSM, the special solutions via GSM

Table 11.2 *The most important GPRS interfaces (i/f)*

i/f	Functioning
R	Interface between MT and TE. Via this interface, the functioning of the MT can be controlled by typing AT commands.
Um	Interface between GSM/GPRS terminal and GSM base station subsystem (BSS).
Uu	Interface between UMTS/GPRS terminal and UMTS radio access network (UTRAN).
Iu	Interface between UMTS radio access network (UTRAN) and SGSN element. UTRAN can be connected both to SGSN and to UMTS core network (CN) via the same *Iu*-interface.
Ga	Interface between GGSN (GPRS gateway support node) and CGF (charging gateway function) or between SGSN (serving GPRS support node) and CGF. The GPRS charging tickets can be collected both from GGSN and SGSN elements. There is further connection defined from CGF to operator's billing system, but the specifications do not define this part.
Gb	Interface between GSM base station subsystem (BSS) and SGSN element. The *A*-interface between BSS and MSC is meant for the circuit switched GSM connections, whereas PCU (packet control unit) separates the GPRS packet connections from BSS to *Gb* interface. *Gb* utilizes Frame Relay network in the beginning of the service.
Gd	Interface between SGSN and short message serving center. The mobile terminated (MT) short message is transferred via SMS-GMSC (SMS-gateway MSC) and the mobile originated (MO) short message is transferred via SMS-IWMSC (SMS-interworking MSC). There is further connection from the transferring MSC to the SM-SC (short message service center).
Gf	Interface between SGSN and EIR (equipment identity register). The working principle of EIR with GPRS is exactly the same as in basic GSM (see Chapter 2, GSM).
Gi	Interface between GGSN and external data network (PDN, packet data network). PDN can be either IP or X.25 packet data network. There are options for adding more protocols on this interface for connecting GPRS to other type of data networks.
Gn	Interface between SGSN and GGSN. The connection is defined according to IP protocol. The above laying network can be in operation with whichever one is suitable, for example, ATM or Ethernet.
Gp	Interface between SGSN and GGSN of external GPRS network. When external GPRS customers roam from their network, their connections can always be routed via their home GPRS network's GGSN. The other option is to route the connections via the roaming network's GGSN.
Gr	Interface between HLR and SGSN.
Gs	Interface between MSC/VLR and SGSN. This interface has been defined optional.

can be served still for a long time. Some examples of the M2M are the automatic delivery of the electricity consumption figures from home of the user, and management of the vending machines.

11.7.6 Energy Saving Functionalities

One important part of the fluent continuum that GSM provides during its remaining lifetime towards more advanced networks requires optimization for the capacity, and also for the energy consumption. There are various innovative solutions presented and under development that lowers the energy consumption of GSM, still maintaining the same quality of service level and capacity of GSM.

11.7.7 Smartphone Signaling Optimization

Along with the introduction of smart devices of 2G, 3G and 4G, the signaling level has increased dramatically. Not only the signaling caused by 2G, but the offloading of the increased data from 3G and 4G to be handled partially via 2G has caused new aspects to be taken into account in the optimal designing of the networks.

11.8 Dual Half Rate

Even if GSM is relatively old system – the first commercial networks were launched already in 1991 – the system has been highly useful due to its robustness, high level of network equipment and user device support, standardized interfaces, among other factors. As GSM is so widely spread system, it is a logical base for voice and basic data services still for years to come.

The GSM has further been developed in order to provide higher spectral efficiency. In emerging markets, GSM is the most useful system due to the low expenses of the network deployment and operation, as well as due to the low-priced user equipment. For the developed markets, as 3G and 4G systems will be growing as for the coverage and number of users in a fast time schedule, GSM can still be operated in order to provide a smooth transition from 2G to more spectral efficient network technologies, via multisystem devices.

The novelty GSM functionalities include capacity increase in the radio interface, for example, via OSC, and optimization of the balance of capacity and interferences, like in case of DFCA. The following chapters describe these functionalities in detailed level.

The capacity utilization is one of the most important optimization items of the GSM radio interface. The recently developed Dual Half Rate (DHR) functionality can be considered as a major step in the 2G evolution path. OSC (Orthogonal Sub Channel) is a solution for offering the pairing of two separate users into a single HR (Half Rate) time slot resource in such a way that no hardware enhancements are needed in the network side. OSC utilizes a SAIC (Single Antenna Interference Cancellation) functionality of the already existing handsets, which means that OSC provides a SAIC penetration dependent capacity gain as soon as it is activated via a software update of the Base Transceiver Stations (BTS) and Base Station Controllers (BSC) of the GSM radio network.

Based on Ref. [37], this chapter presents an extended analysis of the OSC gain as a function of SAIC handset penetration, and complements the analysis by investigating further the radio performance analysis presented in Ref. [38]. These results form a complete OSC gain model that consists of the predicted effect on the capacity and radio performance of GSM voice calls.

During the evolution of the circuit switched (CS) domain of GSM, there have been various capacity enhancement methods applied in the networks like Dynamic Frequency and Channel Allocation [39], Adaptive Multi Rate codecs [40] and Frequency Hopping (FH) [41]. Also several other solutions have been proposed, although these are still under development, like multiple beam smart antennas [42]. At the moment, though, one of the most concrete ways that provide with a notable capacity effect is the SAIC concept. There are various studies available presenting the respective benefit in different environments [43–46].

The first DHR deployments are based on the OSC functionality that has been evaluated and standardized by 3GPP GERAN [47–51]. It utilizes SAIC and provide capacity enhancement without need for modifications to the already existing new mobile station generation.

Despite of the availability of various studies about the SAIC gain, the capacity behavior in a typical OSC network deployment is not clear. This chapter presents thus an analysis of the effect of OSC on the offered capacity by investigating the benefits, for example, in terms of the utilization and possibility for the reduction of the transceivers as a part of the frequency refarming between GSM and other technologies such as HSPA (High Speed Packet Access) and LTE (Long Term Evolution). Based on realistic radio measurements in an indoor environment, this chapter also investigates the effective proportion of the coverage area where OSC can be utilized.

The GSM HR mode is selected as a reference for the investigations as it is assumed to be available in a typical environment. The capacity gain of OSC is obviously lower in the cell edge where the network dependent algorithm switches the OSC first to the HR mode and eventually to the FR (Full Rate) mode assuming that Adaptive Multi Rate (AMR) functionality is utilized.

11.8.1 The Functionality and Usability of OSC

GSM technology is globally most widely deployed cellular mobile communication system. In emerging markets, the increase trend of new-developed subscribers and consequent voice traffic explosion creates a great pressure on network operators, especially for those operators that need to provide service to a large population with only limited bandwidth. In mature markets, GSM frequencies are being refarmed to UMTS and LTE in order to provide a national-wide mobile broadband service with a limited investment and better quality; this action will squeeze GSM bandwidth and increase demand over GSM technology to achieve higher capacity and spectrum efficiency, so as to maintain the current traffic volumes and user experience. Furthermore, considering that the Average Revenue per User (ARPU) continues decreasing, most operators are facing the challenge of improving their hardware utilization efficiency.

In order to cope with the scenarios outlined above, some possible solutions are brought up. Among them, one of the most promising proposals is the OSC feature, a voice capacity enhancement for the GSM networks. OSC is standardized in 3GPP, and the first live network OSC trials have been carried out in 2010. OSC is backwards compatible solution with the original TDMA frame structure of GSM, exploiting the single 200 kHz frame bandwidth by multiplexing more users to its 8 physical time slots. The OSC DHR (Dual Half Rate) mode utilizes a single physical TSL for up to four users compared to the HR mode that can multiplex two users within the same TSL, or only one user in the FR (Full Rate) mode.

In Downlink (DL), OSC DHR distinguishes between two independent GSM users sharing the same HR resource by interpreting the 8-level GMSK modulation as two separate QPSK (Quadrature Phase Shift Keying) modulation diagrams in such a way that the users form pairs within the HR TSL as presented in Figure 11.23.

As presented in Figure 11.24, a MUD (Multiuser Detection) receiver is used in Uplink (UL) in order to identify two users sharing the same physical time slot, and the separation of the users is handled via different training sequences [47]. This requires antenna diversity in the GSM base station with MUD functionality

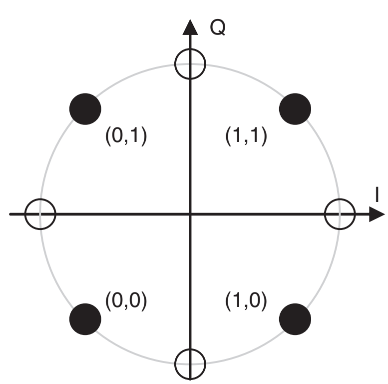

Figure 11.23 *Two separate OSC DHR users can be multiplexed in DL via the pairing based on the QPSK constellation.*

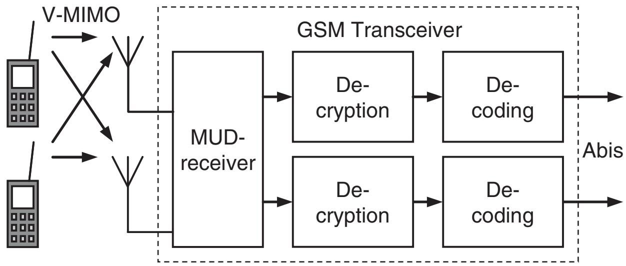

Figure 11.24 *Two OSC DHR users can share the same HR resource in UL via the MIMO functionality of MUD.*

[45]. The applied Multiple-Input, Multiple-Output (MIMO) concept gives benefits especially in Rayleigh fading channel [48]. A GMSK modulation is used in UL.

Figure 11.25 shows an example about the average behavior of the OSC DHR and HR calls in a fully occupied 8-TSL TRX assuming that the OSC penetration is 50% and that the users are uniformly distributed over the investigated area.

As Figure 11.25 indicates, the TSL division for the OSC and HR calls can be estimated to be 1:2 in case of 50% OSC penetration, that is, OSC utilizes 1/3 of the physical TSL resources in this case.

When the radio conditions are ideal, that is, there are no restrictions from the radio network side in the usage of the TSL for OSC DHR, HR or FR calls, it is thus possible to fit 2 times more OSC DHR calls into the same TSL compared to HR calls.

The following analysis shows the more detailed division of the TSL utilization for the OSC DHR and HR users as a function of the OSC penetration, assuming that the network supports both modes in the normal operation.

11.8.2 Effect of OSC on Capacity

This analysis shows the behavior of the GSM cell capacity when both GSM HR and OSC DHR users are found in the area. More specifically, the distribution of the HR and OSC users in the same cell is studied as a function of the OSC capable SAIC mobile station penetration. The capacity enhancement is shown between the extreme values, that is, when only HR users are found, and when the cell is populated with OSC users, in a fully loaded cell.

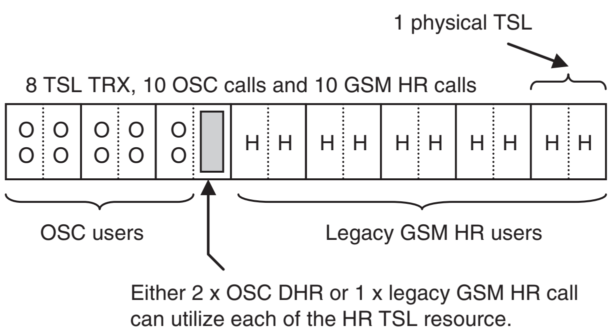

Figure 11.25 *An example of the utilization of the TRX resources for OSC DHR and GSM HR calls with the same OSC (SAIC) terminal penetration.*

In the scenarios presented in this chapter, it is assumed that the functionality of the OSC is ideal, that is, it is transparent for the existing HR traffic, and that there is no impact of the feature on the previous functionality of the network. It should be noted that this analysis concentrates purely on the radio interface time slot capacity, and does not consider the effects of the feature on the radio link budget. The assumption of these calculations is thus the presence of a coverage area where the quality and received power levels are sufficiently high for the functioning of the OSC.

The benefit of OSC is obvious in a fully loaded network where major part of the GSM Mobile Stations (MS) is OSC capable. The presented investigations are divided into three scenarios: The effect of OSC DHR on the TSL distribution in a fully loaded cell, on the number of users per TSL, and on the blocking rate, which can result the possibility to reduce TRX units from the GSM sites.

A fully loaded cell can be taken as a basis for the study. The assumption is that both HR and OSC capable users are uniformly distributed over the investigated area. This means that the OSC capable MS penetration equals to the probability of the OSC MS entering to the cell, the rest of the MSs entering to the cell utilizing the HR mode. It should be noted that the term HR refers to a MS that is capable of utilizing both HR and FR modes.

When the OSC feature is activated in the GSM network, the total number of users, that is, Mobile Stations that are distributed in the investigated area, is a mix of OSC DHR users and GSM HR users:

$$N_{MS}^{tot} = N_{MS}^{OSC} + N_{MS}^{HR} \tag{11.1}$$

The amount of the OSC users depends on the OSC capable MS penetration α with $\{\alpha \in \mathbb{R} | 0 \leq \alpha \leq 1\}$. The number of OSC capable users is thus:

$$N_{MS}^{OSC} = \alpha \cdot N_{MS}^{tot} \tag{11.2}$$

This means that the number of HR users is:

$$N_{MS}^{HR} = (1 - \alpha) \cdot N_{MS}^{tot} \tag{11.3}$$

The total number of physical TSLs is divided to the TSLs utilized by the OSC users and the HR users. The expression for the TSLs occupied by the OSC and HR users is thus:

$$N_{TSL}^{tot} = N_{TSL}^{OSC} + N_{TSL}^{HR} \tag{11.4}$$

The formula contains physical TSLs for OSC and HR. As a single OSC TSL can be utilized by a total of 4 OSC users, we can present the relationship between the TSL and the number of the users in the following way:

$$N_{MS}^{OSC} = 4 \cdot N_{TSL}^{OSC} \tag{11.5}$$

Equally, a total of 2 HR users can utilize the TSL, which can be expressed as:

$$N_{MS}^{HR} = 2 \cdot N_{TSL}^{HR} \tag{11.6}$$

Figure 11.26 clarifies the idea of the notation. In this specific case, the OSC user penetration α is 12/22 = 55%, and the physical TSL utilization for OSC users is 3/8.

Assuming that the users are distributed uniformly in the coverage area, the utilization of the TSLs depends on the OSC penetration. The following analysis represents the best case scenario of a number of HR MSs, which is multiple of 2 and a number of OSC MSs, which is multiple of 4 so that no TSL is occupied by less than 4 OSC MSs or less that 2 HR MSs.

By combining (11.4), (11.5) and (11.6), we can write:

$$N_{TSL}^{tot} = N_{TSL}^{OSC} + N_{TSL}^{HR} = \frac{N_{MS}^{OSC}}{4} + \frac{N_{MS}^{HR}}{2} \tag{11.7}$$

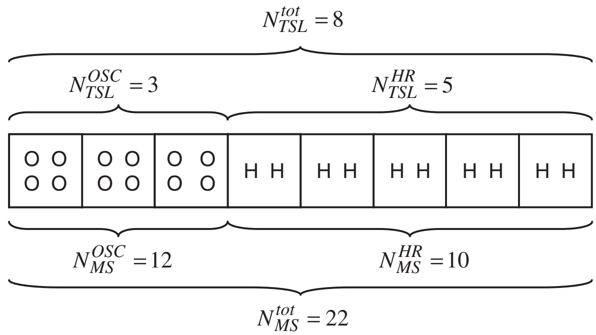

Figure 11.26 *An example of the utilization of a single TRX by OSC (marked as O) and HR (marked as H) users.*

We can then introduce the dependency from the OSC penetration using (11.2) and (11.3):

$$N_{TSL}^{tot} = \frac{\alpha \cdot N_{MS}^{tot}}{4} + \frac{(1-\alpha) N_{MS}^{tot}}{2} \tag{11.8}$$

We have therefore expressed the total number of needed TSLs as a function of the total number of MSs and the OSC penetration coefficient.

11.8.2.1 The Number of Users as a Function of α

Similarly, we can express the total number of MSs that can be served as a function of the total number of available TSLs and the OSC penetration coefficient.

Let's then introduce β with $\{\beta \in \mathbb{R} | 0 \le \beta \le 1\}$ which indicates the percentage of the physical TSLs occupied for the OSC users, the analysis can be carried further. Then we can write:

$$\begin{aligned} N_{MS}^{tot} &= N_{MS}^{OSC} + N_{MS}^{HR} \\ &= \beta \cdot N_{TSL}^{tot} \cdot 4 + (1-\beta) \cdot N_{TSL}^{tot} \cdot 2 \\ &= 2 \cdot N_{TSL}^{tot} \cdot (2\beta + 1 - \beta) \\ N_{MS}^{tot} &= 2 \cdot (1+\beta) \cdot N_{TSL}^{tot} \end{aligned} \tag{11.9}$$

We can rearrange as follows:

$$\begin{aligned} N_{TSL}^{tot} &= \frac{\alpha \cdot N_{MS}^{tot}}{4} + \frac{(1-\alpha) \cdot N_{MS}^{tot}}{2} \\ &= \frac{\alpha \cdot N_{MS}^{tot} + 2 \cdot (1-\alpha) N_{MS}^{tot}}{4} \\ N_{TSL}^{tot} &= \frac{(2-\alpha)}{4} N_{MS}^{tot} \end{aligned} \tag{11.10}$$

The total number of users can then be expressed as a function of the TSLs:

$$N_{MS}^{tot} = \frac{4 \cdot N_{TSL}^{tot}}{2 - \alpha} \tag{11.11}$$

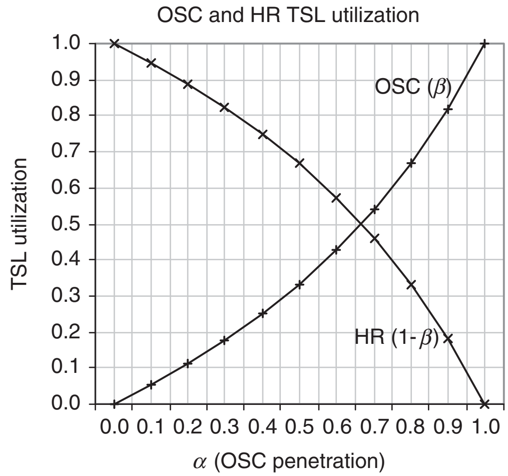

Figure 11.27 *The TSL utilization for the OSC and HR calls as a function of the OSC penetration.*

By comparing (11.9) and (11.11) we obtain:

$$\frac{4 \cdot N_{TSL}^{tot}}{2-\alpha} = 2 \cdot (1+\beta) \cdot N_{TSL}^{tot} \tag{11.12}$$

Therefore the utilization of the TSLs by the OSC users can be expressed as a function of OSC user penetration:

$$\beta = \frac{\alpha}{2-\alpha} \tag{11.13}$$

Figure 11.27 shows the physical TSL utilization for the OSC and HR users as a function of the OSC penetration, that is, the level of the TSL occupation as a function of the percentage of the SAIC capable handset of the all mobiles in the investigated area.

One point of interest of Figure 11.27 is the SAIC handset penetration of 50%, which indicates that the OSC paired connections utilize 33% of the physical TSL capacity of the cell, the rest being occupied by HR users. Another point of interest is the breaking point where both HR and OSC utilize the same amount of TSLs. The graph shows that it is found at 67% of OSC penetration.

It should be noted that the calculation is valid within the functional coverage area where OSC can be used, that is, where the received power level is high enough. For the areas outside the OSC coverage, HR and FR modes are assumed to function normally with their respective ranges of received power levels indicated in Ref. [52].

11.8.2.2 The Effect of OSC on the Number of Users per TSL

As the number of the OSC and HR users is known as a function of α, we can create a function that expresses the usage of a single TSL in terms of the number of users served. It is clear that the values of the served users oscillates within the range of Refs. [38, 40], lowest value indicating 100% HR, and highest value representing the 100% DHR OSC penetration.

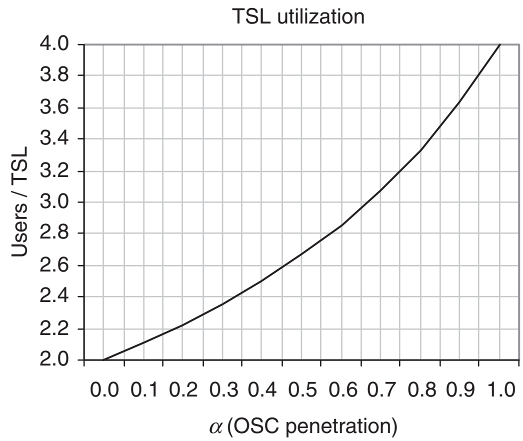

Figure 11.28 *The number of users per TSL as a function of the OSC penetration.*

The number of the users per TSL in a cell that contain mixed HR and OSC users can be obtained by utilizing (11.10) and (11.11):

$$\frac{N_{MS}^{tot}}{N_{TSL}^{tot}} = 2 \cdot \left(1 + \frac{\alpha}{2 - \alpha}\right) \tag{11.14}$$

Figure 11.28 shows the number of users per TSL as a function of the OSC penetration α. As can be noted, the extreme value of 4 is a result of OSC penetration of 100%.

As an example, the 50% OSC penetration level provides an offered capacity for about 2.67 users (which is a mix of OSC and HR users in average) per timeslot. It is again assumed that the OSC functionality can be utilized over the whole investigated area, meaning that the cell border area of the GSM coverage is not considered in this analysis.

11.8.2.3 The Effect of OSC on Blocking Rate

The original situation of the fully occupied cell with only HR users present, that is, when OSC is not activated, can be taken as a reference also for the following analysis. Assuming that the blocking rate of the GSM HR cell in busy hour is B, we can estimate its change via the well-known Erlang B [53] after OSC has been activated:

$$B(N) = \frac{\frac{A^N}{N!}}{\sum_{n=0}^{N} \frac{A^N}{n!}} \tag{11.15}$$

Term B is the blocking rate (% of the blocked calls compared to the number of whole attempts), N is the available amount of time slots, and A is the product of the average call density and average time of reservation, that is, the offered load.

As it is not possible to solve the equation of offered traffic analytically, it can be utilized in a recursive format:

$$\begin{cases} B(0) = 1 \\ B(N) = \dfrac{AB(N-1)}{N + AB(N-1)} \end{cases} \tag{11.16}$$

Furthermore, the offered traffic (Erl) can be expressed as:

$$\bar{x} = A \cdot (1 - B) \tag{11.17}$$

We can establish a reference case with OSC 0% in the following way. The $B(N)$[OSC = 0%] can be set to 2%. This means that, for example, in the case of 7 traffic TSLs for FR GSM, the respective offered load $A =$ 2.94 Erl and the offered traffic is $\bar{x} = 2.88$ Erl. The proportion of the offered traffic over the whole timeslot capacity $\bar{x}/\mathrm{TSL}_{\mathrm{tot}} = 0.41$ Erl/TSL. If HR codec is utilized, there is double the amount of users served within the same TSL number. This can be expressed in Erlag B formula by marking the available radio resources as 14, which results $A = 8.21$ Erl and $\bar{x} = 8.04$ Erl. The $\bar{x}/\mathrm{TSL}_{\mathrm{tot}}$ is now 1.15 Erl/TSL, that is, FR TSL. The difference between these figures shows the Erlag B plus HR gain, that is, the more there is available capacity, the more efficiently the calls can be delivered with the same blocking rate.

When the OSC functionality is activated and the same blocking rate is maintained, the available resources with the same hardware is now $4 \cdot 7 = 28$ for the $B(N)$[OSC = 100%], that is, when $\alpha = 1$. The $\bar{x}$ is now 19.75, and the efficiency 2.82 Erl/TSL, that is, FR TSL. This shows one of the benefits of OSC as it increases clearly the capacity efficiency via the Erlang B gain if the same amount of hardware is maintained.

If instead the amount of users is kept the same, the additional capacity that is liberated via the more efficient usage of TSLs via OSC can be removed partially or totally. This provides the basis for 3G refarming if the operator has both 2G and 3G licenses.

With a lower amount of timeslots, the same number of users can still be served with the same or lower blocking rate like Figure 11.28 indicates. The blocking rate depends on the usage of the TSLs for OSC, that is, on β, which can be expressed as a function of α as shown in (11.13).

11.8.2.4 The Effect of OSC on the TSL and TRX Reduction

It is possible to investigate the dependency of the number of users and Erlang B formula's offered traffic as a function of the OSC penetration. In order to carry out this part of the study, an Erlang B table was created by utilizing (11.15) and (11.16).

A case example with a blocking rate of 2.0% was selected. A table of 1–200 TSLs was created as a basis for the analysis. Figure 11.29 summarizes the behavior of the channel utilization. It can be seen that the performance of the cell increases due to the Erlang B gain, along with the OSC penetration growth compared to the original proportion of the HR capable users.

Next, an analysis of the effect of the SAIC handset penetration on the total reduction of the time slot number is presented. The calculation can be made by assuming that the offered traffic level $\bar{x}$ and the call blocking rate B are maintained in the original level.

Figure 11.30 summarizes the analysis carried out for 20–100 Erl of offered traffic. Figure 11.30 can be interpreted in such a way that when the OSC penetration grows, the needed number of TSLs with the same blocking rate (originally 2%) is now lower.

According to the basic behavior of the Erlang B model, the effect is logically strongest for the higher capacity cells. This can be noted especially from the highest investigated offered traffic class of 100 Erl, which requires only half of the timeslots when OSC penetration is 100%.

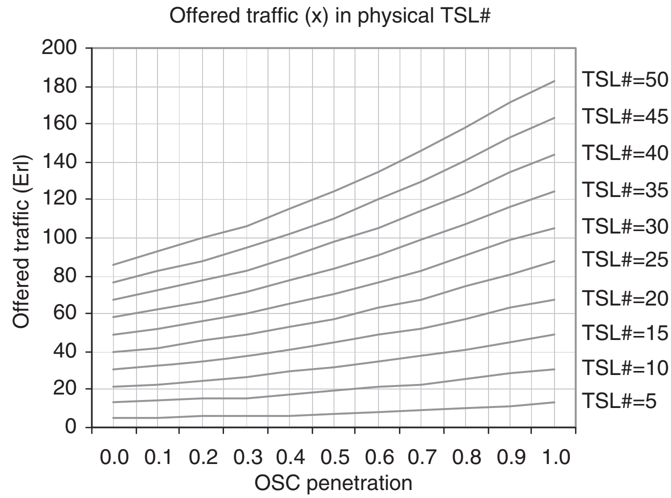

Figure 11.29 *Offered traffic as a function of the OSC penetration, when the maximum number of the available time slots is 50 and B = 2%.*

The effect of OSC can be seen in Figure 11.31 which shows the analysis for the required number of TRXs for a set of offered traffic classes (Erl) when the OSC capable SAIC handset penetration is known.

The assumption in this case is that each TRX contains 8 traffic channels. It is now possible to interpret the benefit of OSC in terms of the transceiver units. Table 11.3 summarizes the calculation as a function of TRX number.

Table 11.3 indicates that the benefit of OSC starts to be significant when the original BTS contains at least 4 TRX elements (per sector or in omniradiating site). In that case, already a relatively low OSC penetration

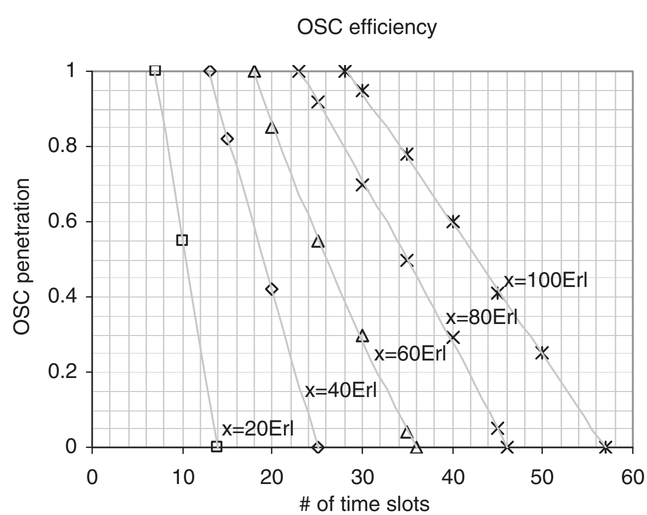

Figure 11.30 *The required time slot number with B = 2% can be observed as a function of the OSC capable user penetration.*

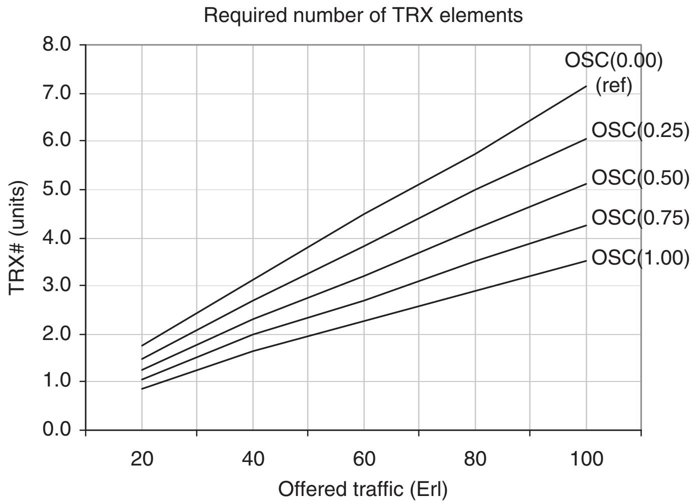

Figure 11.31 *The effect of the OSC penetration on the required amount of TRXs when the blocking rate B is kept the same (2%).*

of 25% provides a possibility to remove a complete TRX element (i.e., one frequency block of 200 kHz) from the cell, yet still keeping the blocking rate below or the same as was observed originally, that is, 2% in this example. If this capacity is removed, it can be utilized for the refarming of 3G frequencies. Alternatively, if no TRX units are removed, the offered capacity can be utilized more efficiently for parallel packet data via the lower circuit switched call blocking rate. There are thus different network evolution scenarios due to OSC.

Figure 11.32 summarizes the amount of TRXs (assuming that each offers 8 physical traffic TSLs) that can be removed from the site depending on the OSC penetration and the offered traffic in such a way that the blocking rate would not change from the original 2% figure. It should be noted that a small part of the TSLs are used also for signaling, and that each TRX unit is a physical 8-TSL equipment, so only integer values rounded up should be considered in the interpretation of the number of TRXs.

The presented analysis provides a base for the GSM network redimensioning in the following scenarios:

- In the case of still growing GSM traffic, the estimation of the already existing penetration of SAIC terminals when the OSC feature is activated, as well as the prediction of the SAIC penetration development within

Table 11.3 *The estimation of the needed TRX element number as a function of the OSC penetration, when the blocking probability B = 2%*

	OSC penetration				
$\bar{x}$ (Erl)	0.00	0.25	0.50	0.75	1.00
20	1.8	1.5	1.3	1.1	0.9
40	3.1	2.7	2.3	2.0	1.6
60	4.5	3.8	3.2	2.7	2.3
80	5.8	5.0	4.2	3.5	2.9
100	7.1	6.1	5.1	4.3	3.5

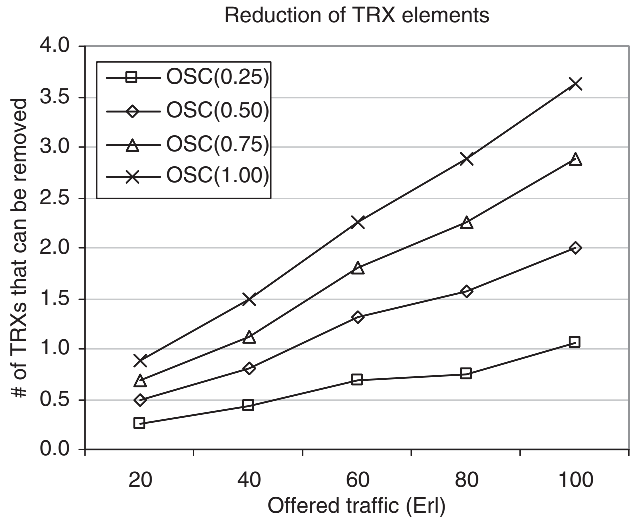

Figure 11.32 *The effect of OSC on the reduction of TRX elements.*

the forthcoming years gives a possibility to adjust the radio capacity plans. This means that the otherwise required TRX extensions can be postponed or rejected as long as the blocking rate can be maintained in the allowed level with OSC.

- In the case of stabilized or lower GSM traffic, the same blocking rate can be maintained with a lower amount of the TRX units. This liberated bandwidth can be reused for the delivery of the growing 3G traffic in the same frequency band, if available for the operator, which makes the frequency refarming more fluent with OSC.

11.8.3 OSC Radio Performance Analysis

11.8.3.1 Methodology for the Analysis

This chapter presents laboratory tests executed at Nokia Siemens Networks Innovation center in Madrid, Spain, 2011. The setup consisted of Base Station Subsystem (BSS) and Network and Switching Subsystem (NSS) with functional elements (Base Transceiver Station BTS, Base station Controller BSC, Mobile Switching Center MSC and Home Location Register HLR) as shown in Figure 11.33. The Abis interface was based

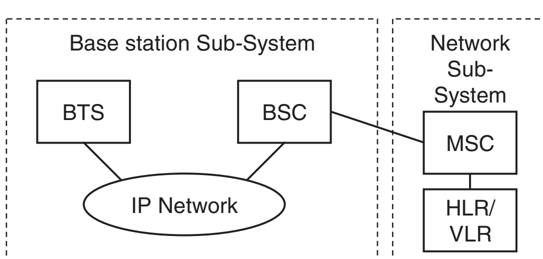

Figure 11.33 *The laboratory setup consisted of a functional GSM network with a single radio cell.*

Table 11.4 *Mapping of quality categories (Q) and Bit Error Rates (BER)*

Q_n category	BER (%)
$n = 0$	[0.0, 0.2]
$n = 1$	[0.2, 0.4]
$n = 2$	[0.4, 0.8]
$n = 3$	[0.8, 1.6]
$n = 4$	[1.6, 3.2]
$n = 5$	[3.2, 6.4]
$n = 6$	[6.4, 12.8]
$n = 7$	[12.8, 100]

on the IP emulation. The BTS was located in Madrid, Spain, and BSC in Finland. The voice calls were established only within the test network. Discontinuous Transmission (DTX) and Frequency Hopping (FH) were disabled during the measurement data collection.

The test network consisted of one cell in the indoors. The link budget was dimensioned by using adjustable attenuators and minimal TRX output power, which resulted in a cell radius of about 40 meters. The indoor radio propagation conditions provided nearly line of sight, with light wall obstacles. This setup represents a noise-limited environment within the whole functional received power level range of GSM.

For the evaluation of the radio performance, the quality level Q and received power level P_{RX} were correlated and stored in Slow Associated Control Channel (SACCH) frame intervals of 480 ms during the active voice calls. The correlated Q_n values with $n = 0 \ldots 7$ and P_{RXm} categories with $m = 0 \ldots 5$ were collected to a matrix. Table 11.4 presents the equivalence of the Q levels and bit error rate (BER), and Table 11.5 shows the P_{RX} categories in terms of dBm ranges [52].

The values for the Carrier-to-Noise ratio (C/N) are based on the thermal noise level and terminal's noise figure. The value for the noise floor can be estimated by applying the formula:

$$\begin{aligned} P_n &= 10\log_{10}\left(k_B TB \cdot 10^3\right) dBm \\ &= -174\ dBm/Hz + 10\log_{10}(B)\,Hz \end{aligned} \tag{11.18}$$

where k_B is Boltzmann's constant ($1.38 \cdot 10^{-23}$ J/K), T is the temperature (290 K is assumed) and B is the bandwidth in Hz. For a single GSM frequency channel of 200 kHz, the thermal noise level is thus

Table 11.5 *PRX categories and respective Carrier-to-Noise ranges*

P_{RXm} category	P_{RX} (dBm)	C/N (dB)
$m = 5$	[−38, −69]	[49, 80]
$m = 4$	[−70, −79]	[39, 48]
$m = 3$	[−80, −89]	[29, 38]
$m = 2$	[−90, −94]	[24, 28]
$m = 1$	[−95, −99]	[19, 23]
$m = 0$	[−100, −110]	[8, 18]

−120.98 dBm. The values of Table 11.5 contain also the receiver noise power of about 3 dB, resulting in total noise level N of −118 dBm.

The OSC functionality was set up to the BTS and BSC by utilizing noncommercial R&D software. DHR mode of OSC was utilized in a single TSL with minimally correlating training sequences.

11.8.3.2 Laboratory Tests with SAIC Handsets

The data collection was carried out over the functional coverage area by applying BSC measurement called RxLevel Statistics. SAIC terminals were used for the test calls. The collection was done in systematic manner by moving the terminals approximately 0.5 m/s with a movable test platform.

The correlated (Q, P_{RX}) matrix was collected for both OSC and reference GSM HR calls. It should be noted that the testing was done only for one physical TSL whilst the other traffic TSLs were blocked. As the environment was noise-limited and no inter-TSL interferences were present, this represents a fully loaded cell. The interval of the (Q, P_{RX}) matrices was 15 minutes.

Figure 11.34 shows the principle of the data collection, which was carried out in three different phases within high, medium and low ranges of received power levels. This method was applied to the investigations in order to collect the data over the P_{RX} categories in as uniform manner as possible.

11.8.3.3 Laboratory Results

The correlated (Q_n, P_{RXm}) results can be organized to a matrix format in such a way that the elements $v_{i,j}$ correspond to the number N of the samples of $Q_{8\text{-}i}$, where $i = 0,1,\ldots,8$, and $P_{RXj\text{-}1}$, with $j = 0,1,\ldots,5$, is indicated as $N(Q_n, P_{RXm})$. The result of (Q_0, P_{RX0}) represents thus the value of the matrix element $v_{8,1}$ and (Q_1, P_{RX1}) corresponds to the element $v_{7,1}$. The last result of (Q_7, P_{RX5}) is found in the element $v_{1,6}$. The following format clarifies the idea.

$$M = \begin{bmatrix} v_{1,1} & v_{1,2} & \cdots & v_{1,6} \\ v_{2,1} & \ddots & & \\ \vdots & & \ddots & \\ v_{8,1} & & & v_{8,6} \end{bmatrix} = \begin{bmatrix} N(Q_7, P_0) & \cdots & \cdots & N(Q_7, P_5) \\ N(Q_6, P_0) & \ddots & & \\ \vdots & & \ddots & \\ N(Q_0, P_0) & & & N(Q_0, P_5) \end{bmatrix}$$

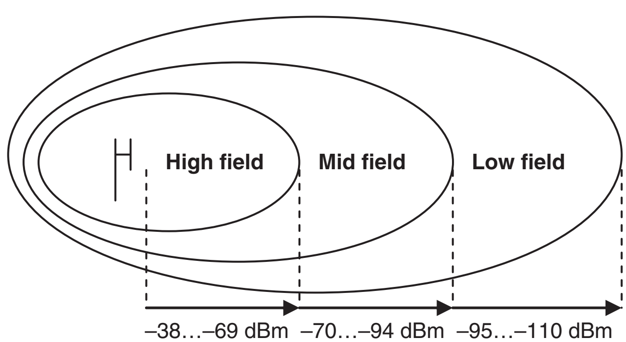

Figure 11.34 *The measurement data were collected in three phases, within high, mid and low field (values of PRX in dBm).*

The following matrices M_{GSM} and M_{OSC} present the normalized results for GSM and OSC laboratory tests, respectively. Each matrix element represents the percentage of the number of samples for that Q and P_{RX} compared to the total number of samples.

$$M_{GSM} = \begin{bmatrix} 1.70 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 2.67 & 0.02 & 0.02 & 0.01 & 0.00 & 0.00 \\ 3.57 & 1.31 & 0.95 & 0.39 & 0.08 & 0.02 \\ 3.81 & 1.96 & 0.71 & 0.23 & 0.02 & 0.03 \\ 2.97 & 1.69 & 0.58 & 0.30 & 0.04 & 0.03 \\ 1.11 & 1.34 & 0.63 & 0.32 & 0.03 & 0.03 \\ 0.66 & 0.66 & 0.18 & 0.23 & 0.06 & 0.02 \\ 1.81 & 5.33 & 10.35 & 14.13 & 11.47 & 28.56 \end{bmatrix}$$

$$M_{OSC} = \begin{bmatrix} 4.44 & 0.03 & 0.00 & 0.00 & 0.00 & 0.00 \\ 2.85 & 5.09 & 0.01 & 0.01 & 0.00 & 0.00 \\ 0.00 & 5.36 & 1.93 & 0.16 & 0.00 & 0.03 \\ 0.00 & 0.80 & 6.95 & 0.70 & 0.00 & 0.05 \\ 0.00 & 0.07 & 6.53 & 1.42 & 0.02 & 0.04 \\ 0.00 & 0.00 & 4.19 & 0.99 & 0.07 & 0.05 \\ 0.00 & 0.00 & 1.97 & 1.19 & 0.05 & 0.07 \\ 0.02 & 0.00 & 2.07 & 18.65 & 24.24 & 9.92 \end{bmatrix}$$

In the original GSM mode and with full load of a single TSL, a total of two HR users consume the whole physical TSL. In the OSC mode, in turn, a total of 4 DHR users utilize the physical TSL in the fully loaded TSL.

The Q_n and P_{RXm} results were collected in three phases, each lasting 15 minutes. One area is characterized as a strong field with main part of the results occurring in P_{RX} categories of P_{RX5} and P_{RX4}. The second represents medium field with values of P_{RX3} and P_{RX2}, and the third area represents low field with values belonging to P_{RX1} and P_{RX0}. The results were then combined in order to present the matrix of the complete field.

All the measurement data samples were collected as uniformly as possible by moving the platform in a constant speed. The data collection in the lowest field was done by observing the Radio Link Timer (RLT) parameter value directly in the mobile phone's engineering mode channel display. When the field reached the critical level, there occurred SACCH frame errors and the retransmission parameter value started to decrease from the original value of 20 towards 0. Before the zero-value was reached (which drops the call), the mobile was moved to the better field in order to raise the value back to 20 without breaking the connection.

There were a total of 20 383 and 19 399 samples collected during the GSM HR and OSC modes, respectively. In average, there is thus more than 400 samples per matrix element, which results in an average margin error Err [95%] = 0.98 / $\sqrt{(400)}$ = 4.9% with 95% confidence level per each matrix element. Table 11.6 shows the Err [95%] for each P_{RX} category.

Table 11.6 The error margin in % of the samples of each PRXm category with 95% confidence level

Mode	P_{RX0}	P_{RX1}	P_{RX2}	P_{RX3}	P_{RX4}	P_{RX5}
HR	1.6	2.0	1.9	1.8	2.1	1.3
OSC	2.5	2.0	1.4	1.4	1.4	2.2

11.8.4 OSC Radio Performance Model

Based on the analysis presented previously, a comparison of the matrices can be now performed. GSM HR matrix M_{GSM} can be selected as a reference in order to produce a matrix M_{diff}, which indicates the difference or change of the original correlated (Q, P_{RX}) distribution in %-units due to the OSC mode. The matrix is:

$$M_{diff} = M_{GSM} - M_{OSC} \tag{11.19}$$

$$M_{diff} = \begin{bmatrix} -2.74 & -0.03 & 0.00 & 0.00 & 0.00 & 0.00 \\ -0.18 & -5.07 & 0.00 & -0.01 & 0.00 & 0.00 \\ +3.57 & -4.05 & -0.98 & +0.23 & +0.08 & -0.02 \\ +3.81 & +1.16 & -6.25 & -0.46 & +0.01 & -0.02 \\ +2.97 & +1.61 & -5.95 & -1.12 & +0.01 & -0.01 \\ +1.11 & +1.34 & -3.56 & -0.67 & -0.05 & -0.02 \\ +0.66 & +0.66 & -1.79 & -0.96 & +0.01 & -0.06 \\ +1.79 & +5.33 & +8.28 & -4.51 & +12.77 & +18.65 \end{bmatrix}$$

It can be noted that in the original M_{GSM}, there are all quality classes $Q_0 \ldots Q_7$ present in the lowest field, that is, in P_{RX0} category, whilst M_{OSC} produces samples only to the quality classes Q_6 and Q_7 in the same field. It can be seen that the collection of samples for different P_{RX} levels is not uniform, indicating that the data collection happened slightly unequally in different fields, in addition to the fact that the ranges associated with each P_{RX} are different. This makes the comparison of the matrices challenging.

In order to cope with this issue, the result matrices can be normalized over each P_{RXm} category (that is by column) instead of the total number of the collected samples, in the following way:

$$M'_{GSM} = \begin{bmatrix} 9.29 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 14.6 & 0.13 & 0.12 & 0.03 & 0.00 & 0.00 \\ 19.5 & 10.7 & 7.07 & 2.48 & 0.71 & 0.05 \\ 20.8 & 16.0 & 5.27 & 1.49 & 0.13 & 0.09 \\ 16.2 & 13.7 & 4.34 & 1.92 & 0.31 & 0.11 \\ 6.08 & 10.9 & 4.69 & 2.05 & 0.22 & 0.09 \\ 3.63 & 5.40 & 1.35 & 1.45 & 0.49 & 0.05 \\ 9.88 & 43.3 & 77.16 & 90.6 & 98.2 & 99.6 \end{bmatrix}$$

$$M'_{OSC} = \begin{bmatrix} 60.7 & 0.26 & 0.00 & 0.00 & 0.00 & 0.00 \\ 39.0 & 44.8 & 0.06 & 0.06 & 0.00 & 0.00 \\ 0.00 & 47.2 & 8.15 & 0.68 & 0.02 & 0.34 \\ 0.00 & 7.04 & 29.4 & 3.01 & 0.02 & 0.48 \\ 0.00 & 0.65 & 27.6 & 6.13 & 0.10 & 0.43 \\ 0.00 & 0.04 & 17.7 & 4.29 & 0.30 & 0.48 \\ 0.00 & 0.00 & 8.32 & 5.14 & 0.20 & 0.72 \\ 0.27 & 0.00 & 8.75 & 80.7 & 99.36 & 97.5 \end{bmatrix}$$

Now, the difference can be calculated as:

$$M'_{diff} = M'_{GSM} - M'_{OSC} \tag{11.20}$$

The following matrix presents the result in %.

$$M'_{diff} = \begin{bmatrix} +51.5 & +0.26 & 0.00 & 0.00 & 0.00 & 0.00 \\ +24.4 & +44.7 & -0.05 & +0.03 & 0.00 & 0.00 \\ -19.5 & +36.5 & +1.08 & -1.80 & -0.69 & +0.28 \\ -20.8 & -8.91 & +24.1 & +1.53 & -0.11 & +0.39 \\ -16.2 & -13.1 & +23.3 & +4.22 & -0.21 & +0.33 \\ -6.08 & -10.8 & +13.0 & +2.24 & +0.08 & +0.39 \\ -3.63 & -5.40 & +6.97 & +3.68 & -0.28 & +0.67 \\ -9.61 & -43.3 & -68.4 & -9.90 & +1.21 & -2.07 \end{bmatrix}$$

It can be seen from the matrices that the proportion of quality classes Q_7 and Q_6 is higher in OSC than in GSM HR.

Figure 11.35 shows the impact of OSC on the quality level per each P_{RX} category. The most affected categories are P_{RX0}, P_{RX1} and P_{RX2} as they include lower quality values.

Figures 11.36 and 11.37 show the difference in the Q class behavior. In GSM HR, the lowest P_{RX} has samples over all the Q classes whilst OSC only causes samples for Q_6 and Q_7. The introduction of OSC feature causes each CDF to shift towards the higher Q classes corresponding to higher BER. This indicates that the useful coverage area produced via the OSC mode is smaller than the one produced by HR.

11.8.4.1 Scenario 1: Difference of the Radio Quality in a Fully Loaded GSM HR and OSC Cell

Table 11.7 summarizes the laboratory test analysis in CDF, showing the differences between HR and OSC per P_{RX} category. As can be noted, the Q_5, that is, BER in the interval [3.1%, 6.4%] can be reached with HR between P_{RX0} and P_{RX1} [−95, −110 dBm]. According to Table 11.7, the OSC moves the *C*/*N* requirement higher, and the Q_5 can now be found between P_{RX1} and P_{RX2} [−90, −99 dBm].

It can be assumed that the practical threshold for the GSM call is at quality level of Q_4, corresponding to a BER in the range [1.6, 3.2]. Figure 11.38 shows the CDF of the combined P_{RX} classes as a function of the Q, indicating that the Q_4 is at 89.5% for GSM HR, and at 80.0% for OSC over the whole investigated area A_{GSM} in this specific case.

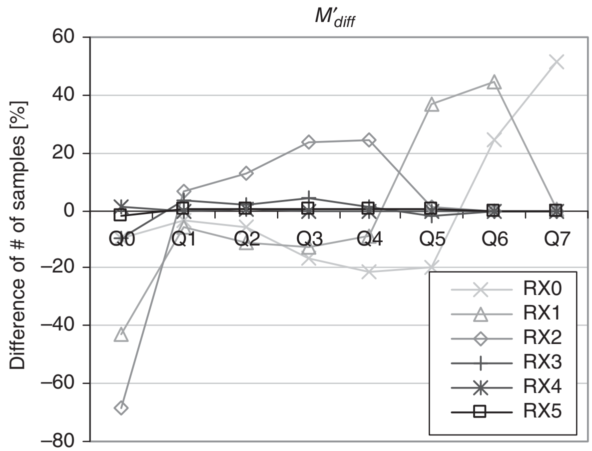

Figure 11.35 *The M′diff presents the change in the distribution of the correlated Q and PRX caused by OSC when HR is the reference.*

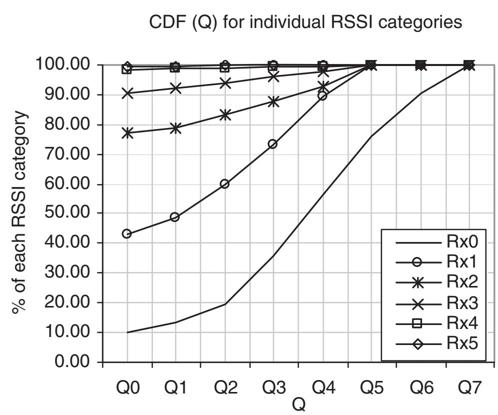

Figure 11.36 *CDF of the GSM HR analysis, scaled individually for each PRX category.*

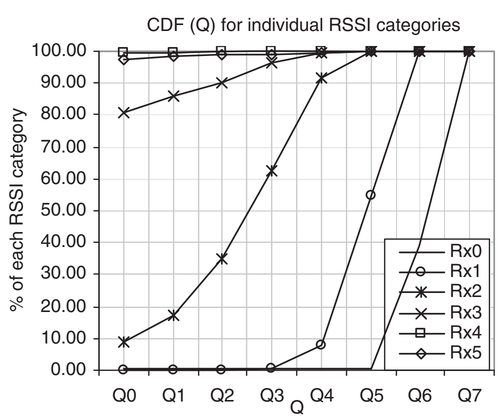

Figure 11.37 *CDF of the OSC analysis, scaled individually for each PRX category.*

Table 11.7 *The 95% Q criterion analysis for the GSM HR calls*

P_{RX} category	Q value for GSM HR (reference)	Q value for OSC / difference with GSM HR (%-units)
P_{RX0}	6.5	6.9 / −0.4
P_{RX1}	4.5	5.9 / −1.4
P_{RX2}	4.3	4.4 / −0.1
P_{RX3}	2.5	2.8 / −0.3
P_{RX4}	0.0* (98.1%)	0.0* (97.5%) / 0.0 (−0.6%-units)
P_{RX5}	0.0* (99.6%)	0.0* (99.4%) / 0.0 (−0.2%-units)

Note (*): When the Q_0 value is achieved more than 95% of the time, the corresponding %-value is shown in brackets.

The CDF of HR and OSC indicates the presence of different Q classes during the measurement, which shows the percentage of Q classes in the investigated area. The effect of OSC can now be observed based on Q_4:

$$\begin{aligned} A_{change}\,[\%] &= 100\frac{A_{HR} - A_{OSC}}{A_{HR}} \\ &= 100\left(1 - \frac{A_{OSC}}{A_{HR}}\right) \approx 10.6\% \end{aligned} \tag{11.21}$$

where A_{change} indicates how much the useful coverage area of OSC is smaller compared to the HR mode. The radius for the OSC mode is reduced from the HR mode by:

$$\begin{aligned} r_{change}\,[\%] &= 100 \cdot \frac{r_{HR} - r_{OSC}}{r_{HR}} = 100 \cdot \left(1 - \frac{r_{OSC}}{r_{HR}}\right) \\ &= 100 \cdot \left(1 - \sqrt{\frac{A_{OSC}}{A_{HR}}}\right) \approx 5.5\% \end{aligned} \tag{11.22}$$

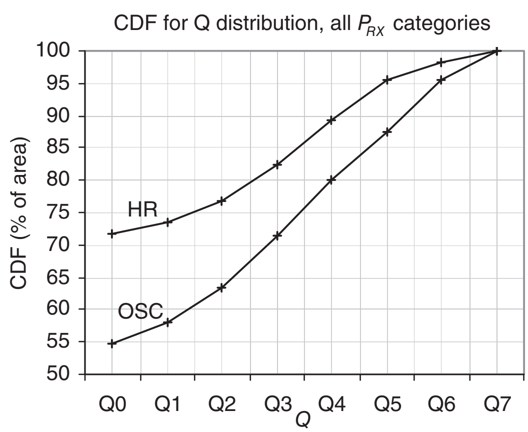

Figure 11.38 *The HR and OSC coverage.*

As the OSC brings up to 100% capacity enhancement compared to the GSM HR, the benefit of OSC is clear even with this level cell size reduction.

In practice, the original cell size remains as determined by the GSM FR and HR limits for the *C/N* and *C/I* levels. In other words, when the OSC feature is activated, the cell contains OSC, HR and FR coverage regions. The FR proportion remains the same also after the activation of OSC, but HR region is divided into OSC and HR regions. The final size of these depends mainly on the original overlapping portions of the neighboring cells and the criteria of the intra/intercell handover algorithms.

Furthermore, the OSC proportion depends also on the time slot usage by OSC, β, which in turn depends on the OSC capable handset penetration α.

11.8.4.2 Scenario 2: Only OSC (SAIC) Handset Penetration is Known

In a practical GSM network, there is a certain penetration of OSC capable terminals in the investigated field. When OSC is activated, there are legacy terminals that are capable of functioning only with the previous GSM codecs (FR, HR, AMR), and SAIC terminals that are also capable to function in the DHR mode.

It is thus important to take the SAIC terminal penetration into account when modeling the effect of OSC in the field. The M'_{GSM} and M'_{OSC} presented previously show the correlated distribution of (Q, P_{RX}) in a fully loaded situation, that is, when all the time slots are occupied either by GMS HR or OSC users. We can still utilize this assumption of a fully loaded cell in order to find the limits of the effect. Let's assume the SAIC handset penetration is α of all the terminals in the investigated area. As the OSC DHR utilizes single TSL for a total of 4 users whilst GSM HR can multiplex two users in the same TSL, the TSL capacity related scaling factor is needed for the performance model.

In a typical case, there are 2 or more TRXs per cell in suburban areas, and 4 or more in dense city environment. The TSL utilization of the OSC vs. HR can be estimated by α. The TSL utilization factor β for the OSC TSL occupancy in a fully loaded cell as a function of the OSC mobile terminal penetration can be formulated as presented in Ref. [37]:

$$\beta = \frac{\alpha}{2 - \alpha} \tag{11.23}$$

Now, when M'_{GSM}, M'_{diff} and OSC penetration α are known, the M'_{OSC} can be obtained by scaling the original GSM matrix element-wise. We can donate the elements of the matrix M'_{GSM} as $v^{GSM'}_{m,n}$ and the elements of the matrix M'_{OSC} as $v^{OSC'}_{m,n}$. The scaling between these corresponding elements can be assumed to be linear according to the principle of the linear interpolation function:

$$y = kx + b, \tag{11.24}$$

where $\{k \in \mathbb{R} | 0 \leq k \leq 1\}$ and represents the coefficient in x-axis of the physical TSL usage, that is:

$$k = v^{OSC'}_{m,n} - v^{GSM'}_{m,n} \tag{11.25}$$

Term b is the value of y when $x = 0$, that is, it equals to $v^{GSM'}_{m,n}$. Term x represents the usage of the TSLs for the OSC:

$$x = \beta \tag{11.26}$$

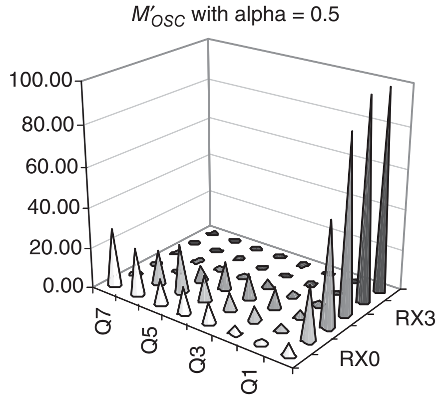

Figure 11.39 An example of the expected (Q, PRX) distribution (normalized over each RXlevel category) when OSC user penetration α is 50% and the OSC TSL usage B is 37.5%.

The scaling of each element for the partially loaded OSC cells can thus be done as follows:

$$y_{m,n} = \left(v_{m,n}^{OSC'} - v_{m,n}^{GSM'}\right) \cdot \beta + v_{m,n}^{GSM'} \tag{11.27}$$

M'_{GSM} and M'_{OSC} obtained from the laboratory can be utilized directly, and the estimate of the new M'_{OSC} (α) can be thus constructed by interpolating linearly the matrix values element-by element basis, as shown in Figure 11.39.

The usage of the laboratory results is most accurate in environment that contains approximately same type of radio channel, that is, in noise-limited environment with almost line-of-sight and light proportion of Rayleigh fading. If the radio channel differs considerably from the laboratory, Scenario 3 should be considered instead. This means that the case results presented in this chapter might vary depending, for example, on the level of the co-channel and adjacent channel interferences as well as on the type of multipath propagation characteristics in the investigated area.

11.8.4.3 Scenario 3: OSC (SAIC) Handset Penetration and New M'_{GSM} are Known

When the new M'_{GSM} is known, showing the correlated (Q_n, P_{RXm}) matrix for HR calls, it can be assumed that the new M'_{OSC} is possible to construct by applying the M'_{diff} that was obtained from the laboratory. The assumption is that the OSC feature performance is independent of the radio conditions, that is, the reference M'_{GSM} obtained from the laboratory already contains the radio channel related performance, whilst M'_{OSC} includes this same radio performance effect and an additional OSC performance specific performance. Assuming this is applicable in varying radio conditions, the new M'_{OSC} can be constructed by taking into account the OSC penetration:

$$M'_{OSC} = M'_{GSM} + \beta \cdot M'_{diff} \tag{11.28}$$

As explained in scenario 2, the activation of OSC is seen within the HR coverage area, the FR area staying unchanged. The utilization of the OSC compared to the original HR region can now be interpreted from the practical measurements or by taking again the Q_4 criterion as a basis. Also another value of the quality classes can be utilized, if that corresponds to the OSC pairing and unpairing criteria of the OSC algorithm.

11.8.5 Complete OSC Model

The strength of the developed model is that it requires only few and basic inputs, which are: (1) Rx Level Statistics tables before the OSC functionality is activated; (2) estimation of the OSC capable penetration in the initial phase of the OSC activation; and (3) optionally the percentage of the utilized HR and FR codecs.

11.8.5.1 OSC Model Process

Figure 11.40 shows the complete process of the model. The model processes the input data in such a way that the expected (Q, P_{RX}) matrix is formed based on the measured and correlated (Q, P_{RX}) table element-by-element according to (27). By default, the scaling of each element can be done by utilizing the already formed difference matrix. It should be noted that the presented difference matrix is valid for the noise-limited environment, so new difference matrix might be needed for the interference-limited environment.

The output of the model indicates the proportion of the OSC usage as a function of the OSC capable handset penetration, with the capacity gain that can be expressed in terms of increased offered load/traffic or of the possibility to reduce TRX elements. The radio performance result is based on the analysis of the changed *C/I* or *C/N* distributions.

When the complete model is applied in the investigation of a realistic GSM network, the steps described in Figure 11.40 can be taken into account for solving first the impact on the radio performance, that is, on the changes of quality.

The actual capacity gain depends on the radio performance, that is, which proportion of the GSM call can be utilized for the OSC, and on the OSC penetration, that is, what share of the handsets can take the advantage of the usable coverage area for the OSC. The estimate of the performance change due to the activation of OSC would be made based on the preformed difference matrices and the SAIC penetration.

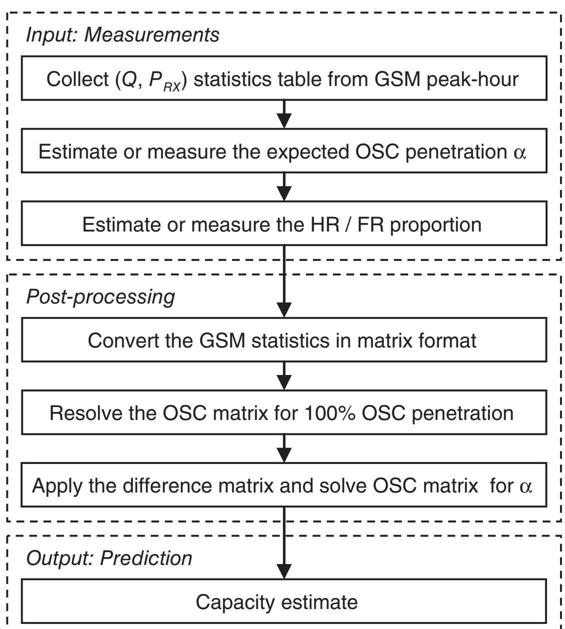

Figure 11.40 *The process chart of the OSC model usage.*

When the Rx level statistics table of GSM is measured from the field (i.e., the table without OSC feature activated), it can be assumed that the corresponding expected Rx level statistics table for the OSC is possible to construct by applying the correction curves that were obtained from the laboratory as shown in Figure 11.35. The assumption is that the OSC feature performance is independent of the radio conditions, that is, the reference table of GSM obtained from the laboratory already contains the radio channel related performance, and that the OSC table includes this same radio performance effect and an additional OSC performance specific performance. Assuming this is applicable in varying radio conditions, the new OSC table can be constructed by taking into account the OSC penetration:

$$M'_{OSC} = M'_{GSM} + \beta \cdot M'_{diff}, \tag{11.29}$$

where M'_{OSC} is the matrix format for the new OSC Rx level statistics table, M'_{GSM} is the matrix of the measured GSM Rx levels statistics table in the field, and M'_{diff} is the correction matrix obtained from the laboratory measurements. The practical way of constructing the OSC Rx level table is to utilize the β factor element-by-element basis for scaling first each of the difference matrix elements. Then, the difference matrix is utilized to scale the GSM Rx level statistics table element-by-element basis.

11.8.5.2 Process Steps

The complete OSC model considers the radio performance, or changes of the performance due to the OSC activation, as well as the resulting capacity gain. The steps for the investigation of the OSC impact are:

Step 1: Storing of the Rx Level statistics measurement and codec utilization measurement via BSC. This gives the reference for the GSM performance without OSC.

From the codec utilization measurement, also the estimate of what is the utilization (percentage) of HR and FR codecs can be included. That information gives the equivalence of the original HR-FR division also areawise as shown in Figure 11.41.

The investigation of the utilization of the codecs can be done for any time of the traffic, but the peak hour as criterion is recommended in order to collect as much data as possible in the given time window, and to make sure that the behavior of the effect of OSC on the capacity is done in the extreme conditions that represents the practical limit for the gain. In order to increase the accuracy of the estimate, the Rx level statistics measurement for the storing of the correlated Q and P_{RX} table can be done at the same time as the codec division.

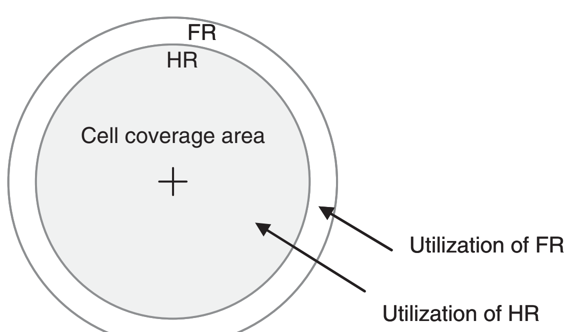

Figure 11.41 *The original division of the FR and HR mode utilization before the OSC feature has been activated.*

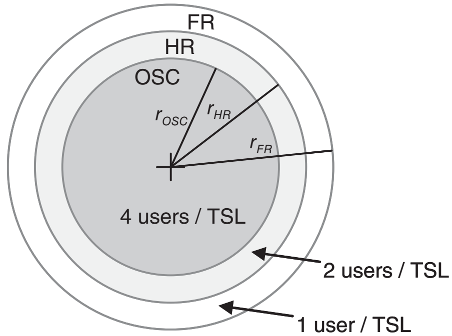

Figure 11.42 *When OSC feature is activated, there will be functional areas for OSC DHR, HR and FR in the cell, each providing different capacity performance.*

Step 2: Estimation of the OSC capable SAIC handset proportion in the field, that is, the factor α. The estimate can be carried out via the network statistics that is based on an IMEI (International Mobile Equipment Identity) analysis which is correlated with the database of the models that support SAIC. The estimate can also be done based on the sales statistics, or other practical "best-effort" estimate. Based on this information, the average TSL utilization for the OSC capable handsets can be estimated by using (23).

Step 3: Estimation of the capacity regions assuming the OSC DHR functions within a part of the HR region. The FR region can be assumed to work unchanged, as the situation was before the OSC activation. The criterion for the OSC percentage can be selected as presented in Figure 11.38, that is, based on the Q_4 level in CDF of the area. If the practical implementation of the OSC pairing and unpairing is based on the other Q levels, that can be utilized instead.

Step 4: Capacity estimate for the OSC and HR regions, as shown in Figure 11.42, and estimate of each region proportions. The division can be estimated by the HR and FR codec utilization statistics of BSC, U_{HR} indicating the utilization for HR (%) and U_{FR} indicating the utilization for FR (%). This division gives the first type of indicator for the capacity gain.

11.8.5.3 OSC Utilization

During the laboratory measurements described above, the codec utilization measurement was not yet activated. In practice, the voice codec statistics measurement can be activated in BSC in the same manner as (Q, P_{RX}) statistics in order to obtain the realistic division of the HR and FR utilization. In this analysis, we can assume case values of $U_{FR} = 10\%$ and $U_{HR} = 90\%$. The comparison is done by calculating first the reference situation, that is, the total capacity utilization via HR and FR, by giving a weight of $\delta_{HR} = 2$ (users/TSL) for HR (δ_{HR}) and 1 (users/TSL) for the FR (δ_{FR}):

$$\begin{aligned} U_{GSM}^{tot} &= \delta_{FR} \cdot U_{FR} + \delta_{HR} \cdot U_{HR} \\ &= 1 \cdot 10\% + 2 \cdot 90\% = 180\% \end{aligned} \tag{11.30}$$

The reference value is for full FR utilization, which results 100%. The HR mode as such gives thus 80% capacity gain in this specific example. When OSC is activated, the utilization of OSC and HR can be estimated accordingly, and the new capacity utilization can be thus calculated by:

$$U_{OSC}^{tot} = \delta_{FR} \cdot U_{FR} + \delta_{HR} \cdot U_{HR} + \delta_{OSC} \cdot U_{OSC}, \tag{11.31}$$

where the weight for OSC users (δ_{OSC}) is 4 (users/TSL). Assuming the utilization is still 10% for FR, we get the new division between the HR and OSC utilization by observing the Q_4 criterion and CDF as shown in Figure 11.38. The respective division between the HR and OSC regions can be thus solved.

In this specific case of the laboratory results, by observing the Figure 11.38 and respective result via (30) and (31), the OSC utilization is 100 − 10.6 = 89.4 (%) compared to the original HR area. The remaining 10.6 (%) share of the area represents thus the new HR region. The total utilization of the capacity when OSC is activated is thus:

$$U_{OSC}^{tot} = 1 \cdot 10\% + 2 \cdot 10.6\% + 4 \cdot 79.4\% = 348.8\% \tag{11.32}$$

The original situation without OSC presenting a 100% reference, this means that the capacity gain of Dual Half Rate OSC is 248.8% compared to the presence of only FR mode, and 348.8 · 100/180 − 100 = 93.8% compared to the presence of both, FR and HR modes.

It is possible to investigate further the dependency of the number of users and Erlang B formula's offered traffic as a function of the OSC penetration as shown in Figure 11.29, which shows the principle of the method with an example of 2% blocking rate. Other blocking rates are possible to be used as a basis for the calculations based on the statistics collected from the network. In order to estimate the capacity gain correctly, the investigation of the blocking rate of the network area of interest during the busy hour is thus needed.

Figure 11.29 summarizes the behavior of the channel utilization. It can be seen that the performance of the cell increases due to the Erlang B gain, along with the OSC penetration growth compared to the original proportion of the HR capable users.

11.8.5.4 TRX Reduction Gain

The calculation for the reduced GSM resources can be made by assuming that the offered traffic level and the call blocking rate are maintained in the original level. Figure 11.30 summarizes the analysis carried out for 20 … 100 Erl of offered traffic when the blocking rate is 2%.

As can be seen from Figure 11.30, along the growth of the OSC penetration, the needed number of TSLs with the same blocking rate (originally 2%) gets lower. According to the behavior of the Erlang B model, the effect is logically strongest for the higher capacity cells. This can be noted especially from the highest investigated offered traffic class of 100 Erl, which requires only half of the timeslots when OSC penetration is 100%.

Two cases can be constructed based on the offered traffic behavior of GSM as a function of OSC penetration. As a first case, the offered capacity can be maintained the same, which means that the blocking rate for users will be lower (gain of lower blocking rate). This case applies also to the situation where future TRX expansions are planned they can be postponed until the original (or separately decided new) blocking rate is achieved.

If, instead, the offered capacity is lowered by keeping the blocking rate the same, we can estimate the OSC capacity gain in terms of the savings in the TRXs. This TRX reduction can be made based on Table 11.8, by taking into account that each TRX element should be informed as integer number that is rounded up (containing always 8 physical TSLs). Again, the blocking rate of 2% (or lower) is utilized as the criterion. The possibility to reduce the TRX elements can be utilized for the additional capacity for 3G and 4G.

Table 11.8 indicates the capacity gain obtained as a function of the OSC penetration in terms of the TRX element reduction within the functional area of OSC. This case applies to the refarming of the 2G, 3G and 4G frequencies.

The reduced need for the GSM bandwidth in order to still deliver the original 2G traffic with an unchanged quality of service level may provide the possibility to add a new UMTS carrier of 5 MHz [54] to the same band. As an example, if the original GSM band is of 10 MHz (50 channels of 200 kHz each), according to Figure 11.31, the 80 Erl traffic can be offered with half of the original amount of TRXs when the SAIC

Table 11.8 The estimation of the number of TRX elements that can be removed as a function of OSC penetration, when the blocking probability B=2%

	OSC penetration				
$\bar{x}$ (Erl)	0.00	0.25	0.50	0.75	1.00
20	—	—	—	—	1
40	—	1	1	2	2
60	—	1	1	2	2
80	—	1	1	2	3
100	—	1	2	3	4

handset penetration is near 100%. This means that the original GSM traffic could be possible to deliver within 5 MHz band still maintaining approximately the same blocking rate, which leaves sufficiently space for adding a complete UMTS carrier as a parallel solution.

The benefit can also be seen with LTE, which provides more freedom to select the bandwidth, compared to the fixed 5 MHz band of UMTS. The narrowest LTE bands, that is, 1.4, 3 and 5 MHz can be utilized to the gradual increasing of LTE, when first, GSM traffic can be offered in smaller band by OSC, and when GSM traffic eventually lowers.

By applying the model and reference measurements for the quality effect of OSC in noise-limited environment, the optimal site configurations can be achieved. According to the results presented in this chapter, the sites that contain at least 4 TRXs with approximately 2% blocking rate during the peak-hour, can benefit from the activation of OSC in such a scale that the 4-TRX cell with OSC penetration of about 25% would give possibility to remove one complete TRX, the blocking rate still being at the same or lower 2% level. Alternatively, if the SAIC penetration is relatively high, in order of 75%, one TRX element can be removed even from a 3-TRX cell without impacts on the blocking.

In the live network, the final effects of OSC depend on the proportion of the overlapping cells, co-channel interference levels, intercell handover algorithms and their parameter values. In any case, the presented analysis gives indication about the behavior of OSC when the OSC-paired HR calls, which can fit a maximum of four users to a single TSL, are switched back to HR mode that allows two users to a single TSL, or to FR mode that occupies the whole TSL for a single connection. The presented model shows how to estimate the proportion of these regions, and what the effect of OSC is on the final capacity compared to the network that supports only the basic mode of FR/HR.

11.9 DFCA

It has been shown that the performance of GSM can be enhanced by the utilization of OSC (Orthogonal Sub Channel) and Dynamic Frequency and Channel Allocation (DFCA) functionalities. So far, the effect of OSC and DFCA on the radio performance and capacity gain has been investigated separately. This chapter shows the performance enhancement when both of the functionalities are utilized at the same time, which results in the higher hardware efficiency and capacity gains. The studies are based on the analysis of the simulations and performance measurements statistics collected from real networks.

11.9.1 Dynamic Frequency and Channel Allocation Principle

DFCA is based on the dynamic assignment of the radio channels – including TSL, Mobile Allocation (MA) list and Mobile Allocation Index Offset (MAIO) – for the incoming circuit switched calls. DFCA is thus a

Radio Resource Management (RRM) functionality of the base station subsystem of the GSM, which results in the dynamic and optimal selection of the radio channel that has sufficiently good, but not unnecessarily high quality level in terms of C/I value, so that each connection fulfil the Quality of Service (QoS) requirements. The algorithm is located to the Base Station Controller (BSC), and it takes care of the radio channel assignments for all DFCA capable Transceiver Units (TRX) that are located within its area.

The DFCA algorithm utilizes MS (Mobile Station) measurement reports in such a way that DFCA dynamically adapts to the varying interference levels. It can be assumed that DFCA is most useful in the interference limited case.

In order to function correctly, DFCA is based on the network synchronization. The functionality utilizes C/I estimations derived from the MS measurement reports, knowledge of time slot and frequency usage in serving and surrounding cells, and information about the used UL and DL power levels. Channel allocation info and statistical C/I data are thus exchanged between the BSCs, which requires a transmission based on the signaling, for example, over an IP network.

The random Frequency Hopping (FH) that has been available since the early days of GSM, is able to spread the potential interferences over a fixed list of hopping frequencies which lowers the peak interference level. DFCA, instead, utilizes cyclic FH over individually selected frequency lists and MAIOs for each connection thanks to accurate C/I estimation. Each connection is assigned with the most suitable radio channel described by the MA list, MAIO, TSL and Training Sequence Code (TSC).

11.9.2 Joint OSC and DFCA Performance

For the estimation of the expected OSC DHR penetration in the field, a SMART system level simulator was used. The following analysis and simulator are described more detailed in Ref. [55]. The following factors were considered: (1) SAIC penetration; (2) HR penetration; (3) RXQUAL DL of both OSC calls; (4) RXQUAL UL of both OSC calls; (5) DL received signal strength of both OSC calls; (6) UL received signal strength of both OSC calls; and (7) UL received signal strength difference between two OSC calls. The joint probability P_j of the OSC DHR penetration can be estimated with the following formula:

$$\begin{aligned} P_j = \alpha^2 \cdot \beta^2 \cdot P\big(&Q_1^{DL} < T_1, Q_2^{DL} < T_1, Q_1^{UL} < T_1,\\ &Q_2^{UL} < T_1, L_1^{DL} \geq T_2, L_2^{DL} \geq T_2, L_1^{UL} \geq T_2, L_2^{UL} \geq T_2,\\ &\left|L_1^{UL} - L_2^{UL}\right| \leq T_3\big) \end{aligned} \tag{11.33}$$

In (11.33), α is the SAIC penetration, β is the HR penetration, T_1 is the RXQUAL threshold, T_2 is the received signal strength threshold and T_3 is the threshold for the difference of the received power levels. Furthermore, Q represents the quality category of the call 1 or 2, and L is the received power level for the call 1 or 2. There are thus 3 independent events for the modeling of the OSC penetration: probability that both mobiles are SAIC capable, probability that both mobiles are in HR mode and probability that for both calls quality and signal strength criteria are simultaneously satisfied.

All the inputs to the estimation method were gathered from the SMART for two different scenarios: (1) with standard RF Frequency Hopping, and (2) with Dynamic Frequency and Channel Allocation.

The RXQUAL and RXLEV samples were collected on cell basis from a simulator's cluster consisting of 123 sectorized macro cells, which represents a real European network topology. 5.6 MHz of available spectrum was divided into 3.6 MHz of BCCH layer and 1.8 MHz of TCH layers with one free guard channel between them. A 2 TRX/cell and 9 TCH frequencies case was analyzed. The BCCH layer was excluded from this evaluation. The signal strength distribution and the best server areas are presented in Figures 11.43 and

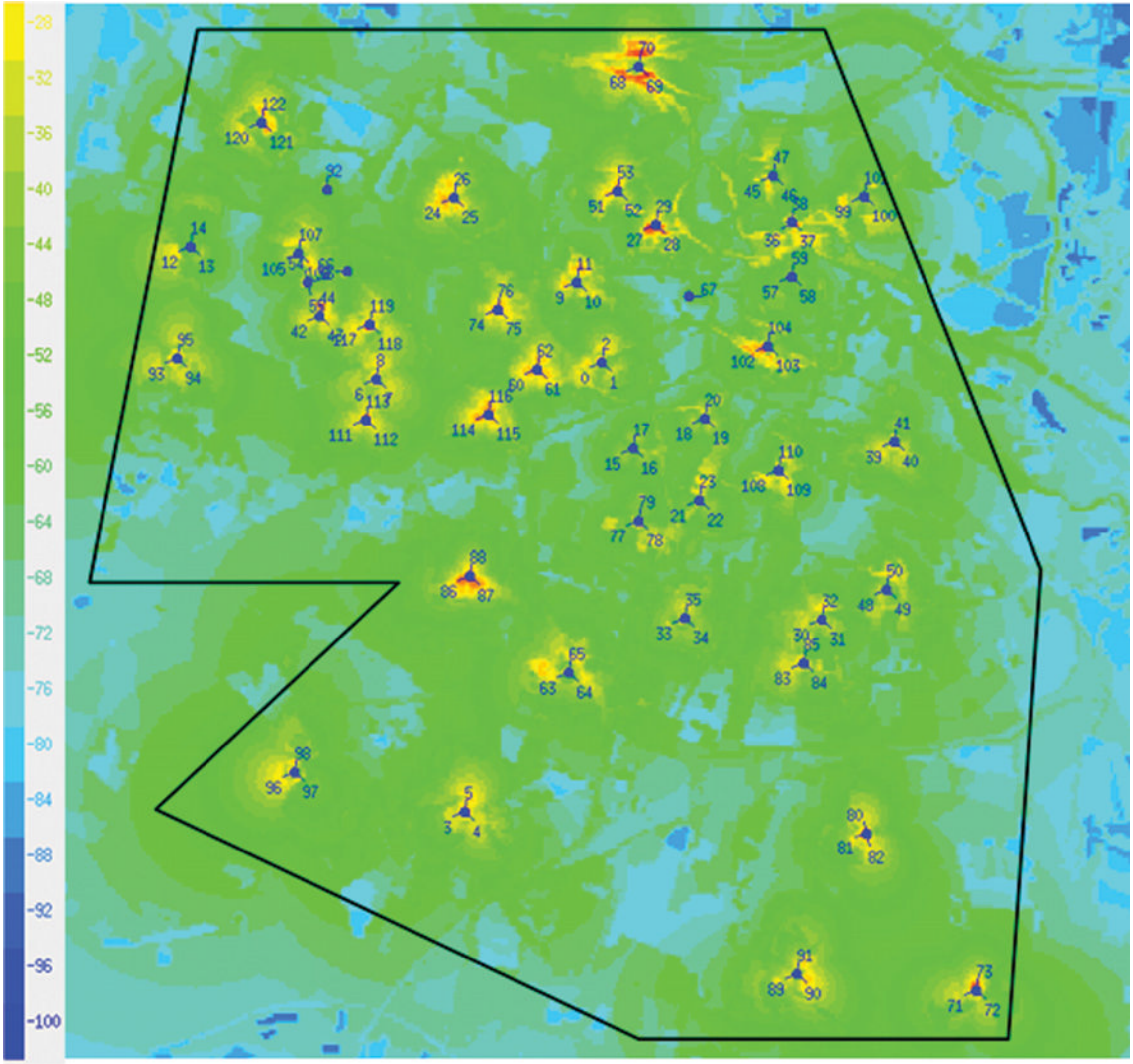

Figure 11.43 *The simulation environment [55].*

11.44. The cell level correlated distribution of RXQUAL and RXLEV samples was collected separately for UL, DL and various load points.

The HR penetration figures were gathered also on the cell level. The SAIC penetration was fixed uniformly for all the cells of the cluster.

The aim of the estimation method was to mimic the behavior of the real pairing and unpairing algorithm in the BSS RRM, and it was tested against the actual OSC penetration figures from the system level simulator for the reference case with a standard RF Frequency Hopping.

It can be assumed that the actual penetration figures, especially for the joint DFCA and OSC case, may be different from the estimated ones due to the lack of modeling of time domain and RRM details that could not be directly covered in the estimation model.

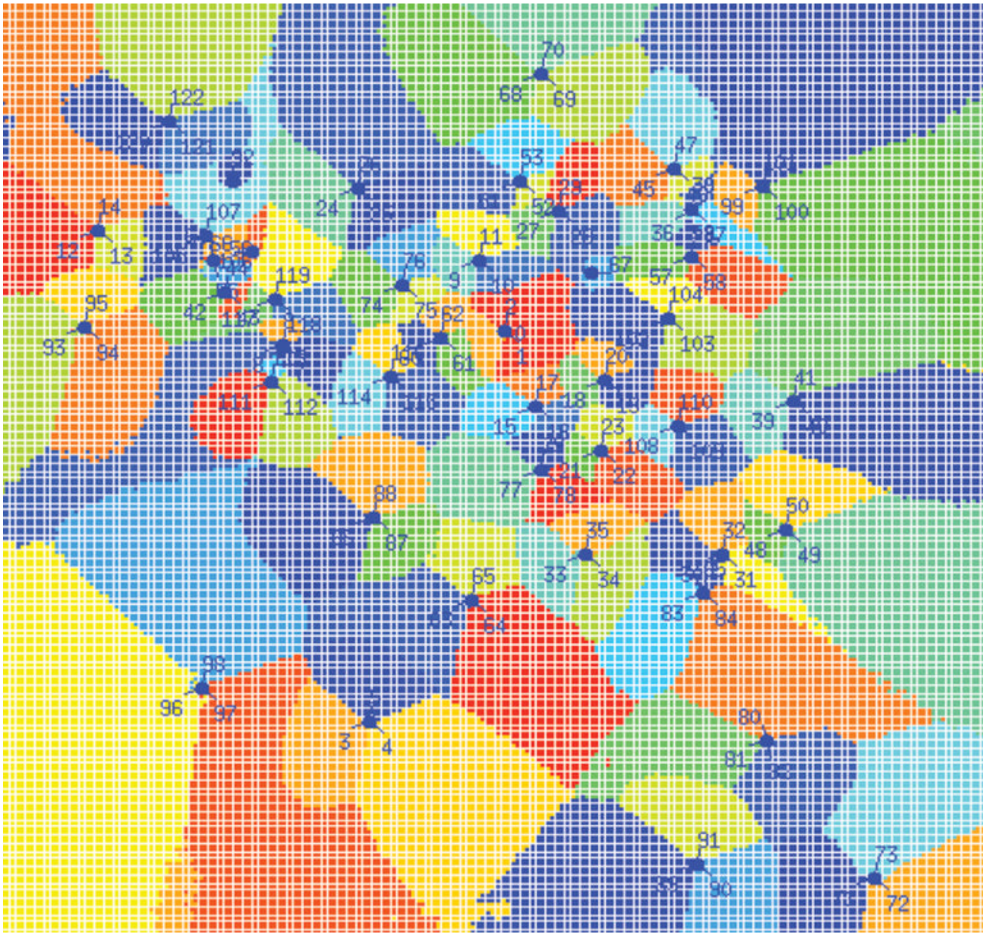

Figure 11.44 The dominating cell areas of the simulator [55].

11.9.2.1 Settings

The following assumptions were taken in the estimation method: (1) Signal strength DL/UL threshold $T_2 = -90$ dBm (for connections whose signal strength is below this limit demultiplexing possibly combined with intercell handover is triggered); (2) UL RXLEV difference $T_3 = 20$ (if the difference between the UL RXLEVs of two calls is higher than 20, demultiplexing is triggered); (3) HR to FR unpacking due to RXQUAL: 5; (4) SAIC penetration: 40, 60 and 80%; (5) RXQUAL demultiplexing threshold $T_1 = 4$; (6) FR to HR packing upper load triggering point: 80%; (7) FR to HR packing lower load triggering point: 40% (if the load of the given exceeds 80% packing process is started and not finished until the cell load is lower than 40%).

11.9.2.2 Results

Figures 11.45 and 11.46 depict the DL and UL RXQUAL distribution for RF Hopping and DFCA case for the highest Erlang Fractional Load point (EFL) in the network. From the quality point of view, the DL direction was the limiting direction in this cluster, and hence UL is not further elaborated. If however,

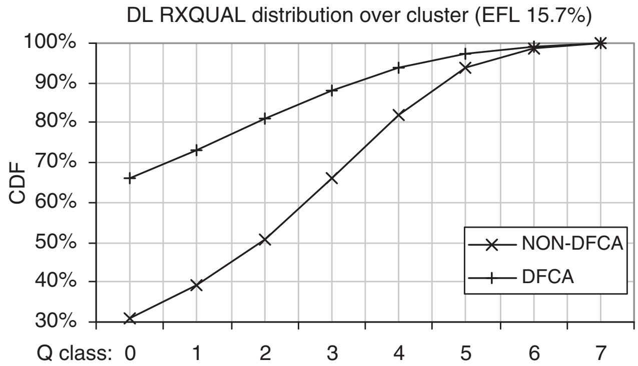

Figure 11.45 *The DL quality distribution.*

UL direction occurred to be the limiting one on the cell level, this was considered in the DHR penetration estimation process.

Due to immanent DFCA interference-aware allocation, visible quality improvement in the network is observed after the implementation of this functionality. As an example, the total amount of the DL RXQUAL 0 value samples is doubled in comparison with the standard frequency hopping scenario. In addition, the ratio of the samples with RXQUAL not exceeding the value 3 rises by approximately 33% due to the DFCA feature.

In order to maintain sufficiently high quality of the OSC connections, that is, for maintaining the OSC paired calls in DHR mode, the criteria of RXQUAL equal to 4 was set as a limiting factor for the multiplexing calls. The UL RXLEV difference threshold T_3 was set to the value of 20 in order to assure proper performance of the BTS receiver with respect to possibly unbalanced MU-MIMO signals coming from two different mobiles. The SAIC ratio was set to 60%.

The estimation of the OSC DHR penetration as a function of load, in terms of absolute OSC penetration, is shown in Figure 11.47. The system load is given as Erlang Fractional Load (EFL), that is, the average voice traffic per cell divided by the product of 8 GSM time slots with the number of hopping frequencies. It should be emphasized that the DFCA capability to assign the most optimal channel and frequency as well as the TSC is crucial for the OSC enabling. According to the results, increase of the OSC penetration due to DFCA introduction is about 20–30% depending on the load.

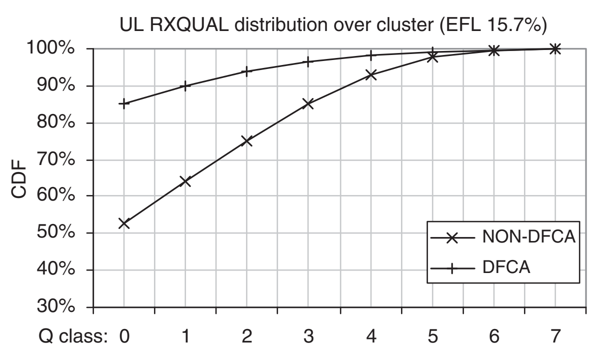

Figure 11.46 *The UL quality distribution.*

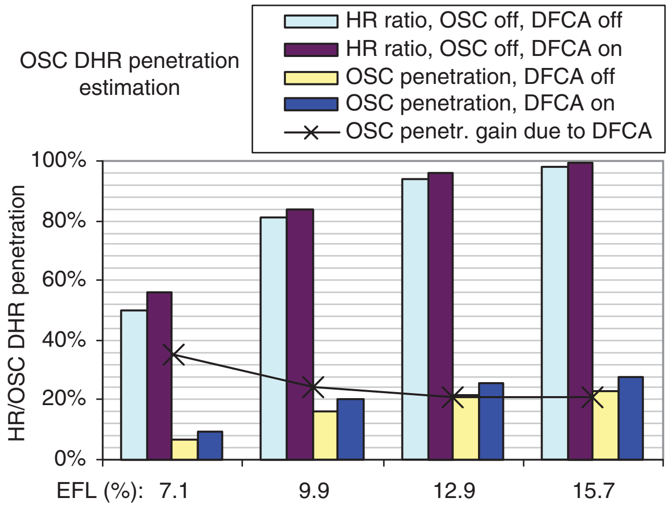

Figure 11.47 *The estimation of OSC DHR penetration as a function of load figures (denoted in EFL). The results on cluster level are based on the absolute OSC penetration.*

The summary of the cluster level results with the penetration gain relative to the SAIC ratio equal to 60% can be seen in Table 11.9.

The cell specific OSC penetration results for two load points are presented in Figure 11.48. As can be seen, the higher the cell load is, the higher is the legacy AMR HR penetration as well as OSC DHR penetration, which reflects the common policy to trigger these two channel modes once the cell load is high enough to justify the HR packing and DHR multiplexing process.

The cell specific DFCA gains connected with relative increase of the OSC DHR penetration for the highest load point are shown in Figure 11.49.

It can be seen from Figure 11.49 that for some cells over 50% gain could be achieved, and that the maximum gain was up to 70% in a single cell. Over the whole set of 123 cells, the average gain is 21%.

11.9.2.3 Influence of SAIC Penetration

In this study, it was assumed that only SAIC-capable mobiles could be paired in the OSC mode. Hence, the SAIC penetration figure directly influences the DHR ratio. The example of this relationship calculated for

Table 11.9 *The results on cluster level showing the results relative to the SAIC ratio*

	OSC penetration		
EFL (%)	Non-DFCA	DFCA	Penetration gain
7.1	11.3	15.3	35%
9.9	26.9	33.3	24%
12.9	35.6	43.2	21%
15.7	38.0	46.1	21%

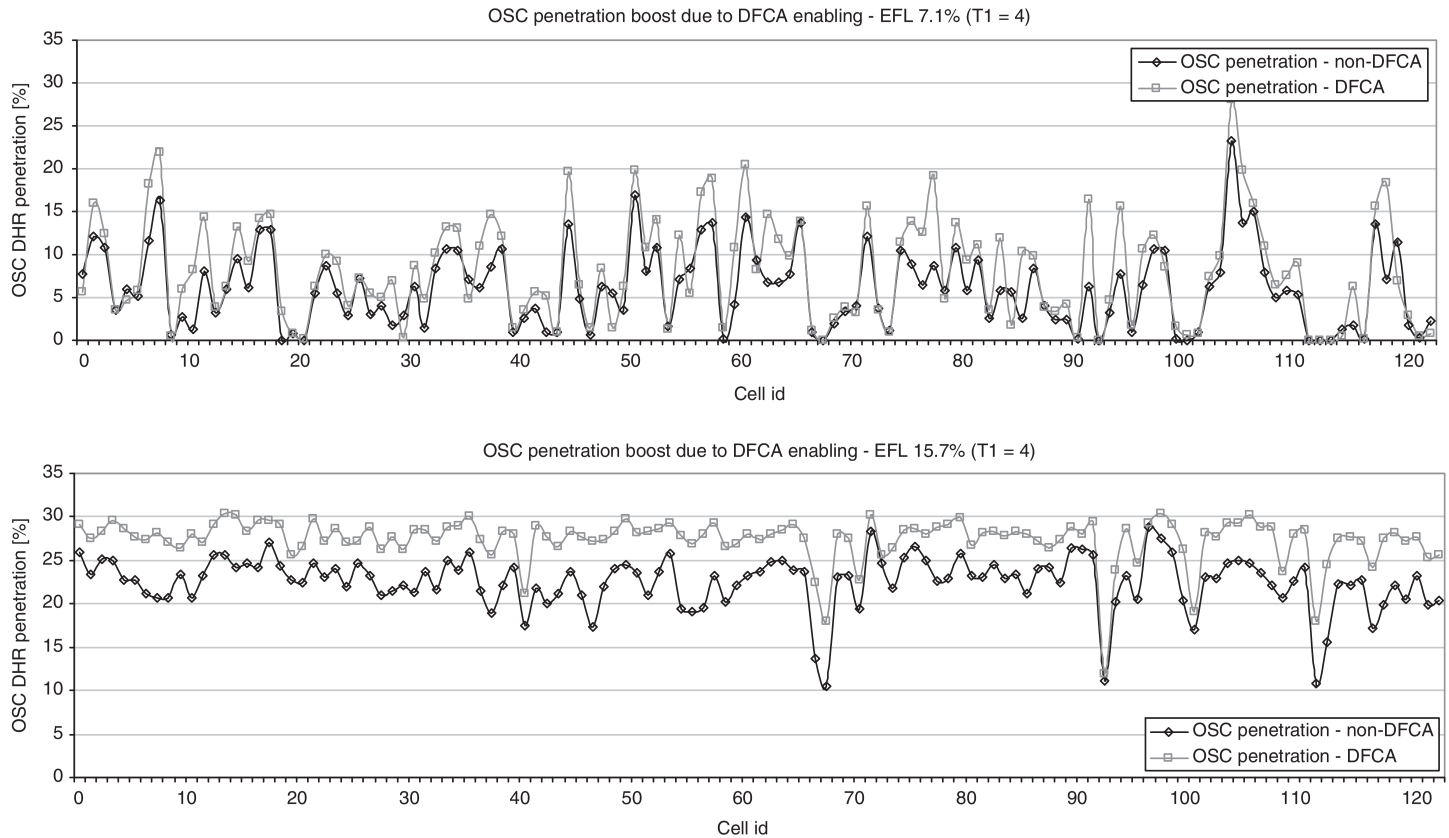

Figure 11.48 *The OSC gain of each studied cell for selected EFL values with the SAIC penetration of 60% [55].*

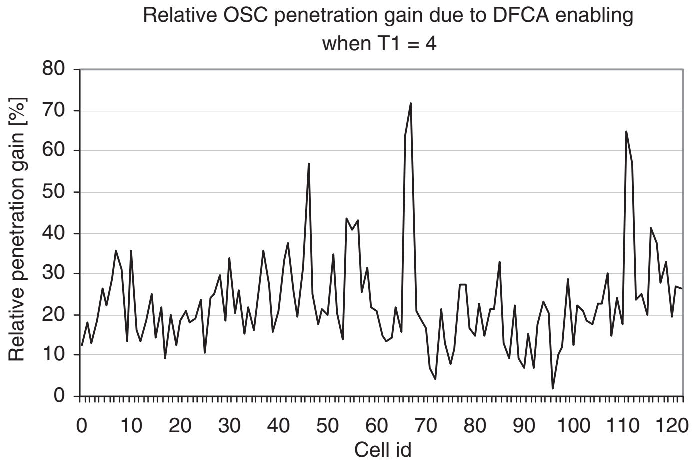

Figure 11.49 *OSC penetration gain due to the DFCA enabling [55].*

the highest load point in the network is presented in Figure 11.50. The results are shown as a function of the SAIC handset penetration.

For the SAIC ratio equal to 80%, DFCA increases the estimated absolute OSC penetration up to the level of 50%.

It should be noted that the influence of the SAIC handset penetration on the OSC pairing (which is the case also jointly with the DFCA functionality) may be even higher in real networks since in order to maintain the same link quality, less power could be transmitted by the BTS for the SAIC MSs in comparison with the non-SAIC handsets. This effect directly improves the overall interference distribution in the network which results in the further increase of DHR penetration in the field as has been noted in Ref. [56].

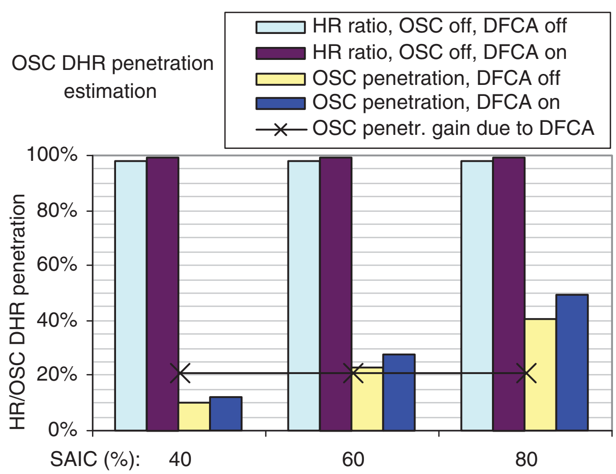

Figure 11.50 *Absolute OSC DHR utilization as a function of SAIC handset penetration, non-DFCA vs. DFCA.* $T_1 = 4$ *[55].*

11.10 EDGE

EDGE (enhanced data rates for global evolution) refers to a new kind of modulation method for the GSM radio interface. The original 0.3 GMSK (gaussian minimum shift keying) of the system actually remains unchanged with EDGE, but 8-PSK-modulation (PSK, phase shift keying) can also be used, if necessary. This means that 8-PSK, which is more error-prone but offers a faster data rate, can be used when the signal-error rate is good enough.

EDGE does not actually refer to an actual service. EDGE has been designed to be used in connection with for example, GPRS, known as EGPRS (enhanced GPRS). It is also possible to use EDGE with HSCSD forming a service known as EHSD (enhanced high speed data). In practice nothing prevents the use of EDGE with regular GSM voice connections to create quality of sound comparable to high level stereophonic music transmission either. This specific subject is known as EAMR (enhanced adaptive multi rate codec). The first specifications based on the EDGE technology are contained in GSM release 99, which was completed during 2000. The expectations for the first phase EDGE networks and terminals range 2002–2003.

The basic idea behind EDGE is to significantly improve the data rate of the GSM system compared to basic GSM data services. In theory, a data rate of over 400 kb/s can be reached with EGPRS using channel coding offering the highest possible data rate and the simultaneous use of all eight TRX time slots. In reality, the same restrictions apply to EGPRS as with basic GPRS – for example, technical restrictions with regards to terminals and network elements and the impossibility of using maximum theoretical network capacity due to other load on the network. However, the actual data rate achieved using EGPRS represents a significant improvement over basic GPRS service data rates.

EGPRS is a relatively attractive alternative for offering a kind of lighter version of third generation mobile systems such as the UMTS, over the GSM network. The theoretical maximum rate of the EGPRS actually meets the 384 kb/s data service specification of the UMTS. However, while it is impossible to reach the UMTS maximum rates with EGPRS, the two are not mutually exclusive techniques.

EDGE changes the principles of the modulation technique, but for example, the radio interface channel specifications including the forms of data bursts and the bits to be transferred remain unchanged. Also the GSM 200 kHz channel division is maintained. Thus EGPRS can share the same radio interface resources with GSM users. The adoption of the technique will, however, be more expensive than adopting basic GPRS services. This is because EGPRS requires the changing of the TRXs, while basic GPRS does not necessarily require actual equipment changes in the base stations. In addition to a change in the modulation technique, EDGE requires a minimum of software-level changes in network elements and protocol stacks. Naturally the terminals are also changed with the new service, as it is unlikely that terminals preceding EDGE could be modified to be EDGE-compliant.

Both the European ETSI and the US-based T1 have participated in the specification work for EDGE. Thus it is likely that EDGE will work in both GSM networks and IS-136 networks. As the GSM technology is also used in the US, EDGE can act as a kind of unifying stage in the European and US mobile systems of the pre-third generation era. EDGE can also be considered a kind of basic technique of the third generation.

The specification term GERAN (GSM EDGE radio access network) refers to a radio network based on the combination of basic GSM and EDGE. GERAN is based on GSM specification release 99, but as with most GSM service elements, it has its own evolutionary path. Thus specifications following release 99 will contain improvements and additions to basic GERAN.

GERAN is also the name for a 3GPP working group focusing on the continuous development of the GSM, GPRS, and EDGE.

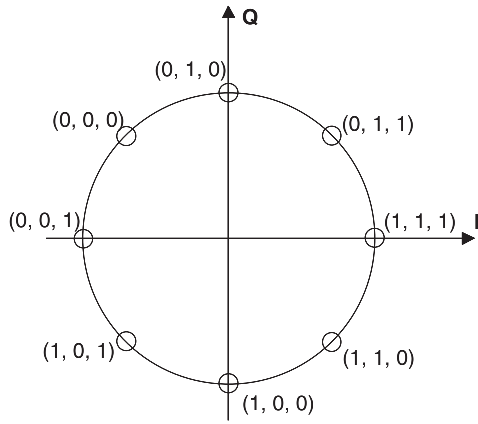

Figure 11.51 *The principle of 8-PSK.*

11.10.1 Technical Features

Figure 11.51 represents the principle of the linear 8-PSK modulation used by EDGE. The same principle also applies to EGPRS and EHSD services, as it can be used with both packet and circuit switched connections. As shown in the figure, the technique consists of eight levels, which in theory enable six times the spectrum efficiency of basic GSM, according to the simulation results of Ref. [57]. According to the levels, the technique is based on rotations of $3\pi/8$ degrees. The traditional 200 kHz channel configuration and TDMA frame structure are used with GSM. The symbol rate of the technique remains 270.8 kb/s.

EDGE is included in the services of release 99. In addition to the change in modulation technique, the automatic link quality control is related to the service. From release 99 onwards, EDGE has been specified into the GSM and US-based IS-136 systems. Releases 4 and 5 will bring enhancements to the EDGE service including UMTS compatibility.

EDGE uses automatic link adaptation (ALA), which allows the optimization of the data rate depending on existing signal-error levels, similar to basic GSM. In practice the fastest data rates are achieved close to a serving base station, whereas data transfer is slower near the cell border as base stations on the same channel are closer.

11.10.2 GERAN-Architecture

The GERAN or GSM EDGE radio access network architecture consists of GSMs base station system (BSS) and its interfaces. The MSs are in contact with the GERAN via the Um-interface. GERAN can use Gb-, A-, Iu-ps- and Iu-cs-interfaces to contact the GSM and UMTS trunk networks, as shown in Figure 11.52.

According to the specifications, an EDGE-MS can only function in the following classes:

- A/Gb-class; this class can function in for example, pre-release 4-compliant terminals, and in release 4-compliant networks, when no Iu-interface has been defined between the network and the trunk network.
- Iu-class; a connection can be formed either via the Iu-ps- or Iu.cs-interfaces. An example of the use of this class would be the situation in a release 4-compliant network, where the Iu-interface has been defined between the network and the trunk network.

EDGE also defines several channel coding methods. With EDGE, channel coding is automatically chosen to optimise the user's data rate, as with the basic GRPS service. EDGE contains the GPRS channel coding

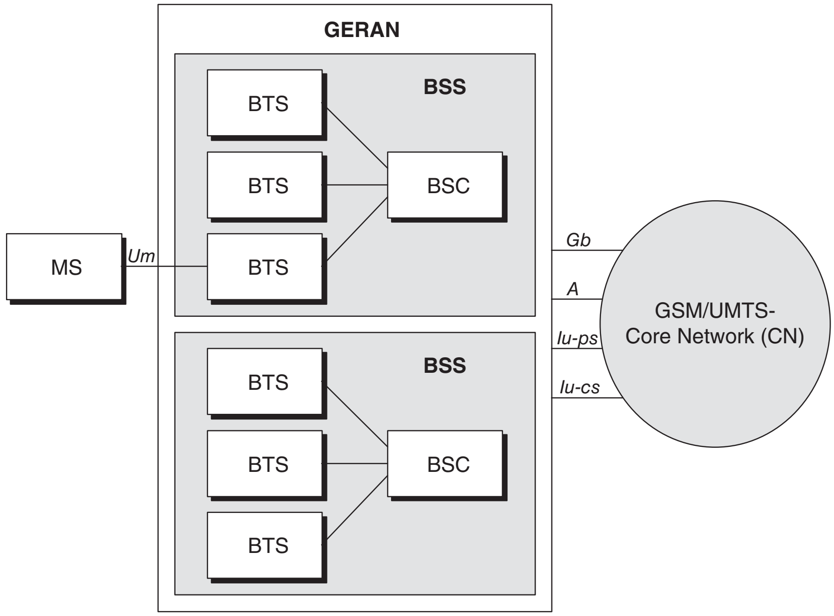

Figure 11.52 *GERAN architecture.*

classes CS-1–CS-4, in addition to which nine new channel coding classes have been defined for the EGPRS service. The new classes are known as MCS-1–MCS-9 (EGPRS modulation and coding scheme). As with the GPRS service, with EDGE the user data rate offered by each channel coding class is saturated at a certain C/I –value. Thanks to automatic link adaptation (ALA), the connection is not interrupted by variances in the C/I ratio. Instead, the system chooses a channel coding class from the envelope curve. The low-protection channel coding class MCS-9 can offer a data rate of approximately 60kb/s, which however requires a very good C/I-value. MCS-1 offers a much lower data rate, but at a relatively low C/I-value.

EDGE channels have been defined somewhat more broadly than with the GSM or GPRS services. EDGE traffic channels can be full rate traffic channels (TCH/F), half rate (TCH/H), or quarter rate (TCH/Q). EDGE also defines a so-called octal traffic channel (O-TCH), used to transfer encoded speech. The O-TCH channel can also be full, half or quarter rate. In addition, EDGE specifications contain an enhanced TCH (E-TCH), used to transfer the user's data, and the packet data traffic channel (PDTCH), which can be either full or half rate.

Of the control channels, the broadcast channel (BCH), cell broadcast channel (CBCH), and the common control channels (CCCH) have been specified into GERAN, similar to the corresponding channels in release 99. However, the dedicated control channels (DCH) for the physical channels of each connection are defined as follows:

- The fast control channel FACCH has been combined with one traffic channel (TCH). The FACCH is of the same type as the traffic channel, so that for example, the FACCH/F channel is used together with the TCH/F channel.
- The packet associated control channel (PACH) has been combined with each packet data traffic channel (PDTCH). PACH type depends on PDTCH type.
- The slow associated control channel (SACCH) has been joined to each TCH-and /or PDTCH channel. SACCH type depends on the types of the channels connected to it.

- The standalone dedicated control channel (SDCCH) is used according to specification 3GPP GSM TS 45.003, as usual. It is used only during initiation of the connection, or for example, for the transmission of short messages, where no voice- or data-connection is established.

The DCH are the following on shared physical channels:

- Bidirectional packet associated control channel (PACCH); called PACCH/U uplink and PACCH/D downlink. PACCH is of the same type as the PDTCH.
- The PTCCH/U intended for the uplink transfer of the TA-parameter, used when the MS sends a PACH-burst. The network can estimate the correct initial value of the TA-parameter using the RACH-burst.
- The PTCCH/D intended for the downlink transfer of the TA-parameter, is used to update the value of the TA-parameter to several MS. Each PTCCH/D channel has a corresponding PTCCH/U channel.

11.10.3 The Functioning of the EDGE

11.10.3.1 Principles

EHSD The EHSD contains both the nontransparent (NT) and transparent (T) transfer modes, like the basic GSM data service. Data rates close to 64kb/s can be reached. As EHSD is a circuit switched data transfer method, the same restrictions apply to it as apply to other circuit switched GSM data services. Thus the data rates of the A-interface and the fixed line network block the use of over 64kb/s connections. The benefit when compared to GSM data and HSCSD connections, is that fast rates can be achieved using only one or two timeslots according to the definitions of the multislot technique.

EGPRS EGPRS is an improvement over the first phase GPRS service. EGPRS defines more channel coding classes than its predecessor. Thus the achieved data rates per individual timeslot are considerably higher than with the basic GPRS service. Like the HSCSD, also EGPRS defines the multislot technique, which further increases data rates. By combining the multislot technique and different channel coding classes, it is theoretically possible to reach data rates of over 400kb/s per user. In practice the rate is likely to be much lower.

EAMR EAMR enables the transfer of high-quality speech and music. The same restrictions apply to EAMR connections as apply to EHSD, as EAMR is also circuit switched.

11.10.4 Channel Coding

11.10.4.1 EHSD

According to specification release 99, the 8-PSK modulation can be used with the EHSD service for user rates 28.8 kb/s, 32.0 kb/s, and 43.2 kb/s per timeslot. The 28.8 kb/s rate can be used both as nontransparent (NT) and as transparent (T). The 32 kb/s rate can only be used as transparent, while the 43.2 kb/s rate can only be used as nontransparent. Also the basic HSCSD rate 4.8 kb/s, 9,6 kb/s, and 14.4 kb/s can be used with EHSD, with the traditional 0.3 GMSK modulation method.

11.10.4.2 EGPRS

EGPRS channel coding classes are known as the modulation and coding scheme (MCS). Classes MCS-1–MCS-4 use the basic GSM 0.3 GMSK modulation, whereas classes MCS-5–MCS-9 use the new 8-PSK

Table 11.10 EGPRS channel coding schemes and some corresponding parameters

Class	Code rate	Code rate of header	Modulation	Data rate (kb/s)
MCS-1	0.53	0.53	0.3 GMSK	8.8
MCS-2	0.66	0.53	0.3 GMSK	11.2
MCS-3	0.80	0.53	0.3 GMSK	14.8
MCS-4	1.00	0.53	0.3 GMSK	17.6
MCS-5	0.37	1/3	8-PSK	22.4
MCS-6	0.49	1/3	8-PSK	29.6
MCS-7	0.76	0.36	8-PSK	44.8
MCS-8	0.92	0.36	8-PSK	54.4
MCS-9	1.00	0.36	8-PSK	59.2

modulation. Table 11.10 contains some of the most important parameters and values relating to the EGPRS channel coding classes.

As with EHSD, EGPRS uses also automatic link adaptation (ALA). Figure 11.53 shows a rough estimate of the user data rates achieved using the different channel coding classes. Each class has been optimized for a certain range of C/I-values, outside of which the data rate no longer increases together with the C/I-value, but saturates. In an ideal situation, the maximum data rate as a function of the C/I curve is achieved by switching channel coding class "on the go," that is, by following the envelope curve. In reality, the link adaptation may not be close to the ideal situation, but the technique is in principle functional.

Another way to choose the optimal channel coding class would be to use the IR-technique (EGPRS Incremental Redundancy), where the best channel coding class is chosen by gradually increasing the level of channel coding. Thus, if the reception of a radio block is unsuccessful, it is resent using a different puncturing technique, and combined with the previous received block. This combination is known as soft combining. However, the IR-technique is only being proposed to be included in specifications.

11.10.5 Multifunctioning

EGPRS and GPRS function according to the same principles. Even if the number of channel coding classes defined by the specifications differs, the specifications define a dynamic GPRS- and EGPRS user multiplexing for the same data channels in the radio interface [57]. When resources are allocated for a GPRS terminal

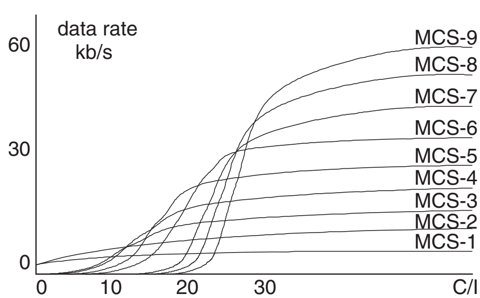

Figure 11.53 The principle of EGPRS channel coding schemes and corresponding C/I values. The values presented in this context are rough estimations; for more detailed information the simulation results of specifications should be examined.

Table 11.11 *EDGE multislot classes*

Multislot Class	DL TSLs	UL TSLs	Maximum active TSLs
1	1	1	2
2	2	1	3
3	2	2	3
4	3	1	4
5	2	2	4
6	3	2	4
7	3	3	4
8	4	1	5
9	3	2	5
10	4	2	5
11	4	3	5
12	4	4	5
30	5	1	6
31	5	2	6
32	5	3	6
33	5	4	6
34	5	5	6

uplink, the network must use GMSK-modulation (channel coding classes CS-1–CS-4 or MCS-1–MCS-4). The multiplexing of GPRS and EGPRS terminals downlink is achieved using the uplink state flag (USF) parameter, as with normal GPRS traffic.

Not only the modulation scheme and the code rate, but also the multislot class determines the final data rate in the uplink and downlink directions. A allocated multislots in downlink and uplink can be different, but they are tight together according to the multislot class table in such a way that the first number of the pair is the number of downlink timeslots and the second is the number of uplink timeslots per single mobile station (e.g., 5 + 4 TSLs). Currently, class 10 is typically supported by GPRS/EGPRS mobile devices, that is, providing 4 TSLs in DL and 2 TSLs in UL. Nevertheless, this class limits the simultaneous maximum number of the DL and UL TSLs to a total of 5. The functionality of the multislots is dynamic and automatic, so the network allocates 3 + 2 or 4 + 1 TSLs depending on the characteristics of the data transfer. Assuming ideal reception conditions combined with the best modulation and coding scheme, a total of, for example, 5 TSLs is able to deliver 5×59.2 kb/s = 296 kb/s data rate. Table 11.11 summarizes the EDGE multislot classes.

11.11 DLDC

11.11.1 Installation Aspects

The DLDC feature of offers a possibility to enhance the data rates of DLDC-capable MSs by increasing the number of radio timeslots that can be allocated for the downlink TBFs of such mobiles. This is achieved by assigning the resources of an EGPRS downlink TBF on two TRXs. The MS receives both radio frequency channels, thus doubling downlink throughput. An uplink TBF, on the other hand, is assigned on one TRX only. However, the UL TBF can be dynamically reallocated between the two TRXs to maximize utilization of uplink resources. This feature doubles downlink peak throughput, bringing rates of up to 592 kb/s. The average throughput can exceed 300 kb/s, enabling for example, the streaming of high quality video.

Typically, DLDC activation is done per BTS. In order to enable DLDC in a BTS, EGPRS must be enabled, and the DLDC license must have enough capacity available. The DLDC functionality might require update to the PCU version. The default EGPRS capacity must be configured on two TRXs which are in the same EDAP in order to serve DLDC-capable mobile stations with a DLDC configuration. Depending on the vendor solutions, DLDC might require the activation of also other features in order to exploit the extended multislot capabilities of MSs with multislot classes 30–45 for DLDC.

11.11.2 Time Slot Allocation

The time slot territory changes due to the increasing and decreasing of the voice and other data users. The territory strategy depends on the network vendor solutions. With the default parameters of the feature, for example, in the case of Nokia Siemens Networks solution of 2 TRX cell, at least 2 idle TSLs needs to be available for CS traffic after the PS territory upgrade, so the PS territory is not upgraded unless there are at least 3 idle TSLs before the territory upgrade. In this example, PS territory downgrade on the other hand is triggered so that at least 1 idle TSL is always available for CS traffic, continuing with the default parameters and 2 TRX per BTS configuration.

As an example, in order to activate the DLDC feature, the respective license key might be needed to be activated. Furthermore, the activation of the feature is done both in BSC and BTS.

The DLDC requires new terminals that support the dual carrier functionality. The supported maximum number of simultaneous time slots depends on the capabilities of both terminals and network, and thus, for example, time slot configurations up to 5 + 5 TSLs for 2 downlink TRXs might be a practical limit.

DLDC feature needs new parameters, for example, for the enabling of the feature in BTS-level. There are also new BTS and PCU level parameters for the managing of the feature. Also new MML commands are needed for the territory downgrade, EDAP table update and territory upgrade for the whole PCU. It is important to configure the suitable EDAPs for DLDC.

In order to maintain the possibility for DLDC configurations from the beginning of the EGPRS territory, the default EGPRS capacity is selected from an EDAP that contains two or more TRXs. In order to maintain the highest amount of TRX pairs for DLDC configurations, when a new TRX is to be upgraded into the EGPRS territory, it should preferably be selected from the same EDAP as the latest (so far) TRX in the EGPRS territory.

11.11.3 Feature Functionality

11.11.3.1 Time Slot Allocation

The DLDC time slot allocation is based on the default CS and PS territories that are varied based on the load. The functionality is the same as in EDGE and GPRS. As an example, in commercial equipment's upgrade and downgrade procedure, the effecting parameters may be the following:

FREE TSL FOR CS DOWNGRADE (CSD)	95%
FREE TSL FOR CS UPGRADE (CSU)	4 s
DEFAULT GPRS CAPACITY (CDEF)	100%

The CSD indicates how much there should be available capacity (time slots) compared to the whole capacity of the cell when downgrade is performed. The margin of idle TCH/Fs that is required as a condition for starting a GPRS territory upgrade is defined by the BSC parameter free TSL for CS upgrade (CSU). The parameter defines how many traffic channel radio timeslots have to be left free after the GPRS territory upgrade. When defining the margin, a two-dimensional table is used. The table columns are for different amounts of available resources (TRXs) in the BTS. The rows indicate a selected time period (seconds) during which probability for an expected downgrade is no more than 5%. The operator can modify the period with the BSC parameter CSU. The default value for the period length is 4 seconds. The default GPRS capacity can be varied in DLDC whilst the capacity reaches also the other TRX. As an example, 60% value can be used.

If fewer timeslots are allocated for a TBF than could be used by the MS, the BSC adds timeslots into the PS territory in order to allow reallocation of the TBF with the desired timeslot configuration. Such upgrades may be prevented by the CS traffic capacity requirements in the cell. The same principle is applied to TBFs created for DLDC-capable MSs. However, the desired timeslot configuration depends on whether single carrier or DLDC resources have been allocated for the TBF:

- If single carrier resources have been allocated for the TBF, the target is to allow reallocation with as many timeslots as the MS can use in a single carrier configuration.
- If DLDC resources have been allocated for the TBF, the target is to allow reallocation with as many timeslots as the MS can use in a DLDC configuration. No upgrade is performed if the PS territory already extends to three or more TRXs.

This approach also means that, following the creation of a DL TBF for a DLDC-capable MS, timeslots may be added to an existing DLDC-territory, but no DLDC-territory is created if one does not already exist.

11.11.3.2 Radio Resource Management for DLDC-Capable MSs

When an MS indicates DLDC capability, the BSC takes this into account when selecting the radio resources for the MS. Resource selection for a TBF involves two main operations: first, which of the BTSs serving the cell would offer the highest capacity is estimated, and then radio timeslots are allocated for the MS from the selected BTS. DLDC is not given absolute priority in either operation, but instead the BTS and the radio timeslots estimated to offer the best throughput for the MS are selected, regardless of whether these offer DLDC or single carrier resources.

11.11.3.3 Channel Allocation

As in the previous releases, the MS-specific and territory-specific factors affecting the number of timeslots that may be allocated for a TBF in a given direction are:

- The multislot class of the MS.
- In case of concurrent TBFs, the number of slots allocated in the opposite direction.
- The size of the PS territory in the selected BTS.

In typical commercial equipment, the BSC may support DLDC for MS multislot classes 8, 10, 11, 12 and 30–45. The BSC may also support so-called “DLDC equivalent multislot classes,” which allow the application of multislot class 30, 31, 32 or 33 characteristics, respectively, in DLDC resource allocation for an MS indicating multislot class 8, 10, 11, or 12, together with an MSCR of 0 or 1.

Table 11.12 summarizes an example of the maximum number of DL timeslots that may be allocated for a DLDC-capable MS. The “Maximum timeslots per DLDC TRX” column indicates the maximum number of DL slots which may be allocated on each of the two TRXs accommodating a DLDC TBF, while the “Maximum timeslots per DLDC TBF with MSCR” column indicates, for each possible value of MSCR, the maximum total of DLDC timeslots on two TRXs. It should be noted that the “Equivalent multislot classes”

Table 11.12 The maximum number of DL timeslots that may be allocated for a DLDC-capable MS

		Max TSLs / DLDC TBF with MSCR						
Multislot class	Max TSLs / DLDC TRX	0	1	2	3	4	5	6
8	5	10	9	8	7	6	5	5
10	5	10	9	8	7	6	5	5
11	5	10	9	8	7	6	5	5
12	5	10	9	8	7	6	5	5
30	5	10	9	8	7	6	6	6
31	5	10	9	8	7	6	6	6
32	5	10	9	8	7	6	6	6
33	5	10	9	8	7	6	6	6
34	5	10	9	8	7	6	6	6
35	4	8	8	8	7	6	6	6
36	4	8	8	8	7	6	6	6
37	4	8	8	8	7	6	6	6
38	4	8	8	8	7	6	6	6
39	4	8	8	8	7	6	6	6
40	5	10	10	10	9	8	7	7
41	5	10	10	10	9	8	7	7
42	5	10	10	10	9	8	7	7
43	5	10	10	10	9	8	7	7
44	5	10	10	10	9	8	7	7
45	5	10	10	10	9	8	7	7

and MSCR are applicable only to DLDC TBFs, and Table 11.12 should not be used in reference to single carrier TBFs.

The restriction on the maximum number of DL slots per TRX arises from a requirement that the UL resources allocated for an MS with a DLDC TBF restrict on both DLDC TRXs (regardless of the TRX accommodating the UL resources) the number of DL timeslots, and the position of these in relation to the slots used for UL traffic, as defined by the MS multislot class restrictions.

For concurrent TBFs, the number of DLDC timeslots that may be allocated is therefore affected on both DLDC TRXs by the number of slots allocated for the UL TBF. For instance, if a two-timeslot UL TBF has been created for an MS class 33 DLDC-capable MS, only four slots per TRX may be allocated for a concurrent DLDC TBF (if MSCR is 0).

Whenever MSCR limits the number of DLDC timeslots that may be allocated for an MS, the BSC restricts the number of slots it allocates on the DLDC TRX that is closer to the border between PS and CS territories. Following the previous example, if the class 33 DLDC-capable MS with a two-timeslot UL TBF indicates MSCR 3, the MS is unable to use more than seven DL slots under any circumstances, and with the UL TBF restricting the maximum number of DL slots per TRX to four, the BSC allocates three slots on the DLDC TRX closer to the PS/CS border, and four on the other.

In DL channel allocation for a DLDC-capable MS, both the best possible DLDC and single carrier resources are identified, after which the better of these is allocated for the MS. As an exception to this, only single carrier DL resources can be allocated concurrent to an EDA UL TBF.

When no UL TBF concurrent to a DLDC TBF exists, the BSC allocates the PACCH on one of the TRXs accommodating the DLDC TBF, and uses this TRX for polling and RLC control message transmission towards the MS. When a UL TBF exists concurrently with a DLDC TBF, the PACCH/PTCCH are allocated from the same TRX as the resources for the UL TBF. In case of a concurrent UL TBF allocation, if suitable

UL resources are available on both TRXs accommodating a DLDC TBF, the TRX offering the higher UL capacity is selected for the UL TBF.

11.11.3.4 DLDC TBF Reallocations

The reallocation triggers applicable to single carrier TBFs are also applied to DLDC TBFs. In addition, a reallocation is triggered for a single carrier TBF created for a DLDC capable MS in a DLDC-activated BTS, if it is estimated that DLDC resources may provide better throughput than the existing resources. Such reallocations are not triggered if the MS has a concurrent EDA UL TBF.

11.11.3.5 Downlink Packet Transfer in DLDC

Common link adaptation functionality is used for both the carriers of a DLDC TBF. This means that the same MCS is used for both carriers. The following formula is used for selecting the MCS:

$$\text{MCS} = \min\,(\text{MCS on Carrier 1},\ \text{MCS on Carrier 2})$$

For DLDC assignments, a new message format is used. The PCU uses the new message format when the assignment message assigns a dual carrier allocation to a DLDC-capable MS. The PCU uses the new message format also when the assignment message assigns a single carrier allocation to a DLDC-capable MS that has a dual carrier allocation. Otherwise, the old message format is used. For DLDC assignments, the extended RLC/MAC control message segmentation mechanism is used.

The basic link adaptation of EDGE has been studied in various publications, including Ref. [40]. As the DLDC link adaptation is based on the same principle, taking into account Formula 1, for example, the C/I mapping of MCSs of DLDC connection can be estimated based on this reference information.

11.11.4 Case Study of DLDC Performance

The following chapter presents investigations carried out in laboratory environment of NSN Innovation Center (NICE) in 2010. The test cases were performed both in static and moving environment, showing the functionality of the DLDC. The following chapter details the setup and outcome.

11.11.4.1 Network Setup

Figure 11.54 shows the network architecture of the test setup. An internal research core network was used for the IP connectivity between the BTS (Madrid) and BSC (Finland). The core network (MSC, SGSN) was also located in Tampere under an internal research packet core network which means that the core network did not create bottlenecks in this specific setup. It is interesting to note that the distance between the BTS and BSC was thus more than 3500 km.

The general DLDC test network setup was made as follows:

- BSC in Tampere, Finland.
- BTS in Madrid, Spain, containing 1–2 cells (depending on the case), 2 TRX in each cell.
- 1800 MHz frequency band.
- Terminals: up to 2 × DLDC, and up to 3 EDGE terminals depending on the case.
- Abis transport: CESoPSN (Circuit Emulation Service over Packet Switched Networks), with throughput over 700 kb/s, Abis EDAP configured for 2 TRX case.
- Indoor coverage in NICE lab area, with 2 partially overlapping cells of about 40 meter radius each.

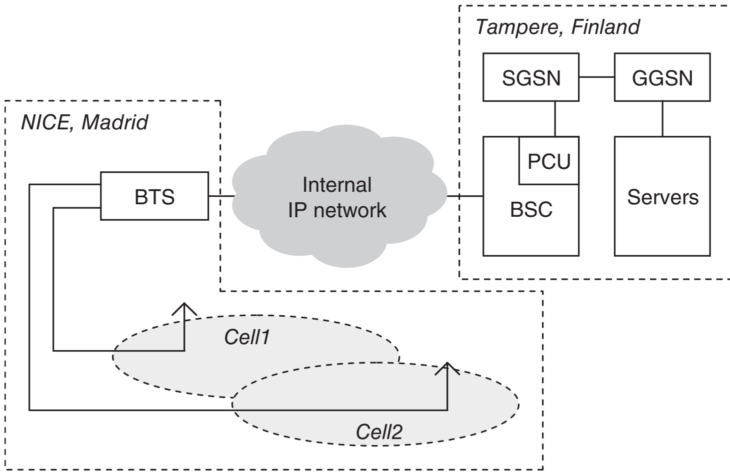

Figure 11.54 *The DLDC test network setup. Cell 2 was partially overlapping with the cell 1, and it was thus utilized in cell reselection related cases.*

11.11.4.2 BTS and BSC

NSN Flexi EDGE BTS was utilized with 2 cells, each containing a dual-TRX configuration in GSM 1800 band. The transmitter power level was set to minimum, with additional 15–20 dB attenuators in the RF output of the BTS.

11.11.4.3 Abis Interface

The DLDC requires the proper dimensioning of the Abis. At least 700 kb/s should be thus reserved for EDAP. Figure 11.55 shows the Abis definition in the laboratory network setup via the BTS Manager display that was utilized specifically for this investigation.

11.11.4.4 Terminals

The DLDC terminals were GPRS/EDGE capable Rel. 7 with DLDC support. Depending on the case, there were either one or two terminals. EDGE terminals were also utilized in the tests to mix the traffic profiles, up to 3 units depending on the case, as well as GSM terminals for generating the CS calls. SIM cards were in the HLR profiles in the test core network.

11.11.4.5 Test Tools

The test tools were used for the monitoring and logging of the measurement data. Following equipment was used:

- WireShark, default tool for the throughput logging of DLDC terminal in IP level.
- Data log software of DLDC terminal that stores the radio interface signaling of the connected terminal, including the time slot usage of the terminal.
- Nokia Application Tester. This tool was used for as secondary option for the LLC level throughput.

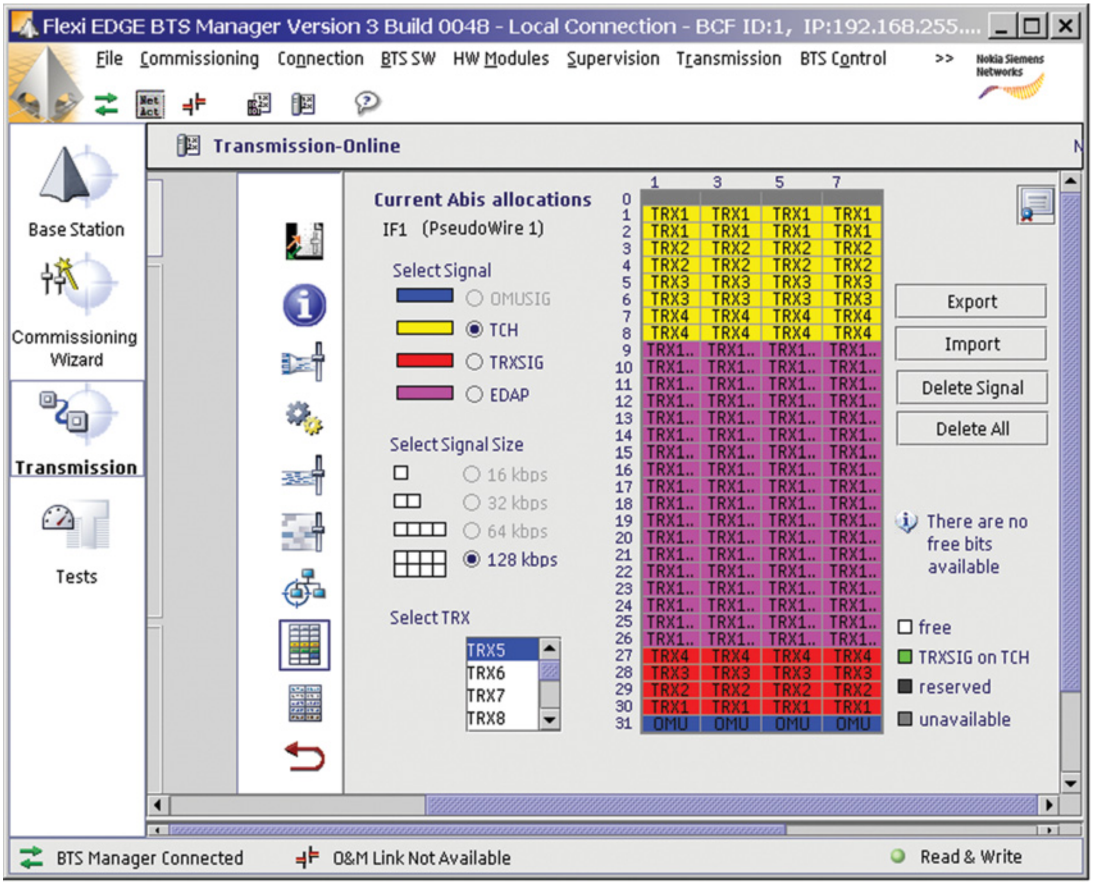

Figure 11.55 *The Abis dimensioning of the laboratory test network.*

- BSC measurement "C_Scheme" shows the distribution of the utilized coding schemes in 15 minute intervals. This measurement was utilized only for assuring that the MCS-9 took place in normal lab conditions.
- BTS monitoring for the time slot usage; separate measurement type (service terminal window) was activated for CS calls and PS calls. This additional tool was utilized during 4–9 February as a parallel method for RIM data log tool in order to log the time slot occupancy information.
- Field Test software of the Nokia E71 terminals for the general received power level revisions and other GSM parameter observations.
- Engineering mode of DLDC terminal was utilized for received power level observations as well as for disabling and enabling DLDC functionality.

11.11.5 Test Cases and Results

11.11.5.1 Data Rates

In a single DLDC cell with 2 TRXs, the data rate was investigated in different configurations. The data rate was studied by downloading a 3 MB FTP file.

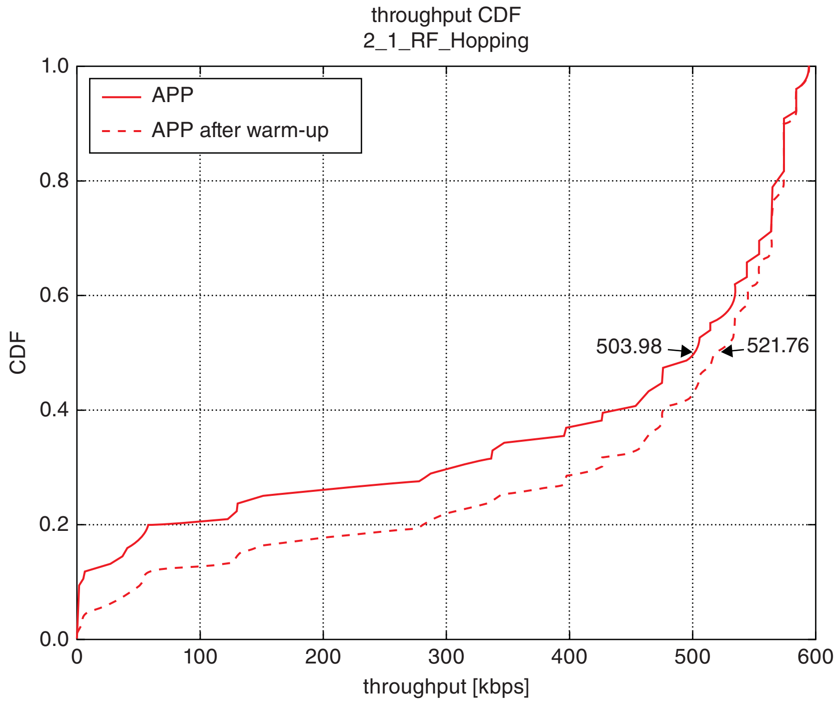

Figure 11.56 *The DLDC throughput in RF hopping case, CDF.*

For the RF hopping, the results as illustrated in Figures 11.56–11.59 show a throughput via 50%-ile CDF:

- 503 kb/s (total)
- 521 kb/s (after warmup).

For the baseband hopping, with Hopping Sequence Numbers of HSN1 = 0 and HSN2 = 2 for the used carriers, and with the same 3 MB FTP download, the results were the following, observing the 50%-ile CDF:

- 542 kb/s (total)
- 556 kb/s (after warmup).

11.11.5.2 Ping Performance

The PING latency performance was executed by utilizing a single cell with two TRXs for DLDC. The PDP context was activated and the PING series was done in MS-DOS prompt. The test case contained the execution of following PING commands with small, medium and big PING packets:

- ping -n 20 -l 32 *IP Address*
- ping -n 20 -l 256 *IP Address*
- ping -n 20 -l 1024 *IP Address*.

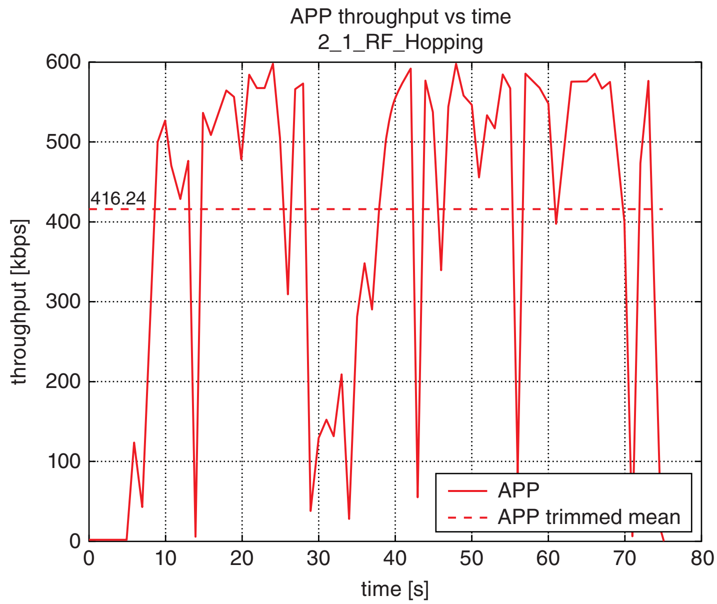

Figure 11.57 *The DLDC throughput in RF hopping case as a function of time.*

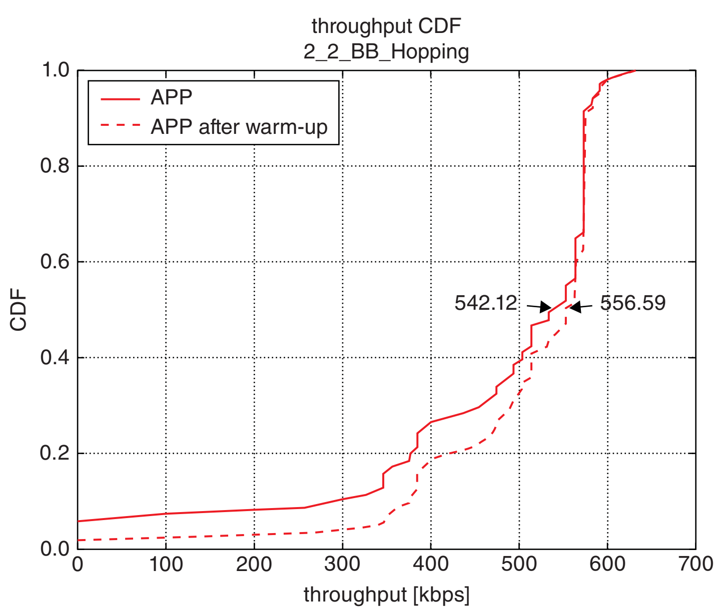

Figure 11.58 *The Base Band hopping case, CDF.*

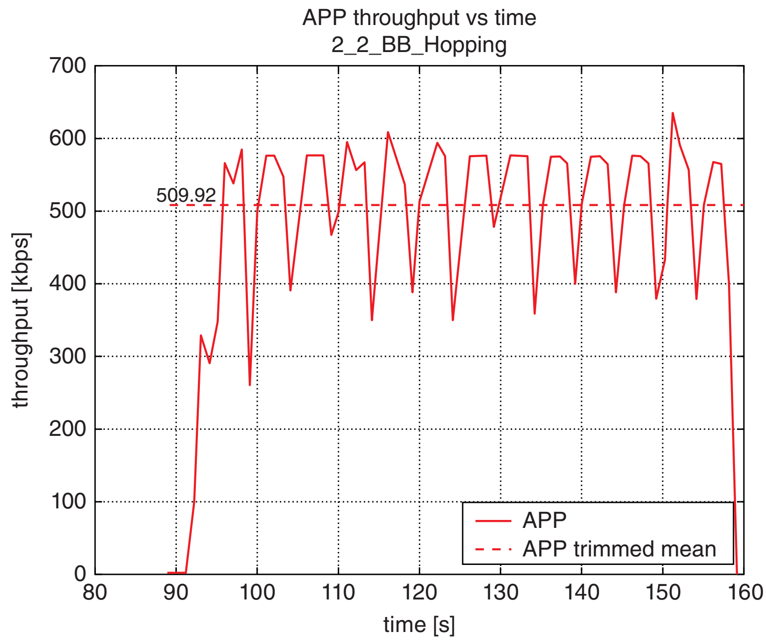

Figure 11.59 The Base Band hopping results as a function of time.

In these cases, the *IP address* was the FTP server utilized in the previous throughput test case. The above shown PING series was repeated 5 times in order to achieve statistical reliability. The same DLDC terminal was defined as EDGE capable via the engineering screen (by disabling the DLDC). The use of the same terminal eliminates the possible unit specific differences in the performance. Then, the PING series was repeated 5 times, as shown in Table 11.13.

The averaged PING response time (ms) of all the 5 repeated series (the first PING of each case has been removed from the calculation due to the establishment period):

11.11.5.3 Voice and Data Timeslot Allocation Algorithm Performance

The functional test with a single cell that contained 2 TRXs shows the dynamic behavior of DLDC in NSN environment. The test was carried out in such a way that there was a DLDC downloading 100 MB file, and meanwhile, CS calls in the same cell: first 1 pair (2 MSs), the 2nd pair added (4 MSs), and so on until the

Table 11.13 PING response times of the experimental setup

	Average (ms)		
Packet size	DLDC	EDGE	Difference (ms)
32 bytes	328	471	143
256 bytes	332	469	137
1024 bytes	515	650	135

cell is full. The cell was defined for 2 TRXs and 1 BCCH TSL in the following way, leaving 8 + 8 TSLs for the PS and CS traffic:

Cell 1 (B=BCCH, X=blocked, Ic=reserved CS idle channel):

TSL:	0	1	2	3	4	5	6	7
TRX1	B	X	X	X	X	X	X	X
TRX2	X	X	X	X	X	X	X	X
TRX3	Ic	Ic	—	—	—	—	—	—
TRX4	—	—	—	—	—	—	—	—

First, 6 CS calls were established, showing the CS territory filling in order:

Cell 1 (B=BCCH, X=blocked, C=CS calls):

TSL:	0	1	2	3	4	5	6	7
TRX1	B	X	X	X	X	X	X	X
TRX2	X	X	X	X	X	X	X	X
TRX3	C	C	C	C	C	C	Ic	Ic
TRX4	—	—	—	—	—	—	—	—

Then, the CS call in the middle was terminated:

Cell 1 (B=BCCH, X=blocked, C=CS call):

TSL:	0	1	2	3	4	5	6	7
TRX1	B	X	X	X	X	X	X	X
TRX2	X	X	X	X	X	X	X	X
TRX3	C	C	Ic	Ic	C	C	—	—
TRX4	—	—	—	—	—	—	—	—

The CS territory was modified as the idle CS TSLs were moved, but the active CS TSLs in TSL 4 and 5 were not moved (it is not necessary in order to minimize the signaling). The 2 idle TSLs for CS are needed due to the minimum requirement of the PS upgrade procedure. Then, the DLDC call was established, which did not trigger the physical TSL rearrangement of the CS territory (still, 2 idle CS TSLs are needed):

Cell 1 (B=BCCH, X=blocked, C=CS call, P=DLDC call):

TSL:	0	1	2	3	4	5	6	7
TRX1	B	X	X	X	X	X	X	X
TRX2	X	X	X	X	X	X	X	X
TRX3	C	C	Ic	Ic	C	C	P	P
TRX4	—	—	—	P	P	P	P	P

Then the CS calls of TSL 4 and 5 was terminated, which triggered the active CS TSL modification, leaving still 2 CS idle TSLs for CS territory (according to the minimum requirement after the PS upgrade):

Cell 1 (B=BCCH, X=blocked, C=CS call, P=DLDC call):

TSL:	0	1	2	3	4	5	6	7
TRX1	B	X	X	X	X	X	X	X
TRX2	X	X	X	X	X	X	X	X
TRX3	C	C	Ic	Ic	P	P	P	P
TRX4	—	—	—	P	P	P	P	P

Then, the last CS call was terminated, liberating sufficiently TSLs for DLDC:

Cell 1 (B=BCCH, X=blocked, C=CS call, P=DLDC call, Ic=CS idle channel):

TSL:	0	1	2	3	4	5	6	7
TRX1	B	X	X	X	X	X	X	X
TRX2	X	X	X	X	X	X	X	X
TRX3	Ic	Ic	—	P	P	P	P	P
TRX4	—	—	—	P	P	P	P	P

Then, during the DLDC still active, 2 CS calls were added:

Cell 1 (B=BCCH, X=blocked, C=CS call, P=DLDC call, Ic= CS idle channel):

TSL:	0	1	2	3	4	5	6	7
TRX1	B	X	X	X	X	X	X	X
TRX2	X	X	X	X	X	X	X	X
TRX3	C	C	Ic	P	P	P	P	P
TRX4	—	—	—	P	P	P	P	P

The CS territory does need in this case only one idle CS TSL due to the PS downgrade requirement of at least one free CS time slot. Then, 2 more CS calls were added, that is, total of 4 CS calls, which resulted 5+3 DLDC configuration and one idle CS TSL. This is again logical and fulfils the requirements of the downgrade procedure.

Cell 1 (B=BCCH, X=blocked, C=CS call, P=DLDC call, Ic=CS idle channel):

TSL:	0	1	2	3	4	5	6	7
TRX1	B	X	X	X	X	X	X	X
TRX2	X	X	X	X	X	X	X	X
TRX3	C	C	C	C	Ic	P	P	P
TRX4	—	—	—	P	P	P	P	P

11.11.6 Analysis

11.11.6.1 Performance

As for the end-user point of view, the usage and functionality of the DLDC terminal is similar with the handling of GPRS/EDGE equipment. The IP throughput of the DLDC was around 500 kb/s with 5 + 5 DL TSL configuration whilst EDGE resulted around 200 kb/s in 4 TSL mode. The PING response times of DLDC were about 130 ms faster in DLDC than in EDGE in all PING sizes.

The DLDC was noted to function as expected in laboratory even with relatively long Abis interface (>3000 km) that was based on IP. The analysis showed that the CS/PS time slot behavior is logical and worked according to the time slot upgrade and downgrade algorithm of NSN.

Time Slot Allocation of DLDC and Impacts The division of the PS and CS territories is done based on the same principle as in GPRS and EDGE of NSN, that is, there are no changes in the functionality due to the DLDC. This means that the CS and PS territory is dynamic. The PS territory should be defined, that is, the TSL capacity that is possible to use for PS calls. As previously, CS calls have always priority over the PS calls. In order to guarantee that the CS calls do have access to the cell even in heavy PS load, idle CS TSLs are defined automatically in CS territory. It is also possible to define dedicated PS TSLs in order to guarantee the respective capacity for the PS users.

Furthermore, PS TSLs can be multiplexed by several users, that is, multiple TBFs can be utilized per TSL, which means that if one or more dedicated TSLs are applied, there is at least some throughput available for PS users even in heavily congested network. Logically, when the maximum number of multiplexed users (that is limited by the number of TBFs and USF) is exceeded, the rest of the PS users experience congestion.

11.12 EDGE2

EDGE2 is the latest evolution for the EDGE. It further increases the delivered data rate by utilizing dynamically modulation methods and additional coding schemes. In fact, there are two variants of EGPRS2:

- Version A (EGPRS2-A DL) contains GMSK, 8-PSK, 16-QAM, 32-QAM, turbo codes, and legacy symbol rate of 271 ksymbols/s. This variant is capable of delivering up to 98 kb/s per timeslot.
- Version B (EGPRS2-B DL) contains GMSK, QPSK, 16-QAM, 32-QAM, turbo codes, higher symbol rate 325 ksymbols/s. This variant is capable of delivering up to 120 kb/s per timeslot.

Furthermore, in the uplink direction, EGPRS2 UL defines QPSK, 16-QAM, 32-QAM and higher symbol rate in uplink. It also has two variants:

- Version A (EGPRS2-A UL) contains GMSK, 8-PSK, 16-QAM, legacy symbol rate 271 ksymbols/s. This variant is capable of delivering up to 81.6 kb/s per timeslot.
- Version B (EGPRS2-B UL) contains GMSK, QPSK, 16-QAM and 32-QAM, higher symbol rate 325 ksymbols/s. This variant is capable of delivering up to 118 kb/s per timeslot.

Both downlink and uplink variants contain new the original channel coding schemes 1–4 that had been introduced since Release 97, and additional DAS-5 to DAS-12 that are able to use 8PSK, 16QAM and 32QAM with turbo codes.

The practical data rates for EGPRS2 can be achieved by utilizing 5 timeslots in downlink (according to class 33). 100–120 kb/s multiplied by 5 times results in about 0.5 Mb/s. Furthermore, there is possibility to combine the dual carrier functionality in downlink which provides about 1 Mb/s data rates. For the uplink direction, class 12 allows the use of 4 time slots. This multiplied by about 80–120 kb/s provides data rates close to 0.5 Mb/s in uplink direction.

References

1. Bjorknas, D., Rautio, J., Penttinen, J. NMTImage. Conference Proceedings. Digital Mobile Radio (DMR) conference, Stockholm, 1994.
2. GSMA. 3 Billion GSM Connections on the Mobile Planet – Reports the GSMA. http://www.gsma.com/newsroom/3-billion-gsm-connections-on-the-mobile-planet-reports-the-gsma/ (last accessed 27 July 2014).
3. 4G Americas sttistics. http://www.4gamericas.org/index.cfm?fuseaction=page&pageid=2042 (last accessed 27 July 2014).
4. Penttinen, J.T.J., Calabrese, F.D., Niemelä, K., *et al.* (2010) Performance model for orthogonal sub channel in noise-limited environment. Conference proceedings. The Sixth International Conference on Wireless and Mobile Communications (ICWMC), Valencia, Spain, Sep. 20–25, 2010, 6 p.
5. Penttinen, J.T.J., Calabrese, F.D., Maestro, L., *et al.* (2010) Capacity gain estimation for orthogonal sub channel. Conference proceedings. The Sixth International Conference on Wireless and Mobile Communications (ICWMC), Valencia, Spain, Sep. 20–25, 2010, 6 p.
6. Ahmed, M.H., Wei, W. and Mahmoud, S.A. (2000) Downlink capacity enhancement in GSM systems with frequency hopping and multiple beam smart antennas. IEEE International Conference on Communications 2000 (ICC), Volume 2, pp. 1015–1019.
7. Barreto, A.N. and Pirhonen, R. (2006) Capacity increase in GSM networks using source-adaptive AMR. IEEE 63rd Vehicular Technology Conference (VTC) 2006-Spring, Volume 2, pp. 553–557.
8. Gimenez, J., Tapia, P., Salmenkaita, M. and Fernandez-Navarro, M. (2002) Analysis of dynamic frequency and channel assignment in irregular network environment. The 5th International Symposium on Wireless Personal Multimedia Communications, Volume 2, pp. 863–867.
9. Barreto, A.N., Garcia, L.G.U. and Souza, E. (2007) GERAN evolution for increased speech capacity. Vehicular Technology Conference VTC2007-Spring, pp. 1287–1291.
10. Moisio, M. and Nikkarinen, S. (2004) Capacity gains of single antenna interference cancellation in GSM. 15th IEEE International Symposium on Personal, Indoor and Mobile Communications, PIMRS 2004, 5 p.
11. EDGE – Evolution of Binary and Quaternary Offset QAM. Tdoc EDGE 007/98. ETSI Joint SMG 1,2,3 and 4 EDGE Workshop #1, Helsinki, Finland, 16–17 February 1998.
12. Evaluation of Continuous Phase Modulation. Tdoc EDGE 04/98. ETSI Joint SMG 1,2,3 and 4 EDGE Workshop #1, Helsinki, Finland, 16–17 February 1998.
13. Holma, H. and Toskala, A. (2009) *LTE for UMTS; OFDMA and SC-FDMA Based Radio Access*. John Wiley & Sons, Ltd, p. 433. ISBN 978-0-470-99401-6
14. ETSI GSM TS 04.60. GERAN GPRS MS-BSS interface, RLC/MAC protocol, release 99 (2001-01). 304 p.
15. ETSI GSM TS 43.05. GSM/EDGE radio access network (GERAN), overall description, stage 2 (release 4).
16. ETSI GSM 02.09. The protection of GSM.
17. ETSI GSM 03.20. Protection
18. ETSI GSM TS 43.059. Positioning services in GERAN.
19. ETSI GSM 03.04. Signalling for the call routing.
20. ETSI GSM 02.79. SOR.
21. ETSI GSM 03.03. Numbering.
22. ETSI GSM 02.17. SIM.
23. Brasche, G. and Walke, B. (August 1997) Concepts, Services, and Protocols of the New GSM Phase 2+ General Packet Radio Service. IEEE Communications Magazine, pp. 94–104.
24. Cai, J. and Goodman, D.J. (October 1997) General Packet Radio Service in GSM. IEEE Communications Magazine, pp. 122–131.

25. Gilchrist, P. and Günther, B. (December 1995/January 1996) General Packet Radio Service on GSM. Mobile Communications International, pp. 50–54.
26. Gilchrist, P. (March 1996) One stop data shop from GSM's GPRS. Mobile Communications International, pp. 62–64.
27. Hämäläinen, J. (1996) Design of GSM High Speed Data Services. Thesis for the degree of Doctor of Technology. Tampere University of Technology, Publications 186. Tampere. 104 p. + publications 1–8.
28. Muthuthamby Sreetharan, Rajiv Kumar. Cellular Digital Packet Data. Artech House, Inc. 1996. ISBN 0-89006-709-0, 315 p.
29. ETSI GSM TS 02.60. GPRS definitions.
30. ETSI GSM TS 03.60. GPRS service description
31. GPRS Conference. 4–5 March 1998, The Walldorf Meridien, London. Centre for EuroTelecomms.
32. Pirhonen, R., Rautava, T., and Penttinen, J. (1999) TDMA convergence for packet data services. *IEEE Personal Communications*, **6**(3), 68–73.
33. ETSI GSM TS 02.33. Lawful Interception, Stage 1.
34. ETSI GSM TS 03.33. Lawful Interception, Stage 2.
35. ETSI GSM TS 04.33. Lawful interception; Stage 3.
36. ETSI GSM TS 03.60. General Packet Radio Service (GPRS); Service description; Stage 2.
37. Penttinen, J., Calabrese, F., Niemelä, K., *et al.* (2010) Performance Model for Orthogonal Sub Channel in Noise-limited Environment. Conference proceedings. International Academy, Research, and Industry Association (IARIA). The Sixth International Conference on Wireless and Mobile Communications, ICWMC 2010, September 20–25, 2010, Valencia, Spain. 6 p.
38. Penttinen, J., Calabrese, F., Maestro, L., *et al.* (2010) Capacity Gain Estimation for Orthogonal Sub Channel. Conference proceedings. International Academy, Research, and Industry Association (IARIA). The Sixth International Conference on Wireless and Mobile Communications, ICWMC 2010, Valencia, Spain, September 20–25, 2010. 6 p.
39. Gimenez, J., Tapia, P., Salmenkaita, M. and Fernandez-Navarro, M. (2002) Analysis of dynamic frequency and channel assignment in irregular network environment. The 5th International Symposium on Wireless Personal Multimedia Communications, Volume 2, 2002, pp. 863–867.
40. Säily, M., Sébire, G. and Riddington, E. (2010) *GSM/EDGE: Evolution and Performance*, John Wiley & Sons, Ltd, p. 504. ISBN 978-0-470-74685-1.
41. Barreto, A.N. and Pirhonen, R. (2006) Capacity Increase in GSM Networks Using Source-Adaptive AMR. Vehicular Technology Conference, 2006. VTC 2006-Spring. IEEE 63rd, Volume 2, pp. 553–557.
42. Ahmed, M.H., Wei, W. and Mahmoud, S.A. (2000) Downlink Capacity Enhancement in GSM Systems with Frequency Hopping and Multiple Beam Smart Antennas. IEEE International Conference on Communications (ICC 2000), 18–22 June, 2000. Volume 2. pp. 1015–1019.
43. Meyer, R., Gerstacker, W.H., Schober, R. and Huber, J.B. (2006) A single antenna interference cancellation for increased GSM capacity. *IEEE Transactions on Wireless Communications*, **5**(7), 1616–1621.
44. Mostafa, A., Kobylinski, R., Konstanic, I. and Austin, M. (2003) Single Antenna Interference Cancellation (SAIC) for GSM Networks. IEEE 58th Vehicular Technology Conference, 2003. VTC 2003-Fall, pp. 1089–1093.
45. Chen, X., Fei, Z., Kuang, J., *et al.* (2008) A Scheme of multi-user reusing one slot on enhancing capacity of GSM/EDGE networks. 11th IEEE Singapore International Conference on Communications Systems, November 2008, pp. 1574–1578.
46. Moisio, M. and Nikkarinen, S. (2004) Capacity Gains of Single Antenna Interference Cancellation in GSM. Symposium on 15th IEEE International Personal, Indoor and Mobile Radio Communications (PIMRC) 2004, Volume 4, pp. 2700–2704.
47. GP-081309. New WID on Voice services over Adaptive Multiuser Orthogonal Subchannels (VAMOS). 3GPP GERAN#39.
48. 3GPP TR 45.903. Feasibility study on Single Antenna Interference Cancellation (SAIC) for GSM networks, V 8.0.0. 18 December 2008.
49. GP-071792. Voice Capacity Evolution with Orthogonal Sub Channels. Nokia Siemens Networks, 3GPP GERAN#36. November 2007.
50. GP-081309. New WID on Voice services over Adaptive Multiuser Orthogonal Subchannels (VAMOS). 3GPP GERAN#39, Florence, Italy, 25–29 August 2008.

51. Technical Specification Group GERAN. Meeting Report. Meeting #42, Shenzhen, P. R. of China, 11–15 May, 2009, 229 p.
52. 3GPP TS 05.05 V8.20.0 (2005–11). Technical Specification Group GSM/EDGE. November 2005. 100 p.
53. Qiao, S. (1998) *A Robust and Efficient Algorithm for Evaluating Erlang B Formula*, Dept. of Comuting and Software, McMaster University, Canada, p. 8.
54. Holma, H. and Toskala, A. (2008) *WCDMA for UMTS; HSPA evolution and LTE*, John Wiley & Sons, Ltd, p. 539. ISBN 978-0-470-31933-8.
55. Lasek, S., Tomeczko, D., Penttinen, J.T.J., Valerdi, D. and Guemes, I. (2011) Performance of Orthogonal Sub Channel with Dynamic Frequency and Channel Allocation. Wireless Telecommunications Symposium (WTS), pp. 1–6.
56. Furskär, A. (2000) GPRS/EDGE overview. Ericsson research, presentation materials.
57. ETSI GSM TS 03.64. GERAN GPRS overall description of the GPRS radio interface, stage 2, release 99. v8.7.0 (2001-01). 57 p.

12

3GPP Mobile Communications: WCDMA and HSPA

Patrick Marsch, Michał Maternia, Michal Panek, Ali Yaver, Ryszard Dokuczał and Rybakowski Marcin

In this chapter, we provide a comprehensive overview on wireless communications according to the 3GPP technologies of Wideband Code Division Multiple Access (WCDMA) and High Speed Packet Access (HSPA). These technologies are used in the Universal Mobile Telecommunication Standard (UMTS) belonging to the 3^{rd} Generation (3G) of mobile cellular solutions.

The chapter is structured as follows: In Section 12.1, the WCDMA and HSPA network architecture is explained, such as the roles of the Radio Network Controller (RNC), base stations (which are known in UMTS as Node Bs; these terms are used interchangeably in this chapter) and the interfaces between these. Section 12.2 looks into physical layer aspects, starting first with CDMA basics such as spreading and scrambling, channel estimation, equalization and power control, after which it is explained how actual transmissions take place in WCDMA and HSPA, and which transport and physical channels are defined for this. Section 12.3 then describes radio link procedures, such as cell search, synchronization, registration of a terminal to a cell, call setup, scheduling and handover. Section 12.4 gives an overview on all major features that were introduced to UMTS after Rel. 5, putting a particular emphasis on multicarrier and multiantenna functionality, multipoint transmission, heterogeneous networks and self-organizing networks. Finally, Section 12.5 covers planning and dimensioning of WCDMA and HSPA systems.

12.1 Network Architecture

UMTS network architecture can be organized in three cooperating parts: User Equipment (UE), UMTS Terrestrial Radio Access Network (UTRAN) and Core Network (CN) [1, 2]. UE is the end-user device that provides the interface for the wireless communication functionality, UTRAN is responsible for the utilization of the air interface while the CN deals with routing the user traffic to external networks and with the overall

The Telecommunications Handbook: Engineering Guidelines for Fixed, Mobile and Satellite Systems, First Edition.
Edited by Jyrki T. J. Penttinen.

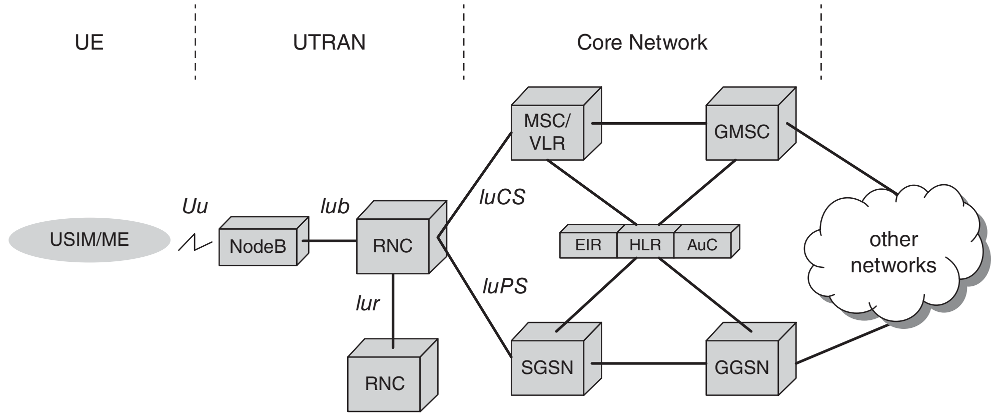

Figure 12.1 Basic system architecture of UMTS with UTRAN interfaces highlighted.

network management. This network architecture with separated network elements and UTRAN interfaces is presented in Figure 12.1.

- UE consist of two modules:
  - Universal Subscriber Identity Module (USIM) which is a smartcard that holds subscriber identity and is responsible for authentication procedures,
  - Mobile Equipment (ME) responsible for radio communication with the Node B over the Uu interface.
- UTRAN consists of two essential network elements: Node B and RNC which are transferring data between each other using Iub interface. Additionally, RNCs are able to communicate with each other over Iur interface.
- CN can be divided into:
  - Circuit switched domain, consisting of Mobile Switching Center (MSC) combined with Visitors Location Register (VLR) and Gateway MSC (GMSC). Circuit switched domain is connected to UTRAN via MSC and IuCS interface.
  - Packet switched domain, consisting of Serving GPRS Support Node (SGSN) and Gateway GPRS Support Node (SGSN). Packet switched domain is connected to UTRAN via SGSN and IuPS interface.
  - Common databases of Equipment Identity Register (EIR), Home Location Register (HLR) and Authentication Center (AuC).

Further explanation of CN elements functionality can be found in Chapters 11 and 25.

The bottleneck of modern wireless network is usually the radio interface, as this over the air connection is in most cases the limiting factor for the overall network performance. From the radio network perspective it is the Node B, managed by RNC, that provides this functionality. In order to properly understand the characteristics of WCDMA and HSPA it is necessary to understand first the functioning of those UTRAN elements. In Sections 12.1.1 and 12.1.2, the high level architectural aspects of Node B and RNC are presented, followed by the description of UTRAN interfaces in Section 12.1.3.

12.1.1 Node B – Base Station

From the functional point of view, the Node B's main functions are radio interface processing and management such as conversion of baseband signals to radio frequency (RF) form suitable for over the air transmission (and

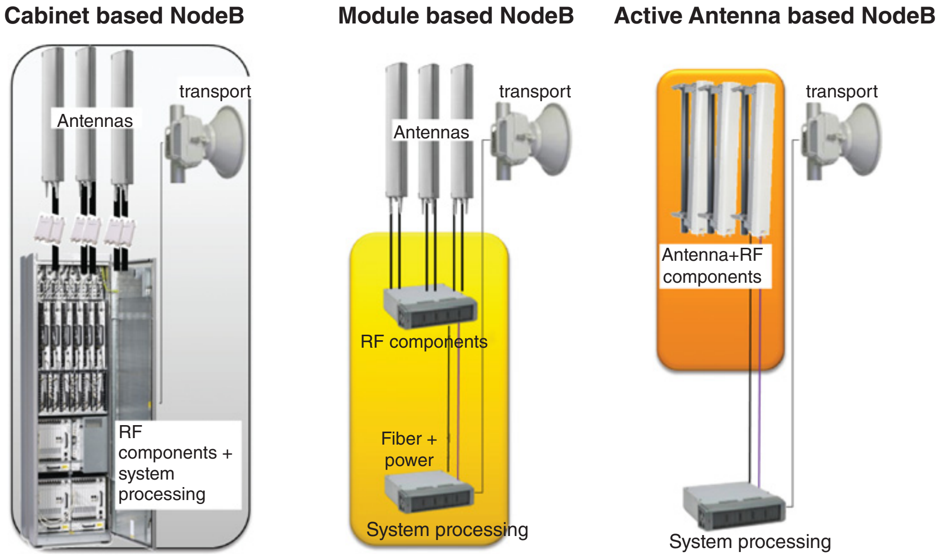

Figure 12.2 *Examples of Node B technology and architecture*

vice versa), encoding and decoding of the data in order to make it robust to the possible errors experienced in the radio interface, spreading and despreading of the data, power control and handling of physical channels. Base station is also responsible for transport interface management, operation and management aspects like interfacing with RNC or Network Management System (NMS).

The examples of NodeB architecture and the technology of the basic building blocks are depicted in Figure 12.2. In the basic configuration of the cabinet based architecture, the RF components and system processing functions are located in the cards installed in rack. This kind of NodeB requires air conditioning.

The next step in node architecture is a module based Node B, where base station functions are integrated in two main blocks. The first one, the system processing block, is responsible for all baseband functions, transport interface, operation and maintenance, and power distribution. The second one, the radio module, is responsible for radio frequency processing and antenna line control. This kind of architecture does not require air conditioning, and the main benefit of this approach is that the radio module can be installed near the antenna which minimizes significantly the radio frequency signal loss and distortion in the feeder cable.

The third and latest architectural solution is based on a so-called Active Antenna System (AAS), where all RF processing components are integrated in the antenna module. The AAS is connected to the system processing module via optical fiber, which means that an RF feeder cable is not required any more. The integration of RF processing inside every antenna element allows among other benefits – direct control of antenna beam pattern tilting.

12.1.2 Radio Network Controller

RNC is a part of the UTRAN which controls and distributes radio resources that belong to its domain comprising of all NodeBs which are connected to it. There are three logical roles of the RNC [3, 4]:

12.1.2.1 Controlling RNC (CRNC)

This is the RNC which is terminating the Iub interface towards the Node B. The illustration of this concept is depicted in Figure 12.3.

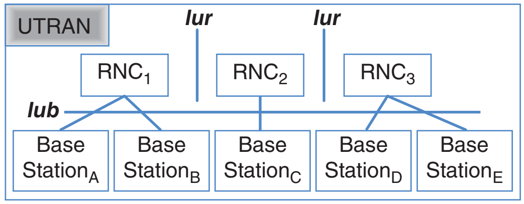

Figure 12.3 *Illustration of the Controlling RNCs.*

RNC1 is a CRNC for Base Stations A and B, RNC2 is CRNC for Base Station C and RNC3 is a CRNC for Base Station D and E. Each CRNC is responsible for load and congestion control for each Node B that it has established an Iub interface with. Additionally, CRNC is managing new radio links which are established in the cells operated by its Base Stations in the matter of admission control and code allocation.

In the mobility situation when resources coming from more than one RNC can be utilized we can differentiate two roles of involved RNCs:

12.1.2.2 Serving RNC

This is the RNC which terminates the Iu (IuPS or IuCS) interface and which is used for transport of user data and signaling to and from the Core Network. The UE which is connected to the UTRAN can only have one SRNC. This concept is depicted in Figure 12.4.

In Figure 12.4, we can see that the UE is simultaneously connected to Base Stations A and C, which are managed by RNC1 and RNC2, respectively. In this case RNC1 is the SRNC due to the fact it has terminated the Iu link. RNC2 is, in this case, a Drift RNC (DRNC).

12.1.2.3 Drift RNC

This is any RNC other than SRNC which is controlling the cells where the UE is utilizing its resources, as depicted in Figure 12.4. In contrary to SRNC there can be zero, one or more DRNCs involved in UE operability. Those cases are depicted on the Figure 12.5:

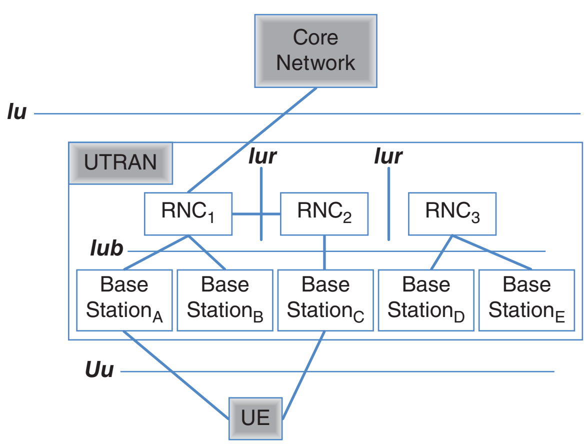

Figure 12.4 *Illustration of the Serving RNC.*

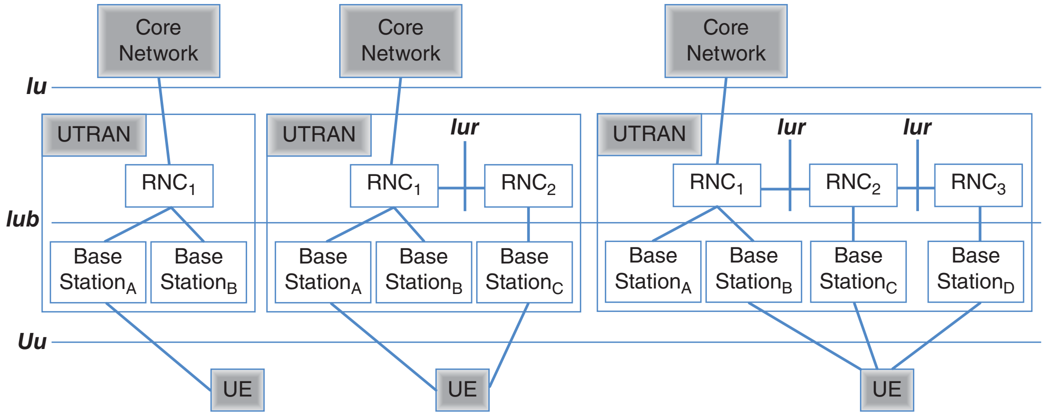

Figure 12.5 *Illustration of zero, one and more DRNCs.*

On the left hand side of Figure 12.5 a situation with only SRNC present and no DRNC is illustrated, in the middle case the UE is connected to the core network using an SRNC (RNC1) and one DRNC (RNC2). Finally, the right hand side depicts a situation where RNC1 is the SRNC and RNC2 and RNC3 are the DRNCs.

12.1.3 UTRAN Interfaces

RNC uses three main interfaces in order to communicate with other parts of the UMTS network: Iub, Iur and Iu (IuPS + IuCS).

Using **Iub** interface, the RNC and Node B decide how the resources should be split among the UEs in order to meet Quality of Service (QoS) requirements of the particular services.

Via **Iur** interface, the RNC can communicate with other RNCs. The Iur protocol called Radio Network Subscription Application Part (RNSAP) can manage four different functionalities which can be implemented independently depending on the requirements for UTRAN. The first, and actually the most important one, is support of the basic inter-RNC mobility, which provides the mobility for the UE connected to the core network via two or more RNCs. The second one provides support for Dedicated Channel traffic, while the third one provides support for Common Channel Traffic. By this functionality handling of the common channels like Forward Access Channel (FACH) and Random Access Channel (RACH) across the Iur interface is possible. The last, optional, functionality provides support of Global Resource Management. If it is implemented, the signaling which supports enhanced radio resource management and operation and maintenance features over the Iur interface is enabled.

The third interface is **Iu** which connects an RNC to the CN. Iu interface was designed as an open interface, i.e. providing the opportunity to connect UTRAN and CN elements from different vendors. Iu interface can operate in two basic instances:

- IuPS connects UTRAN to packet switched CN. There can't be more than one Iu interface towards the PS domain from one RNC;
- IuCS connects UTRAN to circuit switched CN. Similarly to IuPS there can't be more than one Iu interface towards its default CN node within the CS domain;

In addition to the above-mentioned interfaces, IuBC can be added. IuBC is a control plane-only interface that connects the broadcast domain of the CN with the RNC in UTRAN.

The main functionalities of the Iu interface are procedures needed to establish, maintain and release Radio Access Bearers, transferring signaling between UEs and the CN, and enabling simultaneous access to multiple CN domains for each UE. The Iu interface offers Serving Radio Network Subsystem (SRNS, an RNC and its corresponding Node Bs) relocation, intra- and intersystem handovers [5].

The last of the mentioned interfaces, **Uu**, is probably the one that is the most important of all interfaces. It connects the UE to the fixed part of the network and is realized using WCDMA approach. Examples of the most important aspects of this interface like transmission of the radio signals via spreading and scrambling, along with power control and physical channel handling can be found in the section below.

12.2 Physical Layer Aspects

12.2.1 Spreading and Scrambling

A core aspect of WCDMA and HSPA transmission is that of spreading and despreading. *Spreading* means that each symbol to be transmitted (representing a certain number of coded bits) is multiplied with a predefined binary sequence of so-called *chips*, after which each chip is then pulse-shaped and modulated onto the carrier signal. The number of chips per symbol, that is, the ratio between chip rate and symbol rate, is called the *spreading factor* (SF). An example for a baseband signal before and after spreading with SF 8 is illustrated in Figure 12.6. Due to the spreading, the modulated signal occupies an SF times larger bandwidth than the one used if each symbol would be pulse-shaped and modulated onto the carrier directly. For this reason, WCDMA and HSPA are also referred to as communication systems based on *spread spectrum* transmission. The difference in occupied bandwidth for systems without or with spreading is also shown in Figure 12.6.

The benefit of purposely spreading the signal over a larger bandwidth is an improved robustness of the transmission. At the receiver side, the signal is despreaded again, meaning that the down-converted and matched-filtered chip sequence is multiplied with the same binary spreading sequence that was applied at the receiver, yielding the original symbols again, but such that each symbol is replicated in the form of multiple chips, according to the SF. A typical receiver would then sum up each symbol over the corresponding chips, meaning that the desired signal amplitude constructively sums up, while the amplitudes of any noise and

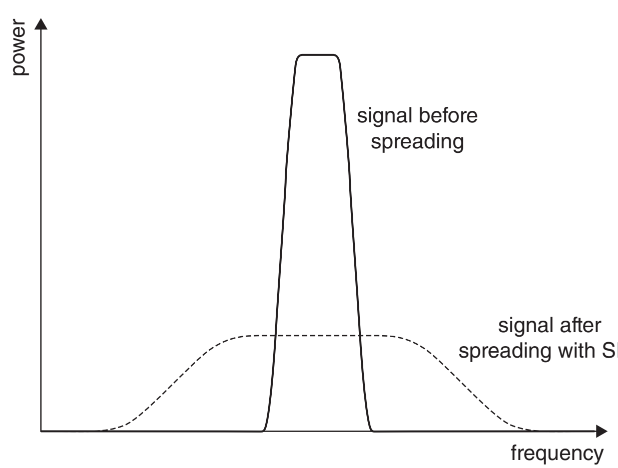

Figure 12.6 *Power spectral density of a signal before and after spreading with SF 8. A raised cosine pulse shape is assumed.*

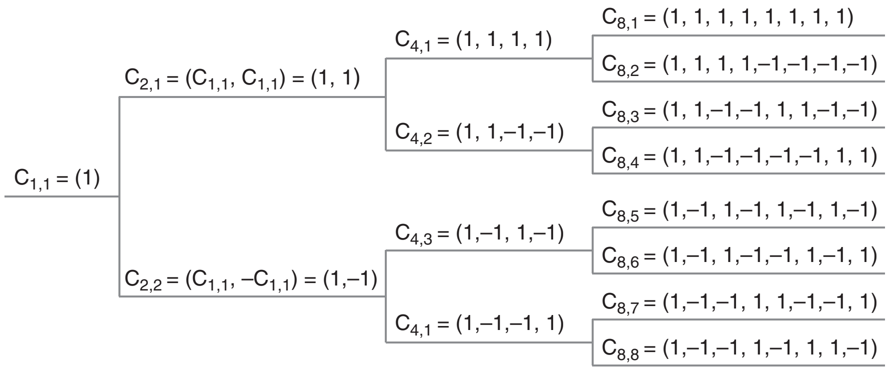

Figure 12.7 *A typical code tree for spreading codes up to SF 8.*

interference terms do not constructively sum up. As a result, the power ratio between the desired signal and any interference and noise terms is boosted by a factor equivalent to the SF[1], a benefit which is commonly referred to as the *processing gain*.

Through the use of spread spectrum communication, one can hence enable reliable communication even under very low *signal-to-interference and noise ratios* (*SINRs*) at the receiver before signal processing, as the processing gain can then lift the SINRs up to decodable levels again.

Clearly, one important aspect of systems based on spreading and dispreading is the choice of appropriate spreading sequences, in particular if multiple transmissions connected to different channels or different users are to be multiplexed. If one is able to synchronize all transmissions to be multiplexed, which is typically the case in the downlink of cellular systems, one should choose the spreading sequences connected to different transmissions such that these are mutually orthogonal. On the other hand, spreading sequences should be designed in a way that for different concurrent transmissions there remains a certain flexibility in the choice of the spreading factor, such that the trade-off between data rate and robustness can be adjusted according to application need.

Most spread spectrum systems fulfill both requirements through the usage of so-called *code trees*, where the spreading codes as such are typically based on Hadamard sequences. An example code tree for spreading codes up to SF 8 is illustrated in Figure 12.7. Here, the left end of the tree provides a spreading code with SF 1, meaning that if this code is used, each chip is connected to the transmission of an individual symbol, providing the highest data rate, but the lowest robustness. Instead of using this code, one could also use two spreading codes of SF 2, for example transmitted to two different users.

As indicated in the figure, one could also use any other selection of codes, as long as each branch of the code tree is only used once, as then the Hadamard sequences avoid mutual interference among transmissions. Note that zero mutual interference only holds at the transmitter side and if transmission takes place over a frequency flat channel. If transmission happens over a frequency-selective channel, the receiver side will experience some extent of interference among originally orthogonal parts of the code tree, as we will see later.

In a cellular uplink, it is typically not feasible and also not desired to synchronize the transmissions of multiple users, and hence it is also not possible to design spreading codes such that mutual orthogonality

[1]More precisely, the ratio between desired signal power and interference plus thermal noise is boosted by SF^2, but as the receiver also picks up SF times more interference and thermal noise due to the larger bandwidth of the transmission, the overall processing gain is equivalent to the SF.

among transmissions is guaranteed. The same holds for concurrent transmissions from multiple base stations in the downlink, where it is also not desirable to enforce synchronization and coordinate the usage of code trees. Hence, for all transmissions originating from the same transmitter, it is possible to use code trees and obtain orthogonality as stated before, but one needs other means to alleviate the impact of mutual interference between transmissions originating from different entities.

In most spread spectrum systems, one applies transmitter-dependent *scrambling*. Scrambling means that the signals after spreading, but before pulse-shaping are multiplied with long, transmitter-specific binary sequences. By doing so, interference among transmissions cannot be avoided, but it is at least randomized such that it appears to unintended receivers as thermal noise, and these can hence exploit the full processing gain for their desired signals. WCDMA and HSPA typically use scrambling codes with a length of 256 chips or 38 400 chips.

12.2.2 Channel Estimation

One disadvantage of spread spectrum transmission is that this is particularly sensitive to frequency selective channels. As the chip rate is very high, transmission even over an only moderately frequency-selective and hence time-dispersive channel already leads to the fact that each transmitted chip is smeared over multiple chip lengths at the receiver side. As an example, let us consider WCDMA, which typically uses a chip rate of 3.84 MHz, hence where the chip length is roughly a quarter of a microsecond. As communication channels in urban environments can easily lead to multipath fading with delay differences of multiple microseconds, this means that the power connected to a transmitted chip is potentially scattered over 5–15 chips at the receiver side. This inter-chip-interference has to be adequately addressed, and for this reason it is first of all necessary that the receiver obtains a reliable estimate of the characteristics of the channel, that is, of the channel transfer function or (equivalently) the channel impulse response.

To facilitate channel estimation, WCDMA employs *reference signals* in uplink and downlink, which are known to the receiver side and transmitted with a spreading factor of 256 (see also Section 12.2.6).

The most straightforward approach to estimate the channel impulse response is to correlate the received chip sequence after down-conversion and matched-filtering with the known spreading and scrambling sequence of the reference signal. This correlation is performed for different timing offsets in order to capture the different taps of the channel impulse response. Note that the receiver can limit the number of applied correlation timing offsets to a very reasonable extent, as it has already obtained synchronization before and hence knows in which range of offsets to expect the dominant tap of the channel impulse response. While a so-called *correlator* estimator is simple to realize (and is in fact the most widely used channel estimator for WCDMA and HSPA systems), it has the downside that it cannot avoid a systematic error in the channel estimation result. This is due to the fact that timeshifted versions of the same spreading and scrambling sequence are not mutually orthogonal, hence the estimation of any tap of the channel impulse response is always subject to some interference from other taps. While this interference is suppressed through the high spreading factor of 256 (and potentially an even longer estimation window used), it can still pose a problem, for instance when weak channel taps are to be estimated, which are then interfered by the stronger taps.

As the autocorrelation between timeshifted versions of the reference spreading and scrambling sequence is known, a more sophisticated channel estimator may exploit this information to suppress the systematic error stated before. A channel estimator which is optimal in the sense of minimizing the average power of the channel estimation error on each tap is a so-called Minimum Mean Square Error (MMSE) channel estimator, which is described in detail in, for example, [6]. The difference in channel estimation performance between a simple correlator and an MMSE channel estimator is that while the mean square error obtained with the MMSE estimator consequently decreases as the signal-to-noise ratio (SNR) increases, the correlator reaches an error floor as at some point the above mentioned systematic estimation error becomes dominant.

12.2.3 Equalization

After the channel impulse response has been estimated, a receiver can now attempt to equalize the received chips in order to minimize the impact of interchip-interference and in fact even exploit the fact that the transmission has seen scattering. In the sequel, we describe three popular forms of equalizers:

A *Rake receiver* is the most widespread equalizer used in today's WCDMA and HSPA chipsets, in particular due to its easy and efficient implementation in hardware. A Rake receiver aims at capturing a major portion of the signal power connected to each transmitted chip by employing a set of "fingers" which sample the received chip sequence under different delays corresponding to the strongest taps of the estimated channel impulse response. In each "finger," the received chip is phase-adjusted and weighted according to the estimated channel, before the signals from all fingers are combined. A typical Rake receiver implementation is shown in Figure 12.8. Such a receiver is equivalent to a so-called Maximum Ratio Combining (MRC) filter (at least if it would employ an arbitrary number of "fingers"), and is able to exploit so-called multipath diversity by coherently combining the signals that have experienced different propagation paths.

A downside of the Rake receiver is that it does not sufficiently address the issue of interchip interference. More precisely, it inherently alleviates the impact of interchip interference to some extent through the phase-coherent alignment of desired signal components, but it does not actively suppress interference. If it is desired to fully suppress interchip-interference, a so-called *zero-forcing* (*ZF*) equalizer can be used. This is a filter applied to a consecutive sequence of received chips, which basically inverts the impact of the channel. While the ZF equalizer may seem appealing at first sight, it has the downsides that it sacrifices some extent of multipath diversity for the sake of interference suppression and it may lead to an amplification of noise. Further, it can become numerically instable if the expression in the equation above to be inverted becomes close to singular. An improved equalizer which is in fact optimal in the sense of minimizing the overall symbol error power after equalization is a so-called MMSE equalizer.

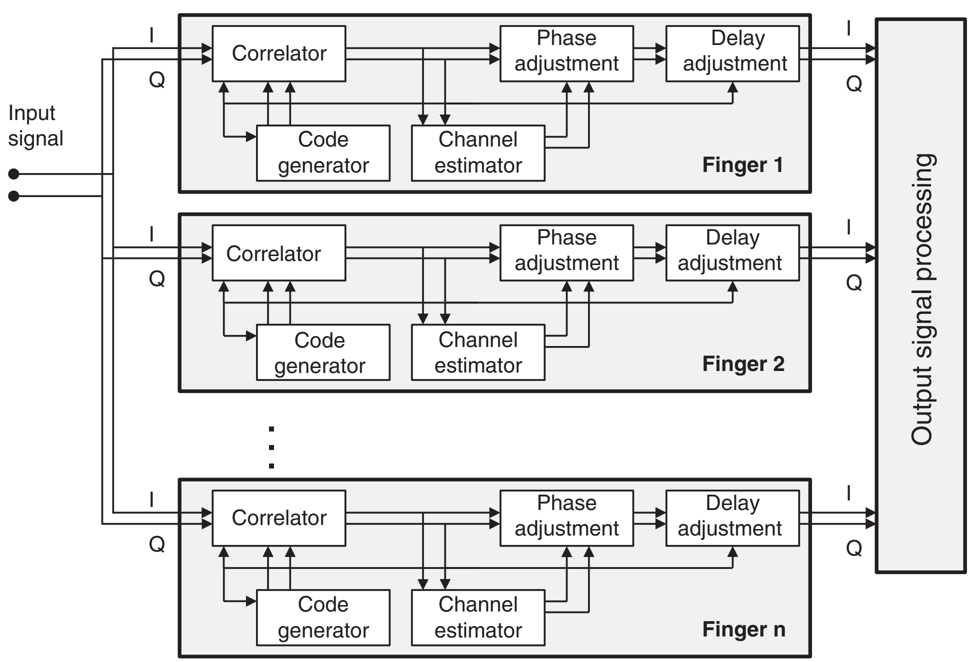

Figure 12.8 *Illustration of a typical Rake receiver.*

Note that we have so far only been concerned with overcoming interchip-interference introduced by transmission over a frequency-selective channel. In the context of multiple-antenna transmission and reception (see Section 12.4.2) or simultaneous transmission from multiple base stations (see Section 12.4.3), however, it may be necessary to use equalization to also alleviate the mutual interference between different concurring transmissions. Here, the same MMSE equalization principle can be applied, with the only difference that the equalizer is fed with the signals from multiple receive antennas, giving it additional degrees of freedom to reject a certain extent of interference. In HSPA Release 7, one refers to a set of receiver categories which differ in the number of receive antennas available and in the knowledge on interference which is used to calculate the MMSE filter coefficients:

- Type 2 receiver: Employs only one receive antenna
- Type 3 receiver: Employs two receive antennas and can equalize up to two spatial streams, but treats interference from other cells as spatially white interference (i.e. it is not able to exploit its spatial properties)
- Type 3i receiver: Employs two receive antennas and can equalize up to two spatial streams, and exploits knowledge on the spatial properties of other-cell interference to reject this as much as possible.

12.2.4 Power Control

Power control is an important aspect in any cellular communications system, but particularly important in the uplink of any CDMA-based system to overcome the so-called *near-far-problem*. This problem occurs when a base station receives multiple transmissions at a very different power levels and consequently cannot correctly decode the weaker signals. Note that this problem does not occur (or is much less pronounced) in OFDM systems, as the transmissions of multiple users in one cell are synchronized in time and kept orthogonal, while in CDMA the uplink transmissions of users in the same cell are not synchronized and not orthogonal, hence with a potentially strong residual interference between concurrent transmissions even after despreading. For this reason, both WCDMA and HSUPA use a very fast uplink power control, to ensure that in each time instant all signals received by a base station are on a similar power level. In addition to this, power control is also used in WCDMA and HSPA in both uplink and downlink to either combat large scale channel properties (i.e. to give more link budget to a user at the cell edge or in otherwise unfortunate link conditions), or to counteract fast fading. In the sequel, we will briefly explain the different forms of power control that are applied in WCDMA or HSPA.

12.2.4.1 Open Loop Power Control

In this case, the choice of a certain transmit power is based only on a transmitter-side estimate of the link quality, for instance based on the previous reception of a pilot signal from the other side. Especially in FDD systems, where different carrier frequencies are used in uplink and downlink, this form of power control is very inaccurate, as the small-scale channel properties in one link direction may differ substantially from those in the other direction. In addition, open loop power control cannot take the interference level at the receiver into account, as the transmitter has no means to obtain this information. For this reason, open loop power control in WCDMA/HSPA is only used in the case where it cannot be avoided, namely when a terminal first tries to gain access to a cell by utilizing the random access channel (RACH). For this, the terminal will use the prior reception of synchronization and reference symbols to estimate the link quality and make the best possible “guess” of a suitable uplink transmit power. In IS-95, open loop power control was used in conjunction with closed loop power control schemes which we will explain later, mainly to be able to better respond to very sudden and strong changes in link quality. As closed loop power control schemes have advanced, this hybrid form of schemes has however become obsolete.

12.2.4.2 Closed Loop Power Control

In the case of closed loop power control, feedback is used from the receiver to the transmitter side to adapt the transmit power over time. One here typically uses a so-called inner and outer loop.

The *inner loop* describes the actual feedback loop and the consequent adaptation of transmit power. In the WCDMA uplink, a 1-bit feedback at a rate of 1500 Hz is used, which tells the transmitter to either increase or decrease transmit power as compared to the previous transmit power level. The default step size used here is 1 dB, but this can also be increased. Smaller step sizes can be emulated.

The *outer loop*, typically operating on a much slower time scale of 10–100 Hz, sets the criterion according to which the inner loop is supposed to operate. A common criterion is the target SINR which is to be obtained at the receiver side. In the uplink, this happens at the RNC, while in the downlink the outer loop resides in the terminal. The target SINR is typically adjusted on the basis of *block error rate* (*BLER*) statistics, that is, it is increased if the BLER in the past has been too high for a particular kind of data transmission, or decreased in case it was too low.

12.2.5 Data Transmission in WCDMA and HSPA

After having explained fundamental physical layer basics of CDMA systems in general, we now explain how data is actually transmitted in WCDMA and HSPA, which differs strongly between different releases, and also between uplink and downlink. Note that we here focus only on the most common channels used for data and corresponding control information transmission in WCDMA/HSPA for terminals in connected state (see the different terminal states in Section 12.3) in order to illustrate key concepts and also the evolution from WCDMA to HSPA. All key aspects are summarized in Table 12.1.

For an overview on other transport or physical channels available for data or control information transmission, please refer to Section 12.2.6, or find a complete overview in, for example, [1]. Before we dive into details, Figures 12.9 and 12.10 depict the general frame structure used in WCDMA and HSPA in downlink and uplink, respectively. A chip frequency of 3.84 MHz is used, 2560 chips are grouped into one slot, and 15 slots are grouped into one 10 ms frame. The slot rate is correspondingly 1500 Hz.

12.2.5.1 WCDMA Downlink (DCH, Introduced in Release 99)

General Information In the WCDMA downlink, the most common form of data transmission is via the *Dedicated Channel* (*DCH*). Here, control information and data are multiplexed in time within each symbol. Fast closed-loop power control is used, where the outer loop resides in the terminal, and data transmission is performed by selecting a spreading factor corresponding to the largest expected data rate needed. If the actual bits to be transmitted are then less than what the data channel could transport, the transmission is gated (i.e. turned on or off) on a slot basis, or, if gating is to be avoided, transmitted bits can also be repeated.

Code Tree and Channel Usage Data transmission takes place through the so-called *Dedicated Physical Data Channel* (DPDCH), which uses variable spreading factors between 4 and 512 and employs QPSK modulation. For coding, one uses either convolutional codes at code rate 1/2 or 1/3, or Turbo codes at rate 1/3. As a maximum of 3 codes of spreading factor 4 can be used concurrently, and QPSK enables to embed two bits in each transmitted symbol, the peak data rate for one user under the assumption of half rate coding

Table 12.1 *WCDMA and HSPA key transmission parameters*

	WCDMA		HSPA	
	Downlink (DCH, Rel .99)	Uplink (DCH, Rel. 99)	Downlink (HS-DSCH, Rel. 5)	Uplink (E-DCH, Rel. 6)
Frame structure	10 ms frame length 15 slots per frame 2560 chips per slot			
Transmit Time Interval (TTI)	10 ms, 20 ms or 40 ms		2 ms	2 ms or 10 ms
Multiplexing of data and control information	Control and data time multiplexed within each slot	Control and data modulated on separate I and Q branch of baseband signal	Control and data code-multiplexed	Control and data modulated on separate I and Q branch of baseband signal
Coding schemes	1/2 or 1/3 rate convolutional coding or 1/3 rate Turbo coding		Turbo coding, rates 1/3–1 in conjunction with puncturing	Turbo coding, rate 1/3
Modulation schemes	QPSK	BPSK	QPSK or 16-QAM	BPSK
Usage of channelization codes/spreading factors (SF)	Flexible usage of channelization codes with SF 4- 512 (max 3 codes with SF 4)	Flexible usage of channelization codes with SF 4- 256 (max 6 codes with SF 4)	Fixed SF 16 (max 15 codes with SF 16)	Flexible usage of channelization codes with SF 2 or 4 (max 2 × SF 2 plus 2 × SF 4)
Power control	Fast closed-loop power control on slot basis, outer loop located in terminal	Fast closed-loop power control on slot basis, outer loop located in RNC	Fixed power	Fast closed-loop power control on slot basis, outer loop located in RNC
Rate matching	Gating of transmissions or repetition	Repetition or puncturing	No	No
Adaptive modulation and coding	No		Yes	No
Retransmission	On RLC level		On PHY level	
Soft handover	Yes	Yes	No	Yes
Peak data rates	2.8 Mbps (for code rate 1/2) (in practice 384 kbps, as only one code used)	2.8 Mbps (for code rate 1/2)	14.4 Mbps (for code rate close to 1)	5.76 Mbps (uncoded)

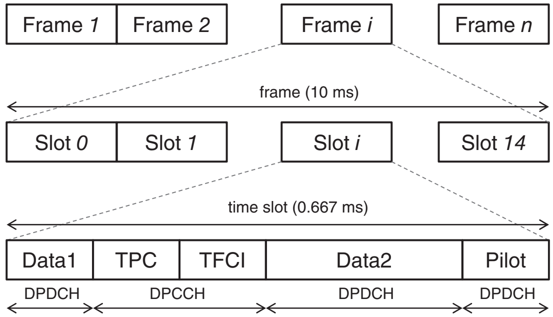

Figure 12.9 *Frame structure used in DL. TPC stands for* Transmit Power Control*, and TFCI for* Transport Format Channel Indicator.

is 2.8 Mbps. As mentioned before, the transmission of data is time-multiplexed within each slot with that of control information in the form of the Dedicated Physical Control Channel (DPCCH). This contains pilots, indicators for the used codes and spreading factors and power control commands for the uplink.

12.2.5.2 WCDMA Uplink (DCH, Introduced in Release 99)

General Information Also in the uplink, the most common form of data transmission in WCDMA is via the *dedicated channel* (*DCH*). One fundamental difference to the downlink is that control information and data are not time-multiplexed, but I/Q-multiplexed by modulating one onto the I-branch and the other onto the Q-branch of the complex-valued baseband signal (see Chapter 10 for details on modulation). The reason is that as opposed to the downlink, where each base station always has some extent of signals to be transmitted at all times (e.g. synchronization channels, common pilot channels etc.), it would happen in the uplink that if a terminal has only control information to send but no data, and these are time-multiplexed, there would be intervals where the transmitter is completely silent.

Especially as WCDMA is operating with a slot interval of 1500 Hz, that is, a frequency that is within the range of what the human ear can perceive, this could lead to *audible interference* if badly protected audio equipment is placed in proximity. A well known example for this is the occasional audible interference resulting from a GSM cellphone being placed close to a radio. Through I/Q-multiplexing of control information

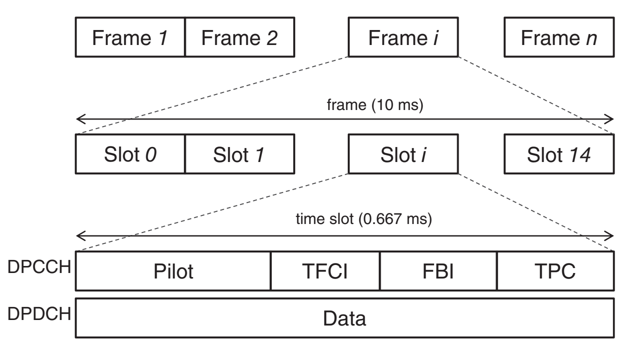

Figure 12.10 *Frame structure used in UL.*

and data, this can be avoided, as then one has a constant transmission, even if no data is to be transmitted. Note that control information and data may be transmitted at different power levels, hence to minimize potential signal envelope variations, a complex scrambling code is applied to the modulated baseband signal before upconversion.

The WCDMA uplink is based on fast closed-loop power control with power control commands in each slot (i.e. at a rate of 1500 Hz), and where the outer loop resides in RNC. As in the downlink, codes and spreading factors are chosen such as to meet the maximum data rate requirements for a terminal, and if less bits have to be carried over the channel, rate matching in the form of repetition coding or puncturing may be applied.

Code Tree and Channel Usage Due to the usage of I/Q-multiplexing, both control and data channels only use BPSK as modulation scheme. In the DCH uplink, control information is transmitted via the *Uplink Dedicated Physical Control Channel* (*Uplink DPCCH*) with a fixed channelization code of spreading factor 256, meaning that one slot is able to transport 10 content bits which contain pilots, information on the chosen data transport format, power control information and feedback information (which is for instance used for downlink transmit diversity, see Section 12.4.2). The actual data transmission takes place via the so-called *Uplink Dedicated Physical Data Channel* (*Uplink DPDCH*) in the form of one or multiple channelization codes from spreading factor 4 to 256.

For coding, convolutional codes at code rate 1/2 or 1/3, or Turbo codes at rate 1/3 are used. As the channelization code used in the control information branch must not be reused in the data branch, the maximum achievable uplink data rate is limited to that obtained through the usage of 6 codes of spreading factor 4 in parallel, which corresponds to 2.8 Mbps if a code rate of 1/2 is considered. Note that it is in general desirable to transmit less channels with lower spreading factor than a higher number of channels with a higher spreading factor, as the latter would lead to an increased peak to average power ratio (PAPR), which reduces power amplifier efficiency.

12.2.5.3 HSDPA (Introduced in Release 5)

General Information High-speed Downlink Packet Access (HSDPA) introduced major changes to the way transmissions are handled, in particular to how the transmission is adapted to changing channel conditions, and how transmission resources are divided among users. While DCH uses fast closed-loop power control to respond to changes in the link quality, and assigns parts of the code tree to terminals according to their maximum data rate needs (with potential repetition coding or discontinuing the transmission if less data is to be transmitted than the code tree allocation allows for), HSDPA keeps the downlink transmit power fixed, but uses *adaptive modulation and coding* (*AMC*) to respond to channel fluctuations, and uses a scheduler located in the base station to redistribute resources according to channel conditions and need on a fast basis (see also Section 12.3. for more details on scheduling). This is particularly beneficial in the context of bursty and best-effort traffic, and also has the advantage that the interference generated towards users in other cells is more constant. Another novelty is that HSDPA uses *Hybrid Automatic Repeat Request* (*HARQ*) functionality on the physical layer, meaning that failed transmissions are repeated and multiple (re)-transmissions are soft-combined at the receiver side.

Adaptive Modulation and Coding in Combination with HARQ As mentioned, HSDPA uses a fixed power for transmission, and as simultaneous transmissions to multiple users are kept moderately orthogonal (depending on how frequency-selective the channel is) through the usage of different parts of the same orthogonal code tree, terminals located in the cell center or at the cell edge will of course experience very different *SINR*. This is exploited in HSDPA through the fact that the modulation order is increased from QPSK to 16-QAM for users with a high SINR, meaning that each transmitted symbol then represents 4 coded bits, and also the code-rate of the used Turbo code can be adapted at a certain granularity within a range from 1/3 to 3/4. In fact, with subsequent puncturing, an effective code rate of close to 1 can be achieved. By choosing different combinations of modulation order and code rate which are roughly equally spaced in terms of required SINR, and by having the option of retransmitting previous failed transmissions and combining these on the physical layer, HSDPA can actually perform rather aggressive scheduling and hence get fairly close to the theoretical channel capacity.

Code Tree and Channel Usage HSDPA uses a fixed spreading factor of 16, where up to 15 channelization codes can be used to form one *High Speed Downlink Shared Channel* (*HS-DSCH*) for data transmission, which is shared among all scheduled users. Due to the maximum number of codes used simultaneously, the maximum modulation and coding scheme of 16 QAM and a code rate after puncturing close to 1, the maximum peak data rate one user can experience is 14.4 Mbit/s. The one remaining channelization code is used for the transmission of control information, e.g., for a so-called *High Speed Shared Control Channel* (*HS-SCCH*), which contains the information for all terminals in the cell which parts of the HS-DSCH are supposed to be despreaded and decoded by which terminals, which modulation and coding schemes have been used, and also whether each scheduled transmission is an initial transmission or a retransmission belonging to a particular previous transmission. As the HS-SCCH has to be decodable by all terminals in a cell, it is transmitted particularly robust with a spreading factor of 128 and a half-rate convolutional code.

Terminals have exactly one slot time to process the control information received via HS-SCCH, meaning that if this contains the information that a data transmission on HS-DSCH is to be decoded, this data transmission starts exactly 2 slots later. Of course, the AMC and HARQ functionality mentioned earlier also requires novel signaling information in the uplink. For this, a *High Speed Dedicated Physical Control Channel* (*HS-DPCCH*) has been introduced which carries both ACK/NACK information on previous downlink transmissions, as well as a 5-bit *Channel Quality Indicator* (*CQI*) which indicates which modulation and coding scheme the terminal would be able to successfully decode under the current channel conditions. This information is reported from each terminal to its assigned base station every 3 slots, and there is the timing relation that the terminals have to provide the ACK/NACK feedback for downlink transmissions around 7.5 slots after the end of these transmissions.

12.2.5.4 HSUPA (Introduced in Release 6)

General Information High-Speed Uplink Packet Access (HSUPA) introduced similar features to the uplink as HSDPA did to the downlink, such as fast and instantaneous need-based scheduling performed by the base station, and *HARQ* on the physical layer. Different from HSDPA, however, HSUPA still uses fast closed-loop power control to overcome the near-far problem (see Section 12.2.4), meaning that all transmissions from multiple users are received at a similar power level, and hence there would not be any benefit from AMC as in the downlink.

Note that in HSUPA, as all users anyway employ different scrambling codes and can hence reuse the same channelization codes, resource scheduling performed by the base station is not about dividing the code tree among multiple users, but about managing the overall uplink interference budget. This is explained in detail in Section 12.3.6. HSUPA also uses shorter transmit time interval (TTI) lengths than uplink DCH, but foresees two different values of either 2 ms or 10 ms TTI length, which different terminals can use concurrently. The reason is that on one hand, one wants to minimize round trip time as much as possible, but for cell edge users the control signaling overhead associated with a short 2 ms TTI length would be too high considering the limited power and interference budget, and hence for these users a 10 ms TTI length is used.

Code Tree and Channel Usage HSUPA introduces a novel physical uplink channel for data transmission called *Enhanced Dedicated Physical Data Channel* (*E-DPDCH*), which uses spreading factors of 2 to 256. A user may simultaneously use up to two codes with SF 2 plus two with SF 4. As the modulation is BPSK as in the case of uplink DCH, the peak data rate a user could experience with HSUPA is 5.76 Mbps. Also, a novel uplink channel for control information called *Enhanced Dedicated Physical Control Channel* (*E-DPCCH*) is introduced, which contains information on the actual code constellation used for any uplink transmission on the E-DPDCH, and bits related to whether a transmission is an initial one or a retransmission of a particular previous transmission. In addition, it contains a so-called *happy bit*, indicating to the base station whether the terminal sees the currently granted uplink data rate as sufficient or would like to demand more.

Note that this novel control channel does not contain any pilots, as these are already contained in the Release 99 DPCCH, which is used in parallel to the new E-DPCCH. Of course, also novel channels had to be introduced in downlink direction to support the new HSUPA functionality. The *HARQ Indicator Channel* (*HICH*) contains information on whether previous transmissions could be decoded successfully. As this is transmitted at a clearly defined delay after the corresponding uplink transmission, the terminals can unambiguously tell which transmission this refers to. The *Enhanced Relative Grant Channel* (*E-RGCH*) and *Enhanced Absolute Grant Channel* (*E-AGCH*) contain information on increasing/reducing the UE's serving grant and on the absolute limit on transmit power, respectively. Here, it is left to the particular base station implementation whether relative grants (i.e. indicating whether a previously granted data rate is to be increased or decreased), absolute grants, or a combination of both is used. See further details on scheduling in Section 12.3.6.

12.2.6 Overview on Transport Channels and Physical Channels

In this section, we provide an overview on the available transport and physical channels in WCDMA and HSPA. Transport channels are logical channels connected to the OSI layers above the physical layer, which are then mapped to one or multiple physical channels in the physical layer. In many cases there is a direct equivalent to a transport channel on the physical layer. There are also physical channels which are not mapped to any transport channel, as they are related only to physical layer procedures.

The following **transport channels** are most relevant in WCDMA and HSPA. All have been introduced in Release 99, except otherwise stated:

- **BCH** (Broadcast Channel). This downlink channel contains information that has to be decoded by each terminal in the cell, such as available random access codes and access slots. This is the first information that terminals will decode after having been powered on. See also Section 12.3.1.
- **DSCH** (Downlink Shared Channel). This transport channel and its corresponding physical channel were foreseen for efficient downlink transmission via code-tree capacity sharing on frame-by frame basis, with scheduling information in the DCH. This channel hence already had part of the functionality provided later in Release 5 in HS-DSCH, and for that reason became obsolete.
- **FACH** (Forward Access Channel). Provides a transport channel for control information or data in the downlink. Multiple FACHs can exist, but one has to be provided which can be decoded by all terminals in the cell.
- **PCH** (Paging Channel). Carries paging information and has to be receivable by all terminals in a cell. See more details on paging in Section 12.3.4.
- **RACH** (Random Access Channel). Uplink transport channel, used by terminals to initially register to a cell after having been powered up, or for connection requests. See also Section 12.3.3.
- **DCH** (Dedicated Channel). Uplink or downlink transport channel used for data communication for terminals in a connected state. See details in Section 12.2.5.

- **HS-DSCH** (High Speed Downlink Shared Channel), introduced in Release 5. Complements DCH in the downlink for data communication for terminals in connected state (starting from Release 7, HS-DSCH can also be used by terminals in state CELL_FACH). See Section 12.2.5.
- **E-DCH** (Enhanced DCH), introduced in Release 6. Complements DCH in the uplink for data communication for terminals in connected state (starting from Release 8 E-DCH can be also used by terminals in state CELL_FACH). See also Section 12.2.5.

Table 12.2 lists the many physical channels that have been defined for WCDMA and HSPA. Note that we here do not list channels that have not been implemented in practice and hence dropped from the standard at some point (e.g. CPCH).

12.3 Radio Interface Procedures

The radio interface is typically the bottleneck of cellular transmission, and wireless bandwidth is usually a scarce and expensive resource. This means that radio resources need to be managed efficiently and sharing them between potential users must be done in an optimal manner. The Radio Resource Control protocol (RRC) [7] is responsible for this task, as shown in Figure 12.11, and in order to efficiently manage available resources, several RRC states are used in WCDMA/HSPA networks.

RRC states are tradeoffs between UE power consumption, location information precision and instantaneously available data rates. In IDLE mode, there is no signaling connection between UE and the network. RRC connection involves a lot of signaling exchange between UE and the network, so the network can be optimized to keep UEs in idle mode as long as possible. However, establishing of RRC connection is also time consuming so new smartphones will try to stay in connected mode to avoid this delay which can influence end-user experience, especially when using applications that need for example, frequent Internet usage.

In the **connected state**, four different RRC states can be used, as presented in Table 12.3. The name of each state is a combination of the location information and available channel that UE and radio network may use to exchange information. These states are:

- Cell PCH. In this state, the location of the UE in the network is known down to cell level. The only way for the network to exchange information with the UE is via paging related channels: PCH and PICH. Paging procedure is further described in Section 12.3.4. This state has a power consumption similar to the idle state with the difference that the RRC connection is established. Low power consumption is obtained via discontinuous reception (DRX) cycles and keeping the transmitter chain of the UE switched off.
- URA PCH. This is a state similar to cell PCH, and the only difference between them is that the UE position in URA PCH state is known in the network on UTRAN Registration Area (URA) level. URA is a cluster of several cells in the same geographical area and its size may be configured by the network operator. URA PCH state is useful in case of frequent cell update procedures. Since the cell update procedure involves signaling exchange between the UE and the network, it creates a potential risk that in case of a fast moving UE (traversing through cells very often) it could generate a large signaling load. To avoid this, a UE may send cell update messages (URA update message) only if it detects that the new cell to camp on belongs to the new URA. A UE may move to this state if it performs a certain number of cell updates in a given time. Both time and number of cell updates necessary to move a UE to the URA PCH state may be configured by the network operator.
- Cell FACH. This is one of the two states when bidirectional communication between the UE and the network is possible. In this particular case, the RACH transport channel may be used to send the uplink information and FACH can be used to send data in the downlink direction. However, since these are shared common channels, data exchange is limited to small portions of information, for example, the signaling exchange when the UE wants to perform a cell update procedure.

Table 12.2 *Physical channels in WCDMA and HSPA*

Rel.	Abbr.	Name	Purpose	Details
Rel. 99	PCCPCH	Primary Common Control Physical Channel	Carries the downlink transport channel BCH with central information that has to be decoded by all terminals in a cell, such as available random access slots and access slots. See also Section 12.3.3.	• No power control • SF 256 • QPSK • 1/2 rate conv coding • Time-multiplexed with SCH, 2304 chips in each slot
	SCCPCH	Secondary Common Control Physical Channel	Carries the downlink transport channels FACH and PCH (containing paging information, see also Section 12.3.4)	• Different SFs and code rates
	DL DPCCH	Downlink Dedicated Physical Control Channel	Carries control information for downlink DCH transport channel, that is, pilots, power control commands for UL and transport format information. See Section 12.2.5.	• Fast closed-loop power control • Time-multiplexed with downlink DPDCH • QPSK • SF 256
	DL DPDCH	Downlink Dedicated Physical Data Channel	Carries data for downlink DCH transport channel. See also Section 12.2.5.	• Fast closed-loop power control • Time-multiplexed with downlink DPCCH • QPSK • SF 4-512
	UL DPCCH	Uplink Dedicated Physical Control Channel	Carries control information for uplink DCH channel, that is, pilots, power control commands for DL, transport format information and feedback (e.g. for downlink transmit diversity). See also Section 12.2.5.	• Fast closed-loop power control • I/Q code multiplexed with uplink DPDCH • BPSK • SF 256
	UL DPDCH	Uplink Dedicated Physical Data Channel	Carries data for uplink DCH transport channel. See also Section 12.2.5.	• Fast closed-loop power control • I/Q code multiplexed with uplink DPCCH • BPSK • SF 256
	CPICH	Common Pilot Channel	Transmitted in downlink to enable channel estimation at terminal side. Note that received CPICH power at terminal side is used for handover, cell (re-) selection. See also Section 12.2.7.	• No power control • SF 256

Table 12.2 *(Continued)*

Rel.	Abbr.	Name	Purpose	Details
	SCH	Synchronization Channel	Primary and secondary SCH used by terminal for coarse and fine synchronization, respectively. After identifying the secondary SCH, terminal has obtained frame and slot synchronization. See also Section 12.3.2.	• No power control • SF 256
	AICH	Acquisition Indicator Channel	Used in the downlink to indicate the successful reception of an uplink transmission on the Physical Random Access Channel (PRACH). Carries ACK for up to 16 signatures. See also Section 12.3.3.	• No power control • SF 256, length 4096 chips
	PICH	Paging Indicator Channel	Carries paging indicators in the downlink which may tell terminals that they are to receive paging information via the Paging Channel (PCH). See also Section 12.3.4.	• No power control • SF 256
	PRACH	Physical Random Access Channel	Used in the uplink for terminals to register to a cell, perform a location update or initiate connection setup. Terminal initially sends PRACH preambles at increasing power, and when acknowledged by the system with an AICH message, a PRACH message may follow. See also Section 12.3.3.	• Open-loop power control • PRACH preamble: SF 256, length 4096 chips • PRACH message: SF 32-256, length 10 or 20 ms
Rel. 5	HS-PDSCH	High Speed Physical Downlink Shared Channel	Carries data for HS-DSCH transport channel. See also Section 12.2.5.	• No power control • Code-multiplexing between terminals • SF 16 • QPSK or 16-QAM
	HS-SCCH	High Speed Shared Control Channel	Carries control information for HS-DSCH transport channel, that is, information on which codes are to be despread and decoded by which terminal, which modulation and coding scheme has been used, and whether a transmission is an initial one or a retransmission for a particular pervious transmission.	• No power control • SF 128 • QPSK
	HS-DPCCH	High Speed Dedicated Physical Control Channel	Carries uplink control information needed for HS-DSCH transport channel, that is, ACK/NACK information for previous downlink transmissions and CQI.	• Fast closed-loop power control • SF 256 • BPSK

(*continued*)

Table 12.2 *(Continued)*

Rel.	Abbr.	Name	Purpose	Details
Rel. 6	E-AGCH	Enhanced Absolute Grant Channel	Downlink control channel supporting E-DCH transport channel and carrying absolute grant information, that is, telling terminals which code constellation and corresponding data rates to use for uplink transmission on E-DPDCH. See also Section 12.2.5.	• No power control • SF 128
	E-DPCCH	Enhanced Dedicated Physical Control Channel	Uplink control channel supporting E-DCH transport channel and carrying information about the used code constellations in E-DPDCH, information on whether an E-DPDCH transmission is new or a retransmission, and a "happy bit" indicating whether the terminal requests a higher data rate on E-DPDCH or not. See also Section 12.2.5.	• Fast closed-loop power control • SF 256 • BPSK
	E-DPDCH	Enhanced Dedicated Physical Data Channel	Uplink data channel for E-DCH transport channel. See also Section 12.2.5.	• Fast closed-loop power control • SF 2 to 256 • BPSK
	E-HICH	Enhanced HARQ Indicator Channel	Downlink control channel supporting E-DCH transport channel and carrying information on whether previous uplink transmissions on E-DPDCH were successful. See also Section 12.2.5.	• No power control • SF 128 • Shares the same code channel as E-RGCH
	E-RGCH	Enhanced Relative Grant Channel	Downlink control channel supporting E-DCH transport channel and carrying relative grant information, that is, telling terminals whether to increase or decrease code usage and corresponding data rates for uplink transmissions on E-DPDCH. See also Section 12.2.5.	• No power control • SF 128 • Shares the same code channel as E-HICH

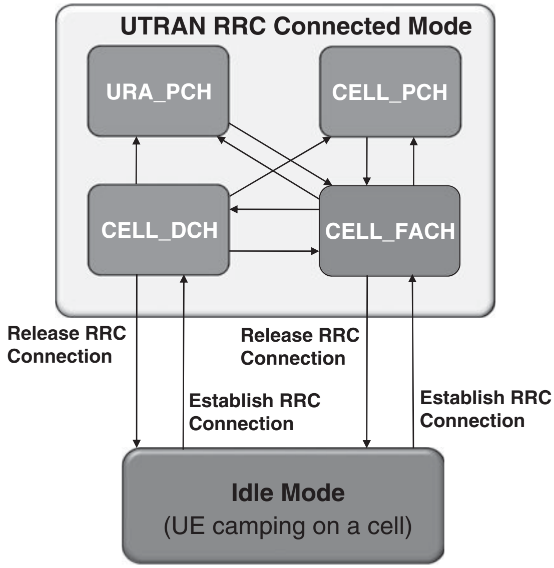

Figure 12.11 RRC states and transition between them.

- Cell DCH. This is the most resource consuming state both on the radio interface and regarding the power consumption level. The benefit of this state is that dedicated channels are used to perform data exchange, so that experienced data rates of the transmission are also the highest and can be guaranteed to a certain extent. Due to the large resource consumption, after some inactivity time (which can be configured by a network operator, usually several seconds) a UE is moved to the less resource consuming Cell PCH state.

It is worth underlining that since support of PCH states is mandatory for the UE and optional for the network, many network vendors do not implement it in their equipment.

12.3.1 Cell Search Procedure

A cell search procedure is the first step in establishing a signaling connection between a UE and the network. It is a necessary step after a mobile has been powered on, and it is entirely performed by the UE. This first cell selection is not intended for finding the best cell to provide wireless service, but rather to find a first suitable

Table 12.3 Comparison of RRC states

	URA PCH	Cell PCH	Cell FACH	Cell DCH
Receiver state	On (periodically)	On (periodically)	On	On
Transmitter state	Off	Off	On	On
Available data rates	Low (reception only)	Low (reception only)	Medium	High
Location knowledge	URA	Cell	Cell	Cell
Power consumption	Low	Low	Medium	High

cell that can be used to obtain network information. Further in the process, if the UE finds a cell which has better radio conditions then the first cell that was found, the UE may move to that cell using a cell reselection procedure. If a UE has successfully completed the cell selection or reselection process and monitors system information broadcasted in a given cell, according to 3GPP terminology such UE is *camping on a cell* [8].

The cell selection procedure can be divided into four parts:

- Public Land Mobile Network (PLMN) selection
- Synchronization
- Acquisition of primary scrambling code
- Obtaining cell and system information from a broadcast channel and taking the decision: camping on a cell or repeating the whole procedure.

In the PLMN selection phase, a UE is responsible for the choice of the optimal network from the point of view of mobile operator and the UE itself. The initial cell selection procedure prioritizes cells from Home Public Land Mobile Networks (HPLMNs) which in general means that after power on, UE will try to connect to the network of its own (i.e. home) operator. An easy way to obtain this is to scan frequencies belonging to the particular operator. In a most common approach, available frequency spectrum for 3G networks is divided between the operators within each country so that one particular frequency is exclusively operated by a single operator only.

If an HPLMN network is not found, for example due to a lack of coverage, the UE searches for other PLMN networks of "friendly operators," that is the ones which have a potential business cooperation agreement with home UE operator. If such PLMNs are not found either, the UE will try to tune to any other PLMN with the signal quality being a selection criterion. Of course both network and user equipment need to support the same technology which means that a terminal that only supports WCDMA and HSPA technology will not be able to connect to a GSM network. Additionally, network operators have an option to enable manual PLMN selection in UE devices.

After successful PLMN selection, the UE performs a cell selection procedure. Two approaches are possible here:

- Initial Cell Selection: UE scans all available carrier frequencies looking for a suitable cell.
- Stored Information Cell Selection: UE may utilize available information about the suitable cells. For example, it may try to camp on a cell that was used for network operations before the device was powered off.

A UE may only camp on a cell that fulfills the cell selection criterion which takes into account both cell signal quality and signal strength in FDD mode. In TDD mode, only a signal strength criterion is used. The signal quality of a cell can be derived from the cell's averaged CPICH signal quality, and the signal strength may be evaluated based on the cell's CPICH Received Signal Code Power (RSCP) in FDD or P-CCPCH RSCP in TDD mode.

After a successful cell selection, the UE reads the system information from the BCH of the cell it camps on. Then it starts performing periodical measurements of signals from other cells that may eventually trigger a cell reselection procedure. This process is necessary in order to find a cell with the optimal radio conditions in a changing environment. Cell reselections may happen in Idle, URA_PCH or CELL_PCH states.

12.3.2 Synchronization

Synchronization is one of the key procedures in the physical layer of cellular networks. Contrary to some systems, such as CDMA 2000, where the transmissions of base stations are synchronized and each one uses identical, but timeshifted scrambling codes, a WCDMA system is based on an asynchronous operation which implies that transmission from different base stations is not time synchronized and every base station marks its own transmission using individual scrambling codes. This approach renders a synchronization procedure necessary. As mentioned before, synchronization is necessary when trying to perform a cell selection/reselection and is also a necessary step in the handover operations described further in Section 12.3.7.

To perform a synchronization procedure to an individual cell, the UE needs to

- Perform slot synchronization.
- Perform frame synchronization
- Find out the cell's scrambling code

Slot synchronization is obtained using the primary Synchronization Channel (primary SCH). This physical layer channel is common for all cells, which means that its sequence is not scrambled by individual cells' scrambling codes. The primary scrambling channel sequence is sent in the first 256 chips of every slot which allows the UE to mark the beginning of each slot using a matched filter. The beginning of a new slot can be determinated by searching for the highest peak in the output of the matched-filter correlating the received signal to the known primary scrambling sequence. Other peaks at the output of the matched filter might be related to other NodeBs' synchronization signals or echoes of the multipath propagation of the strongest NodeBs. To eliminate those ambiguities, the measurements for slot synchronization need to be acquired over some period of time.

After obtaining slot synchronization, the UE needs to determine the beginning of each frame with the help of the secondary SCH. Secondary SCH is sent in the same time as the primary SCH, that is, in the first 256 chips at the beginning of every slot. Contrary to the primary SCH where the same sequence is sent in every slot, one may here send 16 different chip sequences in every slot. These different sequences may be repeated in a slot within a frame, but their different positions form 64 unique sequences which are also known to the UE. Those sequences are chosen in a way that their cyclic shifts are unique, hence it is possible to use these to determine the beginning of the frame. Table 12.4 shows the chip allocation.

The 64 unique sequences have twofold purpose: beside facilitating frame synchronization, they also enable the receiver to distinguish between 64 scrambling code groups. In each code group there can be 8 unique primary scrambling codes, which gives the total number of 512 primary scrambling codes available for the downlink transmission. To determine which one is used, the received CPICH signal consists of a long "all zeros" sequence scrambled with the primary scrambling code. Please note that the whole synchronization procedure is additionally optimized to speed up and simplify the primary scrambling code determination operation that would otherwise be time consuming and complex. With this approach, instead of comparing 512 long sequences with the cell's primary scrambling code, the decision is narrowed down to a selection from 8 available choices.

12.3.3 Cell Update

As stated in the beginning of the section, the UE position is known to the network with the precision of a cell or URA, so if a UE reselects to a new cell it needs to inform the network about it. The message flow, as illustrated in Figure 12.12, used to perform this operation is known as a cell update procedure. The cell update procedure may also be initiated when starting an uplink data transmission, upon radio link and RLC failures, or in the context of a paging response or periodical cell update.

Similarly to a cell update, an URA update may be triggered if the UE reselects to a cell where a new URA identifier is broadcasted. Analog to a cell update, an URA update procedure will also take place in case of a periodical URA update.

Since an exchange of signaling messages is involved, a UE needs to move to a cell FACH state in order to conduct a cell/URA update procedure.

Table 12.4 *Allocation of different chip sequences to S-SCH [9]*

Scrambling Code Group	Slot number														
	#0	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14
Group 0	1	1	2	8	9	10	15	8	10	16	2	7	15	7	16
Group 1	1	1	5	16	7	3	14	16	3	10	5	12	14	12	10
Group 2	1	2	1	15	5	5	12	16	6	11	2	16	11	15	12
Group 3	1	2	3	1	8	6	5	2	5	8	4	4	6	3	7
Group 4	1	2	16	6	6	11	15	5	12	1	15	12	16	11	2
Group 5	1	3	4	7	4	1	5	5	3	6	2	8	7	6	8
Group 6	1	4	11	3	4	10	9	2	11	2	10	12	12	9	3
Group 7	1	5	6	6	14	9	10	2	13	9	2	5	14	1	13
Group 8	1	6	10	10	4	11	7	13	16	11	13	6	4	1	16
Group 9	1	6	13	2	14	2	6	5	5	13	10	9	1	14	10
Group 10	1	7	8	5	7	2	4	3	8	3	2	6	6	4	5
…															
Group 37	2	11	15	3	11	6	14	10	15	10	6	7	7	14	3
…															
Group 58	5	10	10	12	8	11	9	7	8	9	5	12	6	7	6
Group 59	5	10	12	6	5	12	8	9	7	6	7	8	11	11	9
Group 60	5	13	15	15	14	8	6	7	16	8	7	13	14	5	16
Group 61	9	10	13	10	11	15	15	9	16	12	14	13	16	14	11
Group 62	9	11	12	15	12	9	13	13	11	14	10	16	15	14	16
Group 63	9	12	10	15	13	14	9	14	15	11	11	13	12	16	10

Source: Data by courtesy of 3GPP.

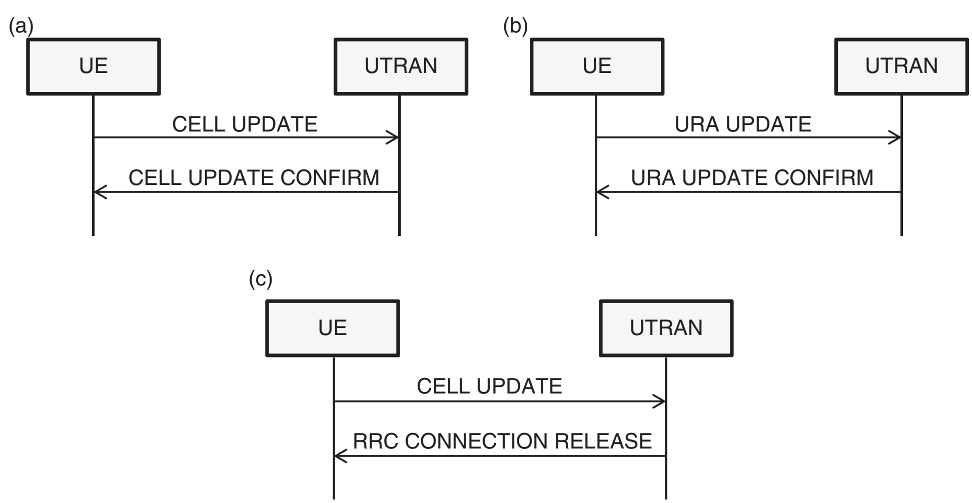

Figure 12.12 *Message flow in case of a (a) successful cell update procedure, (b) successful URA update procedure, (c) failed cell update procedure.*

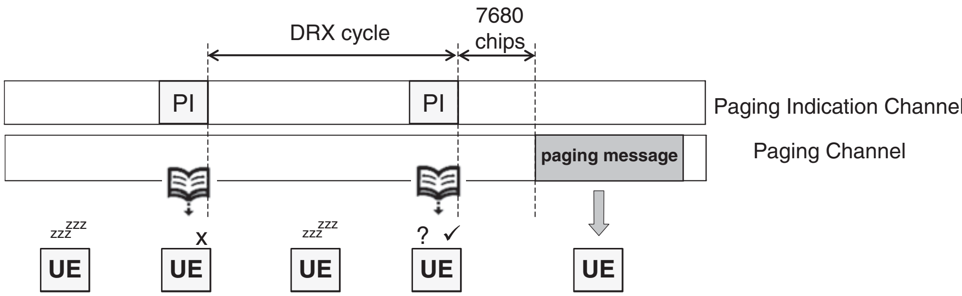

Figure 12.13 *Paging process using Paging Indications.*

12.3.4 Paging

Paging, as illustrated in Figure 12.13, is the only way of sending dedicated data from the network to a particular UE in an IDLE, cell PCH or URA PCH state. It is a unidirectional way of communication which means in this particular case that data can be send only from the network to the UE. There are three situations when a UE is paged:

- In the context of a mobile terminating call establishment procedure, i.e. if the mobile is receiving a phone call
- If the network wants the UE to read the broadcast channel, for instance if important system information has changed
- If the network wants to invoke the UE to perform a cell update procedure, that is, to refresh the information about the UE position.

In order to provide lowest battery consumption in IDLE, cell PCH or URA PCH state, a UE doesn't use its transmitters, and its receivers are switched on and off periodically in so-called discontinuous reception (DRX) cycles.

There are two channels associated with the paging procedure: PICH and PCH. In IDLE, cell PCH or URA PCH states, the UE periodically switches on its receiver in order to read Paging Indication (PI) information on paging occasions associated with this UE on the PICH channel. If the UE finds such information, it will expect paging information on PCH with an offset of 1 transmit time interval after the end of the paging indicator. Later, the UE behavior is dictated by the content of the paging information.

12.3.5 Call Setup

In a call setup, either speech or data connection is established between the UE and the core network. If a UE is in IDLE mode, it will initiate the whole call setup procedure by establishing an RRC connection. To do this, the UE sends an RRC message CONNECTION REQUEST that provides to the radio network information about the reason of establishing this connection (in this case this would be: mobile originated call) and the UE capabilities. The RNC has the possibility to reject such request for instance due to admission control functionalities detecting overloading in that particular cell, but if the request is accepted, the RNC will order the setup of a radio link in the NodeB which operates the cell via which the UE is sending its request. If successful, the RNC also reserves transport resources on the involved Iub interfaces, allocates necessary Radio Network Temporary Identifiers (RNTIs) for the UE and finally responds with a RRC CONNECTION SETUP message that conveys radio link configuration information. Once the establishment of a radio link is acknowledged by the NodeB via NBAP RADIO LINK RESTORE, the UE sends a final RRC connection confirmation in the form of RRC CONNECTION SETUP COMPLETE.

As a next step, signaling exchange between UE and core network is carried out. RANAP and RRC protocol messages are exchanged to perform an authentication procedure and to start optional ciphering and mandatory integrity protection algorithms. If the security procedures are completed, the core network initiates the configuration of Radio Access Bearers (RABs), which are the radio network resources as seen from the core network point of view (for instance a conversational speech data pipe between UE and core network). The procedure involves sending a RANAP RAB ASSIGNEMENT REQUEST that conveys the expected QoS of the call and necessary network addresses, NBAP RADIO LINK RECONFIGURATION PREPARE and RADIO LINK RECONFIGURATION READY that reserves radio resources to carry the requested data. Finally, RRC RADIO BEARER SETUP RESPONSE completes the establishment of a new dedicated transport channel (DCH).

At this moment, all necessary resources are allocated in the radio network, and the RNC sends a RANAP RAB ASSIGNMENT RESPONSE to the MSC/VLR, and this reserves necessary circuits using ISUP INITIAL ADRESS MESSAGE and ISUP ADRESS MESSAGE COMPLETE. The alert message is sent to the UE using RANAP and RRC DIRECT TRANSFER, which result in the end-user hearing the ringing tone in the handset. If the person that we want to call picks up the phone, this is acknowledged by the sequence of ISUP ANSWER MESSAGE, RANAP DIRECT TRANSFER and RRC DL DIRECT TRANSFER. The UE acknowledges this information using RRC UL DIRECT TRANSFER, which is followed by a radio network to core network confirmation in the form of RANAP DIRECT TRANSFER. The conversation may now begin.

This procedure is very similar when the user initiating the conversation would like to call a mobile UE. Then the call setup procedure looks exactly the same, with the exception that before the RRC connection request a paging procedure is performed, as described in Section 12.3.4.

12.3.6 Scheduling

Scheduling is one of the most important aspects in a cellular network in order to provide satisfactory end-user experience. Radio transmission is usually a bottleneck in the whole transmission path, and an optimal assignment of radio resources is therefore necessary, especially in heavily loaded networks. The functional entity that handles those operations is called a scheduler, and its concept has evolved with the evolution of the WCDMA and HSPA standards, but still its main task is to distribute radio resources to users in an operated cell. This is done based on certain priority metrics for each scheduling interval, in order to provide the right balance between maximizing system performance or maximizing the performance of individual users.

Since Release 99, a number of traffic classes are supported in UMTS. However, one of the major divisions in the traffic type is between the real-time and non real-time traffic. Real time traffic needs a constant bit rate and has a certain, usually strict delay requirement. Due to this property, it cannot be controlled dynamically from the RNC (where originally in Release 99 the scheduler entity was located), simply because the control decisions from RNC to NodeB create too much delay. To convey Release 99 real time user data, bidirectional DCH transport channels are used. Those channels can have a different downlink/uplink speed, but they need to be maintained during the whole data transfer. They are power controlled and in case of high load or congestion situations, the RNC, instead of varying the bit rate of already established connections, may simply not allow to setup new connections by means of admission control. However, in some cases, when a new potential connection has a higher priority than ongoing calls (e.g. an emergency call), the RNC may simply drop ongoing connections based on their priorities.

In Release 99, packet scheduling is only applicable to non-real time traffic with less strict delay requirements. To convey such traffic, both DCH and FACH/RACH transport channels can be used. The configuration of a new connection in terms of how many resources (power and spreading codes) can be allocated is based

Table 12.5 Basic differences for scheduling in Rel99 WCDMA and HSPA

	WCDMA	HSPA
Traffic type	Circuit Switched, Packet Switched	Packet Switched
Scheduler location	RNC	NodeB
Spreading Factor	Varying from SF 2 to SF 256 (512)	Fixed SF 16 for HSDPA and varying from SF 2 to SF 256 for HSUPA
TTI duration	10 ms	2 ms or 10 ms (10 ms only for HSUPA)
Adaptation to channel	Power Control	Adaptive modulation and coding, and power control (HSUPA)
Retransmission	Slow (RLC level)	Fast (PHY level)

on the information provided periodically from the NodeB: total transmitted power, load estimation and the threshold load set by the operator. The theoretical limit for peak data rates in the physical layer in Release 99 is 2 Mbps, but using DCHs in this WCDMA release has additional drawbacks. To avoid a frequent and lengthy setup of such channels, these are maintained even up to several seconds after the end of the data transmission. In case of bursty traffic, this causes very inefficient resource utilization in a WCDMA Release 99 cell. Table 12.5 compares the scheduling of Release 99 UMTS and HSPA.

In Release 5, the concept of HSDPA was introduced in UMTS. It brought a tremendous change in the concept of the downlink scheduler. First of all, the downlink scheduler entity was moved from the RNC to the NodeB. This allowed for a faster reaction to the channel changes experienced in a multipath environment. In Release 99, the delay experienced by the information about the current channel state, conveyed from the UE to the RNC, prevented a fast adaptation of the transmission format to the current channel conditions. The data pipe between the UE and radio network had a constant speed, and any changes of provided data rates involved slow and signaling demanding RAB reassignment procedures. Additionally, in Release 99 the transmission time interval (TTI) resolution was set to constant 10 ms. In HSDPA, the TTI value was shortened to 2 ms (equivalent of 3 slots) which allowed more accurate selection of transmission format. Finally, the concept of adjustment to the channel condition changed. Previously, the fast closed-loop power control was used to adapt the link to changing propagation conditions. If the channel state was good, the NodeB sent the data using the same format but simply with less power. If the propagation condition deteriorated, the NodeB would simply increase the transmit power. Figure 12.14 presents examples of the related adaptation functions.

In HSDPA, the NodeB uses a different approach: instead of adjusting the output power level to the channel condition, the NodeB keeps the transmission power constant at maximum level and adapts using different modulation and coding rates of the data. This approach is more efficient, especially for the services which do not require constant bit rate, and maintaining a fixed output power also results in more stable interference characteristics which translate into more predictable and precise scheduling decisions. Another advantage is that with the new solution cheaper NodeB power amplifiers could be used. Since analog amplifiers are one of the most expensive and power consuming parts of the NodeB, this advantage is not to be underestimated.

HSDPA also introduced a significant difference in the sense of code tree utilization. Comparing to Release 99, where the data service speed was determined by the combination of up to two spreading codes with possibly variable spreading factors, the SF of the code tree used for data transmission in HSDPA is constant and equal to 16. 15 out of 16 codes may be used to carry users' data; the remaining part of the code tree is utilized for control channels like CPICH, BCH and so on.

In HSDPA, data transmission is carried over HS-PDSCH. In one TTI, up to 15 HS-PDSCH codes can be used to carry information to one user, which pushed the theoretical peak data rate on the physical layer in Release 5 to 14.4 Mbps. Two channels aid the data transmission. The first one is HS-SCCH which is

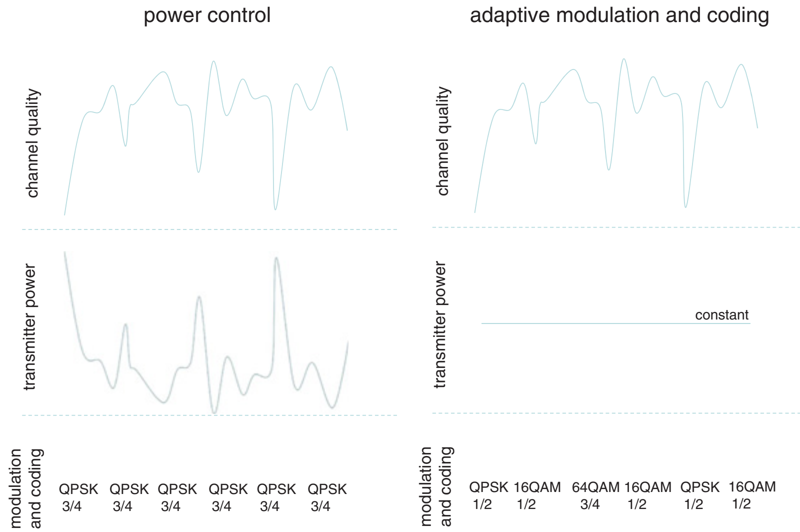

Figure 12.14 *Simplified example of two channel adaptation methods: power control and adaptive modulation and coding.*

transmitted before the HS-PDSCH and it is used to inform the UE about the details of incoming data transmission, e.g. the combination of HS-PDSCHs or transport format necessary for correct decoding of data transmission. There might be several HS-SCCH broadcasted in the cell, but the UE needs to listen to only four of them. If more than one HS-SCCH is used in a cell so-called Code Division Multiplexing (CDM) is possible which means that more than one UE in one TTI can be served with packet data transmission. Since not all UE devices are capable of handling fast data transmission using all 15 codes, this may be a necessary step to fully utilize HSDPA cell capacity. A second channel that complements the HSDPA transmission is an UL HS-DPCCH. This is a dedicated channel that carries Channel Quality Information (CQI) and ACK/NACKs from UE to the NodeB. CQIs calculated by the UE are in the form of indicators to the transport format tables that inform the serving NodeB about the kind of data transmission format that may be handled by the UE at a given error rate probability. The UE calculates it based on the regular CPICH measurements and own receiver capability, for example, based on the number of reception antennas, receiver algorithms or soft bits buffer capability. CQIs are sent to the NodeB as a suggestion, so the final decision on what kind of transport format to use is always up to the NodeB scheduler.

In the uplink direction, packet transmission was optimized by the introduction of HSUPA in Release 6, as described in detail in Section 12.2.5. Also here the scheduler was moved closer to the user and from now on located in the NodeB. However, nonsynchronous uplink transmission requires a different scheduling concept than in the downlink. In the downlink, the resources that could be distributed among the served UEs are time, number of HS-PDSCH codes and the power allocated to each transmission (with a fixed total output power). In the uplink, however, the users are transmitting simultaneously in a nonsynchronized manner under different scrambling codes. This creates the situation where the transmission from one UE is seen as interference during decoding of the transmission from another UE. Hence UL fast transmission still needs to be power controlled, and the resource that can be shared among the UEs in the HSUPA cell is the amount of interference that each transmission introduces, so the other transmissions could be decoded with a given error probability rate. Each UE has its own code tree, so that the code limitation aspect known from HSDPA is not limiting HSUPA transmission.

The amount of interference experienced at the NodeBs' receiver is known as Rise over Thermal (RoT) and is controlled by the NodeB by using new channels introduced in HSUPA: E-AGCH and E-RGCH. E-AGCH is a shared channel and is used for setting maximum power level at the UE side. Absolute grants are not changed often in contradiction to relative grants which can be seen as a fast power control for E-DCH channels. There is only one E-RGCH per each UE with HSUPA transmission in a cell, and it is used to send information like increasing or decreasing the power of E-DCH transmission in the uplink direction. If the UE is in a soft handover state, it listens to the E-AGCH from only one cell. In case that any of the cells in the active set transmits a power down command, the UE needs to decrease its power. To complete the scheduling concept in Release 6 HSUPA, the UE supports the NodeB scheduler by sending a grant request messages informing the NodeB about its power margin and data buffer state.

Another aspect of scheduling is the actual algorithm of resource distribution. By optimizing the scheduler, the cell performance may be adjusted to the network operator demands. Three most popular scheduling algorithms are Round Robin, maximum COverI and Proportional Fair.

The first one, **Round Robin**, is the simplest. It doesn't take CQI into account when considering which user to schedule at a given moment, but of course CQI information still facilitates the choice of a transmission format. Round Robin simply gives the scheduling opportunity sequentially to all potentially served users, which means that if only one UE can be served per TTI and there are 4 UEs in the cell, each of them will be scheduled exactly every 4 TTI. This mechanism is simple, yet its performance in real life scenarios is rather poor.

The second algorithm, **maximum COverI**, takes only CQI information into account while choosing the users to be scheduled in next TTI, meaning that the user that can support highest data rate is chosen in each time interval. This guarantees highest cell throughput, but the UEs which do not have good channel conditions or reception capability will simply be skipped in the scheduling process.

The third **Proportional Fair** algorithm trades fairness against cell performance. The algorithm utilizes both CQI information and each UEs recent throughput history. UEs are scheduled based on priority metrics $M(n)$ that can be calculated as

$$M(n) = \frac{D(n)}{T(n)},$$

where n is the current time step, $D(n)$ is an instantaneous supported data rate which UEs feedback to NodeBs using CQI reports, and $T(n)$ is an average throughput history weighted by a so-called Forgetting Factor (FF), α, that can range from 0 to 1. The average throughput history is calculated as

$$T(n) = (1 - \alpha) \cdot T(n-1) + \alpha \cdot D(n).$$

To illustrate the mechanism of a Proportional Fair scheduler, one can consider the case of two simultaneously active users having an ongoing downlink transmission in an HSDPA cell. If both of them have the same throughput history, the one with better channel conditions will be scheduled in a given TTI. On the other hand, if those users experienced a different throughput history and they report different CQIs, the one which in a given moment has worse reception capability may also be scheduled if the other UE was scheduled more often recently. This fairness of the Proportional Fair algorithm can be adjusted by the FF value. If the forgetting factor in this form is close to 0, then the algorithm gears towards Round Robin behavior. If it is close to 1, then the throughput history is neglected and the scheduler starts to behave close to the maximum COverI algorithm.

One further type of scheduler is the so-called Quality of Service scheduler. Scheduling decisions of those entities are taking into account priorities of the data transmission like delay or minimum data rate. This functionality is placed on top of the algorithms described above. Those schedulers aim at improving individual user experience, not at maximizing the overall cell performance in terms of throughput.

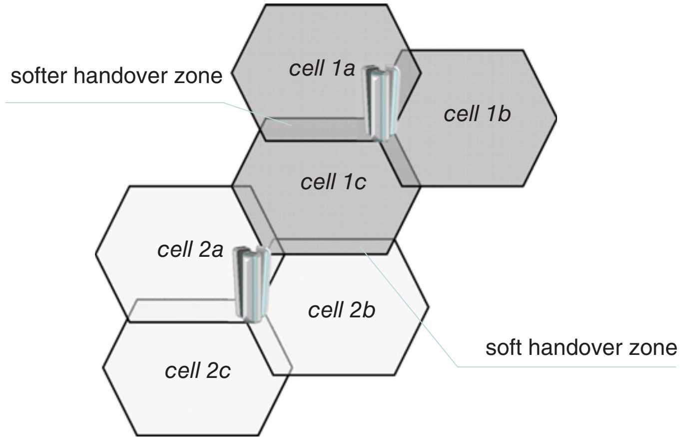

Figure 12.15 *Example of zones with possible soft and softer handovers.*

12.3.7 Handover and Soft Handover

Handover mechanisms are the key tool that provides an undisrupted connectivity for mobile devices in RRC Cell DCH state. Handover (or handoff) in its simplest form is the procedure for changing one serving cell to another while having an ongoing data session. In WCDMA/HSPA they can be divided into three main categories, as indicated in Figure 12.15:

- Hard handovers – the original connection is broken before the new one is set up (so-called "brake before make")
- Soft handovers – original connection is maintained after the new one is set up in a different cell, hence the UE may have two or more ongoing data transmissions from and to different cells at the same time
- Softer handovers – same as the soft handovers, but the original and the new connection is realized via cells belonging to the same NodeB.

Every handover can be realized as a hard handover, but soft and softer handovers have additional advantages. First of all, they can lead to the reduction of interference. As an example one can consider an uplink transmission and UE being at the border of two cells in a softer handover zone. The end result of such diversified packet reception via two cells would be a lower end packet error rate, if the UE transmits at constant total output power, or lower total output power when keeping the same packet error rate. Such an improvement of the radio link performance is caused by a macro diversity principle – one of two radio links is always better than the other one.

By keeping lower output power, the UE creates less interference in the NodeB receiver by lowering overall RoT. Secondly, soft and softer handovers alleviate the near-far-effect in the uplink. The UE is power-controlled from two or more cells, which minimizes the possibility of "overshouting" the transmission of other UEs in either cell. Such a scenario could happen, for example, if a UE that was power-controlled from one NodeB, having a high output power, were to be suddenly handed over to a new cell because of sudden changes in channel conditions. Due to those benefits, both soft and softer handovers are present in Release 99 and in HSUPA. On the other hand the drawback is a higher demand for resources: channelization codes, Iub capacity and additional rake fingers in WCDMA receivers. As a result, soft and softer handovers were not included in the HSDPA standard.

As already mentioned, soft and softer handovers are different in the way that the latter one occurs between the sectors of the same NodeB. Additionally, in softer handovers occurring in uplink direction in cells belonging to the same NodeB, signals received from radio links in different cells are combined like different multipath signals using macro diversity soft combining in the NodeB receiver block. There is only one power control loop and no extra transmissions on Iub are necessary. In case of soft handovers, if two or more NodeBs are involved in the process, frames received in an uplink direction are routed to the RNC, where based on CRC check the best one is selected for further processing. Also two or more power control loops are needed to maintain all connections. In a situation when the UE receives a power up command in one or more links and a power down command on any other, the UE will decrease its transmission power. Since the UE may be in handover among more than 2 cells, it is possible for soft and softer handover to occur together. To facilitate soft handovers in NodeBs belonging to different RNCs, an Iur interface between RNCs is used.

In contrary to cell selection and reselection, in handovers it is the radio network that performs the procedure based on the measurements and indications provided by the UE. The UE should be connected to the best available cell, this is why it constantly monitors the power level of CPICH from neighboring cells. Results of those measurements are sent in so-called measurement events to the RNC which is a handover decision unit. Surrounding cells where the CPICH signal is measured are categorized into:

- Active set – the list of cells where the UE has connection to UTRAN (if more than one entry exists, the UE is in a soft or softer handover state).
- Monitored set – The list of cells that the UE measures, but where their pilot signal is received in too weak level to add them to the active set.

The measurements provided to the RNC are in the form of measurement events. Examples of the three most popular ones are: radio link addition (Event 1A), radio link removal (Event 1B) and radio link replacement (Event 1C). The criteria that should be fulfilled to trigger each event report are listed below:

Event 1A: *New pilot signal quality > Best pilot signal quality − Reporting Range 1A + Hysteresis 1A*
Event 1B: *Pilot signal quality < Best pilot signal quality − Reporting Range 1B − Hysteresis 1B*
Event 1C: *Best new pilot signal quality < Worst old pilot signal quality − Reporting Range 1C + Hysteresis 1C*

The Pilot's Signal Quality may refer to Ec/No measurements of CPICH; Reporting Range is the threshold for a given event; and Hysteresis is the additional offset over a configurable amount of time which is added to the measurement before the decision is reported to the network. This last factor is used to prevent the ping-pong effect, where the UE could be interchangeably adding and removing cells from the active set resulting in excessive signaling load.

As already mentioned, every handover can be realized as a hard handover. However, in some situations only hard handovers are possible. Those handovers are:

- Inter RAT handovers – occurring between two different radio access technologies like GSM and UMTS
- Inter Mode handovers – occurring between UMTS TDD and FDD mode
- Inter frequency handovers – occurring between different frequencies within operator's bands.

One particular challenge related to inter frequency hard handovers is simply the necessity to perform the measurement on different carrier frequencies. This is achieved by interrupting ongoing transmission as the radio chain in the UE has to be tuned to the new frequency. In order to achieve this without the termination of a user's transmission, a compressed mode of transmission may be applied by the network. This can be realized in three different ways. The network may enforce changing the Spreading Factor to the lower one (e.g. from SF 16 to SF 8) which results in higher transmission speed, and as a consequence in potentially freeing some of the time slots that can be used for intra frequency measurements. Another approach is to increase the puncturing rate of the ongoing transmission aiming also at freeing some of the timeslots. The last approach is simply to deliberately skip some of the transmission slots by using appropriate higher layer scheduling.

12.4 WCDMA/HSPA Evolution since Release 5

In this section, we summarize the different features that have been standardized for WCDMA and HSPA since Release 5. These features are not only aimed at boosting the achievable throughput, but contribute to the overall efficiency, manageability and dimensioning of 3G mobile communications systems. An overview on the most important features standardized from Release 5 to Release 11 and their key benefits is provided in Table 12.6, while Figure 12.16 illustrates the corresponding evolution of data rates over time.

In the sequel, we describe the most important of these evolutionary features which have been ensuring fast growth and market dominance of 3G system in recent years. These features include Multicarrier systems, Multiple-Input Multiple-Output (MIMO), Multicell Transmission or Multiflow, Heterogeneous Networks (HetNets) and Self-Organizing Networks (SON).

12.4.1 Multicarrier

In wireless communications, a frequency band is the single most sought after and expensive resource often heavily regulated by government bodies around the globe. However, while frequency pools are limited, the continuous growth in mobile data traffic impose continuous enhancements of the network performance at the lowest possible cost for the network operators. Incremental evolution of a network technology also puts strict requirements for the support of legacy terminals. In the context of a UMTS system, this essentially means that a network employing Release 10 features should be able to serve Release 99 terminals (and everything in between) as well. Multicarrier technologies overcome these constraints by introducing new terminals which are capable of utilizing multiple frequency carriers without affecting the underlying mechanism necessary to ensure the smooth functioning of mobile devices compliant only with the older releases.

A single carrier mobile terminal is designed to operate on one or more frequency bands over a certain carrier with limited bandwidth. Such terminals are confined to the use of only one of the selected carriers. On the other hand, multicarrier terminals are capable of utilizing more than one such carrier at the same time. These individual carriers may or may not be adjacent to each other.

Release 99 WCDMA defines a carrier bandwidth of 5 MHz. Therefore, all the enhancements brought into the system until Release 7 were confined to an overall system bandwidth of 5 MHz. As the demand for data rates grew, which eventually led to data services dominating 3G in comparison to voice services, it became necessary to push this 5 MHz bandwidth limit further. However, simply increasing the carrier bandwidth beyond 5 MHz was not an option since doing so would severely affect the core specifications for legacy terminals.

Figure 12.17 shows the evolution of multicarrier features over time. In Release 8, dual cell HSDPA (DC-HSDPA) was introduced for the first time. A DC-HSDPA capable terminal can receive information from two data channels on adjacent 5 MHz carriers simultaneously. In Release 9, dual cell capabilities were also specified for the uplink while in Release 10, the number of adjacent carriers was increased to 4. Moreover, Release 9 also witnessed the introduction of multiband HSDPA. In multiband operations, the carriers are not required to be adjacent any more, but they can belong to completely different frequency bands. In Release 10,

Table 12.6 Key UMTS evolutionary features from Release 5 to Release 11

Release	Feature	Key aspects	Impact
Rel. 5	HSDPA	• Downlink scheduler moved from RNC to the base station • Reduction of transmit time interval (TTI) length to 2 ms • Fast power control in downlink abandoned • Adaptive modulation and coding (AMC) in conjunction with hybrid automated repeat request (HARQ) introduced	• Significantly increased downlink spectral efficiency, and better handling of bursty best-effort traffic • Increase of downlink peak data rates to 14.4 Mbps
Rel. 6	HSUPA	• Uplink scheduler moved from RNC to the base station • Reduction of transmit time interval (TTI) length to 2 or 10 ms • Codes with lower spreading factor • Hybrid automated repeat request (HARQ) introduced for uplink transmissions	• Uplink spectral efficiency increased • Uplink peak data rate increased to 5.76 Mbps
Rel. 7	DTX/DRX, aka Continuous Packet Connectivity (CPC)	• Allows terminals to stop transmitting control channels UL DPCCH and UL E-DPCCH when there is no data to be transmitted • Allows terminals to only listen to downlink control channels occasionally and shut off the radio completely otherwise	• Significantly increased terminal battery time and talk time
	Downlink 64-QAM	• Allows 64-QAM to be used in downlink transmission on HS-PDSCH	• Downlink peak data rate increased to 21.6 Mbps
	Downlink 2 × 2 MIMO	• Enables downlink 2 × 2 transmission on HS-PDSCH	• Downlink peak data rate increased to 28.8 Mbps
	Uplink 16-QAM	• Allows 16-QAM (or actually 4-PAM on each I/Q-multiplexed code) to be used in uplink transmissions on E-DPDCH	• Uplink peak data rate increased to 11.5 Mbps
	Enhanced FACH	• Allows terminals in state CELL_FACH to utilize HS-DSCH transport channel in downlink • Inherently allows faster transition from terminal state CELL_FACH to CELL_DCH	• Significantly increased downlink data rates possible in CELL_FACH, and correspondingly less need to move terminals into state CELL_DCH • Faster connection setup

(continued)

Table 12.6 *(Continued)*

Release	Feature	Key aspects	Impact
Rel. 8	Enhanced RACH	• Allows terminals in state CELL_FACH to utilize E-DCH transport channel in uplink in addition to RACH transport channel • Inherently allows faster transition from terminal state CELL_FACH to CELL_DCH	• Significantly increased uplink data rates possible in CELL_FACH, and correspondingly less need to move terminals into state CELL_DCH • Faster connection setup
	Fast dormancy	• Allows terminals to quickly go from state CELL-DCH to CELL-PCH upon request from higher layers	• Significant power saving at the terminal side
	CS Voice over HSPA	• Circuit switched (CS) voice is transported via HS-DCH and E-DCH instead of DCH	• Also voice communication benefits from significantly increased spectral efficiency from HSPA
	Downlink 2 × 2 MIMO with 64-QAM	• Enables the usage of 2 × 2 MIMO and 64-QAM in downlink transmissions on HS-PDSCH	• Downlink peak data rate (with 2 × 2 MIMO and 64-QAM) increased to 42 Mbps
Rel. 9	Dual Carrier HSUPA	• Enables uplink transmission on E-DPDCH simultaneously on two carriers in the same band	• Uplink peak data rate (with 16-QAM and Dual Carrier) increased to 23 Mbps
	Multiband Dual Carrier HSDPA	• Enables downlink transmission on HS-PDSCH simultaneously on two carriers in different bands	• Allows much more operators to actually use dual carrier HSDPA
	Dual Carrier HSDPA with 2 × 2 MIMO and 64-QAM	• Enables the usage of 2 × 2 MIMO and 64-QAM in downlink transmissions on HS-PDSCH and on two adjacent carriers	• Downlink peak data rate (with 2 × 2 MIMO, 64-QAM and Dual Carrier) increased to 84 Mbps
Rel. 10	Quad Carrier HSDPA	• Enables downlink transmission on HS-PDSCH simultaneously on up to four carriers in two different bands	• Downlink peak data rate (with MIMO, 64-QAM and Quad Carrier) increased to 168 Mbps
Rel. 11	Downlink 4 × 4 MIMO	• Enables downlink 4 × 4 MIMO transmission on HS-PDSCH	• Downlink peak data rate (with 4 × 4 MIMO, 64-QAM and Quad Carrier) increased to 336 Mbps
	Uplink 2 × 2 MIMO	• Enables uplink 2 × 2 MIMO transmission on E-DPDCH	• Uplink peak data rate (with 2 × 2 MIMO, 16-QAM and Dual Carrier) increased to 46 Mbps
	8 Carrier HSDPA	• Enables downlink transmission on HS-PDSCH simultaneously on up to eight carriers in two different bands	• Downlink peak data rate (with 2 × 2 MIMO, 64-QAM and 8 carriers) increased to 336 Mbps
	Multiflow	• Allows concurrent transmissions from two base stations or two carriers to one terminal in the same transmit time interval (TTI)	• In a single-carrier deployment and under standard 3GPP simulation assumptions [10], average burst rate increased by about 10%, and that of soft handover users increased by about 40%.

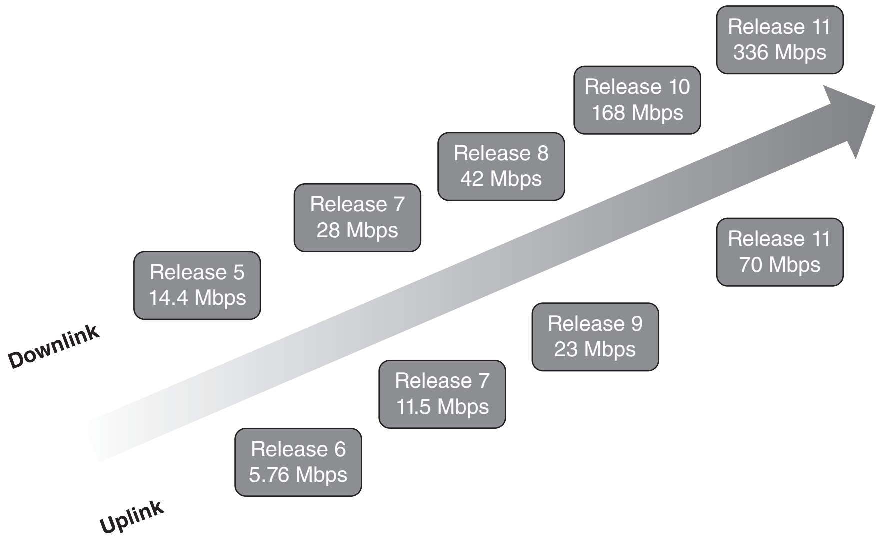

Figure 12.16 *Evolution of HSPA data rates.*

multiband capability was extended to up to 4 carriers on a maximum of two frequency bands in the downlink direction. Multiband HSPA functionality is crucial from an economic perspective since many small operators do not have access to adjacent 5 MHz carriers. However there are many operators who own two 5 MHz bands in different spectrum such as 900 MHz and 2100 MHz. In combination with MIMO (described in the following section), Release 10 has pushed the peak data rates for HSDPA beyond 168 Mbps.

From the performance perspective, multicarrier HSPA offers gains in multiple dimensions. With the availability of two parallel carriers, the achieved throughput can be doubled provided that the full capacity of both frequencies is available to the user. As the system load increases, the possibility of having full access to the frequencies is significantly diminished. However, a combination of frequency scheduling gains, trunking gains and user diversity gains ensures that the achievable multicarrier throughput is still higher than the sum of the throughput of individual carriers. In multicarrier operations, CQI is reported individually for each

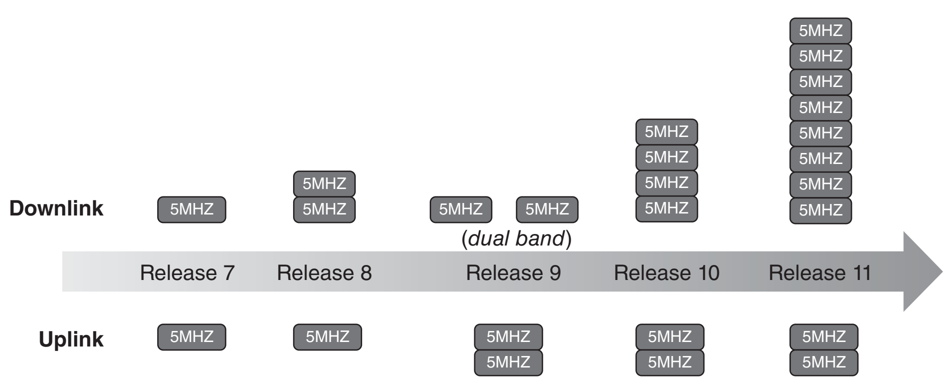

Figure 12.17 *Evolution of multicarrier and multiband feature in HSPA.*

carrier for a single UE. Frequency scheduling gains arise from the fact that the NodeB can always use this CQI information to deprioritize the carrier with deep fades. Since fast fading is uncorrelated among users, it is possible that a different user in a different geographical area is not experiencing such deep fades on some particular carrier. Hence, the overall gains can be maximized by scheduling users on the best reported carrier(s). Moreover, the possibility to balance the load on per TTI basis in multicarrier HSPA gives it an edge over the same number of multiple single carrier HSPA terminals.

12.4.2 MIMO

Multiple-Input Multiple Output (MIMO) is one of the key performance enhancement features in contemporary communication systems. MIMO essentially implies the use of multiple antennas at the transmitter and/or at the receiver side. MIMO systems belong to the wider class of smart antenna technologies which enable faster and more robust communication over the radio link by means of spatial signal processing.

Fading is one of the primary challenges encountered over a radio link. Various diversity techniques have been devised to counter the effects of fading. In principle, diversity implies the use of multiple copies of a wireless signal transmitted to the receiver to combat the effects of fading. These diversity techniques can be broadly categorized into time, frequency and spatial diversity schemes. Time and frequency diversity, although effective, are less efficient than spatial diversity methods. Spatial diversity schemes exploit the separation of signals in space, but in order to achieve maximum benefits out of spatially separated copies, it is essential to have the right amount of physical separation between the transceiver entities.

Along with the application of MIMO for the exploitation of spatial diversity, multiple transmitters can also be used to transmit individual data streams from each antenna to boost the overall data rate, provided that the radio channel experienced by the receiver is good enough to receive multiple streams. This transmission technique is known as spatial multiplexing.

As opposed to SISO (Single-input Single-output), MIMO refers to the general class of multiple antenna systems which can consist of MIMO, SIMO (Single-input Multiple-output) and MISO (Multiple-Input Single-output) systems. In a SIMO system, a single transmission is received by multiple physically separated antennas, allowing to counter the effects of fading through receive diversity. Analogously, MISO system makes use of transmit diversity by letting copies of a single transmission be redundantly transmitted from multiple transmit antennas to a single receive antenna. Although SIMO and MISO have their own pros and cons in terms of implementation complexities, generally for handheld mobile devices it is desirable to have as few antennas as possible due to size limitations. In order to take advantage of spatial multiplexing, however, it is necessary to have multiple antennas at both sides since the number of transmitted streams can at most be as high as the minimum of the number of transmit and receive antennas.

In HSPA, MIMO has been standardized with different antenna configurations. A closed loop feedback system is used to adjust transmission weights at different antennas, allowing to individually phase-rotate the signals transmitted from the different antennas. These transmission weights are necessary to ensure that the signals originating from different transmitantennas overlap constructively in the radio channel. In HSPA, transmission weights are chosen from a pre-defined codebook – the receiver measures the radio channel between each transmit antenna and the receiver antennas, and uses this to chose the entry in the codebook, and correspondingly a set of transmission weights, that provide the best constructive signal overlap. 2×2 or 4×4 spatial multiplexing can be used in order to boost the peak data rates by a factor of two and four, respectively. The transmission weights, also known as PCI (Precoding Control Information) combined with CQI form the overall cycle of the feedback system. CQI information is used to calculate the transport block size and it is also helpful in switching between MIMO (spatial multiplexing) and MIMO fallback mode (spatial diversity).

By using MISO, ordinary mobile terminals with a single antenna standardized with Release 5 specifications and higher can be provided the gains offered by transmit diversity. Such a terminal, however, is not capable of receiving more than one data stream.

Just like MC-HSPA, MIMO also poses transmission challenges to the legacy terminals and non-MIMO channels (for example common channels). Moreover, NodeB is not able to use its second power amplifier for non-MIMO transmissions. Initially dedicated frequency bands were used to tackle such issues, however in order to efficiently utilize the full spectrum, another technique known as Virtual Antenna Mapping (VAM) was standardized.

VAM along with the possibility to limit the precoding weights to two instead of four and the introduction of S-CPICH (Secondary Common Pilot Indicator Channel) enabled the coexistence of MIMO and non-MIMO users on the same carrier. VAM also helps in balancing the power from both power amplifiers at the NodeB.

Together with QC-HSDPA, 64QAM modulation and the use of 2×2 spatial multiplexing, the downlink peak data rate for the downlink HSPA is pushed to 168 Mbps, which is comparable to LTE peak rates with similar antenna configurations.

Further extension of 2×2 MIMO in DL known as 4×4 MIMO took place in Release 11 and allowed for spatial multiplexing of up to four streams simultaneously towards one user. This, along with Quad Carrier functionality pushed the peak data rates up to 336 Mbps.

12.4.3 Multiflow

Multiflow was basically one of the most popular features in HSPA Rel. 11 standardization, as it required fairly limited additions or changes to the standard, but allows us to address a key problem in cellular communication systems, namely the comparatively poor performance of users at the cell-edge.

More precisely, Multiflow allows a terminal in the soft or softer handover regime between two cells to receive downlink transmissions on HS-PDSCH from both cells simultaneously. As both cells of course employ different scrambling codes, this means that the terminal must effectively be equipped two complete receiver chains, as the channel estimation, descrambling, equalization, decoding of control information etc. have to be done individually for both transmissions. A concurrent reception of two transmissions is technically possible, even if in both cells the maximum number of 15 channelization codes on HS-PDSCH is used, if the terminal has two receive antennas and is hence able to spatially separate the two transmissions to a reasonable extent. Ideally, the terminal should here employ a so-called type 3i receiver, meaning that when equalizing one transmission the receiver would explicitly reject the interference from the other transmission. However, Multiflow also works with smaller gains when type 3 receivers are used at the terminal side, which treat the interfering stream as spatially white noise.

A key question related to Multiflow is where the data split takes place, where the principle options would be to have this in the core network, in the RNC on PDCP or RLC layer, or in the NodeB on MAC layer. 3GPP has decided to have the data split located in the RNC on RLC layer, and allows intersite Multiflow across multiple NodeBs, which would not be the case if the split would happen within the NodeBs on MAC layer. See Figure 12.18 where different data split options are illustrated.

Clearly, Multiflow does not increase the system capacity, as for instance under full system load it would be more meaningful from NodeB perspective to schedule transmissions to cell-center terminals than investing resources in two adjacent cells into a cell-edge user that is anyway expected to perform comparatively badly. However, under low system load, the burst data rates of soft or softer handover users can be increased up to about 40% through Multiflow, as such users can benefit from resources in the neighboring cell which would otherwise be left empty. In fact, also the average user population may see a burst rate increase of about 10% under low system load if cell-edge users can be served with Multiflow, as such users can then complete their burst transmissions faster, ultimately leaving more transmit resources also for cell-center users.

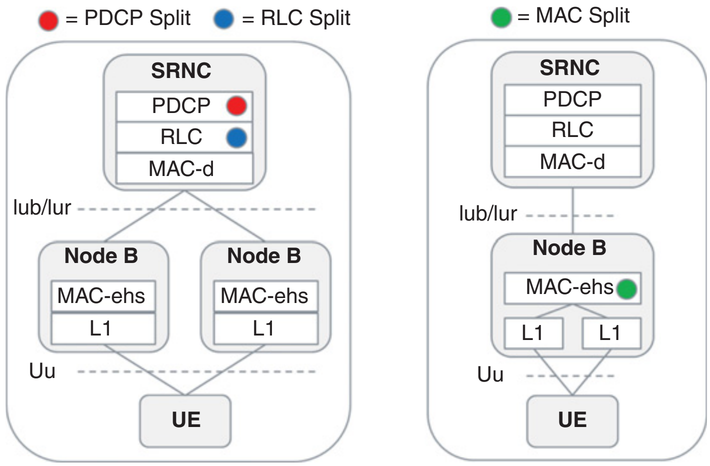

Figure 12.18 *Different data split options considered during the standardization of Multiflow.*

12.4.4 Heterogeneous Networks

The term Heterogeneous Networks (HetNet) can be understood in multiple ways. It can refer to a network that consists of several cellular network technologies, for example, GSM, UMTS and LTE, covering the same geographical area. It can also denote utilization of equipment coming from different vendors. Most commonly, this term is used to describe the coexistence of several nonhomogenous cell types that are different in coverage areas or controlled by different base station types. The difference may come from the total transmit power, antenna configuration, placement, capacity or the application. A very common type of network nodes are macro base stations that are usually deployed to provide coverage. Thanks to the high output power (tens of watts), directional antennas and above the rooftop antenna placements, the macro cells can cover distances of up to several kilometers. This type of network node is typically deployed in the early phases of network deployment in order to provide broad network coverage. The other cell types: micro, pico and femto cells, as presented in Table 12.7, have respectively smaller control area, output power and usually less sophisticated radio components. Initially the main use case of these elements was the provisioning of cellular coverage in the areas where macro stations were not able to provide do so. However, due to increased volume of cellular

Table 12.7 *Main difference between macro, micro, pico, femto nodes*

	Macro	Micro	Pico	Femto
Typical output power	High (>10 W)	Medium (5–10 W)	Medium/Low (1–5 W)	Low (<1 W)
Cell size	Large (>1 km)	Medium (<1 km)	Small (<100 m)	Small (<10 m)
Antenna	Directional, external, usually above rooftop and outdoor	Directional or omnidirectional, external, usually below rooftop and outdoor or indoor	Omnidirectional, integrated, usually indoor	Omnidirectional, integrated, usually indoor
Volume	Large	Medium	Small	Compact

traffic, this type of network elements is more and more often used to provide an extra capacity in hotspot areas, or as a mean to offload macro nodes in, e.g., indoor areas.

A very prominent and promising technology is the last of mentioned cell types, i.e., femto access nodes. Femtos are usually operating with the output power below 1 W, have integrated omnidirectional antennas, low volume and compact form which is suitable for the residential usage. They are connected with the operator's network via the user's private DSL connection (femtos are not directly linked with the core network). The interworking element that facilitates Femto functionality is the so-called Femto Gateway (FGW) and the interface between FGW and femto is standardized as Iuh. Such small cell concepts have been introduced already in GSM, but recently they have gained a lot of attention from the cellular community as one of the easiest forms of macro offloading and a very efficient tool for providing large capacity over a limited area (e.g. a household).

The two possible frequency deployment options assume s deployment on the same carrier frequency as operated by macro (cochannel deployment) or utilization of a separate frequency band (outband deployment). Due to the scarcity of frequency resources the first option is most common. However, it may result in the rise of the interference level (cf. Figure 12.19), which is not the case in the latter approach. The interference limitations may become more visible because of the potentially chaotic deployment of the femtos. They are not foreseen for installation and maintenance by the operator's technical support, but rather by the consumers as plug and play devices. The problem of the interferences may be somehow alleviated by the nature of Femto deployment. For instance, an indoor placement creates a natural isolation of the signal thanks to high signal attenuation of the building walls. Additionally, small cells are bounded by a low output power, hence the impact to the macro (users) is rather limited.

The smaller output power may create another challenge known as uplink/downlink imbalance. This phenomenon is caused by the disproportion of the transmit power of the macro sites (tens of watts) and femto nodes (several hundreds of miliwatts). This difference leads to the fact that a user terminal which is closer to a small cell than a macro cell may still be assigned to the macro cell due to its higher downlink transmit power, while the uplink transmission of the terminal is received more strongly at the small cell than at the macro cell, due to the closer proximity. In this case, the terminal is served by the macro cell, but creates strong uplink interference to the small cell. If one now enforces the UE to be assigned to the small cell (which is

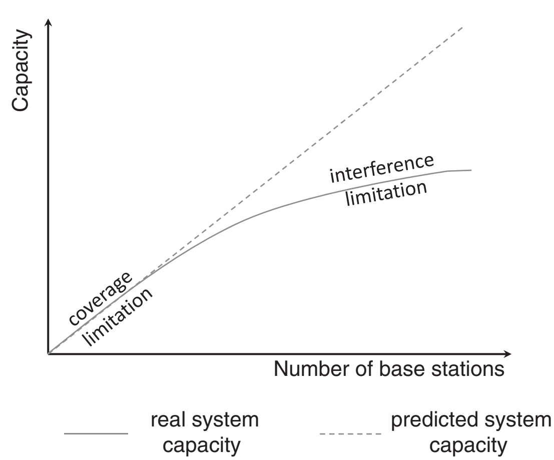

Figure 12.19 *Impact of the interference limitations on system capacity with increased number of deployed network base stations.*

often done in order to offload more terminals to small cells), the uplink interference issue is avoided, but the terminal now receives strong downlink interference from the macro cell. Similar situation can be caused by so-called Closed Subscriber Group functionality. This particular feature allows the owner of the Femto node to the select terminals which can connect to the given small cell. This particular feature allows the owner of the Femto node to select the terminals which can connect to the given small cell. If a particular terminal is not allowed to connect to the Femto node, it may also become a source of heavy uplink interference towards the Femto node, when transmitting to its serving macro node.

12.4.5 Self-Organizing Networks

Self-Organizing Networks (SON) functionality can be divided into three subcategories:

- Self-Configuration – the ability of the network nodes to integrate to the existing cellular environment with existing and operating cells.
- Self-Optimization – the ability of the network nodes to exchange different measurement and configuration information between each other in order to improve the overall network performance, for example, tailoring the handover parameters between adjacent nodes in order to reduce the signaling overhead and limit the number of failed handovers.
- Self-Healing – the ability of network to compensate for the unexpected events or breakdowns, e.g., network nodes failures.

All of the above-mentioned features are aiming to limit and eventually to eliminate the human effort required for manual setting of configuration parameters.

The term SON is rather linked with LTE technology. Despite the fact that SON was already discussed during GSM rollouts, the first standardized implementations took place during the introduction of LTE to the market. For WCDMA and HSPA networks, the term SON usually appears in the context of HetNets and Femto node deployments. This small cell technology is foreseen for the residential usage and may potentially result in the chaotic deployment of additional network nodes. This is in contrary of the typical operator triggered node deployment which is planned, coordinated and based on the network analysis. SON features can help to address some of those problems by automatized adjustment of radio network parameters.

12.5 Planning and Dimensioning of WCDMA/HSPA Networks

The main idea behind planning and dimensioning of WCDMA and HSPA networks is to minimize the number of sites required to provide given coverage and capacity. Initial network rollouts usually focus on coverage, as the main goal of those early phase deployments is to provide a large (usually country wise) availability of the new technology. In the later stage, as the number of user devices supporting a given technology is increasing, the capacity issues are becoming more and more burning and the first layer of macro nodes is complemented with the layer of micros, picos and femtos. On top of capacity and coverage calculations, the most commonly addressed aspects of planning and dimensioning, there are many other aspects that need to be taken into account, for example coverage areas of different Modulation and Coding Schemes or MIMO.

12.5.1 Typical Frequency Usage

3GPP has specified several operating bands which can be used by WCDMA/HSPA. The ones specified up to Release 10 together with corresponding duplex separations are presented in Table 12.8. What is worth to notice

Table 12.8 UMTS operating frequencies for FDD mode [11]

Operating Band	UL Frequencies UE transmit, Node B receive	DL frequencies UE receive, Node B transmit
I	1920–1980 MHz	2110–2170 MHz
II	1850–1910 MHz	1930–1990 MHz
III	1710–1785 MHz	1805–1880 MHz
IV	1710–1755 MHz	2110–2155 MHz
V	824–849 MHz	869–894 MHz
VI	830–840 MHz	875–885 MHz
VII	2500–2570 MHz	2620–2690 MHz
VIII	880–915 MHz	925–960 MHz
IX	1749.9–1784.9 MHz	1844.9–1879.9 MHz
X	1710–1770 MHz	2110–2170 MHz
XI	1427.9–1447.9 MHz	1475.9–1495.9 MHz
XII	699–716 MHz	729–746 MHz
XIII	777–787 MHz	746–756 MHz
XIV	788–798 MHz	758–768 MHz
XV	Reserved	Reserved
XVI	Reserved	Reserved
XVII	Reserved	Reserved
XVIII	Reserved	Reserved
XIX	830–845 MHz	875–890 MHz
XX	832–862 MHz	791–821 MHz
XXI	1447.9–1462.9 MHz	1495.9–1510.9 MHz
XXII	3410–3490 MHz	3510–3590 MHz
XXV	1850–1915 MHz	1930–1995 MHz
XXVI	814–849 MHz	859–894 MHz

Source: Data by courtesy of 3GPP.

is that due to higher propagation attenuation at higher frequencies, frequencies foreseen for transmission in the UL direction from low power user devices are always lower than carrier frequencies in DL direction.

Multiplicity of available operating bands is dictated by the different frequency usage regulations in the countries around the world. It is also up to the state regulation authorities to distribute the available bandwidth between interested parties like mobile network operators, usually by means of spectrum auctions. Spectrum auctions can be really lucrative to the state, for example, during first spectrum auctions in 2000 in UK, Mobile Network Operators (MNOs) spent $35 billion [12]. Also in Germany the total income of the state was 50.8 billion Euros [13]. In the latter case, six companies Viag Intercom (O2), Mobilcom, E-plus, T-Mobile D, Group 3G and Vodafone shared the whole available Band I (2×60 MHz) among each other, so that each of the operators obtained 2×10 MHz paired spectrum. For the initial network deployment, when both the UE traffic and the number of UE devices was low, this amount of spectrum was more than enough to provide satisfactory user experience, but with the forthcoming high data rate services like HD Youtube and with the growing number of 3G device populations, two carriers might be not sufficient to provide high data rates to the end-users. As an example, the quad carrier feature of HSPA Release 10 requires 20 MHz bandwidth in the DL in total, but in return it boosts the peak user data rate on the physical layer up to 84.4 Mbps (168.8 Mbps in quad carrier when combined with 2×2 MIMO mode).

The most popular frequency allocation is the one using Operating Band I, where 2×60 MHz bandwidth are available for WCDMA/HSPA operations in FDD mode. In this operating band, UL frequencies are ranging from 1920 MHz to 1980 MHz, whereas the NodeB transmission is foreseen for the frequencies between 2110 and 2170 MHz. This is the most popular band in Europe and Asia, however in Northern America, the 1900 MHz frequency band is the most common one. Additionally, since 3G networks are superseding 2G networks, various 2G frequencies are also becoming more and more popular among the 3G operators (a trend called 2G frequency reuse). The additional advantage of low frequency bands used previously in GSM (e.g. in the 800 MHz region) is the fact that due to the lower signal attenuation they may be used to provide better network coverage in, for example, rural areas. Also digital dividend frequencies released in digital television transition are of particular interest to operators.

3GPP specified several unpaired bands which are aimed mainly at so-called UMTS TDD Low Chip Rate (LCR). This particular UMTS variant is based on Time Division Synchronous CDMA (TD-SCDMA) which uses TDD method for separation of reception and transmission. Despite the fact that several operators deployed their service using this technology, it never really gained much momentum outside of China where it was declared as a standard 3G technology in 2006.

Another spectrum allocation reserved by 3GPP is the Mobile Satellite Services (MSS) spectrum. In Europe and most of the world it comprises from 1980 MHz to 2010 MHz for Earth to space communications and from 2170 to 2200 MHz for space to Earth direction. MSS aimed at providing a truly global 3G coverage, even when the users run out of the coverage provided by terrestrial base stations. Such services are available commercially, although usually they come at lower data rates and higher pricing comparing to their terrestrial counterparts. Its key advantage is, as already mentioned, global or almost global coverage.

The typical single carrier channel width is 5 MHz, however the spacing between carriers can vary from 4.4 MHz to 5.2 MHz. Additionally, in order to aid the UMTS and GSM coexistence, a channel raster of 200 kHz was chosen, which means that the carrier's center frequency must be a multiple of 200 kHz. Different spacing between carriers gives some freedom for the operators in configuring their network.

12.5.2 Capacity and Coverage Optimization

Capacity and coverage optimization may take a different form depending on whether they are applicable for a planned network or already existing one. It also heavily depends on whether we want to optimize uplink or downlink transmission. Theoretically both transmission directions need to be streamlined, but due to the heavy traffic asymmetry (users tend to download more data than they upload) only one transmission direction may be of particular interest.

Coverage and capacity optimization, even in a network planning phase, is not a trivial task and usually requires usage of advanced software tools. As an input to such calculations a lot of input data is required including the topography of the coverage area, available technology features, possible site locations, propagation conditions, estimated user number and their traffic type, and so on.

Usually the coverage of the cell is UL limited, due to the fact that mobile stations transmit with less power than the stationary NodeBs. The easiest way of estimating cell coverage is calculation of link budget. Its purpose is to estimate maximal pathloss from transmitter to the receiver assuming specific type of traffic, transmitter and receiver parameters in a given radio condition. Example of link budget parameters and the calculation formula for the uplink Rel99 maximum mean pathloss is given below.

$$\begin{aligned} L_{\max\_UL} = {} & P_{UE} + G_{ant,UE} - L_{feeder,UE} + G_{ant,NB} - L_{feeder,NB} - Information\_rate \\ & - Thermal\_noise\_density - NF_{NB} - \frac{E_b}{N_O} - M_{Interference} + G_{SHO} \\ & + G_{ASH} - M_{fastfading} - L_{body} - M_{shadowing} - L_{penetration}. \end{aligned}$$

where:

$L_{max,UL}$	=	maximum mean pathloss in UL direction for Release 99
P_{UE}	=	maximum mobile transmission power = 21 dBm (125 mW)
$G_{ant,UE}$	=	antenna gain of UE = 2 dB
$L_{feeder,UE}$	=	feeder loss at UE side = 0 dB
$G_{ant,NB}$	=	antenna gain at the base station = 14 dBi
$L_{feeder,NB}$	=	feeder loss at NB side = 2 dB
$Information\_rate$	=	processing gain for AMR 12.2 kbps voice service = $10\log\left(\frac{3840}{12.2}\right) = 25dB$
$Thermal\_noise\_density$	=	for the 5 MHz WCDMA carrier and room temperature equal to 174 dBm/Hz
NF_{NB}	=	NodeB's noise figure = 9 dB
$\frac{E_b}{N_O}$	=	required $\frac{E_b}{N_O}$or for AMR 12.2 kbps voice service – 5 dB
$M_{Interference}$	=	interference margin = 3 dB
G_{SHO}	=	soft handover gain = 2.5 dB
G_{ASH}	=	gain against shadowing = 2.5 dB
$M_{fastfading}$	=	fast fading margin = 0 dB
L_{body}	=	body loss = 3 dB
$L_{penetration}$	=	penetration losses, depending on the scenario, for example, 10 dB.

One significant aspect of coverage planning in the UL direction of an UMTS network is that, in contradiction to, e.g., GSM networks, the coverage is also dependent on the load in the cell, that is, it decreases as the traffic load increases. In WCDMA technology, users are able to operate on the same carrier frequency, so the number of simultaneously served users affects the receivers' noise level in the UL. Therefore, in UMTS cellular systems based on a WCDMA radio interface, coverage planning is directly related to capacity planning. Additionally, in contrary to GSM where typical network planning focused on coverage and blocking probability for voice services, UMTS provides a variety of different data services for which also Quality of Service requirements needs to be taken into account during planning phase.

In modern data networks high traffic asymmetry is observed, as users tend to download more than they upload, hence in this chapter the focus will be on HSDPA capacity. This topic is more complex than the calculation of coverage for a given service, since capacity can be considered with respect to many factors contributing to overall system performance like blocking probability, outage probability, code tree utilization, UL/DL load or maximum number of VoIP connections and so on, but in the scope of this book we will consider pole capacity as the maximum cell throughput of a single site for the user plane. Capacity is influenced by multiple aspects: given service type, available bandwidth, network features and topology, distribution of users among the cell and their speed, number of sectors in a given site, power settings, used scheduler or coverage area are just few of the most important contributors to the final output. Such studies are also helpful in the network planning phase. In the early network rollouts the biggest focus is on providing sufficient coverage, but in the densely populated areas or hotspots with large penetration of user devices, estimation of capacity can translate into the numbers of sites in a given region and their configuration.

Conclusions of capacity studies, together with real network measurements, can help the operators to predict possible congestion situations, which can severely affect the Quality of Experience perceived by their clients. This subjective measure is the sum of all factors experienced by a user while using a service that results in an overall satisfaction level. One example of such a criterion may be a delay between the time when user clicks on a hyperlink in a web browser installed on his device and the time when the content of a web page is actually available. Various researches showed that in order to provide reasonable quality of experience this delay can't be longer than 2 seconds [14], but modern users are even more demanding and usually tend to expect the performance similar to the one experienced in fixed networks.

Usually two methods can be used to obtain pole capacity: analytical calculations and so-called system level simulations. Analytical approach is based on simplified assumptions and statistical processes. System level simulation approach additionally considers dynamic processes like fading, mobility or handovers. Despite the fact that it may be more accurate in providing the estimation for the pole capacity of the real network site, it takes more time to derive the results using this method, and the implementation of any changes related to simulation assumptions requires more time than in the other method. Typically both approaches are used and quickly derived analytical values which are calculated more often are cross checked with the system level numbers.

The capacity planning of UMTS systems, especially for the cells in urban areas, is done not for the averaged data demand of a given zone, but for the so-called busy hour, the 60 minute period of the day where operators experience the most heavy traffic in their system.

As already mentioned, capacity but also coverage studies can influence Node B configuration, but of course it is a mutual relation. A typical HSPA feature that can influence the site capacity is Dual Carrier HSDPA, which roughly doubles the site's air interface pole capacity comparing to a single carrier operation (assuming the same configuration for power, codes and antennas on both carriers). Another good example is the support of Higher Order Modulations (64 QAM) or Dual Stream 2×2 MIMO. Of course the real life capacity improvement will also depend on the penetration of the UE devices that are capable of supporting those features.

Typically, coverage upgrades are linked with capacity upgrades. Features like new receiver types or MIMO fallback mode aim at improving the SINR on UE side which can lead to higher supported throughput by the UE. Coverage can be also extended by lowering the noise factor in the user device. In interference limited scenarios of HSDPA systems, the coverage of a given cell can be artificially changed by applying so-called range extension (signal quality biasing). In this method, the UE is informed by the network that it should perceive the pilots' signal quality as stronger or weaker by a given factor. Since the pilots' signal quality is used among others to estimate which cell it should camp on, the UE can, for example camp on a given cell despite the fact that the signal quality of a given cell is worse than its neighbor. Such an approach is mainly used for load balancing, for instance to push some UEs that are originally served by the heavily loaded macro stations to the neighboring pico stations or other less loaded cells.

Coverage optimization is typically important in the early network rollouts, but as the network is becoming more mature and its traffic load is increasing, capacity improvements are becoming essential to provide sufficient user experience. In many practical cases, at some point the limitation is not only the air interface but also the backhaul capacity. With the advent of HSPA, the network operators experienced a huge traffic upsurge and a lot of sites had to be upgraded with respect to their backhaul capacity. This problem was very often seen on the Iub interface due to the fact that many 3G sites were collocated with 2G stations and they shared the same backhaul based on the E1/T1 links. In some cases, an upgrade from copper wires to optical fibers technology was necessary to prevent data throttling on the backhaul.

Another aspect of the network capacity optimization is the densification of the network and HetNet technology. At this point it is worth to mention that due to interference limitations, the network capacity does not grow linearly with respect to network densification.

The capacity and coverage planning phase is usually followed by the optimization phase linked with drive tests. The first one aims at optimization of various Radio Access Network parameters like tilt or antenna direction of deployed Node Bs. The goal of those actions is the detailed optimization of already placed access nodes. Drive tests are real life measurements done by the service teams of the operators or network equipment vendors that verify the correctness of the aforementioned actions.

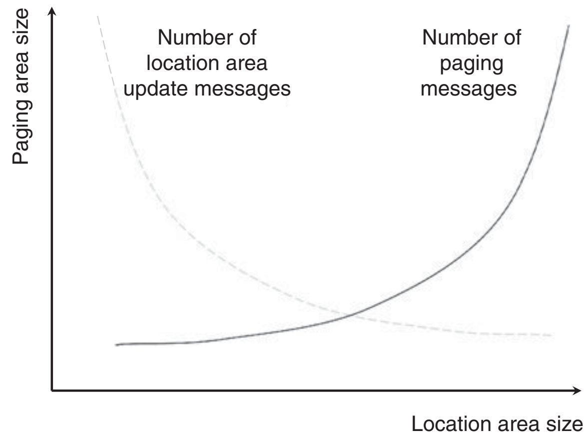

Figure 12.20 *Paging load and location area optimization.*

12.5.3 Location Areas Size vs Paging Load

Another example of network dimensioning and planning may be the decision on the Location Area (LA) size. LAs are created by several cells that are adjacent to each other and are sharing the same Location Area Code (LAC). The purpose of such cell grouping is the fact that when UE is not in the cell DCH state, its location is known up to the level of Location Area in CS domain or Routing Area (RA) in PS domain. If the location of UE needs to be known down to the cell level (e.g. there is a pending voice connection from another users and the network needs to know which NodeB should handle the transmission) the network sends a paging information to all cells belonging to a given Location Area in order to determine the UE position. The size of the Location Area, that is the number of clustered cells, will reflect the overall signaling load caused by the paging procedure, as the paging information is sent in every cell. The larger the LA is, the higher signaling load is generated.

On the other hand we can imagine a situation when UE is in connected mode in URA PCH state and no paging is ongoing. The user is moving quite fast, as for instance he is traveling in a car on a highway. Whenever the UE detects that it has crossed the border between different LAs, and the new cell operates on a different LAC, it needs to perform a Location Update Procedure to inform the network about its new location. This operation not only drains the user's device battery, but also results in signaling load generated in the network. With small LA sizes and users with large mobility, this may lead to a situation when such signaling will become a large part of the total load of the UMTS system, draining the resources that could otherwise be used for the user's data transmission.

The consensus between those two contradictory tendencies is depicted in Figure 12.20 and is the goal of network dimensioning engineers.

References

1. 3GPP TS 25.420 v.11.0.0, *UTRAN Iur Interface: general aspects and principles (Release 11)*, 2012.
2. 3GPP TS 25.430 v.11.0.0, *UTRAN Iub Interface: general aspects and principles (Release 11)*, 2012.
3. Holma, H. and Toskala, A. (2010) *WCDMA for UMTS: HSPA Evolution and LTE*, John Wiley & Sons, Ltd.
4. Ahson, S.A. and Furht, B. (2011) *HSDPA/HSUPA Handbook*, CRC Press.
5. 3GPP TS 25.410 v11.0.0, *UTRAN Iu Interface: general aspects and principles (Release 11)*, 2012.

6. van Trees, H.L., Bell, K.L. and Tian, Z. (2013) *Detection, Estimation, and Modulation Theory. Part 1 – Detection, Estimation and Filtering Theory*, John Wiley & Sons, Ltd.
7. 3GPP TS 25.331 v.11.7.0, *Radio Resource Control (RRC); Protocol specification (Relese11)*, 2013.
8. 3GPP TS 25.304 v11.4.0, *User Equipment (UE) procedures in idle mode and procedures for cell reselection in connected mode (Release 11)*, 2013.
9. 3GPP TS 25.213 v11.4.0, *Spreading and modulation (FDD) (Release 11)*, 2012.
10. 3GPP TR 25.872 v11.0.0, *High Speed Packet Access (HSDPA) multipoint transmission (Release 11)*, 2011.
11. 3GPP TS 21.101 v12.0.0, *User Equipment (UE) radio transmission and reception (FDD) (Release 12)*, 2013.
12. ITU, *Spectrum Management for a Converging World: Case Study on the United Kingdom*, [Online]. Available: http://www.itu.int/osg/spu/ni/spectrum/UK-RSM.pdf (last accessed 27 July 2014).
13. CESifo Working Paper No. 584, *The third generation (UMTS) spectrum auction in Germany*, V. Grimm, F. Riedel, E. Wolfstetter, October 2001.
14. Galletta, D.F. (2004) *Web Site Delays: How Tolerant are Users*, vol. **5**(1), JAIS.

13

3GPP Mobile Communications: LTE/SAE and LTE-A

Jacek Góra, Krystian Safjan, Jarosław Lachowski, Agnieszka Szufarska, Stanisław Strzyż, Szymon Stefański, Damian Kolmas, Jyrki T. J. Penttinen, Francesco D. Calabrese, Guillaume Monghal, Mohammad Anas, Luis Maestro, Juha Kallio and Olli Ramula

13.1 Introduction

Long Term Evolution (LTE) and System Architecture Evolution (SAE) are the next, evolutionary steps in the development of mobile communication systems for both radio access (LTE) and core network (SAE). They are often commonly referred to as the "4th generation" systems (4G) though such terminology is over-simplified as within LTE itself one can distinguish a number of releases with gradually more advanced features and better performance.

The work on 4G mobile communication systems began in 2004 when there was already a mature concept for evolving WCDMA in a more mature state – HSPA. It was clear that the 3rd generation (3G) systems might still be competitive and guarantee bit rates required by end-users for the next few years – yet 3GPP decided to go for resource-consuming development of the 4G system with an entirely new air interface as well as significant changes in system architecture.

The motivation and requirements can be found in the publicly available 3GPP specification [1] – one of opening sentences there reads: "(. . .) to ensure competitiveness in an even longer time frame, that is, for the next 10 years and beyond, a long-term evolution of the 3GPP radio-access technology needs to be considered. Important parts of such a long-term evolution include reduced latency, higher user data rates, improved system capacity and coverage, and reduced cost for the operator."

The Telecommunications Handbook: Engineering Guidelines for Fixed, Mobile and Satellite Systems, First Edition.
Edited by Jyrki T. J. Penttinen.

The above-mentioned requirements were stated in a more specific way in further sections of Ref. [1] as targets for the evolution of the radio-interface and radio-access network architecture:

a. Improved data rates:
   - Increased peak data rate (initial numbers: 100 Mb/s for downlink and 50 Mb/s for uplink).
   - Increased throughputs in cell edge regions.
   - Significantly improved spectrum efficiency (initially 2–4 higher than in UMTS Release 6).
b. Reduced latency:
   - In radio-access network (user plane) < 10 ms.
   - In control plane < 100 ms (from camped-state to user data exchange, excluding downlink paging delay).
c. Other aspects:
   - Scalable bandwidth.
   - IP-optimized transport.
   - Reduced all cost components.
   - Assured interworking with other technologies.
   - Enhancements in MBMS and IMS.

For full list and more details please refer to the source specification [2–25]. References [2], [15] and [23] in particular are often referred to in this chapter.

13.2 Architecture

The evolved 3G system consists of radio and core networks. The core side of the new network is called SAE (System Architecture Evolution), and the radio part is called LTE (Long Term Evolution). It should be noted that the SAE was utilized at the beginning of the standardization, but nowadays the core part is called EPC (Evolved Packed Core). As the term SAE has been established in common terminology, it can be utilized in parallel with EPC. This book utilizes both the official standard term EPC as well as the nonstandard SAE.

This chapter presents the functional blocks and interfaces of LTE/EPC. The new architecture is illustrated with comparisons of solutions with earlier mobile communications systems.

Also a protocol layer structure and functioning of the LTE/EPC protocols are described, and examples are given in order to clarify the principles of each protocol. The LTE architecture is defined in 3GPP specification [26] (E-UTRAN Architecture description).

Similarly as the LTE system is the next step in the evolution of the radio communication interface, the system architecture evolution (SAE) is the next step in the evolution of the system architecture. The drivers for evolution of the SAE are similar to those for the LTE. The key aspects are: (1) optimization for packet switched services, (2) support for higher data rates and (3) reduced latencies (packet setup and round-trip time). The main step taken to meet those targets was to flatten the system architecture. The flat architecture model means fewer nodes, simplified procedures and less signaling in comparison with previous versions of the CN.

This section presents details of the system architecture evolution. Firstly, the details of the evolved packet core (EPC) are presented, followed by presentation of the evolved universal terrestrial radio access network (E-UTRAN). In both parts, the nodes as well as interfaces, are described with focus on their functionalities.

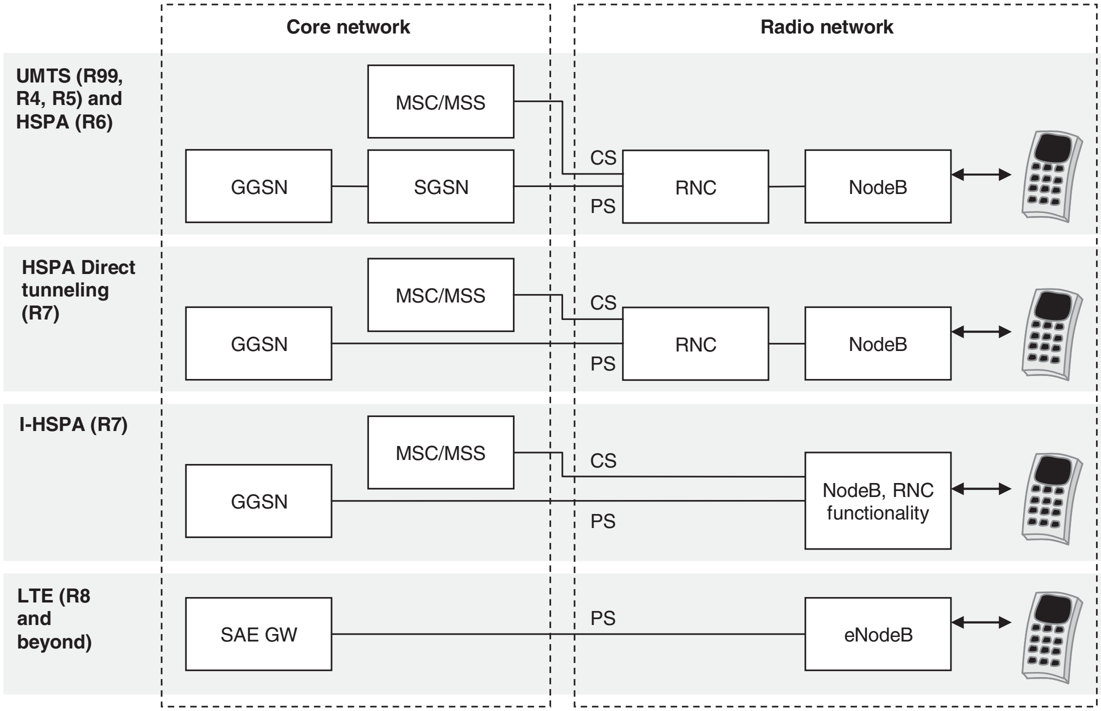

Figure 13.1 *The evolution of network architectures. LTE simplifies the general layout of the network via the flat architecture by providing only packet switched connections.*

13.3 Elements

LTE is based on flat architecture, meaning there is only one element type for the radio network, and one element type for the core network. Figure 13.1 shows the high-level architecture of LTE and compares it with the packet switched domain of the earlier 3GPP mobile systems.

The overall LTE/EPC division is the following: LTE refers to the E-UTRAN (Evolved UMTS Radio Access Network), whilst EPC means the evolved core network. Figure 13.2 clarifies the division.

The EPC is the new version of core network (CN) developed to support the LTE radio interface. Following the general requirements for the SAE the EPC is strictly packet switched oriented. Furthermore, it follows the flat architecture model. The EPC involves five types of nodes. The EPC nodes are (Figure 13.4):

- Mobility Management Entity (MME)
- Serving Gateway (S-GW)
- Packet Gateway (P-GW)
- Policy and Charging Resource Function (PCRF)
- Home Subscription Server (HSS).

The nodes cooperate with each other on control and/or user planes to enable users to communicate over the operator's network as well as access to external networks (e.g. the Internet).

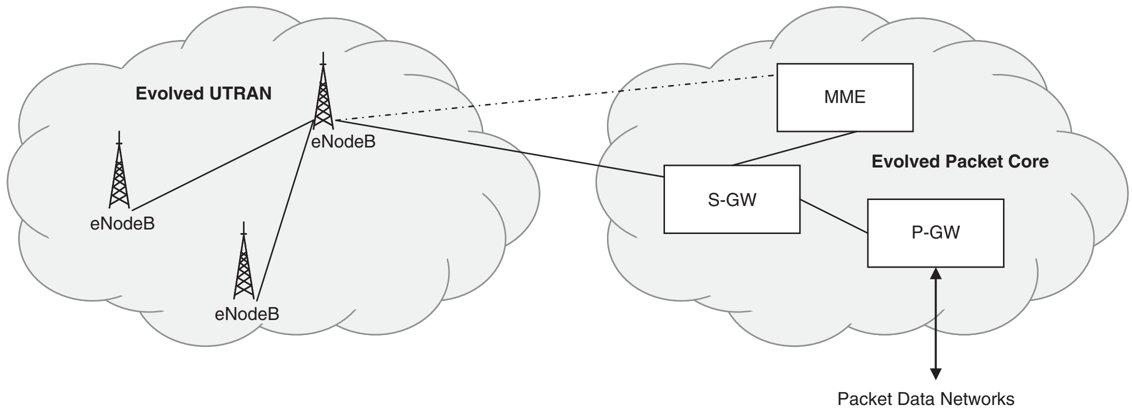

Figure 13.2 *The Evolved Packet System (EPS) consists of Evolved UTRAN (LTE) and EPC.*

Figure 13.3–13.5 show the elements with respective interfaces. As the figure indicates, the LTE can be connected to the packet core of GERAN and UTRAN as a new additional radio technology via the EPC. The SGSN acts as a centralized connection point for all of these technologies.

E-UTRAN contains only one type of node, eNB, which provides the air interface to UE. The eNB elements are connected to MMEs and S-GWs via the S1 interface. Furthermore, eNBs can be connected to each other via the X2 interface. It is worth noting that unlike in previous solutions, in LTE the single eNB can be

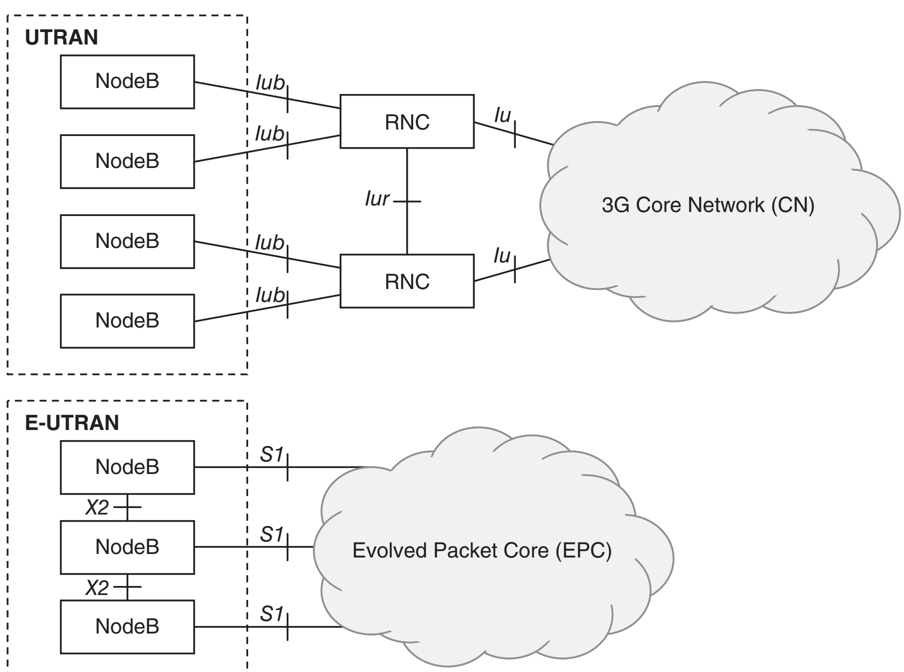

Figure 13.3 *The architectural difference between UTRAN and E-UTRAN.*

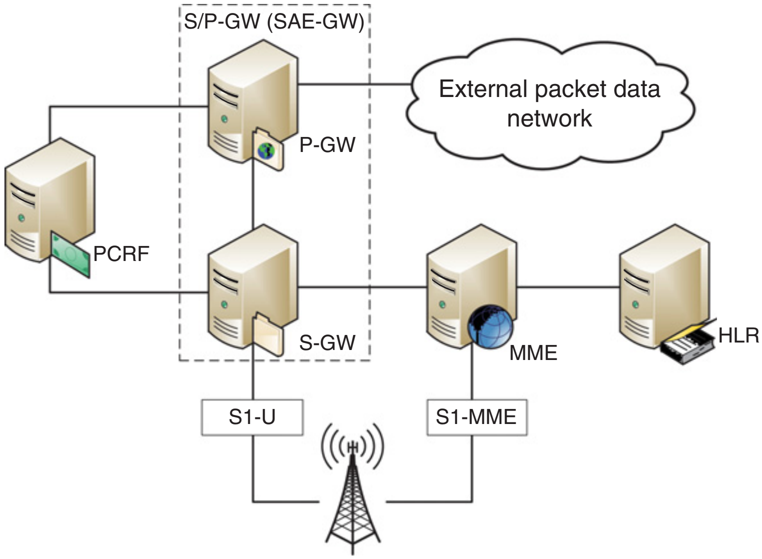

Figure 13.4 Structure of the EPC including nodes and interfaces.

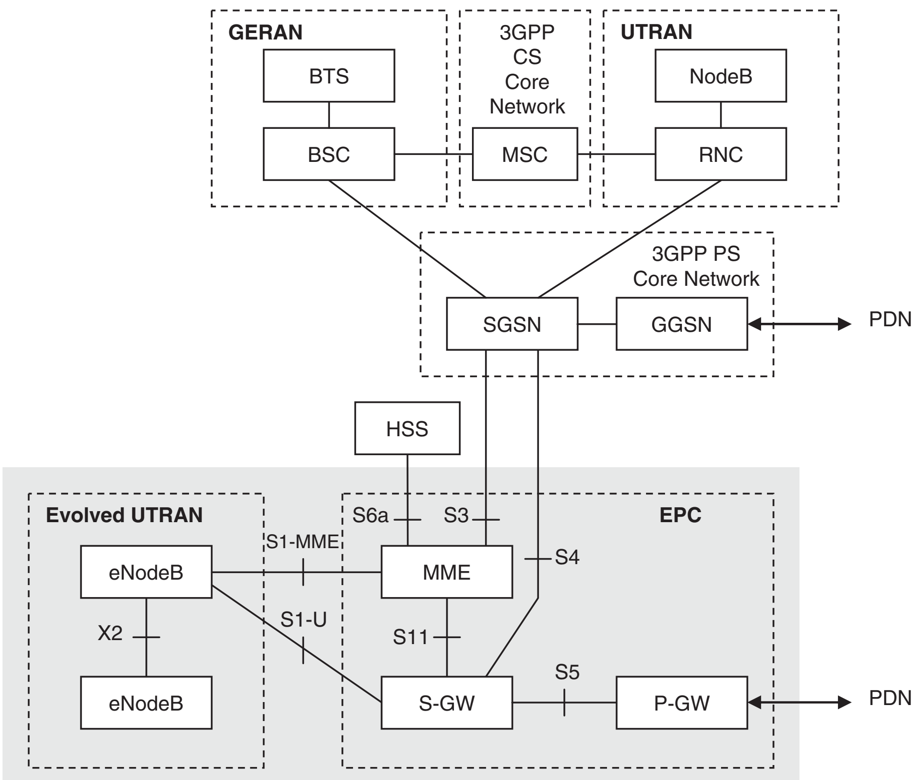

Figure 13.5 LTE/SAE elements and interfaces.

connected to multiple MMEs and multiple S-GWs. This ability provides additional reliability in a flexible way, and is referred to as S1-flex.

13.4 Evolved Universal Terrestrial Radio Access Network

While the EPC is not visible for an end-user of a cellular system, he/she is typically at least to some extent familiar with the radio access part of the network. Next the building blocks of the LTE Radio Access Network (RAN) – the nodes and the interfaces (see Figure 13.6) – are described in detail.

13.4.1 eNodeB

The base station in the LTE system is called the evolved Node B (eNodeB, eNB). The eNodeB is the only infrastructure element of the E-UTRAN. It controls all radio interface communication with UEs. On the other side, it has a direct connection to the EPC (to S-GW for user plane and to MME for control plane communication). Compared to the UMTS system architecture the eNodeB is a combination of the base transceiver station (BTS) and the radio network controller (RNC). The merging of BTS and RNC functionalities follows the flat architecture principle. By this step more intelligence has been given to the base station which allows more effective resource management (especially in terms of timing). The merging of RNC with BTS functionalities also, however, has a drawback. In the UMTS system one of the RNC roles was to provide coordination over operation of multiple base stations. In the LTE system eNodeBs lack this centralized coordination (to some extent the operator's operation and maintenance, OAM, functionality can play this role, but only with very slow adaptation). For this reason complex coordination schemes have been developed based on the inter-eNodeB information exchange over the X2 interface.

The eNB element of LTE is responsible for radio transmission and reception with UE. eNB provides the needed functionality for radio resource management (RRM, including admission control, radio bearer control, scheduling of user data, and control signaling over the air interface. In addition, eNB takes care of ciphering and header compression over the air interface.

The clearest difference between UTRAN and E-UTRAN can be seen in the role of the base station. The eNodeB of LTE now includes basically all functionalities that were previously concentrated on the RNC of the UTRAN system. In addition, the traditional tasks of the NodeB are still included in the new eNodeB element. eNodeB works thus as the counterpart of the UE in the radio interface, but includes procedures for decision making related to the connections. This solution results in the term "flat architecture" of LTE, meaning that there are less interfaces and only one element in the hierarchy of the architecture.

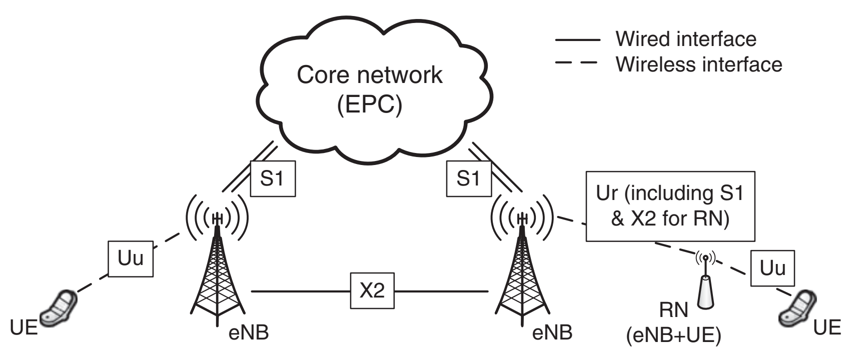

Figure 13.6 *Nodes and interfaces of the E-UTRAN.*

As the controlling has been moved closer to the radio interface, the respective signaling time has also been reduced. This is one of the key issues for the reduction of the latency of LTE compared to previous solutions of 3G.

More specifically, the eNB element handles the following tasks:

- Radio Resource Management (RRM).
- Radio Bearer Control.
- Radio Admission Control.
- Connection Mobility Control.
- UE scheduling (DL and UL).
- Security in Access Stratum (AS).
- Measurements as a basis for scheduling and mobility management.
- IP header compression.
- Encryption of the user data.
- Routing of the user data between eNB and S-GW.
- Handling of the paging that originates from MME.
- Handling of the broadcast messaging that is originated from MME and Operations and Management System (OMS).
- Selection of the MME element in case UE does not provide this information.
- Handling of PWS messages, including ETWS and CMAS.

It is also possible to utilize an additional element set which is called Home eNB and Home eNB Gateway. Specific aspects for the Home eNB, that is, HeNB, are the following:

- HeNB are equipment that can be utilized in the customer's premises and that use the licensed operator's spectrum.
- HeNB is meant for the enhancing of the network coverage and/or capacity.
- Includes all the eNB functionalities, added by the HeNB-specific functions that are related to the configuration and security.

Related to the HeNB, the Home eNB Gateway, that is, HeNB GW solves the problem of the support of possibly a very large number of S1 interfaces. It is thus an additional element that can be utilized for the balancing of the interfaces.

Figure 13.7 shows the principle of the HeNB concept, together with HeNB GW elements.

Furthermore, the HeNB concept can be utilized in the following access scenarios:

- In the closed access mode, only predefined Closed Subscriber Group (CSG) members can access the respective HeNB.
- In the hybrid access mode, both members and nonmembers of the closed subscriber group can access the HeNB, but the members are prioritized over nonmembers, for example, in the case of the congestion.
- In the open access mode, HeNB is seen by members and nonmembers exactly as a normal eNB.

The correct functioning of the closed and hybrid modes requires additional parameters in order to support the identification and search of the HeNB by UE. Also mobility management should be aware of the HeNBs in order to perform handovers.

The Release 9 of LTE includes enhancements for the HeNB concept. Some of the most important radio access additions are Ref. [27]:

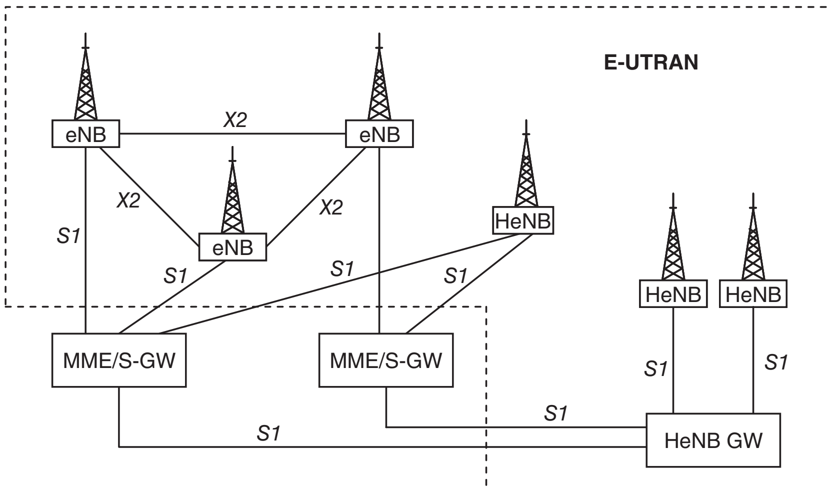

Figure 13.7 The idea of Home eNB concept.

– Inbound Mobility.
– Access Control.
– New Hybrid Cell concept.
– Management of out-of-date CSG info.
– Operation, Administration and Maintenance for HeNB elements.
– Operator controlled CSG list.
– RF Requirements for TDD and FDD HeNBs.

13.4.2 User Equipment

The user equipment (UE) is the user terminal device that allows him/her access to the E-UTRAN. This can be either a mobile telephone or a data modem. The UE contains an interface for the operator to control behavior of the device and an E-UTRAN radio transceiver. In addition, one of the key functional elements of the UE is the Universal Subscriber Identity Module (USIM) that contains basic information for identifying the subscriber (e.g. International Mobile Subscriber Identity – IMSI, phone number).

13.4.3 S-GW

The Serving Gateway (S-GW) provides the user plane connectivity, the UE being on the other side, and Packet Data Network Gateway (P-GW) on the other side of the physical S-GW element. Depending on the network provider's approach, these elements can be separate ones, or they can be combined physically as a single element.

It should be noted that no control messaging goes between UE and S-GW, as the control plane is taken care of by the MME element.

The S-GW element takes care of the following functionalities:

– S-GW is the local anchor point for the inter-eNB handover procedure.
– S-GW is also an anchor point for the inter-3GPP network mobility.
– Lawful Interception (LI).

- Packet routing and forwarding.
- S-GW makes the packet buffering in the E-UTRAN idle mode.
- S-GW handles the network initiated/triggered service request procedure.
- Packet marking in the transport level for both DL and UL.
- Charging Data Record (CDR) collection, which can identify the UE, PDN and QCI.
- Accounting on user and QCI granularity for the interoperator charging processes.

The S-GW is thus a local router for user plane transmissions. Its main functionality is to manage the GTP (GPRS Tunneling Protocol) data tunnels for user transmissions within EPC. In case of user transmissions to an external network (e.g. the Internet) the user plane data is routed from the S-GW to the P-GW, and vice versa. The S-GW is also involved in the user handover procedure. It is responsible for establishing packet forwarding between base stations during the handover procedure and path switching (i.e. redirecting GTP tunnels to the new serving base station) after successful handover. While the S-GW operates mainly on the user plane (UP), it also receives control plane (CP) requests from other EPC nodes related to GTP tunnel establishment and management.

13.4.4 P-GW

The PDN Gateway (P-GW) provides, equally as the S-GW, user plane connectivity in the chain of UE, S-GW and P-GW. The P-GW element interfaces with the S-GW, and on the other side, with the external packet data network (PDN). In addition, P-GW includes GGSN (GPRS Gateway Support Node) functionality.

More specifically, P-GW includes the following functionalities:

- UE IP address allocation.
- Packet filtering that can be done in user-based level. The other term for this functionality is a deep packet inspection.
- Lawful Interception (LI).
- Packet marking in the transport level, in DL.
- Service level charging in DL and UP, as well as gating and rate enforcement.
- Rate enforcement in DL based on APN-AMBR.
- Online charging credit control.

The P-GW (or PDN-GW) is thus the router between the operator's EPC and the external Packed Data Networks (PDNs). It assigns IP addresses to user terminals and, if needed translates IP packets coming from the external network into GTP tunnels used within the EPC. The P-GW is also the highest level anchor point for a user terminal mobility, that is, it is unchanged during the connection session of the terminal. The P-GW communicates on the control plane with the PCRF from which it receives messages related to bearer creation and management. On the user plane the P-GW communicates with the S-GW, with which it exchanges user plane data related to communication with the external network. Due to the strict relation and similar functionalities of the S-GW and the P-GW, the two entities are commonly integrated into a common node called the S/P-GW or SAE-GW.

13.4.5 MME

The Mobility Management Element (MME) is meant for control plane signaling between the UE and other network elements like HSS. Equally, as the user plane LTE/SAE messaging does not go through MME, the control plane signaling does not go through S-GW or P-GW of LTE/SAE.

MME handles the following functionalities:

- Signaling in the Non Access Stratum (NAS).
- Security of the NAS signaling.
- AS security control.
- The selection of the P-GW and S-GW elements.
- The selection of other MMEs in the case of the handover.
- The selection of SGSN in the case of handovers between LTE and 3GPP 2G/3G access networks.
- Inter-CN node mobility signaling between different 3GPP 2G/3G access networks.
- The management of the Tracking Area (TA) lists.
- International and national roaming.
- User authentication.
- The establishment and management of bearers.
- The support of PWS message transmission, including ETWS and CMAS.
- The management of the paging retransmission of UE and other functions related to the finding the UE in the idle state.

In other words, MME is the main control element in the core network. It is involved in session establishment and management for user terminals within its coverage area (corresponding to a list of base stations connected to the MME). Exact responsibilities of the MME are:

- Security procedures – the MME authenticates user terminal when it tries to attach to the network. Authentication is performed on the basis of authentication keys provided from the HSS. After successful verification the user terminal is assigned with a Global Unique Temporary Identity (GUTI) that masks its real identity for security reasons. Finally, after the connection session is established for the user, the MME manages ciphering and integrity protection.
- Mobility management – the MME tracks user terminal position and provides the location information to the HSS in the user's home network. The tracking is done with per cell resolution for active terminals (the serving cell is known) or with per tracking area (TA) resolution for the idle terminals (idle terminals signal current TA code, TAC, periodically or on TA change). Finally, the MME is also involved in the user terminal handover process, by controlling resource reservation and release at base stations and at S-GW.
- Session management – upon establishment of a connection session the MME creates a register (the user context) for the connecting user terminal. The register holds information about the subscription profile for the user provided from the home network HSS, as well as information about the running session. The MME also sets up the basic bearer for the user terminal and may be involved in creation of dedicated bearers.

13.4.6 Policy and Charging Resource Function (PCRF)

The PCRF has two main responsibilities. The first is policy control. Based on service requests provided from the S/P-GW or from the external network, the PCRF makes decisions on how the network should handle the requested service in terms of quality-of-service (QoS). This leads to defining the Policy and Charging Control (PCC) rules that control bearer creation and management for the service. Secondly, the PCRF is responsible for charging subscribers for using the operator's network. The PCRF is in direct connection with the S/P-GW. This allows the PCRF to monitor and control user plane data flow to and from the user terminals, and to charge the user accordingly.

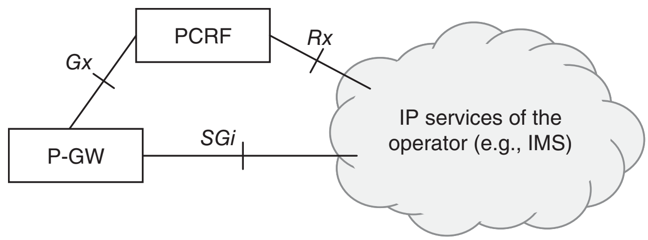

Figure 13.8 The location of the PCRF element in the LTE/SAE network.

13.4.7 Home Subscription Server (HSS)

The HSS is a server containing a database of all subscribers of the operator. The database contains information about the subscribers' identities and details of their subscriptions (Home Location Register, HLR, functionality). Furthermore, the HSS is responsible for generating security keys used by MME for authenticating the subscribers, and for encryption and integrity protection (Authentication Center, AuC, functionality).

During attachment of a user terminal to the network the corresponding MME communicates with the central HSS to validate the subscriber identity and to determine the enabled communication options (e.g. quality-of-service profile). For this reason the HSS needs to have connections to every MME in the operator's network and can track the user terminal position with the per MME resolution.

13.4.8 GSM and UMTS Domain

As Figure 13.5 indicates, the SGSN can be utilized for the centralized point for connecting the PS domains of GERAN, UTRAN and LTE. LTE creates the packet switched connections to the external packet data networks via P-GW, whilst the GERAN and UTRAN utilizes the traditional GGSN. These legacy network elements that are also relevant to LTE are the following:

- Gateway GPRS Support Node (GGSN) is responsible for terminating Gi interface towards the PDN for legacy 2G/3G access networks. For the LTE/SAE network, this node is only of interest if provided as parts of P-GW functionality and from the perspective of inter system mobility management.
- Serving GPRS Support Node (SGSN) is responsible for the transfer of packet data between the core network and the legacy 2G/3G Radio Access Network (RAN). For the LTE/SAE network, this node is only of interest from the perspective of inter system mobility management.
- Home Subscriber Server (HSS) is the IMS Core Network entity which is responsible for the management of the user profiles, and performs the authentication and authorization of the users, including the new LTE subscribers. The user profiles managed by HSS consist of subscription and security information as well as details on the physical location of the user.
- Policy Charging and Rules Function (PCRF) is responsible for brokering QoS Policy and Charging Policy on a per-flow basis. Figure 13.8 shows the position of PCRF in the LTE/SAE architecture.
- Authentication, Authorization & Accounting function (AAA) is responsible for relaying authentication & authorization information to and from non-3GPP access network connected to EPC.

13.4.9 Packet Data Network

Packet data network (PDN) is the IP based network to which LTE/SAE connects via the P-GW element. PDN can be, for example, Internet or operator's IP Multimedia Subsystem (IMS).

13.5 Interfaces

The LTE/SAE system consists of several interfaces internally and between the other 3GPP 2G/3G networks, as is shown in Figure 13.5.

13.5.1 Uu Interface

The LTE radio interface LTE-Uu is defined between the eNodeB and UE. The eNodeB provides PS connectivity in such a way that previous 3G RNC functionality is integrated into the eNodeB. This flat architecture approach makes separate RNC equipment unnecessary.

Uu is the radio interface between an eNodeB and a UE. It is used for making user plane transmissions as well as for exchanging control plane information. The Uu control plane signaling covers all layers of protocol stack starting with the Non-Access Stratum (NAS) Mobility and Session Management signaling exchange with the MME, down to the physical layer control signaling (radio resource scheduling, power control, etc.).

The Uu interface has been also adapted in LTE-Advanced standard to support communication between a relay node (UE-functionality) and its donor eNodeB. In such cases the interface is called the Ur interface. In case of relay nodes the control plane transmissions on the Ur interface correspond to control information targeting the UE-functionality of the relay, similarly as on the Uu interface of a UE. The user plane transmission, however, may include the user and control plane data targeting the UEs connected to the relay node as well as control plane information targeting the eNodeB-functionality of the relay node (S1 and X2 forwarding).

13.5.2 X2 Interface

The X2 interface defines the connection between the eNodeB elements. It is meant for the inter-eNodeB handover procedures, as well as for the intercell radio resource management signaling and interface management signaling.

Physically, the interface can be for example, fiber optics or any other solutions whilst the required capacity can be delivered with the required maximum delay and jitter.

The X2 interface is thus a logical interface interconnecting eNodeB located in close vicinity (neighboring eNodeBs). The logical characteristic means that the X2 interface can have the form of either a direct physical connection or (in case of its absence) it can be established over the EPS (forwarding over the S1 interface). The main functionalities of the X2 interface correspond to: (1) user mobility and (2) inter-eNodeB cooperation.

The mobility support function of the X2 interface relates to inter-eNodeB communication during UE handover preparation and execution. The source serving eNodeB may ask the target eNodeB, if it is ready to accept a new UE. If the target eNodeB can accept the new UE a handover procedure is executed, during which the source eNodeB forwards to the target eNodeB the downlink user plane packets incoming from the S-GW. Packet forwarding is terminated after the handover procedure is successfully finalized (late path switch procedure).

The cooperation support signaling in the basic form includes information related to the status and optimization procedures of the eNodeBs. One example of signaling of this type are the intercell interference coordination (ICIC) messages. The ICIC X2 messages enable an eNodeB to inform its neighbors about the level of interference received or generated in downlink and uplink. Other, commonly utilized, type of X2 signaling are the messages related to load control. By the means of this type of signaling the eNodeBs can indicate traffic load status and negotiate mobility parameters for users.

For details on the X2 interface please refer to the 3GPP specifications [28–32].

13.5.3 S1 Interface

The S1 interface is divided into S1-MME and S1-U. S1-MME connects eNodeB and MME elements, whilst S1-U is utilized between the eNodeB and S-GW elements.

S1-MME interface is defined for control plane signaling between eNodeB and MME. S1-U interface is defined between the eNodeB and S-GW in order to carry the user plane data. Data is transferred over GTP.

The S1 interface is thus the interface interconnecting the E-UTRAN with the EPC. Specifically, the S1 interface connects eNodeB with MME and S-GW. Each eNodeB can be connected to multiple MMEs and S-GWs. The same applies for each MME and each S-GW can be connected to multiple eNodeBs. This provides higher reliability of the network as well as lower latencies in communication.

From the functional point of view the S1 interface can be characterized as the S1-U interface for the user plane communication (eNodeB – S-GW) and the S1-MME for control plane communication (eNodeB – MME). The S1-U interface mainly supports exchange of user plane data packets. The S1-MME interface is used for control signaling related to user mobility, but also configuration of the eNodeB and inter-eNodeB communication (in case a direct physical connection is not available). For details on the S1 interface please refer to the 3GPP specifications [33–37].

13.5.4 S3 Interface

The S3 interface is meant for the signaling between the MME and SGSN.

13.5.5 S4 Interface

The S4 interface is defined between the S-GW and SGSN. This provides a GTP based tunnel in user plane during the intersystem handover.

13.5.6 S5 Interface

The S5 interface is located between the S-GW and P-GW elements. Depending on the vendor solution, these elements can be integrated in the same physical element.

13.5.7 S6a Interface

The S6a interface carries the subscription and authentication information between the HSS (Home Subscriber Server) and MME.

13.5.8 S11 Interface

The S11 interface handles the signaling messages between the S-GW and MME.

13.5.9 SGi

The SGi interface is defined between the P-GW and packet data network (PDN), which can be an external public or private, IP packet network. It also can be internal IP network, like IP Multimedia Subsystem.

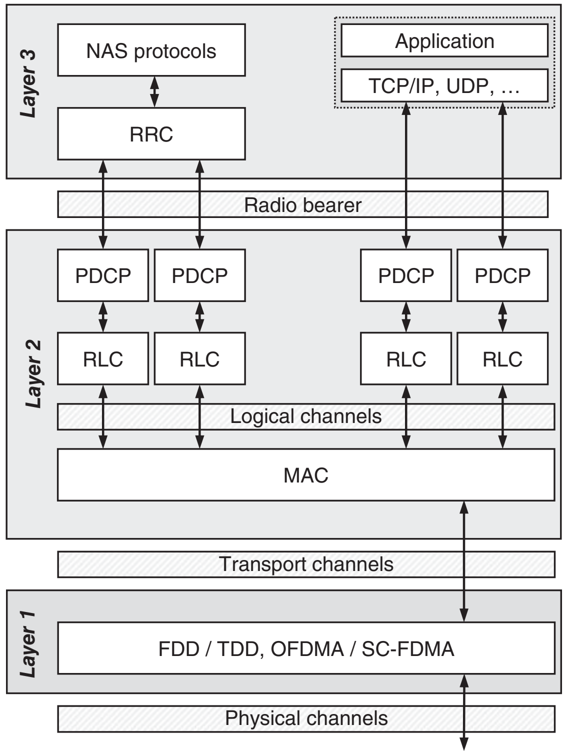

Figure 13.9 *The role of LTE layers.*

13.5.10 Gn/Gp

As an alternative to the SGi interface, also legacy Gn/Gp interface is supported in EPS in order to create the connection to the packet data networks.

13.6 Protocol Stacks

LTE protocol stacks between different elements are divided into user and control plane. In general, protocol stacks are similar to those utilized in the WCDMA of UMTS.

Figure 13.9 presents the overall idea of the roles of each LTE protocol layer. The following subchapters describe functionalities in more detail.

The LTE/SAE radio protocol stacks are defined in layers of 1, 2 and 3, which overlap partially with definitions of the OSI layer structure of ISO. The LTE/SAE layer 1 is related to the physical realization of the interface (like radio interface or optical fiber), whereas layer 2 is related to data link and access, and layer 3 is related to hosting of the access stratum protocols or nonaccess stratum signaling protocols. In LTE/SAE, the application level is included in this 3rd layer.

13.6.1 User Plane

Figure 13.10 shows the complete protocol stuck structure for the user plane between the UE, eNB, S-GW and P-GW.

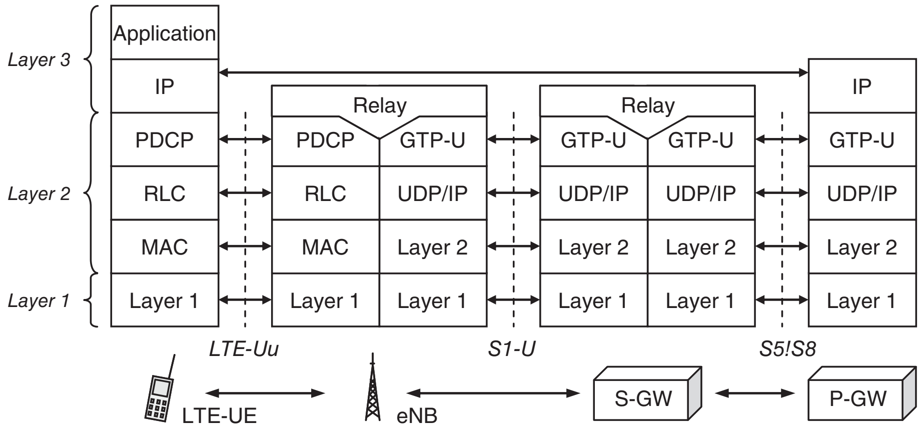

Figure 13.10 *The LTE/SAE protocol layers for the user plane.*

Additionally, the user plane is defined as shown in Figure 13.11 in case of the direct communications between two eNBs.

The functionalities of the essential user plane entities are the following:

MAC takes care of the mapping between the logical and transport channels, multiplexing and demultiplexing, reporting of the scheduling information, HARQ functions, priority handling and transport format selection.

RLC takes care, among other tasks, the ARQ functions, segmentation concatenation, resegmentation concatenation, in-sequence delivery, duplicate packets detection and reestablishment.

PDCP layer takes care of the ciphering of the user and control plane, header compression (ROCH), in-sequence delivery of the upper layer packet data units (PDU), duplicate elimination of the lower layer SDUs, integrity protection for the control plane, and timer based discarding.

13.6.2 Control Plane

The control plane protocol stack structure is shown in Figure 13.12.

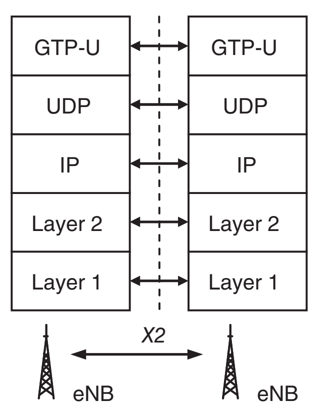

Figure 13.11 *The user plane protocol stack structure in the case of the communications between two eNBs.*

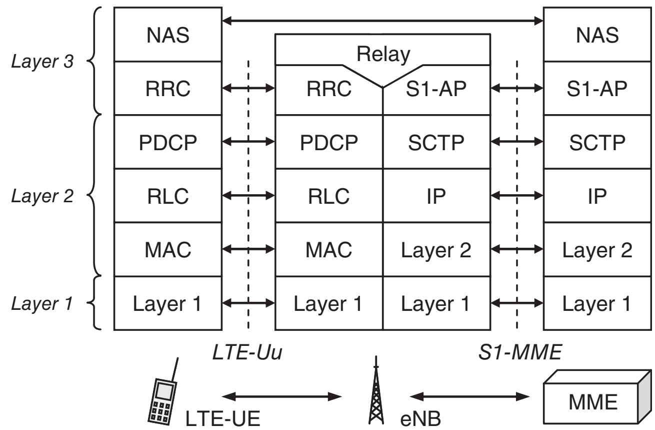

Figure 13.12 The control plane protocol layer structure of LTE/SAE.

Again, in the case of direct communication between two eNBs, the protocol stack is as presented in Figure 13.13. This communication can happen, for example, when the handover procedure takes place.

13.6.3 Layer 1

The LTE/SAE radio protocol layer 1 describes the physical layer. In general, it provides the means and basic functionality in order to deliver the bits over the air interface both in downlink and uplink directions.

The radio interface of LTE is based on two separate access techniques: OFDMA (Orthogonal Frequency Division Multiple Access) in the downlink direction, and SC-FDMA (Single Carrier Frequency Division Multiple Access) in the uplink direction. The functionality of OFDMA and SC-FDMA is described more detailed in Chapter 7.

There is a set of LTE channels defined for signaling and data delivery. The channel definitions have been simplified compared to previous 3G solutions of UMTS and HSPA, including removal of dedicated channels.

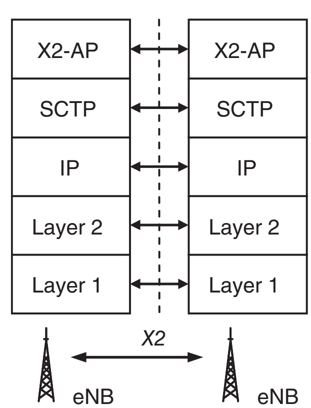

Figure 13.13 The control plane protocol stack in the case of direct communication between two eNBs.

Instead, the shared channels are utilized for signaling and data delivery. In the LTE solution, the physical channels are mapped dynamically to the resources (meaning the physical resource blocks and antenna ports) that are available at the moment. This is done with the help of the scheduler.

The physical layer communicates and handles data transmission with the higher LTE/SAE layers via the transport channels that is a block oriented service that takes into account the bit rate, delays, collisions and reliability of the transmission.

13.6.4 Layer 2

13.6.4.1 MAC

The MAC (Medium Access Control protocol) is the first one, that is, the lowest protocol in layer 2. The main functionality of MAC is related to management of transport channels. On the other hand, MAC is fed from the higher layers with the logical channels, which correspond to certain radio bearers.

MAC multiplexes the data of the logical channels onto the transmission of the transport channels, and demultiplexes it in the reception, according to the priority level of the logical channels.

MAC includes HARQ functionality (Hybrid Automatic Retransmission on reQuest). MAC also takes care of handling of collisions, and identifies the UEs.

13.6.4.2 RLC

The Radio Link Control (RLC) is next to the MAC protocol in the second LTE/SAE protocol layer structure. There is a one-to-one relationship between each Radio Bearer and each RLC instance.

RLC enhances the radio bearer quality via the ARQ (Automatic Retransmission on reQuest) by utilizing the data frames that contains sequence identities, and via the status reports in order to trigger the retransmission mechanism.

RLC also segments and reassembles the data with the higher layer data, and on the other hand, concatenates the higher layer data pieces into blocks that are suitable for the transport over the transport channels as they allow limited transport block sizes.

13.6.5 Layer 3

Layer 3 radio protocols consist of the following:

- PDCP (Packet Data Convergence Protocol)
- RRC (Radio Resource Control)
- NAS Protocols.

13.6.5.1 PDCP

Each radio bearer always uses a respective PDCP (Packet Data Convergence Protocol) instance. PDCP manages the header compression, which is called ROHC (Robust Header Compression) according to the RFC 3095. PDCP also manages ciphering and deciphering functionalities.

It should be noted that header compression is useful for IP datagram delivery, but the effect is not so significant for signaling. This means that for signaling, PDCP will usually only do the ciphering and deciphering without header compression.

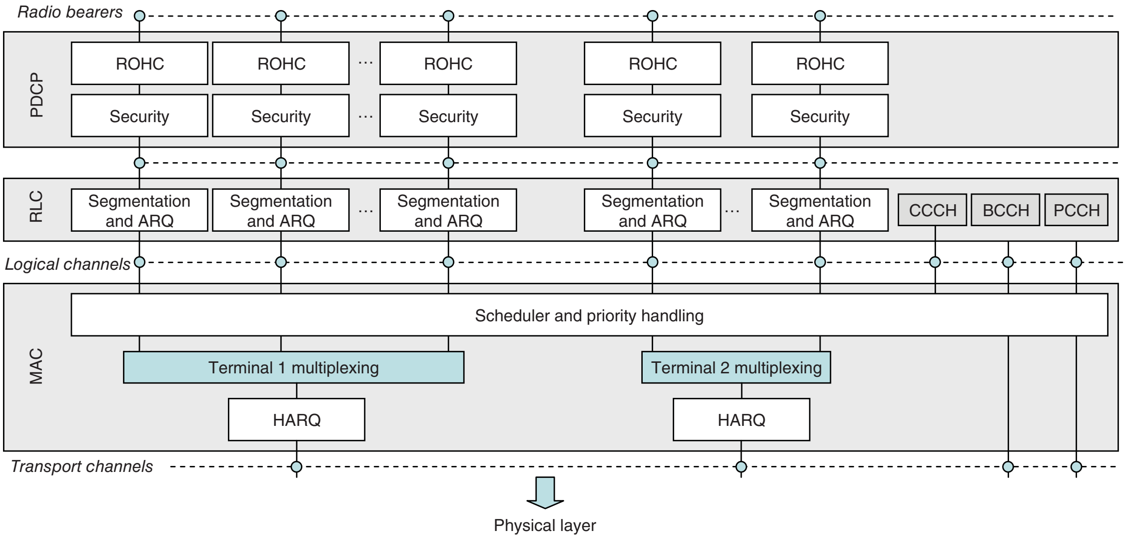

Figure 13.14 *Layer 2 structure for downlink.*

13.6.5.2 RRC

RRC (Radio Resource Control) is an access stratum specific control protocol for E-UTRAN. It provides the required messages for channel management, measurement control and reporting. The control plane of RRC is a multitask entity that takes care of, for example, broadcast and paging procedures, RRC connection setup, radio bearer control, mobility functions and LTE-UE measurement control.

RRC functionality is examined in more detail in Chapter 7.

13.6.5.3 NAS Protocols

The NAS protocol runs between UE and MME. It is located on top of RRC, which provides the required carrier messages for NAS transfer. Some of the most important tasks of NAS are authentication procedure, security control, EPS bearer management, EMC_Idle mobility handling, and paging origination in the EMC_Idle state.

13.7 Layer 2 Structure

As seen in previous chapters, the MAC layer delivers information towards the radio interface via the transport channels, and on the other hand, MAC delivers information to the RLC layer above via logical channels. RLS in turn, delivers information for PDCP functionalities above, which then contains radio bearers in top of it.

When looking in more detail at the tasks within MAC, RLC and PDCP, we can see that there are several *service access points* between the MAC and physical layers, referring to individual transport channels. The service access points between the MAC and RLC layers refer to the logical channels. It should be noted that the multiplexing of various logical channels to the same transport channels can be done. Figure 13.14 shows the downlink structure of the layer 2, and Figure 13.15 shows the uplink structure.

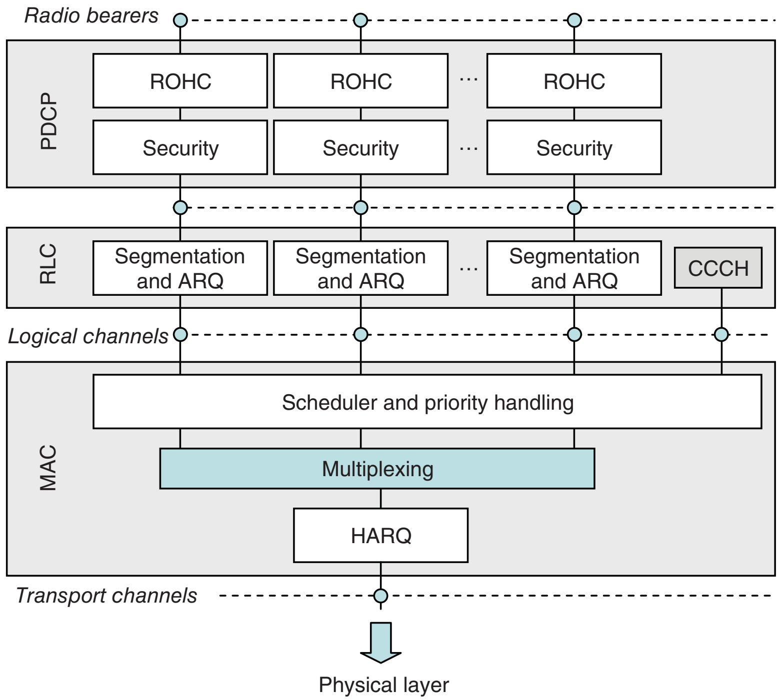

Figure 13.15 *Layer 2 uplink structure.*

13.8 LTE Radio Network

13.8.1 Introduction

This chapter presents a set of LTE radio interface related topics. First, an LTE spectrum with defined and practical bands is described. Then, a description is given of the LTE multiplex in downlink and uplink, that is, OFDM and SC-FDMA respectively. Both Frequency Division and Time Division variants of LTE are shown. LTE link budget is explained with practical examples. The hardware solutions of eNodeB and the antenna systems are explained in detail, as well as the practical points of view in the hardware reuse in LTE deployment.

13.8.1.1 LTE radio Interface

The LTE radio interface is based on the frequency division multiplexing technique. In the downlink direction, OFDMA (Orthogonal Frequency Division Multiplex) is used whereas in the uplink direction, SC-FDMA (Single Carrier Frequency Division Multiple Access) is used. OFDMA gives a good protection against the fast varying radio conditions, including the fast fading and multipath propagated radio components. Nevertheless, it is not an efficient transmitter solution because the peak-to-average power ratio (PAPR) behavior causes challenges in the equipment's circuit design. For that reason, SC-FDMA is selected in the uplink as the terminal can handle these challenges more easily.

LTE supports both FDD (Frequency Division Duplex) and TDD (Time Division Duplex). In the FDD mode, the uplink and downlink transmission happens in separate frequency bands, whereas TDD mode utilizes timeslots of the same frequency band for downlink and uplink transmission. Both these modes can be utilized efficiently in such a way that the total bands are the same, varying between 1.4 and 20 MHz. Depending on the bandwidth and other functionalities as MIMO variants and modulation scheme, the maximum data speed that LTE provides is up to about 300 Mb/s in downlink and 75 Mb/s in uplink.

Due to the flexible bandwidth definitions of LTE, it can be deployed in many scenarios of mobile operators. The smallest bands are applicable in situations where the operator does not have too many extra frequencies due to other systems of the operator already existing. Although the smallest LTE band can provide lowest data speeds and capacity, it is justified as an interim solution in the frequency refarming. In this way, the same band can be shared between GSM/UMTS and LTE. Along the growth of the LTE subscriber penetration, the 2G/3G band can be reduced whilst LTE takes bigger share of the band. As an example, there are LTE terminals available for the 900 MHz in the beginning of the service, and as the LTE matures, there will be terminals also for the other bands offering multisystem functionalities. LTE provides more flexible rearrangement of the frequencies than UMTS, which is fixed to a figure of 5 MHz (or, in some cases 3.8 or 4.2 MHz, depending on the vendor solutions).

The LTE bandwidth dictates how many subcarriers can be utilized in that band. This, on the other hand, dictates the number of radio resource blocks, that is, how much capacity can be offered in that area. A single radio resource bock corresponds to 12 consecutive subcarriers.

13.9 LTE Spectrum

There are various frequency band options available for the LTE system, depending on the country and continent. The LTE networks can be deployed in the existing and new frequency bands. The most popular bands since the launch of the LTE networks have been:

- The 1800 MHz band 3 that is currently widely utilized for the GSM system. According to Ref. [38], the Band 3 is used in 43% of tall the deployments.
- A new 2600 MHz band 7 is becoming available globally in various parts and is one of the most popular for the LTE deployments. According to Ref. [38], Band 7 is used in 27% of the deployments.
- The 700 MHz band has typically been used for analog TV broadcast networks and are refarmed due to digitalization of TV systems. The most utilized 700 MHz LTE bands are 12, 13, 14 and 17.
- The 800 MHz band 20 is increasingly popular for LTE deployments. According to Ref. [38], Band 20 is used in 13% of the deployments.
- The 850 MHz band 5 and 1900 MHz band 25 that are widely utilized for GSM in the Northern America region.
- The 2100 MHz band outside America, and the combined 1700 MHz and 2100 MHz bands in America, that are widely utilized for the previous 3G systems, that is, UMTS/WCDMA and HSPA. This AWS band 4 is also highly popular. According to Ref. [38], AWS Band 4 represents 9% of the world's deployments.

For the TDD deployments the most popular bands up to February 2014 are [38]:

- Band 40 in 2.3 GHz (15 networks)
- Band 38 in 2.6 GHz (9 networks)
- Band 41 in 2.6 GHz (5 networks)
- Band 42 in 3.5 GHz (3 networks)
- Band 39 in 1.9 GHz (1 network).

The initial LTE deployment has happened especially in FDD bands in 1800 MHz, 2600 MHz, 700 MHz, 800 MHz and combined 2100 MHz / 1700 MHz bands. Up to February 2014, there have been a total of 244 FDD-LTE operators, 17 TDD-LTE operators, and 13 operators having both FDD and TDD modes. Please note that China Mobile uses LTE-TDD bands 39, 40 and 41.

Please note that the LTE FDD band 6 is not applicable according to Ref. [39]. Also, the band 29 (former MediaFLO DL for mobile TV) is restricted to E-UTRA only with carrier aggregation, the uplink being paired with the respective carrier aggregation frequency. Please also note that in addition to the bands shown in Table 13.1, 3GPP has identified FDD frequency bands of (a) 1915–1920 (UL) and 1995–2000 (DL) as well as (b) 1755–1780 (UL) and 2155–2180 (DL). These bands have not been assigned, though, by the time of writing

Table 13.1 *FDD frequency bands for LTE. Please note that bands 15 and 16 are not used. Term "all" refers to all bandwidths 1.4, 3, 5, 10, 15 and 20 MHz*

Band nr	Band name, common	Uplink frequency (MHz)	Downlink frequency (MHz)	DL/UL Bandwidth, tot (MHz)	Bandwidths supported (MHz)
1	IMT (2100)	1920–1980	2110–2170	60	5, 10, 15, 20
2	PCS (1900)	1850–1910	1930–1990	60	all
3	DCS (1800)	1710–1785	1805–1880	75	all
4	AWS-1 (1.7/2.1 GHz)	1710–1755	2110–2155	45	all
5	CLR (850)	824–849	869–894	25	1.4, 3, 5, 10
6	UMTS, Japan (800)	830–840	875–885	10	5, 10
7	IMT Extended (2600)	2500–2570	2620–2690	70	5, 10, 15, 20
8	E-GSM (900)	880–915	925–960	35	1.4, 3, 5, 10
9	UMTS 1700 / Japan DCS (1800)	1749.9–1784.9	1844.9–1879.9	35	5, 10, 15, 20
10	Extended AWS (1.7/2.1 GHz)	1710–1770	2110–2170	60	5, 10, 15, 20
11	Lower PDC, Japan (1500)	1427.9–1447.9	1475.9–1495.9	20	5, 10
12	US lower SMH A, B, C (700)	699–716	729–746	17	1.4, 3, 5, 10
13	US upper SMH C (700)	777–787	746–756	10	5, 10
14	US upper SMH D (700)	788–798	758–768	10	5, 10
15	N/A	N/A	N/A	N/A	N/A
16	N/A	N/A	N/A	N/A	N/A
17	US, lower SMH B, C (700)	704–716	734–746	12	5, 10
18	Lower 800, Japan (800)	815–830	860–875	15	5, 10, 15
19	Upper 800, Japan (800)	830–845	875–890	15	5, 10, 15
20	EU Digital Dividend (800)	832–862	791–821	30	5, 10, 15, 20
21	Upper PDC, Japan (1500)	1447.9–1462.9	1495.9–1510.9	15	5, 10, 15
22	3500	3410–3490	3510–3590	80	5, 10, 15, 20
23	US S-band, or AWS-4 (2000)	2000–2020	2180–2200	20	1.4, 3, 5, 10
24	US L-band (1600)	1626.5–1660.5	1525–1559	35	5, 10
25	Extended PCS (1900)	1850–1915	1930–1995	65	all
26	Extended CLR (850)	814–849	859–894	80	1.4, 3, 5, 10, 15
27	SMR (850)	807–824	852–869	17	1.4, 3, 5, 10, 15
28	APAC (700)	703–748	758–803	45	5, 10, 15, 20
29	Lower SMH D, E (700), former MediaFLO DL	N/A	717–728	11	5, 10
30	WCS A, B (2300)	2305–2315	2350–2360	10	5, 10
31	450 (Brazil driven)	452.5–457.5	462.5–467.5	5	

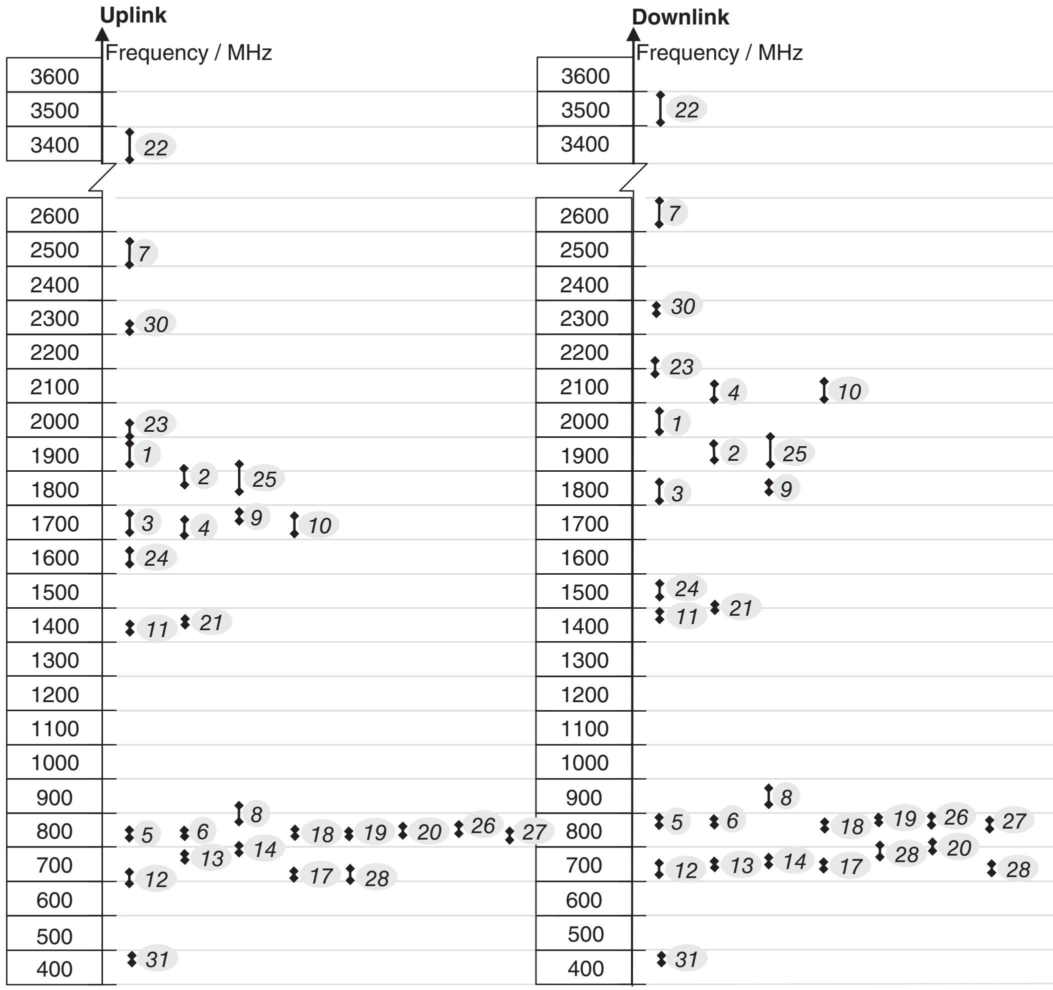

Figure 13.16 *The LTE spectrum for FDD downlink and uplink. The references are found in Table 13.1.*

this book. Please also note that the band 26 definition has varied along with the time, the latest bandwidth being 17 MHz based on the source [39].

Figure 13.16 and Figure 13.17 show the graphical presentation of the FDD and TDD frequency bands of LTE.

13.10 Physical Layer

13.10.1 Principles of OFDMA and SC-FDMA

LTE utilizes OFDM (Orthogonal Frequency Division Multicarrier) in downlink, that is, in the direction from the eNodeB to the UE. This direction is sometimes also referred as a forward link. OFDM complies with LTE requirements for spectrum flexibility and enables a cost-efficient base for wide frequency bands that

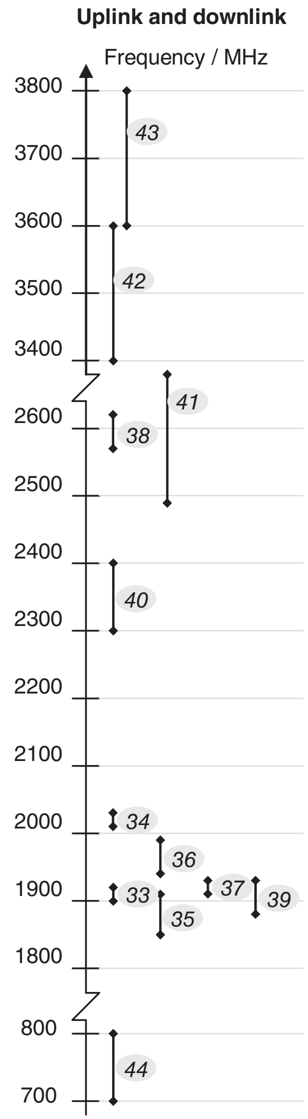

Figure 13.17 *The frequency bands of TDD LTE. The references are found in Table 13.2.*

provide high peak data rates. The LTE downlink physical resource can be seen as a time-frequency grid. In the frequency domain, the spacing between the adjacent subcarriers, Δf, is 15 kHz. In addition, the OFDM symbol duration time is 1/Δf + cyclic prefix. The cyclic prefix is used to maintain orthogonality between the subcarriers even for a time-dispersive radio channel. One resource element carries QPSK, 16QAM or 64QAM.

Opposite to WCDMA, LTE uses OFDMA as the access method in downlink what enables it to scale the occupied bandwidth from 1.4 MHz to 20 MHz. The Orthogonal Frequency Division Method (OFDM) employs a large number of narrow subcarriers for multicarrier transmission (see Figure 13.18).

Table 13.2 TDD frequency bands for LTE

Band nr	Band name, common	Uplink frequency (MHz)	Downlink frequency (MHz)	DL and UL Bandwidth (MHz)	Bandwidths supported (MHz)
33	Lower TDD 2000 (2000)	1900–1920	1900–1920	20	5, 10, 15, 20
34	Upper TDD 2000 (2000)	2010–2025	2010–2025	20	5, 10, 15
35	Lower PCS 1900 (1900)	1850–1910	1850–1910	60	all
36	Upper PCS 1900 (1900)	1930–1990	1930–1990	60	all
37	PCS duplex spacing (1900)	1910–1930	1910–1930	20	5, 10, 15, 20
38	IMT-E (2600)	2570–2620	2570–2620	50	5, 10, 15, 20
39	TDD, China (1900)	1880–1920	1880–1920	40	5, 10, 15, 20
40	2300	2300–2400	2300–2400	100	5, 10, 15, 20
41	US BRS / EBS (2600)	2496–2690	2496–2690	194	5, 10, 15, 20
42	3500	3400–3600	3400–3600	200	5, 10, 15, 20
43	3700	3600–3800	3600–3800	200	5, 10, 15, 20
44	APAC (700)	703–803	703–803	100	5, 10, 15, 20

Figure 13.18 shows the basic principle of the difference between earlier 3G bandwidth, which is fixed to 5 MHz, and the flexible LTE bandwidth. The dynamic definition of the bandwidth is actually one of the main benefits of LTE over WCDMA and HSPA. A smaller band allows the efficient frequency band refarming between LTE and other systems, for example, WCDMA and GSM, which is beneficial especially in cases when there is not too much band in use. On the other hand, the largest LTE bandwidths provide the highest data rates, which is the main differentiator compared to the WCDMA and HSPA data rates within their fixed 5 MHz band.

Orthogonal Frequency-Division Multiplexing (OFDM) is a modulation technique for data transmission which has been known since the 1960s. Nowadays, OFDM is used in many standards such as European Digital Audio Broadcasting (DAB), Terrestrial Digital Video Broadcasting (DVB-T), Asynchronous Digital Subscriber Line (ADSL) and High-bit-rate Digital Subscriber Line (HDSL). It can also be found in the IEEE 802.11a Local Area Network (Wi-Fi) and the IEEE 802.16 Metropolitan Area Network (WiMAX). Furthermore, it has been defined as the medium access technology for downlink in Long Term Evolution (LTE), chosen among other candidates such as Multicarrier Wideband Code-Division Multiple Access (MC-WCDMA) and Multi-Carrier Time-Division Synchronous-Code-Division Multiple Access (MC-TD-SCDMA).

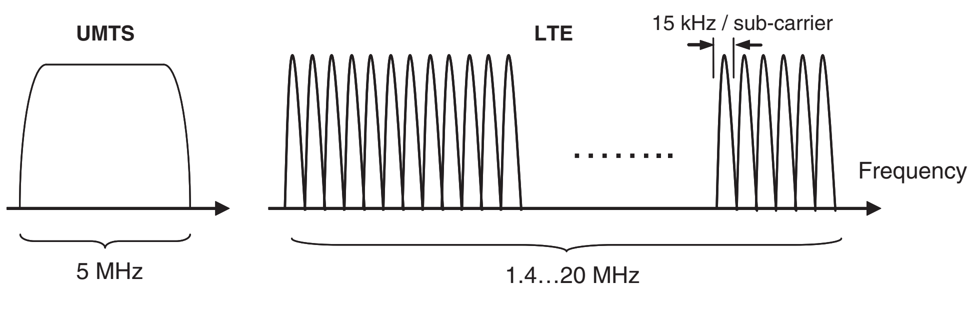

Figure 13.18 *The frequency band of LTE consists of several subcarriers whilst UMTS utilizes one complete carrier for all the traffic of a single cell.*

OFDM is based on splitting the data stream to be transmitted onto several orthogonal subcarriers and thereby allowing an increased symbol period. Since low-rate modulation schemes are more robust to multipath, it is more effective to transmit many low-rate streams in parallel than one single high-rate stream. The goal of using these subcarriers is to obtain a channel that is roughly constant (flat) over each given sub-band, making equalization much simpler at the receiver. Furthermore, OFDM allows the use of low-complexity Multiple-Input Multiple-Output (MIMO) techniques. Finally, OFDM provides a flexible use of the bandwidth and can achieve high peak data rates.

OFDM is based on the well-known Frequency Division Multiplexing (FDM) technique. In FDM different streams of information are mapped onto separate parallel frequency channels. OFDM differs from traditional FDM in the following ways:

– The same information stream is mapped onto a large number of narrowband subcarriers increasing the symbol period compared to single carrier schemes.
– The subcarriers are orthogonal to each other in order to reduce the Intercarrier Interference (ICI). Moreover, overlap between subcarriers is allowed to provide high spectral efficiency.
– A guard interval, often called *Cyclic Prefix* (CP), is added at the beginning of each OFDM symbol to preserve orthogonality between subcarriers and eliminate Intersymbol Interference (ISI) and Intercarrier Interference (ICI).

Figure 13.19 depicts the above introduced concepts. In the frequency-domain the overlap between subcarriers, plotted with different colors, is easily seen as well as the fact that they are orthogonal to each other. On the other side, in the time-domain, the presence of the guard interval at the beginning of each OFDM symbol has to be pointed out.

The orthogonality of the subcarriers is thus assured thanks to the single subcarrier bandwidth and subcarrier spacing in the frequency domain being equal (15 kHz). Because of this, while power of the given subcarrier reaches its maximum, the power of all other subcarriers is equal to zero (see Figure 13.19). Another benefit from using the OFDM is the opportunity to select a subset of subcarriers based on the channel conditions, thus

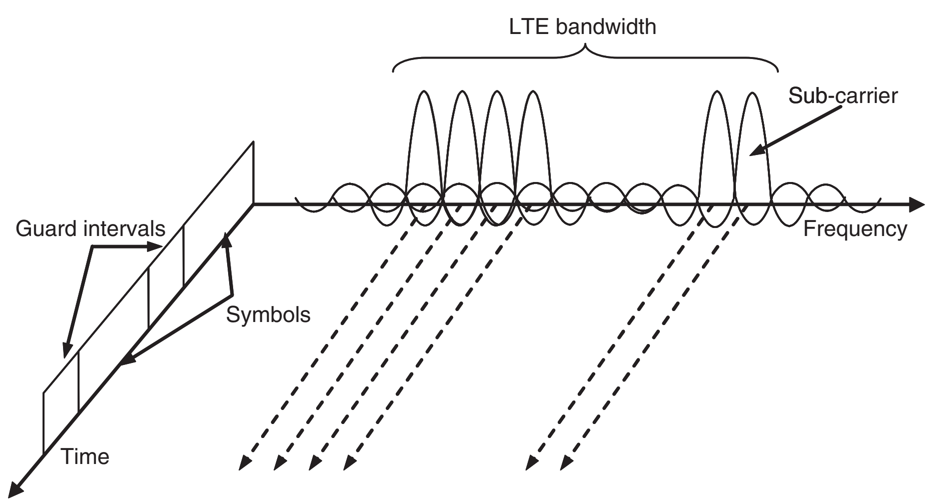

Figure 13.19 *Frequency-time interpretation of an OFDM signal.*

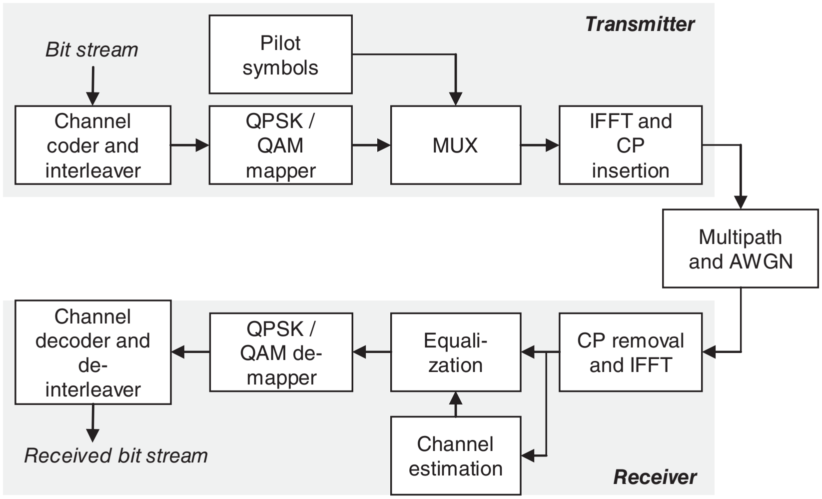

Figure 13.20 *SISO OFDM simplified block diagram.*

avoiding those with large fading. In order to eliminate time spreading guard intervals with cyclic repetition are introduced between groups of symbols.

13.10.2 OFDM Transceiver Chain

Figure 13.20 presents a simplified block diagram of a single-input single-output (SISO) OFDM system. On the transmitter side, the modulated (QAM/PSK) symbols are mapped onto N orthogonal subcarriers. This is accomplished by means of an Inverse Discrete Fourier Transform (IDFT) operation. Most commonly, the IDFT is performed with an Inverse Fast Fourier Transform (IFFT) algorithm which is computationally efficient. Next, the CP is inserted and a parallel-to-serial conversion is performed prior to transmission over the air.

At the receiver end, the reversal operations are performed. Once the received signal reaches the receiver, the CP which is potentially interfered by previous OFDM symbols is removed. Then, a Fast Fourier Transform (FFT) operation brings data to the frequency domain. This way, channel estimation and equalization is simplified. Note that in order to be able to carry out the latter operations, known symbols called pilots are to be inserted in certain frequency positions/subcarriers at the transmitter side. At the end of the chain, the equalized data symbols are demodulated yielding the received bit stream.

13.10.3 Cyclic Prefix

As mentioned before, a guard interval is added at the beginning of each OFDM symbol to mitigate some of the negative effects of the multipath channel. If the duration of the guard interval T_g is larger than the maximum delay of the channel $\tau_{\max}$, all multipath components will arrive within this guard time and the useful symbol will not be affected avoiding Intersymbol Interference (ISI) as can be seen in Figure 13.21.

One particularization of the guard interval is the so-called cyclic prefix. In this case the last N_g samples of the useful OFDM symbol with N samples in total are copied to the beginning of the same symbol. Since the number of cycles of each orthogonality function per OFDM symbol will be maintained as an integer,

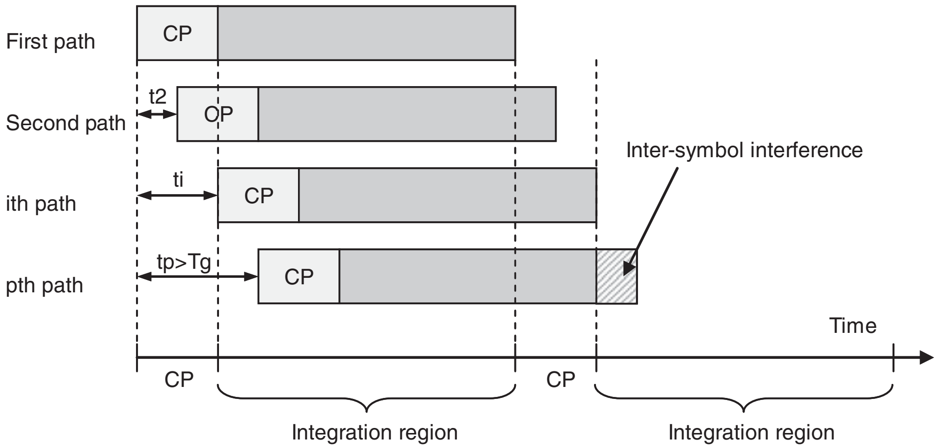

Figure 13.21 *Cyclic prefix (CP) avoiding ISI.*

this strategy also allows the orthogonality properties of the transmitted subcarriers to be kept, avoiding ICI. Figure 13.22 shows the cyclic prefix concept where

$$T_u = N \times T_0 \tag{13.1}$$

$$T_g = N_g \times T_0 \tag{13.2}$$

$$T_s = \left(N + N_g\right) \times T_0. \tag{13.3}$$

The symbols means the following: T_u is the useful OFDM symbol where data symbols are allocated, T_g is the duration of the cyclic prefix and T_s is the total duration of the OFDM symbol.

The insertion of CP results in a not so important Spectral Efficiency Loss (SEL) compared to benefits provided in terms of ISI and ICI robustness. The SEL can be interpreted as the loss of throughput that the OFDM transmission system will suffer by the addition of the cyclic prefix. It can be written as:

$$SEL = \frac{T_g}{T_g + T_u}. \tag{13.4}$$

It can be seen that the loss of spectral efficiency is directly related to the ratio between the duration of the CP and the total duration of an OFDM symbol.

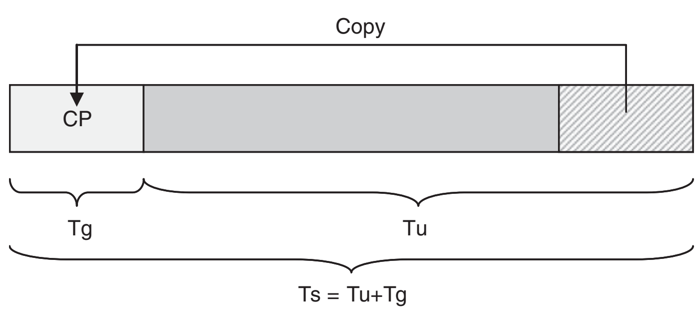

Figure 13.22 *Cyclic prefix as a copy of the last part of an OFDM symbol.*

13.10.4 Channel Estimation and Equalization

In wireless OFDM systems, the received symbols have been corrupted by the multipath channel. In order to undo these effects, an equalization of the received signal that somehow compensates for the variations introduced by the channel must be performed.

Assuming that the CP is longer than the maximum delay of the channel and a nonvariant channel over the duration of an OFDM symbol (slow-fading channel), each subcarrier symbol is multiplied by a complex number equal to the channel transfer function coefficient at this subcarrier frequency.

In other words, each subcarrier experiences a complex gain due to the channel. In order to undo these effects a single complex multiplication is required for each subcarrier yielding low complexity equalization in the frequency domain:

$$y[k] = \frac{z[k]}{h[k]} = d[k] + \frac{w[k]}{h[k]} \tag{13.5}$$

where $y[k]$ is the equalized symbol in the k^{th} subcarrier, $z[k]$ is the received symbol at the k^{th} subcarrier after FFT and $h[k]$ is the complex channel gain at subcarrier k. $w[k]$ represents the additive white Gaussian noise at subcarrier k.

Note that this equalization has been performed assuming a perfect knowledge about the channel. However, in most systems that employ equalizers, the channel properties are unknown a priori. Therefore, the equalizer needs a channel estimator that provides the equalization block the required information about the channel characteristics.

Different approaches have been proposed to estimate the channel in OFDM systems but *pilot-aided channel estimation* is the most suitable solution for the mobile radio channel. Furthermore, in LTE it is the proposed solution. This technique consists in transmitting symbols, often called pilot symbols, known by both the transmitter and receiver in order to estimate the channel at the receiver. This approach presents an important tradeoff between the number of pilots used to perform the estimation and transmission efficiency. The more pilots are used, the more accurate the estimation will be, but also the more overhead will be transmitted reducing the data rate.

Figures 13.23–13.26 depict the mapping of cell-specific reference signals in LTE or different number of antenna ports and with normal CP. These pilot symbols are distributed in both frequency- and time-domain and they are orthogonal to each other in order to allow for accurate channel estimation.

Figure 13.23 shows the idea of the LTE radio resource block, and following figures show the mapping of the reference signals.

13.10.5 Modulation

LTE can use QPSK, 16-QAM and 64-QAM modulation schemes as shown in Figure 13.27 and Figure 13.28. The channel estimation of OFDM is usually done with the aid of pilot symbols. The channel type for each individual OFDM subcarrier corresponds to the flat fading. The pilot-symbol assisted modulation on flat fading channels involves the sparse insertion of known pilot symbols in a stream of data symbols.

The QPSK modulation provides the largest coverage areas but with the lowest capacity per bandwidth. 64-QAM results in smaller coverage, but it offers more capacity.

13.10.6 Coding

LTE uses Turbo coding or convolutional coding, the former being more modern providing in general about 3 dB gain over the older and less effective, but at the same time more robust convolutional coding.

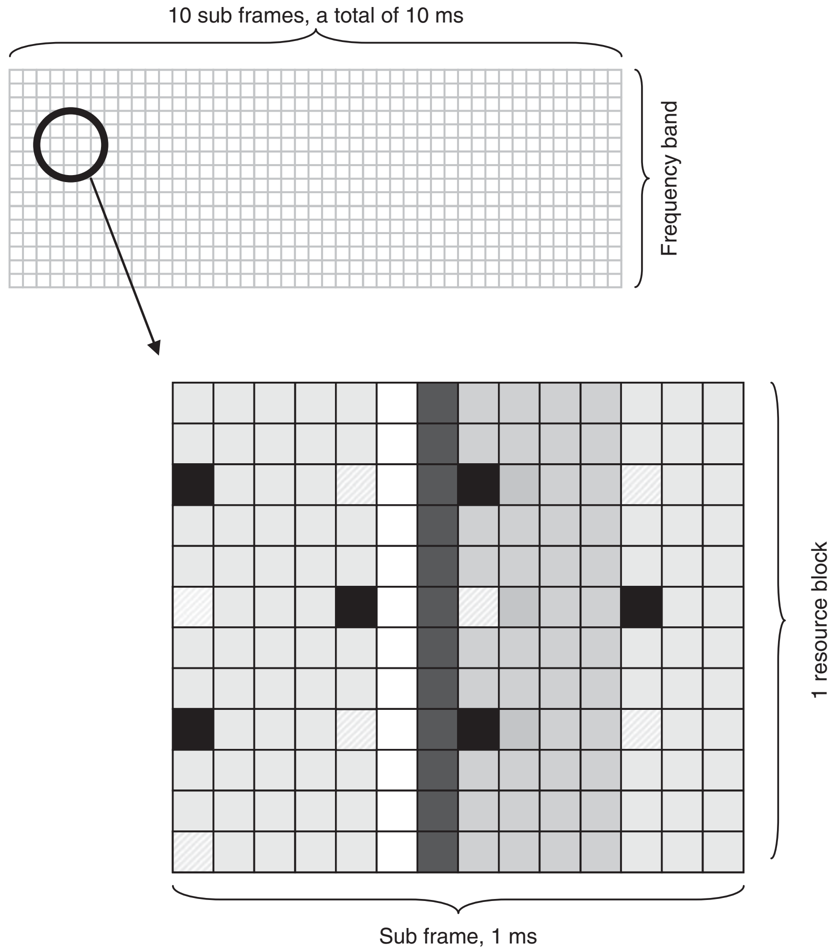

Figure 13.23 *The forming of the LTE radio resource block.*

The creation of the OFDM signal is based on the Inverse Fast Fourier Transform (IFFT), which is the practical version of the Discrete Fourier Transform (DTF) and relatively easy to be deployed as there are standard components for the transform calculation. The reception utilizes, on the other side, the FFT for combining the original signal.

13.10.7 Signal Processing Chain

After the coding and modulation of the user data, the OFDM signal is formed by applying serial-to-parallel conversion. This is an essential step in order to feed the IFFT process. Before bringing the parallel subcarriers of the user data, the subcarrier mapping also takes the needed amount of parallel subcarriers from other users, that is, ODFMA is applied. All these streams are fed into the IFFT input in order to do the Inversed Discrete Fourier Transform in a practical way. It is important to note that the process from the serial symbol stream to S/P conversion, subcarrier mapping process and N-point IFFT process happens in the frequency domain, whereas the process from the IFFT conversion happens in the time domain.

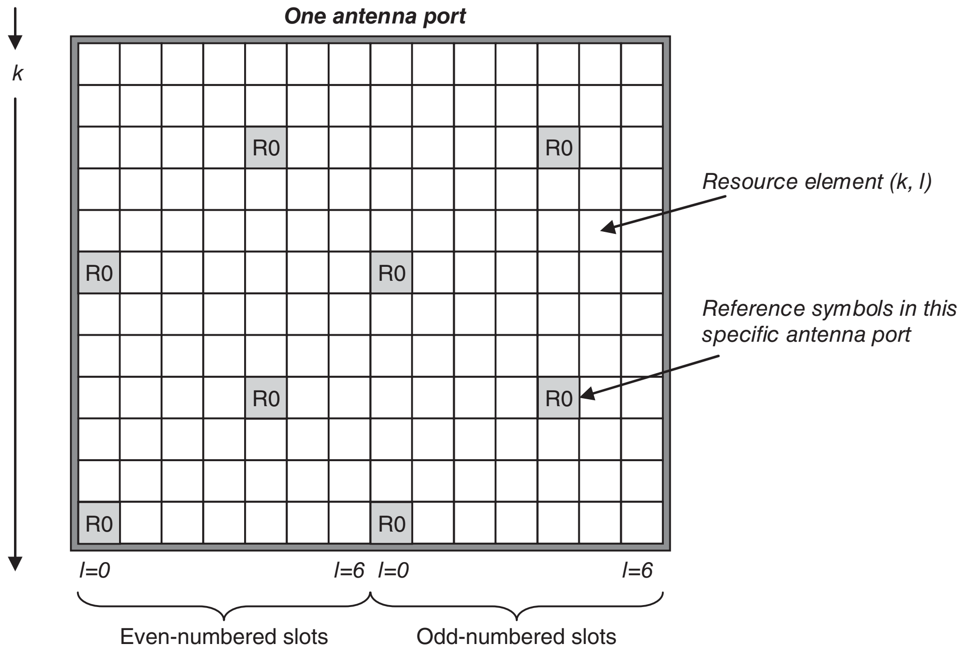

Figure 13.24 *Mapping of downlink cell-specific reference signals in LTE with normal CP, that is, in one antenna port setup of LTE.*

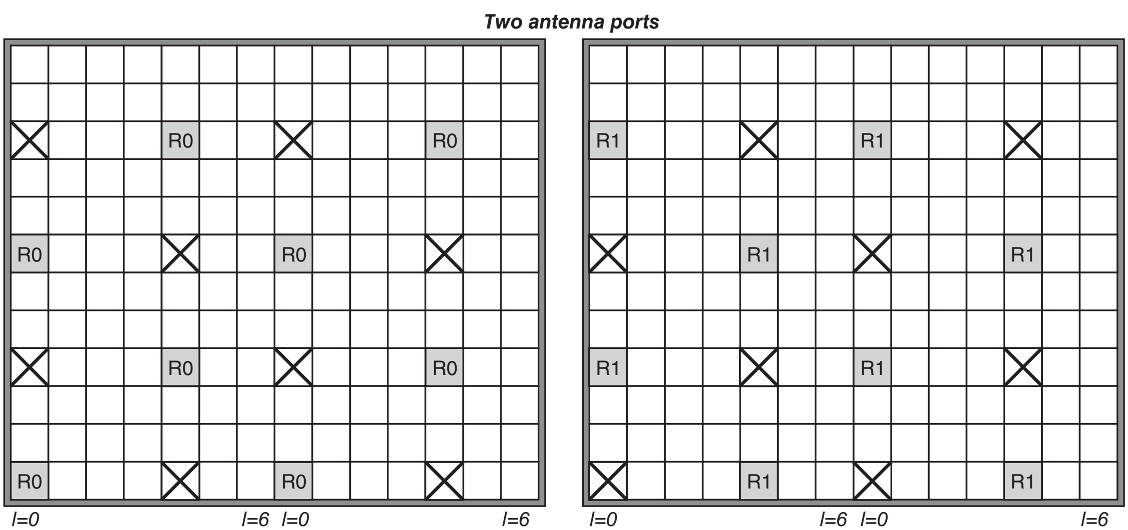

Figure 13.25 *Two-port MIMO in LTE. The cross indicates the resource elements that are not used in the respective antenna port.*

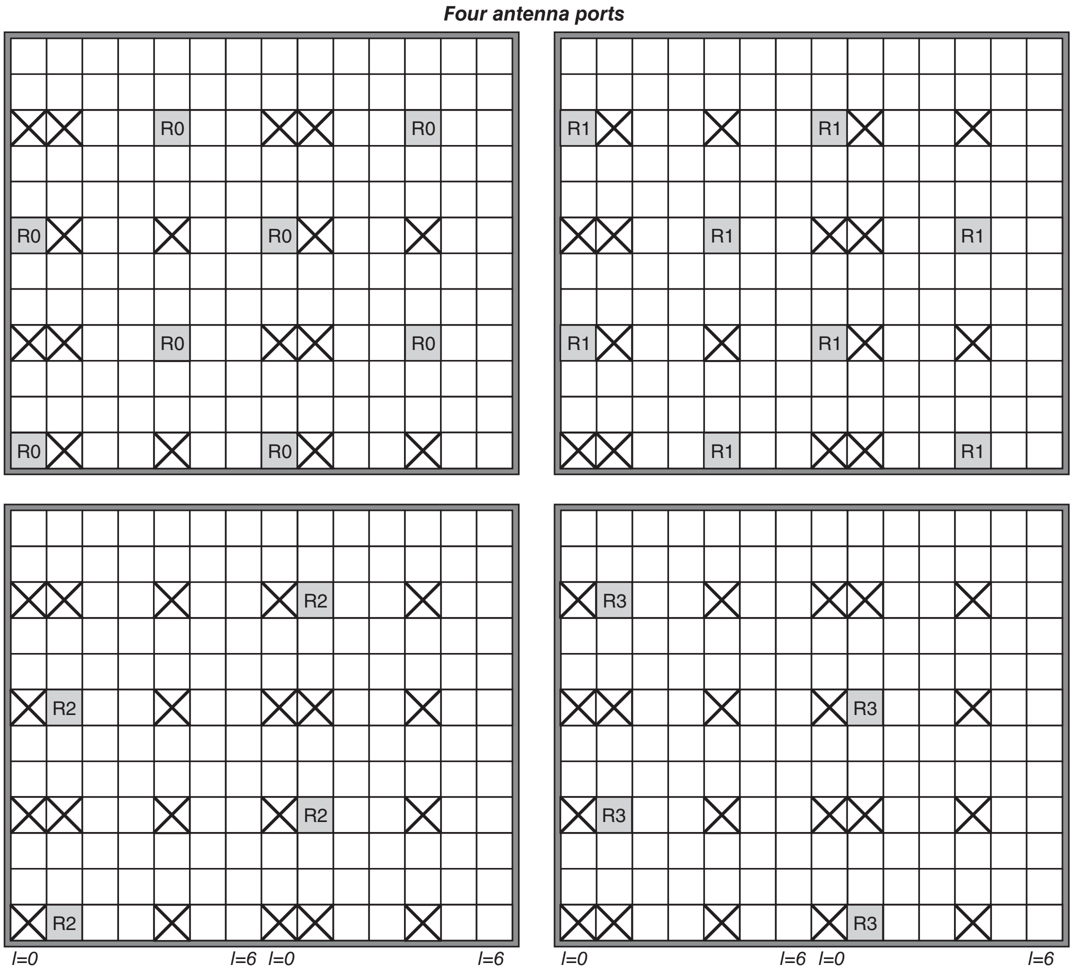

Figure 13.26 Four antenna port setup.

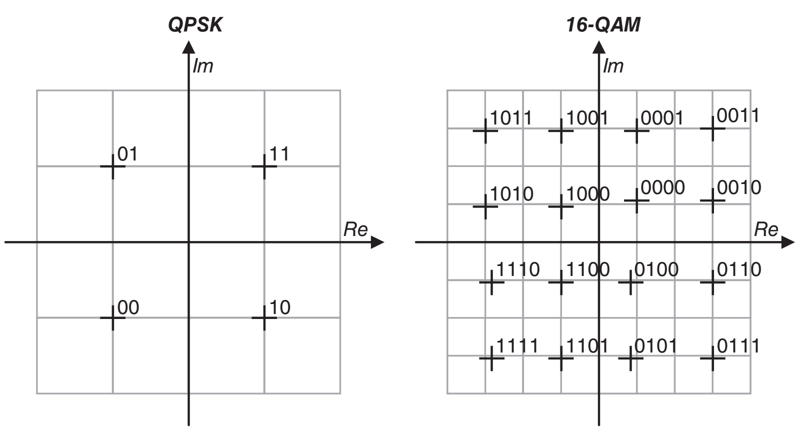

Figure 13.27 The I/Q constellation of the QPSK and 16-QAM modulation schemes.

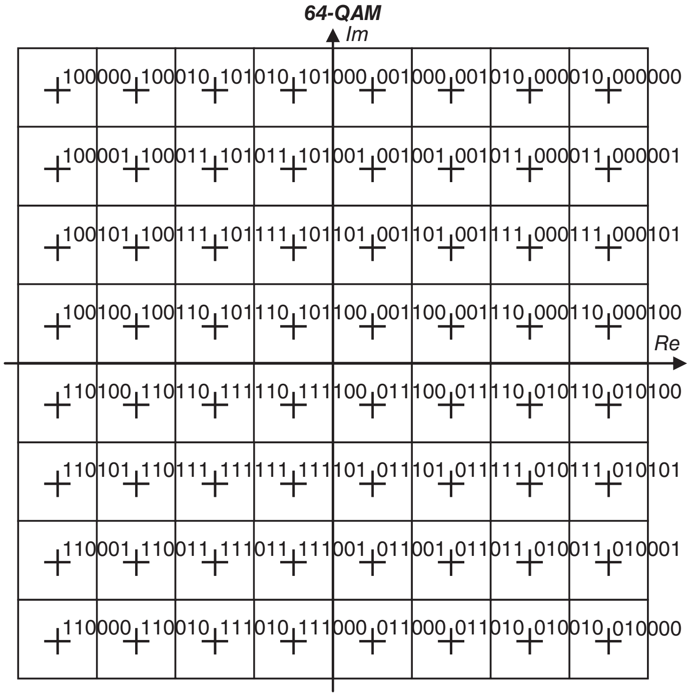

Figure 13.28 *The I/Q constellation of the 64-QAM modulation scheme.*

The OFDM symbols are formed by adding the cyclic prefix in the beginning of the symbols in order to protect the signal against multipath propagated components. Then, the windowing, digital to analog conversation, frequency up-conversion, RF processing and finally the actual radio transmission are performed in the transmitter of the eNodeB. The OFDM transmission is only used in the downlink, so the LTE-UE does have the OFDM receiver and SC-FDMA transmitter.

13.11 SC-FDM and SC-FDMA

Single-carrier frequency-division multiplexing (SC-FDM), sometimes referred to as discrete Fourier transform (DFT)-spread OFDM, is a modulation technique that, as its name indicates, shares the same principles as OFDM. Therefore, its same benefits in terms of multipath mitigation and low-complexity equalization are achievable.

The difference though is that a discrete Fourier transform (DFT) is performed prior to the IFFT operation at the transmitter side, which spreads the data symbols over all the subcarriers carrying information and produces a virtual single-carrier structure. Figure 13.29 shows the principle of the SC-FDMA transmission.

As a consequence, SC-FDM presents a lower peak-to-average-power ratio (PAPR) than OFDM. This property makes SC-FDM attractive for uplink transmissions, as the user equipment (UE) benefits in terms of transmitted power efficiency.

Furthermore, DFT spreading allows the frequency selectivity of the channel to be exploited, since all symbols are transmitted in all the subcarriers. Therefore, if some subcarriers are in deep fade, the information can still be recovered from other subcarriers experiencing better channel conditions. On the other hand, when DFT despreading is performed at the receiver, the noise is spread over all the subcarriers and generates an

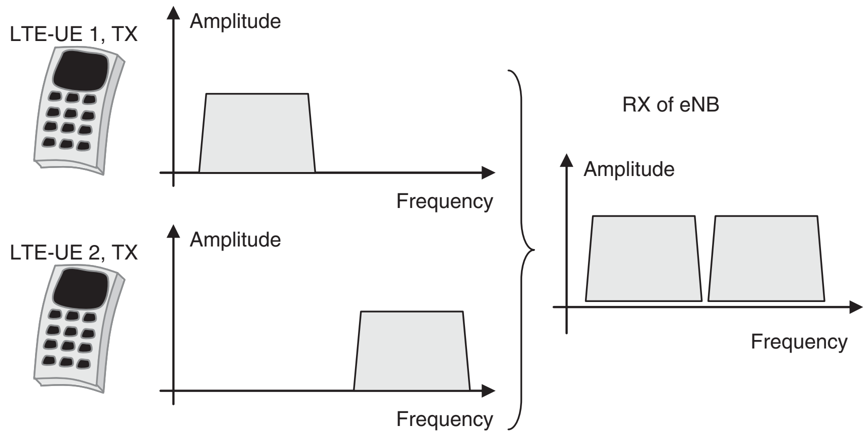

Figure 13.29 *The principle of the SC-FDMA transmission.*

effect called noise enhancement, which degrades the SC-FDM performance and requires the use of a more complex equalization based on a minimum mean square error (MMSE) receiver.

The single carrier frequency division multiplexing (SC-FDM) is thus a variation of the OFDM, and its modulation scheme follows the same steps as for OFDM with the addition of an inverse FFT. The iFFT converts the frequency signal, shifted to the desired subcarriers, to time domain. The fact that the constellation mapping is being done in time domain results in a single carrier transmit signal. The advantage of such solution is the reduced peak-to-average ratio (PAPR) of the signal compared to plain OFDM.

13.11.1 SC-FDM Transceiver Chain

Figure 13.30 presents the block diagram of a SISO SC-FDM system. It can be seen that the main difference compared to the OFDM diagram is the FFT/IFFT block which spreads the data symbols over all the subcarriers prior to the IFFT operation. The rest of the blocks remain the same as in OFDM.

13.11.2 PAPR Benefits

As mentioned before, OFDM shows large envelope variations of the transmitted signal. The different subcarriers carrying parallel data could constructively add in phase leading to instantaneous peak power compared to the average. Signals with a high PAPR require highly linear power amplifiers in order to avoid excessive intermodulation distortion. Therefore, the power amplifiers have to be operated with a large back-off from their peak value. This eventually translates into low power efficiency. This effect is particularly critical for uplink transmissions at the UE side.

Since SC-FDM is spreading the data symbols over all the subcarriers, then an averaging effect is achieved and thus transmission peaks are diminished resulting in a lower PAPR (see Figure 13.31).

13.12 Frame Structure and Physical Channels

The LTE frame structure differs between FDD and TDD version of LTE. The basic FDD LTE frame structure, where downlink and uplink are separated in frequency, comprises 20, 0.5 ms, slots with the total length of 10 ms (see Figure 14.5). Each consecutive pair of slots forms a TTI – a subframe of 1ms which is the lowest schedulable time interval. For TDD, where downlink and uplink are separated in time, a subframe is allocated

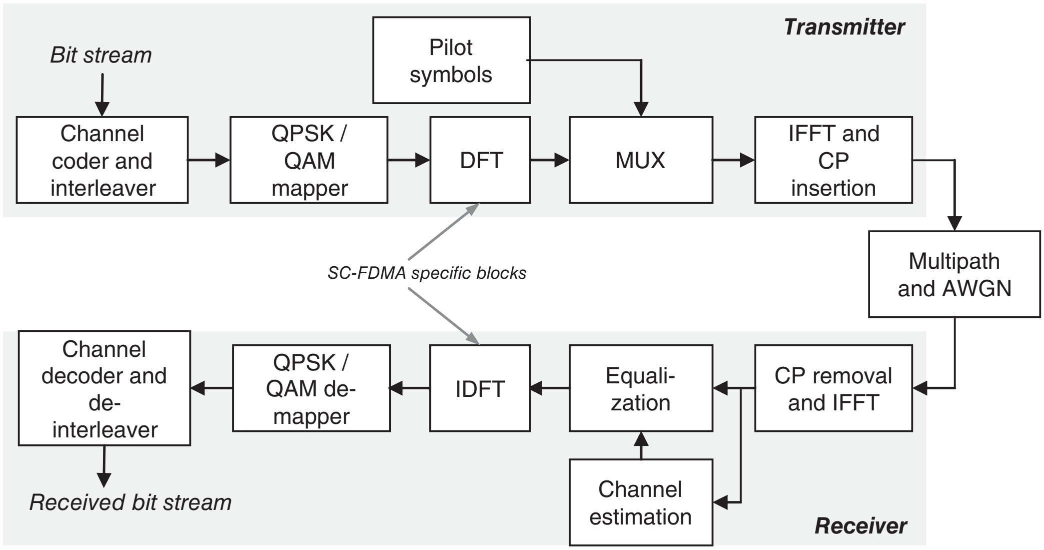

Figure 13.30 SISO SC-FDM simplified block diagram.

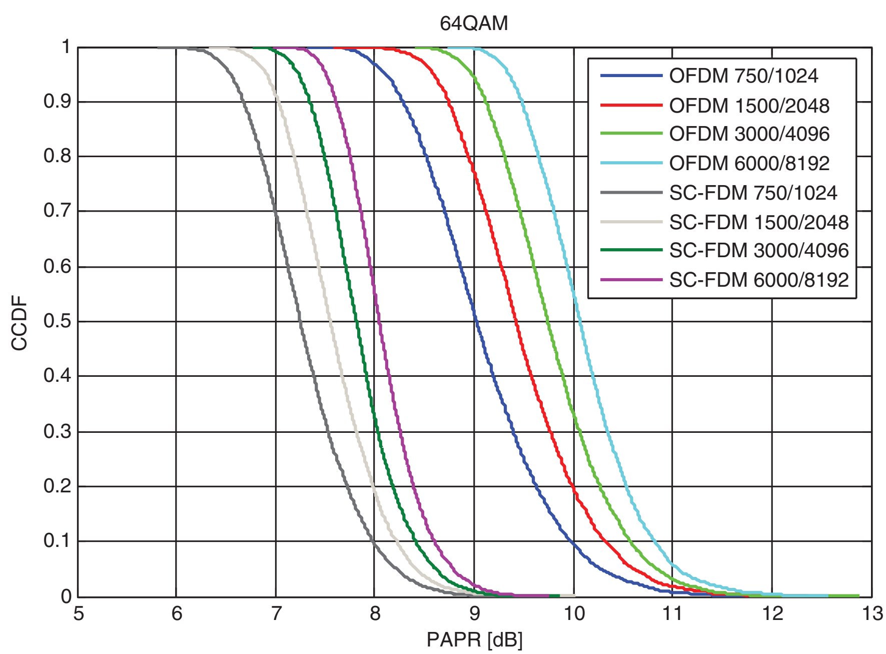

Figure 13.31 PAPR for 64QAM modulation and different OFDM/SC-FDM bandwidths [2]. Figure by Luis Maestro.

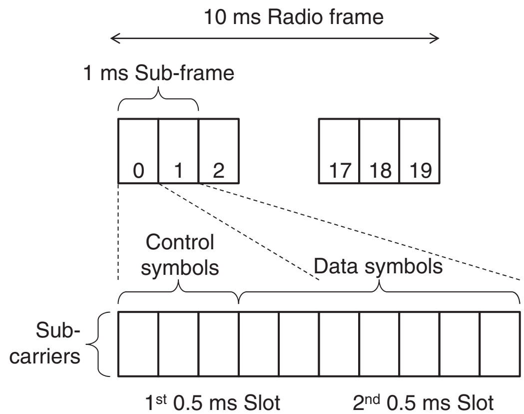

Figure 13.32 *LTE frame structure.*

either for downlink or uplink. It is worth mention that subframes no. 0 and 5 are always allocated for downlink transmission. Figure 13.32 shows the LTE frame structure.

13.12.1 Downlink

Downlink resource allocation, as indicated in Figure 13.33, can be done in noncontinuous blocks of 180 kHz bandwidth (Physical Resource Blocks – PRBs). The downlink data payload is reduced by control information, reference signals, synchronization signals and broadcast data.

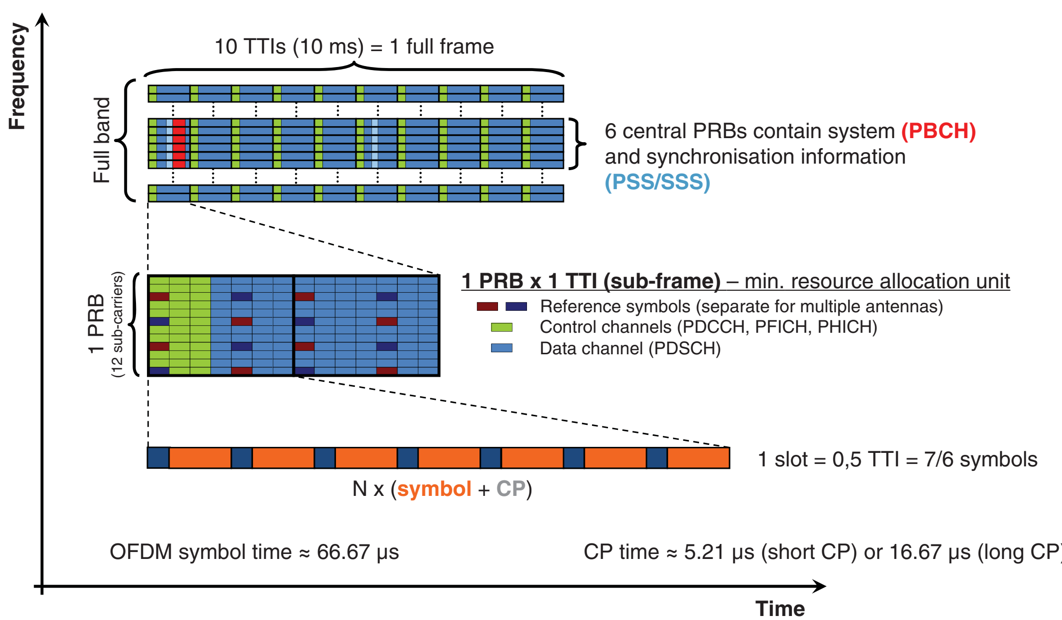

Figure 13.33 *DL frame structure.*

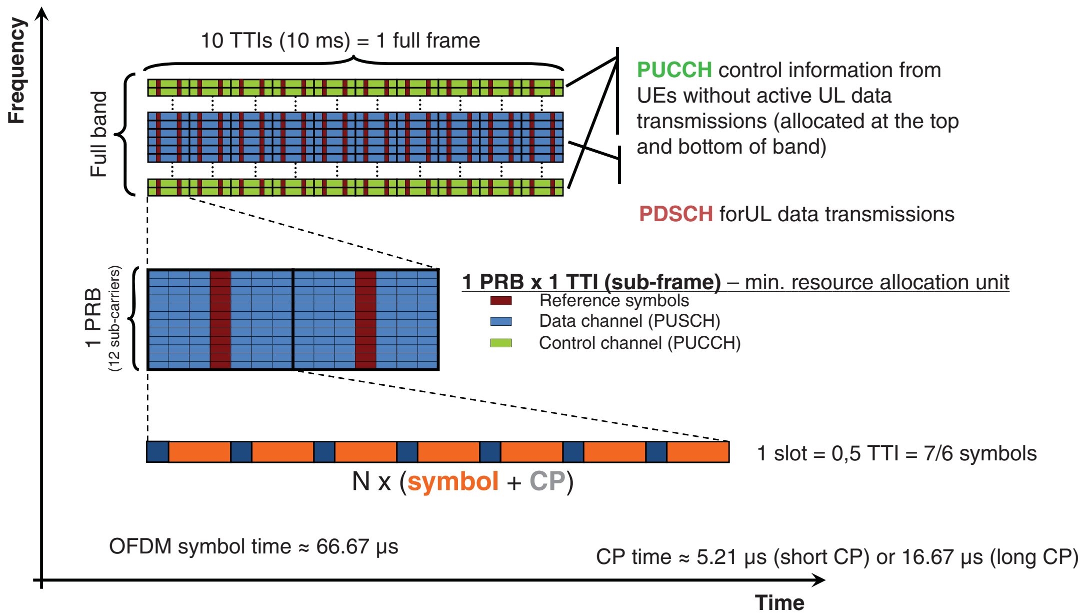

Figure 13.34 *UL frame structure of LTE.*

Downlink Physical Channels:

- *Physical Data Shared Channel (PDSCH)* – Carries downlink data
- *Physical Control Format Indicator Channel (PCFICH)* – Indicates the number of OFDMA symbols for control information. Based on this channel the UE is able to determine which symbols should be treated as control information
- *Physical Downlink Control Channel (PDCCH)* – Carries information on both UL and DL resource allocations for a given UE, UL power control commands, modulation and coding scheme used in downlink transmission, HARQ process number and so on.
- *Physical HARQ Indicator Channel (PHICH)* – Provides uplink HARQ feedback (ACK/NACK)
- *Physical Broadcast Channel (PBCH)* – Carries the system information needed to access the system (e.g. RACH parameters)
- *Synchronization Signals (PSS/SSS)* – Consist of Primary (PSS) and Secondary (SSS) synchronization signal which enable cell synchronization and identification.

13.12.2 Uplink

Since uplink transmission, as indicated in Figure 13.34, utilizes SC-FDMA the resource (bandwidth) allocation can be only in continuous PRBs. The actual data payload is reduced by control information, reference signals and cyclic prefix.

A UE may use the following uplink physical channels and signals:

- *Physical Random Access (PRACH)* – Enables establishment of a network connection or request resources for uplink transmission

- *Physical Uplink Shared Channel (PUSCH)* – Carries the uplink data and control signals if the UE has been scheduled for data transmission for example, data transport format, HARQ information, downlink CQI (Channel Quality Indicator) and so on.
- *Physical Uplink Control Channel (PUCCH)* – Carries the downlink HARQ feedback (ACK/NACK), downlink CQI if there is no UL data transmission or a Scheduling Request Indicator (SRI).
- *Uplink Reference Signals (RS)* – Send by the UE in order to enable demodulation (DM) of PUCCH or PUSCH; Send by the UE as sounding reference signals (SRS) in order to enable channel quality estimation in the eNodeB. This is one of the pre-requisites of a channel aware frequency scheduling.

13.13 Physical Layer Procedures

Main LTE physical layer procedures are random access, timing advance, power control and HARQ.

13.13.1 Random Access

The Random access procedure can be used for:

- transition from idle to connected mode,
- timing advance,
- initial resource assignment for uplink transmission.

The random access procedure begins with a preamble (PRACH) transmission. In order to minimize the probability of collisions (random access procedures from two UEs with the same RACH configuration) LTE provides up to 64 different RACH configurations which the UE can select. Once UE selects a RACH configuration a preamble is send. In case of no response UE transmits the preamble again with higher transmission power.

The eNodeB sends the Random Access Response with Timing Advance Info, UL grant and a temporary C-RNTI upon preamble reception. The UE uses the grant to send UL messages for example, connection request and indicate that it requires more resources. The eNodeB sends the Contention Resolution message which contains the C-RNTI used for future transmissions.

13.13.2 Timing Advance

The timing control procedure is essential to maintain uplink transmission orthogonality between users. The initial timing advance is set base on the random access procedure described above. This includes the UE to transmit a random access preamble base on which the eNodeB estimates the uplink timing. Based on this estimate, eNodeB sends the timing advance command using the preamble response message (see figure). The accuracy of the timing advance is up to 0.52 ms and the maximum value supports cells up to 100 km radius. The accuracy is sufficient to set the timing advance within the cyclic prefix and maintain the uplink transmission orthogonality. Figures 13.35–13.36 show the random access procedure.

13.13.3 Power Control

LTE power control controls uplink power spectral density which means that the total uplink transmit power scales linearly with bandwidth used for transmission as seen in Figure 13.37. The fractional power control can compensate for all or only part of path loss thus allows a tradeoff between intracell fairness and intercell interferences. The correction factor is provided by the eNodeB. Additionally the eNodeB can fine tune the

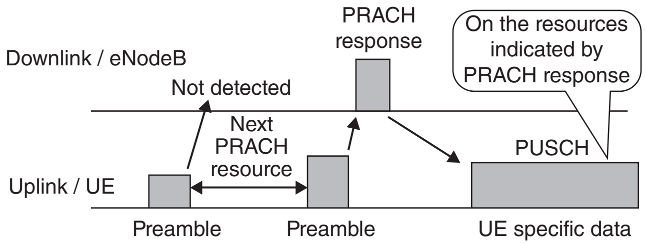

Figure 13.35 *Random Access procedure of LTE.*

UE transmit power using closed loop power control commands. Separate power control is used for PUCCH and PUSCH channels.

13.13.4 HARQ – Hybrid Automatic Repeat Request

The HARQ used in LTE is a "stop and wait" algorithm with forward error correction and ARQ error control, as shown in Figure 13.38. It uses 8 HARQ process for both downlink and uplink – asynchronous in downlink and synchronous in uplink. In case of retransmission (negative acknowledgement – NACK), the original transmission is combined with the retransmission and decoded again. This enables the network to operate at higher initial BLER point thus reducing the transmit power requirement or the coding rate (increase the data payload). The LTE HARQ supports two combining methods:

- soft combining and
- incremental redundancy.

Soft combining enforces identical as original transmission parameters whereas incremental redundancy enables use of a different set of coded bits than the previous transmission which gains the receiver extra information with each retransmission.

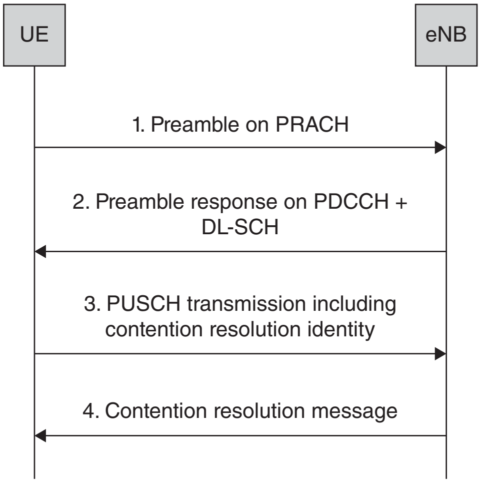

Figure 13.36 *Flow chart of random access procedure.*

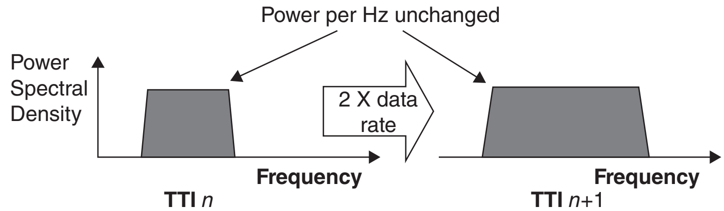

Figure 13.37 *Power control of LTE.*

13.14 User Mobility

User mobility in LTE is covered among others by tracing area update, cell reselection and handover.

13.14.1 Tracking Area Update

Tracking Area, as shown in Figure 13.39, is the same concept as Location Area in CS mobile networks (UMTS) or Routing Area in GPRS network: it is a set of cells. UEs position in the network is known by MME on TA level when UE is in idle state. Tracking Area Identifier (TAI) is used to identify a Traffic Area uniquely in a network. The traffic area update procedure is initiated whenever the UE moves between cells in different traffic areas. The setting of a correct size traffic area is important step in network planning phase as a large TA means less Tracking Area Updates (TAUs), and small TA reduces the paging signaling load for incoming packet calls.

13.14.2 Handover

The LTE handover, as shown in Figure 13.40, is the basic mobility in connected state. It is network-controlled – the target cell is selected by serving node, UE assisted – based on measurements made by the UE and lossless – although it is a hard handover no packets are lost thanks to packet forwarding on the X2 interface. Additionally, LTE utilizes a late path switching concept for its handovers. This means that the core network S1 connection is updated only once the radio path is fully changed.

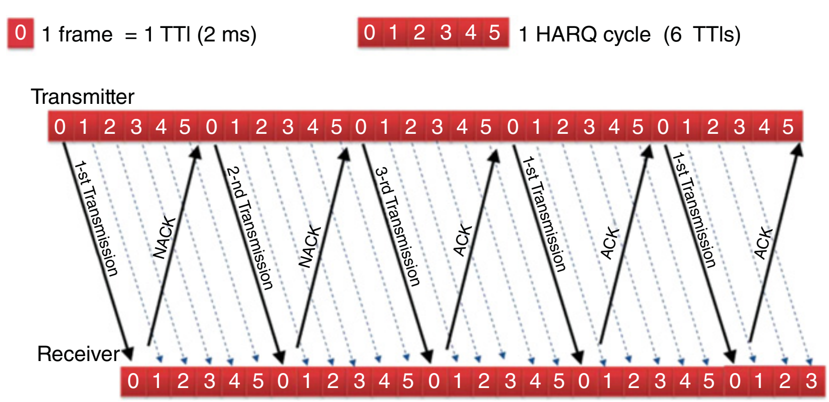

Figure 13.38 *HARQ of LTE.*

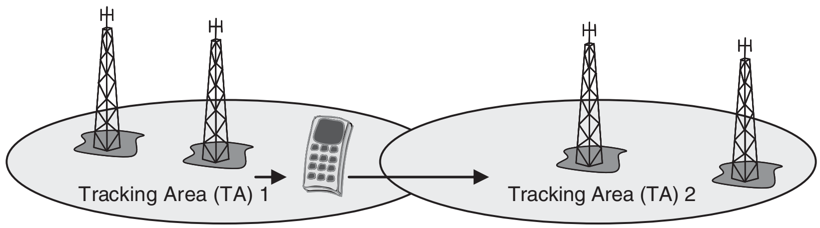

Figure 13.39 *Tracking areas of LTE. When the UE moves from one TA to another in RRC Idle mode, it has to perform a Tracking Area Update (TAU) by signaling with MME element.*

The handover procedure itself can be described as follows:

1. Handover decision – based on UE measurement reports of own and neighboring cell signal strength the serving eNodeB decides to initiate handover.
2. Handover request – using the X2 interface the serving eNodeB requests radio resource reservation from the target eNodeB.
3. Handover acknowledgement – the target eNodeB acknowledges the radio resources reservation.
4. Handover command – the serving eNodeB sends the handover command to the UE.

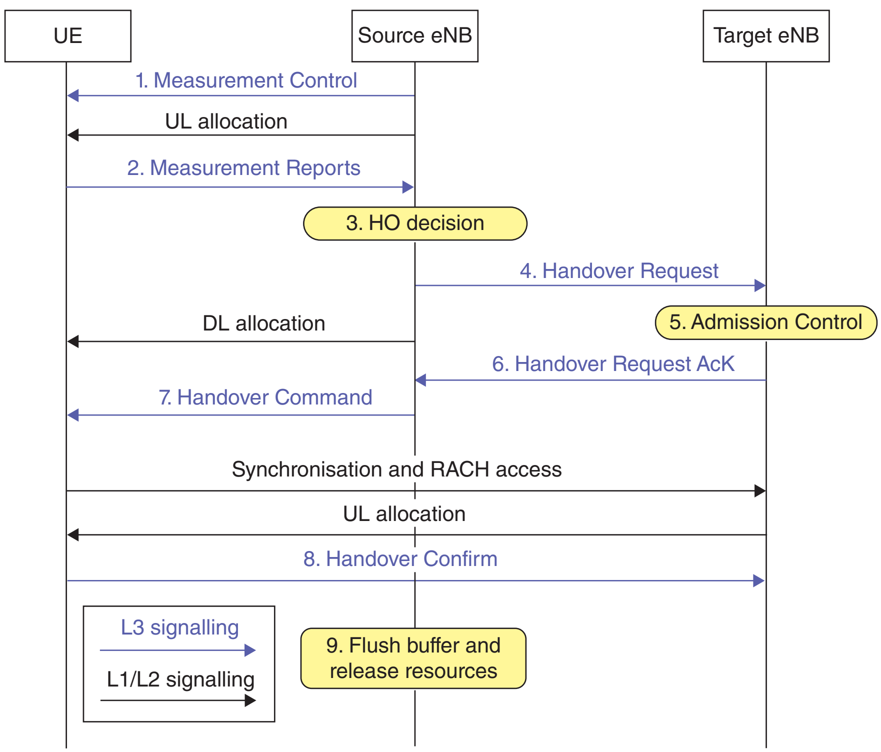

Figure 13.40 *Handover procedure in LTE.*

5. Handover – upon reception of the HO command the UE makes the final synchronization to the target eNodeB and access it using RACH procedure. The serving eNodeB forwards the data packets (RLC SDU – Service Data Units) towards the target eNodeB via the X2 interface.
6. Path switch request – the target eNodeB sends the path switch request to the core network (Mobility Management Entity – MME).
7. Path switch execution & acknowledgment – the core network (Serving Gateway – SG) switches the data path to the target eNodeB and MME acknowledges the path switch.
8. Resource release request – the target eNodeB sends the resource release request to the source eNodeB.
9. Resource release – the source eNodeB releases the radio and control resources associated with the UE.

13.15 Radio Resource Management Procedures

13.15.1 Packet Scheduling

The role of packet scheduling is to allocate resources of RAN for each active user (user being in active state) in the cellular network. Scheduling algorithms for uplink and downlink are independent and different scheduling policy can be applied. In real networks exact scheduling algorithm is a company confidential solution (it is not a subject of 3GPP standard). Moreover, the efficiency of the scheduler has direct impact on spectrum utilization efficiency. On the service level, scheduling algorithms are designed to provide data throughput and delays appropriate for the service requested by the user.

In order to improve user experience scheduler should support selected standardized features. For example by supporting LTE-A features like CoMP or Multi User MIMO (Virtual MIMO) it can significantly increase user throughput. Why not supporting all standardized features and use optimal selection of features in a given network deployment? Unfortunately the cost of advanced scheduler is limited by processing power of eNodeB equipment. Therefore, the final scheduling algorithm is a compromise between desired set of supported 3GPP features and limited processing power available for complex calculations. Additionally all calculations should be done for all supported features with minimum delay (e.g. one TTI) after receiving for example, measurements report. Scheduling algorithm should perform calculation over whole available bandwidth and for all active users. Both of them (bandwidth and number of supported users) increase, creating very demanding constrains for final implementation.

In LTE networks the scheduling process is two steps: in the first step there are resources granted in time domain, in the second step resources in spectrum domain are assigned to the users. In Figure 13.41 the steps of scheduling process are depicted.

In the process of time domain scheduling the number of users is selected to be allowed to transmit (uplink) or receive (downlink) packets. Time domain scheduling allows implementation of the admission control mechanism, and control of packet delays. Frequency domain scheduling allows performance of channel aware scheduling, that is, takes into account channel information when assigning resource blocks.

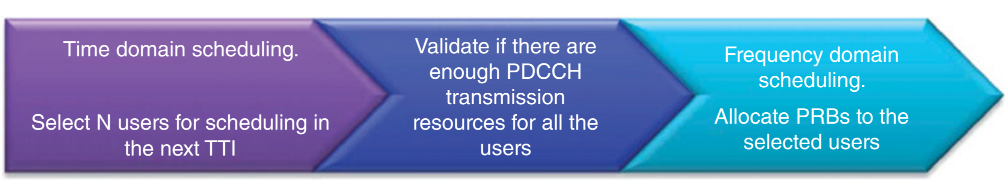

Figure 13.41 *Packet scheduling of LTE.*

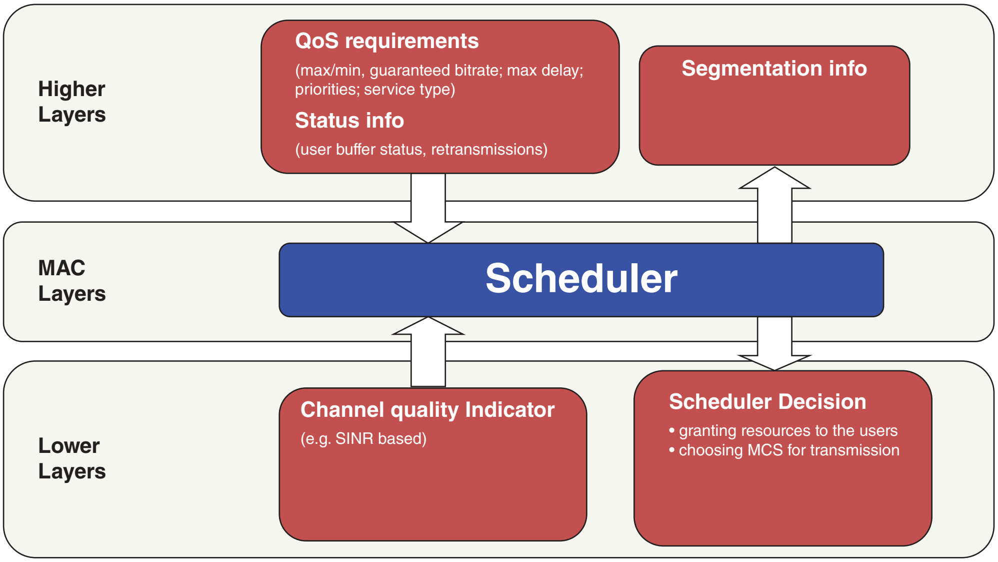

Figure 13.42 *Information flow between layers seen from scheduler perspective.*

Both time and frequency domain scheduling can be used to mitigate or coordinate interferences in the network.

If we are considering network layers model, the packet scheduler is located in the MAC layer but exchange of information is performed with both higher and lower (than MAC) layers as presented in Figure 13.42.

In order to perform channel aware scheduling the channel state information should be received from lower layers. The scheduler sent information on resources granting to users and corresponding information on modulation and coding that has to be used during the transmission. Selection of MCS has an impact on amount of information that can be sent with given resources grant – the amount of information that can be carried with each grant is provided to higher layers in order to allow segmentation of data stored in the user buffer. Buffer information from higher layers as well as quality of service parameters are valuable input for scheduler operation.

13.16 Link Adaptation

The term of link adaptation refers to system ability to adjust transmission mode to channel conditions by adjusting modulation, coding and choosing the right time for transmission. In the LTE system the link adaptation is based on Adaptive Modulation and Coding (AMC). Basically if SINR is high enough the high order modulations (e.g. 64QAM) are used, otherwise lower order modulations are used in order to provide correct reception of the symbols. On top of proper modulation selection there is a process of code rate selection. If there if high quality of the channel the high code rate is used – this allows higher data rates to be achieved compared to low code rates but by cost of weaker code protection of transmitted data.

The need for link adaptation originates from variability of channel conditions both in time and frequency domain. To make best use of resources continuous adaptation to radio channel conditions is performed. In a real system this is not a trivial task since channel information is limited, that is, there are several SINR inaccuracies that impact efficiency of link adaptation. The channel information used to link adaptation is measured instantaneous SINR which is the base for selecting modulation and coding scheme. The SINR measurement errors decrease SINR information accuracy; moreover, measured value is quantized and sent with delay, thus nonideal channel knowledge leads to suboptimal adaptation decisions and in consequence to packet transmission errors. A pair of modulation and code rate is selected in order to assure the desired BLER. Since there are mechanisms for handling retransmissions in case of error (HARQ) it has been proved that the most efficient is to target BLER in order of 0.1, that is, use relatively high MCS and accept some risk of unsuccessful transmission, hoping that in the end HARQ will succeed with transmission of the given block of data.

In DL UE measure the quality of the channel based on received reference symbols (known at UE) and send to the eNB information called CQI (Channel Quality Indicator). The CQI is the measurement that takes into account not only channel, interference and noise but also receiver quality (characterized by parameters such as noise figure).

In UL UE performs wide band channel sounding by sending reference signal, the eNB estimate SINR based on a received signal. In the resources that UE was scheduled for transmission, the SINR can be measured on Demodulation Reference Symbols (DRS) received at eNB.

13.17 ICIC

Modern wireless networks are based on multicell deployment scenarios. With such a deployment users can be served by different BTSs depending on their location. Therefore UE moving away from serving BTS observe two main effects:

- Signal from serving BTS becomes weaker.
- Signal from other BTSs becomes stronger. If serving BTS use the same resources (frequency, time, spreading code) other BTSs their signal is seen by the UE as interference decreasing the link quality.

Those effects lead to decrease in SINR experienced by the UE, especially in case all BTS use the same band (cochannel deployment). Unlike the SNR that can only be improved by increase of signal level as the noise level in a given bandwidth remains constant, the SIR can be improved by several methods, called intercell interference coordination.

The main idea behind ICIC is to orthogonalize resources in time or frequency in order to avoid severe interference to cell-edge UEs. However this leads to decrease in the average amount of resources per BTS, which limits the cell capacity. Therefore optimal tradeoff is to be found for various systems and network deployments.

13.17.1 Hard Frequency Reuse

The simplest method of resource orthogonalization is to assign different frequency reuse to different cells. This approach is called hard frequency reuse and the frequency reuse factor denoted k means that each frequency band is used in every k-th cell. In order to fully benefit from frequency reuse the network is planned in such a manner that the cells operating on the same frequency are located as far away from each other as possible. Sample layouts are shown in Figure 13.43 [40].

By means of adaptive modulation and coding schemes modern systems can utilize the available spectrum close to the Shannon law. The adaptation, however, is performed for limited SINR range. If the SINR falls below the minimum one required the connection is not possible and thus throughput equals zero. For very high values of SINR the highest MCS is used, so the SINR increase does not bring any benefits for throughput.

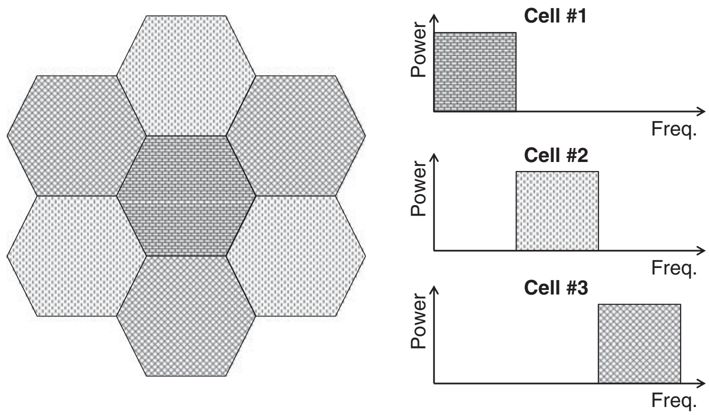

Figure 13.43 *Hard frequency reuse of LTE [40]. Figure by Jacek Góra.*

In order to estimate network performance for various frequency reuse factors, the following formula for throughput calculation can be used:

$$\mathrm{THP_k} = \begin{cases} 0 & \mathrm{SINR_k} < \mathrm{SINR_{MIN}} \\ \dfrac{\mathrm{BW}}{k} * \log(1+\mathrm{SINR_k}) & \mathrm{SINR_{MIN}} \le \mathrm{SINR_k} \le \mathrm{SINR_{MAX}} \\ \dfrac{\mathrm{THP_{MAX}}}{k} & \mathrm{SINR_k} > \mathrm{SINR_{MAX}} \end{cases}.$$

Here $\mathrm{SINR_k}$ is the effective SINR achieved for given frequency reuse, BW is the system bandwidth, $\mathrm{THP_{MAX}}$s the maximum possible throughput for full system bandwidth, $\mathrm{SINR_{MIN}}$ nd $\mathrm{SINR_{MAX}}$re the boundaries of range in which MCS adapts to channel quality.

It can be noted that due to the nature of log function, the benefits of SINR increase (for the cost of available bandwidth) will be highest for low SINR values. Therefore it makes no sense to limit the bandwidth in case the achieved SINR is already high. What is necessary is to ensure that for given network deployments the SINR will not fall below$\mathrm{SINR_{MIN}}$. Considering the two above-mentioned facts other frequency reuse schemes were proposed: soft frequency reuse and fractional frequency reuse.

13.17.2 Fractional Frequency Reuse

Figure 13.44 shows the idea of fractional frequency reuse. In such a scheme the available spectrum is divided into two parts: One is shared by all BTSs and one is further divided and used with hard frequency reuse. Cell-center UEs are served using a shared part of the spectrum, while cell edge UEs are served on a dedicated part of the spectrum. Thanks to such an approach the average part of total bandwidth available per one BTS is increased, while cell edge UEs can experience lower interference.

13.17.3 Soft Frequency Reuse

Another approach aiming on merging advantages of cochannel deployment and frequency reuse is soft frequency reuse, as seen in Figure 13.45. When used, all BTSs can use whole available spectrum; however, only one part of the spectrum full transmission power is used. Comparing to a hard frequency reuse:

- Frequency band used by given BTS in case of HFR is used with full power in case of soft frequency reuse.
- Frequency bands unused in case of HFR are used with decreased power.

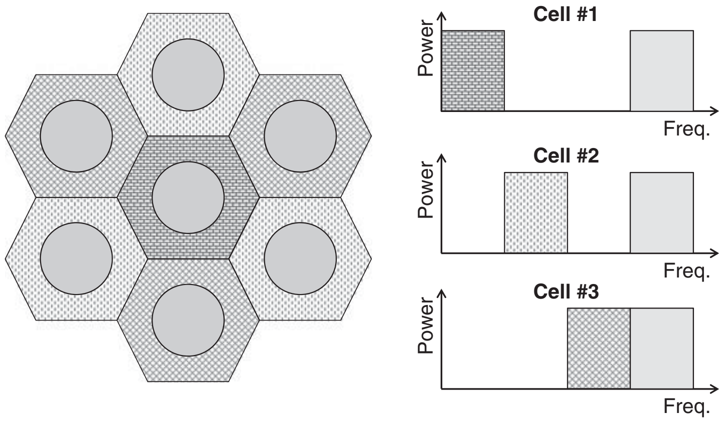

Figure 13.44 *Fractional frequency reuse in LTE [40]. Figure by Jacek Góra.*

Cell center UEs are served with limited power, therefore they experience lower signal quality as compared to HFR. However, due to a higher amount of available resources, the performance remains high. Cell-edge UEs are served with full power and because neighboring cells transmit with lower power on the frequency band used to serve cell-edge UEs, the observed interference are not severe.

13.17.4 LTE Context

In LTE, similar as in case of UMTS, the frequency reuse factor k = 1 is used for homogeneous network deployments. However, introduction of heterogeneous networks caused a need of additional methods of ICIC. Two most popular scenarios can be highlighted, as seen in Figures 13.46 and 13.47:

a. Macro + HeNB scenario

   It is assumed that only authorized users will be allowed to connect to HeNBs. Therefore unauthorized UEs being in close proximity of active HeNB will experience severe interference from that cell. It should be noted that outdoor UEs in that scenario will not be harmed by HeNB interference due to low DL transmission power of HeNBs and attenuation of a signal on buildings' walls.

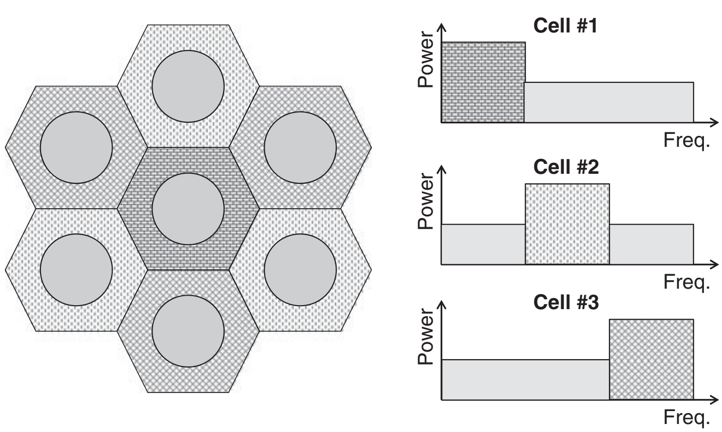

Figure 13.45 *Soft frequency reuse in LTE [40]. Figure by Jacek Góra.*

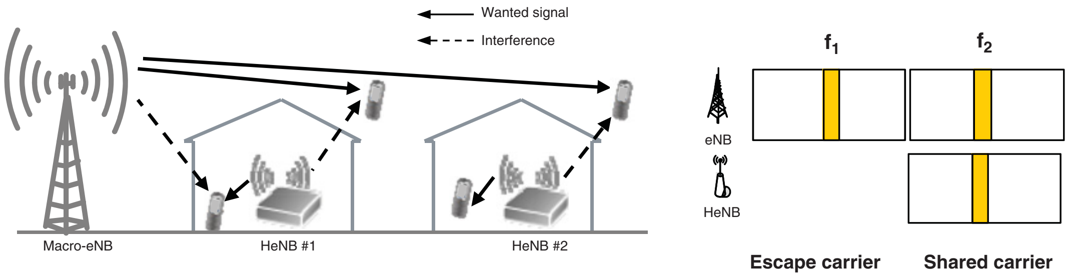

Figure 13.46 *LTE context.*

One of the possible solutions in macro+HeNB scenario is to use the escape carrier concept. This concept is based on fractional frequency reuse and assumes that one carrier frequency is to be shared by all the eNBs (both macro and HeNBs) while other carrier frequency is reserved only for macro UEs. By default macro UEs are served on a shared carrier. When the UE enters the proximity of a HeNB it starts to report interference. When the level of interference exceeds predefined threshold the UE receives the command to change the carrier frequency to an escape carrier, free of HeNB interference. The threshold should be set in such a way that the load on both carriers is similar.

b. Macro + pico scenario.

It is desired that in order to increase the range of pico cells for better offload some of the UEs should be served by pico eNBs even though the signal from macro eNB is higher. It can be done for example, by means of Cell Range Expansion, but creates a challenge how to alleviate macro interference for UEs offloaded towards the pico layer. Unlike in scenario (a) here the macro eNB is an aggressor to some of the UEs. Therefore the fractional frequency reuse scheme is set such that there is one carrier frequency used in both layers and the second one available only for pico eNBs.

13.17.5 TDM eICIC

ICIC based on various frequency reuse schemes has several limitations and drawbacks in real systems. In case of LTE the main are:

- Limited component carrier bandwidths (1.4 MHz, 3 MHz, 5 MHz, 10 MHz, 20 MHz) limits the flexibility in setting dedicated resources. Moreover, often the operators can use only one 5 MHz spectrum, which cannot be divided into smaller component carriers without loss of total available bandwidth.
- Decreasing the bandwidth available per one eNB limits the peak data rate.

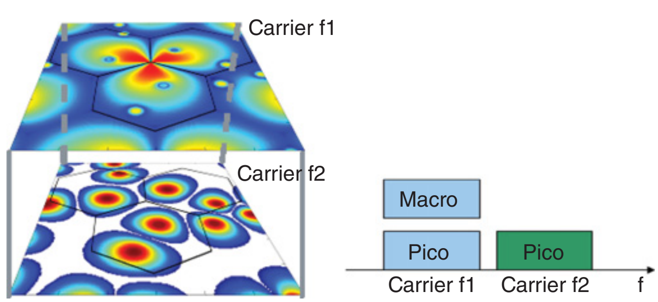

Figure 13.47 *Macro and pico scenario of LTE.*

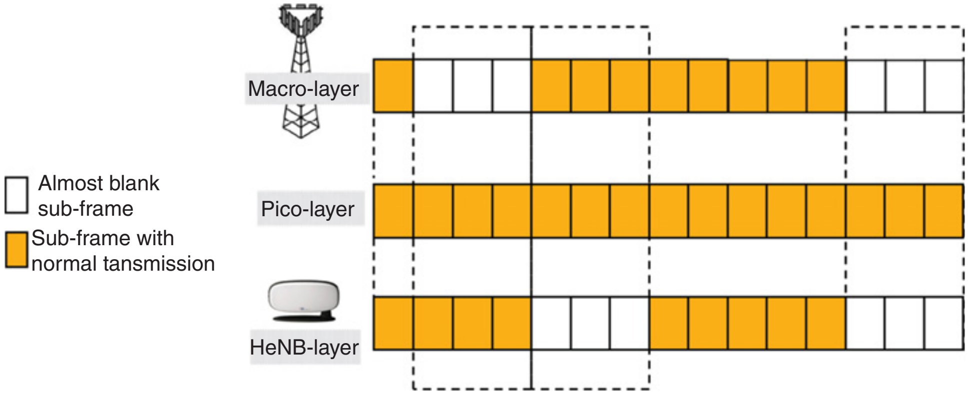

Figure 13.48 TDM eICIC of LTE.

- UEs have to perform interfrequency measurements to report the channel state on component carriers different than the currently used one. Frequent intrafrequency measurements increase the power consumption and decrease the performance, as the UE is not able to receive DL transmission during measurements.

To overcome the above-mentioned drawbacks of frequency based ICIC the time-domain (TDM) enhanced-ICIC (eICIC), as indicated in Figure 13.48, was introduced in rel'10 LTE. Unlike in the case of frequency domain ICIC all eNBs use the same carrier frequency (cochannel deployment). The concept is based on so-called Almost Blank Subframes (ABS), which are subframes without transmission on PDCCH and PDSCH channels. Other signals, like reference symbols and broadcast transmission are not muted due to backwards compatibility reasons. The aggressor eNB (i.e. eNB causing severe interference to some UEs) use the ABSs to reduce the interference in given time instances. The ABSs are organized in patterns with 40 ms periodicity. An exemplary sample of ABS pattern is shown in Figure 14.21. The pico UEs observing high interference from an overlaying macro eNB (e.g. due to CRE) are served in the subframes when the macro transmits ABS, while macro UEs harmed by HeNB are scheduled when HeNB mutes its subframes. When a pico UE is harmed both by macro and HeNB layer it has to be scheduled in the moments when both macro and HeNB transmit ABS.

In order to be able to exploit the benefits of TDM eICIC the eNBs in a network have to be synchronized on a subframe level and have an information exchange to know which subframes will be muted in neighboring cells. Also, the scheduling decision and link adaptation has to be performed based on knowledge during both the periods when neighboring cell transmit ABS and non-ABS subframes.

On the UE side it is necessary to introduce Radio Link Monitoring (RLM) measurement restrictions in order to avoid a risk of triggering Radio Link Failure (RLF) due to for example, too low RSRQ measurements. For example, a macro UE being in close proximity of unauthorized HeNB would perform RLM only during the periods when HeNB transmits ABS. It should be noted that due to the fact that some signals remain present during ABS the interference is only limited to a certain extent. However, for more advanced UE receivers it is possible to further suppress residual transmission in ABS and achieve higher performance.

13.18 Reporting

13.18.1 CSI

In LTE, the LTE-UE reports to the network via UE Channel state information (CSI). Some of the key feedback types in LTE are CQI, RI and PMI. The CSI feedback is meant to deliver information for eNodeB about DL channel state. This is the base for the eNodeB to decide the scheduling. The principle of the channel feedback

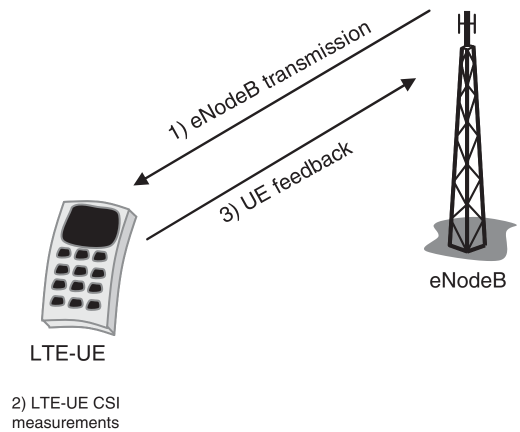

Figure 13.49 *The principle of the UE measurements.*

of LTE is quite similar to the WCDMA/HSPA, the most important difference being the frequency selectivity in the case of LTE reporting

LTE-UE measures the CSI during the call, and sends it for eNodeB via PUCCH or PUSCH channel depending on the situation. The three types of channel state information are the following: CQI (Channel Quality Indicator), RI (Rank Indicator) and PMI (Precoding Matrix Indicator). It should be noted, though, that CSI that LTE-UE sends to eNodeB is meant as general information for decision making. In fact, eNodeB is not obligated to follow it.

In the uplink direction, there is a procedure called channel sounding that delivers information about the UL channel state The information is carried with Sounding Reference Symbols (SRS).

Figure 13.49 shows the principle of the method.

13.18.2 CQI

The most intuitive channel feedback is the Channel Quality Indicator (CQI). CQI contains 16 levels (0–15), from which the level 0 is out of the range. The CQI value, that is, the index, indicates the utilized modulation and coding scheme (MCS) at the time as indicated in Table 13.3. During the LTE data call, LTE-UE reports for eNodeB the highest CQI index corresponding to the MCS for which the transport block BLER does not exceed 10%. This, in turn, can be interpreted directly as the quality of the connection at a given time. The CQI value can vary as fast as the TTI interval. In practice, when measuring the CQI values, for example, via the radio field test equipment, the statistics might be shown in such a way that during the selected period, for example, 1 second, the statistics show all the occurred CQI values and their respective percentages. This information can further postprocessed in order to create histograms over the investigated area.

LTE-UE always has a minimum of 2.33 ms for the processing of the CQI measurement. This is due to the synchronization of downlink and uplink in such a way that the CQI report transmitted in the uplink subframe $n+4$ corresponds to the reference period of the downlink subframe n for FDD. Figure 13.50 clarifies the synchronization of the reporting.

13.18.3 RI

Rank Indicator (RI) is utilized as a reporting method when LTE-UE is operating in MIMO modes with spatial multiplexing. For single antenna operation or TX diversity it is not used.

Table 13.3 The CQI values

Index	Modulation scheme	Code rate (x 1024)	Efficiency
1	QPSK	78	0.15
2	QPSK	120	0.23
3	QPSK	193	0.38
4	QPSK	308	0.60
5	QPSK	449	0.88
6	QPSK	602	1.2
7	16-QAM	378	1.5
8	16-QAM	490	1.9
9	16-QAM	616	2.4
10	64-QAM	466	2.7
11	64-QAM	567	3.3
12	64-QAM	666	3.9
13	64-QAM	772	4.5
14	64-QAM	873	5.1
15	64-QAM	948	5.6

RI is in reality a recommendation of LTE-UE for the number of layers to be used in spatial multiplexing. The RI can have a value of 1 or 2 in the case of 2-by-2 antenna configuration, and value of 1, 2, 3, or 4 in the case of 4-by-by antenna configuration. The RI is always associated to one or more CQI reports

13.18.4 PMI

Pre-coding Matrix Indicator (PMI) gives a set of information about the preferred Precoding Matrix. It should be noted, though, that like in RI, PMI is relevant only when the MIMO operation is active. The MIMO operation combined with the PMI feedback forms a closed loop MIMO.

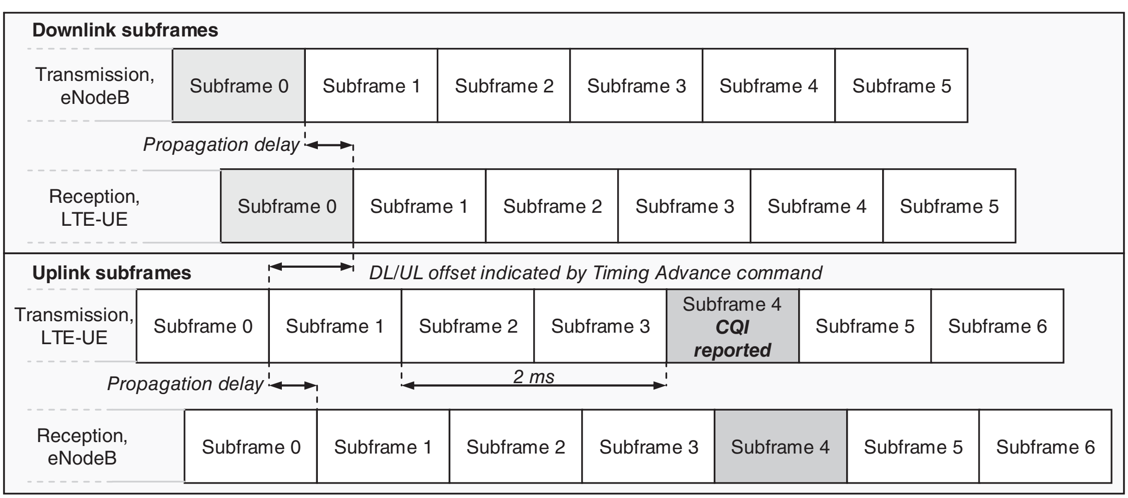

Figure 13.50 *The synchronization LTE-UE reporting.*

Table 13.4 The MIMO precoding matrix indicator table (PMI)

Codebook	1 layer	2 layers
0	$\frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\end{bmatrix}$	N/A
1	$\frac{1}{\sqrt{2}}\begin{bmatrix}1\\-1\end{bmatrix}$	$\frac{1}{2}\begin{bmatrix}1 & 1\\1 & -1\end{bmatrix}$
2	$\frac{1}{\sqrt{2}}\begin{bmatrix}1\\j\end{bmatrix}$	$\frac{1}{2}\begin{bmatrix}1 & 1\\1 & -j\end{bmatrix}$
3	$\frac{1}{\sqrt{2}}\begin{bmatrix}1\\-j\end{bmatrix}$	N/A

13.19 LTE Radio Resource Management

13.19.1 Introduction

The term Radio Resource Management (RRM) generally refers to the set of strategies and algorithms used to control parameters like transmit power, bandwidth allocation, Modulation and Coding Scheme (MCS), and so on. The aim is to utilize the limited radio resources available as efficiently as possible while providing the users with the required QoS (Quality of Service).

The uplink and downlink RRM functionalities, while sharing the same general objective of efficiently utilizing the available radio resources, face different problems and are limited by different conditions. For this reason, after a common introduction, details will be given separately.

13.19.2 QoS and Associated Parameters

As operators move from single to multiservice offering, the tools for subscriber and service differentiation become increasingly important. The EPS QoS concept in LTE comes with a set of parameters and functionalities to enable such differentiation.

The lowest level for QoS control in LTE is represented by the bearer. A bearer uniquely identifies a set of packet flows receiving a common forwarding treatment in the nodes encountered from the terminal to the gateway. A packet flow is uniquely identified by the 5-tuple: source IP address and port number, destination IP address and port number, protocol ID.

Bearers can be classified as GBR or non-GBR and as default or dedicated. Table 13.5 shows some examples of bearers and their classification. It is worth noting that a dedicated bearer can be GBR or non-GBR while a default bearer can only be non-GBR.

Table 13.5 Bearers classification

GBR type	Default bearers	Dedicated bearers
Non-GBR bearers	Bearer setup at terminal attachment	E.g., Internet browsing, chat, email
GBR bearers	N/A	E.g., VoIP, streaming

Table 13.6 QCI mapping table and typical services

QCI	Resource Type	Priority	L2 packet delay budget	L2 packet loss rate	Example services
1		2	100 ms	10-2	Conversational Voice
2	GBR	4	150 ms	10-3	Conversational Video (Live Streaming)
3		3	50 ms	10-3	Real Time Gaming
4		5	300 ms	10-6	Non-Conversational Video (Buffered Streaming)
5		1	100 ms	10-6	IMS Signaling
6		6	300 ms	10-6	Video (Buffered Streaming) TCP-based (e.g., www, e-mail, chat, ftp, p2p file sharing, etc.)
7	Non-GBR	7	100 ms	10-3	Voice, Video (Live Streaming) Interactive Gaming
8		8	300 ms	10-6	Video (Buffered Streaming) TCP-based (e.g., www, e-mail, chat, ftp, p2p file sharing, etc.)
9		9			

There exists one default bearer per terminal IP address. The default bearer is setup when the terminal attaches to the network and a serving GW is selected for it. Dedicated bearers are required to provide a different QoS to different flows belonging to the same IP address of a terminal.

GBR bearers require reserving transmission resources when the user is admitted by an admission control function. Such bearers are chosen, based on operator polices, for services for which it is preferred to block a service requests rather than degrading the performance of an already admitted service request. Non-GBR bearers, instead, may experience congestion-related packet loss, which occur in case of resource limitations.

Each EPS bearer (GBR and non-GBR) is associated with the following bearer-level QoS parameters signaled from the Access Gateway (aGW) (where they are generated) to the eNode-B (where they are used):

– Quality Class Identifier (QCI): scalar that is used as a reference to access node-specific parameters that control bearer level packet forwarding treatment (e.g. bearer priority, packet delay budget and packet loss rate), and that have been preconfigured by the operator owning the eNode-B. A one-to-one mapping of standardized QCI values to standardized characteristics is captured in specifications.
– Allocation Retention Priority (ARP): the primary purpose of ARP is to decide whether a bearer establishment / modification request can be accepted or needs to be rejected in case of resource limitations. In addition, the ARP can be used by the eNode-B to decide which bearer(s) to drop during exceptional resource limitations (e.g. at handover).

Additionally, for GBR bearers, the maximum bit rate (MBR) and GBR are defined. These parameters define the MBR, that is, the bit rate that the traffic on the bearer may not exceed, and the GBR, that is, the bit rate that the network guarantees (e.g., via admission control) it can sustain for that bearer. There also exists an aggregate MBR (AMBR) which sets a limit on the maximum bit rate that can be consumed by a group of non-GBR bearers belonging to the same user.

3GPP specifications define a mapping table for nine different QCIs, as shown in Table 13.6.

13.20 RRM Principles and Algorithms Common to UL and DL

13.20.1 Connection Mobility Control

Connection mobility control is concerned with the management of radio resources in connection with idle (RRC_IDLE) or connected (RRC_CONNECTED) mode mobility. In idle mode, the cell reselection algorithms are controlled by setting of parameters (thresholds and hysteresis values) that define the best cell and/or determine when the UE should select a new cell. Further, LTE broadcasts parameters that configure the UE measurement and reporting procedures. In connected mode, the mobility of radio connections has to be supported. Handover decisions may be based on UE and eNode-B measurements. In addition, handover decisions may take other inputs, such as neighbor cell load, traffic distribution, transport and hardware resources, and operator defined policies into account. Connection mobility control is located at L3 in the eNode-B.

13.20.1.1 Handover

The intra-LTE handover in RRC_CONNECTED state is UE assisted and network controlled. One of the goals of LTE is to provide seamless access to voice and multimedia services with strict delay requirements which is achieved by supporting handover from one cell that is, source cell, to another that is, target cell. The decentralized system architecture of LTE facilitates the use of hard handover. Hard handover (break-before-make type) is standardized for LTE while soft handover (make-before-break type) is not included, which makes the problem of providing seamless access even more critical.

The handover procedure in LTE can be divided into three phases: Initialization, Preparation, and Execution as shown in Figure 13.51. In the initialization phase UE does the channel measurements from both source and target eNode-Bs, followed by the processing and reporting of the measured value to the source eNode-B. The channel measurements for handover are done at the downlink and/or uplink reference symbols (pilots). The downlink reference symbols structure in an E-UTRA FDD frame is illustrated in Figure 13.52.

In the preparation phase the source eNode-B makes a handover decision, and requests handover with target eNode-B. Further, the Admission Control (AC) unit in target eNode-B makes the decision to admit or reject the user, which is sent to the source eNode-B using handover request ACK or NACK. Finally, in the execution phase source eNode-B generates the handover command towards UE, followed by which the source eNode-B forwards the packet to the target eNode-B. After this UE performs synchronization to the target eNode-B and accesses the target cell via Random Access Channel (RACH). When UE has successfully accessed the target cell, the UE sends the handover confirm message to the target eNode-B to indicate that handover procedure is complete. Further, target eNode-B sends a path switch message to the aGW to inform that UE has changed the cell, followed by a release resource message the source eNode-B is informed of the success of handover. After receiving the release resource message the source eNode-B releases radio as well as user-plane and control-plane related resources associated with the UE context.

13.20.2 Admission Control

The task of Admission Control (AC) is to admit or reject the establishment requests for new radio bearers. In order to do this, AC takes into account the overall resource situation, the QoS requirements, the priority levels and the provided QoS of in-progress sessions and the QoS requirement of the new radio bearer request. The goal of AC is to ensure high radio resource utilization (by accepting radio bearer requests as long as radio resources available) and at the same time to ensure proper QoS for in-progress sessions (by rejecting radio bearer requests when they cannot be accommodated).

AC is located at Layer 3 (network layer) (L3) in the eNode-B, and is used both for setup of a new bearer and for handover candidates. Hence a QoS aware AC is a requirement for GBR bearers in LTE. The AC for non-GBR bearers is optional. The QoS aware AC determines whether a new UE should be granted or denied access based on if QoS of the new UE will be fulfilled while guaranteeing the QoS of the existing UEs.

Further, due to the fact that AC is located in L3 in the eNode-B, it will utilize the local cell load information to make an admission/rejection decision. The eNode-B could also interact on X2 interface sharing load information in neighboring cells and make AC decision based on the multicell information.

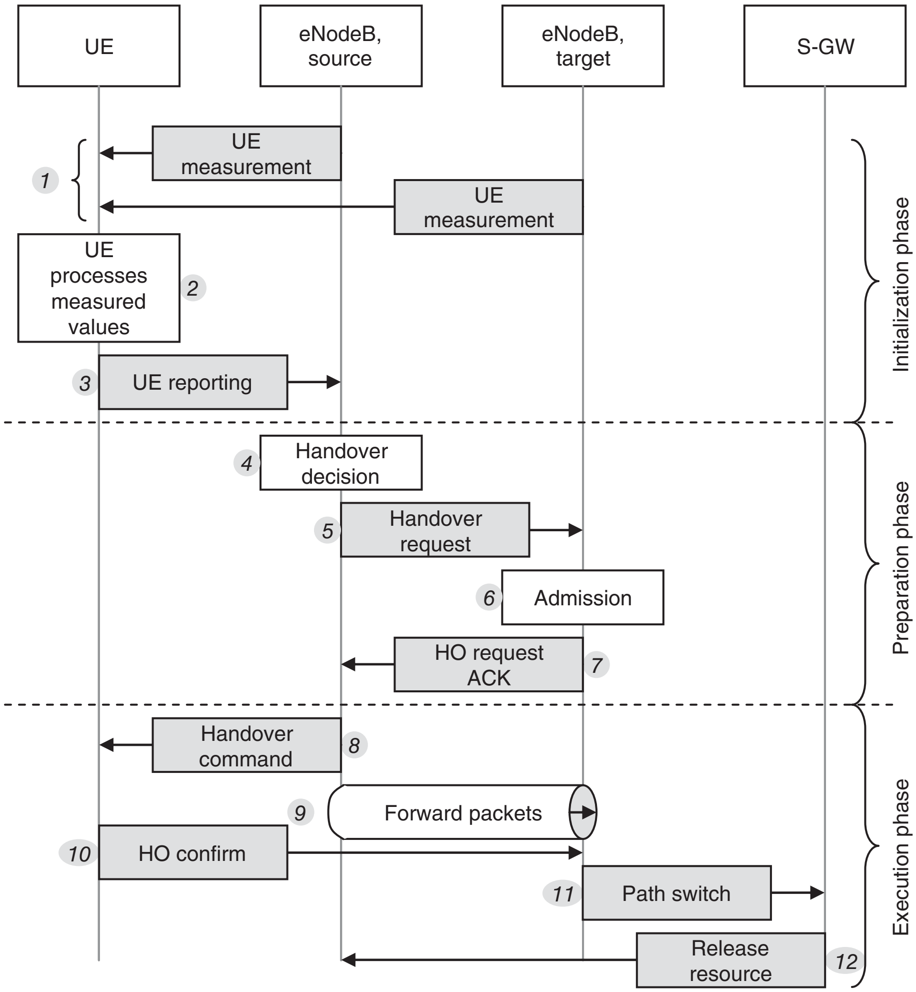

Figure 13.51 *Intra-LTE handover procedure.*

13.20.2.1 The Role of Admission Control

Admission Control (AC) is performed on the UEs that request a bearer establishment with the eNodeB. This occurs at handovers, or simply when a new bearer connection is being created. The AC decides whether the bearer can be established or not. The responsibility of the AC functionality can be seen as twofold:

– Ensure that the eNodeB has enough free resource in order to accommodate the incoming bearer.
– Ensure that the eNodeB will be able to maintain the overall expected level of QoS with the introduction of the new bearer.

In order to evaluate these two conditions, the eNodeB can take into account the QoS parameters of all the UEs and of the incoming bearer. Besides, it can take into account the channel conditions of the connected UEs via their CQIs. However no CQI is available for the incoming UE as it is not connected yet. Therefore

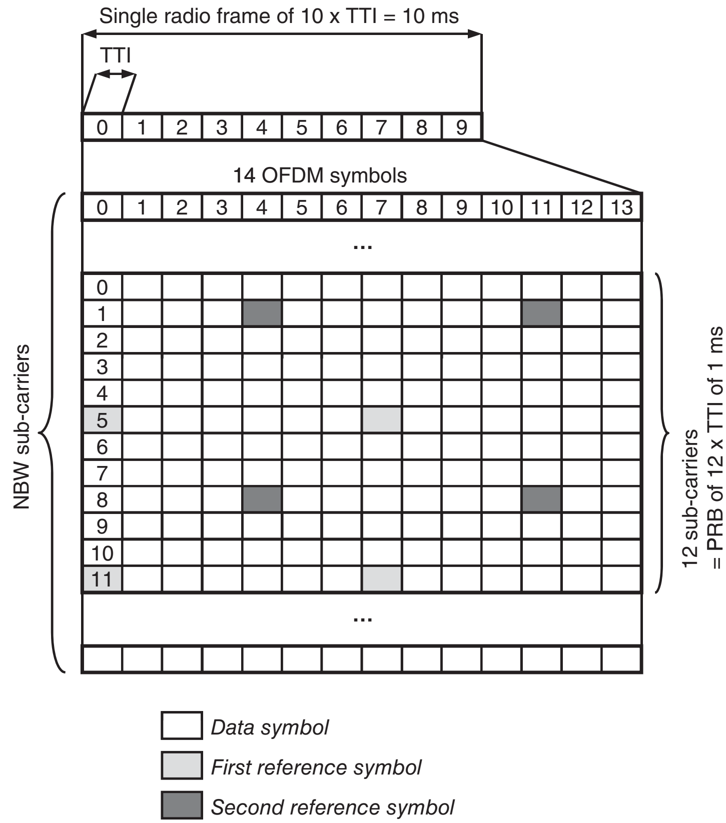

Figure 13.52 *Frame structure of E-UTRA FDD containing 14 OFDM symbols per TTI including downlink sub-carrier structure with reference signal (pilot) structure for one eNode-B transmit antenna port.*

in order to take into account the channel conditions of the incoming user, the eNodeB has to rely mostly on the following layer 3 measurements:

– The RSRP which indicates the wideband received pilot power.
– The RSSI which indicates the wideband received power including interference.

13.20.2.2 Examples of Algorithms

In this section, we give an overall list of standard algorithms.

Number of Connections: The simplest AC algorithm is simply to accept up to an arbitrary number of bearers. This method is of course overly simple and the main drawback is obviously the complete disregard to the QoS constraints of the users. This type of algorithm could, however, be found in early LTE eNodeBs as the early roll out of LTE is focused on best effort services. Best effort services are concerned by AC only since there should not be so many that only a low throughput can be provided to them.

Fixed Capacity Based: Capacity based algorithms consist of assuming a certain capacity for the system and of seeing that the sum of all the GBRs does not exceed the capacity. The main drawback of such an algorithm is that it does not take into account the channel conditions of any user. Instead, it simply assumes that the cell can accommodate a certain throughput.

Average Required Resource Based: Average required resource based algorithms calculate:

- The average percentage of resource needed for each UE to fulfill their GBR. In order to calculate this value, the CQI can be used as well as the resource allocation history of the UE.
- The expected average percentage of resource needed by the incoming UE to fulfill its GBR.

The sum of these numbers should not exceed the total available resource for the incoming bearer to be admitted.

13.20.3 HARQ

In LTE both retransmission functionalities Automatic Repeat reQuest (ARQ) and HARQ are provided. ARQ provides error correction by retransmissions in acknowledged mode at the Radio Link Control (RLC) sublayer of Layer 2. HARQ is located in the MAC sublayer of Layer 2 and ensures delivery between peer entities at Layer 1.

In case a data packet is not correctly received, the HARQ ensures a fast Layer 1 retransmission from the transmitter (UE). In this way the HARQ provides robustness against LA errors (due, for example, to errors in CSI estimation and reporting) and it improves the reliability of the channel.

In case of HARQ retransmissions failure, the ARQ in the RLC sublayer can handle further retransmissions using knowledge gained from the HARQ in the MAC sublayer.

13.20.4 Link Adaptation

13.20.4.1 The Role of Link Adaptation

The Link Adaptation (LA) is a fundamental functionality for a radio channel. It is the mechanism that chooses the appropriate Modulation and Coding Scheme (MCS) of a transmission in order to maximize the data transmitted over the channel. In LTE, Link Adaptation is also referred to as fast Adaptive Modulation and Coding (AMC) as the MCS can be changed every TTI (every 1 ms).

In LTE, the Physical Uplink Shared CHannel (PUSCH) supports BPSK, QPSK and 16 QAM at various coding rates, while the Physical Downlink Shared CHannel (PDSCH) supports QPSK, 16QAM and 64QAM with various coding rates.

In order to optimize resource use, AMC usually aims at maintaining a BLock Error Rate (BLER) on the order of 10%, while relying on HARQ to provide a packet error rate significantly smaller than 1% to the RLC sublayer. This relatively high BLER target allows the system to use high MCS thus taking full advantage of the link capacity.

13.20.4.2 Outer Loop Link Adaptation

AMC can use various channel state information (CSI in UL and CQI reports in DL) in order to determine the MCS with an appropriate block error probability. However, due to the various possible channel evaluation errors, it is unlikely that the expected block error rate occurs.

In order to maintain the BLER at first transmission as close as possible to the target, an OLLA algorithm is needed to offset the channel measurements, as shown in Figure 13.53, for a user i and a bandwidth bw.

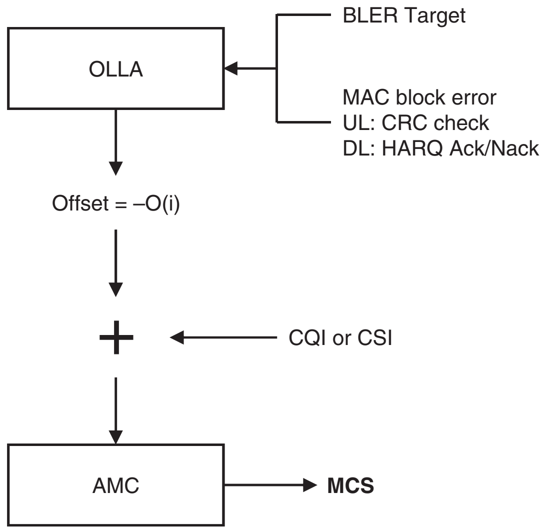

Figure 13.53 Interaction of OLLA and AMC.

The offset $O(i)$ is adjusted following the same rules of outer loop PC in WCDMA:

1. if a 1st transmission on PUSCH or DSCH is correctly received, $O(i)$ is decreased by $OD = S \cdot BLERT$;
2. if a 1st transmission on PUSCH or DSCH is not correctly received, $O(i)$ is increased by $OU = S \cdot (1 - BLERT)$,

where S represents the step size and BLERT the BLER to which the algorithm will converge if the offset $O(i)$ remains within a specified range $Omin <= O(i) <= Omax$.

13.20.5 Packet Scheduling

The PS is an entity located in the Medium Access Control (MAC) sublayer which aims at utilizing efficiently the downlink and uplink shared channel resources. The main role of the PS is to multiplex users in the time and frequency domains. Such multiplexing takes place via mapping of users to available physical resources. The PS has the possibility of performing a mapping of users to Physical Resource Block (PRB)s on a Transmission Time Interval (TTI) (1ms) basis and is therefore referred to as fast scheduling.

13.20.5.1 Multiuser Diversity

One of the reasons for using fast scheduling in LTE is the possibility to use Multiuser Diversity. Indeed, if the system is affected by time and frequency selective fading the PS entity can exploit multiuser diversity by allocating users to the portions of the bandwidth which exhibit favorable channel conditions. In this way the radio channel fading, which used to be a limitation to the performance of wireless system, is turned into an advantage.

13.20.5.2 The Packet Scheduling Problematic in Short

The packet scheduler can operate using the following information:

– The inputs from link adaptation regarding achievable MCS
– The Quality of Service (QoS) parameters associated to each UE.

The QoS parameters of a UE give, among others, explicit constraints to the eNodeB as to how much throughput should be delivered. On the other hand, the channel quality reports of a UE give indication to the eNode-B on how resources should be allocated in order to provide a certain throughput.

The packet scheduler should allocate resources with the aim of satisfying as best as possible the QoS constraints of the UEs. Meanwhile it should also maximize the throughput available for best effort services (services with no specific QoS constraints). In reality, the scheduling decisions need to take into account a large set of factors including also payloads buffered in the UE, HARQ retransmissions, UE sleep cycles, and so on.

13.20.5.3 The Role of the PDCCH

In LTE Downlink, one of the physical resources division units is the Time Transmission Intervals (TTI). From a time domain prospective, a TTI consists of 11 or 14 OFDM symbols (depending on the cyclic prefix settings), which corresponds to 1ms. The first OFDM symbols of the TTI (up to three) consist of the Physical Downlink Control Channel (PDCCH), while the remaining of the OFDM symbols can be used for data channels. Every TTI, the PDCCH is read by all connected UEs. It contains both UL and DL allocation information: sets of allocated PRBs, MCS and antenna diversity modes. Each UE will in turn read the DL TTI and send data over the UL TTI accordingly.

The PDCCH is designed to be a robust, low error channel, as a pure signaling channel, it is also limited in capacity, thus limiting both in DL and UL the number of users that can be scheduled per TTI.

13.20.5.4 Some Packet Schedulers

It is important to note that the packet scheduler is not standardized by 3GPP. The main consequences are that:

– Different eNodeB products are very likely to use different packet schedulers, which can be tuned and set in very different ways, using very different parameters.
– Packet scheduling is a way for an eNodeB manufacturer to differentiate itself from its competitors. Therefore, packet scheduling is an extensive source of research.

The present section presents important principles for packet scheduling in LTE based on the current state of research. The aim of the present section is to provide a deeper understanding of the packet scheduling problematic. Those principles generally apply to both UL and DL. The specificities and extra constraints of packet scheduling in DL and UL are detailed in later chapters.

PDCCH and Complexity Limitations The PDCCH introduces a limit as to how many users can be scheduled per TTI. For this reason, it is proposed in the multiple to adopt a decoupled time/frequency domain packet scheduling structure where:

- In a first step, the time domain scheduler selects a certain number of UEs to be scheduled out of all the connected UEs.
- In a second step, the frequency domain scheduler determines which PRBs will be allocated to which UE.

Another important limitation of the packet scheduler is that it operates on a per TTI basis: every 1ms. In such a short time, the computing capacity of even an eNodeB is limited and therefore simplicity for a packet scheduler must be preferred. For example, complex iterative algorithms may be prohibitively demanding. In order to cope with the complexity issues, the literature is mainly centered around proposing simple metric based algorithms as in Refs. [1, 28]. Metric base algorithms consist simply in:

- For the time domain packet scheduler: selecting the N_TD with the highest metrics
- For the frequency domain scheduler: allocating the UE with the highest metric of each PRB.

Diversity Algorithms One of the key features of packet scheduling in LTE is that it enables multiuser diversity gain. While usually, fading caused by multipath propagation is seen as a drawback, the LTE flexible resource allocation capabilities allows multipath propagation to be turned fading into potential gain in a very simple way. Indeed, multiuser diversity gain consists simply in scheduling only users at an SINR peak for each PRB. The following frequency domain packet schedulers are examples of diversity based packet schedulers:

- The Maximum Throughput algorithm schedules on each PRB the UE with the highest CQI, thus maximizing the overall received throughput. This scheduling strategy is very unfair as it prioritizes mainly the UEs that are close to the base station.
- The Proportional Fair algorithm schedules on each PRB the UE with the highest ratio: CQI to average CQI. Assuming that all the UEs undergo the same fading statistics, the Proportional Fair algorithm schedules on average the same amount of resource to each UE. This scheduling strategy is interesting as it introduces the notion of fairness (each user gets the same amount of resource), while each UE gets the best of its channel conditions. It has been shown that the PF algorithm provides a system throughput gain of up to 40% depending on the transmission diversity scheme used.

GBR Aware Algorithms Simple diversity algorithms may be appropriate mechanisms to cope with best effort traffic, which, by definition, does not require any specific delivery quality. However, QoS users require a more sophisticated scheduling to comply with their constraints.

Still within the decoupled frequency/time domain scheduling framework, the algorithms proposed in the literature generally focus on:

- Enforcing the GBR through the time domain scheduler.
- Provide the multiuser diversity benefits, for example, with a scheduler like proportional fair, via the frequency domain packet scheduler.

Following that method, a lot of packet schedulers inherited from time domain division multiplexing systems can simply be reapplied to LTE. An example of time domain packet scheduler that enforces GBR constraints is the barrier function based packet scheduling which consists in prioritizing:

- First the GBR UEs which don't comply with their GBR constraints. Within that set, the UEs which are furthest from their GBR target have higher priority.
- Then the GBR UEs that comply with their GBR constraint and the best effort users. Within that set, a time domain proportional fair metric is used.

This overall GBR control principle is further refined in 3GPP specifications where two principles are shown. Firstly, in order to provide an appropriate control to the time domain scheduler, the frequency domain

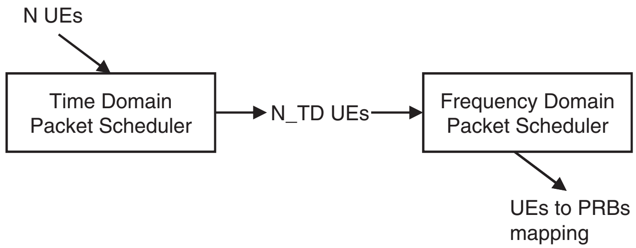

Figure 13.54 *Split time and frequency packet scheduling.*

scheduler should be independent from the time domain scheduler in the sense that its average behavior should not be influence by the nature of the time domain scheduler. Secondly, it has be shown that the time domain scheduler cannot necessary insure the GBR if this one is too high. In that case, it is necessary to include throughput control as well in the frequency domain scheduler, as seen in Figure 13.54.

Delay Aware Algorithms Many applications like video, sound streaming or VoIP may have tight delay budget in their QoS constraints. Video and sound streaming UEs, can be delivered their requested QoS using a barrier function based scheduling in the very same way than for GBR UEs. It consists simply in prioritizing with the time domain scheduler the UEs which are getting close to their delay budget compared to other users.

In practice, one of the challenges with delay budget UEs is that the packet scheduler must insure the transmission of the full amount of data that is close to expiry before expiry. For example in a traffic type with large packet, if a large packet is getting close to expiry, the packet scheduling must ensure that the full packet will have been delivered in time. For this reason, time domain scheduler must be aware of the packet size and predict how much in advance a UE must be handled to deliver the full packet in time.

VoIP UEs present some further challenges associated with their typically small packet size and very tight delay constraints (typically 50 ms). With VoIP, small packets need to be scheduled very often, which tends to overload the PDCCH when too many VoIP users are present in the cell at the same time. If VoIP UEs are overprioritized, this can result in the shared channel being underused through lack of PDCCH resource.

In order to avoid PDCCH overloading, it is necessary, as described in Ref. [32] to enforce VoIP packet bundling. It simply consists in scheduling several consecutive VoIP packets of one user in a single TTI. From the time domain packet scheduler prospective it simply means that only UEs with several packets in their buffer should be scheduled. The consequence is that VoIP users are scheduled less often but with more data at a time, thus releasing PDCCH resource for other UEs, which enable a better use of the overall shared channel.

Persistent and Semipersistent Scheduling: A Little Help for VoIP Another way to cope with the VoIP traffic is to use persistent or semipersistent scheduling. Those two modes give the possibility (with various degrees of flexibility) for the eNodeB to schedule VoIP UEs on predetermined PRBs in predetermined TTIs without having to specify it in the PDCCH every time. The main advantage of those modes is that it saves PDCCH resource and therefore allows other traffic types to better use the shared channel. However the downside is that it cannot benefit from diversity gain.

13.20.6 Load Balancing

Load balancing (or load control) has the task to handle uneven distribution of the traffic load over multiple cells. The purpose of load balancing is thus to influence the load distribution in such a manner that radio resources remain highly utilized and the QoS of in-progress sessions are maintained to the extent possible and

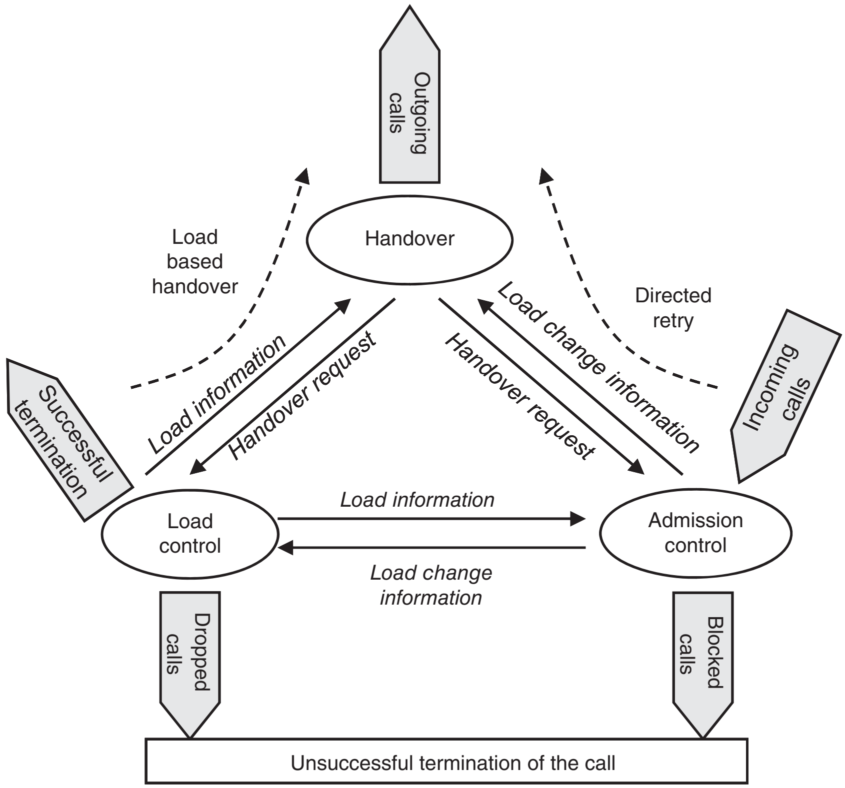

Figure 13.55 *Interaction between handover, load control, and admission control.*

call dropping probabilities are kept sufficiently small. Load balancing algorithms may result in the handover or cell reselection decisions with the purpose of redistributing traffic from highly loaded cells to underutilized cells. Load balancing functionality is located in the eNode-B.

Figure 13.55 shows that the AC, handover, and load control are closely coupled RRM functionalities. Handover is made when an active user in the source cell could be best served in the target cell. AC with the feedback from the load control functionality decides whether an incoming call (new or handover call) should be accepted or blocked. AC then informs the load control about the change in load conditions due to admission of a new or handover call. If an incoming call cannot be served in the originating cell, and if the call can be served by an adjacent cell, the call is immediately handed over to the adjacent cell. This is called directed retry which is a well-known concept used in Global System for Mobile Communication (GSM) and could potentially be used for LTE as well.

The load control keeps track of the load condition in a cell and in case of overloaded situation it drops a Best Effort (BE) call to maintain the QoS of the active calls in the cell. One way to decrease call dropping probability is to make a handover to an adjacent cell if this call could be served in the adjacent cell with the required QoS. This is called load based handover. Beside call dropping or handover of lower priority calls the QoS of lower priority calls can be degraded to free resources. This is especially useful in situations where no appropriate adjacent cell is available and therefore call drops can be avoided at the expense of degraded quality of lower priority calls.

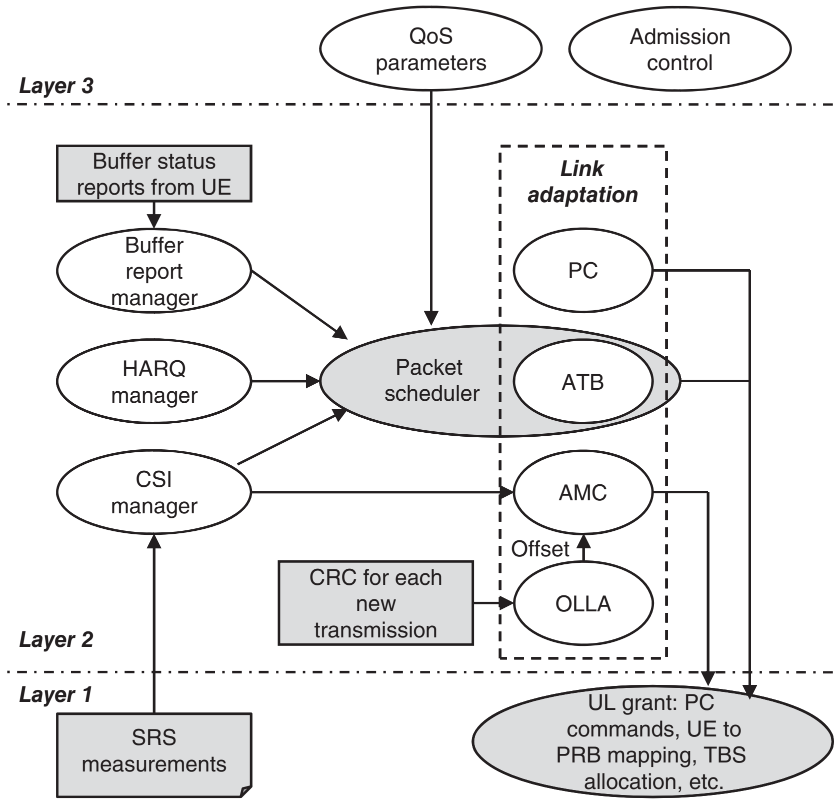

Figure 13.56 *Interaction between RRM functionalities with focus on scheduling and adaptation.*

13.21 Uplink RRM

An overview of the functionalities of UL RRM, their interaction and location in the protocol stack is shown in Figure 13.56.

After a description of some uplink-specific scheduling and link-adaptation aspects, a description of the signaling mechanisms needed to support them will be given.

13.21.1 Packet Scheduling: Specific UL Constraints

13.21.1.1 PRB Contiguity Constraint

In uplink the complexity of packet scheduling mainly arises from the fact that the PRBs allocated to the same user have to be adjacent in frequency. This constraint is connected to the physical access transmission multiplexing mode used in LTE UL: SC-FDMA. The PRB contiguity constraint greatly limits the flexibility of the scheduling and therefore negatively affects the multiuser frequency and diversity gain which can be derived from it. On the other hand, this puts less stress on the PDCCH as the PDCCH UL allocation field is shorter: it only consists of the first allocated PRB (PRBs are sorted by ascending frequency) and the number of PRBs allocated.

13.21.2 Link Adaptation

In the following, the mechanisms that control the adaptation of the main transmitting parameters, that is MCS, bandwidth and power, are described.

13.21.2.1 Adaptive Modulation and Coding

It is well known that Adaptive Modulation and Coding (AMC) can significantly improve the spectral efficiency of a wireless system. The MCS selection algorithm is based on mapping tables which return an MCS format (and hence a Transport Block Size TBS) after having received an SINR value and, optionally, the BLock Error Rate (BLER) target at first transmission as input. In LTE uplink the supported data modulation schemes are QPSK, 16-QAM and 64-Quadrature Amplitude Modulation (64-QAM).

The expected instantaneous throughput per TTI for a given MCS and SINR can be defined as:

$$T(MCS, SINR) = TBS(MCS) \cdot (1 - BLEP(MCS, SINR)) \tag{13.6}$$

where the Block Error Probability (BLEP) represents the probability that the transmitted block is going to be in error.

A possible algorithm for the selection of the MCS consists in selecting the MCS which maximizes the throughput under the constraint that the estimated BLEP is smaller or equal than the BLER target at first transmission.

The AMC can be performed on a slow basis, for example with the same rate of the power control commands to exploit the slowly changing channel variations, or on a faster basis, for example every TTI, to exploit the high instantaneous SNR conditions.

13.21.2.2 Adaptive Transmission Bandwidth

In a SC-FDMA system, which inherently enables bandwidth scalability, the adaptability of the transmission bandwidth represents a fundamental feature given the variety of services that an LTE system is called to provide. The ATB, therefore, becomes a necessary technique to cope with different traffic types, varying cell load and power limitation in the UE.

Some services, for example, VoIP, require a limited amount of bandwidth while a user with BE type of traffic may receive as much bandwidth as available as long as there are data in the buffer and power available at the UE. The power limitations also represent a constraint which highlights the importance of the ATB: The PSD a user is required to transmit with may be as high, due to adverse channel conditions, as to limit the user to support only a limited bandwidth. Additionally a varying cell load also calls for the adaptability of the transmission bandwidth as the bandwidth a user can receive depends also on the number of other users in the system.

The ATB is ultimately a functionality which allows the allocation of different portions of bandwidths to different users and therefore offers a significant flexibility when exploited as part of the scheduling process. The integration of the two functionalities, indeed, gives the possibility to better exploit frequency diversity by limiting user bandwidth allocation to the set of PRBs which exhibit the largest metric value.

13.21.2.3 Power Control

In an OFDM-based system like LTE, where orthogonality removes the intracell interference and the near-far problem typical of CDMA systems, the role of PC is changed into providing the required Signal-to-Interference-plus-Noise Ratio (SINR) while at the same time controlling intercell interference. The classic idea of PC in uplink is to modify the user transmit power as to receive all the users with the same SINR at the Base Station (BS). Such an idea is known as full compensation of the path-loss. In 3GPP the idea of Fractional Power Control (FPC) has been introduced. In this scheme users are allowed to compensate for a fraction of the path-loss so that users with higher path-loss will operate with a lower SINR requirement and will likely generate less interference to neighboring cells.

The agreed FPC scheme to set the power on PUSCH is based on an Open Loop Power Control (OLPC) algorithm aiming at compensating for slow channel variations. In order to adapt to changes in the intercell interference situation or to correct the path-loss measurements and power amplifier errors, nonperiodic close-loop adjustments can also be applied. The user transmit power is set according to the formula (13.7) expressed in *dB*m.

$$P = \min\left\{P_{\max}, P_0 + 10 \cdot \log_{10} M + \alpha \cdot L + \Delta_{MCS} + f\left(\Delta_i\right)\right\}\ [dBm] \tag{13.7}$$

where $P_{\max}$ is the maximum user transmit power, P_0 is a user-specific (optionally cell-specific) parameter, M is the number of PRBs allocated to a certain user, α is the cell-specific path-loss compensation factor that can be set to 0.0 and from 0.4 to 1.0 in steps of 0.1, L is the downlink path-loss measured in the UE based on the transmit power P_{DL} of the reference symbols.

Δ_{mcs} is a user-specific parameter (optionally cell-specific) signaled by upper-layers, Δ_i is a user specific close-loop correction value and the $f(\cdot)$ function performs an absolute or cumulative increase depending on the value of the UE-specific parameter *Accumulation-enabled*.

If the absolute approach is used the user applies the offset given in the PC command using the latest OLPC command as reference. If the cumulative approach is used the user applies the offset given in the PC command using the latest transmission power value as reference. In the latter case Δ_i can take one of four possible values: −1, 0, 1, 3 dB.

In case the Closed Loop Power Control (CLPC) term is not used, the formula is simplified to include only the open loop terms as indicated in (7.8).

$$P = \min\left\{P_{\max}, P_0 + 10 \cdot \log_{10} M + \alpha \cdot L\right\}\ [dBm] \tag{13.8}$$

The exchange of the different signals related to PC is exemplified in Figure 13.58.

13.21.3 Uplink Signaling for Scheduling and Link Adaptation Support

The PS and LA entities rely on the CSI gathered via Sounding Reference Signals (SRSs) to perform channel-aware scheduling and AMC (Figure 13.57). Similarly, the allocation of time-frequency resources to users requires knowledge of their buffer status to avoid allocating more resources than are needed. Finally, the knowledge of how close the user is to its maximum transmit power is especially relevant for ATB operations. For this reason it is worth describing in more details the signaling needed to support such operations as simplified in Figure 13.59.

13.21.3.1 Channel State Information

The uplink CSI can be described as the SINR measurement of the SRS. CSI measurements are used to gain knowledge of the channel and perform fast AMC and Frequency-Domain Packet Scheduling (FDPS).

The SRS is transmitted over the full scheduling bandwidth or over a fraction of it. Users in the same cell can transmit in the same bandwidth without interfering with each other thanks to the orthogonality provided by Constant Amplitude Zero AutoCorrelation (CAZAC) sequences and the uplink synchronous transmission. In reality there exists a constraint on the number of users in one cell that can simultaneously sound the same bandwidth without interfering with each other. The PSD on the pilot channel is the same as the one used on the

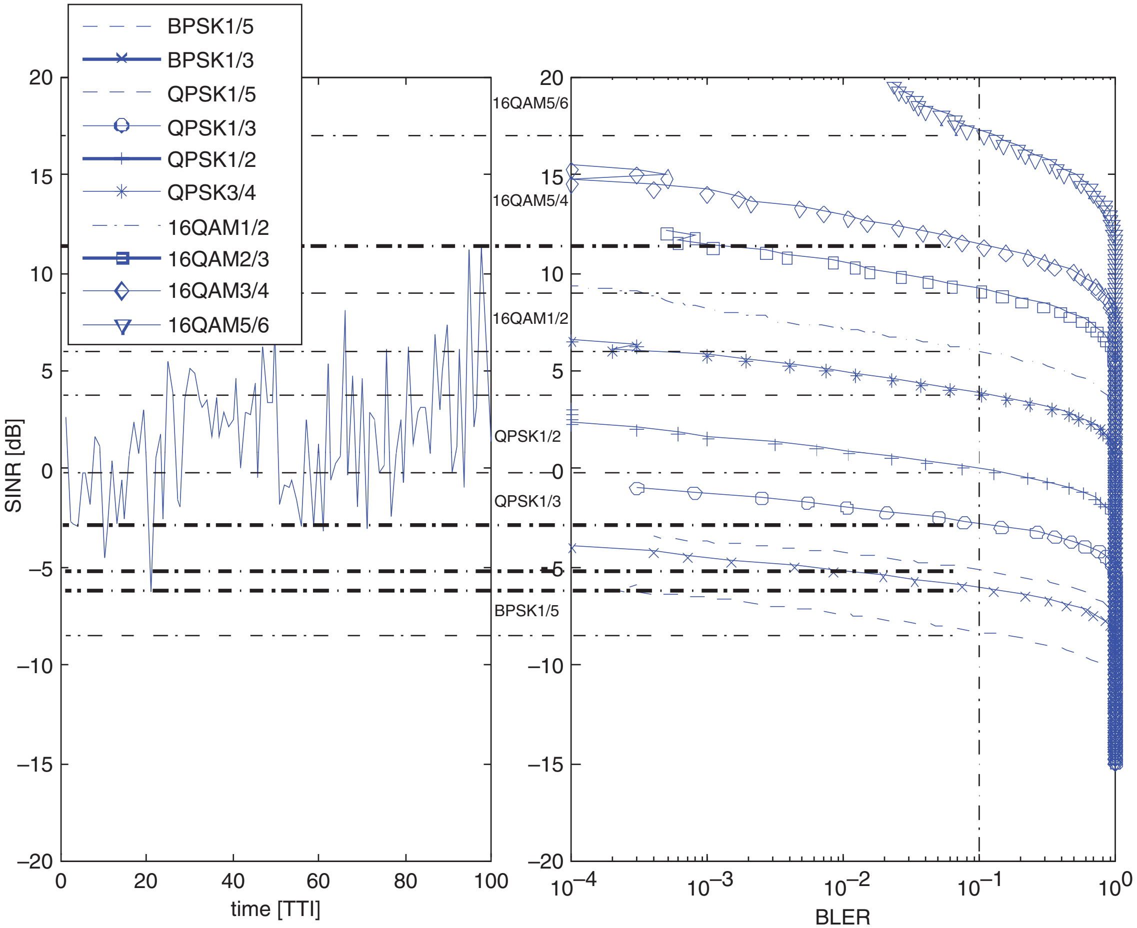

Figure 13.57 *AMC mechanism: MCS selection based on estimated SINR.*

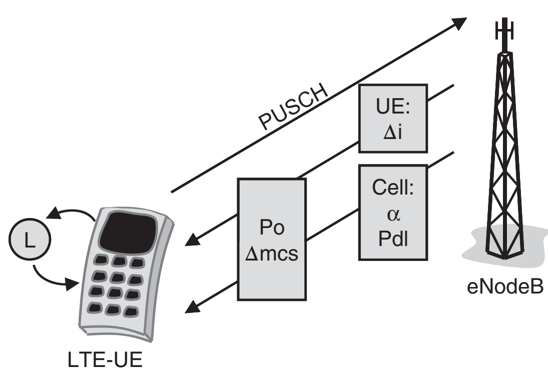

Figure 13.58 *Power control signaling.*

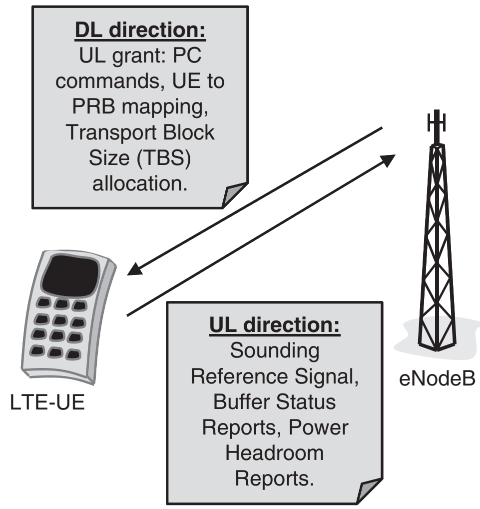

Figure 13.59 *Signaling exchange between UE and eNode-B.*

data channel. UE power capabilities typically impose a limit on the sounding bandwidth, or, alternatively, on the level of accuracy of the corresponding SINR measurements. Typically, due to the dynamic scheduling and the variability of the instantaneous interference conditions in uplink, the interference component is averaged over a certain time window. This is shown to be beneficial for channel estimation and consequently for an improvement of average cell throughput and outage user throughput.

13.21.3.2 Buffer Status Reports

The Buffer Status reporting procedure is used to provide the serving eNode-B with information about the amount of data available for transmission in the UL buffers of the UE.

A BSR can be triggered in one of three forms:

- "Regular BSR": The UE buffer has to transmit data belonging to a radio bearer (logical channel) group with higher priority than those for which data already existed in the buffer (this includes as special case the situation in which the new data arrive in an empty buffer) or in case of a serving cell change.
- "Padding BSR": UL resources are allocated and number of padding bits is equal to or larger than the size of the BSR MAC control element.
- "Periodic BSR": issued when the periodic BSR timer expires.

BSR are reported on a per Radio Bearer Group (RBG) basis as result of a compromise between the need of differentiation of data flows based on QoS requirements and the need of minimizing the resources allocated for signaling. Each RBG groups radio bearers with similar QoS requirements.

13.21.3.3 Power Headroom Reports

Due to the open-loop component of the standardized PC formula, the eNode-B cannot always know the PSD transmitted by the UE. Such information is important for different RRM operations including the allocation

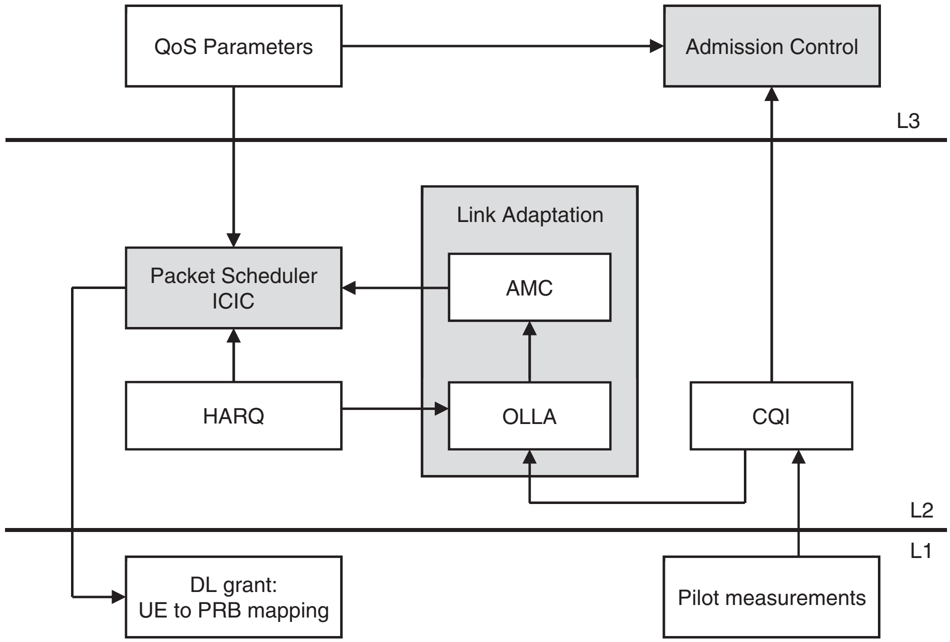

Figure 13.60 *Downlink RRM system overview.*

of bandwidth, modulation and coding scheme. Assuming that the eNode-B knows the user bandwidth, the transmission power can be derived from the information on the PSD. For this reason the power headroom reports have been standardized.

The Power Headroom reporting procedure is used to provide the serving eNode-B with information about the difference between the nominal UE maximum transmit power and the estimated power for UL-SCH transmission. A Power Headroom Report (PHR) is triggered if any of the following criteria is met:

– A predefined timer expires or has expired and the path-loss has changed more than a predefined threshold since the last power headroom report when the UE has UL resources for new transmission.
– The predefined timer expires.

13.22 Downlink RRM

Figure 13.60 presents the high-level principle of the RRM systems. The following subchapters present more detailed functionality.

13.22.1 Channel Quality, Feedback and Link Adaptation

The power levels received by each UE from the signaling eNodeB (the eNodeB to which the UE is connected) and the various interfering eNodeBs of a network affect directly the Signal to Interference and Noise Ratio (SINR) variations in both time and frequency domain. Each UE provides a CQI to the eNodeB that indicates the channel state along these two dimensions. This allows for the DL packet scheduling decisions to be made with channel information.

13.22.1.1 Channel Quality Indicator

In order to allow UEs to provide CQIs to the eNodeB, The 3GPP standard specifies a Reference Signal (RS) scheme. Every TTI, The eNodeB broadcasts reference pilots spread over the whole frequency bandwidth.

The CQI for a given sub-band is the index of the MCS supported with a block error probability not exceeding 0.1. The sub-band size depends on the transmission bandwidth. In that sense, the link adaptation in DL is very simple as the UE directly provides a MCS. The eNodeB is of course free to use a value different from the CQI. This can happen for example if the OLLA scheme described in Section 10.3.4.2 is used.

According to the 3GPP standard, the CQI can be reported in a periodic or aperiodic fashion. For periodic reporting, the reporting frequency can be set by the eNodeB. Besides, there exist several CQI transmission modes, which offer various tradeoffs between CQI codeword size and CQI frequency accuracy. The choice of the CQI transmission periodicity and mode directly affect the performance of the system.

With the CQI reports, the eNodeB has information on how much throughput is supported for each UE on each PRB. However this information is not fully reliable for mainly two reasons:

- The CQI is not instantaneous: due to the reporting delay and the reporting frequency, the CQI information does not match the exact scheduling time. An older CQI is more like to provide less accurate information. According to the remark in the present section regarding the change of the radio channel in time, the CQI will tend to be outdated faster if the UE moves at a higher speed and if the carrier frequency is higher.
- The CQI is prone to various types of estimation error related to the imperfect nature of receivers in general.

13.22.2 Packet Scheduling

The challenges exist in the design of a DL packet scheduler mostly arise from the following elements:

- The channel quality reports are not fully reliable.
- Advanced transmission techniques like Multiple Input Multiple Output (MIMO).

13.22.2.1 Resource Allocation Types: A Downlink PDCCH Related Limitation

The LTE DL TTI structure comes with an important constraint: The limited size of the PDCCH allows for only a limited number of UEs to be scheduled every TTI. Besides, in order to keep the number of UEs that can be scheduled every TTI as high as possible, LTE specifies PDCCH resource allocation code words which allow only a limited flexibility as to the PRB configurations that can be allocated to any UE. There are three types of resource allocation field referred to as:

- Resource Allocation type 0, where the allocation granularity is the Resource Block Group, a set of consecutive PRBs the size of which depends on the transmission bandwidth.
- Resource Allocation type 1, which enables distributed allocations but where the minimum distance between two allocated PRBs is of the Resource Block Group size.
- Resource Allocation type 2, where the allocation simply consists of a set of consecutive "Virtual Resource Blocks," where Virtual Resource Blocks to PRB mapping changes in a pseudo random fashion on a time slot basis.

13.22.2.2 Asynchronous HARQ

In DL, HARQ is asynchronous. In that case, the packet scheduler must take care of scheduling the retransmissions in the same way as the rest of the data as the UE does not know ahead of time when the retransmissions are scheduled. A retransmission scheduling strategy is described in Ref. [28]. In this example, the time domain scheduler treats retransmission UEs equally to others, however the frequency domain scheduler:

- First schedules the nonretransmission UEs and leave enough PRBs for the retransmissions UEs.
- Then schedules the retransmission UEs.

13.22.2.3 Packet Scheduling and MIMO Spatial Multiplexing

One of the major features of LTE is MIMO spatial multiplexing. Spatial multiplexing allows for transmitting several streams of data over the same PRBs. The number of streams that a UE can support at the same time depends on the antenna configuration at the receiver and transmitter and to a great extent on the instantaneous channel state. Together with the CQI, the UE feeds back a "rank indicator" which informs the eNodeB of how many data streams can be supported by the UE. When spatial multiplexing is enabled in the eNodeB, it is the role of the Packet Scheduler to decide whether spatial multiplexing is used or not for a UE. Besides, for close-loop spatial multiplexing, the UE feeds back a Precoding Matrix Indicator.

More complex is MU-MIMO, which consists in sending several data streams on the same set of PRBs but to various UEs. Research is still largely ongoing on MU-MIMO packet scheduling.

13.22.3 Intercell Interference Control

13.22.3.1 The Role of ICIC

LTE offers the possibility to work as a reuse 1 system: a cellular system where all the frequency resource is fully used in all cells. The main problem of a reuse 1 cellular system is the poor radio conditions at cell edge due to high interference. The role of ICIC is to control the downlink power allocation in frequency in order to provide some frequency resource with limited interference.

13.22.3.2 Fixed Power Frequency Allocation Schemes

One simple way to perform ICIC is by allocating fixed frequency power profile to each eNodeB so that:

- The part of the frequency band that has higher transmit power corresponds to low power transmission in adjacent base stations so to create a frequency allocation zone for this eNodeB with higher SINR mostly destined to cell edge users.
- The part of the frequency band that has lower transmit power corresponds to high power in adjacent base stations in order to avoid creating interference in the high SINR frequency allocation zones of adjacent eNodeBs. The low power frequency band is destined mainly to UEs closer to the eNodeB.

Figure 13.61 shows an illustration of this type of scheme.

Such schemes can be easily be mapped in a hexagonal cellular grid for example. There is almost no evidence in the literature that these techniques can achieve a significant cell edge throughput gain as most of the time, the packet scheduler can compensate the low SINR of cell edge users by allocating them more resource.

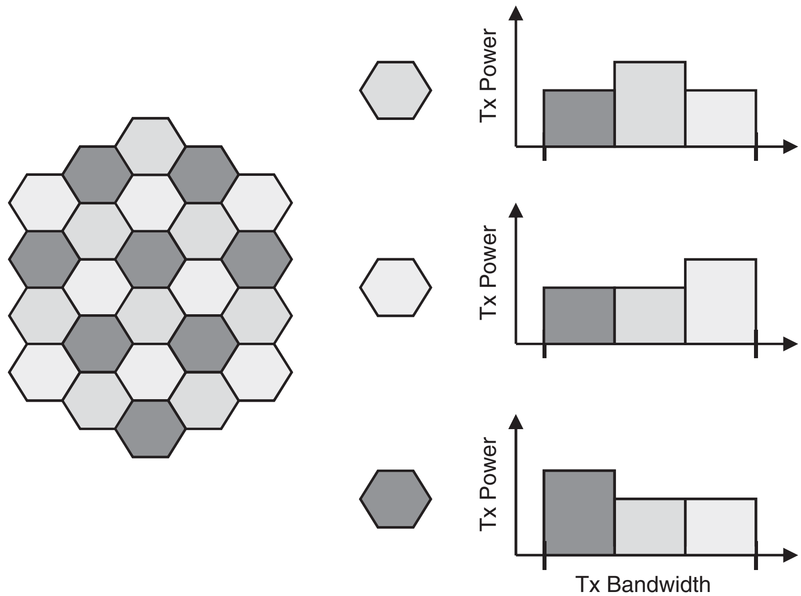

Figure 13.61 On the left-hand side: an example of cell layout with three types of cell. On the right-hand side: the transmitting power patterns associated to each cell type. Each cell has a high power / low interference frequency zone.

13.22.3.3 Dynamic ICIC

Dynamic ICIC consists in changing the power allocation based on the current power allocation schemes of neighboring cells. Dynamic power allocation schemes can use the following information:

- The resource usage status from neighboring eNodeBs.
- The CQIs indirectly contain information about the power allocation in neighboring cells as they contain some information about interference on the various PRB of each UE.

Again, there is no evidence in the literature that Dynamic ICIC can bring any significant gain for cell edge UEs. Note that Dynamic ICIC must be used cautiously as fast power variations can diminish the relevance of the CQIs and therefore create large block error rates in the system [33].

13.23 Intra-LTE Handover

The LTE uses scalable bandwidth up to 20 MHz (1.4, 3, 5, 10, 15, 20 MHz) based on the number of used subcarriers. The use of scalable bandwidth in LTE allows doing the handover measurement on different bandwidths. Hence, measurement bandwidth is a parameter of L1 filtering and should be optimized for different environments for example, user speeds. The frequency selective multipath fading will have an impact on handover performance depending on the measurement bandwidth. Handover decisions are typically based on the downlink channel measurements standardized in 3GPP which consist of RSRP, RSRQ, and so on, as seen in Figure 13.62. These measurements are filtered using a L3 filter before using it for evaluation of

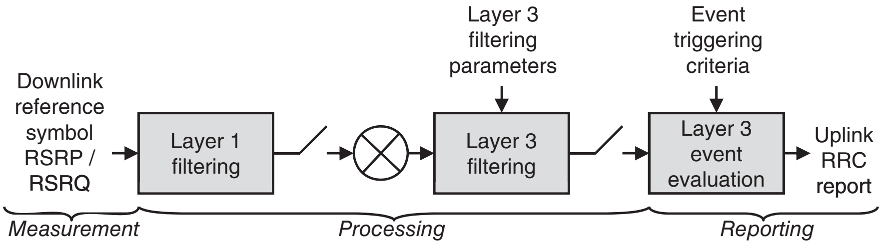

Figure 13.62 *Handover initialization phase including handover measurement, filtering and reporting in the UE.*

reporting criteria or for measurement reporting, as:

$$F_n = (1 - a) \cdot F_{n-1} + a \cdot M_n \tag{13.9}$$

where

- M_n is the latest received measurement result from the physical layer;
- F_n is the updated filtered measurement result, that is used for evaluation of reporting criteria or for measurement reporting;
- F_{n-1} is the old filtered measurement result, where F_0 is set to M_1 when the first measurement result from the physical layer is received; and
- $a = 1/2^{(k/4)}$, where k is the *filterCoefficent* for the corresponding measurement quantity received by the *quantityConfig*.

The relative influence on the updated filtered measurement result of the latest received measurement and the old filtered measurement result is controlled by the factor a.

The handover measurement report triggering is standardized in 3GPP as events A1, A2 ... A5. The handover decision is based on the filtered measurement result, $\overline{F_n}$, and is executed if for example the condition in Equation (13.10) is satisfied, where *Hys* is hysteresis margin. Additionally, time-to-trigger (TTT) window is introduced as the time to wait before making a handover decision during which the same cell remains the potential target cell.

$$\overline{F_{nT\arg etCell}}[n] \geq \overline{F_{nSourceCell}}[n] + Hys[\mathrm{dB}] \tag{13.10}$$

Introducing TTT window is one way to suppress the number of unnecessary handovers called ping-pong handovers. The ping-pong handover is defined as a handover to one of the neighboring cells that returns to the original cell after a short time. Each handover requires network resources to reroute the call to the new eNode-B. Thus, minimizing the expected number of handovers minimizes the signaling overhead. Another solution to reduce the number of handovers is to introduce a handover avoidance timer which allows handover only after the timer expires.

One of the key performance indicators (KPI) of handover performance is the reduced number of handovers which would mean reduced signaling overhead. The use of larger measurement bandwidth makes significant improvement in the performance in terms of number of handovers in the low Doppler environments. For example at 3 kmph by increasing the measurement bandwidth from 1.25 to 5 MHz a decrease of 30% in average number of handovers is noticed.

Although higher measurement bandwidth provide performance gain but in situations when different cells are operating at different transmission bandwidth an idea is to limit the measurement bandwidth to a well defined fixed value. Additionally, with the high Doppler shift, larger measurement bandwidth does not provide

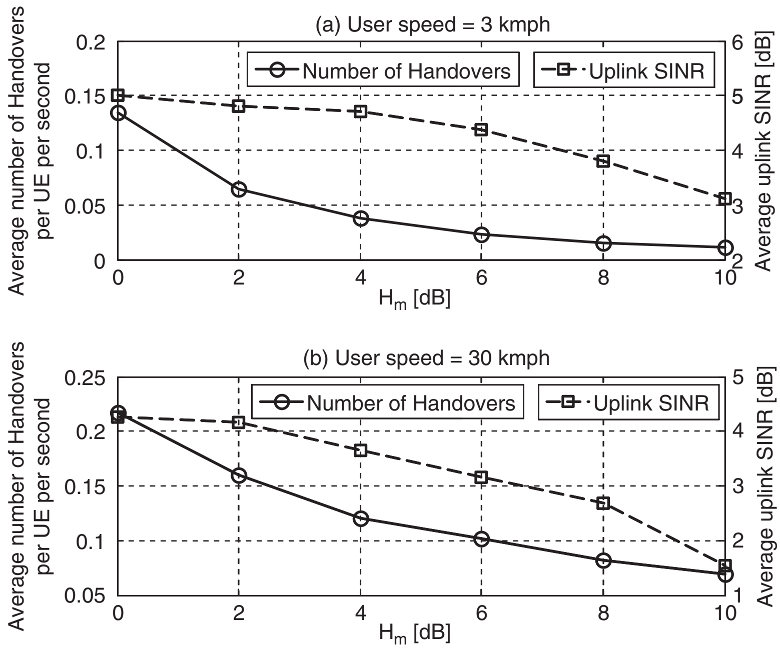

Figure 13.63 *Effect of varying Hys for the intra-LTE handover on average number of handovers and average uplink SINR at the user speeds of 3 and 30 kmph.*

any significant performance gain in terms of number of handovers. Further, it is noticed that for an adaptive choice of filtering period, depending on user speed, the gain for using larger measurement bandwidth can be made negligible for a penalty on signal quality. Hence, it is recommended to use 1.25 MHz of measurement bandwidth for a good choice of L3 filtering period.

The average number of handovers decreases for increase in the hysteresis margin as seen in Figure 13.63. The reduction in average number of handovers is a desired criteria but it as well leads to reduction in uplink quality, which is not desired.

13.24 LTE Release 8/9 Features

13.24.1 MIMO

The multiantenna transmission techniques, under the common name of multiple-input/multiple-output (MIMO), have been one of the key elements of the LTE system from the very beginning of its definition. In Release 8 of the standard support of up to 4 transmission (Tx) and 4 reception (Rx) antennas has been defined with a variety of transmissions modes (TM). Due to complexity of the related signal processing most of the transmission modes are available only in downlink. Nevertheless, although strongly limited the uplink MIMO is also available in the LTE standard in form of the so-called virtual-MIMO (V-MIMO). Details of the specific LTE MIMO transmission modes are described in this section.

For a correct reception of transmissions propagated over MIMO channel precise information on the channel state is required. For this reason the LTE Release 8 standard defines up to four sets of reference symbols (Figure 13.64) corresponding to four Tx antennas. The reference symbols allow estimation of impulse response

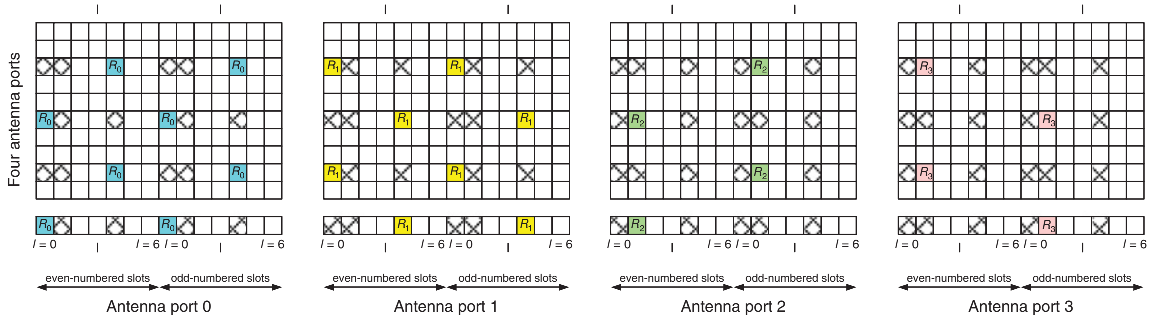

Figure 13.64 *LTE reference symbols for MIMO transmission [31]. Figure adapted from original of 3GPP.*

of each of the MIMO links. Further, this allows estimation of three basic measurements of the channel that can be used for selection and configuration of the transmission modes. The three channel state information (CSI) measurements are:

- Channel Quality Indicator (CQI)
- Rank Indicator (RI)
- Precoding Matrix Indicator (PMI).

The CSI measurements are done by the user terminal and may be reported to the base station for channel aware transmission scheduling. The CQI corresponds to the signal quality on specific physical resources. Specifically it indicates the modulation and coding scheme (MCS) recommended by the user terminal considering the observed radio conditions. The RI indicates rank of the channel, that is, the number of data streams that can be transmitted per resource element over the MIMO channel. Finally, the PMI is a recommendation of the transmission precoding optimal for a specific MIMO transmission mode.

The LTE Release 8 standard defines seven transmission modes (TM). An eight transmission mode has been added in LTE Release 9. All the TMs can be categorized into three general MIMO schemes: (1) diversity transmission, (2) spatial multiplexing and (3) beamforming.

13.24.2 Diversity MIMO

The goal of the diversity MIMO schemes is to increase robustness of a transmission. This is achieved by utilizing multiple independent transmission channels to deliver the same data. In the ideal case probability of an erroneous reception of the transmission is reduced proportionally to the number of independent channels in use. The ideal case assumes here zero correlation and no cross-interference between the individual channels.

A simple example of the diversity gain is reception of transmission in the presence of fast fading. When N transmission channels are used, the probability that all N channels will face simultaneous fade is significantly (1/N times) lower than in case of a single channel transmission. This provides that the transmission will be correctly received on at least one of the channels thus reducing block error rate (BLER) and allowing usage of higher order MCS at lower SINRs.

The diversity can be provided at the transmitter and/or at the receiver side. The former case is actually the most common case in present networks with base stations using multiple (e.g. 2) Tx antennas and mobile terminals having a single Rx antenna. This is the so-called multiple-input/single-output (MISO) transmission scheme. Alternatively, the base station can use a single Tx antenna and the terminal could use multiple Rx antennas, thus, creating the single-input/multiple-output (SIMO) transmission channel (receive diversity gain scenario).

In the LTE system the downlink transmission mode #2 (TM2) is designed explicitly for the transmit diversity. Additionally, the downlink transmission mode #1 (TM1) assumes a single Tx antenna, but could be used in combination with multiple Rx antennas, thus, providing receive diversity gains. In uplink, on the other hand, the default transmission mode is from a single Tx antenna. If the UE is equipped with multiple Tx antennas it is possible to dynamically switch the antennas to achieve the best performance. This is not the classical case of transmit diversity, yet still providing diversity gains.

In the downlink TM2 copies of the same data are transmitted from the multiple transmission antennas. Mapping of the data copies into antennas and physical resources is defined by block coding schemes, such as the Alamouti's code [3]. The Alamouti code typically defines mapping of data symbols onto two continuous time slots (space-time block code, STBC). In the LTE system it has been decided to utilize the adjacent subcarriers rather than time slots to provide the diversity (space-frequency block code, SFBC).

$$\begin{bmatrix} \hat{x}_1^0 & \hat{x}_2^0 \\ \hat{x}_1^1 & \hat{x}_2^1 \end{bmatrix} = \begin{bmatrix} x(1) & x(2) \\ -x(2)^* & x(1)^* \end{bmatrix}, \tag{13.11}$$

where $\hat{x}_c^a$ denotes the symbol transmitted from antenna a on time slot (Alamouti STBC case) or subcarrier (LTE SFBC case) c, and $x(s)$ denotes the symbol s of the data stream x.

The LTE SFBC is defined only for 2 Tx antenna case. For higher order MIMO configurations (4 Tx in LTE Release 8/9) the frequency switched transmit diversity (FSTD) modification of the SFBC is used.

$$\begin{bmatrix} \hat{x}_1^0 & \hat{x}_2^0 & \hat{x}_3^0 & \hat{x}_4^0 \\ \hat{x}_1^1 & \hat{x}_2^1 & \hat{x}_3^1 & \hat{x}_4^1 \\ \hat{x}_1^2 & \hat{x}_2^2 & \hat{x}_3^2 & \hat{x}_4^2 \\ \hat{x}_1^3 & \hat{x}_2^3 & \hat{x}_3^3 & \hat{x}_4^3 \end{bmatrix} = \begin{bmatrix} x(1) & x(2) & 0 & 0 \\ 0 & 0 & x(3) & x(4) \\ -x(2)^* & x(1)^* & 0 & 0 \\ 0 & 0 & -x(4)^* & x(3)^* \end{bmatrix}. \tag{13.12}$$

13.24.3 Spatial Multiplexing

The usage of multiple Tx and multiple Rx antennas provides the so-called spatial diversity. The spatial diversity provides that, having NTx Tx antennas and NRx Rx antennas, up to R separate transmissions can be simultaneously conducted on a resource element without loss of orthogonality. The R parameter is called the rank of the MIMO channel.

$$R = \min(N_{Tx}, N_{Rx}). \tag{13.13}$$

To correctly receive and separate the individual transmissions complex signal processing is required at the receiver side, supported by exact knowledge of instantaneous impulse response of all MIMO connections. Considering computational limitations and energy consumption constraints of a typical user terminal, it has been decided to limit in LTE Release 8/9 the maximum number of data streams to be used to 2. This means that even in case of 4x4 MIMO (rank 4) up to 2 data streams can be used per resource element.

To improve robustness and simplify the receiver side processing related to spatial multiplexing MIMO a precoding scheme has been proposed in the LTE standard. The precoding is based on a set of weights. The weights are applied on the transmitter side and map individual data streams to multiple Tx antennas. Each weight corresponds to a specific phase offset. All the available weight values are collected in a standardized codebook known to both base stations and user terminals.

$$[\hat{x}]_{(N_{Tx},1)} = [w]_{(N_{Tx},N_{Str})} * [x]_{(N_{Str},1)}. \tag{13.14}$$

where $[x]_{(N_{Tx},1)}$s vector of symbols corresponding to N_{Str} transmission streams, $[w]_{(N_{Tx},N_{Str})}$s the precoding matrix mapping N_{Str} streams to N_{Tx} Tx antenna ports, and $[\hat{x}]_{(N_{Tx},1)}$s the vector of precoded symbols to be transmitted.

The purpose of the precoding is to achieve at the receiver a specific distribution of phases of signals propagated from different Tx antennas. In case of spatial multiplexing the goal is to increase orthogonality between the individual data streams, that is, maximize difference in phases between the signals after they are propagated over the radio channel.

$$[y]_{(N_{Rx},1)} = [H]_{(N_{Rx},N_{Tx})} * [\hat{x}]_{(N_{Tx},1)} = [H]_{(N_{Rx},N_{Tx})} * \left([w]_{(N_{Tx},N_{Str})} * [x]_{(N_{Str},1)}\right). \tag{13.15}$$

where $[y]_{(N_{Rx},1)}$s vector of the signals received at N_{Rx} Rx antenna ports, and $[H]_{(N_{Rx},N_{Tx})}$s the impulse response of the MIMO channel.

In LTE Release 8 downlink there are three transmission modes that fall into the spatial multiplexing category. Those are:

- TM3 – open loop special multiplexing (OL-SM)
- TM4 – closed loop special multiplexing (CL-SM)
- TM5 – multiuser MIMO (MU-MIMO).

In the OL-SM (TM3) mode a user terminal provides to the base station the RI information, but no PMI. Depending on the RI message the number of streams can be selected. If the terminal reports Rank-1, the MIMO channel can support only one data channel. In the case a single data stream is transmitted with Tx diversity (similar to TM2). If Rank-2 is reported by the terminal, two data streams are transmitted in the cyclic delay diversity (CDD) mode. With the CDD the same data is transmitted on multiple subcarriers with a time delay added. Considering different frequencies of the subcarriers, each copy of the data will receive a different phase shift, which provides additional diversity at the receiver. As the PMI message is not provided in the TM3, the precoding weights cannot be optimally configured. In the CDD mode the precoding weights are either fixed (e.g. always the first set of weights from the codebook is used), or the weights are periodically permutated.

In the CL-SM (TM4) both the RI and PMI are provided from the terminal to the base station. This allows optimal adaptation of both the rank and precoding matrix for the transmission. This way a user terminal can achieve the higher performance than in case of OL-SM. The problem is, however, validity of the PMI feedback. Considering fast fading, this makes CL-SM not suitable for fast moving users. Additionally, transmission of multiple streams reduces the effective SINR for each of the streams (less signal power per stream plus interstream interference). For the reason the multistream transmission (Rank>1) is only used for users with very good signal quality (i.e. placed close to a base station).

The third option with special multiplexing is the multiuser MIMO (MU-MIMO). This transmission mode is available in both downlink and uplink. In downlink the MU-MIMO is very similar to CL-SM, with the difference that the two data streams transmitted per resource element are directed to two different users. Such operation requires detailed PMI information from multiple users. Based on the PMI information the users can be paired so to minimize interstream interference. In uplink, on the other hand, two user terminals transmit to one base station on a resource element. The task of the base station is to correctly distinguish the two transmissions basing on channel estimations. As the uplink MU-MIMO transmissions come from different devices, this transmission mode is also called the virtual MIMO (V-MIMO). While the MU-MIMO does not increase performance for a single user, the main benefit of its usage is in increasing capacity of the system.

13.24.4 Beamforming

In TM4 and TM5 the PMI information is used to provide maximum orthogonality between two data streams. With the beamforming transmission scheme the goal is to maximize the positive combining of copies of the same data at the receiver location. In the LTE Release 8 standard there are two modes for beamforming transmissions:

- TM6 – CL-SM Rank-1
- TM7 – antenna array beamforming.

Additionally in LTE Release 9 TM8 has been added being TM7 with multistream transmission capability.

The TM6 is in fact the CL-SM with number of transmitted data streams fixed at one (Rank-1). When only one data stream is transmitted from the base station the SINR is not degraded as with CL-SM (no interstream interference), and with optimal precoding it can be even increased. This makes TM6 suitable for transmissions to cell-edge users that typically face low SINRs. With 2x2 MIMO the theoretical gain with TM6 is 3 dB.

The TM7 and TM8 are significantly different from TM2-6. While TM2-6 require availability of multiple individual transmission antennas at the base station, TM7-8 require availability of an antenna array. The antenna array can be in general considered as a single antenna, however, containing multiple correlated Tx/Rx elements. All the antenna elements transmit the same data streams, however, phase and amplitude modulation can be applied do shape the antenna pattern. Specifically, the beam-width and the bore-sight direction can be controlled. To optimally shape the transmission beam the TM7-8 assume that channel estimation can be done based on UE-specific reference symbols or that the base station can localize UEs based on uplink angle-of-arrival (AoA) measurements. This allows the signal power to be delivered at exact UE position with low interference spread. As result, the observed SINRs can be significantly increased. To utilize the high SINRs without MCS saturation in TM8 the dual-stream transmission (similar as in CL-SM) is enabled.

Although the TM7 and TM8 can provide very high performance, the disadvantage is that they require the antenna arrays to be used. The precision of the beamforming depends directly on the antenna array size (halving beam width requires doubling array size) and its signal processing capabilities. Therefore the solutions involving complex antenna arrays are expensive and not yet common.

For details of the MIMO technique implementation in the LTE system we recommend the reader to go to the LTE system specifications [4, 31].

13.24.5 Self-Organizing Networks

A self-organizing network (SON) is a term used to describe a communication network that supports techniques enabling the automation of network operational processes that minimize human intervention, improve their efficiency and enhance performance. The first documents reflecting the common understanding of the operators concerning self-organizing functionalities have been published by the Next Generation Mobile Networks (NGMN) alliance in 2007 and followed in 2008 by a document on requirements for the implementation of SON functionalities. This SON concept has been adopted by the 3GPP to the LTE standard from the very beginning, initiated with self-configuration features that have been specified in 3GPP Release 8 and self-optimization in 3GPP Release 9.

13.24.6 Self-Configuration Rel. 8

Scope of LTE self-configuration functions is to automate procedures associated with a new eNodeB installation and its integration with existing network environment with a human intervention reduced to the minimum.

The only required manual effort is physical installation of eNodeB in the field and operator's IP network connection provisioning. The rest of the configuration processes is carried out automatically in two phases: self-configuration in pre-operational state and self-optimization in operational state.

13.24.7 Self-Configuration in Preoperational State

At the beginning of the self-configuration process (assuming that a base station is now connected to the backhaul and powered on) a new base station must establish connection via S1 interface to the OAM and download required software as well as configuration parameters. This is basic setup stage of the eNodeB self-configuration procedure and includes following steps: (a) initial IP address acquisition by new eNodeB and basic information of OAM self-configuration Subsystem; (b), authentication of eNodeB; (c) association of a gateway; and (d) downloading of eNodeB software.

In the second stage of self-configuration process when the eNodeB is physically and logically connected to the core network, the initial radio configuration is performed. Parameters to be configured in this phase refer to the group of radio transmission parameters for example, transmission power or electrical tilt and to those describing neighbor relations. According to the parameter classification in Ref. [5] the radio transmission parameters belong to the class of the parameters being in one-to-many relationship between cells, hence impact of their configuration must be considered in at least a cluster of directly surrounding cells, whereas neighbor relationship parameters belong to the class that refers to one-to-one relationship between cells what significantly reduces initial configuration process complexity.

13.24.8 Physical Cell Identifier Selection

The Physical Cell Identifier (PCI) of different LTE cells enables user terminal to recognize source of received signal and determines DL reference signal sequence. Proper PCI assignment to a new base station is crucial in the deployment and initialization process because it must be unique between the surrounding cells to avoid collisions and confusion when the target cell for handover is determined, see Figure 13.65 (one cell cannot have more than one neighbor with the same PCI sequence). There are only 504 PCI values available in the LTE system split into 168 groups of 3 PCIs each, hence the same identifiers are reused across the network. The PCI might be assigned either by the OAM entity as a definite value or randomly selected by eNodeB from the set of potential IDs provided by the OAM.

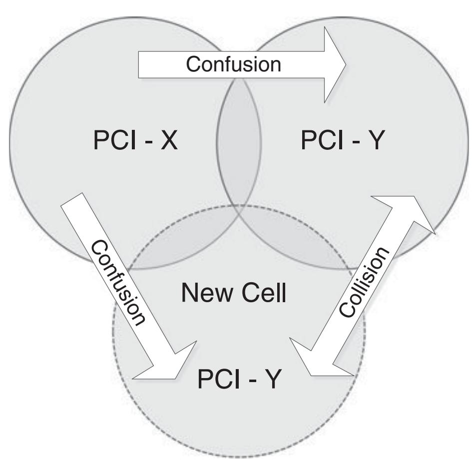

Figure 13.65 *PCI assignment with collision and confusion.*

The problem of the collision and confusion free PCI assignment is well recognized and different solution proposals can be found in literature for example, [5, 6] but neither one has been selected nor standardized, thus strategy for PCI assignment remains vendor-specific implementation.

13.24.9 Automatic Neighbor Cell Configuration and X2 Setup

Mobile networks (by a definition) allow users to switch between neighboring base stations by means of handovers, hence well-defined neighbor relation is a prerequisite. Manual neighbor relation configuration is prone to errors and costs considerable human effort. Hence, Automatic Neighbor Relation (ANR) mechanism has been introduced into LTE system that allows eNodeB to perform initial configuration to integrate with existing network environment (after rollout) as well as their optimization in a mature operational state (remove outdated or creates new relations). The LTE eNodeB starts to operate without predefined neighbor list which is dynamically created based on user terminal reports. Neighbor relationships are represented in neighbor relation table (NRT) that include all relevant information for handover purposes, for example, target cell identifier (TCI) with attributes. Whenever UE carried report indicates unknown PCI, the eNodeB orders to read E-UTRAN global cell id (EGCI) from new base station BCCH signal. The eNodeB queries OAM for relevant information of new base station, updates its NRT and establishes X2 connection if needed. The ANR procedure is shown in Figure 13.66.

13.24.10 Self-Optimization Rel. 9

The self-configuration procedures are executed mainly in the very first phase of NE life cycle, whereas self-optimization network mechanisms accompany it for the rest. Self-optimization mechanisms perform parameters tuning in a fully operational state based on information from the network as well as UEs' reports. Various algorithms can be applied to the same optimization problem and their choice might be driven by operator policy; thus optimization algorithms are not part of the standards. The 3GPP standards provide basic LTE SON use cases description, goals, limitations, list of parameters but the selection of algorithmic solutions belongs to equipment vendors.

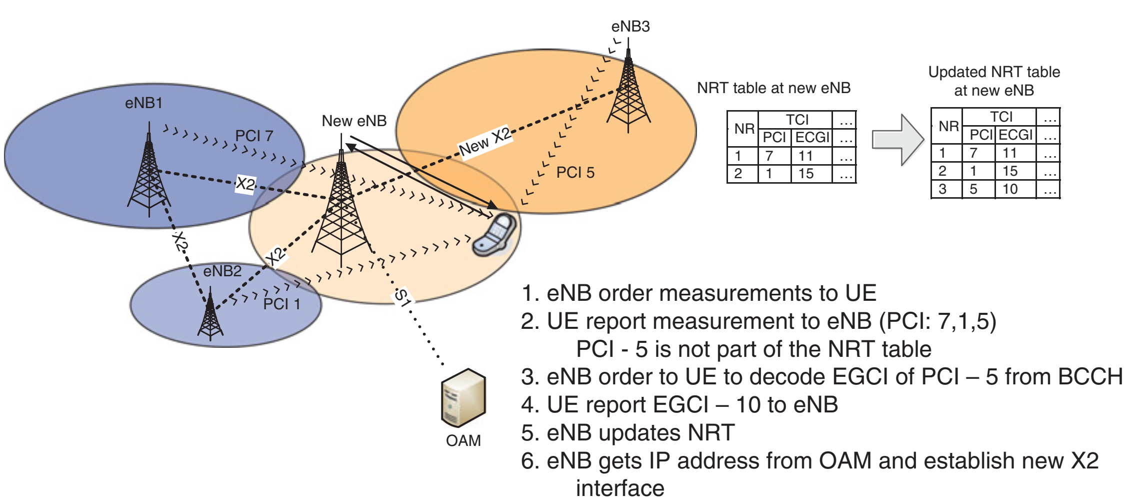

Figure 13.66 *Automatic Neighbor Relations procedure.*

13.24.11 Mobility Robustness Optimization

Mobility Robustness Optimizations (MRO) functionality aims at HO related parameters optimization in order to increase HO execution effectiveness, improve users' experience and resources utilization. The main focus of the MRO algorithms is put on reduction of Radio Link Failures (RLF) caused by Too Early, Too Late or To Wrong Cell HOs, Ping-Pong HOs and misalignments HO parameters between Idle and Active modes. The MRO mechanism operates in a distributed architecture (function is located in eNodeB), thus detected failures are reported through X2 interface to the other eNodeBs involved in HO process (related messages can be found in Ref. [7]). From the analysis of the received messages as well as locally available information (signal strength measurements, PCI etc.) the MRO algorithm makes a decision of new parameters settings, for example, Time-To-Trigger (TTT), Cell individual offset or reselection parameters and so on. The reselection-related parameters can be found in Ref. [35]. If the same parameter is linked between two eNodeBs, information about the new settings is exchanged through X2 interface.

13.24.12 Mobility Load Balancing

The Mobility Load Balancing (MLB) function aims at network parameters optimization in situation of unexpected traffic increase causing cell congestion (e.g. number of users concentrated in a small area covered by one cell) to avoid cell overload (and keep the QoS at the required level) by traffic sharing amongst the neighboring less loaded cells. In order to achieve this goal users are forced to handover from serving eNodeB (SeNodeB) to target eNodeB (TeNodeB) in intra-LTE scenario or to other RAT if overlay (and user terminal enables this). Three main phases in MLB procedure can be distinguished [8]:

- Load reporting: load related information for example, PRBs usage or cell capacity class are exchanged between eNodeBs through the X2 interface [7] or with different RAT base stations through S1 interface [9].
- Load balancing action execution: based on the obtained load information, the SeNodeB selects the best TeNodeBs (e.g. criterion of amount of the load that can be accommodated), utilizes users measurements to find best candidates and calculates new mobility settings. Designated users are forced to handover to the neighboring target cell and offload source cells.
- HO parameters readjustment: The HO process is performed due to the load purposes and against radio condition. Hence, in order to avoid users comeback new HO parameters must be adjusted between SeNodeB and TeNodeB. For this purpose a virtual border between cells might be created and shifted according to the HO point displacement by additional HO offset. Handed over users start to generate load at the TeNodeB side and release resources at SeNodeB side. This process is presented in Figure 13.67.

13.24.13 Energy Savings

Reduction of the energy consumption becomes nowadays one of the main objectives of every industry sector and telecommunication as well. LTE automated Energy Savings (ES) mechanism reduces energy consumption by switching off these eNodeB that are not needed at the time. In order not to impair the network quality by energy saving actions, as a target eNodeBs for ES algorithm are selected only these that create additional capacity, since the coverage must remain unchanged [34]. Neighboring cells must be aware of the switched off cell status in order to avoid unnecessary actions, for example, HO attempts, and hence the ES functionality must distribute this information to all relevant neighbors. The ES algorithm provides time information when to switch eNodeB off/on, but eNodeB that is switched off must be available for on demand reactivation.

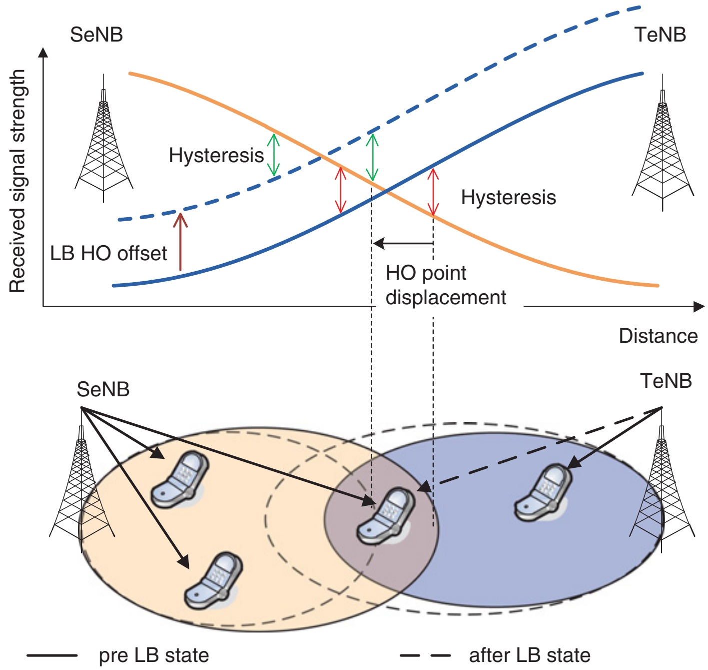

Figure 13.67 Intra-LTE load balancing principles.

13.24.14 RACH Optimization

The general task of the Random Access Channel (RACH) optimizations functionality is to increase access probability and reduce number of attempts (preambles that must be sent) in a call setup process. Number of the preamble sequences is limited to 64 per cell. This set is divided into three groups where only one is contention free with dedicated preambles (e.g. aligned during HO) while in two remaining groups preambles are common. The preamble sent by UE in access procedure might not be detected by eNodeB due to for example, high interferences or collision with the same sequence, and so on. In this case UE repeats preamble few times but after several unsuccessful attempts UE ramps-up the transmission power. Each attempt costs system resources and higher power leads to increased interferences and hence RACH optimization is crucial for UL transmission. Parameters that might be optimized by RACH optimization algorithm can be found in Ref. [36]. Random access performance can be read directly from UE reports on preamble attempts and neighboring eNodeBs.

13.24.15 Heterogeneous Networks

Traffic growth is one of the main factors that drive network evolution. Mobile communication is supporting us in more and more daily activities thus number of applications and use cases is growing. The consequence of this evolution is that additional diversification is observed in how we use the mobile communication. The difference in use cases has important meaning from network planning and radio access perspective.

This diversification was notified by research focused on mobile communication and diversification in network architecture is the response for the observed trend. The solution that takes aforementioned diversification into account is called heterogeneous networks and the basic concept can be referred as introduction of low power nodes that are deployed to complement existing network spanned by the net of macro base stations.

In heterogeneous networks macrolayer is providing basic coverage and the main goal of introducing new nodes is offloading existing macrolayer and proving high speed access to services. The new base stations are typically nodes that are simplified, better fit to the deployment environment and have lower transmit power. They are cheaper in manufacturing, operation (low energy consumption) and installation.

In this section three types of low power nodes will be described: femto-, pico- and relay-nodes.

13.24.16 Node Types (LTE Context)

Femto base stations are typically deployed indoor, have low transmit power (up to 20 dBm) and are meant to be connected with core network via user-owned Internet connection. The femto base stations are deployed by the users, so their exact location is not known to the O&M – what makes it difficult to control and manage interferences in the network. In most cases the access to the femto nodes will be granted to the Close Subscriber Group (CSG), which means that only a restricted group of users will be allowed to connect to the femto-node. In case of cochannel deployment (both macro and femto nodes operate on the same carrier) such a mode of access causes specific interference conditions in close vicinity of femto base station, that is, UEs that are not allowed to connect to a given femto experience an increased level of interference, which leads to degradation of radio link to the macro base station. Another problematic scenario occurs on the macro cell edge in UL when macro UEs transmit with high power (a large distance from the base station requires high transmit power due to path loss) and being in close vicinity of femto cell causes high interference for users served by this femto node.

In the real networks, the spatial distribution of the traffic is not uniform. Especially in the metropolitan scenarios there are locations with high concentration of active users that are called hotspots. It was proved to be beneficial for the network performance and user experience when pico-nodes are deployed in such a hotspot. The pico-nodes are similar to macro base stations – the main difference is maximum allowed transmit power (not exceeding 30 dBm compared with max. 46 dBm in case of macro base station). The pico base station is meant to be installed outdoors, has dedicated backhole and its main role is to offload macro layer in areas with high demand for the traffic. In case of pico-cells disproportion between pico and macro node transmit power cause suboptimal user to cell assignment, that is, since serving node selection is based on reference signal received power (RSRP), UE can select macro eNodeB as a serving node even if path loss to nearby pico cell is smaller than towards macro.

Third type of low power nodes are relay-nodes. The main role of these stations is to improve coverage, therefore typical location of relay-node are the macro cell edges and areas with thought conditions for radio propagation (e.g. tunnels, public transport vehicles). Those nodes are similar to pico-nodes but the main difference is that backhole is provided via radio-link, therefore part of resources (in time or spectrum domain) is dedicated for backhole. Such a flexible backhole makes relay-nodes cheap in deployment and maintenance. The flexibility is obtained by scarifying performance since part of the resources is used for the backhole link.

The remote radio head is the solution that from functional perspective is similar to relay node – provides coverage in areas with thought conditions for radio propagation. The base station radio unit is deployed in remote location and connected to the base station with fiber optics link.

13.25 LTE-Advanced Features (Rel. 10)

13.25.1 Requirements for LTE-Advanced

Mobile systems according to the LTE Release 8 specification can provide peak data rates of more than 300 Mbit/s on the downlink (DL), and 75 Mbit/s on the uplink (UL) using a frequency band of up to 20 MHz.

Currently, major performance enhancements are studied in 3GPP under the item LTE-Advanced (LTE-A, also called LTE Release 10), which will even exceed the International Mobile Telecommunications-Advanced (IMT-A) requirements [3GPP TR 36.913 2009]. Peak data rates of more than 1 Gbit/s in the downlink, and 500 Mbit/s in the uplink are the targets. Consequently, the peak spectral efficiency has to be increased up to 30 bit/s/Hz (DL), and 15 bit/s/Hz (UL) at 40 MHz bandwidth. The available bandwidth shall be expanded up to 100 MHz. Furthermore, the cost per bit has to be further lowered and the flexibility of the network enhanced. LTE-A has to be considered as an evolution of LTE with the strong requirement of backwards compatibility to Release 8 user equipment (UE) [12].

13.25.2 Motivation and Targets

After defining specifications for the initial releases of IMT-2000 (defined in ITU-R M.1457) ITU-R to meet increasing demands for wireless communication the ITU Radiocommunication Assembly approved Question ITU-R 229/8 on the future development of IMT-2000 and systems beyond IMT-2000. In 2002 strategic vision for 4G was laid out by ITU(ref) and later in 2003 ITU-R Recommendation M.1645 on IMT-Advanced was issued. The expectations for the new systems were high and with existing technologies and available spectral efficiency it was not possible to fulfill IMT-Advanced requirements in existing frequency bands; therefore in 2004 on 3GPP LTE workshop in Toronto downscaled IMT-Advanced system was proposed by NTT DoCoMo (Japan) and that was a starting point for 3GPP LTE development. The breakthrough was in 2007 where WRC-07 finally identified and allocated global spectrum for 4G. As a follow-up of spectrum allocation, in April 2008, after receiving the circular letter, the 3GPP organized a workshop on IMT-Advanced where it was decided that LTE Advanced, an evolution of current LTE standard, will meet or even exceed IMT-Advanced requirements following the ITU-R agenda. In November 2008, ITU-R established the detailed performance requirements of IMT-Advanced, by issuing a Circular Letter calling for candidate Radio Access Technologies (RATs) for IMT-Advanced as defined in M.2134. On 6 December 2010, at the ITU World Radiocommunication Seminar 2010, the ITU stated that LTE, WiMAX and similar "evolved 3G technologies" could be considered as defined in ITU-Y Q.1702.

Requirements related to technical performance for IMT-Advanced radio interfaces were defined in ITU-R M.2134. The Table 13.7 provides selection of LTE-A targets compared with IMT-Advanced minimum requirements.

13.25.3 Advanced MIMO

The Release 8/9 LTE standard defines MIMO in up to 4x4 configuration, with multitude of downlink transmission modes (TM1-8), but rather limited uplink MIMO support (single antenna transmission with optional MU-MIMO and transmit diversity). The Release 10 LTE-A standard extends the MIMO support up to 8x8 configuration in downlink and 4x4 configuration in uplink. The general focus with the Release 10 MIMO enhancements was on MU-MIMO and feedback optimization. The specific enhancements are:

- New precoding codebook design supporting up to 8 transmission antennas
- New channel state information reference symbols (CSI-RS)
- Optimizations to downlink MU-MIMO, including dynamic switching between single/multiuser modes
- Optimizations to uplink MU-MIMO providing more flexibility in user grouping.

The increase of the maximum transmission antennas up to eight imposed a significant pressure on the channel estimation and feedback mechanisms. Basically, the signaling overhead corresponding to the backwards compatible feedback mechanisms would consume big portion of the gains possible to achieve with the

Table 13.7 *Comparison of LTE-A targets with IMT-A requirements*

Criterion		LTE-Advanced Projected Capability	IMT-Advanced Requirement
Peak Data Rate for low mobility	Downlink Uplink		1 Gbit/s 0.5 Gbit/s for low mobility
Spectrum Allocation		Up to 100 MHz	Up to 40 MHz
Latency	User Plane Control Plane	10 ms 50 ms	less than **10 ms** in unloaded conditions less than **100 ms**
Peak Spectral Efficiency	Downlink Uplink	30 15	15 6.75
Cell Edge spectral efficiency[a]	Downlink Uplink	0.09 0.07	0.06 0.03
VoIP capacity * (Active users/sector/MHz)		?	40

[a]Requirements for Base coverage Urban Test environment described in M.2135.

additional antennas. For the reason the new codebook design for the eight transmission antennas case defines the two-stage feedback. The two-stage feedback includes sparse long-term precoding weights that correspond to general position of the UE and frequency selective weights for accurate precoding per resource element. In addition new, more sparsely allocated, CSI reference symbols have been defined. Such configuration enables that less feedback information can be provided from UE top the eNodeB, yet still providing sufficient channel state information.

Regarding MU-MIMO optimizations, in downlink a new user-specific reference symbols (URS) have been introduced. Usage of the new URS in principle allows UE grouping for MU-MIMO to be done independently on each resource block, thus providing increased flexibility and frequency diversity gains.

In uplink MU-MIMO the so-called orthogonal cover codes (OCC) have been introduced. Purpose of the OCC is to introduce additional orthogonality between UEs, thus enabling more proactive UE grouping for MU-MIMO, compared to Release 8 opportunistic MU-MIMO usage.

13.25.4 Carrier Aggregation

One of the requirements of ITU-R for IMT-A systems is support of downlink peak data rates of 1 Gbps. Considering spectral efficiency limitations the peak data rate can be provided only by bandwidth extension. However, according to Release 8/9 standards, the LTE system supports maximum carrier size of 20 MHz. This is insufficient to support the 1 Gbps data rate, even considering highest order MIMO schemes. For the reason, in LTE-A Release 10 the concept of carrier aggregation (CA) has been introduced. The principle of CA is enablement for a user terminal to utilize for data transmission resources of more than one frequency carrier (here called the component carrier, CC). Figure 13.71 show practical CA case.

The CA technique, as defined in the Release 10 standard, provides that, if a base station operates on more than one frequency carriers, a UE can be configured to communicate with the base station on multiple CCs at the same time. For backwards compatibility the UE is logically attached to only one CC at a time (i.e. the primary component carrier, PCC), that corresponds to the serving cell according to Release 8/9 nomenclature. For a CA-supporting UE, additional, that is, secondary component carriers (SCC) can be configured by the

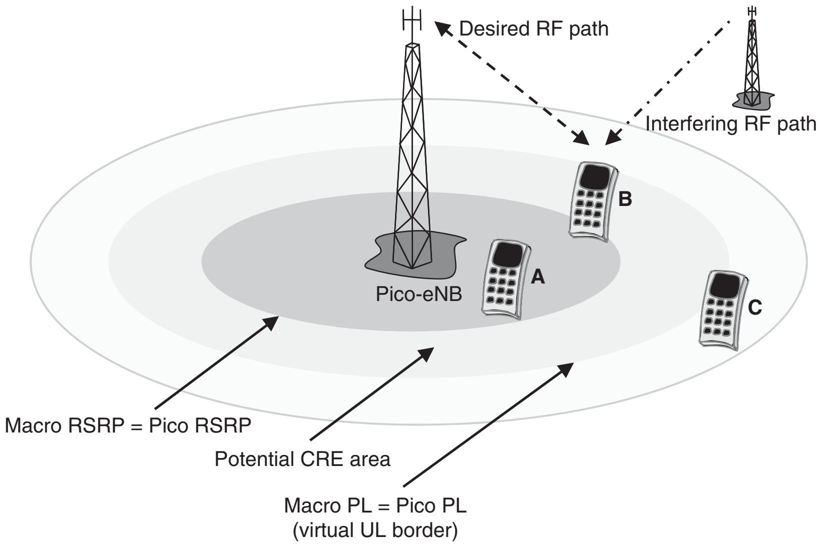

Figure 13.68 *UL-DL cell border impairment problem.*

eNodeB on top of the PCC. Both the PCC and the SCC(s) are configured per UE basis. This means that one CC can be a PCC for one UE and a SCC for a different UE.

The CA operation is controlled per UE on the radio resource control (RRC) and Mediun Access Control (MAC) levels. By means of RRC signaling an eNodeB can control the CA feature for each UE. Specifically, aside of its normal functionality the RRC layer can control configuration, activation and deactivation of the SCCs for a UE. Further, for the UEs having activated CA the MAC layer decides on allocation of radio resources from the PCC and the active SCC(s). At this stage functionalities such as carrier load balancing (Figure 13.67) and cross-carrier scheduling may be employed. Details of the user plane structure with carrier aggregation are depicted in Figure 13.69.

The Release 10 standard defines that the CA feature support aggregation of up to 5 CCs per UE (up to 100 MHz). The CCs can be allocated in the same frequency band (continuous or discontinuous aggregation) or in different frequency bands (discontinuous aggregation) (Figure 13.70). What is also important is that the coverage regions of the PCC and SCC(s) do not need to overlap. Furthermore, in general sense they do not even need to belong to the same eNodeB. The aggregation of carriers is, however, possible only in common coverage areas of the CCs. The general configurations are, however, restricted by capabilities of the UE radio transceiver. The extended (Release 10) UE capabilities define:

- Maximum number of supported CCs (PCC + SCCs)
- Supported frequency bands
- Support for intra- and interfrequency band CA

An interesting feature of the LTE-A CA is the possibility of intercarrier provision of control information, for example, scheduling related signaling, ACK/NACK feedback, power control commands, and so on. This feature will allow in the future releases of the standard support of the so-called nonbackwards compatible (NBC) carriers. An NBC carrier in general definition is an LTE carrier that for any reason does not support operation of UEs of release lower that the one in which the carrier type has been defined. Currently there are three types of NBC carriers envisioned:

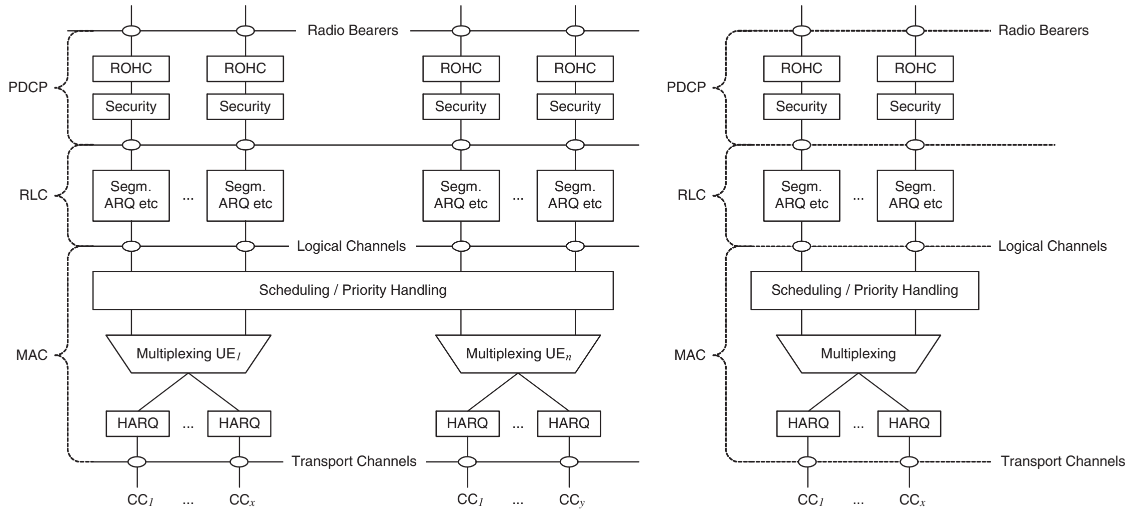

Figure 13.69 *User plane structure with carrier aggregation.*

- Extension carriers, that is, carriers dedicated for data transmissions only (i.e. not supporting all or some of the standard control channels)
- FDD carriers with nonstandard duplex spacing or carriers dedicated for only one way communication (i.e. downlink or uplink only carriers)
- Carrier segments, that is, additional resource blocks adjacent to a backwards compatible carrier, for example, in the frequency guard bands between backwards compatible carriers.

The purpose of the NBC carriers is to provide additional bandwidth for data transmission with low control channel overhead. Their introduction to the LTE-A standard is currently under discussion.

13.25.5 Relaying

The principle of the relaying technique is introduction of a new communication node, the relay node (RN), in a traditional two-node communication channel. The purpose of the RN is to support communication between the source and the target nodes by either extending connection range and/or enhancing the connection quality (Figure 13.72). Additionally, in the context of the LTE-A system, RNs are envisioned as low cost devices

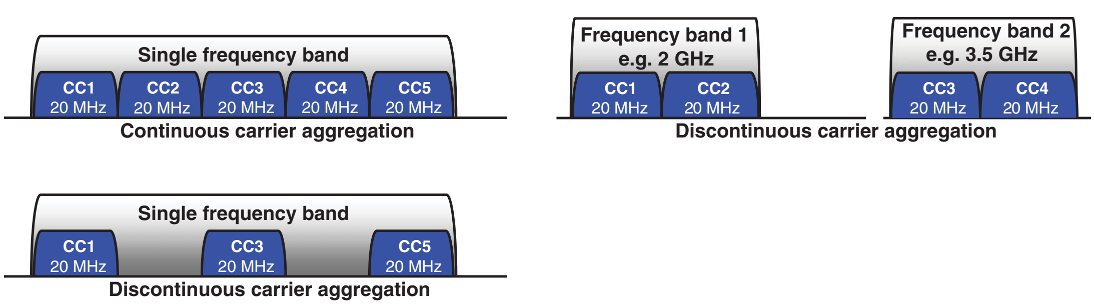

Figure 13.70 *Basic carrier aggregation configurations.*

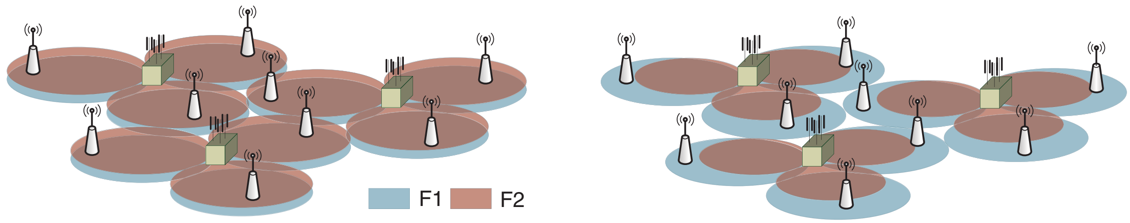

Figure 13.71 Exemplary scenarios for carrier aggregation.

(compared to traditional base stations) that could be used for providing ad hoc temporary coverage, for example, for mass events or during network failure recovery.

Over the years many concepts for the RN functionality have been developed. From the wide selection of configurations there are two that can be currently used in the LTE-A system:

- Amplitude-and-forward (AF) repeaters
- Decode-and-forward (DF) relays.

From the two configurations the DF relays are explicitly supported by the LTE-A system standard since Release 10 [1, 34], while the AF repeaters can be used without significant impact on the standard. The AF repeaters are, however, defined for compatibility reasons.

The AF model assumes simple analog amplification of radio signals on physical layer (Layer-1, L1). As the AF repeater does not perform any complex processing, it applies the same power gain to all signals, useful and interference. For the reason, the AF repeaters provide marginal gains in the presence of interference, yet are recommended for coverage limited scenarios (e.g. cell coverage extension or coverage hole patching). Due to lack of complex signal processing the AF repeaters are practically transparent for the source-target communication including negligible relaying delay.

In contrast to the AF repeaters, the DF relays are advanced digital communication devices with specific relations with other network nodes. Specifically, a LTE-A DF RN includes all functionalities of a traditional base station, lacking only fixed connection to the operator's core network (backhaul link). For full operability, the backhaul link is established by the DF RN dynamically over the LTE radio interface to another base station. On the basis of these characteristics, the DF RNs are also called the "self-backhauling" base stations. Figure 13.73 presents an example of DF RN.

The DF RNs perform a full signal reception and retransmission procedures up to Layer-3 (L3) of the communication stack. The signal processing provides regeneration of useful signals including: error correction, remodulation and re-encoding. This way the DF RNs can be used in interference-limited scenarios,

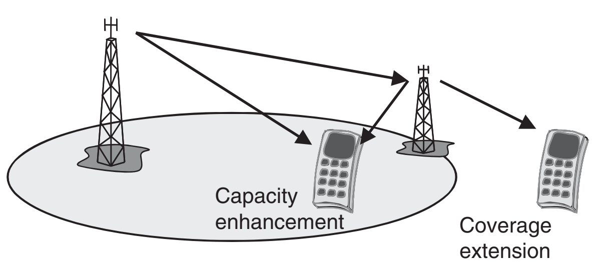

Figure 13.72 Relaying scenarios in cellular networks.

for example, to enhanced cell-edge performance. The signal processing, however, also introduces additional delay, which may decrease quality-of-service (QoS) for delay sensitive traffic, for example, voiceover-IP (VoIP). Optimization of relaying for such services is still ongoing [10].

For backwards compatibility reasons a LTE-A DF RN is defined as a combination of the eNodeB and UE functionalities (RN-eNodeB and RN-UE respectively). The RN-UE functionality is used for RN communication with a serving base station (in this context called the donor eNodeB, DeNodeB), while the RN-eNodeB functionality is used for communication with UEs (Figure 14.31). Communication wise, the RN-eNodeB supports all the interfaces and messages of a traditional eNodeB (including X2 and S1), and the RN-UE supports the signaling of a traditional UE. In case of RN-UE, however, the interface between a DeNodeB and RN-UE supports additional signaling and is called the Un interface. The additional Un signaling corresponds mainly to configuration and operation of the RN-eNodeB functionality, for example, tunneling X2 and S1 interface signaling.

In order to support RN operation the DeNodeB is equipped with a proxy and S/P-GW functionalities. The DeNodeB proxy functionality masks existence of the RN to the network and vice versa. From the network perspective the RN cell is considered to be one of the DeNodeB cells. On the other hand the DeNodeB is seen as the core network for the RN. The S/P-GW functionality of the DeNodeB takes care of routing of the data packets of the RN-served UEs and of the control signaling targeting the RN-eNodeB (e.g. X2 and S1 packets).

On the physical layer two configurations are considered for relaying: in-band and out-band. In the in-band configuration the DeNodeB-to-RN and RN-to-UE links (backhaul and access respectively) are operated in the same frequency carrier. To avoid self-interference at the RN the two links have to be time multiplexed (exception mentioned in the standard is a case when sufficient separation is provided on hardware basis). The time multiplexing is done by reusing the multimedia broadcast over single-frequency network (MBSFN) subframes defined in LTE Release 8. The MBSFN subframes are explicit subframes in which the RN-served UEs should not expect any transmissions (data and control plane). The transmission gaps created this way on the access link can be used by the RN to communicate with the DeNodeB on the backhaul link.

The alternative RN operation scheme is out-band. In this configuration the RN backhaul and access links are operated on separate frequency carriers. This requires at least two frequency carriers to be available in the system, thus, the out-band operation is so far not as popular as the in-band operation. However, considering introduction of such features as carrier aggregation the popularity of the out-band relaying is expected to increase.

Works on the LTE-A relaying standard are ongoing. Based on both 3GPP internal and external studies (e.g. in Ref. [11]) a list of new relaying configurations is developed. The most probable future relaying configurations include:

- Moving relays, that is, RNs installed in public transport vehicles
- Multi-hop relaying, that is, relaying with more than one consecutive RN in source-to-target node connection
- Cooperative relaying, that is, relaying with more than one RN supporting in parallel source-to-target node connection.

Some of the configurations may be included in the Release 12 of the LTE-A system standard.

13.25.6 Cooperative Multipoint

One of the common problems of cellular systems such as LTE is the intercell interference as seen in Figure 13.68. The LTE system has been developed with the assumption of resource reuse-1. This means that the full set of radio resources is used in each cell of the network. Such an approach increases resource availability; however, it reduces connection quality, especially close to cell borders. Overall, this leads to high unfairness of performance available at different locations in the cell.

To overcome the intercell interference several concepts have been proposed over the years under the general name of the intercell interference coordination (ICIC). The ICIC concepts consider coordinated resource partitioning and reuse between neighboring cells to reduce the negative effect of the intercell interference, with possibly minimal impact on the resource availability. Initially the ICIC concepts considered only resource coordination in frequency domain (Release 8), which further developed into time domain coordination (the so-called enhanced-ICIC, eICIC, introduced in Release 10).

All the (e)ICIC solutions can be characterized as long-term adaptation procedures. Considering the dynamics of modern networks the semistatic nature of the (e)ICIC solutions seems not be adequate. A dynamic solution for the intercell interference problem is the cooperative multipoint (CoMP) transmission concept proposed recently. In CoMP a group of cells cooperates on per transmission time interval (TTI) basis to maximize the overall performance of the system. The cooperation, as currently defined, can take one of the two forms:

- Coordinated scheduling and beamforming (CS/CB)
- Joint processing (JP).

In the CS/CB scheme the cells in a cooperation set coordinate scheduling and beamforming so as to avoid intercell interference. Thanks to the coordination in space domain (beamforming) UEs from different cells can be served on the same resources without excessive signal quality degradation. The effect is similar to a multiuser MIMO with the difference that each of the UEs receives data only from its own serving cell.

A different view on the intercell interference problem presents the JP CoMP. This concept is based on utilization of the intercell interference, rather than avoiding it. Although, this may seem as a revolutionary idea, on some level the WCDMA soft-handover may be considered as its first implementation. In LTE-A CoMP, similarly as with the WCDMA soft-handover, the user plane data need to be provided to all cells participating in the JT cooperation. Secondly, there are two options for the cooperation itself:

- Dynamic Cell Selection (DCS) – only one cell transmits to a UE at each time, however, the transmitting cell is dynamically selected for each TTI
- Joint transmission (JT) – multiple cells transmit to a UE creating a virtual MIMO effect. Additionally, multiple users can be served per resource element (virtual multi-user MIMO).

A common requirement of all CoMP schemes is provision of accurate and detailed channel state information (CSI). The CSI data is required for all links between UEs and eNodeBs in a CoMP set. Additionally, the JP CoMP requires that also the user plane data is delivered to all cooperating base stations. This imposes high requirements on the capacity and reliability of the interfaces interconnecting the cooperating base stations. Typically, the more links that are considered, the higher is the flexibility for the coordination procedures. On the other hand more links mean more data to be distributed, because the performance of CoMP technique is currently significantly downgraded by the performance of the data redistribution.

Considering the currently available interfaces the following CoMP implementations are available:

- Intrasite CoMP, that is, cooperation of cells of one eNodeB. The cooperating cells belong to one physical device, thus, redistribution of CSI data is not an issue. User plane data is provided to the eNodeB over a common S1 interface. The problem with the intrasite CoMP is low overlapping of the cooperating cells (small fraction of cell area can take advantage of the CoMP).

- Intersite CoMP, that is, cooperation of cells belonging to neighboring eNodeBs. The CSI data is redistributed between eNodeBs over X1 interface and the user plane data is provided to all the cooperating nodes from core network over the S1 interface. The overlap area is high. The bottleneck of the CoMP scheme is performance of the X1 interface.
- CoMP in relay-enhanced networks, that is, cooperation between a relay node and its donor eNodeB. The CSI and user data redistribution done over the RN wireless backhaul link. Overlap area rather high. The bottleneck is capacity of the RN backhaul link.
- Remote radio head (RRH) based CoMP. In this scheme the cooperation preprocessing is done in a central entity (eNodeB) that has a direct fast connection to distributed transmitting elements – radio heads. This scheme is to some extent similar to the WCDMA configuration with RNC that controls multiple base stations.

From the above described CoMP implementations the last one seems to be most promising. Its extension is the so-called baseband pooling concept. The baseband pooling proposes introduction of a central entity carrying on full signal preprocessing for a group of base stations on a given area. The role of base stations would then be just reduced to performing transmission and reception of radio signals. The concept is contradiction to the baseline LTE flat architecture model, however, fits perfectly into the CoMP technique.

13.26 LTE Transport and Core Network

This chapter presents the functional blocks and interfaces of LTE. The new architecture is illustrated with comparisons of the solutions with the earlier mobile communications systems. Also a protocol layer structure and functioning of the LTE protocols are described, and examples are given in order to clarify the principles of each protocol.

13.26.1 Functionality of Transport Elements

The following chapters give a practical description of the MME, S-GW, P-GW and respective transport modules. The information below is a snapshot of typical functionalities that can be applied in the core network of LTE/SAE. The complete list of functionalities depend on the vendor and the commercialization time schedules according to the road maps, so more specific information should be investigated via each vendor directly.

13.26.2 Transport Modules

An example of a transport module by Nokia Siemens Networks is characterized by the following aspects [2]:

- $4 \times$ E1/T1/JT1
- IPSec support
- Ethernet switching
- ToP (IEEE1588-2008), Sync Ethernet.

In this specific case of NSN transport element solution, there are two options. The FTLB contains $3 \times$ GE, divided into $2 \times$ GE electrical $+ 1 \times$ GE optical via SFP module, whilst FTIB contains $2 \times$ GE.

13.26.3 LTE Transport Protocol Stack

This basic item of the LTE/SAE network solution contains the IPv4 based protocol stacks for the user, control and management planes, and should be available from the first day of the LTE deployments. The LTE could also be a logical base for the support of IPv6 in order to drive the evolution path towards the evolved IP solutions.

13.26.4 Ethernet Transport

The basic LTE/SAE solution includes electrical and optical Ethernet interfaces which provides the operator with the lowest transport cost with high offered transport capacity. More specifically, the physical solution can be logically a Gigabit Ethernet 100/1000Base-T with electrical connectivity via the RJ-45 standard, and 1000Base-SX/LX/ZX with optical connectivity. Furthermore, the logical functionality includes the automatic negotiation of the mode and data rate.

13.26.5 IP Address Differentiation

This solution provides different IP address for each one of the LTE/SAE planes, that is, user, control, management and synchronization (U, S, M, S). The eNodeB applications can utilize either interface addresses or virtual addresses. In the address sharing option, the single address is shared between all the planes whilst in multiple interface address solution each plane utilizes separate addresses. In the virtual address allocation, the applications are bound to the separate, virtual addresses of each plane.

13.26.6 Traffic Prioritization on IP Layer

This functionality ensures a reliable system control in such a way that it supports different user service classes. More specifically, the DiffServ Code Points (DSCP) are possible to be configured, and also the user plane DSCPs are configurable based on the QCI of the associated EPS bearer.

13.26.7 Traffic Prioritization on Ethernet Layer

This functionality ensures the quality of service if the transport network is not QoS-aware in IP domain. One way of making this functionality is to use Ethernet priority bits in the Ethernet layer.

13.26.8 VLAN Based Traffic Differentiation

This functionality supports virtually separated networks for all the planes, that is, U, C, M and S planes. This is based on the possibility to configure the VLAN identities via the IEEE 802.1q definitions (Figure 13.75).

13.26.9 IPSec

This functionality is related to the security of the transport. Typically, IPSec can be supported in all the planes over the transport network. Practical deployments may have the security gateway and firewall integrated into the same element. Figure 13.76 presents an example of IPSec.

For more information, please refer to security related Chapter 11 and the specifications TS 33.210 (network domain security), TS 33.310 (authentication framework) and TS 401 (security architecture).

13.26.10 Synchronization

A straightforward and practical solution for synchronization is the introduction of GPS. It is a functional solution for the synchronization in such a way that no additional requirements are needed to take into account form the transport network side. GPS supports both frequency and phase synchronization. The practical limitations are result of the maximum length of the data and power cable. If the GPS receiver is integrated into the antenna element, the synchronization interface down to the eNodeB element in a commercial solution can be based on the optical fiber that does not have transmission losses. Also surge protector can be utilized between the GPS antenna and receiver and system module which minimizes the damages in case of, for example, thunderstorms.

Alternatively to the GPS synchronization, synchronization can also be done from the 2.048 MHz signal of the TDM infrastructure provided by the colocated equipment such as Base Transceiver Station of 2G, or NodeB of 3G. Still another alternative is synchronization from PDH (Plesiocronic Digital Hierarchy) interfaces.

13.26.11 Timing Over Packet

A more advanced method for synchronization is timing over packet (ToP), which provides the synchronization from the Ethernet interface and thus makes it unnecessary to utilize previous mentioned GPS or TDM link for the synchronization. This solution is defined in the IEEE 1588-2008 documentation. The solution contains a ToP Grandmaster element, which is the root source for the synchronization data delivered for the eNodeB elements over the IP / Ethernet network. The reference clock is connected to the ToP Grandmaster, and the eNodeB recovers the clock signal over the Ethernet via a ToP slave. The requirement for this type of synchronization is logically a sufficiently high-quality packet data network.

13.26.12 Synchronous Ethernet

Other synchronization method is based on the synchronous Ethernet concept, as presented in Figure 13.74. It provides accurate frequency synchronization over Ethernet links in such a way that accuracy does not depend on network load. This functionality is based on the G.8261, G.8262 and G.8264 definitions of ITU. It distributes the frequency via the layer 1 by applying an SDH-type of mechanisms. The challenge of the solution is that it must be implemented in all the nodes found in the synchronization path.

13.27 Transport Network

The deployment of LTE also means that the transport network should be designed accordingly in order to support increased maximum radio interface data rates. This means that the existing backhaul, aggregation and

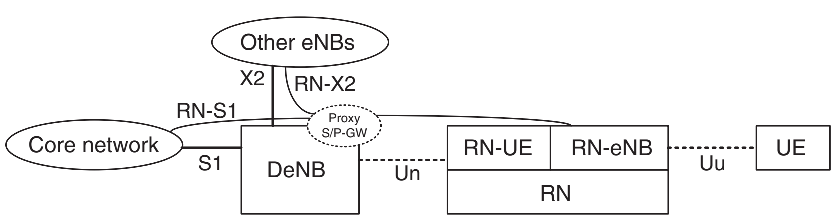

Figure 13.73 *LTE-A model of DF RN [1]. Figure adapted from original of 3GPP.*

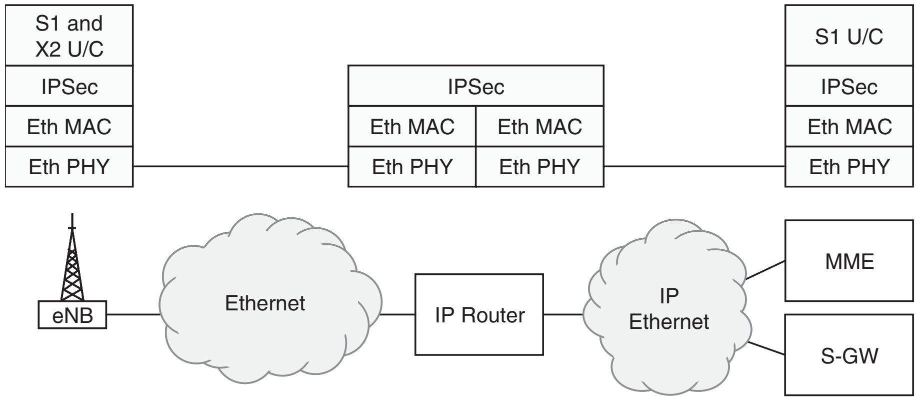

Figure 13.74 *The Ethernet solution for the LTE/SAE transport.*

backbone networks might need considerable redimensioning, that is, new hardware for enhanced capacity to guarantee the data delivery from the radio network to SAE elements, and further to the external packet data networks.

The traditional backhaul of operators, that is, based on TDM connectivity, can also be updated for the support of the packet data via the Ethernet connectivity. This type of hybrid backhaul network is a logical option for fluent enhancement of the already existing infrastructure. The Ethernet provides connectivity from eNB elements towards the MME and S-GW of EPC whilst the combination of the TDM and Ethernet provides connectivity from 2G BTS and 3G NodeB elements to the BSC and RNC, respectively.

If there is no full IP transport in the interface between the base station site and controller (in case of 2G and 3G) or S-GW/MME (in case of LTE), there is an alternative solution for base station connectivity. The LTE traffic, together with previous traffic types of 2G BTS and NodeB of WCDMA and HSPA, can actually be delivered over the IP packet infrastructure by using Carrier Ethernet Transport together with a pseudo wire transport concept. This means that if, for example, the Iub interface of 3G RAN is based on ATM transport, the traffic is carried over ATM pseudo wire connections, for example, in such a manner that one connection is reserved for the circuit switched traffic and the other is reserved for packet switched traffic. Similarly, 2G

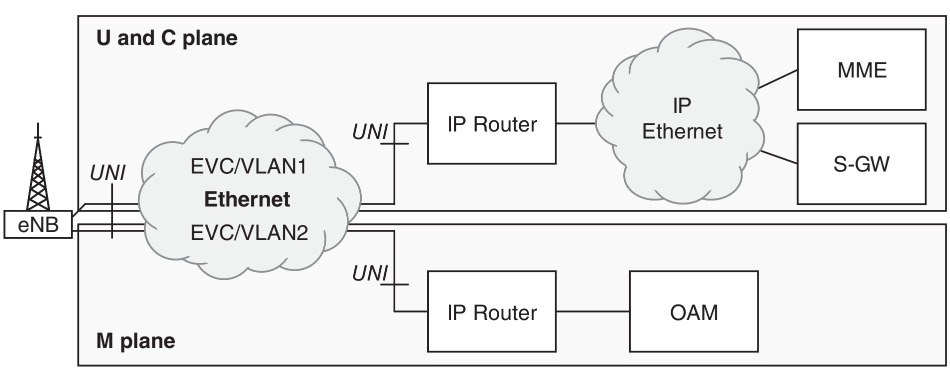

Figure 13.75 *The VLAN ID can be defined separately for different planes.*

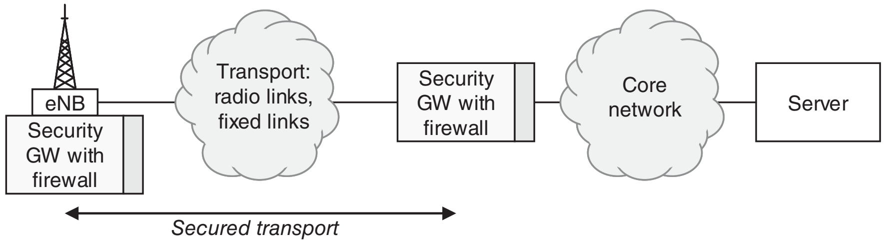

Figure 13.76 The IPSec can be utilized in the transport network.

traffic can be delivered over a TDM pseudo wire connection between BTS and BSC in such a way that TDM signals are carried transparently over the radio access network.

13.27.1 Carrier Ethernet Transport

The Carrier Ethernet Transport (CET) technology can be utilized for the deployment of new backhaul networks. For connectivity, so-called pseudo wire solutions can be applied for the emulation of the TDM and ATM, if native solutions are not available. Figure 13.77 presents the principle of CET.

The CET concept is a cost-effective solution that can replace the traditional time division multiplex (TDM) transport solutions such as SDH/PDH. It is possible to deploy CET for both access and aggregation networks. Basically all of LTE, 3G and 2G traffic types can be delivered over the packet-based backhaul infrastructure. The main benefits of CET solution are: Support of standardized services of variety of physical infrastructures, wide scalability of the bandwidth (from 1 Mb/s to over 10 Gb/s), high reliability, support of

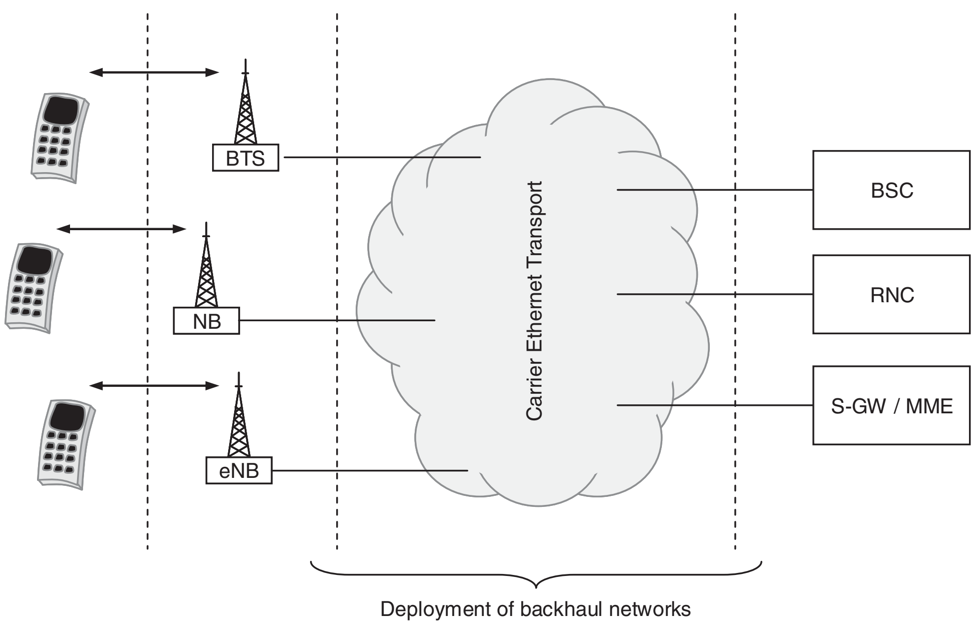

Figure 13.77 The principle of CET.

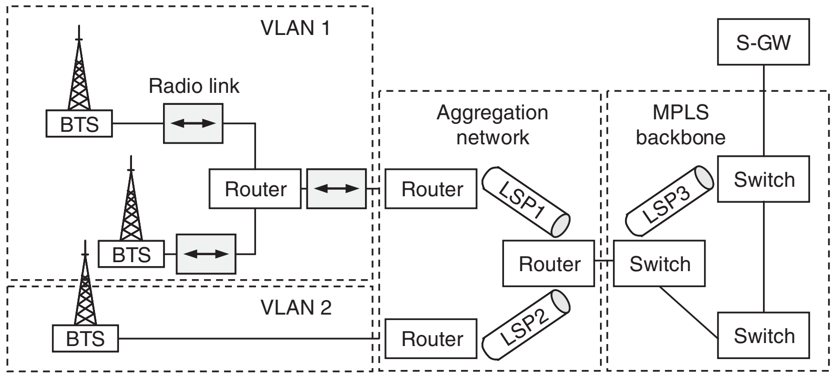

Figure 13.78 *An example of the traffic delivery between LTE eNB elements and S-GW.*

the Quality of Service options. It also offers the possibility to monitor, diagnose and manage the network in a centralized way. CET has been standardized by the Metro Ethernet Forum, so it provides a vendor-independent implementations.

13.27.2 Transport for S1-U Interface

The network between eNodeB elements as well as between the eNodeB and S-GW elements includes typically access network itself, aggregation networks and MPLS (Multi-Protocol Label Switching) backbone network. There can be microwave radio links within the access network in order to provide wireless interconnection especially in the areas where fiber optics is not available. In this way, the handovers between eNodeB elements can be designed offering sufficiently high capacity and low delays.

The LTE *access network* can consist partitions of several Virtual Local Area Networks (VLAN) in such a way that each partition contains one or more eNodeB elements, as shown in Figure 13.78.

The LTE *aggregation network* can be designed by applying, for example, ring topology combined with a Virtual Private LAN Service transport (VPLS). VPLS in turn is based on the MPLS backbone concept. The aggregation network reserves a single Label Switch Path (LSP) for respective single VLAN connection towards the LTE access network as shown in Figure 13.78. The figure also shows the option for the actual MPLS backbone network, that is based on the layer 3 routers installed in a mesh-topology, and is connected to S-GW. Like in aggregation network, a corresponding LSP is utilized also in this MPLS network in order to deliver the IP data traffic between the aggregation network and S-GW.

13.28 Core Network

The logical solution for multiple radio access technologies is to utilize a common packet core concept. This is possible as far as the S4 interface is defined between the SGSN and Serving Gateway network entities.

The common core provides an optimized interworking functionality and Quality of Service handling between the LTE network and non-LTE access networks defined by 3GPP. It handles both LTE and 2G/3G bearers in a similar manner. Furthermore, it provides a common interface with the Home Subscriber Server (HSS). Figure 13.79 illustrates the idea of the common core concept. The QoS of each radio access network can be handled via the common Policy and Charging Rules Function (PCRF).

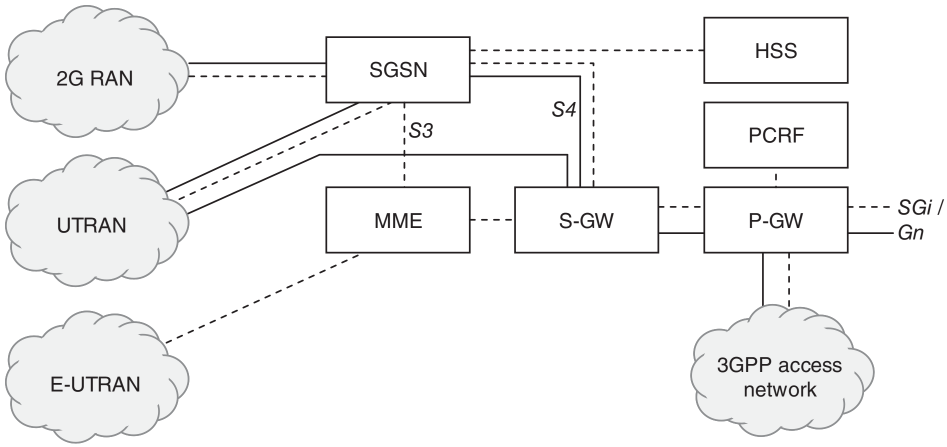

Figure 13.79 *The idea of common core concept, which can be shared between different 3GPP radio access networks.*

The class-based Quality of Service (QoS) concept specified for LTE networks in 3GPP Release 8 provides network operators with effective techniques to enable service or subscriber differentiation at the application level, and to maintain the required QoS level across the end-to-end system.

13.29 Charging

Charging procedures are essential in the commercial telecommunications networks. The functionality includes the creation of the charging data based on the principles that the operator has adopted. There is baseline legislation for the general charging principles in internal network operations as well as for the internetwork connections.

The high-level principle of charging is the collection of the charging data records (CDR) from the connections the users are creating. In the circuit switched plane, the charging is straightforward as a function of time that the connection is active. In packet switched domain, the charging is typically based on the transferred data.

The network elements are able to collect and store various events that can be charged after the CDR has been transmitted to the typically centralized charging system. The format of the CDR depends on the network.

The charging can be done in real-time (online charging) or later (offline charging). In case of online charging, Policy and Charging Enforcement Function (PCEF) is utilized in order to create a real-time interaction with the bearer, session or service control. Online charging is utilized for the prepaid customers, and it offers the possibility for real-time event based or session-based charging. In the case of offline charging, the charging information does not affect the service that is charged, and PCEF is thus not needed. Offline charging relates to postpaid subscriptions, and provides them with an event based or session based charging in such a way that the cost is charged, for example, on a monthly basis. Figure 13.80 clarifies the idea.

13.29.1 Offline Charging

The principle of offline charging is that the CDR arrives at the billing domain after utilization of the network resource. The billing domain then postprocesses the CDR and creates the charging summary to be utilized in the postpaid customer's billing, for example, on a monthly basis.

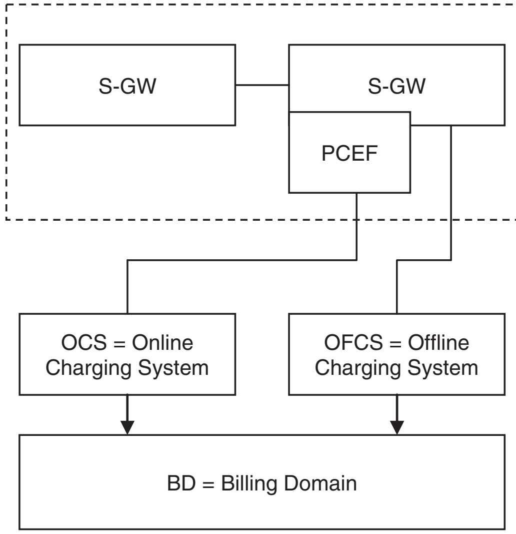

Figure 13.80 *The charging of LTE connections can take place in offline and online modes.*

The PDN Gateway (P-GW) includes a Charging Trigger Function (CTF) which detects the suitable events for the charging purposes. CTF transforms then each of these identified events into separate *charging events*. The charging events are forwarded to the Charging Data Function (CDF) which in turn creates the CDRs with a standard format. CDFs are then forwarded to the Charging Gateway Function (CGF) which compiles the separate CDR information into the single file and sends it to the billing domain. Figure 13.81 shows the data flow of the procedure with the elements and interfaces. It is worth noting that the implementation of the

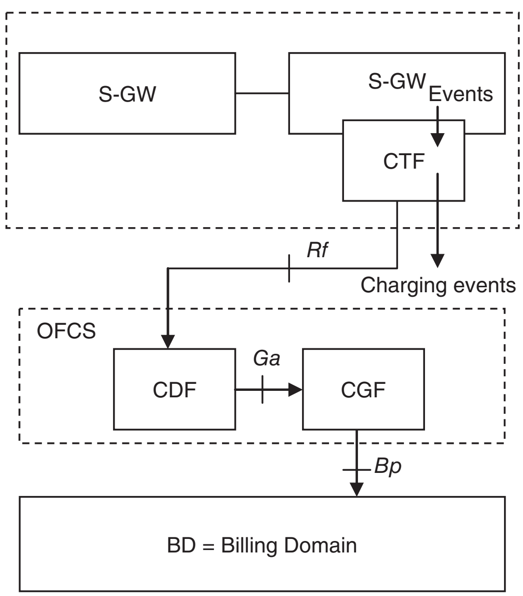

Figure 13.81 *The offline charging.*

charging functions of this procedure is left flexible in the standardization, and they may thus be integrated, for example, into the S-GW element partially or totally.

13.29.2 Charging Data Record

CDR (Charging Data Record) is a set of information about the chargeable events. CDR is formatted in a specific way in order to be understood by the centralized billing domain (BD) where it is sent by the element that collected it. There are various events that can be charged, including bearer utilization like call duration, the amount of the received or sent data and duration for the call setup. The CDRs that LTE generates are based on the packet switched domain. The CDR is transferred via GTP' protocol. The format for packet-based CDR is standardized in the 3GPP technical standard TS 32.251, and GTP' is defined in 3GPP TS 32.295.

13.29.3 Online Charging

The special characteristics of online charging are that the charging information may affect the service in real time. Online charging information is delivered from P-GW to the online charging system (OCS), which in turn takes care of the real-time control of the credits of the users. Figure 13.82 shows the online charging principle.

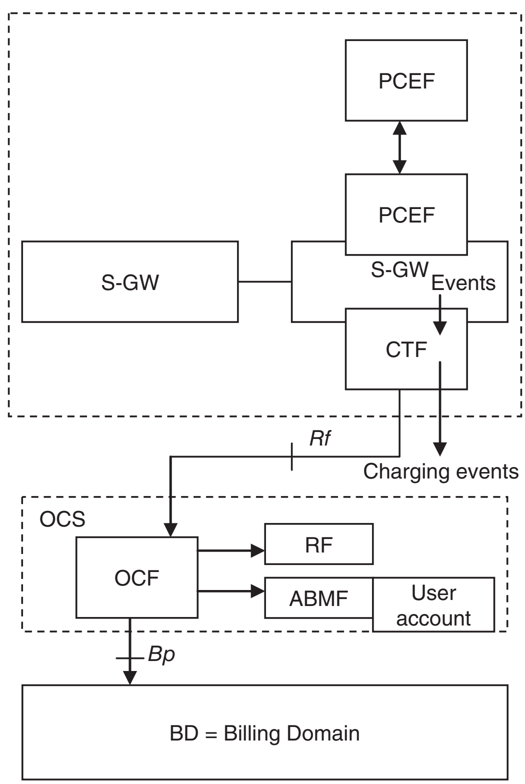

Figure 13.82 *Online charging.*

There is the CTF (Charging Trigger Function) in P-GW, detecting the events that should be charged, that is, events that include bearer resource usage. P-GW then converts these events into charging events, which P-GW forwards to the OCF (Online Charging Function). This function is revised if the network resource can actually still be utilized by UE. To do this, OCF interchanges messages with the ABMF (Account Balance Management Function). The task of ABMF is to store available credits on the user account, and updates this information along the utilization of the credits. OCF also interacts with RF (Rating Function), which contains the means to decide the costs of the services according to defined tariffs. Whilst there are sufficiently credits, OCS allows usage of resources. In contrast, whenever credits are running out, OSC may interrupt the utilization of resources, for example, by terminating calls.

The Gy interface between the Policy and Charging Enforcement Function (PCEF) and the online charging system is defined in 3GPP Technical Specification 23.203. The signaling is based on the IETF Diameter Credit Control Application (DCCA) framework as defined in RFC 4006.

References

1. 3GPP TS 25.913. Technical Specification Group Radio Access Network. Requirements for Evolved UTRA (E-UTRA) and Evolved UTRAN (E-UTRAN), Release 9.
2. Penttinen, J. (2011) *The LTE/SAE Deployment Handbook*, John Wiley & Sons, Ltd.
3. 3GPP, TS 36.331: Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Resource Control (RRC); Protocol specification, 2012.
4. 3GPP, TR 36.806: Evolved Universal Terrestrial Radio Access (E-UTRA); Relay architectures for E-UTRA (LTE-Advanced), 2010.
5. Toskala, A. and Holma, H. (2009) *LTE for UMTS – OFDMA and SC-FDMA Based Radio Access*, John Wiley & Sons, Ltd.
6. 3GPP, TS 36.300: Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2, 2012.
7. 3GPP, TS 36.106: Evolved Universal Terrestrial Radio Access (E-UTRA); FDD repeater radio transmission and reception, 2012.
8. 3GPP, TS 36.143: Evolved Universal Terrestrial Radio Access (E-UTRA); FDD repeater conformance testing, 2012.
9. Mueller, A. and Frank, P. (2010) Cooperative interference prediction for enhanced link adaptation in the 3GPP LTE uplink. VTC Spring.
10. 3GPP, TS 36.213: Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures, 2012.
11. 3GPP, TS 36.216: Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer for relaying operation, 2012.
12. Frank, P. (2010) LTE-Advanced by Prof. Joachim Speidel (University of Stuttgart), Heinz Droste and Philipp Frank (Deutsche Telekom Laboratories).
13. Internet Telephony or Circuit Switched Telephony. http://www2.sis.pitt.edu/~mweiss/papers/itelv3b.pdf (last accessed 28 July 2014).
14. From Circuit Switched to IP-based Networks. http://lyle.smu.edu/~tchen/papers/circuits-to-IP.pdf (last accessed 28 July 2014).
15. 3GPP, TS 32.511: Telecommunication management; Automatic Neighbour Relation (ANR) management; Concepts and requirements, 2012.
16. 3GPP, TS 36.113: Evolved Universal Terrestrial Radio Access (E-UTRA); Base Station (BS) and repeater Electro-Magnetic Compatibility (EMC), 2012.
17. 3GPP, TS 36.211: Evolved Universal Terrestrial Radio Access (E-UTRA); Physical channels and modulation, 2012.
18. 3GPP, TS 36.304: Evolved Universal Terrestrial Radio Access (E-UTRA); User Equipment (UE) procedures in idle mode, 2012.
19. 3GPP, TR 36.902: Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Self-configuring and self-optimizing network (SON) use cases and solutions, 2011.

20. ARTIST4G, D3.5a: Enhancements to Type-1 Relay Implementation, ARTIST4G, 2012 (https://ict-artist4g.eu/projet/deliverables) (last accessed 28 July 2014).
21. ARTIST4G: Advanced Radio Interface Technologies for 4G Systems, https://ict-artist4g.eu (last accessed 28 July 2014).
22. Ramiro, J. and Hamied, K. (eds) (2011) Self-Organizing Networks (SON): Self-Planning, Self-Optimization and Self-Healing for GSM, UMTS and LTE.
23. NGMN Alliance, "Recommendation on SON and O&M Requirements," version 1.1, July 2008.
24. Alamouti, S.M. (1998) A simple transmit diversity technique for wireless communications. *IEEE Journal on Selected Areas in Communications*, **16**(8), 1451–1458.
25. Hamalainen, S., Sanneck, H. and Sartori, C. (2011) LTE Self-Organising Networks.
26. 3GPP TS 36.401. Technical Specification Group Radio Access Network. Evolved Universal Terrestrial Radio Access Network, E-UTRAN), Architecture description, Release 11.
27. Sesia, S., Toufik, I. and Baker, M. (2009) *LTE Umts Long Term Evaluation*, John Wiley & Sons, Ltd.
28. 3GPP, TS 36.420: Evolved Universal Terrestrial Radio Access Network (E-UTRAN); X2 general aspects and principles, 2012.
29. 3GPP, TS 36.421: Evolved Universal Terrestrial Radio Access Network (E-UTRAN); X2 layer 1, 2012.
30. 3GPP, TS 36.422: Evolved Universal Terrestrial Radio Access Network (E-UTRAN); X2 signalling transport, 2012.
31. 3GPP, TS 36.423: Evolved Universal Terrestrial Radio Access Network (E-UTRAN); X2 Application Protocol (X2AP), 2012.
32. 3GPP, TS 36.424: Evolved Universal Terrestrial Radio Access Network (E-UTRAN); X2 data transport, 2012.
33. 3GPP, TS 36.410: Evolved Universal Terrestrial Radio Access Network (E-UTRAN); S1 general aspects and principles, 2012.
34. 3GPP, TS 36.411: Evolved Universal Terrestrial Radio Access Network (E-UTRAN); S1 layer 1, 2012.
35. 3GPP, TS 36.412: Evolved Universal Terrestrial Radio Access Network (E-UTRAN); S1 signalling transport, 2012.
36. 3GPP, TS 36.413: Evolved Universal Terrestrial Radio Access Network (E-UTRAN); S1 Application Protocol (S1AP), 2012.
37. 3GPP, TS 36.414: Evolved Universal Terrestrial Radio Access Network (E-UTRAN); S1 data transport, 2012.
38. Lazhar Belhouchet, M. and Hakim Ebdelli, M. (2010) ITU/BDT Arab Regional Workshop on 4G Wireless Systems. LTE technology. Session 1: Evolution path towards 4G. Technology evolution and regulation. Tunisia, 27–29 January 2010. 40 p.
39. 3GPP TS 36.104 V12.2.0 (2013-12). Technical Specification Group Radio Access Network. Evolved Universal Terrestrial Radio Access (E-UTRA). Base Station (BS) radio transmission and reception, Release 12. 149 p.
40. Góra, J. (2013) Radio Resource Management for Multi-Carrier Relay-Enhance Networks. PhD Thesis, Poznan University of Technology, Poznan.

14
Wireless LAN and Evolution

Jyrki T. J. Penttinen

14.1 Introduction

Wireless LAN (Local Area Network) with a number of variants is currently one of the most popular Internet access methods. Along with the general development of wired Internet access methods and packet core networks, WLAN solutions have also gone through major enhancements. As a result, the bit rate has increased exponentially since 1990s, and the functional area of the networks grows constantly. The first-phase WLAN has been formed by the early IEEE 802.11 standards, which are being complemented constantly.

14.2 WLAN Standards

Table 14.1 presents the most utilized WLAN standards, which are defined via IEEE.

Especially IEEE has been active in defining wireless networks, and there is thus a variety of IEEE WLAN variants. In addition, there are various versions under standardization.

14.3 IEEE 802.11 (Wi-Fi)

The base for the WLAN solutions is formed by the development of Wi-Fi, that is, IEEE 802.11. Its development still continues actively. The most widely spread variants of the wireless networks are currently IEEE 802.11a, 802.11b and 802.11g. In additions to these, almost all alphabets are in use to describe different aspects of WLANS, including the development of WLAN, like optimal routing, enhancements for the security, and dynamic channel selection up to the varying of the quality levels. In general, Wi-Fi standards define interoperable implementations of the IEEE 802.11 Wireless LAN standards certified by the Wi-Fi Alliance.

The Telecommunications Handbook: Engineering Guidelines for Fixed, Mobile and Satellite Systems, First Edition.
Edited by Jyrki T. J. Penttinen.

Table 14.1 The most popular WLAN IEEE 802.11 standards by IEEE as for the access method of large audience

Version	Name	Frequency band	Bit rate (maximum theoretical)
IEEE 802.11 (legacy)	WLAN	2.4 GHz	1 Mb/s–2 Mb/s
IEEE 802.11a	WLAN (Wi-Fi)	5 GHz	54 Mb/s
IEEE 802.11b	WLAN (Wi-Fi)	2.4 GHz	11 Mb/s
IEEE 802.11g	WLAN (Wi-Fi)	2.4 GHz	54 Mb/s
IEEE 802.11n	WLAN (Wi-Fi)	2.4 / 5 GHz	300 Mb/s
IEEE 802.11ac	WLAN (Wi-Fi)	5 GHz	1 Gb/s (total for area) and 500 Mb/s (single station)
IEEE 802.11ad	WiGig	60 GHz (and backwards 2.4 / 5 GHz)	7 Gb/s
IEEE 802.15.1	Bluetooth	2.4 GHz	1 Mb/s
IEEE 802.15.3/3a	UWB	Various bands	10–500 Mb/s
IEEE 802.15.4	ZigBee	2.4 GHz, 915 MHz (America), 868 MHz (Europe)	250 kb/s
IEEE 802.16	WiMAX	10–66 GHz	120 Mb/s
IEEE 802.16a/e	WiMAX	2–11 GHz	70 Mb/s
IEEE 802.20	WMAN/WAN	3.5 GHz	1 Mb/s
IEEE 802.22	Wireless Regional Area Network	VHF/UHF TV bands	

The basic idea behind the definitions of 802.11 standards has been to provide a short-range wireless coverage whereas the 802.16 is meant to be a wide-range system for either fixed areas (comparable to Wireless Local Loop concept, up to about 50 km radius), or to mobile environment providing less coverage. Another important difference between these network types is the possibility of IEEE 802.16 to provide the variation of the quality of service level (QoS) for each user separately. In general, the idea of standardization is to offer systems that support each other in such a way that IEEE 802.11 is suitable for unlicensed bands of WLANs whereas IEEE 802.16 is closer to the WMAN concept (Wireless Metropolitan Area Network) that functions both at licensed and unlicensed bands.

The basic Wi-Fi has been the most popular type of WLAN in both households as well as for the business environment. Along with wider utilization, the expense of the user equipment has not been an issue any more after 2003–2004, and Wi-Fi has thus been a default access method for Internet via home laptops for years.

The principle of Wi-Fi provides various ways to deploy and use the telecommunications services. The most utilized cases are home or company hotspots which are either public or password protected. It is also possible to hide the SSID of the networks. In a shared case, a household or community can obtain Wi-Fi contract in such a way that it is also provided to the neighboring community. In these cases, the contract with the actual access provider may limit the operation technically or commercially. In any case, it is possible that, for example, in a community, there is a single access via xDSL or cable modem for several users. This could be an especially interesting option in remote locations where telecommunications infrastructure is limited, and a single line to service provider facilitates sufficiently high capacity for several simultaneous users. If the antenna of the AP (Access Point) is located correctly, coverage can be over 100 meters.

14.3.1 Wi-Fi Variants

In addition to the most utilized names in the public, that is, IEEE 802.11a, b and g, there are various other supporting standards in this set.

14.3.1.1 IEEE 802.11 Legacy

The original IEEE 802.11 standard was published by IEEE in 1997, which is now called the legacy 802.11 variant. It defines two theoretical data rates of 1 and 2 Mb/s, the physical layer being infrared (IR). Although IR is still included in the IEEE 802.11 standard, the practical implementations are rarer all the time due to advances in the radio interface of WLANs.

The legacy variant has also already defined for the first time the CSMA/CA access technique in order to detect and manage collisions in the multiuser environment in a controlled way. The legacy standard had problems, though, in some cases, for example, with the interoperability of the equipment from multitude of manufacturers. These disadvantages were taken into account in the following IEEE standard 802.11b, which is the first WLAN variant that had a large adaptation in the commercial markets, and is still working as a part of the WLAN equipment.

14.3.1.2 IEEE 802.11a

The IEEE variant 802.11a was standardized in 1999. The technical base, including the protocol layers, is in general the same as in the legacy variant, the main differences being the data rate and frequency band. IEEE 802.11a works in 5 Ghz frequency band. The maximum theoretical data rate of 54 Mb/s is achieved by utilizing 52 OFDM (Orthogonal Frequency Division Multiplex) subcarriers. In practice, the data rate per user in a moderately loaded environment is of the order of 20 Mb/s. The dynamic scalability of the offered ODFM capacity provides the data rate classes of 6, 9, 12, 18, 24, 36, 48 and 54 Mb/s.

Physically, this variant has 8 channels reserved for the WLAN network, and 4 channels for point-to-point connections. As the technology differs from IEEE 802.11 legacy (and its enhanced variant IEEE 802.11b), the IEEE 802.11a is not backwards compatible. In practice, the WLAN equipment typically has support for both IEEE 802.11a and b, allowing the user freedom to select the one they want depending on the support of the AP, the load conditions of the area and so on. The practical maximum radius of the IEEE 802.11a coverage area is in the same order as IEEE 802.11b, around 100–200 meters in LOS (line of sight).

14.3.1.3 IEEE 802.11b

The enhanced variant was ready couple of years alter the legacy standards was published, that is, 1999. In IEEE 802.11b, the maximum theoretical data rate is 11 Mb/s. It still uses the original CSMA/CA as defined in the legacy standard, which provides the practical data rates of around 5–6 Mb/s for TCP and around 7 Mb/s over UDP.

The 802.11b is defined to 2.4 GHz frequency band, which is the widely utilized nonlicensed band.

14.3.1.4 IEEE 802.11c

IEEE 802.11c is the less utilized variant. It is designed for communication between two different types of networks and is thus suitable for interconnecting the communications networks of, for example, buildings in different locations. This variant can also be used over fiber optics. Due to special characteristics of this variant, it is not utilized nor operated directly by end-users. Furthermore, IEEE 802.11c is a modified variant of IEEE 802.11d in such a way that it is possible to utilize OSI layer 2 components of both IEEE 802.11d and IEEE 802 legacy variants.

14.3.1.5 IEEE 802.11d

IEEE 802.11d is an extension to the IEEE 802.11 legacy variant. The aim of IEE 802.11d is to provide international utilization of IEEE 802.11 local area networks in such a way that the data is interchangeable regardless of the frequency ranges each country is permitted to use for WLAN terminals.

14.3.1.6 IEEE 802.11e

The IEEE 802.11e standard provides interoperability between public networks, business networks and household locations. It also adds capacity to previous variants, and introduces new support of multimedia into 802.11b and 802.11a yet maintaining backwards compatibility to these variants. IEEE 802.11e also adds new mechanisms to the MAC layer in order to provide guaranteed QoS.

In addition, IEEE 802.11e includes FEC mechanism (Forward Error Correction). It contains adaptation interfaces for audio and video to enhance control and integration of mechanisms to manage lower range networks. The concept of centralized management and control, which is integrated to QoS mechanisms, lowers the data packet collisions. This in turn increases the packet data rate and enhances response time. As a result, IEEE 802.11e combined with IEEE 802.11 provides real-time traffic.

IEEE 802.11e brings the concept of Hybrid Coordination Function (HCF). It contains two access types which are: EDCA (Enhanced Distributed Channel Access) which is comparable to DCF, and HCCA (HCF Controlled Access) which is comparable to PCF. Furthermore, IEEE 802.11e defines four different types of categories for the medium access: the highest priority Voice (AC_VO), second priority Video (AC_VI), third priority Best Effort (AC_BE), and lowest priority Background (AC_BK).

14.3.1.7 IEEE 802.11f

IEEE 802.11f is a recommendation about access points in order to provide better compliance for equipment interoperability. The recommendation uses IAPP protocol as a basis so that users can roam, that is, switch the connection between different access points without worrying about the differences in functionality of different access point equipment in the network.

14.3.1.8 IEEE 802.11g

IEEE 802.11g is the third version of the actual WLAN access for the public. It was accepted in 2003 to offer an alternative for IEEE 802.11a and b. In practice, it has evolved as a variant of IEEE 802.11b. IEEE 802.11g uses the same nonlicensed band of 2.4 GHz yet providing the same theoretical data rate of 54 Mb/s as in IEEE 802.11a. IEEE 802.11g is backwards compatible with IEEE 802.11b as it supports both DSSS and OFDM. The IEEE 802.11b and 802.11g can thus be utilized under the same cell, but as the 802.11b variant is not based on OFDM, the overall performance of the cell is lowered to a theoretical single data rate of around 20–22 Mb/s in the compatibility case. In the case when both IEEE 802.11b and 802.11g users share the same AP, as the IEEE 802.11b communication does not understand the OFDM collision mechanisms of IEEE 802.11g, the retransmissions increase as a function of the load, which further decreases performance in the compatibility mode. Logically, when IEE 802.11g is utilized in standalone configuration, the problem does not exist.

It should be noted that there are also proprietary variants of IEEE 802.11g in the markets, as the adaptation of the standards was faster than typically. This 802.11g+ provides in theory double data rate of 108 Mb/s, but functions only with the access point and terminal of the same provider due to proprietary protocols.

14.3.1.9 IEEE 802.11h

IEEE 802.11h is an addition to IEEE 802.11, and was published in 2003. As a result of the requirements of ITU and ERO, the aim of IEEE 802.11h is to enhance the performance of IEEE 802.11 under the presence of radars and satellites within the same area. IEEE 802.11h aims thus to minimize the 5 GHz frequency band impacts as there are other important systems near or sharing the band, for example, military solutions. IEEE 802.11h offers solution to enhance performance of IEEE 802.11a in such a way that there are more dynamics in channel selection, as well as for radiated power.

The solutions for these functionalities are DFS (Dynamic Frequency Selection) for avoiding cochannel interferences with radars and for coordinating the use of channels, as well as TPC (Transmitter Power Control) which manages the transmitted power in such a way that interference with satellite systems is minimized.

14.3.1.10 IEEE 802.11i

IEEE 802.11i has been developed to enhance the otherwise vulnerable performance of the WLAN protocol security in the authentication and coding. In practice, IEEE 802.11i introduces new methods to IEEE 802.1x, which are: TKIP (Temporal Key Integration Protocol) and AES (Advanced Encryption Standard) which are implemented under WPA2.

14.3.1.11 IEEE 802.11j

The contents of IEEE 802.11j are basically the same as in 802.11h, but 802.11j is meant specifically to offer a solution for the Japanese regulations to minimize interferences.

14.3.1.12 IEEE 802.11k

IEEE 802.11k further enhances performance of WLAN in such a way that the exchanges and wireless access points can take part in the evaluation of the WLAN radio resources. The idea of IEEE 802.11k is to implement the functionality into SW so that the addition is possible by executing a SW update in the WLAN clients (WLAN cards and other user equipment) as well as the infrastructure (WLAN access points and exchanges).

14.3.1.13 IEEE 802.11n

IEEE 802.11n is meant to offer a new version for the IEEE 802.11 standard. The aim has been to offer realistic data rates of around 300 Mb/s. The high-level goal has been to provide up to 10 times faster networks that is possible to offer via IEEE 802.11a and 802.11g, and up to 40 times faster networks that IEEE 802.11b is capable to offer.

To achieve the goals, techniques like MIMO (Multiple Input Multiple Output) are considered. The standard has been ready since 2008 for implementation, but there have been still further adjustments in the contents. The IEE 802.11n was finally ratified in 2009, with the physical layer maximum data rate of 600 Mb/s. At the moment, there are practical solutions in the commercial market that provide peak data rates of up to 300 Mb/s and average rates of around 100 Mb/s. It can be estimated that due to backwards compatibility (single equipment is capable of connecting all the previous networks) and superior data rate offering, the IEEE 802.11n is becoming the default variant in the WLAN in order to access ADLS in private households as well as the public hotspot areas.

Unlike previous versions, IEEE 802.11n makes it possible to utilize two separate and simultaneous frequency bands, which are the previously defined 2.4 GHz of 802.11b and 802.11g, and GHz utilized in 802.11a. In this way, IEEE 802.11n is backwards compatible with all previous variants of WLAN. The compatibility provides the possibility to optimize the resource utilization which results in maximum data rates due to the still relatively low utilization of the 5 GHz frequency band.

14.3.1.14 IEEE 802.11p

IEEE 802.11p is specifically designed for the automotive environment, in the frequency bands of 5.90 GHz and 6.20 GHz. This variant is meant principally for very short distances (DSRC, Dedicated Short-Range Communications) in the North Americas.

Some of the practical applications of this technology allow the exchange of information between cars, as well as between cars and communications solutions implemented along highways.

14.3.1.15 IEEE 802.11r

IEEE 802.11r aims to provide optimal change of access points in such a way that when the handover is executed, the authentication procedures are already performed in advance in the new access point whilst the communication is still continuing in the old access point. After the security procedures are ready in the new access point, the rest of the handover is done with fluent continuum of the actual data transmission. For this reason, IEEE 802.11r is also known as Fast Basic Service Set Transition.

The IEEE 802.11r functionality provides very low breaks in connection, with less than 50 millisecond transition time between the nodes. This is sufficiently fast for the supporting of VoIP type of services as the break is practically not noticeable for the end-user.

14.3.1.16 IEEE 802.11v

IEEE 802.11v provides means for the centralized configuration and management of the terminals, making the functionality of WLANS a bit similar as in the case of mobile networks. The functionality is provided via the OSI layer 2 mechanisms. In addition, the IEEE 802.11v provides mechanisms for energy saving of handheld VoIP devices, localization of devices, timer functionality for precise calibration applications, and coexistence that optimizes interference levels of different wireless technologies that are supported in the same device.

14.3.1.17 IEEE 802.11w

The aim of IEEE 802.11w is to enhance the medium access control plane of IEEE 802.11 so that the security level of the authentication and coding protocols is increased. In the basic WLAN solutions, the information is sent over the radio interface as such in the frames without additional protection, which gives possibilities for fraudulent activities. IEEE 802.11w aims to protect the networks against malicious attacks that imitate the original packets of the valid users. The intention of IEEE 802.11w is in fact to enhance functionality of IEEE 802.11i to also cover the management frames. This extension has implications in the interaction with IEEE 802.11r as well as with IEEE 802.11v.

14.3.1.18 IEEE 802.11ac

The next step in the IEEE 802.11 development results in IEEE 802.11ac. The aim of this standard is basically to replace IEEE 802.11n. The goal of IEEE 802.11ac is to offer theoretical data rates of at least 1 Gb/s in multistation environment, and at least 500 Mb/s for a single station.

IEEE 802.11ac is a wireless computer networking standard based on 802.11. The technology is still under development, the standards approval is expected to happen by the end of 2013. IEEE 802.11ac functions at WLAN 5 GHz band. It can be expected that equipment with IEEE 802.11ac support will be commonly available in 2015.

In order to achieve planned performance values for the data rate, IEEE 802.11ac extends the air interface functionality of 802.11n by providing frequency bandwidth up to 160 MHz. IEEE 802.11ac standards define a mandatory 80 MHz channel bandwidth for stations (STAs) whereas IEEE 802.11n provides a maximum of 40 MHz. Furthermore, IEEE 802.11ac has an optional 160 MHz bandwidth available. IEEE 802.11ac also gives possibility to apply a maximum of 8×8 MIMO antennas whereas IEEE 802.11n defines a maximum of 4×4 MIMO. Multiuser MIMO (MU-MIMO) can also be utilized in IEEE 802.11ac. The modulation scheme is selected dynamically, and the fastest data rates are possible with optional 256-QAM combined with code rates of $^3/_4$ and 5/6. As a comparison, IEEE 802.11n provides maximum data rates up to 64-QAM with code rate of 5/6.

As other additional features, IEEE 802.11ac is capable of beam forming with standardized sounding which provides vendor compatibility. This functionality has not standardized in 802.11n which makes the functioning of beam forming challenging in a multivendor environment. There are also enhancements to the MAC layer, and coexistence mechanisms are included to take into account the channel bandwidths of 20, 40, 80 and 160 MHz so that there is possibility for backwards compatibility with IEEE 802.11ac, 802.11a and 802.11n equipment.

The mandatory and optional features of IEEE 802.ac equivalent to the ones in IEEE 802.11a and 802.11n are the following:

- 800 ns guard interval, mandatory
- BCC (Binary Convolutional Coding), mandatory
- Single spatial stream, mandatory
- 2 to 4 spatial streams, optional from IEEE 802.11n
- Low-density parity-check code (LDPC), optional from IEEE 802.11n
- Space-Time Block Coding (STBC), optional from IEEE 802.11n
- Transmit Beam forming (TxBF), optional from IEEE 802.11n
- 400 ns Short Guard Interval (SGI), optional from IEEE 802.11n.

The mandatory and optional features that are new in IEEE 802.11ac are the following:

- 80 MHz channel bandwidth, mandatory
- 5 to 8 spatial streams, optional
- 160 MHz channel bandwidths (continuous 80 + 80), optional
- 80 + 80 MHz channel bonding (discontinuous 80 + 80), optional
- MCS 8/9 (256-QAM), optional.

In the beginning of the commercial phase, the IEEE 802.11ac capable devices are wireless routers and access points as well as external wireless cards based on PCI or USB. In later phases, IEEE 802.11ac will be supported also in-built to laptops, tablets and smartphones. The commercial phase begun with the introduction of first equipment by the end of 2012, and in larger scale by mid-2013.

14.3.1.19 IEEE 802.11ad (WiGig)

The Wireless Gigabit Alliance [1] is also known as the WiGig. The organization promotes the adoption of wireless communications technology operating at the unlicensed 60 GHz frequency band and which is capable of handling several gigabit per second of data rates. This technology is called IEEE 802.11ad, or WiGig.

The history of WiGig initiated in 2009 when the first phase of technology plan was completed, and the specification was published in 2010. By the publishing of the version 1.0 of the specification, WiGig organization also informed about the opening of the Adopter Program, and about the liaison agreement with the Wi-Fi Alliance for the cooperation of Wi-Fi technologies. The version 1.1 that includes certification procedure was published later in 2011.

The aim of WiGig is to provide high bit rate and high performance wireless data transmission. It also promotes display and audio applications that are suitable for modern wireless LAN devices.

The most advanced variant of the WiGig devices support all the 3 frequency bands of WLAN technologies, that is, 2.4, 5 and 60 GHz bands. As a result of the advanced technology, WiGig is capable of providing data rates up to 7 Gb/s in such a way that it still is backwards compatible with the previous WLAN variants.

The version 1.1 of WiGig MAC and PHY specifications include the following items:

- Support of data rates up to 7 Gbit/s
- Complements and extends the 802.11 Media Access Control (MAC) layer in such a way that it is backward compatible with the IEEE 802.11 standard
- Physical layer enables low power WiGig devices
- Protocol adaptation layers will support system interfaces such as data buses for PC peripherals and interfaces for HDTV displays
- Support for beamforming, which provides high-performance communication over 10 meters of distance
- Definition of high-performance wireless implementations of typical computer interfaces over 60 GHz
- Enabling of high-speed wireless connectivity between any two devices.

WiGig technology overlaps partially with WirelessHD which is an industry-led initiative to specify a new digital network interface for wireless high-definition signal transmission for consumer electronics products.

14.3.2 Wi-Fi and Other Wireless Networks

The main categories of wireless networks are seen especially via IEEE 802.11, 802.16 and 802.20 systems. Table 14.2 presents the most important differences between these definitions.

IEEE 802.20 aims to provide wireless access in metropolitan environment in such a way that its principal comparable technologies are xDSL and cable modem with the data rate of over 1 Mb/s. The maximum useful radius of the radio interface of 802.20 is over 15 km. An important differentiator of IEEE 802.20 compared to previous variants is its support for high terminal speed. This is suitable for reasonable commutation environments up to, for example, midlevel bullet-trains.

Instead, the maximum terminal speed of IEEE 802.16 is suitable for lower speeds, but nevertheless they are also suitable for all practical highway speeds as the speed is defined for 120–150 km/h.

In general, IEEE 802.16 and 802.20 offer at least partially the same type of platform for wireless access customers, but the differences are important enough to justify the parallel existence of the technologies in the commercial markets. More specifically, IEEE 802.16e is meant to be used in PDA and smartphone use within city coverage areas whilst IEEE 802.20 is suitable up to bullet-train type of environments.

Figure 14.1 presents typical environments for the main IEEE 802 standards.

Figure 14.2 presents the relationship and evolution between different wireless IEEE standards. For each network type, there is an optimal environment for the usage. The role of WiMAX would be most suitable in a metropolis type of networks in wide areas.

Table 14.2 *Main differences between some of the IEEE WLAN standards*

	802.11 (Wi-Fi)	802.16 (WiMAX)	802.20 (WMAN)
Frequency bands	2.4 / 5.8 GHz (nonlicensed)	10–66 / 2–11 GHz (nonlicensed and licensed)	<3.5 GHz (licensed)
Mobility management	Portable	Mobile and fixed	Fast mobile (max. 250 km/h)
Requirements for radio interface	NLOS	LOS (10–66 GHz), NLOS (2–11 GHz)	NLOS

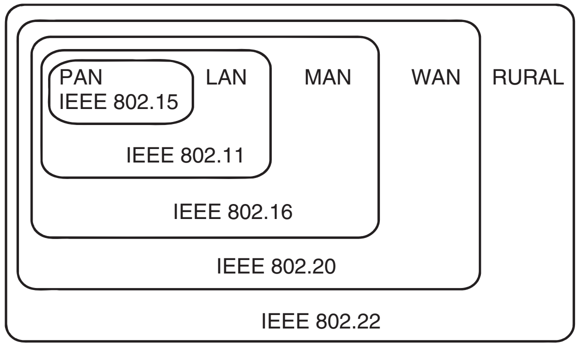

Figure 14.1 *The role of different IEEE 802 variants.*

14.3.3 Security Aspects

Similarly as with the mobile networks, the security is an important aspect also for WLAN environment both in household and business use. In practice, if the radio interface of the WLAN is not protected by password, it is basically available for public use. In the works case, the AP can be thus used for illegal purposes. Setting up the server as a delivery element for music or video contents without the permission of the content owners is only one of the countless examples how the open access can be misused.

The most elemental and easy option for increasing the security level of the WLAN networks is to activate access code. WEP has been in use for the initial WLAN access restriction, and it has been enhanced via

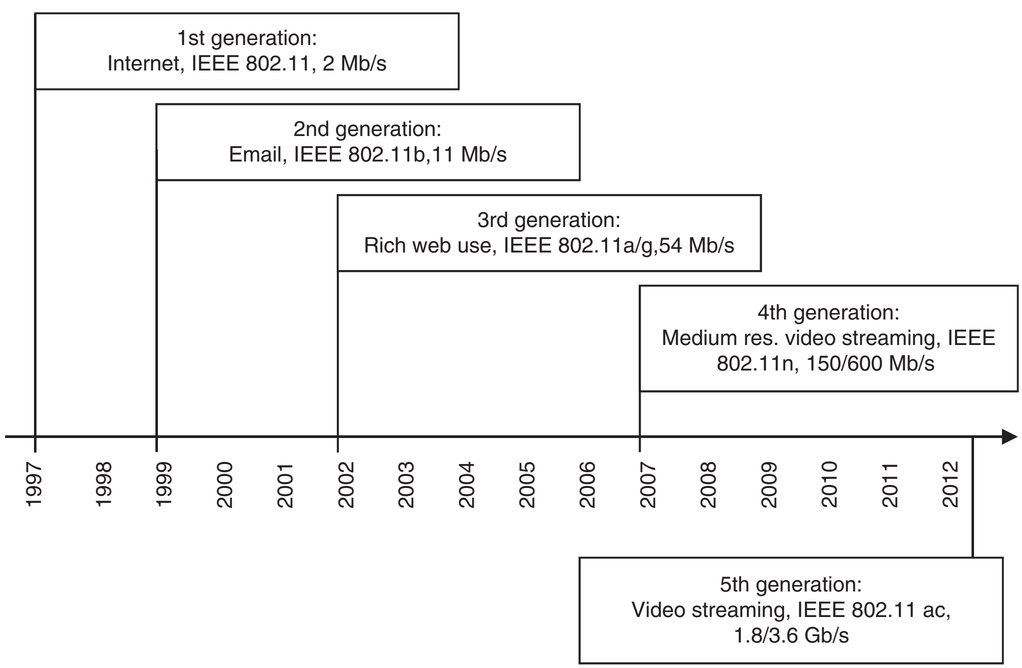

Figure 14.2 *The evolution of Wi-Fi.*

several versions. Other solution is to hide the SSID of the network in such a way that in order to access the WLAN, the user must know the SSID. In addition to the actual Wi-Fi standards, there also exist various solutions in equipment and application level.

In all the cases, regardless of the access point offered via public Wi-Fi hotspot, mobile networks or fixed Internet access, the security level of the applications can be increased by using VPN (virtual private network). This provides a point-to-point protection so that the user data cannot be monitored via the parallel equipment connected to the same access point.

14.4 IEEE 802.16 (WiMAX)

WiMAX (Worldwide Interoperability for Microwave Access), that is, IEEE 802.16 is a wireless communications standard defined by IEEE. It provides rata rates of about 30–40 Mb/s. The enhancements published in 2011 have increased the maximum theoretical data rate up to 1 Gb/s for fixed stations, which complies with the 4G requirements of ITU-R.

The term WiMAX is a result of the work of the WiMAX Forum. The Forum was initiated in 2001 in order to promote conformity and interoperability of the standard. According to the WiMAX Forum, WiMAX is a standards-based technology that enables the delivery of last mile wireless broadband access as an alternative to cable and DSL.

The IEEE 802.16e-2005 added Scalable-Orthogonal Frequency Division Multiple Access (S-OFDMA) to the previous definition of IEEE 802.16-2004 Air Interface Standard which, together with other features for support of mobility, created the basis for WiMAX System Release 1. Further enhancements were added via IEEE 802.16e-2009 [2].

The first WiMAX System Release 1 deployments took place in 2006. According to WiMAX Forum WiMAX service providers have covered more than 600 million people with over 500 deployments in over 140 telecommunications markets worldwide. The air interface enhancements approved for WiMAX, designated as WiMAX Air Interface Release 1.5, that is, Air Interface R1.5, were ready for certification testing in 2010 [3], and the advanced version, Release 2.0 is already compliant with the ITU-R performance requirements of 4G.

The basic idea of IEEE 802.16 is to offer a wide area WLL (Wireless Local Loop), which is defined to function within 10–66 GHz frequency band. The maximum data rate of basic version of 802.16 is 120 Mb/s. The actual WiMAX is defined in IEEE 802.16-2004, or more officially, 802.16 REVd, and it is designed for fixed usage. The enhanced version of WiMAX called Mobile WiMAX is defined in IEEE 802.16e (IEEE 802.16-2005) and it also includes methods for mobility management. It supports portable and mobile communication methods. WiMAX frequency range is in 2–11 GHz. The maximum data rate of WiMAX is 70 Mb/s.

The portable communication category means in practice an extension of the fixed communication in such a way that the connection can be established from various access points. Nevertheless, the portable category does not support seamless handover between access points. This possibility of changing access points for communications is called nomadic access method. It requires initiation of the connection separately via each access points.

The mobile communications category in the WiMAX environment means that the communications channel can be changed relatively quickly so that the amount of lost packets is minimal during the procedure. The mobile category WiMAX thus practically supports real time VoIP calls in mobile environment.

The standardization of IEEE 802.16 has been divided into various subgroups just as is the case with 802.11. 802.16 Maintenance Task Group takes care of the functioning of the service definitions and corrects upon need the definitions. 802.16e takes care of mobility management. It concentrates on the enhancements of MAC layer, including procedures for channel change and calls, as well as for scalability aspects of OFDMA (Orthogonal Frequency Division Multiplex Access). 802.16f takes care of information management of fixed services, whereas 802.16g handles management procedures and services, and 802.16h handles permissions. In general, WiMAX defines interoperable implementations of the IEEE 802.16 family of wireless-networks standards ratified by the WiMAX Forum.

It should also be noted that Bluetooth and UWB (Ultra Wide Band) defined by IEEE 802.15 can act as wireless networks, although the first Bluetooth deployments and definitions have been based on pairing the communications part between two devices, like mobile phone and Bluetooth head set.

The IEEE:n 802 series has addition from 2004, 802.22, which extends the functional areas of the previously defined systems. It is meant especially for the rural and very wide areas with low population density. It is a broadcast technology which would be suitable for TV broadcast type of contents. The term for 802.22 is thus cognitive radio.

14.4.1 WiMAX Standardization

WiMAX can be classified to FWA (Fixed Wireless Access), especially in the initial phase of the deployments. Both ETSI and IEEE have been active in the specification work of the system.

The WiMAX definitions have been taken care of by IEEE and WiMAX Forum. The latter has a separate subgroup designing regulation, that is, RWG (regulatory working group). This group has a goal to facilitate the functioning of WiMAX solutions in such a way that they are not interfering with other existing systems. The special task of the group is to take care of planned frequency band allocation for WiMAX equipment and to identify additional frequency bands.

14.4.2 WiMAX Frequencies

The first phase WiMAX frequencies are 3.400–3.600 GHz (for both TDD and FDD modes) and 5.725–5.850 GHz (TDD-mode). These frequencies have been defined in 2004. Nevertheless, the additional allocation of the frequencies is important for the future functioning and expansion of the system.

The Radio communication Sector of the International Telecommunication Union (ITU-R) has decided in 2007 that WiMAX technology is included in the IMT-2000 set of standards. This facilitates the use of the respective spectrum owners especially at 2.5–2.69 GHz band in order to use WiMAX equipment in countries that recognize the IMT-2000.

The allocation of frequencies depends on the country basis even within the European Union, which means that the uniform utilization of the system is not possible in all the regions. As an example, the availability of frequency band of 5.8 GHz is only limited in Europe. CEPT has defined 3.5 GHz as a preferred and for the WiMAX deployments, but also this band needs to be planned and deployed block-by-block basis.

The additional allocations of WiMAX are possible for 2.500–2.690 GHz which is licensed, for example, in the USA, Mexico, Brazil, and Philippines. The regulation subgroup of WiMAX Forum has also identified other possibilities for extensions in frequency bands of 2.3–2.4 GHz, 3.3–3.4 GHz and 3.6–3.8 GHz. WiMAX Forum is also seeking possibilities to have blocks from below 1 GHz frequencies, for example, 700–800 MHz. These frequencies are especially suitable for the large coverage areas due to the propagation characteristics of the bands.

14.4.3 Technology for WiMAX Deployments

WiMAX is based on OFDM multicarrier technology. OFDM is highly tolerant against the problems of wireless interfaces, which means that WiMAX also functions partially in NLOS environments (nonline of sight). The possibility of having obstacles in the communication path greatly increases the utilization of WiMAX in the mobile environment.

The additional benefit of OFDM is that it allows flexible bandwidth utilization for the connections. The first phase WiMAX solutions have been using bandwidths of 3,5 MHz, 7 MHz and 10 MHz. From the three physical layer solutions of WiMAX, especially the scaling OFDM is most suitable for practical solutions. In general, OFDM provides the scaling of WiMAX bands in the range of 1.75–20 MHz. As is characteristic to OFDM systems, also WiMAX includes both FDD (Frequency Division Duplex) and TDD (Time Division Duplex) components.

Unlike the previous wireless networks, WiMAX is meant as a platform for systems that use long-haul connections. The radiating power of 3.5 GHz band of WiMAX does not have specific limitations apart from regulative ones, but the deployment of 5.8 GHz band may have some problems due to the power limitations of some countries. Combined with the lower coverage area due to higher frequency, the useful areas of this band might be limited to less than 5 km whereas the theoretical radius is in the range of several tens of km. With the lowest power levels WiMAX requires thus LOS, as the buildings and other obstacles create too much attenuation for functional connections.

The Medium Access Control (MAC) of WiMAX includes methods for broadcast and multicast transmissions (point to multipoint) as well as QoS (Quality of Service) definitions. The broadcast of WiMAX is based on CSMA/CA (Collision Sense Multiple Access/Collision Avoidance) and packet scheduling of different transmission directions.

The WiMAX service has been defined for 4 quality classes which are: UGS (unsolicited grant service), rtPS (real time polling service), nrtPS (non-real time polling service) and BE (best effort).

Other technical aspects related to WiMAX are header compression, encapsulation and fragmentation of packets, advances security measures like PKM (Privacy Key Management) and EAP (Extensible Authentication Protocol), procedures for fast channel changes and variation of power levels (normal, sleep, idle).

The technical development of WiMAX is taken care of in WiMAX Forum TWG (Technical Working Group). Its most important tasks are the assurance of interoperability with other systems as for the interference levels and interworking functionality. It also takes care of application development in such a way that the system also functions in the future.

14.4.4 Architecture of WiMAX

WiMAX has been designed in such a way that it provides hotspots as in previous WLAN versions. Furthermore, the coverage areas of WiMAX can be together with Wi-Fi hotspots a part of 2G, 3G and 4G service area so that each one complements the service availability. The seamless handover between the mobile networks and WiMAX is part of the Offloading concept.

The basic idea of WiMAX is to facilitate the interworking between mobile networks and IP networks. WiMAX supports the evolution of IPv4 and IPv6 as well as the real-time and non-real-time applications via TCP and UDP. The principles also include the support of cooperation between 3G and Wi-Fi in one or another way. Figure 14.3 clarifies the idea.

WiMAX standards include several interfaces and functionalities for the evolution path. The architecture of WiMAX consists of three main blocks which are: 802.16 SS/MSS (Subscriber Station/Mobile Subscriber Station, i.e., the user), 802.16 RAN (Radio Access Network) and interface for the connection towards other systems (I, interoperability). Users connect to WiMAX system via radio interface and APs (access point).

Mobility management includes methods for channel changes of WiMAX. In the simplest form, the active connection can be handed over between the sectors of a single AP. In the nomadic use case, the user can swap the connection point but only in such a way that the connection must be initiated again.

Figure 14.4 shows the idea of WiMAX for replacing “last mile” connectivity. Equally, WiMAX coverage areas can be overlapping and giving additional service within the coverage areas of 2G, 3G, 4G and Wi-Fi.

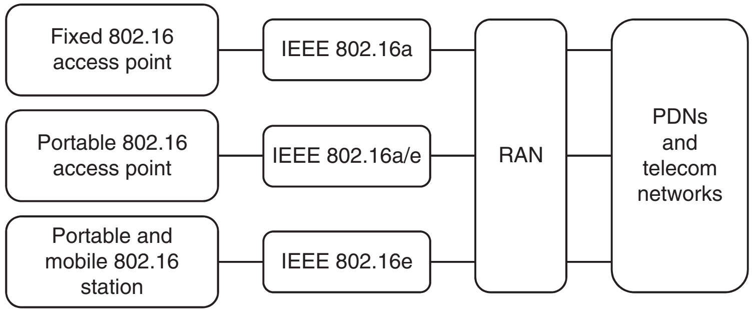

Figure 14.3 *The connectivity of WiMAX can be done according to WLL in fixed, portable or mobile categories.*

WiMAX variant IEEE 802.16e is also suitable for mobility management which provides seamless handovers between cells.

14.4.5 Marketing Aspects

The marketing section of WiMAX Forum has created basic principles for promoting WiMAX technology. One of the potential dangers in marketing is the creation of hype as was noted in the initial phase of UMTS. WiMAX as such is a logical part of the complete picture to support other telecommunications networks, and its performance is optimal especially for fixed, long-distance wireless solutions yet also providing mobility.

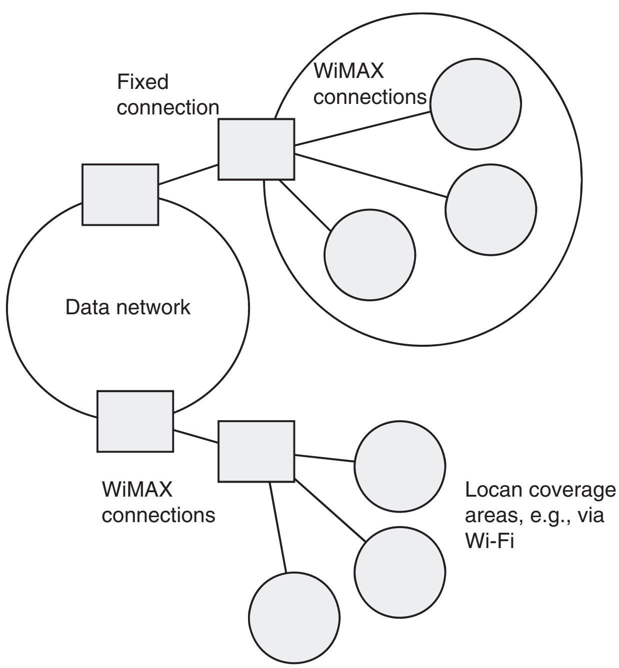

Figure 14.4 *WiMAX deployment can be done in various manners. One option is to use WiMAX to replace the "last mile" for each user separately, or by using WiMAX as one part of the transmission network.*

For the latter, WiMAX provides additional service areas to the 3G and 4G mobile networks, although the current market seems to favor the latter as a default method for mobile communications, especially along with the deployment of LTE, and its evolved variant, LTE-Advanced. In any case, the aim of WiMAX Forum is to promote WiMAX as a significant part of mobile communications, in such a way that WiMAX offers high-quality service levels for demanding applications.

Mobile WiMAX was a replacement candidate for cellular phone technologies such as GSM and CDMA, or can be used as an overlay to increase capacity. Fixed WiMAX is also considered as a wireless backhaul technology for 2G, 3G, and 4G networks in both developed and developing nations [1, 4].

WiMAX is a suitable technology to enhance the backhaul connectivity in areas where the transmission is provided via copper wire line connections satellite and microwave links. WiMAX provides substantially higher bandwidth than legacy cellular applications require.

The architecture for the connection of WiMAX network with an IP based core network has been designed in the WiMAX Forum. This case is typical for operators that are Internet Service Providers (ISP). On the other hand, the WiMAX BS facilitates integration with other types of architectures as with packet switched Mobile Networks.

The WiMAX Forum-based solution includes a set of components and reference points or interfaces between the elements (R1-R5 and R8). The elements are:

- SS/MS (Subscriber Station/Mobile Station)
- ASN (Access Service Network)
- BS (Base station) which is part of the ASN
- ASN-GW (ASN Gateway) which is part of the ASN
- CSN (Connectivity Service Network)
- HA (Home Agent) which is part of the CSN
- AAA (Authentication, Authorization and Accounting Server) which is part of CSN
- NAP (Network Access Provider)
- NSP (Network Service Provider).

It is possible to design the functional architecture into several hardware configurations. The architecture allows remote and mobile stations of varying scale and functionality. Also the Base Stations can have different sizes and can thus be, for example, femto, pico, and macro cells. Figure 14.5 shows the WiMAX architecture of WiMAX Forum.

14.4.6 Applications

WiMAX is a logical part in the complete WLAN offering. It is suitable especially in WLL type of environments and also in highly challenging conditions and in long distances. One example of this is the deployment of communications lines to remote areas where public telecommunications infrastructure is lacking capacity or service areas. Some logical deployment for WiMAX can be considered, for example, in some Latin American, African and Asian rural areas which are otherwise difficult to access and that have been covered previously by special solutions like GSM-based public telephones up to VSAT satellite phones and Internet connections. WiMAX could thus provide a relatively straightforward manner to offer Internet access and IP based voice calls in areas that might not have otherwise any infrastructure for fixed telephony, mobile networks nor Internet access.

The vision of WiMAX Forum is to offer the system for data and video transmission both in business and private environments. The potential WiMAX operators are found most logically amongst existing fixed line operators, new local loop operators, wireless Internet providers as well as current mobile communications operators.

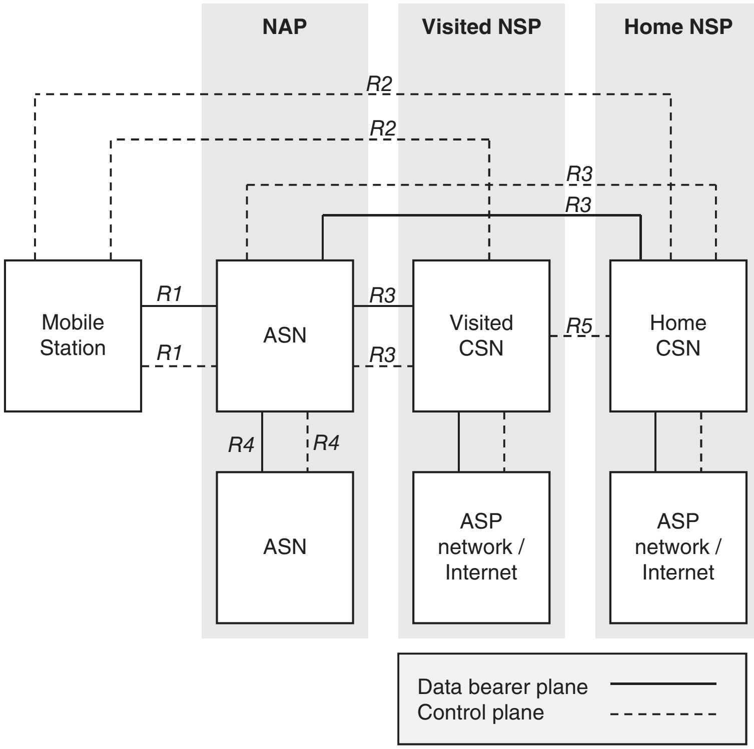

Figure 14.5 *The WiMAX architecture as defined by WiMAX Forum.*

14.5 Evolved IEEE 802.16 (4G)

14.5.1 General

The Mobile WiMAX (IEEE 802.16e-2005) mobile wireless broadband access (MWBA) standard is also known as WiBro in South Korea. In the commercial offering, Mobile WiMAX is sometimes referred to as 4G technology. As it provides practical peak data rates of 128 Mb/s in downlink and 56 Mb/s in uplink by using 20 MHz bandwidth, it does not fulfill in its basic version the performance requirements of ITU-R for the IMT-4G systems. According to the ITU-R principles on the matter, IMT-Advanced (International Mobile Telecommunications Advanced) cellular systems can be called as 4G if they are able to deliver a peak download speed of up to 100 Mb/s in a high mobility environment and up to 1 Gb/s in a low mobility environment [2].

Nevertheless, in addition to LTE-Advanced, also WiMAX2 (802.16m), known also as Wireless MAN-Advanced solutions do qualify as 4G. Both these technologies are capable of theoretical peak performance dictated by ITU-R. The development of both technologies is in progress. The improvements of IEEE 802.16m include support for new service classes, enhanced mobility and guaranteed Quality of Service (QoS).

The goal of IEEE 802.16m is to provide an advanced air interface for licensed bands. The technology is planned to be backwards compatible with legacy devices of earlier variants of the 802.16 standard whilst the performance is notably higher for the new devices. 802.16m devices represent various categories, which are: AMS (Advanced Mobile Stations), ABS (Advanced Base Stations) and ARS (Advanced Relay Stations). They all support 802.16m. The legacy devices for IEEE 802.16m are called R1MS (Revision 1 mobile station) and R1BS (Revision 1 base station) [5].

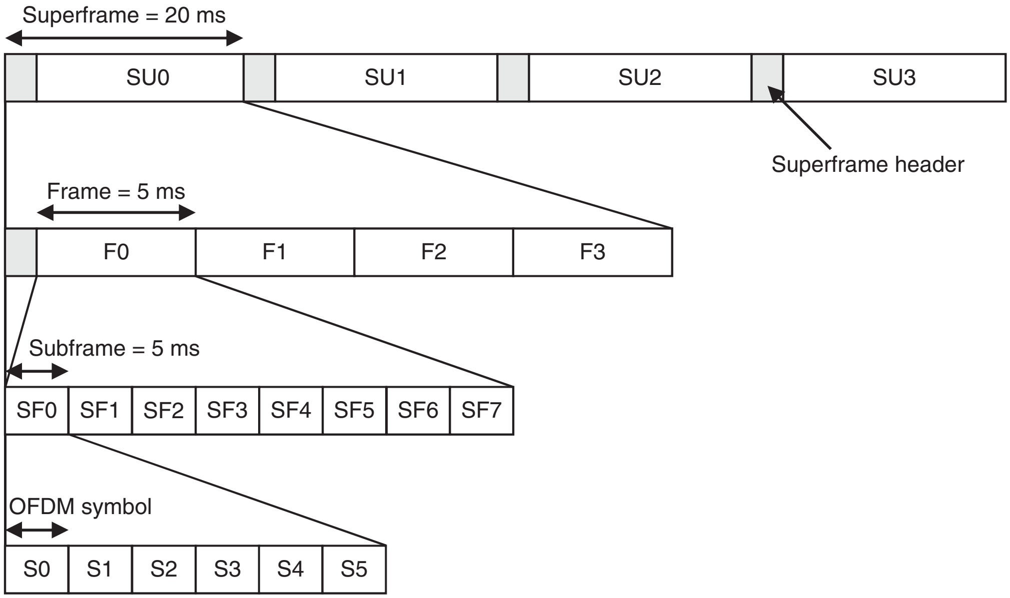

Figure 14.6 *The frame structure of IEEE 802.16m. This example shows the Type 1 subframe.*

IEEE 802.16m has been planned to support frequency bands of 450–470 MHz, 698–960 MHz, 1710–2025 MHz, 2110–2200 MHz, 2300–2400 MHz, 2500–2690 MHz, and 3400–3600 MHz. It should be noted that from this set, the bands of 450–470 MHz, 1710–2025 MHz, and 2110–2200 MHz bands are new in IEEE 802.16 standards.

Figure 14.6 shows the cyclic frame structure of IEEE 802.16m. The structure consists of superframes, frames, subframes and OFDM symbols.

The duration of superframe is 20 ms. The superframe contains four consecutive frames of 5 ms each, filling the 20 ms superframe. Furthermore, each frame is structured to 8 consecutive subframes. There are two exceptions, though, for the 7 MHz bandwidth that can contain 6 subframes, and for the 8.75 MHz bandwidth that can contain 7 subframes.

The subframe can have one of the four types in such a way that the Type 1 has 6 ODFMA symbols, Type 2 has 7 symbols, Type 3 has 5 symbols, and Type 4 has 9 symbols. Types 1, 2 and 3 belong to the actual IEEE 802.16m standards, and Type 4 is meant for the legacy operation of 802.16-2009 devices that use 8.75 MHz bandwidth.

One of the most important targets of IEEE 802.16m is to minimize latency for the complete system. The enhancements are applied to, for example, delay of radio interface, state transition delay, access delay and handover interruption time. This enhances and makes it possible to offer guaranteed QoS for the IMT-Advanced services. According to Ref. [6], some of the key latency objectives are:

- Link layer and user plane: less than 10 ms for both DL and UL.
- Handover interruption: less than 30 ms.
- Control plane, stage change from idle to active: less than 100 ms.

These values are suitable for, for example, high-quality VoIP connections.

Table 14.3 The peak and average spectral efficiency figures of IEEE 802.16m for low user mobility scenario [6]

Performance item	Antenna	Performance value
Peak DL spectral efficiency	• 2×2 MIMO • 4×4 MIMO	• 8.5 b/s/Hz • 17.0 b/s/Hz
Average DL spectral efficiency	• 4×2 MIMO	• 3.2 b/s/Hz, or 0.32 b/s/Hz/user
DL cell-edge user throughput	• 4×2 MIMO	• 0.09 b/s/Hz/user
Peak UL spectral efficiency	• 1×2 SIMO • 2×4 MIMO	• 4.6 b/s/Hz • 9.3 b/s/Hz
Average UL spectral efficiency	• 2×4 MIMO	• 2.6 b/s/Hz, or 0.26 b/s/Hz/user
UL cell-edge user throughput	• 2.4 MIMO	• 0.11 b/s/Hz/user

Source: Data by courtesy of WiMAX Forum.

IEEE 802.16m provides support for FDD (Frequency Division Duplex) and TDD (Time Division Duplex) modes. Also MIMO concept (Multiple Input Multiple Output) is supported as a mandatory functionality, although the legacy mode can use single antenna configuration. ABS device must support minimum 2×2 MIMO, and can optionally support for up to eight transmit antennas. AMS device must support at least 1×2 MIMO. IEEE 802.16m allows the utilization of SU-MIMO (Single User MIMO) and MU-MIMO (Multi User MIMO).

The achieved data rate depends heavily on the selected MIMO configuration. Table 14.3 shows the target performance of some MIMO cases and Table 14.4 presents a summary of the expected performance figures of MIMO in IEEE 802.16 m.

Table 14.5 compares the target performance of WiMAX 1.0 and WiMAX 2.0 [7]. Furthermore, Table 14.6 shows the target performance of IEEE 802.16m according to Ref. [3].

14.5.2 Impacts of IEEE 802.16m on Network Planning

According to Ref. [2], IEEE 802.16m provides an improvement in the link budget over WiMAX

System Release 1 of at least 3 dB when the same antenna configuration is utilized. This means that there may be a 20–30% increase in cell coverage area in a typical non-line-of-sight environment. If the same area is considered instead, IEEE 802.16m enhances the radio link budget by providing increased cell edge user throughput. According to Ref. [2], this may result in a two times improvement over WiMAX System Release 1.

IEEE 802.16m also includes interworking functionalities with other wireless networks at the global level. First, WiMAX System Release 2 of IEEE 802.16m provides backwards compatibility with WiMAX System

Table 14.4 Some of the expected performance figures of MIMO in IEEE 802.16 m

Requirement type	Direction	MIMO	Peak efficiency
Baseline	UL	2×2	8.0 b/s/Hz
	DL	1×2	2.8 b/s/Hz
Target	UL	4×4	15 b/s/Hz
	DL	2×4	6.8 b/s/Hz

Table 14.5 *Comparison of WiMAX 1.0 and WiMAX 2.0*

Item	IEEE 802.16e (WiMAX 1.0)	802.16m (WiMAX 2)
Multiplexing	Time	Time and frequency
Channels	3.5, 7, 8.75, 10 MHz	5, 10, 20 MHz
Typical MIMO scenarios	2 × 2 (DL); 1 × 2 (UL)	2 × 2 / 2 × 4 / 4 × 2 / 4 × 4 (DL); 1 × 2 / 1 × 4 / 2 × 2 / 2 × 4 (UL)
Latency	20 ms (access); 35–50 ms (handover)	<20 ms (access); <30 ms (handover)
Peak spectral efficiency per sector	6.4 b/s/Hz (DL); 2.6 b/s/Hz (UL)	15 b/s/Hz (DL); 6.75 b/s/Hz (UL)
Average spectral efficiency per sector	1.55 b/s/Hz (DL); 0.9 b/s/Hz (UL)	2.6 b/s/Hz (DL); 1.3 b/s/Hz (UL)
Simultaneous VoIP users per sector / MHz	25	>60

Release 1 of IEEE 802.16e-2009. It also improves coexistence by taking into account optimization of interferences, and enhances interworking with other Radio Access Technologies (RAT) via seamless handover procedures. One part of this is seen in concurrent operation of IEEE 802.16m and non-802.16m technologies via a single device. These other RATs may include IEEE 802.11 Wi-Fi Networks, 3GPP networks like HSPA, LTE and LTE-Advanced, and 3GPP2-defined 1x-EVDO of America.

FFR (Fractional frequency reuse) and segmentation functionality are supported via the Release 2 subchannelization mechanism of IEEE 802.16m. This provides the possibility to deploy Release 2 BSs via various sectorization schemes like 3 and 4 sector configurations in such a way that also Reuse 1 operation is supported at the same time [8].

For network transition from Release 1 to Release 2 of IEEE 802.16m, in many cases, the migration is logical to plan when additional spectrum is available providing more capacity. The simplest migration scenario can be executed as soon as an additional frequency band of 10 MHz becomes available, making it possible to replace Release 1 configuration with a new Release 2 configuration with, for example, enhanced reuse 3 scenario [8]. Figure 14.7 clarifies this example.

Table 14.6 *Target performance of IEEE 802.16m according to Ref. [3]*

Item	Value
RF frequency band	Various, typically 450 MHz-3.6 GHz
Duplex	TDD, FDD, H-FDD
Scalable channel bandwidth	5, 7, 8.75, 10, 20, 40 MHz
Multicarrier support	Up to 100 MHz, with Carrier Aggregation
Mobility	Up to 500 km/h
Superframe	20 ms
Frame	5 ms
Subframe	0.625 ms
User plane latency	<10 ms
Control plane latency	<100 ms (from idle stage to active stage)
DL peak user data rate	1 Gb/s, low user mobility 100 Mb/s, high user mobility

Source: Data by courtesy of WiMAX Forum.

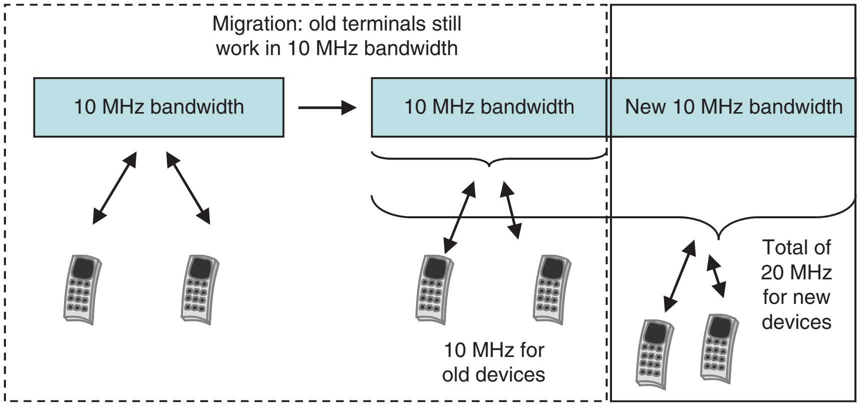

Figure 14.7 *Example of the transition from IEEE 802.16e Release 1 to IEEE 802.16m Release 2 [8]. Figure adapted from original of WiMAX Forum.*

14.5.3 Applications and Key Functionalities

IEEE 802.16m provides a high-quality VoIP experience for end-users along with increased capacity and enhanced performance figures for the delays. Thanks to persistent and group scheduling, faster HARQ retransmissions, rate matching, optimized QoS support, among with other enhancements for spectral efficiency, the VoIP capacity of IEEE 802.16m is significantly increased compared with 802.16m [6].

IEEE 802.16m provides sufficient bandwidth and advanced network traffic management tools to support a cost-efficient variety of applications that work best in the broadband environment. According to [4], there will be over 20 million vertical connections worldwide by 2014, from which 24% are estimated to be utilized via WiMAX retail connections. Source [4] further claims that the revenue stream generated by vertical applications will accelerate the path to profitability for WiMAX operators, and that the need to preserve the high profitability of voice services and limited network capacity are the main reasons for cellular operators' limited interest in vertical applications. Furthermore [4], concludes that widespread presence and the provisioning of crucial services make utilities one of the segments with the strongest and most varied demand for vertical applications, and that WiMAX is well positioned to meet the demand in this market.

Other features and services that IEEE 802.11m WiMAX system release 2 is capable of providing are [6]:

- E-MBS (Enhanced Multicast Broadcast Service). This makes it possible to use the broadcast service in optimized way as the spectral efficiency is enhanced from the previous solutions. It is also possible to switch between broadcast and unicast services on the same or different frequencies.
- Advanced Location Based Services (LBS) via either GPS or other positioning systems. More specifically, IEEE 802.16m contains triangulation schemes in such a way that the latency for the location solving is less than 30 seconds.
- Support of SON concept (Self-Organizing/Optimizing Network). This provides means for the automatic configuration and optimization of the WiMAX network that has positive impact on the service availability, QoS, network efficiency, and throughput with highly dynamic circumstances of the traffic profiles and radio conditions.
- Enhanced security via advanced encryption schemes. This provides additional benefits in location privacy and user identity protection.
- Enhanced mobility via the support of maximum terminal velocity of up to 350 km/h or 500 km/h as a function of the frequency band.

14.6 Comparison of Wireless Technologies

Table 14.7 summarizes some of the key parameters and performance indicators of wireless technologies from where IEEE 802 solutions form part. It should be noted that the information is based on assumptions and peak data rate estimates.

It is important to note that all of the values mentioned in Table 14.7 give only a rough idea and may not be compared directly between the others. This is due to the fact that each technology may consist of different setups, network deployment solutions, parameters with interdependencies, and additional solutions that can largely vary the final values. Logically, the systems that share the same bandwidth are affected on the capacity utilization of other users. As another example, the deployment of active antenna system or high-order MIMO antennas can considerably enhance the performance where applicable. Also practical values depend on many factors in the utilization of the device. As an example of this, the link budget might change notably when the device is held in the hand instead of on the table via Bluetooth headset.

14.6.1 Other Connectivity Methods

The WLAN variants of IEEE 802.11, as well as IEEE 802.16 form the current basis for Internet access methods. IEEE 802.20 (Mobile Broadband Wireless Access) and 802.22 (Wireless Regional Area Network) are part of the wireless access methods.

The other wireless variants belong under IEEE 802.15, which describes the Wireless Personal Area Networks (Wireless PAN). It contains the following substandards:

- IEEE 802.15.1 Bluetooth certification
- IEEE 802.15.2 IEEE 802.15 and IEEE 802.11 coexistence
- IEEE 802.15.3 High-Rate wireless PAN
- IEEE 802.15.4 Low-Rate wireless PAN (e.g., ZigBee, WirelessHART and MiWi)
- IEEE 802.15.5 Mesh networking for WPAN
- IEEE 802.15.6 Body area network.

From this list, Bluetooth and UWB are the most relevant ones for short range connectivity of users.

In addition to IEEE-defined wireless local area networks, there are many solutions that might partially share the same ideas for short range connectivity. One example of these is NFC (Near Field Communications) which is presented in more detailed level in Chapter 5.

14.6.2 The Future

WLAN is without doubt the most useful and widely utilized base for any Internet communications. The original IEEE 208.11legacy standard is a logical base for all developed IEEE 802 variants, in order to enhance the coverage of WLANs constantly. Each IEEE 802 WLAN solution justifies their position in the markets, and for each environment there is an optimal technology available from short to long distance environments, to fixed, portable and mobile use. In addition, WLAN solutions complement and enrich the functionality and performance of mobile communications networks of all generations. As a most concrete example, Wi-Fi Offloading provides seamless handovers between Wi-Fi and mobile networks and thus provides smooth user experiences even in highly congested situations.

The current limitations of the WLANs can further be solved, for example, by the deployment of smart antennas or adaptive active antennas (AAS). AAS systems provide highly advanced methods for enhancements of coverage areas, minimizing interference levels, which increases capacity accordingly.

Table 14.7 *Comparison of wireless technologies*

Technology	Radio access	DL data rate (Mb/s)	UL data rate (Mb/s)	Additional information
3GPP HSPA+	WCDMA	21–672	5.8–168	Widely deployed technology which in practice has been interpreted as 4G technology.
3GPP LTE	OFDMA / SC-FDMA	100 (Cat 3), 150 (Cat 4), 300 (Cat 5) in 20 MHz channel.	50 (Cat 3), 75 (Cat 4), 75 (Cat 5) in 20 MHz channel.	LTE-Advanced as defined as of Release 10 will increase DL peak data rate up to 1 Gb/s (fixed location) and 100 Mb/s (mobile).
WiMAX rel 1.1 (IEEE 802.16)	SOFDMA	37 in 10 MHz TDD channel, 2 × 2 MIMO	17 (in 10 MHz TDD channel, 2 × 2 MIMO	Wireless MAN
WiMAX rel 1.5 (IEEE 802.16-2009)	SOFDMA	83 in 20 MHz TDD channel; 141 in 2 × 20 MHz TDD channel	46 in 20 MHz TDD channel; 138 in 2 × 20 MHz TDD channel	Wireless MAN
WiMAX rel 2 (IEEE 802.16m)	SOFDMA	2 × 2 MIMO 110 (20 MHz TDD) 183 (2 × 20 MHz FDD) 4 × 4 MIMO 219 (20 MHz TDD) 365 (2 × 20 MHz FDD)	2 × 2 MIMO 70 (20 MHz TDD) 188 (2 × 20 MHz FDD) 4 × 4 MIMO 140 (20 MHz TDD) 376 (2 × 20 MHz FDD)	These values are for 2 × 2 MIMO
Flash-OFDM	Flash-OFDM	5.3 10.6 15.9	1.8 3.6 5.4	Mobile Internet with up to 200 mph / 350 km/h speeds. Mobile range 30 km / 18 miles, and extended range 55 km, 34 miles.
HiparMAN	OFDM	56.9	56.9	Mobile Internet
Wi-Fi, IEEE 802.11n	OFDM	288.8 with 4 × 4 MIMO and 20 MHz BW, or 600 with 4 × 4 MIMO and 40 MHz BW.	288.8 with 4 × 4 MIMO and 20 MHz BW, or 600 with 4 × 4 MIMO and 40 MHz BW.	
iBurst, 802.20	HC-SDMA / TDD / MIMO	95	36	Cell Radius up to 12 km, speed up to 250 km/h.
GSM EDGE	TDMA / FDD	1.6	0.5	Mobile Packet Access for GSM according to 3GPP Rel 7.
UMTS FDD and HSPA evolution	WCDMA	0.384 / 14.4 / more	0.384 / 5.76 / more	Packet data evolution of UMTS is advancing fast in the markets.
UMTS TDD	WCDMA	16	16	Especially Asian markets driven mode.
EV-DO Rel. 0 / A / B	CDMA 2000	2.45 / 3.1 / 4.9	0.15 / 1.8	Mobile Internet for CDMA evolution, North America driven.

References

1. WiFi Alliance. WiGig: http://www.wigig.org (last accessed 28 July 2014).
2. Wimax Forum general presentations. http://www.wimaxforum.org/ (last accessed 28 July 2014).
3. WiMAX Forum. White Paper. WiMAX, HSPA+, and LTE: A Comparative Analysis. November 2009. 36 p.
4. WiMAX Forum. White Paper. WiMAXTM Applications for Utilities. A case study of how WiMAX can enable vertical enterprise applications. October 2008. 18 p.
5. Washington University, the Department of Computer Science. WiMax presentations. http://www.cse.wustl.edu/~jain/cse574-10/ftp/wimax2/index.html (last accessed 28 July 2014).
6. Wimax Forum. White Paper. WiMAX and the IEEE 802.16m Air Interface Standard – April 2010. 26 p.
7. Comparison of WiMAX 1 and 2: http://bandaancha.eu/articulos/wimax-2-completara-noviembre-redes-2012-7422 (last accessed 28 July 2014).
8. WiMAX Forum. White Paper. Migration from WiMAX Release 1 to WiMAX Release 2. 12 p.

15

Terrestrial Broadcast Networks

Jyrki T. J. Penttinen

15.1 Introduction

Television and audio broadcast networks are of the utmost importance in the world's communication. They are both most suitable for real-time information sharing in wide areas, as well as for entertainment.

The analog systems are disappearing gradually whilst digital systems are taking over both audio and television broadcasting. In addition, new methods for global distribution of radio and video contents are being utilized rapidly via the Internet [1].

For a more thorough study of the terrestrial broadcast technologies, references [1–35] provide detailed information based on standards and publicly available studies.

15.2 Analog Systems

15.2.1 Radio

Along with the digitalization of radio and television systems, the older analog systems are disappearing in a fast time schedule. Many countries have completely replaced analog systems with digital from the beginning of the 2000s.

Nevertheless, the era of the analog radio systems will not be over completely for a long time. Low-frequency radio transmissions are still resisting global digitalization, and despite the growing popularity of Internet radios, there is still demand for traditional, "old-fashioned" radio which is useful especially as part of weather alarm systems. It is also a logical way of communication in many rural and marginal areas due to the fact that terminals cannot be assumed to be replaced by digital equipment in a fast time schedule.

The local oscillator of the receiver forms the difference between oscillator frequency and normal received frequency. This difference is always constant and is called interfrequency. In FM bands, interfrequency of 10.7 MHz is used, whilst AM systems use 460 kHz. When the modulated carrier frequency is presented in

The Telecommunications Handbook: Engineering Guidelines for Fixed, Mobile and Satellite Systems, First Edition.
Edited by Jyrki T. J. Penttinen.

interfrequency, it is possible to achieve better frequency selectivity compared to some other solutions like directly connected receivers (an early example of these being a crystal receiver).

Stereophonic audio can be created via stereo receiver. The audio information of the left and right channels is formed by the sum and reduced components. The left audio channel of FM system is transmitted as such whilst the right channel is transmitted above it. In between the channels, at 19 kHz, there is a pilot signal transmitted which indicates the right channel. Each stereophonic transmission happens in 300 kHz channel in radio interface as well as in the cable networks. In each solution, the typical frequency band for the FM is in the range of 87.5–108.0 MHz.

The evolved variants of the FM transmission include RDS (radio data system), which is capable of transmitting added information along with the actual audio. It also provides automatic handover of the channel for RDS receivers, for example, when driving a car. RDS contains the possibility to use various different program types which ease the finding of wanted content. The display of RDS receiver is capable of showing related textual information about, for example, a song currently playing on the radio. The system also provides the possibility to set the automatic playing back of different announcements. These attenuate the audio of the ongoing program. It is also possible to transfer any other type of data via RDS. As an example, in the early days of GPS, RDS was utilized for delivering differential correction data of the GPS.

15.2.2 Television

The analog television system is already close to its life cycle. The same basic principles of television were utilized for decades; the only really significant addition having been color, which still utilized the very same base for black and white signals – the color components were transferred simply on top of the original signal, and the amplitude of the additional signal indicated the color schemes. The base for analog system was CRD (cathode ray tube). A beam was formed within the tube which was directed from one edge to another by sweeping the ray, and by repeating the sweep for each line from top to bottom. The number of lines depended on the system. The sweeps formed single pictures on the display of the cathode ray tube, and when repeated with new pictures, the human eye interpreted the series of the pictures as moving without considerable delays between pictures.

There were three main systems in global markets: the European PAL (phase alternation line), American and Japanese NTSC (national television system committee) and French system that was also utilized in Eastern Europe and Russia, SECAM (*séquentiel couleur avec memoire*).

The PAL variant was based on 625 cross lines. In order to attenuate the flickering effect, the odd lines were swept first, and then the paired lines. The repetition rate of the complete pictures was 25 Hz, that is, 50 Hz for the incomplete pictures (odd and paired separately).

The analog television system has been utilized typically in VHF via 7 MHz bands (Table 15.1) and in UHF via 8 MHz bands. The transmission was formed in such a way that the television pictures are produced in a 5 MHz block, and the audio is transmitted on top of that. More details about this principle can be seen in the book chapter about modulations.

Table 15.1 *Frequency bands of analog television*

Band	Frequency
VHF I	47–68 MHz
VHF III	174–230 MHz
UHF	470–790 MHz
S (cable channels)	105.25–463.25 MHz

In its basic format, that is, black and white television system, the television picture is formed in the display of the cathode ray tube according to the vertical and horizontal deviation plates. The beam is visualized by a fluorescent material of the display. This material itself activates the displaying of the ray per each spot and keeps the fluorescent activated for a sufficiently long time even when the ray moves to other spots. Thanks to this delay, the human eye will not interpret flickering in the consecutive pictures. The strength of the beam is possible to adjust by the user, in order to adjust the brightness.

Color television is deployed by utilizing exactly the same principles as in the black and white variant. It is thus possible to receive color programs with older black and white television sets. The color picture is formed by utilizing three color components, which are blue, red and green. The color components are added to the BW transmission and are separated in the reception side in such a way that each component is formed via separate electron devices.

The evolution of analog television contains the introduction of text television signaling. It takes advantage of the wait period between single pictures, and sends low-bit rate data for the receiver. The basic variant of text television provides transmission of 4 text pages per second, to be browsed by users. The original variant of text television is capable of showing 40 characters in 24 lines.

The further evolution of the television system contains NICAM (near instantaneous compounding audio multiplexing). This addition was made for delivery of high quality stereo sound, or alternatively, transmission of two separate mono audio channels. The latter is useful, for example, for delivery of two different languages in a parallel way for the same video contents like a sport event.

In areas where cabling is possible, CATV (cable television) is a logical solution. The cable operators have to comply with special regulations. One or several households and communities can be covered via MATV solution (master antenna television), which means that it is possible to deliver the contents from a centralized point, by combining different delivery systems as cable, terrestrial and satellite systems for the delivery chain.

For single family houses, the most logical solution is typically a single point of receiver either via own antennas, or via the connection to the cable television infrastructure. The single point reception can also be done via satellite. In that case, the technique is referred as DTH (direct to home). In SMATV solution (satellite master antenna television system), it is possible to deliver reception of the contents via a single antenna to a group of users connected to the delivery chain.

The analog and digital transmission of cable television networks is delivered typically in bands of 47−606 MHz or 862 MHz. For the possible return channel, there may be own frequency band of 5−65 MHz. In this case, the reception frequency bandwidth is set to 87.5−606 MHz or 862 MHz. The channels of cable television are rearranged in such a way that they are not coinciding with the terrestrial system, which lowers the effects of, for example, multipath propagation.

15.3 Digital Radio

15.3.1 Principle

For the digital variant of the radio system, there are various options available. One of the early solutions has been European DAB (digital audio broadcasting) which was meant to enhance FM radio quality and usability. Although, DAB was not commercially successful at the beginning of the 2000s, Germany and UK have been the main drivers for this technology.

DAB is based on SFN (single frequency network), which functions differently from the previous MFN (multi frequency network) that has been used traditionally for FM radio. Nevertheless, the deployment of each technology requires planning. The most limiting factor in MFN is cochannel interference, which requires sufficiently safe distance between cochannel sites. This is no issue for SFN as transmitting of the same contents happens in the same frequency for all sites defined for the SFN area. Nevertheless, the propagation delay of SFN transmission must be taken into account as signals propagating from sites located too far away start to act as interfering sources. With careful planning, by utilizing directional antennas and dimensioning transmitter power levels, antenna heights and by taking into account the topology of the surrounding environments, it is possible to construct large, even country-wide SFN networks.

The bandwidth of DAB is 1.5 MHz which provides with 1.2 Mb/s data rate. DAB is delivered via VHF III band.

The total bit rate of DAB is divided between different sources of contents. This can be done in a flexible way so that different channels can be transmitted with different bit rates. This provides the operator with a way to differentiate quality of channels. As an example, there can be high-fidelity channels for classical music with, for example, 240 kb/s per channel, and low bit rate speech channels for news with, for example, 64 kb/s. The mid-quality can be delivered, for example, with 192 kb/s. In addition, it is possible to deliver the same contents via DAB channel but by varying the accompanying voice with different languages.

Typically, a single multiplex, that is, a bunch of DAB channels, contains 6 stereophonic channels whereas the number of channels meant for speech programs can be substantially higher.

The audio bandwidth of DAB is 22 kHz, compared to the 15 kHz bandwidth of FM systems. The acoustic dynamic range of DAB is 90 dB, FM dynamics being 60 dB. This results in a clearly higher audio quality of DAB.

15.4 Digital Television

The analog era of television systems lasted several decades. It can be generalized that during that time, there was only one major evolutionary step taken, that is, when color was added on top of the black and white system, in a backwards compatible way.

There were intentions to deploy digital television service already in 1990s by developing HDTV (high definition TV) but it did not take off by that time. Nowadays in Europe, digital television has been deployed via other variant, DVB (digital video broadcasting).

The global distribution of digital television deployments is mainly based on the following standards: DVB (Europe, Africa, Asia and Australia), ATSC (North America), ISDB-T (South America, Japan) and DTMB (China).

15.4.1 DVB

The terrestrial digital television system defined by ETSI is DVB-T (DVB, terrestrial). It is meant to be used in fixed locations, for example, in homes. It also includes definitions for mobile environment, but is not optimal for small handheld devices.

In addition to systems presented in Table 15.2, there is also a development path for each system: DVB-NG for Hand Held, DVB-T2 for terrestrial and DVB-S2 for satellite environment.

15.4.2 DVB-T

The real commercial outbreak of television happened in the 1950s. Since then, television has been a permanent piece of furniture in every household. During the long life of television, there have been basically only two major evolutionary steps: introduction of color and digital transmission. The next logical step is mobility. Although portable analog TV sets have been available for a long time, it is challenging to maintain their reception quality in the mobile environment.

Table 15.2 *DVB variants*

System	Information
DVB-H	Operator can choose an optimal modulation.
DVB-C	Modulation does not allow strong reflections or attenuations, but provides high bit rates.
DVB-S	The demand is high for the frequency bandwidth but the modulation is resistant against low signal-to-noise ratios.
DVB-T	Resistant against reflected signals.
DVB-IPTV	A/V services via IP networking. Formerly known as DVB-IPI.

It is not an overstatement to say that, although ISDB-T was the first standard in introducing Television service, that is, MobileTV to handheld devices, DVB-H and especially OMA-BCAST over DVB-H have been the pioneers in the global outbreak of MobileTV. Today DVB-H has spread all the way from Europe to Africa, Asia, Australia and Latin America, leaving out only the North America and some other markets like Japan and South Korea. Italy was the first country to launch commercial DVB-H services in 2007. Several commercial launches have followed since then. DVB-H has also been supported by the European Union as the dominant digital television system for handhelds.

Whilst the broadcast technologies have emerged to the stage where consumers can view broadcast transmissions through their handhelds, such as mobile phone, also the cellular systems have reached the point, where almost the same content and services can be offered through for example 3G and also soon over further advanced systems such as Long Term Evolution (LTE).

DVB-T is European driven terrestrial digital TV standard, developed by DVB (Digital Video Broadcasting). The system is designed to deliver audio and video via MPEG-2. The modulation is based on COFDM.

The DVB-T, along with other digital TV systems, is a logical continuum for analog systems that are being ramped down globally. In Europe, transition towards digital systems has been advancing in a fast time schedule and major part of the countries are already closing down analog TV networks.

15.4.3 DVB-H

The Japanese ISDB-T was the first digital television standard which offered possibility to view television with the handsets such as mobile phones with satisfactory end-user experience. The European Digital Video Broadcasting (DVB) standard was developed soon after this. Today, the DVB standards consist of several different standards covering terrestrial, satellite and cable systems. The first complete DVB-H specification was published in 2004 which generated various trials and pilots. One of the first DVB-H trials was set up in Finland during 2004. Approximately at the same time, the DVB-CBMS sub-group was founded in DVB to define the IPDC over the DVB-H standard, which would specify a complete end-to-end solution for the delivery of multimedia services over DVB-H. The DVB-IPDC systems layer specifications cover such aspects as Electronic Service Guides, Content Download Protocols and Service Purchase and Protection.

DVB-H (Digital Video Broadcasting, Handheld) is based on the terrestrial digital television standard DVB-T. The DVB-T system is designed for a static environment where a rooftop mounted receiver antenna is installed providing line-of-sight (LOS) or nearly-LOS with the transmitting site. Although it is possible to use the DVB-T receiver to some extent in a mobile reception, according to related experiences, the use of DVB-T is not optimal in an environment where multipath propagation, impulse noise and Doppler shift are present [2].

The initiation of DVB-H standardization work was a result of the noted need for a sufficiently high-quality mobile TV reception when small portable or mobile terminals are used. An EU-sponsored Mobile Television and Innovative Receivers (MOTIVATE) project studied the item and stated in the conclusion in 2000 that although DVB-T could be used in the mobile environment, it was not an optimal solution. During standardization of DVB-T, TV subscribers had not yet used the service significantly in the mobile environment. Even if DVB-T also contains definitions for the mobile channel, the experience has shown that the usage of home settop box solutions in the mobile reception is not feasible in practice due to constantly changing radio conditions. Among the general mobile communications development, the need for optimized broadcast system for the mobile environment increased and finally triggered the standardization work of the DVB-H [2].

There is a set of special aspects in TV reception when the terminal is moving. The main differences between fixed TV and the mobile TV are related to the characteristics of the radio interface. Whilst DVB-T utilizes the fixed roof-mounted high-gain directional antenna, the mobile receiver antenna is normally an internal one. Mobile users are typically at street level where the radio wave propagation conditions are challenging to cope with as the radio interface consists of time and space dependent variations both in outdoor and indoor environments. In addition, power consumption is an important issue with handheld devices.

A considerable power saving can be achieved in receiver functionality of DVB-H [2] which provides sufficiently long usage time in the typical moving environment compared to the receiver that is switched on permanently.

It can be noted that TV is developing towards the use of multiple niche channels. DVB-H would be a suitable platform for this as it contains several audio and video channels with varying quality settings.

The first version of the DVB-H standard with the system name of Digital Video Broadcasting – Handheld, was published as an ETSI standard number EN 302 304 in November 2004. The standard actually complies with the existing DVB-T standards in such a way that the DVB-H system can be created based on those. In practice, the DVB-H includes the fixed and in-car standards of DVB-T, adding new functionalities that take into account the above mentioned specialties in the mobile environment.

The most important additional parts in the link layer include the Time Slicing and MPE-FEC (Multi-Protocol Encapsulation, Forward Error Correction) functionalities as described in [3]. The DVB-T is defined in ETSI EN 301 744, and now includes an annex for DVB-H.

DVB Data is defined in ETSI EN 301 192, which embeds the encapsulation mechanism to the DVB-H data. The DVB service information related information is defined in ETSI EN 300 468, with updated signaling definitions for the DVB-H handheld terminals.

The Time Slicing functionality provides the possibility to transmit data within cyclic, high-capacity bursts. After the reception of a single burst, the terminal switches its receiver to a sleep-mode until the next burst is transmitted. According to DVB-H implementation guidelines [4], Time Slicing reduces average power consumption of the DVB-H receiver front-end by approximately 90–95%. Nevertheless, as the terminal also includes other functionality in addition to the DVB-H receiver (video processing capability with related audio and video player, separate mobile phone functionality of GSM and/or 3G as well as other functionalities, each reserving part of the processing power), the overall saving of power consumption is lower than the savings of the receiver sleep-mode provides, but anyway extends battery life in the mobile environment. In addition to the receiver's power saving, Time Slicing can be used for seamless frequency handover when moving from one site cell area to another.

DVB-H contains the forward error correction (FEC) mechanism of DVB-T. There is also an additional error correction functionality included in DVB-H, that is, MPE-FEC. It improves the performance under impulse interference and Doppler shift. MPE-FEC is defined as an optional functionality in DVB-H.

As Time Slicing and MPE-FEC are defined in the link layer, already existing DVB-T receivers are not disturbed due to the existence of DVB-H. On the other hand, DVB-H is backwards compatible with DVB-T, so both of these broadcasting methods can be multiplexed into a single transmitter antenna.

Also the physical layer of DVB-H has some important additions. According to DVB-H Guideline documentation, these are transmitter parameter signaling (TPS), 4K OFDM (Orthogonal Frequency Division Multiplex) mode, new interleaving depths and additional bandwidth of 5 MHz.

The DVB-T Transmission Parameter Signaling (TPS) is upgraded for DVB-H. The DVB-H system's TPS thus includes two additional bits that indicates the presence of the DVB-H services and the presence of the possible MPE-FEC. The signaling enhances the service discovery process speed. TPS bits also carry the cell identifier in order to support a fast frequency scanning of the mobile receiver and to perform a frequency handover. TPS is mandatory in the DVB-H system.

In addition to the already existing OFDM FFT modes of 2K and 8K of DVB-T, a new 4K mode is specified in order to optimize mobility of the DVB-H terminal and the size of the single frequency network (SFN). This allows single-antenna reception in a medium-sized SFN with a relatively high mobile speed. 4K mode is not mandatory for DVB-H, though.

The DVB-H standard defines new symbol interleaving options. In DVB-T, the 8K mode has a native interleaver in order to spread the bits over the time domain and minimize the bursty errors over a complete symbol. In DVB-H, it is possible to interleave data also in 2K mode, which results in interleaving over four OFDM symbols. In case of 4K mode, the interleaving is done accordingly over two OFDM symbols. This DVB-H functionality is called in-depth interleaving which reduces the effects of impulse noise up to the level that is possible to obtain with the native interleaver of 8K mode.

It is worth noting that neither the 4K mode nor the in-depth interleaver are mandatory for DVB-H, but TPS is obligatory as well as Time Slicing and the cell identifier. As DVB-H is backwards compatible with DVB-T, all of its modulation schemes, that is, QPSK, 16-QAM and 64-QAM, are also possible to use in DVB-H.

In addition to the already existing bandwidth definition for 6, 7 and 8 MHz of DVB-T, the DVB-H standard also defines a new 5 MHz bandwidth for areas where this value is possible to utilize according to the regulation. In practice, this new definition was added to specifications to ease the potential DVB-H activities in the Americas.

It can be generalized that DVB-H defines the radio functionality of the system whilst the remaining part of the network, that is, the leg from the encoders up to the IP Encapsulator (IPE which is the interface towards DVB-H) is called DVB-IPDC (DVB IP Datacast) network. DVB-IPDC delivers the video, audio and/or file content in the form of data packets by applying standard routing principles of Internet.

The benefit of IP delivery is the possibility to use standard components and protocols for the content transmission, storage and manipulation [Dig05]. In addition, DVB-CBMS (Convergence of Broadcasting and Mobile Services) defines video and audio formats, Electronic Service Guide (ESG) and content protection on top of DVB-H.

15.4.3.1 Standardization

DVB-H belongs to the DVB family, the other variants being DVB-T (terrestrial), DVB-C (cable) and DVB-S (satellite). DVB-H is based on the DVB-T and is backwards compatible with it. Handheld mobile TV service motivated a creation of a new DVB standard. During DVB-T standardization mobility was already taken into account so that basic mobility capability was already there, but some additions made the standard more beneficial. It can be said that the DVB-H is the DVB-T standard tailored to meet the requirements and needs of handheld mobile devices. In the DVB-H band, which would be useful for a single DVB-T channel at the time, it is possible to transmit several DVB-H subchannels containing video and/or audio due to the lower capacity need per DVB-T channel and smaller screen size of the terminal.

The DVB-H specification was planned in technical and commercial modules, that is, subgroups, of DVB project, which is a co-operative effort of ETSI (European Telecommunications Standardization Institute), EBU (European Broadcasting Union) and CENELEC (European Committee for Electrotechnical Standardization).

The system specification was included into ETSI specifications, and the first version was published in 2004, resulting in various trials and pilots. One of the first trials was initiated in Finland during 2004, and Italy was the country where the first launch of commercial DVB-H service took place in 2007. The system is also supported by European Union. The system has not stayed specific to Europe, but it has been evaluated as a strong candidate in other continents. As the telecommunications systems tend to evolve for higher capacities, performance enhancement is planned also for the DVB-H. The next generation is called DVB-NGH, previously referred as DVB-H2. The main goal is to provide more capacity, and more flexibility for the service bit rate.

It can be said that DVB-H is the DVB-T standard tailored to meet the requirements and needs of handheld devices. Similarly, like DVB-T and all other standards of the "DVB family of standards," DVB-H supports transport streaming. The main differences in DVB-H vs. DVB-T are in the power consumption, error protection and signaling. Another fundamental difference between DVB-T and DVB-H is that the terminal specifics are taken into account, light weight, portable and battery-powered devices.

The DVB-H specification, which is limited to radio transmission, is not sufficiently complete to enable global interoperability for different networks and devices. In the "DVB project," which created the DVB standards, the DVB-CBMS (Convergence of Broadcasting and Mobile Service) subgroup was founded to define IPDC (IP datacast) over DVB-H standard, which specified a complete end-to-end solution for the delivery of multimedia services over DVB-H. The DVB-IPDC systems layer specifications cover such aspects as Electronic Service Guides, Content Download Protocols and Service Purchase and Protection. The DVB-H thus forms radio layers, that is, physical and media access layers, for larger entity of Mobile Broadcast system.

While the DVB-H standard defines the OSI layer 1 and 2 protocols, the IPDC over DVB-H standard defines the protocols up to the OSI layers 3–7. In addition, the IPDC over DVB-H standard defines also some extensions and rules for the DVB-H specific Program Specific Information (PSI)/Service Information (SI) signaling. During the development process of IPDC over DVB-H standard, an alternative solution for the OSI layers 3–7 over DVB-H was defined by the Open Mobile Alliance (OMA). This OMA BCAST solution adopts a great part of the IPDC over DVB-standard, including PSI/SI as such. The other way round, the IPDC over DVB-H standard has also adopted some parts from the OMA BCAST. Today the roles of the OMA BCAST and IPDC over DVB-H standards have become clearer. The OMA BCAST defines the Upper layer solution and the IPDC over DVB-H standard defines adaptation for the DVB-H bearer. This "OMA BCAST over DVB-H" is today supported by the most of the mobile operators and device vendors which provide products for the DVB-H based mobile TV systems.

15.4.3.2 DVB-H Network

As the core network of DVB-H is based on packet data transmission, the elements are connected to each others via IP network. The capacity of the distribution network should thus be dimensioned sufficiently high in order to avoid possible bottlenecks in the transmission.

The complete end-to-end DVB-H service chain consists of the DVB-H domain which delivers the signal for the DVB-H terminal, and IP network domain, that transports the contents from the encoder to the DVB-H domain. Figure 15.1 clarifies the division.

As can be noted for the Figure 15.1, the IP encapsulator (IPE) acts as an interface between the IP networks and DVB-H broadcast network. The DVB-H standards define the latter one, leaving the practical solution designing of the IP multicasting for the related network element and service vendors.

DVB-H has been designed in such way that it is backwards compatible with the digital terrestrial broadcast network DVB-T. This means in practice that the contents of both DVB-T and DVB-H can be offered via the same physical site and radio bandwidth by multiplexing the contents of each one.

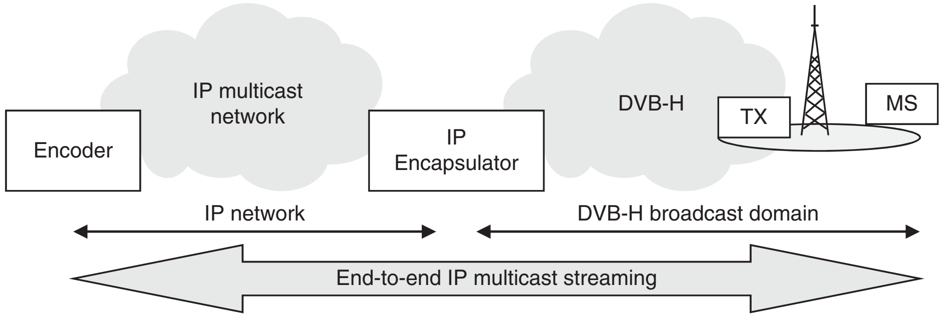

Figure 15.1 *The end-to-end DVB-H service consists of IP network and the actual DVB-H broadcast network.*

Each DVB-H stream, that is, transport stream (TS), is multiplexed for the IP multicast network domain. The basic principle of this can be seen in Figure 15.2.

In the practical DVB-H setup, the core network consists of elements (servers) that manage the streams (IP Encapsulator and its management element), and take care of customer interactions (authentication, authorisation, billing).

As an example, Figures 15.2 and 15.3 shows a high level DVB-H architectural solution. The architecture can be divided roughly into four parts. The content is the original source of the streams, for example, video and/or audio programs or data files. The content streams are converted as IP packet streams, and handled by the media delivery platform. As a third part, there is infrastructure for the supporting functions as charging, operations and maintenance, and customer care. The final part of the architecture consists of DVB-H devices which can be either standalone or integrated with 2G and/or 3G terminals. The former provides the means of using interaction channels via for example, GPRS.

Figure 15.4 shows a more in-depth view to the DVB-H network with the respective elements. The division can be made between the actual DVB-H network and the supporting mobile communications network. As can

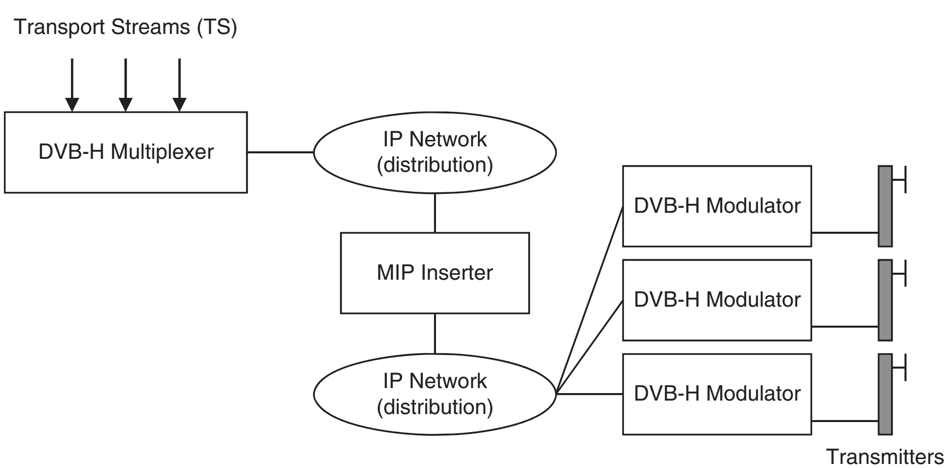

Figure 15.2 *High level DVB-H architecture. The source streams are multiplexed and distributed over the DVB-H IP network, and delivered to the terminals via DVB-H transmitters. If Single Frequency Network is applied, the correct synchronization of the transmitters must be taken care of by using MIP inserter.*

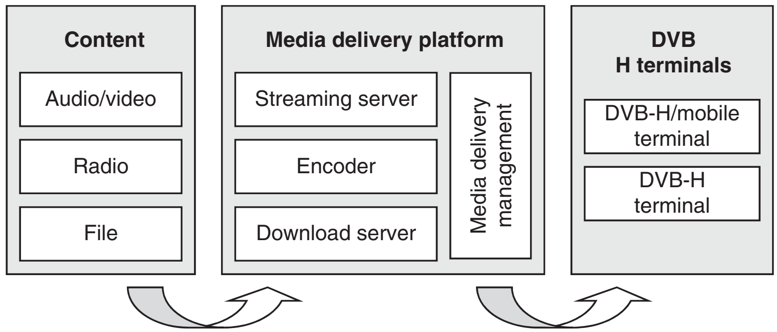

Figure 15.3 *The block diagram of the DVB-H architecture.*

be seen from the figure, the source contents are handled by a set of encoders, that is, each encoder handles a single program. The encoder pool can be set up in a static or dynamic way, that is, the encoder can use either fixed bit rates (FBR), Constant Bit Rate (CBR) or Variable Bit Rate (VBR). Depending on the equipment and capacity requirements of each captured program, the used bandwidth may vary also in dynamic way balancing the load between the encoders. This provides enhanced quality with the same total bandwidth of the core network as the effect is the same as in fixed and mobile telecommunications networks behaving according to the Erlang B model.

DVB-H consists of both DVB-T as well as new DVB-H specific functionalities. The idea of the main functionality of DVB-H network is presented in Figure 15.5. As the convolutional coding of the radio interface is the same for DVB-T and DVB-H, the system is backwards compatible providing the possibility to multiplex both DVB-T and DVB-H streams in the same radio frequency band.

In the receiving end of Figure 15.5, the DVB-H receiver consists of the demodulator and terminal. The DVB-H demodulator gets the DVB-T signal via the RF input either from internal or optional external antenna. The DVB-H demodulator block is in fact a DVB-T demodulator with the DVB-H specific 4k and TPS functionality added on it. The demodulator block also contains time slicing with power control functionality as well as optional MPE-FEC. The DVB-H terminal part obtains the streaming data (IP datagrams and TS data) from the demodulator block.

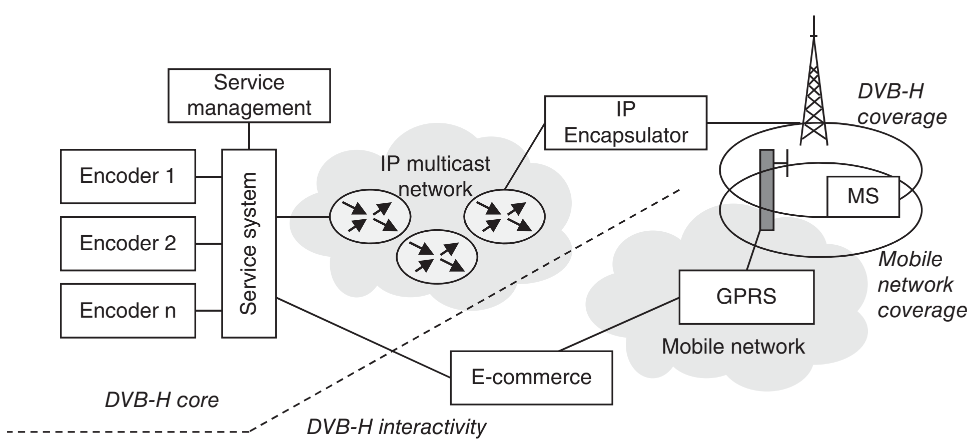

Figure 15.4 *Main elements of the DVB-H network.*

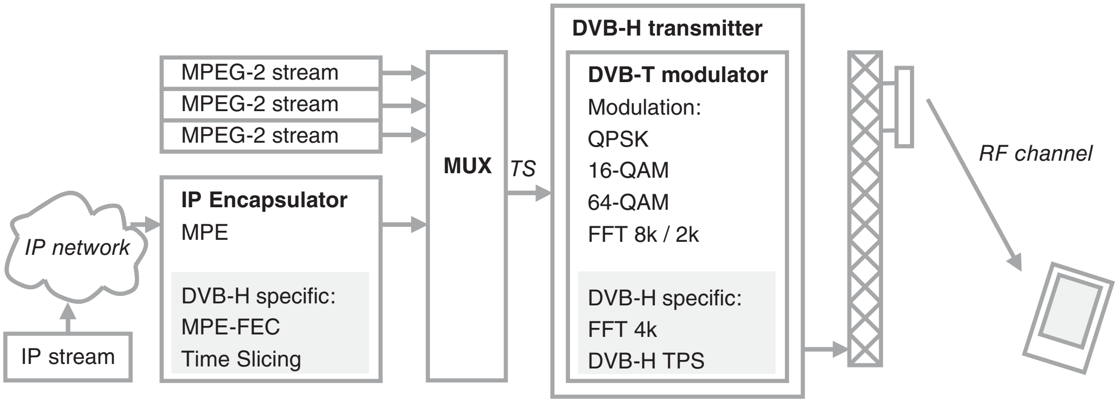

Figure 15.5 *It is possible to multiplex the DVB-H IP streams with the DVB-T MPEG-streams. The streams can be delivered via the common infrastructure for the DVB-H and DVB-T terminals.*

15.4.3.3 Core Network

The following sections provide an overview of the main elements within the core DVB-H network. Figure 15.6 illustrates generic layout of the delivery chain within the core DVB-H network, where the main elements are: *Operations and management, Head-end, Distribution network and Radio network.*

Figure 15.7 presents an example of the DVB-H network design, which is based on the Nokia MBS Mobile Broadcast Solution. As can be observed from Figure 15.7, the broadcast network is unidirectional, sending data streams in downlink. The possible return channel can be done by using already existing mobile communications networks, GPRS being the most logical solution for the interactions.

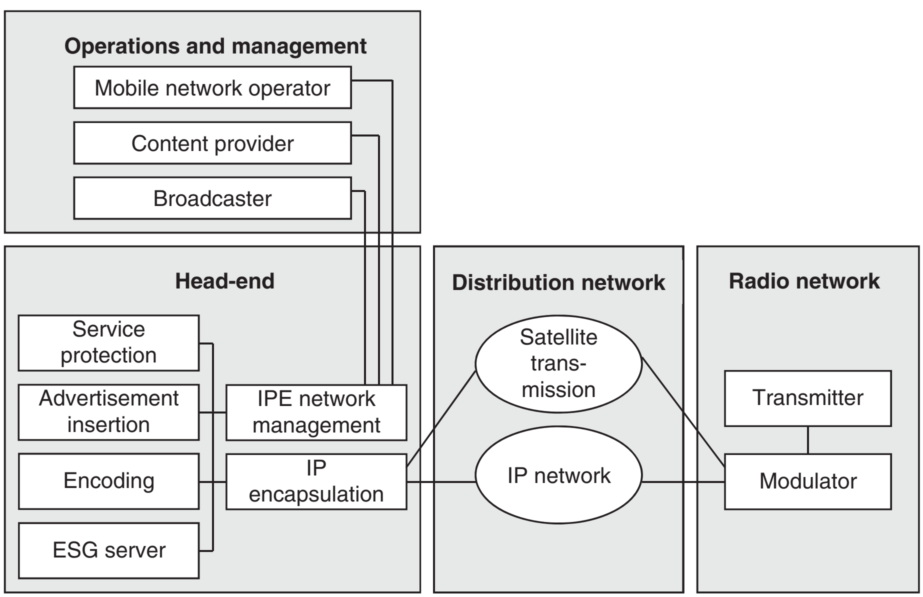

Figure 15.6 *The layout of the delivery chain within the core DVB-H network.*

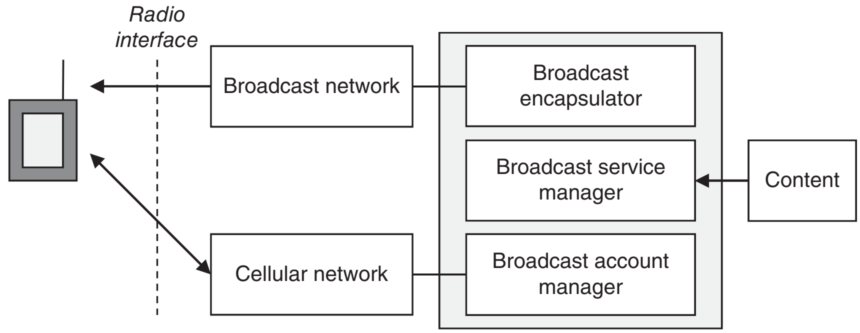

Figure 15.7 *An example of the DVB-H network design according to the Nokia MBS Mobile Broadcast Solution.*

15.4.4 ISDB-T

In Japan, there has been deployed ISDB-T (integrated services digital broadcasting for terrestrial television broadcasting) network. It is functional with commercial terminals in the market already since 2003. The principle of the system is comparable with DVB-H, but its architecture and functionality differ from each others. The system is useful for both broadband television ISDB-T, as well as for the ISDB-T_{SB}, that is, ISDB for terrestrial sound broadcasting.

ISDB-T consists of 13 segments, that is, frequency blocks, whose total bandwidth is 5.6 MHz. The bandwidth of ISDB-T_{SB} is either one or three segments in a total of 429 kHz or 1.3 MHz band.

The commercial deployment of ISDB-T has been initiated by 2003 in Tokyo, Nagoya and Osaka. At that time, ISDB-T_{SB} was already in test use. The first phased of deployment contained the urban city centers.

ISDB-T provides HDTV transmission (high definition TV), ordinary TV transmission (SDTV, standard definition TV), high-quality audio transmission, still pictures and data, as well as light television program transfer for handheld equipment. ISDB-T is thus a set of various modes for many environments.

15.4.5 ATSC

Also the USA has its own digital television system, ATSC (Advanced Television Systems Committee). The mobile version of it is not compatible with the fixed network variant, though, so they can be considered as separate systems.

ATSC system uses the same 6 MHz bandwidth as the analog NTSC television channels. The terrestrial broadcasters apply 8VSB modulation which provides maximum data rate of 19.39 Mb/s. This data stream is used to carry various video and audio programs as well as metadata.

Due to the more controlled interferences, the cable television stations are able to operate at a higher signal-to-noise ratio levels and apply 16VSB (defined in ATSC) or 256-QAM (defined in SCTE). This provides data rates of 38.78 Mb/s within the same 6 MHz channel.

15.4.6 MBMS/eMBMS

The Multimedia Broadcast and Multicast Service (MBMS) is an efficient solution for the distribution of data streaming according to the broadcast principles, but via cellular network. The benefit of the solution is that the infrastructure of the underlying cellular system is completely possible to use for the coverage and capacity the MBMS requires. In other word, MBMS takes a part of the capacity of the system and delivers the contents via the same sell for the customers. As there is no return channel as such, the transmission of MBMS can be received by customers without specific capacity limitations. As an example of use-case of MBMS, the service could be received by all the audience in a large football stadium in order to see, for example, repetition of the goal via the mobile devices. MBMS has been standardized in 3GPP Release 6 specifications.

15.4.6.1 MBMS Services

MBMS contains two data transmission modes which are broadcast mode and multicast mode. The main principle of the MBMS broadcast service is that it is a unidirectional point-to-multipoint service. This service delivers the data stream from the network to all capable 3GPP User Equipments (UE) within the respective broadcast service area. In this way, broadcast services can be received by all users within the service area if they have activated the reception of MBMS service in their UE.

MBMS utilizes the radio and core network resources in an optimal way as the data is transmitted over common radio channel.

MBMS has been planned flexible as for the data rate. The transmission thus adapts to variable radio access network capabilities as well as for varying availability of the resources. The adaptation happens by adjusting the MBMS streaming bit rate accordingly. As there is no return channel for the error recovery purposes, the correct reception of MBMS can not be guaranteed. Nevertheless, as FEC functionality is incorporated, the receiver may be able to recognize data loss and intent to recover the data according to FEC principles.

In addition to the broadcast mode, MBMS also contains Multicast mode. In this case the data is transmitted over a unidirectional point-to multipoint connection from a server to multiple subscribers that are present in the multicast service area. As a difference compared to the broadcast mode, multicast services can only be received by users that have been subscribed to a certain multicast service and have joined the multicast group associated with the specific service. The basic principle is thus that the broadcast mode is open by default whereas the multicast mode typically requires a subscription to multicast groups prior to the possibility for the user to join the multicast group.

MBMS user services can be divided into three groups:

- Download delivery service delivers file (binary data) via MBMS bearer. The user equipment, that is, the MBMS client activates the related application and consumes the received data. In this solution, it is important that the data is received in a reliable way which makes the forward error correction method essential.
- Streaming delivery service provides a continuous stream of, for example, audio and video. Supplementary information via text or still images can also be important in this service. This provides an interactive way of using additional services. As an example, the received text may include WEB links that lead to additional information related to the contents. In this scenario, the link received via broadcast delivery, the user can access the additional content simply by clicking the link, and by creating a dedicated point-to-point connection via any available access method.
- Carousel service can repeat a broadcast transmission of, for example, a file or set of files repeatedly in such a way that the customer can receive the same contents every now and then. This can be, for example, an update to the map data of location based service. The carousel service combines actually aspects of the download and streaming services.

To provide MBMS bearer services the cellular network elements like GGSN, SGSN, RNC or BSC take part in various MBMS related functionalities. From these, part is specific only to MBMS. Figure 15.8 shows the reference architecture of MBMS.

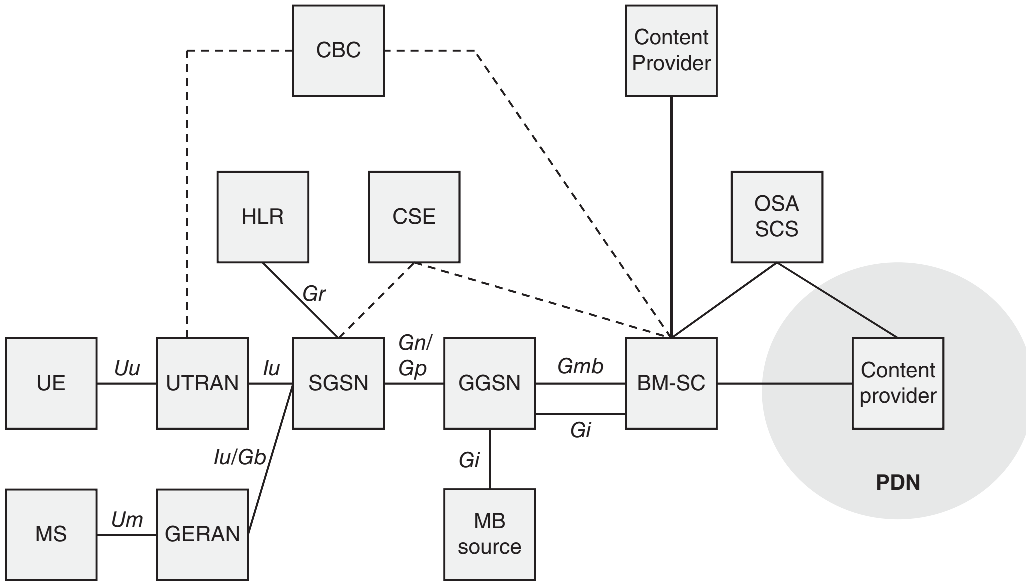

Figure 15.8 *MBMS reference architecture.*

The functionality of the MBMS elements is the following: Broadcast – Multicast Service Center (BM-SC) provides functions for MBMS user service provisioning and delivery. It may serve as an entry point for content provider MBMS transmissions, used to authorize and initiate MBMS bearer services within the Public Land Mobile Network (PLMN) and can be used to schedule and deliver MBMS transmissions. The BM-SC shall be able to:

- Generate charging records for content provider transmitted data.
- Supply the GGSN with transport associated parameters such as quality-of-service and MBMS service area, to Initiate and terminate MBMS bearer resources for following transmission of MBMS data.
- Accept content from external sources and transmit it using error resilient schemes.
- Schedule MBMS session retransmissions, and label each MBMS session with an MBMS Session Identifier to allow the UE to distinguish the MBMS session retransmissions.
- Provide service announcements for multicast and broadcast MBMS user services.
- Transfer data on separate MBMS bearer services for 2G or 3G coverage, typically having different QoS.

The User Equipment supports functions for the activation and deactivation of the MBMS bearer service and security functions as appropriate for MBMS. The UE should be able to receive MBMS user service announcements, paging information (non-MBMS-specific) or support simultaneous services. The MBMS Session Identifier contained in the notification to the UE shall enable the UE to decide whether it ignores the forthcoming transmission of MBMS session.

UTRAN/GERAN network is responsible for efficiently delivering MBMS data to the designated MBMS service area. Efficient delivery of MBMS data in multicast mode may require specific mechanisms in the UTRAN/GERAN. Intra-RNC/BSC, inter-RNC/BSC mobility of MBMS receivers shall be supported. The UTRAN/GERAN shall be able to transmit MBMS user service announcements, paging information and support other services parallel to MBMS.

Serving GPRS Support Node (SGSN) within MBMS architecture performs user individual MBMS bearer service control functions and provides MBMS transmissions to UTRAN/GERAN. The SGSN provides support for intra-SGSN and inter-SGSN mobility procedures. The SGSN generates charging data per multicast MBMS bearer service per user. The SGSN is also able to establish Iu bearers (SGSN-UTRAN signaling) and Gn bearers (SGSN-GGSN signaling) shared by various users on demand when data is transferred to the users. This should be done upon notification from the GGSN.

Gateway GPRS Support Node (GGSN) within the MBMS architecture serves as an entry point for IP multicast traffic as MBMS data. Upon notification from the BM-SC the GGSN shall be able to request the establishment of a bearer plane for a broadcast or multicast MBMS transmission. Furthermore, upon BM-SC notification the GGSN shall be able to tear down the established bearer plane. Bearer plane establishment for multicast services is carried out towards those SGSNs that have requested to receive transmissions for the specific multicast MBMS bearer service. The GGSN shall be able to receive IP multicast traffic from BM-SC and to route this data to the proper GPRS Tunnelling Protocol (GTP) tunnels setup as part of the MBMS bearer service. The GGSN may also receive IP multicast traffic from other sources than the BM-SC. However, the MBMS Bearers are not used to forward this traffic from non-BM-SC sources. The GGSN shall collect the charging data.

15.4.6.2 Flute

FLUTE (File Transport over Unidirectional Transport) is a transport protocol that is designed to deliver files from a set of senders to a set of receivers over unidirectional systems (Figure 15.9). FLUTE has been developed by Internet Engineering Task Force (IETF). FLUTE serves for the file transmission over wireless and point-to-multipoint systems, such as MBMS. FLUTE is specified on top of the Asynchronous Layered Coding (ALC) protocol instantiation of the Layered Coding Transport (LCT) building block. FLUTE uses UDP/IP layer and is transported via respective packets. It is thus independent of the IP version and underlying link layers. FLUTE uses File Delivery Table (FDT) to provide a dynamic index of files. A Forward Error Correction (FEC) can be used optionally for improving reliability of the downlink data transfer. Congestion Control (CC) can also be selected optionally in order to enable Internet friendly bandwidth use on the Internet.

15.4.7 CMAS

CMAS refers to Commercial Mobile Alert System. It has been designed for emergency alerts that can be distributed as quickly as possible over cellular networks. The Federal Communications Commission (FCC), along with the Federal Emergency Management Agency (FEMA) and the wireless industry have all provided the CMAS in the North America, and it is also adopted in the rest of Latin and North America with possibly some technical variations.

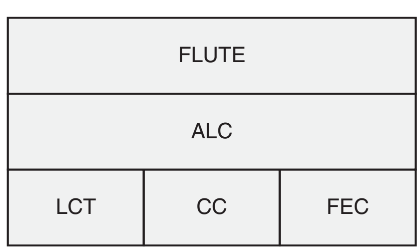

Figure 15.9 *The protocol layers of FLUTE.*

CMAS is a relatively new public safety information system. It provides mobile device to receive location-based text messages via cell broadcast message service (CB) about local safety threats. CMAS functions also in highly congested areas. CMAS complements the earlier Emergency Alert System (EAS) in the USA. CMAS is public and private partnership between the FCC, FEMA and the wireless industry in order to enhanced public safety. On practice, authorized government officials may send emergency alerts upon the justifications, like in the case of tornados or other threats. There is a procedure included into CMAS to authenticate the alert and to verify that the sender is authorized prior to forwarding the message to participating wireless carriers.

There are 3 categories for CMAS alerts: Alerts issued by the President, alerts involving imminent threats to safety or life, and amber Alerts. A CMAS alert is accompanied by an attention signal and vibration.

References

1. Penttinen, J., Jolma, P., Aaltonen, E. and Väre, J. (2009) *The DVB-H Handbook*, John Wiley & Sons, Ltd, 270 pp.
2. Faria, G., Henriksson, J.A., Stare, E. and Talmola, P. (2006) *DVB-H: Digital Broadcast Services to Handheld Devices*, IEEE, 16 pp.
3. ETSI EN 301 192, V1.4.1 (2004-11). Digital Video Broadcasting (DVB); DVB specification for data broadcasting. European Telecommunications Standards Institute 2004. 78 pp.
4. DVB-H Implementation Guidelines (2006) Draft TR 102 377 V1.2.2 (2006-03). European Broadcasting Union. 108 pp.
5. CMAS, FCC. http://www.fcc.gov/guides/commercial-mobile-alert-system-cmas (last accessed 1 August 2014).
6. Aparicio, M. (ed.) (2006) Wing TV. Services to Wireless, Integrated, Nomadic, GPRS-UMTS&TV handheld terminals. D6: Common field trials report. November 2006. 86 pp.
7. Aparicio, M. (ed.) (2006) Wing TV. Services to Wireless, Integrated, Nomadic, GPRS-UMTS&TV handheld terminals. Wing TV Country field report. November 2006. 258 pp.
8. Bouttevin, T. (ed.) (2006) Wing TV. Services to Wireless, Integrated, Nomadic, GPRS-UMTS&TV handheld terminals. D8 – Wing TV Measurement Guidelines & Criteria. Project report. 45 pp.
9. IP datacast over DVB-H : Implementation guidelines for mobility, DVB document A117, DVB organization, July 2007, pp. 18–19.
10. Erceg, V., Greenstein, L.J., Tjandra, S.Y. *et al.* (1999) An empirically based path loss model for wireless channels in suburban environments. *IEEE Journal on Selected Areas in Communications*, **17**(7): 1205–2011.
11. ETSI, "Digital Video Broadcasting (DVB); Framing Structure, channel coding and modulation for digital terrestrial television." ETSI standard, ETSI EN 300 744, November 2004.
12. ETSI, "Digital Video Broadcasting (DVB); Transmission System for Handheld Terminals." ETSI standard, ETSI EN 302 304, June 2004.
13. ETSI, "Digital Video Broadcasting (DVB); Transmission System for Handheld Terminals (DVB-H)," EN 302 304 v1.1.1, Nov. 2004.
14. ETSI, "IP Datacast over DVB-H; Set of Specifications for Phase 1," TS 102 468, 2006.
15. ETSI, "IP Datacast over DVB-H; PSI/SI," TS 102 470, 2006.
16. ETSI, "Digital Video Broadcasting (DVB); Specification for Data Broadcasting." ETSI standard, ETSI EN 301 192 v1.4.1, November 2004.
17. Gardikis, G., Kokkinis, H., and Kormentzas, G. (2007) Evaluation of the DVB-H data link layer. European Wireless 2007, Paris, France, 1–4 April 2007, 6 pp.
18. Henriksson, J. (2005) DVB-H standard, principles and services. HUT seminar T-111.590, Helsinki, 24.2.2005, Presentation material. 53 pp.
19. Himmanen, H. (2008) *Studies on Channel Models and Channel Characteristics for Mobile Broadcasting*. Department of Information Technology, University of Turku, Turcu Centre for Computer Science, Finland, 9 pp.
20. Jeong, M. (2001) *Comparison Between Path-Loss Prediction Models for Wireless Telecommunication System Design*, IEEE, 4 pp.

21. Joki, H. and Paavola, J. (2006) *A Novel Algorithm for Decapsulation and Decoding of DVB-H Link Layer Forward Error Correction*. Department of Information Technology, University of Turku, 6 pp.
22. Jokela, T. and Lehtonen, E. (2007) Reed-Solomon Decoding Algorithms and Their Complexities at the DVB-H Link-Layer. Turku Centre for Computer Science, Department of Information Technology, University of Turku. 5 pp.
23. Mazenc, J. (2007) Multi protocol Encapsulation – Forward Error Correction (MPE-FEC). ETUD/Insa. Toulouse, France, January 9, 2007, 7 pp.
24. Limits of Human Exposure to Radiofrequency Electromagnetic Fields in the Frequency Range from 3 kHz to 399 GHz. Safety Code 6. Environmental Health Directorate, Health Protection Branch. Publication 99-EHD-237. Minister of Public Works and Government Services, Canada 1999. ISBN 0-662-28032-6. 40 pp.
25. OMA: "Mobile Broadcast Services," Draft Version 1.0 – 03 Nov 2008, Open Mobile Alliance, OMA-TS-BCAST_Services-V1_0-200801103-D.
26. OMA: "Service Guide for Mobile Broadcast Services," Draft Version 1.0 – 28 Oct 2008, Open Mobile Alliance, OMA-TS-BCAST_Service_Guide-V1_0-20081028-D.
27. Paavola, J., Himmanen, H., Jokela, T. *et al.* (2007) The performance analysis of MPE-FEC decoding methods at the DVB-H link layer for efficient IP packet retrieval. *IEEE Transactions on Broadcasting*, **53**(1), 263–75.
28. Penttinen, J. (2008) Field Measurement and Analysis Method for DVB-H Terminals. The Third International Conference on Digital Telecommunication, ICDT 2008, 29.6.-5.7. 2008 – Bucharest, Romania. 6 pp.
29. Penttinen, J.T.J. (2008) The Simulation of the Interference Levels in Extended DVB-H SFN Areas. The Fourth International Conference on Wireless and Mobile Communications, IEEE 2008, pp. 223–8.
30. Penttinen, J.T.J. (2008) The SFN gain in non-interfered and interfered DVB-H networks. The Fourth International Conference on Wireless and Mobile Communications, IEEE 2008, pp. 294–9.
31. Penttinen, J.T.J. (2009) DVB-H Performance Simulations in Dense Urban Area. The Third International Conference on Digital Society, IEEE 2009, 6 pp.
32. Plets, D., Joseph, W., Verloock, L. (2008) Influence of reception condition, MPE-FEC rate and modulation scheme on performance of DVB-H. *IEEE Transactions on Broadcasting*, **54**(3), 590–8.
33. Silvennoinen, A. (2006) DVB-H –lähetysverkon optimointi Suomen olosuhteissa (DVB-H Network Optimization under Finnish Conditions). Master's Thesis, Helsinki University of Technology, 15.5.2006, 111 pp.
34. Väre, J. and Puputti, M. (2004) Soft Handover in Terrestrial Broadcast Networks. Proceedings of 2004 IEEE International Conference on Mobile Data Management, Berkeley, CA, USA, Jan 2004, pp. 236–42.
35. Väre, J., Alamaunu, J., Pekonen, H. and Auranen, T. (2006) Optimization of PSI/SI Transmission in IPDC over DVB-H Networks. Proceedings of 56th Annual IEEE Broadcast Symposium, Washington, DC, USA, Sept. 2006.

16

Satellite Systems: Communications

Jyrki T. J. Penttinen

16.1 Introduction

Via satellites, it is possible to deliver various types of data, including voice and data of mobile communications, broadcast channels of radio and television, delivery of telemetry, and location based services.

The clear benefit of satellite systems is the large operational area as a single satellite antenna's beam covers whole nations and large areas that are not covered necessarily by any other means of communications.

A satellite telephone, or terminal, is typically a mobile phone that connects to satellites. The functionality of satellite telephones is comparable with that of terrestrial mobile telephones. Thus voice calls as well as a short message service and Internet access are supported through most systems. Depending on the satellite system, the service area offered may cover the whole globe, or part of it.

The benefit of the satellite telephone is that it is designed to work in the most remote locations where no other commercial telecommunications systems are available. The size of early satellite telephones was relatively large, with typically a large retractable antenna, but along with the general development of the telecommunications systems, the size has been approaching to the typical cellular mobile phones which make it a more feasible solution for remote expeditions.

There is also fixed installations available for the satellite telephony, in maritime environment as well as for the terrestrial communications. For the former, there is typically a directional satellite antenna system that dynamically and automatically tracks the available satellites. The terrestrial installations may use VoIP solution, VSAT being an example with the additional benefit of typically lower CAPEX and OPEX.

The common phenomena for satellite telephony and satellite Internet service are that indoor reception is problematic, the satellite connection requires line-of-sight (LOS) or near-LOS connection in order to function correctly. In practice, the phones typically have external antenna connectors for the antenna installation in the external parts of buildings or vessels. In addition, there is the possibility to use repeaters for enhanced coverage.

The Telecommunications Handbook: Engineering Guidelines for Fixed, Mobile and Satellite Systems, First Edition.
Edited by Jyrki T. J. Penttinen.

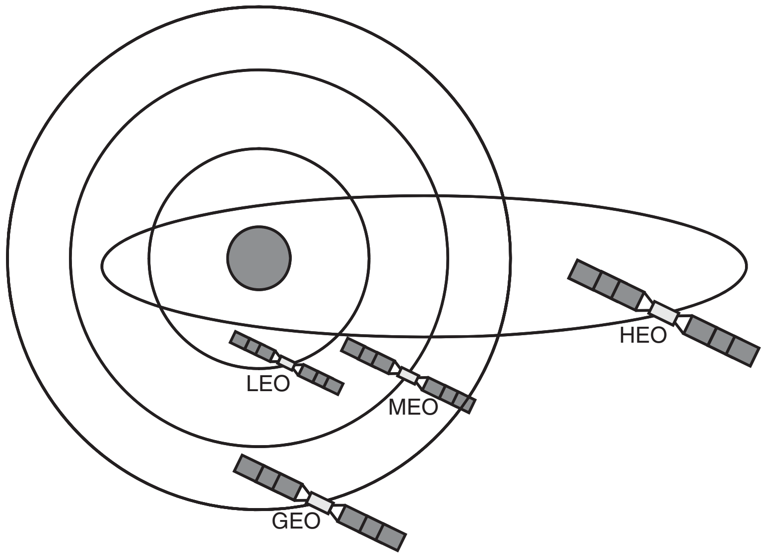

Figure 16.1 The orbits of satellite systems.

16.2 Principles of Satellite Systems

16.2.1 General

Satellite system refers to a set of wireless services, like radio and television broadcast, location based services, data transfer services and delivery of telemetry data both from Earth and space. In addition to the publicly available services, there are also various satellites and services for closed groups like military environment and scientific space investigation.

16.2.2 Orbits

The main solutions for placing the satellites into the Earth's orbits are LEO, MEO and GEO as shown in Figure 16.1. Each orbit has specific order of the orbits as well as special characteristics as explained below.

16.2.2.1 LEO

LEO (Low Earth Orbit) refers to the lowest position satellites can operate compared to the Earths surface. The satellites located into this orbit provide broadband data services with the lowest possible delay in the radio transmission path. LEO satellites typically are below 5000 km height from the Earth's surface in such a way that the most of the satellites are located between 500 and 1600 km. Typical solutions for the LEO satellites are related to the rescue missions, voice calls and data transfer.

At an altitude of 200 to 300 km, LEO is used for certain types of scientific and observation solutions in such a way that the satellite is able to provide a view of the Earth as the satellites fly over both hemispheres.

Typical solutions for the LEO satellites are location based services like GPS, GLONASS and GALILEO. This chapter describes the communications solutions.

16.2.2.2 MEO

MEO (Medium Earth Orbit) is located in between the lowest and highest orbits, with distance from the Earth of about 10 000 and 20 000 km. Logically, the delay values of the radio transmission are longer than in the case of LEO.

16.2.2.3 GEO

GEO (Geostationary Earth Orbit) refers to a geosynchronous satellite angle with zero inclination. This means in practice that the satellite apparently looks like staying over a single point on the Earth's Equator. In the professional unofficial terminology, the satellite that is located to the geosynchronous orbit can be called a "bird."

As for the more official terminology, Geostationary Transfer Orbit refers to the Equatorial plane. It has an elliptical form, and its perigee is at 200 km and apogee at 35 870 km. Furthermore, term geosynchronous refers to the Clarke circular orbit above the Equator. For Earth, the point is at 22 237 miles above the Earth's surface. Characteristics for satellites placed in these orbits are that, although their velocity compared to the Earth is several thousands of miles per hour, they appear to be stationary when observed from a certain fixed point on the Earth. This phenomenon is explained by the rotation of Earth along its own axis at the same angular rate that the satellite is traveling around the Earth.

GEO is the highest circular orbit, with the distance from the Earth of 35 848 km. Unlike in the case of previously mentioned orbits, the position of the GEO satellites is fixed compared to the Earth's surface, that is, they seem to stay still by the observers at the surface of Earth. The benefit of this solution is that it provides a permanent coverage area without variations during utilization of the services the satellite provides to end-users.

Due to the height of GEO satellites, they provide larger coverage areas compared to previously mentioned solutions. In theory, three GEO satellites would be sufficient to cover the whole surface of the Earth.

The drawback of GEO is that the delay of the radio transmission is considerably higher than the other solutions have. The value is about 0.24 seconds between the satellite and earth station. If two-way communication is applied, this value is doubled due to the complete leg of round trip resulting in about half second total delay which can already be noted in two-way voice communications lowering slightly the user experience for the fluent call.

In order to stay at the GEO orbit, the satellites should be located approximately at the level of the Equator. The distance between neighboring satellites should be around 800 km which corresponds to about one degree angle.

Geostationary satellites are limited as for the latitude which typically ranges between the maximum of 70 degrees north of the Equator and maximum of 70 degrees south of the Equator. Another disadvantage of GEO systems is that many times the LOS between terminals and satellites is distracted by obstacles due to the topology and vegetation of the surrounding environment. In order to avoid the lost connection, users need to search for sufficiently open areas prior to the utilization of the terminal. LEO systems are easier to use in this sense because even if the signal is attenuated by an obstacle, another satellite typically appears in the area in some minutes. On the other hand, moving LEO satellites may also drop calls during the time period when LOS is not available.

16.2.2.4 HEO

A special orbit is HEO (Highly Elliptical Orbit), which is, unlike any other variant mentioned previously, elliptical, Earth being in one of the two focal points of the ellipse. It is also referred to as Extremely Elliptical Orbit (EEO). HEO is used, that is, by the Russian Molniya Satellite system.

Table 16.1 Radio frequencies of typical satellite systems

RF band	Radio frequency
HF	3–30 MHz
VHF	50–146 MHz
P	230–1000 MHz
UHF	430–1300 MHz
L	1.530–2.700 GHz
S	2.700–3.500 GHz
C	3.700–4.200 GHz (downlink), 5.925–6.425 GHz (uplink)
X	7.250–7.745 Ghz (downlink), 7.900–8.395 GHz (uplink)
Ku	10.700–12.750 (downlink), 14.000–14.800 / 17.300–18.100 (uplink)
Ka	18–31 GHz

Helio-synchronous Orbit is at an altitude of 600–800 km and it is in a quasi-polar plane. Satellite orbiting this plane is always visible from that part of the Earth in sunlight. Helio-synchronous orbits are used for Earth observation or solar-study satellites.

16.2.3 Frequencies

The frequencies utilized in the satellite systems are typically in GHz ranges, or up to tens of GHz. In some special cases the satellites can also operate in lower frequencies.

Table 16.1 summarizes typical satellite radio frequency ranges.

Some characteristics of the above mentioned frequencies are:

- High Frequency (HF) refers to radio frequency range of 3–30 MHz. HF radio is commonly known as shortwave, and it has varying propagation characteristics depending on the more specific frequency. In terrestrial communications, HF signals may bounce from the ionosphere making worldwide communications possible, depending on the characteristics of ionosphere which in turn vary depending on the hour, temperature (presence of sun) and sunspot activity level. The bouncing is beneficial for long-distance communications when both transmitter and receiver are on Earth, but the satellite communications might suffer from the varying characteristics of ionosphere, which attenuates the received signal useless in the worst case.
- Very High Frequencies (VHF) ranges between 30 and 300 MHz. It contains, that is, TV channels 2–13. VHF already requires clearer LOS although the communications still allow scattering and diffractions.
- L-Band refers to the frequency band of about 0.5–1.5 GHz. This band is commonly used for terrestrial mobile communications. The higher bands of this range are already more clearly attenuated lowering the useful service area, that is, when utilized in terrestrial communications, the cell ranges are smaller than in lower frequency systems.
- C Band refers to the RF band of about 4–8 GHz from which the bands in 4 and 6 GHz are typically dedicated to satellite communications. More specifically, the band of 3.7–4.2 GHz is used for downlink communications together with 5.925–6425 GHz band in uplink, respectively. This range requires already clear LOS communications in order to work properly.

- Ka Band refers to the frequency band of about 18–31 GHz.
- Ku Band refers to the frequency band of about 10.9–17 GHz.
- Ultra-high Frequency (UHF), According to the ITU principles for the RF radiation, the HF band ranges in 300–3000 MHz. As an example, there is terrestrial TV in this area beginning at 470 MHz, The UHF channels are numbered from 14 up to 70.
- Superband refers to the frequency range of 216–600 MHz. It is used for fixed and mobile radio communications, as well as for additional television channels on cable systems.
- X-Band refers to the frequency range of about 7–8 GHz. Its primary use is meant for military satellite communications.

16.2.4 Characteristics of Satellite Systems

The useful lifetime of the satellites depend on the orbit in such a way that higher the satellite is located, more the life time is. A rough average for the satellite life time is in the range of 5–15 years.

Excluding GEO satellites, the satellites are in constant movement compared to the surface of Earth. In order to avoid service breakdowns, there are typically handovers applied between satellites. In other words, the satellites are typically communicating both with other satellites, as well as with the earth stations. One example of this type of system with channel awareness and possibility to have handovers is Iridium that is located into LEO. As a detail, Iridium is compatible with GSM systems.

Another challenge of the satellites with constant movements is the setting of the beam direction of the communications link between satellites or between satellite and end-user's equipment. One solution for this challenge is a set of phased antennas.

The communications via satellites can be generalized by the term Satcoms (Satellite Communications), which refers to the transmission of radio signal from the earth station or user terminal via uplink to satellite, and back via downlink to be received by an earth station or user equipment. Typical use cases for satellite communications are telephony, television and radio broadcasting, Internet access, weather information, location based services, space investigation and special communications like military data and voice service.

The delay of time to send the signal to satellite and back to Earth is referred to turnaround time. In practice, the functional satellite communications require line of sight between the earth station or user terminal and satellite in the uplink direction, and again between satellite and earth station or user terminal in downlink direction. Slant Range is the term used for the length of the path between a communications satellite and the respective earth station.

Earth station's transmitting frequency is different from the receiving frequency as the satellite makes the conversion in the relaying the signal. The difference between these signals is called translation frequency.

16.2.5 Functionality

Some of the most relevant terms in satellite communications are presented in the following.

Footprint is the area on the Earth's surface where the signal strength is functional for the successful communications. The level of the signal can be seen via EIRP contours of equal signal strengths as they cover the Earth's surface. It should be noted that different satellite transponders even within the same satellite may produce varying footprints and thus sets of received signal strength on Earth's surface.

Apogee refers to the farthest point of elliptical satellite orbit. Even with geosynchronous satellites that are based on circular Earth orbits are launched into elliptical orbits in the initial phase of the deployment, with apogee of 22 237 miles, as soon as the satellite has reached the planned apogee the satellite is placed onto its circular orbit. Once in orbit, the satellite has an ***orbital period*** which refers to the time for satellite to complete one complete circle of its orbit.

Azimuth refers to the horizontal rotation angle that a ground-based parabolic antenna must be rotated through to point to a specific satellite in a geosynchronous orbit. The azimuth angle for any particular satellite can be determined for any point on the surface of the Earth given the latitude and longitude of that point. It is defined with respect to due north as a matter of easy convenience.

Station keeping refers to the small scale orbital adjustments for maintaining satellite's orbital position within acceptable region (geostationary arc).

Spin Stabilization is needed to keep the satellite in controlled position. This technique makes the satellite spinning from the exterior of the spacecraft around its axis at a fixed rate. Furthermore, three-axis stabilization makes the body of the satellite maintain a fixed position compared to the orbital track and the Earth's surface. The reference axes are called roll, pinch and yaw, which are familiar from nautical terminology.

The time standard of satellites typically utilizes ***Zulu Time***, that is, Greenwich Meridian Time (GMT). As an example, INTELSAT and INMARSAT are based on Zulu Time for obtaining global synchronization.

Threshold Extension refers to technique which applies to satellite television receivers in order to increase SNR. The value may be around 3 dB which means 50% increase in terms of received power level. The threshold extension is especially useful for small antennas that are utilized only for receiving the satellite signals.

Subcarrier refers to the signal that is embedded onto main signal. Subcarrier is utilized for carrying additional information like audio along with main signal's video contents. In this typical case in satellite communications, the accompanying audio contents are based on the frequency modulated subcarrier. The amount of subcarriers can be in practice, that is, up to 4 for handling audio or data transmissions. It should be noted that these subcarriers may not always be related to the respective main carrier.

The ***turnaround frequency*** can be calculated by subtracting the satellite translation frequency from the uplink transmit frequency. This is used to calculate the Rx downlink frequency that the transmit signal will return at; this is also known as the return link or downlink frequency. The turnaround frequency can be thus marked as: *RxFrequency* (T/R) = *TxFrequency* – *TranslationFrequency*.

As an example, if the transmitting frequency of the earth station in uplink is 10.5 Ghz, and the satellite translation frequency is 650 MHz, the receiving frequency of the earth station in downlink is 10.5 GHz – 0.65 GHz = 9.85 GHz.

16.2.6 Equipment

Transponder is a set of equipment that consists of receiver, frequency converter and transmitter. This set forms physically part of the whole communications satellite. The power level of transponder is typically in range of few up to 10 watts. Transponders work typically in range of 36–72 MHz bandwidth in the L, C, Ku and Ka Bands. The number of transponders within satellite oscillates in range of 12–24 although it can also be considerably higher like in the case of INTELSAT VI which carries 50 transponders.

Transmitter is typically based on SSPA (Solid state power amplifier). It is gradually replacing older techniques based on travelling wave tubes of satellite communications systems because of lighter weight and higher reliability. The transmitted power of the satellite can be categorized into 3 levels: low, mid and high:

- Low-Power Satellite has transmitting RF power levels below 30 watts.
- Medium-Power Satellite has power levels between 30 and 100 watts.
- High-Power Satellite has power levels of 100 watts or more.

Satellite transmission is based on Mux, or ***multiplexer*** which combines various signals onto a single communication channel (transmitter with separate power amplifiers) and transmits them within the desired band. In the receiver side, demultiplexer is used for separating the individual signals.

Hub refers to a master station which handles complete communications between micro terminals. It should be noted that MESH networks can replace the "old-fashioned" hub principle as they provide connectivity between all relevant points of the network via onboard processing.

Engineering Service Circuit (ESC) refers to the narrowband voice service (300–3400 Hz band), as well as to the teletype service which is used for system maintenance. The communications of these services happens between earth stations, as well as between earth station and operations center. In case of analog satellite systems (FDM/FM), there are two S+DX channels dedicated within 4–12 kHz range of the baseband, whereas in digital systems, there is 1–2 maintenance channels with 32 kb/s or 64 kb/s data rate embedded to the general communications data stream of the earth station. It should be noted that typical ESC is able to communicate with both analog and digital satellite carriers and backhaul terrestrial links to the local switching center.

16.2.7 System Architecture

Earth Station refers to the whole equipment that consists of antenna, low-noise amplifier (LNA), down-converter and receiver. Earth station is used for receiving signals transmitted from satellite.

Backhaul refers to the satellite system's terrestrial communications part between earth stations and local switching elements.

Subsatellite Point refers to the assigned location over the Earth's Equator for geostationary satellite. The location of the satellites in coordinated way is needed in order to avoid any potential dangers of collisions.

16.2.8 Satellite Antennas

Antenna is a device for transmitting and receiving radio waves, that is, they act in the radio interface transforming the signals into electromagnetic radio waves at the transmitting end, and collecting the electromagnetic energy back to the signal for processing at the receiver end.

16.2.8.1 Antenna Types

Depending on the use and frequency, antennas may have variable forms. The simplest variant of the antenna is wire which can be, that is, vertical pole, dipole and longwire. There are countless forms for the wire antennas. The wire antenna may be close to omniradiating variant or directive. The rule of thumb for directive wire antenna is that the longer it is the more directive it is both in horizontal and vertical planes. Another typical variant is an antenna grid which can be, that is, Yagi or log-periodic. More elements they have, more directive the antenna is. The polarization of the antenna grid is dictated by the orientation of the elements. For satellite communications, typical antenna forms are horn, a helix and parabolic dish. More advanced antenna variants are based on phase array of active electronic elements that can also be flat or convoluted surfaces.

For dish antennas, which are the most typical for transmission and reception of broadcast systems, the downlink direction experiments strong signal path loss due to the relatively low radiated power level, which requires high-gain receiver antenna in the earth station side, which in practice is a large parabolic dish. The dish reflects the signal to the focal point which has feedhorn installed with brackets. The feedhorn is in practice a flared front-end of a section of waveguide gathering signals in the proximity of the focal point and transmits the signals to an element (probe) which, in turn, is connected to a low-noise block down-converter (LNB.) The task of LNB is to amplify the low-level signals, make the band bypass filtering and to convert the remaining frequency band to a lower band.

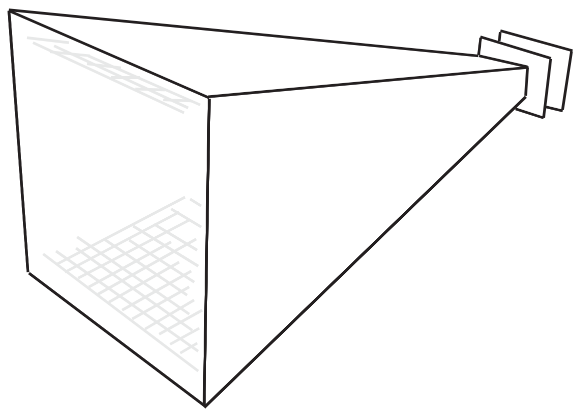

Figure 16.2 *A principle of typical satellite's horn antenna element which is highly directive.*

16.2.8.2 Mounting

Satellite antennas form ***Spot Beams*** which cover a certain geographical service area depending on the antenna pattern characteristics. Spot beams are useful to cover geographically isolated areas like big islands and important parts of remote areas.

On the satellite side, the ideal antenna has some important characteristics: it should be durable, lightweight, and highly directive. Horn antenna is one of the typical solutions due to suitable characteristics in the special operations. Figure 16.2 shows one example of advanced horn antenna element.

Antenna mount refers to the mechanism that supports ground-based satellite antenna [1]. aim of the earth station's antenna is that the mount should provide precise pointing angles towards the satellite. Furthermore, the mount should hold the antenna firmly so that only minimum deviations occur to the physical direction. For keeping track on the satellite's physical direction, the earth station's antenna mount should be constructed in such a way that the antenna element can move via two or more axes, in order to assure correct pointing towards the satellite in question.

Antenna mounts can be constructed as fixed, or in such a way that the steering of the antenna is possible. The fixed mount type has limitations because once it is done it can only be readjusted by making the change physically. The antenna with possibility for steering has adequate mechanism for moving the antenna physically so that the beam spot can be changed from one satellite to another, that is, via motorized or hydraulic mechanics. Typically, the steering solution provides possibility to remotely control the antenna direction either manually or automatically via preprogrammed schedule for the antenna directions.

The satellite systems are based on two different kind of antenna mounting. One type is EL/AZ mount which allows the antenna adjusting via the azimuth and elevation axes. The other type is Polar mount which provides the possibility to adjust the antenna in the hour angle and declination axes.

Figure 16.3 presents the EL/AZ mounting. In this case, the azimuth can be adjusted by rotating the antenna by the vertical azimuth axis. When the antenna is rotated along the vertical axis, the antenna beam follows a line parallel to the horizon. The elevation, on the other hand, can be adjusted by rotating the antenna along with the horizontal elevation axis. The rotation of the antenna via horizontal axis moves accordingly the antenna beam along a vertical line.

The adjustment of EL/AZ-mounted antenna is challenging if it needs to be pointed towards another geostationary satellite due to the need for adjustment of both azimuth and elevation, that is, there is a need to have two motions along the horizontal and vertical axis.

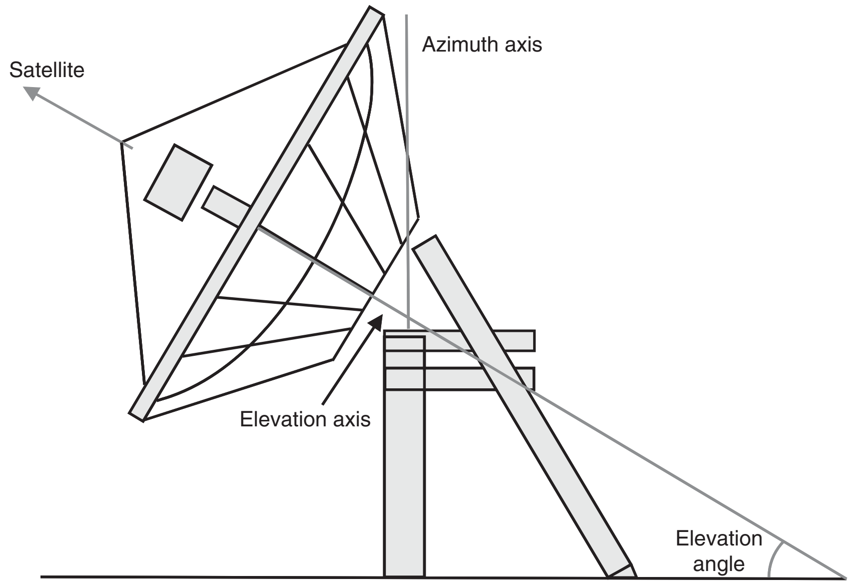

Figure 16.3 *The principle of EL/AZ mount of earth station's antenna element.*

The polar mount is sometimes referred to as equatorial mount. Polar mount is even more complicated to adjust compared to the EL/AZ mount. Figure 16.4 shows the principle of Polar mount.

Polar-mounted antenna should be aligned with respect to the Earth's polar axis. This happens in such a way that the hour angle is adjusted by rotating the antenna over the polar axis, that is, hour angle axis which is parallel with the polar axis of Earth. The rotation in this way moves the antenna beam along a line which is parallel to the Celestial Equator. On the other hand, the declination can be adjusted by rotating the antenna over the declination axis which is perpendicular to the hour angle axis, that is, parallel to the Equatorial plane. The antenna rotation over the declination axis moves the antenna beam along a line which is perpendicular to the Celestial Equator. When the hour angle is 0°, the antenna is pointed directly to the physical south or north of the globe. The relation of the declination related to satellite elevation and local latitude can now be calculated via $EL + DEC + LAT = 90°$.

The selection of the antenna mount depends on different factors. In general, polar mount has an important benefit over EL/AZ mount which is the fact that when the declination angle is correctly set, the earth station's antenna is able to track the Clarke Belt relatively accurately. This provides the possibility to change the antenna direction from one geostationary satellite to another by varying only the hour angle axis.

16.2.8.3 Antenna Characteristics

Polarization refers to antenna technique for increasing the capacity of the satellite communications by reusing the satellite transponder frequencies. In the linear cross-polarization case, the principle is simple: half the transponders can transmit the signals to downlink via vertical polarization whilst the rest of the signals are transmitted via horizontal polarization. The 90 degree difference in the transmitted signal's phase assured that the signals, although transmitted in the same frequency, will not interfere with each other. In order for earth station to decode the signals, the respective receiver antenna (feedhorn) must be properly adjusted to take advantage of the vertically or horizontally polarized signals. There are also solutions that provide simultaneous

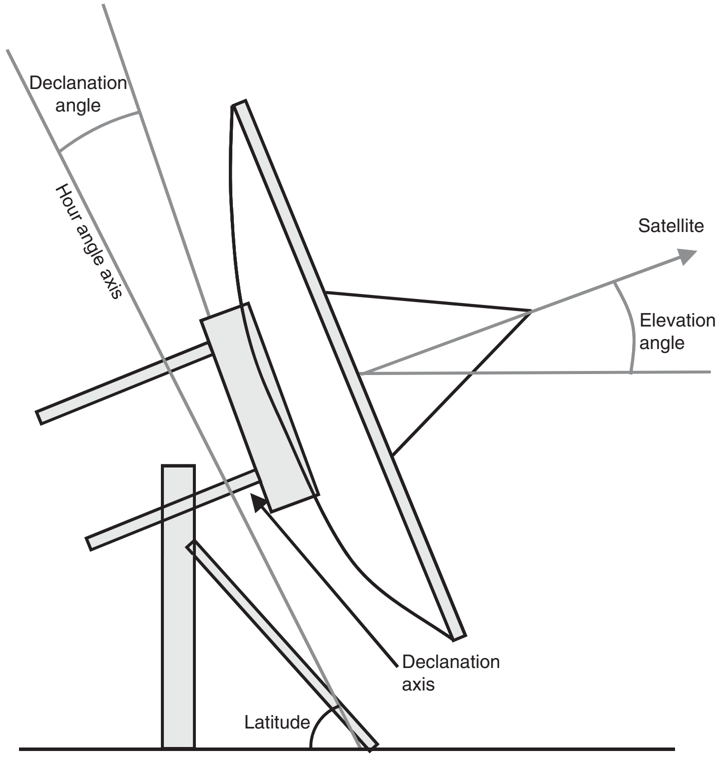

Figure 16.4 *The principle of the earth station's satellite Polar mounting of the antenna.*

reception of vertical and horizontal transponder signals which are then fed into separate receiving circuits. As an example, Intelsat system is based on left-hand and right-hand circular polarization, whereas many other systems are based on horizontal polarization. Circular polarization means that the antenna polarity is rotating constantly either clockwise (right-handed rotation) or counterclockwise (left-handed rotation). Sometimes right-hand and left-hand rotation is possible to transmit via the same frequency which in theory increases channel capacity by 100%.

The antenna diameter of earth stations vary typically from approximately 2 feet up to 12 feet, or around 60 cm up to 4 meters for typical TV signal reception. The size can also be considerably larger, up to around 100 feet or 30 meters for international communications systems. The typical antenna used for INTELSAT is around 13 to 18 meters or 40 to 60 feet.

As for the terminology, the aperture of satellite antenna refers to its cross-sectional area which is visible for the signal. Beam width refers to the angle or conical shape of the beam that the antenna projects. In general, the larger antenna is the narrower the respective beam width is, making it more efficient as energy can be concentrated more precisely towards the satellite, which in turn increases the transmitted and received power levels in that specific direction.

There are various forms of satellite antennas. The following lists some of the most typical and their characteristics.

Feedhorn is a receiver antenna for satellite TV. It functions in such a way that it collects the signal reflected from the main surface reflector. The signal is further directed towards low-noise amplifier (LNA).

Focal length of the antenna refers to the distance from the center feed to the center of the dish.

Focal point of the antenna refers to the area toward which the primary reflector directs and concentrates the signal received.

Elevation of the antenna refers to the uplink tilt toward the satellite. It is informed in degrees. At the elevation angle of zero, the antenna points to the horizon. On the other hand, the 90 degrees refers to the point directly upwards.

Cassegrain antenna refers to a method based on subreflector at the focal point. The reflector directs energy to or from a feed located at the apex of the main reflector.

Global beam refers to antenna's downlink pattern used by, that is, Intelsat satellites as the beam can create useful service area within about one-third of the Earth. Because of the relatively wide coverage area of global beam, the transponders of the satellites are able to create only lower radiating power levels. For this reason the earth stations that are receiving contents via global beams require considerably large antennas compared to the highly concentrated narrow beams. For the TV systems, typical diameter of such receiving antenna can be in order of more than 10 meters, that is, larger than 30 feet.

Gregorian antenna is a dual-reflector antenna system which uses parabolic main reflector and a concave ellipsoidal subreflector.

An example of early antennas is the Telstar transmitting antenna which was very small. However, the earth station's receiver antennae were considerably larger. As an example, the Andover horn-type receiver antenna was seven stories high and weighed 340 tons. It was extraordinarily much compared to modern satellite antennas with the size of an umbrella installed in the household roof-tops for receiving current commercial TV satellite contents. That is basically the difference between the technology in the initial phase of Telstar and current systems [2].

16.2.8.4 Antenna Gain

The antenna gain G can be calculated in the following way:

$$G = \frac{4\pi A}{\lambda}, \tag{16.1}$$

where A is the efficient area of the antenna. It is calculated as a product of antenna reflector area and antenna efficiency. The antenna efficiency, in turn, refers to the level of radiation of the power fed into it. In theory the maximum level would be 100%, though in practice, a typical efficiency figure may be in the range of 60–70%. The rest of the power is lost due to reflected power and other impurities of the matching from the antenna feeder to the antenna radiator. The wavelength can be generalized via the following equation:

$$\lambda = \frac{c}{f} = \frac{3 \cdot 10^8}{f}, \tag{16.2}$$

where f is frequency in Hz and c is the speed of light in m/s.

16.2.9 Challenges in Satellite Communications

The modern digital transmission has certain level of bit errors due to the practical effects on the radio interface combined with relatively long link distances. The logical solution for satellite systems is Forward Error Correction (FEC) which adds known codes to the original digital signal in the transmitting side at the source. For this reason, the occurred bit errors can be detected and corrected in the receiver side by comparing the received expected bits and the received bits.

There are various reasons for the increased bit error rates. One of those is solar outage which can occur in the situation when antenna is pointed towards satellite whilst the sun passes behind or near the satellite, within the functional radiation pattern of the antenna. It should be noted that this field of view is usually wider than the antenna's beam width dictated by the 3 dB attenuation points. Nevertheless, for fluent operations, solar outages can be calculated in advance.

There are also small variances between identical senses of polarity generated by two or more satellites. This can be corrected by skew which refers to an adjustment that compensates these variances in angles.

One challenge of the satellite position is that each satellite needs to be in a certain slot. It is longitudinal position in the geosynchronous orbit to which communications satellite is located in such a way that it does not harm the functionality of other satellites. As an example, above the United States, communications satellites have a typically three degree separation.

In communications, cross-modulation interference may occur. This refers to signal distortion, which is a result of modulation from other RF carriers (one or more). This problem is similar as in the case of terrestrial mobile communications systems when the neighboring cell is too close to another and has the same frequency in use. To avoid the problem correct frequency planning is required.

Blanking is related to visible distortion in television reception. The analog television signal is constructed by showing 30 individual frames per second. As the refresh rate is relatively high compared to the processing of the pictures by human brain, the outcome is blurred and gives a feeling to the observer of continuously moving pictures, which is basically the original idea of the television systems. Now, the blanking interval refers to the time slot of the television signal when one frame has been sent out to the receiver, and the next frame is still being transmitted.

Satellite transmission is sometimes affected by the weather conditions of the Earth. One example of this is rain. During the heavy rain, there might be outages in the service levels due to additional radio signal loss especially in Ku and Ka bands because of the effect of the absorption. The attenuation can be seen as a loss in the received power level of the electromagnetic signal, and if it exceeds the planned link budget margin, the receiver will not be able to receive the signal as the input signal is below the sensitivity level of the receiver.

Spillover, on the other hand, refers to the satellite signals that propagate to locations outside the antenna pattern. These signals outside the planned coverage areas may also cause interferences if they are not coordinated accordingly.

Van Allen radiation belts refer to two propagation areas in space that have highly charged particles and high-energy neutrons, and thus cause high interference levels to the signals of communications satellites.

In general, satellite communications interference refers to any energy levels which interfere with the reception of the desired signals. These interferences may be caused, that is, by transmitted signal levels of adjacent and neighboring channels, or by reflections of the radio signals from the surrounding areas such as from mountains or buildings. Microwave interference, in turn, refers to interference which is present when earth station with the beam directed at satellite receives another, typically stronger signal, from a nearby terrestrial microwave link. In addition, microwave interference can result from nearby radar systems, and even from the radiation of the sun. The most efficient way to attenuate microwave interference in practice is to move the antenna by some tens of centimeters up to about one meter which normally helps to enhance the signal-to-interference level sufficiently for functional connection.

The satellite receiver has certain limits from which one of the most important is noise figure (NF) which dictates the signal level at which the equipment still can receive correctly signals. The noise figure depends basically on the quality of the electrical components of the equipment. Noise figure is expressed in dB which indicates the device performance compared to theoretically ideal device.

Another important factor that limits the received signal levels is noise that is the result of the surrounding area, including cosmic noise level. Noise refers basically to any unwanted and unmodulated energy that is present in any signal. Johnson–Nyquist noise is the term for thermal noise which is a type of electrical noise resulting from thermal effects on the charged carriers like electrons within electrical conductor at the point of equilibrium. It should be noted that this effect does not have correlation on the voltage level. In general, this type of noise is called fluctuation-dissipation theorem, and its medium value is obtained via general impedance or susceptibility.

In the ideal environment, thermal noise is white which refers to the practically constant power spectral density over the whole selected frequency spectrum. In addition, the characteristic for white noise is that the signal amplitude is close to the Gaussian probability density function (PDF).

The thermal noise in terms of dB can be obtained via the following equation:

$$\mathrm{MDS[dB]} = 10\log_{10}\left(\frac{kT}{1 \cdot 10^{-3}W}\right) + NF + 10\log_{10} B + C/N. \tag{16.3}$$

In this equation, kT is the noise power level (dBm) when the bandwidth is 1 Hz. T is the environmental temperature in Kelvins (K), and k is the Boltzmann's constant (1.38×10^{-23} J/K) which corresponds −228 dBW/kHz. It is typically assumed that the temperature is 290 K which results in noise power of −174 dBm in 1 Hz channel. This refers to the noise floor of the system's input which means that any signal that has lower received power level can not be identified in typical situations. An exception for this principle is spread spectrum technique which is able to transmit and receive signals below the noise floor. It should be noted that once the noise floor value is obtained for 1 Hz channel in certain temperature (generalized typically to 290 K), all the other values of the noise floor can be obtained to all the power levels and bandwidths respectively. As an example, for the 1 Hz channel and 290 K, the noise power is −174 dBm. According to the principle described here, the noise level, that is, the thermal noise or Johnson noise for the 1 kHz channel can be calculated via the equation:

$$P\,[\mathrm{dBm}] = -174\,\mathrm{dBm} + 10\log_{10}(1000\,\mathrm{Hz}) = -144\,\mathrm{dBm} \tag{16.4}$$

The power of the signal is typically informed in dB values to ease and clarify the calculations. The commonly used method is the comparison of the power levels with 1 mW reference which can be informed in dBm values, and can be expressed in the following equation:

$$P\,[\mathrm{dBm}] = 10\log_{10}\left(k_B T \Delta f \cdot 1000\,\mathrm{Hz}\right). \tag{16.5}$$

In this equation, the constant 1000 is referring to the mW conversion from the Watt value. Furthermore, the equation can be expressed as:

$$P\,[\mathrm{dBm}] = 10\log_{10}\left(k_B T \cdot 1000\right) + 10\log_{10}(\Delta f). \tag{16.6}$$

In typical room temperature of T = 290 K, the equation becomes:

$$P\,[\mathrm{dBm}] = -174 + 10\log_{10}(\Delta f). \tag{16.7}$$

In this equation, Δf is the noise bandwidth and is expressed in Hz.

The noise floor for typical telecommunications systems can thus be calculated directly based on the respective bandwidth (Table 16.2).

In addition, in the space environment, there is also cosmic noise present. Its characteristics are comparable with thermal noise. It can be generalized that cosmic noise takes place in frequencies above 15 MHz, that is, when directional antennas points towards the sun or other spots that generate noise, like the center of the galaxy of the Milky Way. There are also many other sources of cosmic noise, like quasars which emit electromagnetic radiation in wide frequency spectrum, including the nonionizing radio frequencies. Also meteorites generate cosmic noise as a result of the friction they cause whilst falling down in the atmosphere of Earth, and respective ionized gazes which is possible to observe in radio frequencies. The most basic form of cosmic noise is wide spectrum CMBR, Cosmic Microwave Background Radiation which may be a result of the initial Big Bang.

Table 16.2 *Noise floor values as a function of noise bandwidth, and examples of respective telecommunications systems*

BW Δf	Noise floor [dBm]	Typical systems
1 Hz	−174	N/A
10 kHz	−134	FM radio
15 kHz	−132	Single LTE subcarrier
180 kHz	−121	Single LTE radio resource block
200 kHz	−121	GSM radio channel of 8 timeslots
1 MHz	−114	Bluetooth radio channel
2 MHz	−111	GPS radio channel
384 MHz	−108	UMTS radio channel
6 MHz	−106	Analog TV channel
20 MHz	−101	WLAN 802.11 radio channel
40 MHz	−98	WLAN 802.11n radio channel
80 MHz	−95	WLAN 802.11ac radio channel of 80 MHz
160 MHz	−92	WLAN 802.11ac radio channel of 160 MHz
1 GHz	−84	UWB radio channel

In addition to this noise level, there may also be external noise generators like transmitters of the other systems nearby which contribute to the total noise level at the receiver end. Nevertheless, the limiting factor as for the equipment itself is the noise figure (NF). It is noise factor (F) which is expressed in dB. F refers to the ratio of input signal-to-noise ratio (SNRi) and the signal-to-noise ratio of output (SNRo). The theoretical receiver has a noise factor NF of 1 which yields a noise figure NF of 0. This totally noiseless receiver is not possible in practice, and typical noise figures of, that is, mobile communications equipment are in range of 2 (excellent) and 5 (typical). The noise figure needs to be taken into account in the estimate of the minimum required received power level, that is, the reference level of the receiver is the sum of thermal noise and the noise figure of the equipment. It should be noted that if the bandwidth of the measured signal is not 1 Hz, the estimate needs to be calculated via the equation:

$$NF = 10\log_{10}(BW). \tag{16.8}$$

In this equation, BW refers to the bandwidth of the receiver.

In the actual communications, there is challenge with delay factor. It refers to the time for signal to propagate from the transmitter via satellite back to Earth, to the receiver. The transmission delay for a single hop depends on the height of the orbit of the satellite system, and possibly there is also additional delay in the digital processing of the signal (transmitting side may buffer the signal due to the interleaving and forming of the FEC correction).

The satellites are constantly moving and thus correction movements are needed. As for the physical functioning of the satellite, there may be dual spin applied. It means that the main construction of the satellite can be spun in order to provide altitude stabilization, and at the same time the antenna system is despun in order to maintain the beam direction correctly towards the planned coverage area of Earth. This dual-spin configuration thus serves to create a spin stabilized satellite.

In the actual communication, there might also be different radio interface related phenomena like deviation in the modulation level of an FM signal determined by the amount of frequency shift from the frequency of the main carrier. There may also be an echo present in the voice communications which refers to the reflection of the voice. The echo may be caused, that is, due to the reflecting signals along the path basically in the core network as the elements are interfacing with others and there may occur nonidealities in the connection points. The echo can typically be attenuated sufficiently by digital echo cancellers. In practice, echo canceller of satellite system is an electronic circuit which attenuates the echoing effect on satellite telephony links. Echo cancellers are largely replacing obsolete echo suppressors.

In the moving of the satellite along the orbit, there is eclipse present when the satellite travels through the line between the Earth and the sun or the Earth and the Moon. As for the functioning of the satellite, the implication of this is that the solar power charging does not function when the sun is not visible. Eclipse protected system refers to a transponder that can remain powered during the period of an eclipse.

There are also sun outages in cases when the sun is aligned behind the geostationary satellite towards what the reception antenna points to. This phenomenon happens twice per calendar year approximately at noon, and the effect lasts two weeks. It affects C-and Ku-band reception in such a way that the reception is disturbed for some minute time frame because the sun emits microwaves at the same frequencies that are used in transponders of the satellite. The phenomenon takes place in spring and fall.

Finally, the satellite has only a limited lifetime until it reaches the end of life (EOL) of a satellite. The challenge is thus how to remove the nonfunctional satellite from the system in a controlled way. Not only the satellite as such, but the related space debris is one of the most important issues in the satellite systems. According to Ref. [3], there are over 13 000 satellites and other large objects in orbit around the Earth, and countless smaller pieces of debris generated by spacecraft explosions and by collisions between satellites.

Due to the challenges of the end-of-lifetime management of satellites so far, the practice has been to place satellites into their orbits and simply leave them there. As the importance of the satellite systems has grown considerably also the orbital debris is growing alarmingly fast. This debris is a threat to all the other equipment as well as personnel of space as the orbital speed of the debris may be 7–8 km/s. The ones in constant danger are the Space Shuttle, the International Space Station, and the many satellites in Earth's orbit. The concrete problem is that the retired satellites, as well as the final parts of the launch vehicles and other related parts due to the separation of the launch vehicles, and also smaller parts like fragments of paint and frozen water, remain orbiting in Earth's atmosphere. When these fragments collide with each other and other debris, it results in even more debris. The amount of such debris is considerable at the moment, the estimates of the amount of this space junk being around half a million bits of particles with the size of 1–10 cm [4].

There are ideas for reducing the problem as presented, that is, in Ref. [5]. This is about the Swiss Space Center development of a nanosatellite with a size of 30 x 30 x 30 cm with the aim of cleaning up debris within low-Earth orbit. Several new technologies are being designed and multiple industrial partnerships are being established.

16.3 Voice and Data Services

Table 16.3 summarizes typical telecommunication satellites and respective orbits currently operating for public use. For more detailed information about satellite systems, please refer to [6–30].

As for the general naming, AsiaSat refers to the satellite systems covering the Asia mainland. Arabsat is in turn an Arabsat Satellite Organization providing regional telecommunications services for the Middle East region. Apstar (Asia-Pacific Star) is Chinese satellite system which carries commercial video services in the region. Furthermore, COMSAT refers to the Communications Satellite Corporation which serves as the US Signatory to INTELSAT and INMARSAT whilst ANIK is the Canadian domestic satellite system transmitting Canadian Broadcasting Corporation's (CSC) contents. COMSAT also delivers long distance voice and data services in Canada, to the USA and Mexico.

As of the closed systems, there are also military satellite systems for the largest nations. Defense Satellite Communications System (DSCS) is one example of the military communications for the use of the USA. DSCS is being replaced by the Wideband Global SATCOM system.

Table 16.3 *Examples of telecommunications satellites*

System	Orbit	Description
ACeS	GEO	Regional operator in East Asia, single satellite for voice and data.
Inmarsat	GEO	The oldest satellite telephony operator offering services since 1979. As the name suggests, it was meant mainly for the maritime solutions by international maritime satellite organization. Nevertheless, even if it initiated the services for ships, it currently also offers handheld terminals in a joint venture with ACeS. The system consists of 11 satellites. Wide coverage with exception of polar areas. Inmarsat can be used for voice, data and facsimile services.
Thuraya	GEO	Established in 1997. Covers about two-thirds of the world's population and with large satellite terminal penetration. Operates in Europe, Africa, the Middle East, Asia and Australia.
MSAT / SkyTerra	GEO	Is based on the similar terminals as Inmarsat uses. Plans to deploy service for handheld terminals in the Americas.
ICO Global Communications	GEO	Satellite telephony system which has launched a single geosynchronous satellite which is not yet active.
Intelsat	GEO	International Telecommunications Satellite Organization was an intergovernmental consortium that owned and managed a constellation of communications satellites and provided international broadcast services.
Eutelsat	GEO	French-based satellite provider.
ICO	MEO	Voice, data, facsimile, pager.
Teledesic	LEO	Broadband data network.
Globalstar	LEO	GSM-compatible voice, data, facsimile, pager and GPS.
Iridium	LEO	GSM-compatible voice, data, facsimile, pager.
Odissey	LEO	GSM-compatible voice, data, facsimile, text messaging.
SkyBridge	LEO	Broadband data network.
GE Americon	GEO	US corporation which provides satellite systems for domestic communications. It also has ownerships in international satellites.
VSAT	GEO	Very Small Aperture Terminal is based on a small antenna communication of data via satellites. Has point-to-point and broadcasting services.
XM Satellite Radio	GEO	XM Radio and Sirius have merged into a single company, providing satellite radio broadcast services.
TerreStar	GEO	A geosynchronous network offered to AT&T users.
SkyTerra	GEO	A company developing telecommunications systems which integrates satellite and terrestrial radio communication technologies into a single system.
Solaris Mobile	GEO	Solaris Mobile Ltd. is a next generation Mobile Satellite Service (MSS) operator providing access to satellite and terrestrial network infrastructure which support enhanced mobile communications across Europe.
Sirius Satellite Radio	GEO	Sirius Satellite Radio is a satellite digital audio radio service (SDARS) operating in North America, owned by Sirius XM Radio.
O3b	GEO	Provides satellite connectivity.
AMSAT	HEO	The Radio Amateur Satellite Corporation was formed as an educational organization. Its goal was to foster Amateur Radio's participation in space research and communication.

16.4 Broadcast Satellite Systems

16.4.1 Principles

Broadcast refers to the delivery of single transmission to multiple end-users. This is in practice the same as unicast method which transmits the contents (copies of original IP packets) to all the receivers in the service area in downlink direction.

There are public transmissions as well as closed services like pay TV. Furthermore, there are special services like Business Television which offers video and audio transmission via satellite that facilitates, that is, meetings and training.

ITU terminology contains BSS (Broadcast Satellite Service) although in practice, DBS or Direct Broadcast Service is more commonly used term in the satellite industry.

Satellite TV can be delivering the contents, that is, to CATV which refers to Community Antenna Television. In this concept, one business model is that separate companies may organize a large antenna in a strategic spot like in high mountain area in order to receive a relatively weak TV signal from a distant urban area. By amplifying the transmission and by modulating it onto television channels and sent along a coaxial cable strung from house to house.

Television and radio programs can be delivered via satellites that typically operate in Ku band (10.70–12.75 GHz) or in C band (3.7–4.2 GHz). The frequencies are converted in the reception side to intermediate frequencies of 0.950–2.150 GHz, and the signal is also amplified.

The delivery technique of the radio and television satellite systems into households can be done via several means. The signal can be delivered, that is, directly into the rooms, meaning that each household requires own satellite receiver equipment, whist the reception is done via individual antennas. Satellite Terminal refers to a receiver earth station which includes antenna reflector which is typically parabolic, feedhorn, low-noise amplifier (LNA), a down converter and receiver equipment. Furthermore, especially TVRO refers to television receive terminals using same type of antenna reflectors and equipment for receiving television and audio contents. In the smallest equipment case, USAT refers to ultra small aperture terminal. It is small terminal that are used typically in small home systems.

The delivery can also be done via centralized antenna which further distributes the signal to several receiver points as presented in Figure 16.5. The way for the delivery of satellite signal after the reception is to convert the intermediate frequency into lower band via transmodu-lation and deliver the signal within a cable system like DVB-C to the end-users.

Parabolic antenna is the most typical satellite antenna type for TV reception in both individual as well as for centralized reception. The term parabolic is due to the shape of the dish. The aim of the parabolic antenna shape is to focus optimally the relatively low energy level focusing it via the parabolic reflections to the unique focal point located in front of the dish, which at the same time typically has the actual feedhorn.

The third way to distribute the received satellite signal is to modulate the signal into some known TV system like PAL before the actual delivery of the signal to the end-users. This technique is called remodulation which refers to the demodulation of the received signal and modulation of the same signal in order to achieve compatibility with the end-user's television receivers.

16.4.2 Formats

For color television systems, the following lists typical, noncompatible formats of analog era:

- SECAM. This system was developed by the French and has been used also in Russia. The system is based on 625 vertical lines per received frame with a refreshing frequency of 50 Hz.

Figure 16.5 *An example of parabolic satellite antenna that is fed via the head located at the front of the antenna.*

- NTSC (National Television Standards Committee) is a video standard developed by the United States, and which is utilized also in various other countries. The system is based on 525 vertical lines and 60 Hz refreshing rate.
- CIF (Common Intermediate Format) is a compromise of the television display format developed by CCITT. It is possible to derive it from either PAL or NTSC.
- PAL (Phase Alternation System) is European TV standard which is based on 625 vertical lines and a 50 Hz refresh rate.

It should be noted that previously utilized analog methods are being heavily replaced by digital transmissions from which the following lists some commonly used ones:

- DVB (Digital Video Broadcasting) is a European driven standard to harmonize adoption of digital video, and to replace the former analog systems PAL and SECAM. For the satellite variant of DVB, there is DVB-S (DVB-Satellite) defined.
- ATSC (Advanced Television Standards Committee) made the ATSC standards which defines respective ATSC TV system in North America and South Korea. This is evolved version of the previous, analog NTSC (National Television Standards Committee) system.
- Japanese: ISDB-T (Integrated Services Digital Broadcasting, Terrestrial). For the satellite TV services two satellite systems are operating in Japan: BSAT and JCSAT. Former is utilized by digital WOWOW Broadcasting Satellite service, and the latter variant of digital TV broadcasting is utilized by SKY PerfecTV!
- DMB-T/H is local standard in China, and is used also in Hong Kong and Cuba.

16.4.3 Satellite TV

The satellite TV system contains uplink facility and respective antenna subsystem which by default is based on large dish antennas with several meters of diameter. The uplink beam is directed to respective satellite where transponders receive the contents. The task of the satellite is to forward the contents via downlink beam spots to specific areas to be received by earth stations via other frequency.

Typically, TV satellite has up to 32 transponders in Ku band, and up to 24 in C-band. The bandwidth of the transponders is typically up to 50 MHz.

For the location of the satellites in the geostationary orbit, the spacing between neighboring satellites needs to be dimensioned in such a way that the frequencies does not interfere the transmitters and receivers of each others. In practice, the satellites operating in Ku band, the physical spacing of the satellites require 1 degree spacing, resulting in 360 satellites in the geostationary orbit. For the C-band, the respective values are 2 degrees and 180 satellites. It should be noted that each band has special radio propagation characteristics, Ku band being sensible to rain, ice crystals and thunderstorms, and C band being sensible to terrestrial interferences.

In the downlink direction, the satellite transmission is relatively weak die to the small transmitting antennas. Again, earth stations use thus typically satellite dish antennas.

Satellite TV is typically based on the LNB. The advantage of LNB is the possibility to use economical cable-type to connect the indoor receiver with the satellite TV dish and LNB. Also, in the initial phase, the technology based on L-band and UHF band was considerably more economical compared to the technology based on C-band. This resulted in early adaptation of satellite TV receivers which in fact were modified UHF TV tuners for satellite TV reception, via down-conversion to lower intermediate frequency centered on 70 MHz where it was demodulated. As an example, in the United States, intermediate frequency of about 950–2150 MHz is typically used for delivering signal to the receiver.

The satellite receiver, or alternatively a separate settop box, demodulates the received signal. If the receiver is also able to decrypt the closed (payment) channel contents, it is called integrated receiver and decoder (IRD).

The analog TV broadcast formats are typically NTSC, PAL and SECAM television broadcast standards. Frequency modulation is used for the analog signal. The FM baseband signal contains video contents as well as audio subcarrier or several subcarriers. Furthermore, the audio subcarrier is demodulated. For the digital TV transmission, QPSK and its variants are typically used.

The TV broadcast which is possible to observe without additional payments of the usage is called free-to-air (FTA). The closed channels can be observed by using conditional access. For the encryption of the TV contents, conditional access encryption and scrambling methods are applied. There are various commercial solutions available in the market.

16.4.4 Satellite Audio and Radio

16.4.4.1 Audio

Satellite transmission via, that is, TV broadcasting, can contain audio subcarriers. It refers to the transmission of audio via a separate analog or digital signal carrier along with the main contents, which typically is video. Audio signal is thus modulated already once, and after that, it is further modulated into another signal in higher frequency and bandwidth.

The utilization of analog subcarriers for the delivery of audio is old method. Analog SCPC subcarrier audio can be received via dedicated satellite receivers in 50–90 MHz band. A digital SCPC (Single Channel Per Carrier) and MCPC (Multiple Channels Per Carrier) are examples of digital subcarrier audio encoding and modulation methods. The utilization of this started to increase along with the adaptation of VSAT technology, and further via the offering of capacity of DBS systems for commercial systems. Currently, the typical methods for digital satellite audio are digital single channel per carrier and multiple channels per carrier multiplexing.

Digital radio systems deliver variable bit-rate (VBR) data streaming. Different audio contents are multiplexed into a single transport stream. A statistical multiplexing optimizes the capacity utilization, and thus various video and audio channels can be transmitted simultaneously on a single RF channel.

Some examples of the digital SCPC and MCPC subcarrier-capable systems are DVB-S and the evolved variant DVB-S2 which are using MPEG-2 and MPEG-4, respectively. DVB-S, as well as many other digital satellite systems, are based on digital QPSK (Quadrature Phase Shift Keying).

16.4.4.2 Radio

The actual satellite radio refers to a high-quality broadcast audio service which is typically consumed in locations where other transmission methods are not available, like rural areas while driving a car. Satellite radio is typically commercial which requires a subscription. Characteristic to the satellite radio is that the contents has fewer advertisements, and may have considerably more available channels than terrestrial radio systems.

Sirius Satellite Radio is one example of various commercial systems. Sirius operates in the S-band which FCC has allocated to digital audio broadcasting in the USA. Sirius uses the 2.3 GHz S band in North America for nationwide digital audio broadcasting (DAB). Outside of the USA, Sirius is based on the DAB service in 1.4 GHz L-band. The Sirius receiver adapts to the varying conditions of the radio interface. Sirius receiver input frequency for the satellite and terrestrial reception is at 2.315 GHz. The tuner then down-converts the received signal to intermediate frequency (IF) at about 75 MHz. The signal is further digitized, demodulated, error-corrected, de-interleaved and decrypted. The streaming of the audio contents is based on buffering it for a few seconds in order to facilitate the switch of the reception from one satellite to another without notable interruptions. It also is possible to store contents of longer audio streams to the receiver equipment. The base band signal is finally formed to digital audio which goes to Digital-to-Audio (D/A) converter and is formed to notable analog audio signal.

The satellites of the Sirius are named as Radiosat 1-4. The first three were deployed in 2000. Radiosat 4 was built as a ground spare and was removed in 2012. All three operational satellites broadcast simultaneously, but only two of them are visible at the time due to the elliptical orbit.

16.5 Standardization

International cooperation is required for the coordination of the satellite orbit and frequency utilization. Frequency coordination is one of the key processes in order to control the RF interferences between satellite systems or between terrestrial microwave systems and satellites. As an example, in the USA, the frequency coordination is based on a computer service which utilizes a database in order to calculate and analyze potential microwave interference cases between satellite organizations planning the operation in same frequency bands. These frequency coordination studies are needed to determine potential problems before the actual deployment. As a base of the operation, frequency reuse technique is utilized maximizing the capacity of the links as the beams are isolated from each others and do not interfere with each others when the locations are coordinated properly. Also the varying of vertical and horizontal polarization further helps in the isolation of the interfrequency interferences.

Some of the key organizations for international standardization and regulation with direct influence or supportive role for the development or operations is listed in Table 16.4 [31–48].

Table 16.4 *Key organizations that have role in the satellite standardization, regulation or operation*

Name	Description
WRC (World Radio Conference) [31].	World radiocommunication conferences are held every three to four years, and are sponsored by the ITU. WRC reviews and revise the radio regulations, the international treaty governing the use of the radio-frequency spectrum and the geostationary-satellite and nongeostationary-satellite orbits. This job is executed based on agenda determined by the ITU Council by taking into account recommendations made in previous WRC.
FCC (Federal Communications Commission) [32].	Regulatory body in the USA, which controls the national RF band allocations, including satellite frequencies.
IFRB (International Frequency Registration Board) [33].	IFRB is an administrative body to regulate the use of frequencies. It acts under ITU and regulates also the allocation of satellite orbital locations.
ITU (International Telecommunication Union) [34].	ITU is the United Nations specialized agency for information and communication technologies, that is, ICTs. ITU allocates global radio spectrum and satellite orbits, develops the technical standards that ensure networks and technologies seamlessly interconnect, and strives to improve access to ICTs to underserved communities worldwide.
INTERSPUTNIK (international organization of space communications [35].	International body which is initiated by the Soviet Union. It provides global communications services via Russian satellite network.
NTIA (The National Telecommunications and Information Administration) [36].	NTIA belongs to the Department of Commerce. Its focus areas are the government telecommunications policies, standards creation and radio spectrum allocation in the USA. NTIA is an Executive Branch agency that is responsible by law for advising the President on telecommunications and information policy issues. NTIA's programs and policymaking focus on expanding broadband Internet access and adoption in America, expanding the use of spectrum and ensuring that the Internet remains an engine for continued innovation and economic growth.
TSS (Telecommunications Standardization Sector) [37].	Organization for global telecommunications standards creation. It is a combination of CCITT (Consultative Committee on Telephony and Telegraphy) which is today ITU-T, and the CCIR (Consultative Committee on International Radio) which is now ITU-R.
PTT (Post Telephone and Telegraph Administration)	PTT is global setup which is formed by operating agencies that are either directly or indirectly controlled by governments that have power to regulate telecommunications services. PTT refers to postal, telegraph, and telephone service and is a government agency responsible for postal mail, telegraph, and telephone services. Various PTTs have been partially or completely privatized. As a part of privatization, many PTTs were renamed or modified.
OFTEL (The Office of Telecommunications of the United Kingdom government), nowadays OFCOM (Office of Communications) [38].	OFTEL was a unit under the Department of Industries which regulates telecommunications in the United Kingdom. In 2003, the tasks of Oftel were transferred for Ofcom which is consolidated entity of the British telecommunication and broadcasting regulators. Ofcom is the communications regulator which egulates the TV and radio sectors, fixed line telecoms, mobiles, postal services, and the airwaves over which wireless devices operate.
NAB (National Association of Broadcasters) [39].	The National Association of Broadcasters is the voice for the US radio and television broadcasters. Among other tasks, NAB advances the interests of the members in federal government, industry and public affairs, and improves the quality and profitability of broadcasting.

(*continued*)

Table 16.4 *(Continued)*

Name	Description
ITSO (International Telecommunications Satellite Organization) [40].	ITSO is an intergovernmental organization with the mission to ensure that Intelsat Ltd. provides public telecommunications services, including voice, data and video, on a global and nondiscriminatory basis. ITSO has 149 member countries.
INTELSAT (The International Telecommunications Satellite Organization) [41].	INTELSAT is in charge of a network of international satellites. Intelsat is global provider of satellite services. During decades, Intelsat has been delivering information and entertainment for leading media and network companies, multinational corporations, Internet Service Providers and governmental agencies. Intelsat has satellite, teleport and fiber infrastructure for transmissions of video, data and voice services.
NASA (National Aeronautics and Space Administration) [42].	NASA is an US agency which administers the American space program. It includes the planning and deployment of commercial and military satellites. For the deployment, NASA utilizes a fleet of space shuttle vehicles.
JAXA (Japan Aerospace Exploration Agency) [43].	The Institute of Space and Astronautical Science (ISAS), the National Aerospace Laboratory of Japan (NAL) and the National Space Development Agency of Japan (NASDA) were merged into one independent administrative institution in 2003 to be able to perform all their activities in the aerospace field as one organization, from basic research and development to utilization. The independent administrative institution is the Japan Aerospace Exploration Agency (JAXA).
NCTA (National Cable Television Association) [44].	NCTA is the principal trade association for the U.S. cable TV industry. It represents cable operators which serve major part of the nation's cable households. It also represents over 200 cable program networks, and various equipment suppliers and providers of other services to the cable industry. NCTA provides its members with unified voice on issues affecting the cable and telecommunications industry. NCTA promotes the growth of the cable industry and manages the industry's regulatory and legislative priorities.
ACTS (Advanced Communications Technology Satellite) [45].	An experimental satellite project of NASA for evaluating the use of the Ka-Band (30/20 GHz) services. ACTS was launched in 1993. It offers breakthroughs in satellite communications. ACTS was conceived as a partnership project between NASA and industry. ACTS was the first satellite with the ability to carry digital communications at standard fiber optic data rates with the same quality of transmission. ACTS technology integrated with existing ground fiber optic systems for transmission of high speed data rates to very remote locations.
ISO (International Standards Organization) [46].	ISO develops International Standards. ISO was initiated in 1947, and since then it has published more than 19 500 International Standards covering almost all aspects of technology and business. The topics vary from food safety to computers, and agriculture to healthcare. ISO creates global standards like JPEG and MPEG. ISO works in cooperation with the CCITT.
EBU (European Broadcasting Union) [47].	EBU is global alliance of public service media organizations. It has members in 56 countries in Europe and other parts. EBU defends the interests of public service media. It also takes care of industry knowledge and expertise. EBU operates EUROVISION and EURORADIO. The EUROVISION and EURORADIO satellite and fiber network are meant for delivering global public service media.
Eutelsat [48].	The European Telecommunications Satellite Organization which is headquartered in France. It provides a satellite network for Europe and parts of North Africa and the Middle East.

16.6 Commercial Satellite Systems

16.6.1 ACeS

ACeS (Asia Cellular Satellite) is a regional company operating satellite telecommunications. It is based in Indonesia, and it offers GSM-type of satellite telephony services to Asian market. The coverage area includes local areas, including Indonesia, Malaysia, Thailand, Philippines, Sri Lanka, Vietnam, China and India. The company operates the Garuda 1 satellite which was deployed in 2000.

The ACeS satellite and most network operations are controlled by ACeS Network Control Center which is located to Indonesia. The ground stations are operated by NSPs which in turn offers communications between terrestrial telephone networks.

ACeS uses virtual country code +88 220. ACeS has international roaming agreements with various GSM networks which are used via dual mode GSM / satellite handheld terminals. It is possible to use a major set of typical standard services, so in addition to the basic calls, there is also possibility to use call forwarding and call waiting. Nevertheless, ACeS terminals currently do not have short message service (SMS) in satellite mode. In addition to the handheld device, ACeS also has a land phone which has been useful for covering rural and remote areas where no other telecommunications infrastructure is available.

The ACeS network customer base has been estimated as about 2 million subscribers, but in practice, the amount of subscribers was minimum. Together with other challenges in the failures of the functioning which lowered the offered capacity of the satellite, the company started cooperation with Inmarsat in 2006. Nevertheless, the company has ceased the operation in Indonesia during 2011, and Inmarsat currently has the remaining assets of ACeS.

The communications satellite Garuda 1 which has been the base for ACeS is developed in Indonesia. It contains two antennas with diameter of 12 meters. It is placed in geostationary orbit at 123° East which is suitable location for communications services in Asia regions (Table 16.5). It has a total of 88 transponders which is sufficient for serving the whole continent with 140 spot beams. There also were plans to include second satellite, Garuda 2, in 1999, for backup functions to Garuda-1 and gradually for expanding satellite communications coverage also to western and central Asia, the Middle East, Europe and northern Africa. Nevertheless, due to the economical challenges and lower business than was originally planned, the launch of Garuda 2 did not happen in practice. Nevertheless, with the given maintenance, the expected life time of Garuda 1 is until 2015.

Garuda-1 satellite is based on the A2100AXX bus. The initial power level of the satellite was 14 kW, and by the end of the lifetime the power is estimated to be 9 kW. This results in Garuda-1 being one of the most powerful communications satellites so far. Garuda-1 weighs 4500 kg, and its planned lifetime is 12 years.

The basic principle of ACeS has been to offer advanced mobile telephony and data service communications via handheld terminals, providing voice, data and fax services. The RF interface towards the terrestrial gateways is based on links in C-band link. The communications with the mobile terminals happens via links

Table 16.5 *Characteristics of ACeS*

Name	ACeS
System	GEO
Inclination / degrees	123° East
Satellites, main	1
Satellites, spare	1 spare planned but cancelled
Operational lifetime / years	12

in L-band, that is, via the 140 spot beams. It should be noted that the coverage area contains about 3.5 billion people.

16.6.2 Telstar

The AT&T Corporation maintains Telstar and currently operates its domestic satellite system under the Telstar name. Telstar is actually term for describing a set of communications satellites. The first two Telstar satellites were deployed merely for experimental purposes. Telstar 1 was launched in 1962, and it could be used for the first television pictures, telephone calls and fax pages. It also made possible the first live transatlantic television transmission making it the pioneer in the initial satellite broadcast and telecommunications era. Telstart-1 weighed 171 pounds (77 kg), and its spherical size was 34.5-inches (0.87 m), with power system based on solar panels which produced 14 watts. Telstar-1 was spin-stabilized. Telstar-2 continued, and it was launched in the following year 1963. Although Telstar-1 and Telstar-2 are not operational any more, they remain orbiting.

The initial Telstar setup was based on cooperation between US companies/organizations of AT&T, Bell Telephone Laboratories, NASA, Brittish GPO and the French National PTT with the aim of development of experimental satellite communications between Europe and Americas. The US ground station was build by Bell Labs and it was located in Andover. There was also a British ground station located in southwestern England.

The original Telstar had a single transponder for data relaying that could be either a TV channel or multiplexed set of telephone lines. The antennas of Telstar-1 were omnidirectional, forming an array of antenna elements around the satellite belt. Te RF band was in 6 GHz for the reception, and the transponder converted the uplink frequency to 4 GHz downlink frequency. There was also a helical antenna element for telecommands from a ground station.

Telstar-1 was medium-altitude satellite in an elliptical orbit with time for complete round of 2 hours and 37 minutes. The inclination was 45 degrees to the Equator, and perigee of 952 kilometers (592 miles) from the surface of Earth. The apogee was about 5933 kilometers (3687 mi). It should be noted that currently major part of the communications satellites are based on circular geostationary orbits instead, unlike the nongeosynchronous orbit of Telstar-1. The latter had the drawback of communications being limited to 20 minutes at a time in each orbital round as the satellite moved over the Atlantic Ocean.

Due to the relatively low transmitting power level of Telstar-1, the ground station antennas had to be respectively very large in order to maintain the radio link budget functional. The ground station was thus based on conical horn antenna with a parabolic reflector. The antennas were 177 feet (54 m) long and weighed 380 tons (340,000 kg).

The second wave of Telstar satellites includes Telstar 301 in 1983, Telstar 302 in 1984 and Telstar 303 in 1985.

The following wave includes Telstar 401 in 1993 and Telstar 402 in 1994, although their lifetime was short due to accidents. The following Telstar was 402R in 1995 which was later renamed as Telstar 4. Telstar 10 followed in 1997, and in 2003, Telstars 4–8 and 13 became part of Intelsat. Telstar 4 was damaged, and the others were renamed the Intelsat Americas 5. Telstar 18 followed in 2004.

One example of the Telstar services as a base for a commercial solutions is the Q-KON que-Vi satellite. It provides broadband Internet access from the Europe backbone to Africa via satellite beams. The que-Vi integrates the following:

- Telstar 11N Ku-band satellite space segment.
- Linkstar DVB-S2 VSAT network.
- European teleport facilities.

- 1st Tier Internet access.
- Terrestrial connectivity networks.
- The Q-KON NOC for network and customer quality-of-service management.

The que-Vi is suitable for providing IP connectivity, that is, for the following:

- Business applications which need communications link between offices.
- Internet hot spots.
- IP telephony.

More information about Telstar history can be found in Ref. [49].

16.6.3 Globalstar

Globalstar represents LEO satellite constellation. Globalstar is designed for satellite telephony and low data rate data communications. Technically, it is thus similar system with Iridium Orbcomm. Businesswise, Globalstar project was started in 1991 as a result of joint initiative of Qualcomm and Loral Corporation. The base of the system advanced along with the spectrum allocation of FCC for the USA operations in 1995. Finally, after difficulties in the launches, the complete set of 48 satellites and four spare satellites were in orbit by 2000, along with the full commercial service in North America, Europe and Brazil.

Even with economical challenges, Globalstar is the most significant provider of mobile satellite voice and data services measured by the number of users which was over 300 000 by 2008, with services in more than 120 countries.

Globalstar system devices include portable and fixed telephones, simplex data modems, as well as duplex voice and data modules. The Globalstar services are voice telephony, one way mobile terminated text messaging, CS and PS data services up to 9,6 kb/s, one way mobile originated "short-burst messages" for simplex devices, and device geolocation service.

Globalstar consists of 48 operational satellites, and the four spare satellites are located also on the orbit (Table 16.6). Globalstar network contains ground gateway stations that link the satellites with PSTN and Internet. The telephony number plan of the system is compatible with the international numbering plan of ITU, with the exception of Brazil which has a special country code prefix of +8818.

It should be noted that Globalstar does not have intersatellite linking. This means that the satellites use gateway ground stations in all the connectivity. This, on the other hand, provides customers with regional phone numbers. The drawback of this solution is that the service may not be possible if the gateway station is missing in certain remote area like North Pole, regardless of the possible visibility of the satellite in these areas.

Table 16.6 *Characteristics of Globalstar*

Name	Globalstar
System	LEO
Inclination / degrees	52
Satellites, main	48
Satellites, spare	4
Operational lifetime / years	7.5
Height, approximately from Earth's surface / km	1400

The radio interface is based on CDMA. For authentication, there are telephones based on standard GSM SIM as well as on the IS-41. In practice, the Globalstar gateways are required to support both core solutions.

It is also possible to use Globalstar terminals in roaming with a set of cellular network operators in limited areas. The benefit of this solution is the possibility to use the same subscriber number via both satellite and cellular networks in cases where roaming is valid.

Globalstar is a mobile satellite system that has a network of 48 satellites. Globalstar offers practically global footprint for delivering voice and data service, except in certain low-population areas like north and south poles. It also offers handheld satellite messaging and tracking personal safety device, SPOT Satellite Messenger. The services include voice telephony, one-way MT SMS text messaging, CS data of 9,6 kb/s, 9,6 kb/s PS connections to Internet, one-way MO "short-burst" messages for simplex devices, as well as device geolocation service.

The system is a result of a Globalstar project which was initiated in 1991. FCC granted Globalstar a US spectrum in January 1995. Consequently, the first satellites were placed on the orbit in 1998, and after some drawbacks in the launches, the rest of the satellites were in place. The system uses 48 active satellites and 4 spare satellites placed in orbit. There also have spare satellites launched and later also a set of second generation satellites in order to keep the system service level adequate. The Globalstar second-generation constellation is planned to include 32 LEO satellites.

The system was ready for the full commercial service in 2000, covering North America, Europe and Brazil. As many other satellite companies, also Globalstar experienced financial difficulties, but was restructured.

The system architecture of Globalstar satellite module includes a communications system in S and L bands. The body of the satellite has trapezoidal form, and it contains two solar arrays. The Globalstar satellites are located at an altitude of 1414 km (876 miles).

For the actual calls, in the downlink direction, various satellites transmit the signaling of the A-subscriber to the antenna of a gateway which routes the call via terrestrial telecommunications system. As Globastar does not have possibility to communicate between satellites unlike, that is, in Iridium, the communications of each satellite needs to be routed via a gateway station. One result of this system architecture is that the respective gateway ground stations require the use of respective regional phone numbers for the Globalstar handsets. It should be noted that due to this restriction, if there is lack of gateway stations in certain remote areas, as is the case in polar regions, the Globalstar service is not possible to use regardless of the visibility of the satellites.

The radio interface of satellite communications is based on CDMA. The satellites contain the hardware whilst the SW of the system is on the elements of the Earth which lowers the complexity of maintenance. Globalstar satellite constellation is controlled by the Satellite Operations Control Center (SOCC).

Globalstar satellites are thus relatively simple repeater elements. The life expectancy of the first generation satellites is about 7.5 years whilst the second generation satellites have planned life expectation of 15 years. The Globalstar satellite orbits have an inclination value of 52 degrees. This means that Globalstar does not provide coverage in polar areas.

A set of ground gateway stations handle the connectivity from the Globalstar satellites to PSTN and Internet in such a way that users have fixed line telephone numbers based on the North American Numbering Plan, or alternatively based on telephone numbering plan of the country where the gateway is located to, in case it is outside the USA. An exception is Brazil which has a dedicated country code +8818 for Globalstar users.

The Globalstar radio interface is based on the CDMA of Qualcomm. There are also handsets available including cellular mobile system support of GSM or IS-41 CDMA. This also means that the Globalstar gateways must support both of these technologies, also for the authentication of the customers. In addition, there are roaming agreements for Globalstar customers in some regions via multimode Globalstar handsets. Figure 16.6 shows the overall idea of Globalstar call.

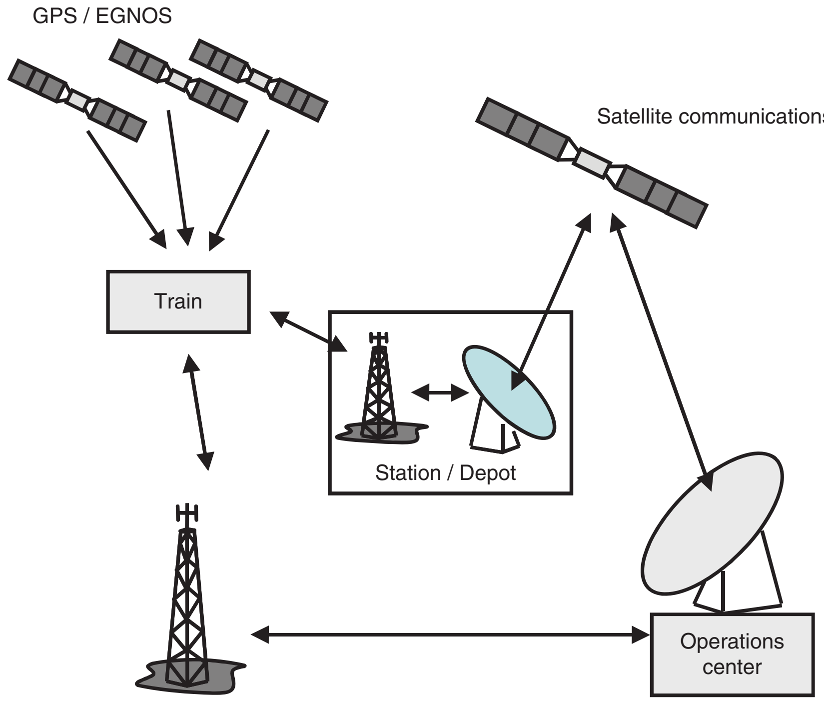

Figure 16.6 The principle of the Globalstar satellite call.

The major set of the second-generation satellites were launched in 2010–2012, and the last ones were placed on the orbit in 2013, to complete the system of 24 new satellites.

More information about Globalstar can be found in Ref. [50].

16.6.4 ORBCOMM

ORBCOMM is a provider of data communications via a network of LEO satellites. Typical use cases include monitoring and alerts without the added cost of broadband or voice, in addition to the two-way data communications. ORBCOMM also offers cellular data subscriptions for assets located or traveling within coverage in order to reduce the cost associated with monitoring or controlling the assets [51]. ORBCOMM thus provides M2M global asset monitoring and messaging services via 29 LEO satellites (Table 16.7).

Table 16.7 ORBCOMM satellite system characteristics

Name	ORBCOMM
System	LEO
Satellites, main	35 (originally), 29 (currently)
Height, approximately from Earth's surface / km	775

Communication between subscriber units and satellites is based on SDPSK modulation with data rates of 4.8 kb/s in DL and 2.4 kb/s in UL.

ORBCOMM satellites have 56 kb/s backhaul which is based on TDMA multiplexing and QPSK modulation. In addition, the satellites are equipped with GPS receiver. The data payload of the system is 6–30 bytes which is sufficient for transmitting GPS position data or sensor readings.

ORBCOMM is used for satellite data services. It is suitable for transferring relatively low amount of data suitable for emails and messages.

ORBCOMM has also been designed to provide Automatic Identification System (AIS) for tracking ocean vessels, although it is not currently in use.

The user equipment transmission time is restricted to a maximum of 1%, meaning that they are permitted to send 450 ms data burst two times within 15 minute time windows.

ORBCOMM has control centers in the United States, Brazil, Japan and Korea. ORBCOMM also has ground stations in New York, Georgia, Arizona and Washington States in the USA, and in Curaçao, Italy, Australia, Kazakhstan, Brazil, Argentina, Morocco, Japan, Korea and Malaysia.

The system operates in the VHF band, in the 137–150 MHz frequency range. The benefit of the low frequency is that it is possible to use the equipment with lower power levels compared to the other satellite communications systems. The drawback is that the system requires larger antennas which in practice can be one meter or more for obtaining typical performance.

16.6.5 Mars Odyssey

Mars Odyssey is a robotic spacecraft designed to orbit the planet Mars. The source Ref. [52] shows the project plan of Odyssey. The focus of the Mars Odyssey is to explore the past and present water and volcanic activity of the surface of Mars by utilizing spectrometers and electronic imagers. In addition to the actual scientific investigation instrument, Mars Odyssey functions for relaying data communications between the Mars Exploration Rovers, Mars Science Laboratory and the Phoenix Lander to Earth.

The journey of Odyssey was initiated in 2001, and it reached Mars orbit in 2001. The actual scientific activity was initiated in 2002. The instruments the Odyssey carries are Thermal Emission Imaging System, Gamma Ray Spectrometer and Mars Radiation Environment Experiment.

Odyssey's teleco subsystem includes a radio system which operates in the X-band microwave frequency band, and a system that operates in the UHF band. This subsystem provides technical base for all communications throughout the planned phases of the mission of Odyssey. The X-band is used for communications between Earth and the orbiter, and the UHF band is used for communications between Odyssey and any landers present on the Martian surface at any given time [52]. The telecommunication subsystem weighs 23.9 kilograms (52.7 pounds).

The objectives of 2001 Mars Odyssey mission are related to the continuing of the global reconnaissance of Mars. For this, the mission produces high spatial and spectral resolution images of surface mineralogy, images of surface morphology, and studies near-Mars radiation, global map of elemental composition, and shallow surface abundance of hydrogen.

16.6.6 SkyBridge

SkyBridge is a satellite broadband access system which provides services like high-speed Internet and video conferencing with global coverage [53]. Skybridge offers global telecommunications services for business and residential users including interactive use of multimedia applications as well as LAN and ISDN applications. The focus of the system is to offer services to fixed terminals, though including the ones that provide user mobility and terminal portability. The system is suitable interactive, real-time applications due to the relatively low delay of the signal path from terminals to satellites and back to Earth. Skybridge provides also means for interconnection of satellite cells to be used as wireless local loop, and it offers enhanced narrowband voice services, videoconferencing and data transmission.

The system focuses on urban, suburban and rural areas lacking broadband terrestrial communications infrastructure or which are not sufficiently economical to cover via other type of telecommunications networks. It can thus be generalized that SkyBridge is one type of broadband wireless local loop system.

SkyBridge is based on 64 LEO satellites linking professional and residential users to terrestrial gateways via low-cost satellite terminals. It should be noted that the user equipment is not specific to the system. The system architecture of the receiving sites may thus be based on the following configurations:

- Individual reception.
- Community reception (a SkyBridge terminal is shared between several subscribers).
- Professional configuration where the SkyBridge terminal is connected to a LAN or a PBX.

Figure 16.7 shows the architecture of the professional configuration.

SkyBridge spot beams have about 350 km radius and they serve the SkyBridge terminals connected to one or several user terminals and one or several SkyBridge gateways. In order to utilize the services, end-user is registered in a single gateway that concentrates the incoming and outgoing traffic. The system supports various multimedia-capable user terminals providing terrestrial applications.

A single Skybridge satellite provides a coverage area with 3000 km radius which consists of fixed spot beams of 350 km radius each as shown in Figure 16.8. When initiating the communications, users attach to a single SkyBridge gateway which acts as a concentrator of the traffic for each user registered to it. The role of gateway is to provide access network independently from the other gateways.

Skybridge satellites are placed on circular orbit which has an altitude of 1457 km above the Earth. The system consists of two symmetrical Walker subconstellations, each having 32 satellites. LEO constellation provides relatively low signal propagation delay, in order of 20 ms, which is compatible with TCP/IP protocols. This assures that SkyBridge links are not disturbed by the performance requirements of typical terrestrial protocols. These similarities in the performance between Skybridge and terrestrial communications provide sufficiently high quality of service which results in seamless integration into terrestrial networks. The SkyBridge system is thus suitable for interactive applications.

SkyBridge is planned for asymmetrical broadband use with fixed telecommunications networks. It provides a maximum of 60 Mb/s data rates in the downlink, and up to 2 Mb/s in uplink direction. Professional terminals have been planned to offer also higher data rates. It should be noted that Skybridge takes into account from the initial phase the characteristics of Internet communications in order to cope with the bursty nature of the IP communications and asymmetric data transmission. As the data rates of Skybridge are defined with 16 kb/s steps, that is, relatively small differences of the data rate categories, the system provides a functional base for the concept of bandwidth on demand.

SkyBridge is suitable for interactive real-time applications such as:

- High-speed satellite access to Internet
- Efficient utilization of online services with comparable QoS as via terrestrial access networks.
- Access to business servers and LANs, email and file transfer services.
- Video conference and video telephony.
- Telemedecine.
- Entertainment services, including interactive VOD (Video On Demand) services and video games.

As for the positioning of Skybridge compared to the terrestrial networks for the offering of broadband access, the benefit of SkyBridge can be seen as a complementary solution rather than competing solution. Technically,

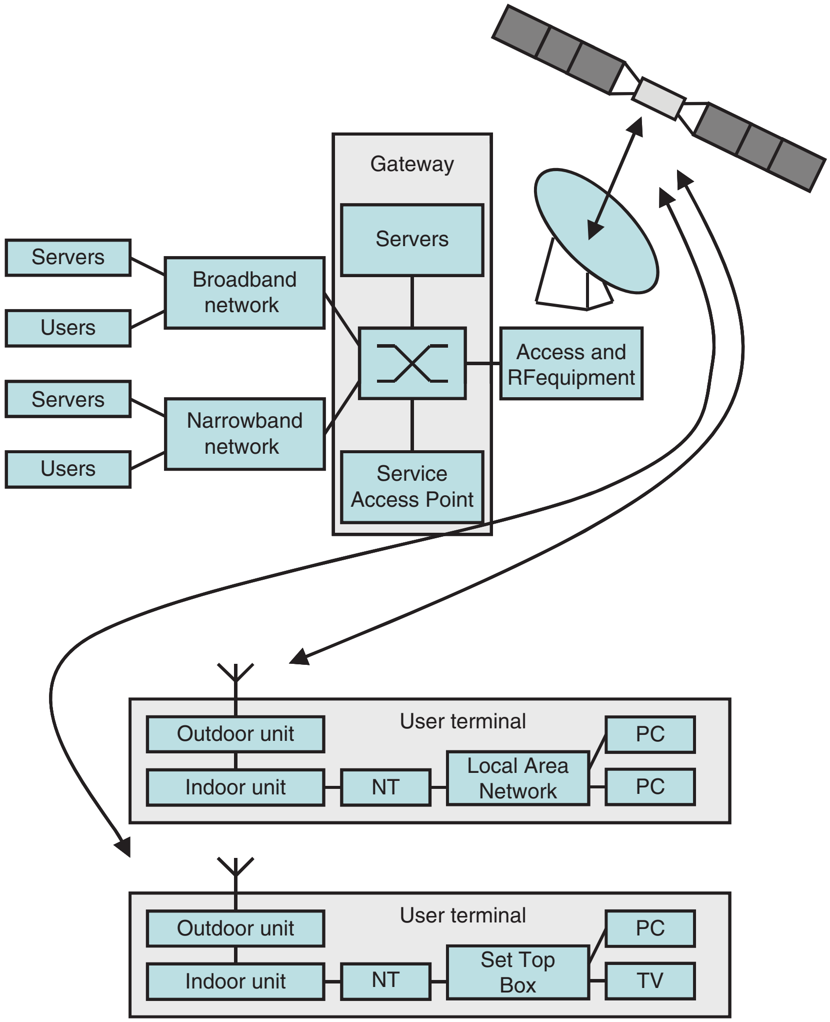

Figure 16.7 *Example of the Skybridge architecture for the professional service category. The gateway has ATM switch which can handle narrowband and broadband transmissions, as well as leased lines.*

the performance of SkyBridge is comparable to ADSL (asymmetric digital subscriber line) and HFC/FTTC (hybrid fibre coax / fibre to the curb) terrestrial solutions. The difference is about markets and location where access can be provided.

SkyBridge is designed to operate in the Ku band which provides efficient frequency reuse and thus minimizes the interfrequency interferences, including with the colocation of other satellite systems in the same band. In order to minimize the interferences with other geostationary systems, SkyBridge includes a nonoperating zone around geostationary arc. When the SkyBridge earth station sees the satellite entering the nonoperating zone, the transmitted satellite beam is switched off and there is a handover of connection to another available Skybridge satellite. This handover procedure is transparent to the users. Equally the same principle is applied to minimize interferences from SkyBridge gateway or terminal to geostationary terminal. International Telecommunication Union (ITU) defines the rest of the cases via international regulation.

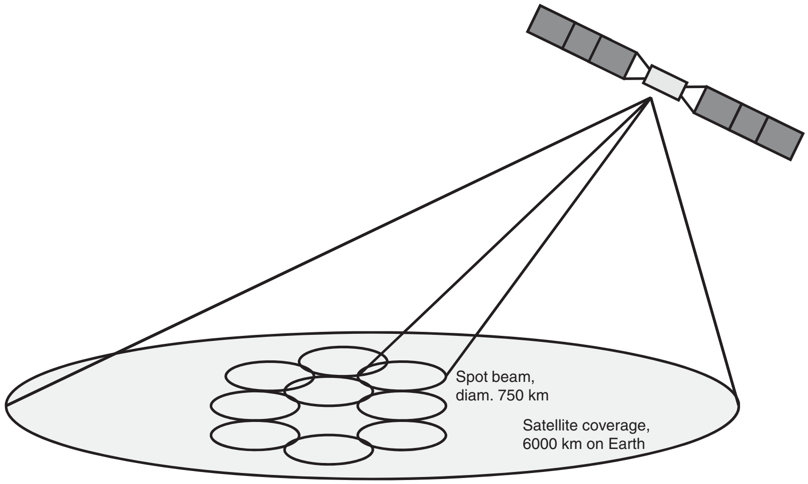

Figure 16.8 *The principle of the spot beams of SkyBridge.*

The principle of SkyBridge is to offer cost-effective access for the end-users for high-quality broadband and narrowband services. The aim of the systems is thus to provide universal access to communication services via satellite links [54].

16.6.7 Iridium

Iridium is based on constellation of 66 interconnected LEO satellites as shown in Figure 16.9 [55, 56]. The core network of Iridium includes a gateway located to Arizona, satellite network operations center in Virginia, a technical support center in Arizona and a total of five tracking, telemetry and control (TTAC) stations in Canada, Alaska, Norway and Arizona. All elements are interconnected via fiber optics and broadband satellite links.

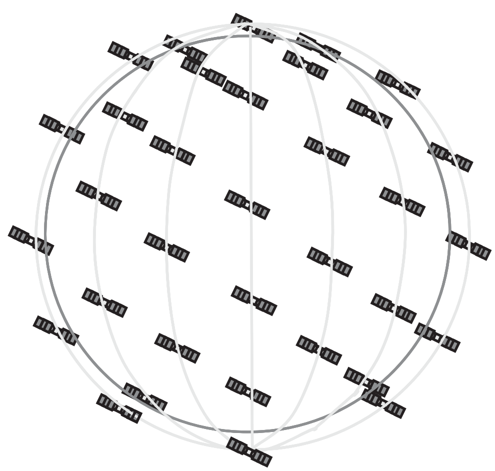

Figure 16.9 *Iridium constellation [55, 56]. Figure adapted from the original of Iridium.*

The Iridium's services include voice and data traffic, and satellite backhaul data links. Iridium also has a gateway for closed use of US government communications.

The Iridium satellite constellation is a large group of satellites providing voice and data coverage to satellite phones, pagers and integrated transceivers over Earth's entire surface. Iridium Communications Inc. owns and operates the constellation and sells equipment and access to its services. It was originally developed in 1992, and subsequently implemented in October of 1999.

In addition to the 66 active satellites, there is also a set of additional spare satellites for the cases of failures. Iridium operates in LEO constellations with height from Earth's surface of about 781 km (485 miles). The inclination of Iridium satellites is 86.4°.

Iridium has intersatellite communication links satellites via Ka band. For it, satellites may have up to four links with other satellites. Two of these links are meant for the satellites orbiting on the same plane, and the other two links are defined with satellites located to neighboring planes.

As the satellites move from pole to pole, and due to the amount of the satellites, the visibility and thus the service level of the satellites is good at all times also at the North and South poles. Iridium is based on the satellites rotating to different directions in planes next to one another. The active satellites are located to a total of 6 orbital planes with 30 degrees separation between the planes, resulting in a total number of 11 satellites per plane. As a detail, the original plan of Iridium would have consisted of 77 satellites (Table 16.8), which was the original reference for the name of the system (Iridium referring to the atomic order of 77).

Iridium Satellite System is meant for mobile telephony. The key characteristics of Iridium network are:

- Pole-to-pole coverage which would require only one gateway without reliance on regional infrastructure or ground routing.
- Included satellite diversity which is meant to provide high access probability.
- Security based on digital solutions.
- Minimal call setup time and low latency.

The satellite constellation of Iridium was designed based on 66 operational satellites, and additional in-orbit spare satellites. The system would offer global real time coverage at any time. The idea is to get Iridium constellation in operational service during mid-2014. The terrestrial gateways would at the same time provide terrestrial interconnection via redundant architecture, whist the network management would be done via Satellite Network Operations Center.

Although commercially the system is not available for public use, technically the Iridium satellite constellation is in working condition, and the system is able to provide a high level of service availability, with voice and data services to government and commercial users in global level. According to the analysis, 66 satellites are predicted to remain operational through 2014.

New partnerships are being discussed with commercial entities and governments to advance with Iridium utilizations, including renewed ground infrastructure and updated technology. All Iridium avionics are based

Table 16.8 *Iridium key figures*

Name	Iridium
System	LEO
Satellites, main	77 (originally), 66 (currently)
Height, approximately from Earth's surface / km	781

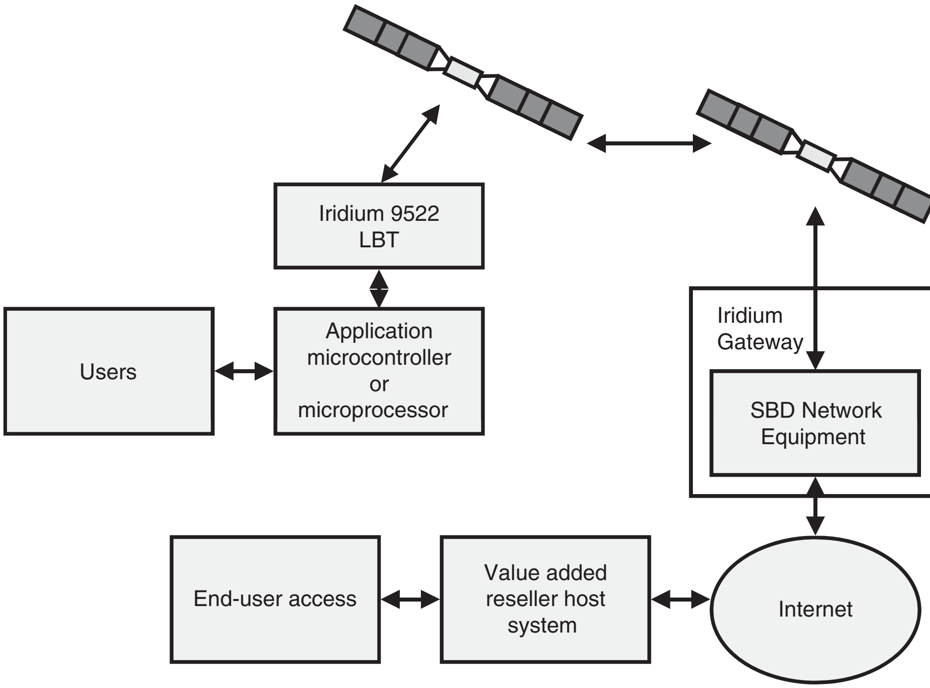

Figure 16.10 SBD architecture of Iridium.

on equipment developed and manufactured by Iridium, that is, L Band Transceiver (LBT) and L Band Transceiver Data Modem. The avionics include single channel voice units as well as multichannel voice and data units. The antenna solution of iridium is omnidirectional. Iridium satellites are low weight.

For the constellation, first launches are expected 2014. The system is designed as backward compatible for existing customers. The new setup includes improved data speeds. The focus of the system is to maintain global coverage, security and low latency.

16.6.7.1 SBD

Commercial example of Iridium services is Iridium Short Burst Data (SBD). This is a new data service that enables value-added applications to send and receive short data transactions efficiently over the Iridium network. SBD is new capability for vertical market applications in the oil, gas, railway, maritime, aeronautical and utility industries as well as applications in the government and military sectors.

The Iridium Satellite network enables SBD in global environment through a single network point of presence. The SBD architecture can be seen in Figure 16.10.

As can be seen in Figure 16.10, the Remote Applications may send mobile originated SBD (MO-SBD) messages by utilizing Iridium L-Band Transceiver (LBT). The maximum size of MO-SBD is 1960 bytes. In order to initiate this procedure, the application microcontroller sends AT commands to LBT. Next, the application transfers the messages into LBT which, in turn, sends them across the Iridium satellite network. The network is based on intersatellite links in order to route the messages to the Iridium Gateway [57].

In the other direction, the mobile terminated SBD (MT-SBD) messages are transferred to the Iridium Gateway via email format originated from host computers. The maximum length of a MT-SBD message is 1890 bytes. MT-SBD messages are thus delivered to the LBT.

As for the delay of the messaging from originating until destination via the global network structure of Iridium, the typical values range between 5 and 20 seconds depending on the length of the message.

16.6.7.2 IRISS

Related to the commercial potential of IRIDIUM system, the IRISS project has identified a set of applications and analyzed their feasibility and requirements to be used in integrated information, communication and navigation gateway [58]. The commercial sectors in question are related to rail transport area which use terrestrial and satellite communications as a part of reporting systems and satellite navigation services.

This capability provides suitable means of communications to Train Operating Companies (TOC) regardless of location. The solution makes the data transfer possible in real time as a part of decision making processes. The solution eases the handling of increased capacity demand, and also has positive effects on the optimization of the fuel consumption and expenses of the operations.

These types of environment involve special points of view which do not have such a high importance in personal or typical professional communications. The rail transport area depends on reliable, real-time, accurate information for assuring primarily the safety of the passengers, and also to maintain the designed performance and punctuality of the transport network.

In addition to the communications directly related to the operations, there is also increasingly high demand for services like instant access to video data (CCTV), telemetrics like location and speed of the cars, as well as enhanced diagnostics of the equipment and services. Also enhanced access methods to special services related, that is to incident management. In the typical case, the rail network is not capable of providing these advanced communications services.

The enhanced services of IRISS for railway environment can be offered via various technologies and solutions. The most complete system consists of various sub-areas, like:

- On-train segment.
- Cellular network segment.
- Station / Depot segment based on Wi-Fi hotspots. This segment allows the uploading of nonpriority and non-real-time data from trains.
- Back-office segment, including Central Server and Application Server.
- Satellite segment.

The IRISS service architecture is based on the Iridium communications, as well as on the general satellite navigation which complements the terrestrial communications systems. The combination of the different technologies and services assures that the most logical means of data transfer is applied with the most reasonable expenses.

Figure 16.11 presents the functional architecture of IRISS.

IRISS development estimates that there are, that is, the following key benefits on the concept:

- Improved performance which helps in understanding better the train, driver and network performance.
- Expense savings due to enhanced operations and logistics.
- Enhanced user-friendly operations interfaces which benefits in lowering the misinterpretations.

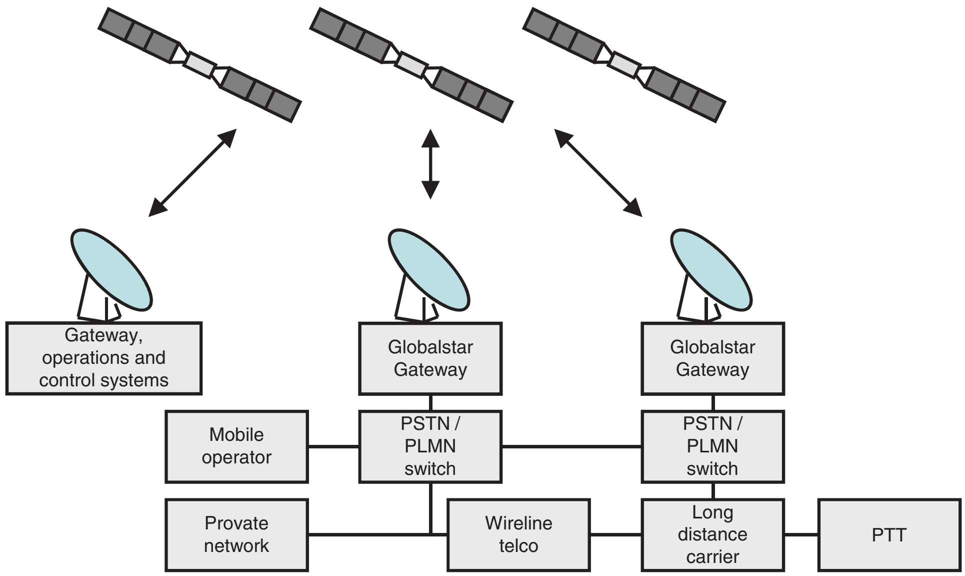

Figure 16.11 *IRISS architecture [59]. Figure by IRISS.*

The concept of the IRISS service includes train Dataset Download Service, Train Tracking Service, On-Train GNSS Feed Service, Staff Paging and Messaging Service, and Train Dataset Analysis Service. Especially for the space segment, the services would be the following:

- GNSS based on any of the commercial satellite location systems like GPS, GLONASS and Galileo. This service provides positioning, navigation and time (PNT) solution. As the GNSS is not tight to a single system, it offers optimal service level.
- Satellite Communications which is based on Iridium. This solution assures there is data link in locations with low or nonexisting service level of terrestrial communications systems. This also enables seamless communications capability over the whole rail network.

The benefits if IRISS are related to:

- Assuring better driving style.
- Delivering more efficiently information especially for supporting decision making.
- Improving the reliability of the service of trains.

16.6.8 Molniya

Molniya is a Russian satellite system designed for domestic market areas. Unlike typical commercial satellite systems, Molniya is based on elliptical orbit with relatively high altitudes.

Molniya refers to a Russian term for lightning, and its root is in military environment where it was utilized for satellite communications. The orbits that the system utilized were known as Molniya orbits. They had inclination of 63.4 degrees with round time period of approximately 12 hours. The benefit of the elliptical orbits is that the satellites are visible in polar regions relative long time periods. This is the difference with geostationary system.

The first Molniya satellite was placed on its orbit in 1965 for military commands, and in later phase, Molniya system has also been used for national television broadcasting. The Molniya constellation included four pairs of satellites. The mass of Molniya satellites oscillate between 1500 and 1800 kg.

16.6.9 Teledesic

Teledesic refers to the proposed US LEO satellite system, Teledesic having been the respective commercial company in charge of the Internet services the satellites were aimed to offer. Originally, it would be based on a total of 840 satellites designed for global telecommunications services in 700 km altitude. The idea has been that with relatively small antennas, the satellite uplink would reach 100 Mb/s data rates whilst downlink would offer up to 720 Mb/s. Nevertheless, die to the high costs, the plan was changed to include 288 active satellites with altitude of 1400 km, and later the figures were still rationalized. Microsoft has been in important role for the funding of the project, and the idea was already concrete level as for the allocation of the Ka band for nongeostationary services. The merger of Teledesic with ICO further helped in the plan which resulted in launching of test satellite. Nevertheless, as has been the case with many other satellite service programs, also Teledesic had to be suspended the activities in 2002.

The satellites were planned to be stabilized over three axis. Remarkable in the plan was the possibility to utilize several launch platforms in a dynamic way. The elevation would have been near polar, 98.2 deg sun synchronous orbit. The initial phase was planned to use 12 orbit planes with 24 satellites placed on each one, resulting footprint per satellite of approximately 700 km.

16.6.10 ICO/Pendrell

ICO is planned to offer global mobile communications services via MEO constellation and international ground telecommunications network [6]. The idea of ICO is to provide services via 10 active satellites. There also are 2 spare satellites located to the constellation with 45 degree inclination orbit. The system has handheld devices as well as special models and fixed phones for the utilization of voice calls, data connections and fax. The other services include voicemail, messaging, three-way calling and call forwarding.

The call initiation happens by signaling with satellite which relays the signal to a satellite access node (SAN), and further via a terrestrial communications network to the destination.

ICO Global Communications initiated the plan for deploying the system in MEO by using two orthogonal planes with 45°inclination. Again, like in many other cases, ICO had economical challenges in advancing with the operations. As a result, ICO Global Communications was renamed to Pendrell Corporation in 2011 [7]. The satellites have in any case been launched into the orbits, the G1 satellite aiming to support a variety of mobile services. The base is in voice calls, video and data services. ICO / Pendrell has coverage in the United States continent, Alaska, Hawaii, Puerto Rico and the US Virgin Islands.

16.6.11 Inmarsat

INMARSAT is the International Maritime Satellite Organization. It is operating a network of satellites meant for international mobile communications mobile services including maritime, aeronautical and land mobile.

Inmarsat is a British satellite telecommunications company. It provides telephone and data services via portable and mobile devices. The system consists of ground stations and 11 geostationary satellites. Typical users of Inmarsat include governments, aid agencies, media, shipping, governments, airlines, media, oil and gas industry, mining and construction and businesses personnel who need mobile communications in

Table 16.9 The Inmarsat satellites

Satellite	Operative	Service area	Longitude
Inmarsat-2 F1	1990	Removed 2013	N/A
Inmarsat-2 F2	1991	POR (Pacific Ocean Region)	143° east
Inmarsat-2 F3	1991	Removed 2006	N/A
Inmarsat-2 F4	1992	Removed 2012	N/A
Inmarsat-3 F1	1996	IOR (Indian Ocean Region)	64.5° east
Inmarsat-3 F2	1996	AOR-E (Atlantic Ocean Region, East)	15.5° west
Inmarsat-3 F3	1996	POR (Pacific Ocean Region)	178° east
Inmarsat-3 F4	1997	AOR-W (Atlantic Ocean Region, West)	54° west
Inmarsat-3 F5	1998	I-3 Europe, Middle-East and Africa	25° east
Inmarsat-4 F1	2005	I-4 Asia-Pacific	143.5° east
Inmarsat-4 F2	2005	I-4 Europe, Middle-East, Africa	25° east
Inmarsat-4 F3	2008	I-4 Americas	98° west

remote regions. Inmarsat launched the first satellite in 1982 for global mobile communications, and continues actively in the provisioning of advances telecommunications services, that is, by offering world's first global 3G network in 2009. The new Global Xpress network delivers download speeds of up to 50 Mb/s.

Each Inmarsat satellite has a beam that covers up to one-third of the Earth's surface. The poles are not covered though. In practice, the Inmarsat coverage is between the latitudes of −82 to +82 degrees everywhere.

Inmarsat has both regional beams and global beams, the regional beams operating with smaller antennas. Regional beams are formed by satellite in such a way that each I-3 satellite may form four to six spot beams and each I-4 satellite produces 19 regional beams.

Furthermore, there are narrow beams provided by three Inmarsat-4 satellites. The narrow beams are clearly smaller than the global or regional beams, and require thus even smaller antennas than regional beams. The narrow beams form the backbone of the system's handheld (GSPS, Global Satellite Phone Service) and broadband services (BGAN, Broadband Global Area Network). Each I-4 satellite provides about 200 narrow spot beams. Table 16.9 summarizes the development of Immarsat.

Inmarsat has developed a series of networks providing certain sets of services (most networks support multiple services). They are grouped into two sets, existing and evolved services, as well as advanced services. Existing and evolved services are offered through land earth stations which are not owned nor operated by Inmarsat, but through companies which have a commercial agreement with Inmarsat. Advanced services are provided via distribution partners but the satellite gateways are owned and operated by Inmarsat directly.

Inmarsat is capable of providing voice calls, data services ad additional services. The voice service categories of Inmarsat are as listed below:

- IsatPhone Pro refers to the mobile satellite phone service of Inmarsat for voice telephony as well as for data transmission up to 20 kb/s, short message service (SMS), short message emailing and location based service (GPS lookup).
- IsatPhone refers to a fixed satellite phone service which includes voice service as well as possibility for data transmission. IsatPhone offers voice calls at 4.8 kb/s as well as data and fax services at 2.4 k/s. It serves Africa, the Middle East, Asia and Europe, as well as maritime in the EMEA and APAC areas.
- FleetPhone is the most basic option of the voice services. It refers to a fixed phone service designed primarily for voice calls.

Selected Inmarsat satellites offer a set of advanced services via Broadband Global Area Network (BGAN) technology. It contains IP-based services according to the following list:

- BGAN for land use (i.e., not maritime nor aerial). BGAN uses the latest I-4 satellites. They all are capable of BGAN, providing IP packet switched data service with maximum data rate of 492 kb/s. They also offer on-demand IP streaming service with a maximum data rate of 450 kb/s. as well as narrowband circuit switched ISDN services (N-ISDN) with 64 kb/s data rate which also can be downgraded.
- FleetBroadband (FB) is a service designed for maritime use. FB uses BGAN technology and infrastructure as a base and is thus capable of providing similar services for ships.
- SwiftBroadband (SB) is a service designed for aeronautical use. As is the case for FB, SB is also based on BGAN technology.
- Machine-to-machine communications (M2M) is a service designed for unmanned environment. The focus of this category is to offer low data-rate IP services in wide service areas, and with high service level. Typical in this category is the need for data transmission with low latency. M2M is suitable for environments without personnel, like in the remote areas for utilizing telemetry.
- IsatM2M refers to short-burst data service which is functioning in storing and forwarding principle. These short-burst messages have lengths of 10.5 or 25.5 bytes, and 100 bytes, in transmitted and received ends, respectively.

The third type of Inmarsat services represents older technologies and belongs to a category of “Existing and Evolved services” as presented below:

- Aeronautical services offer voice, data and fax connectivity for aircraft. This service category includes 3 user equipment types. Aero-L is focused for packet data transfer, Aero-H for medium-quality voice calls as well as for data and fax calls with data rates up to 9.6 kb/s, whereas Aero-I is meant for low-quality voice calls as well as for data and fax calls with maximum data rate of 2.4 kb/s. Aero-L, H and I refer to Low Gain Antenna, High Gain Antenna, and Intermediate Gain Antenna, respectively.
- Inmarsat-B is meant for voice calls, telex, medium-speed data and fax services up to 9.6 kb/s, and high-speed data services of 56–128 kb/s.
- Inmarsat-C is designed for store-and-forward services.
- Inmarsat-M offers voice calls and medium-speed data and fax services at 2.4 kb/s.
- Mini-M offers voice calls and medium-speed data and fax services at 2.4 kbit/s.
- GAN (Global Area Network) offers voice calls, data and fax at 2.4 kb/s, narrowband ISDN-type of services at 64 kbit/s and shared-channel IP PS data services at 64 kb/s. The latter is referred as Mobile Packet Data Service (MPDS). GAN is also called M4.
- MPDS is an IP data service offered in such a way that multiple users may share the capacity of a single 64 kb/s carrier.
- Fleet refers to a set of networks that includes Inmarsat-Fleet33, Inmarsat-Fleet55 and Inmarsat-Fleet77 members. It provides various services including voice calls data and fax at 2.4–128 kb/s depending on the member and terminal capabilities, narrowband ISDN-type of services at 64 kb/s as well as shared-channel IP PS data services at 64 kb/s.
- Swift 64 reminds GAN, but it typically is used as a multichannel variant which is capable of utilizing multiple data rates of 64kb/s.
- Inmarsat D, D+ and IsatM2M refer to pager service within Inmarsat. The most advanced D+ devices provide two-way pager signaling. Instead of actual pager systems as know from 1990s, the main focus of the service is in tracking information of vehicles like trucks, and in SCADA (Supervisory Control and Data Acquisition) applications as a part of industrial control system (ICS). These are automatically

controlled systems that are used for monitoring and controlling industrial processes. Specifically SCADA systems are typically large-scale processes in multiple sites and long communications distances which makes satellite communications a logical option as a base.

More information about Inmarsat can be found in Ref. [8].

16.6.12 Thuraya

Thuraya is a mobile satellite communications company that connects organizations and communities. The coverage of the system is about two thirds of the world's population via satellites and GSM roaming capabilities.

Thuraya is based in the United Arab Emirates. The company claims to operate in more than 140 countries across Europe, the Middle East, North, Central and East Africa, Asia and Australia. It also has partnership with T-Mobile USA to provide roaming services in the United States. The organization provides data and voice satellite communications solutions in remote locations for the energy, media, government, NGO and maritime sectors. In addition to the dual-mode satellite and GSM devices, Thuraya provides IP data modem for a satellite broadband communications with data rates up to 444 kb/s.

The services of Thuraya include:

- SIM card for Thuraya satellite phone.
- Voice communications with satellite phones.
- Short message service.
- 9.6 kb/s data and fax service.
- 60 kb/s downlink and 15 kb/s uplink "GMPRS" mobile data service.
- 144 kb/s high-speed data transfer via a notebook-sized terminal.
- GPS.
- Value-added services like news, call back, call waiting, missed calls, voicemail and WAP.
- One-way "high power alert" capability. It notifies users about incoming call when the signal path to the satellite is attenuated.
- Marine Services: a combination of fixed base station and subscription that provides voice, fax, data and Internet-access services, as well as an emergency service which is capable of sending SMS messages with alarm status and location information to predefined destinations.

Thuraya's country calling code is +882 16 belonging to ITU-T International Networks numbering group. The satellite terminals have maximum output power of 2 watts. The system is based on QPSK modulation. All Thuraya user devices have GPS receiver, and the terminals transmit the location information to the Thuraya gateway periodically.

More information about Thuraya can be found in Ref. [9].

16.6.13 MSAT/SkyTerra

SkyTerra, which was formerly Mobile Satellite Ventures, is a US-based telecommunications system integration company for satellite and terrestrial radio communications. During 2010, the ownership of the company changed and as a result it became part of LightSquared. The first satellite of the company, SkyTerra-1, was placed on its orbit in 2010.

The major part of the services is focused on emergency services, law enforcement and transportation. In addition, MSV and Boeing are working jointly on satellite telephony network.

The Boeing's GeoMobile platform provides coverage to United States with only one satellite. The antenna of MSV satellite is of 22 meters in diameter, making it the largest special antenna so far. The large size provides sufficiently good link budget for the communication with hand held devices with size and transmitter power comparable to typical terrestrial mobile phones. The interworking with terrestrial networks is possible due to the new FCC rules. The devices are compatible with upcoming MSV satellite phone network and regional mobile networks in such a way that the use of satellites happens in areas where cellular coverage is not provided.

The SkyTerra space segment has MSAT geostationary satellites MSAT-1 and MSAT-2, which were launched 2012 and 2012, respectively.

16.6.14 TerreStar

TerreStar Networks was a US-based company that operated both terrestrial and satellite telecommunications systems. The company declared bankruptcy in 2010. As an example, XM Satellite Radio has been a spinoff of this activity.

The first launched satellite was TerreStar-1 in 2009. Remarkably, this satellite was larger than any other commercial telecommunications satellite so far, with mass of 6910 kg. The planned services for TerreStar-1 were mobile voice, messaging and data communications services, the service area being North America. The system was based on S-band.

After the bankruptcy of the TerreStar, Dish Network requested permission to use the spectrum of TerreStar in 2011 for offering terrestrial wireless broadband services.

16.6.15 VSAT

VSAT refers to a very small aperture terminal. It includes small-sized earth stations, typically with antennas of 1.2 to 2.4 meter diameters. Small aperture terminals under 0.5 meters are sometimes referred to Ultra Small Aperture Terminals (USAT).

VSAT is a terminal that is meant for the satellite system and that delivers voice, data video and broadband television contents via an earth station. VSAT refers thus, as a general term, to any satellite terminal that can be utilized either for receiving, or for receiving and transmission.

Via VSAT, it is possible to offer a variety of services to end-users. VSAT provides a straightforward solution for many environments, including the most remote areas of the globe. It is possible, among other functionality of VSAT, to connect to the Internet, and it can be applied in the remote learning, telemetry, maritime communications, communications in wide geometrical areas and communications in the remote missions of, that is, international crisis management. The benefit of VSAT is a clearly wider coverage area compared to, that is, terrestrial cellular systems. The drawback is typically a higher expense for communications compared to terrestrial solutions.

Compared to the benefits, VSAT solutions are flexible and cost-efficient. The setup of a VSAT station is straightforward and fast, and it can be used by individuals as well as for the companies and organizations. The VSAT concept is applied in over 120 countries. Figure 16.12 presents an example in a remote area of Nicaragua.

The VSAT equipment consists of parabolic antenna located outdoors in line of sight with the satellite. The diameter of the antenna is typically in the range of 0.5–3 meters. The setup also contains the user interface which is typically a PC. There is also a centralized transmission and control station (HUB) included in the VSAT system. It includes the respective equipment and a parabolic antenna with a typical diameter of several meters.

Figure 16.12 An example of VSAT station in Nicaragua, Laguna de Perlas, which is isolated from other cities as there are no car routes, only a river leading to the village. The parabolic antenna located outside, and the actual VSAT equipment with respective computer located indoors can be used to connect to Internet with over 100 kb/s data rate. Photo: Jyrki T.J. Penttinen.

16.7 Radio Link Budget

16.7.1 Principle of the Link Budget

The radio link budget of satellite systems is comparable with that of cellular networks, with only few differences. The major difference is that by default, the satellite communications link requires a line of sight (LOS) connection in order to provide sufficient energy level for reception.

Satellite link budget is basically the sum of the losses between the transmitter and receiver, and vice versa, that is, the link budged is done both for uplink (transmitter of user equipment and receiver of satellite) and downlink (transmitter of satellite and receiver of user equipment). The losses are reduced from the sum of the components that contribute gains in the line. The resulting level of received power level needs to be above the critical level for the functional connection, in terms of signal-to-noise (S/N), signal-to-interference (S/I), or in a more complete form, signal-to-noise and interference ratio (S/(N+I)).

The typical parameters that can be taken into account in the complete satellite link budget are listed in the following.

For satellite systems, the RF transmit frequency f_{TX} is typically in Ku band, or sometimes in C band, and nowadays increasingly often also in the higher frequency Ka band for both uplink and downlink due to its available capacity. The benefit of Ka band is the possibility to use a smaller uplink antenna dish, that is, less than 2 m compared to over 2 m of the antenna of Ku band.

The EIRP (Effective Isotropic Radiated Power) is the realistically radiating power level after the gains and losses between the transmitter output and radiating element of the satellite antenna.

Gain refers especially to the gain of the receiver and transmitting antennas. As the physical antenna is a reciprocal element, the same gain is typically valid for satellite's transmitting and receiving antenna of link budget, as well as for the user equipment's transmitting and receiving antenna of link budget, unless there is a considerable difference between the receiving and transmitting frequencies. The basic rule of thumb is that for tuned antenna, the larger the antenna is, also the higher the gain is.

The G/T indicates the quality of the transmitter in terms of noise level. It is calculated by the total gain G of the transmitter (including amplifier, other equipment and antenna) divided by the total noise in the uplink direction. The complete uplink route refers to the set of stages the signal passes prior to radiating from the antenna element.

Latitude and Longitude indicates the location of the user terminal equipment on the Earth's surface. This set of parameters is informed in degrees. Please note that for the satellite, as the satellite is typically located above the Equator at latitude of 0 degrees, solely the longitude is enough for identifying the location of the satellite.

The elevation indicates the angle compared to the horizontal reference angle of 0 degrees in order to have a direct line between the user equipment and satellite. The azimuth, in turn, is the angle between the magnetic north and the direction to the satellite.

Radio signal path loss refers to the level of attenuation between the radiating power level of the transmitter antenna (after antenna element) compared to the received power level at the receiver's antenna (before antenna element). The level of the attenuation can be estimated by utilizing several path loss prediction models, the simplest being a theoretical LOS Equation:

$$L_{FS} = \left(\frac{4\pi d}{\lambda}\right)^2 = \left(\frac{4\pi df}{c}\right)^2, \tag{16.9}$$

where L_{FS} is the path loss in free space (dB), d is the distance between the receiver and transmitter (m), λ is the wavelength of the signal (m), f is the frequency (Hz), and c is the speed of light (m/s).

The free space path loss (FSPL) refers to the loss of signal strength in case of an electromagnetic wave traveling in LOS path in free space with no obstacles that could cause additional attenuation in the path or reflections.

16.7.2 Link Budget Forming

The practical radio wave propagation models take into account also the realistic effects like Earth's atmosphere and possible partial obstacles depending on the terrestrial types.

Loss between transmitter or receiver and antenna elements refers to the cable and connector losses. Typically the cable loss in handheld receiver device's side can be estimated close to zero due to the short lengths, whilst fixed receiver with outdoor antenna could have long antenna cable which contributes several dB losses to the link budget.

The margin takes into account additional attenuation caused by the atmosphere, including the effects of, that is, rain and snow.

In the satellite's side, the translation frequency refers to the frequency that is used to convert the transmitted signal to another frequency in order to have difference between the received and transmitted signals in order to avoid interfrequency interferences. In other words, the translation frequency is the value in MHz for the difference of the receiving and transmitting frequencies.

Transponder gain refers to the transponder's contribution to the gain in satellite's side. Transponder refers to a single channel in a set of channels that form frequency bands of satellite. It is possible to control transponders independently which has effect on the transponder gain.

Required EIRP refers to the power level that the satellite needs in order to be received successfully by user equipment.

"%-EIRP" refers to the proportion from the available satellite power for a single signal, and "Pwr (sat)" is the real transmitted power level of the signal at satellite.

In the user equipment side, the Rx Freq refers to the receiver frequency determined by the satellite translation frequency.

In the modem side, data rate is informed in bits per second. In case of broadcast satellite, it is sufficient to take the data rate into account only in downlink. For two-way communications, the data rate may vary depending on the mode and direction. It should be noted though that for two-way voice communications the data rate can be assumed the same regardless of the direction.

"Required Eb/No" is the minimum energy per bit compared to noise level that still provides functional level for the designed data service with the maximum target bit error rate.

Link margin is sum of all the contributions for the attenuation and gain along with the complete link so that the modem can still handle for locking to the service correctly. It should be noted that in the satellite systems, the loss of lock refers to the attenuation of the received signal below the functional level which means that the connection is basically lost.

G/T refers to the performance of the combination of satellite antenna and low noise amplifier, and is expressed in dB. G refers to the net gain of the system and T refers to the noise temperature of the system. As a general rule of thumb, the higher the number the better the system is.

F/D refers to the ratio of antenna focal length to antenna diameter. A higher value indicates shallower dish.

EOC (Edge of Coverage) means the limit of the service area provided by the satellite. Typically, EOC refers to the limit of 3 dB lower than the signal level at the main direction of the beam. It should be noted that this −3 dB criterion is for network planning purposes, and that the practical functioning of the system may be possible also beyond the point.

EIRP (Effective Isotropic Radiated Power) indicates the signal strength in the radiating antenna elements of the satellite or the transmitting earth station. The transmitting power can be expressed, that is, in dBW and dBm which is a result of the product of the transponder output power and the gain of the satellite transmit antenna.

C/No, C/kT and C/kTB refer to the carrier-to-noise ratio measured either at the Radio Frequency (RF) or Intermediate Frequency (IF). Carrier to Noise Ratio (C/N) is the ratio of the received carrier power and the noise power in certain bandwidth, and it is expressed in dB. This figure is related to G/T and S/N.

16.7.3 Example of the Link Budget

Based on the link budget parameters presented above, the link budget can be constructed. The outcome of the budget is the link margin. It is thus possible to investigate immediately via its value if the link is functional. If the link margin is too small, it is important to know that additional loss is always possible especially due to the changes in atmosphere. Example of such an event is heavy rain, which can significantly contribute to the link thus lowering the signal level below functional level if the margin is not dimensioned correctly.

The task of the dimensioning is thus to balance the gains and losses in such a way that the best estimate for the additional margin should be included. Estimating that rain fall could reduce the power level to half, typical margin could be around 3 dB level.

The idea of the link budget calculation is to investigate the end-to-end transmission and reception as detailed level as reasonably possible in order to understand the maximum path loss that is still allowable for functional connection between the transmitting and receiving ends. The complete chain of the satellite communications thus includes the uplink of the transmission of the earth station to satellite, and downlink of the transmission from the satellite to the earth station. The final received power level depends on the transmitted power level,

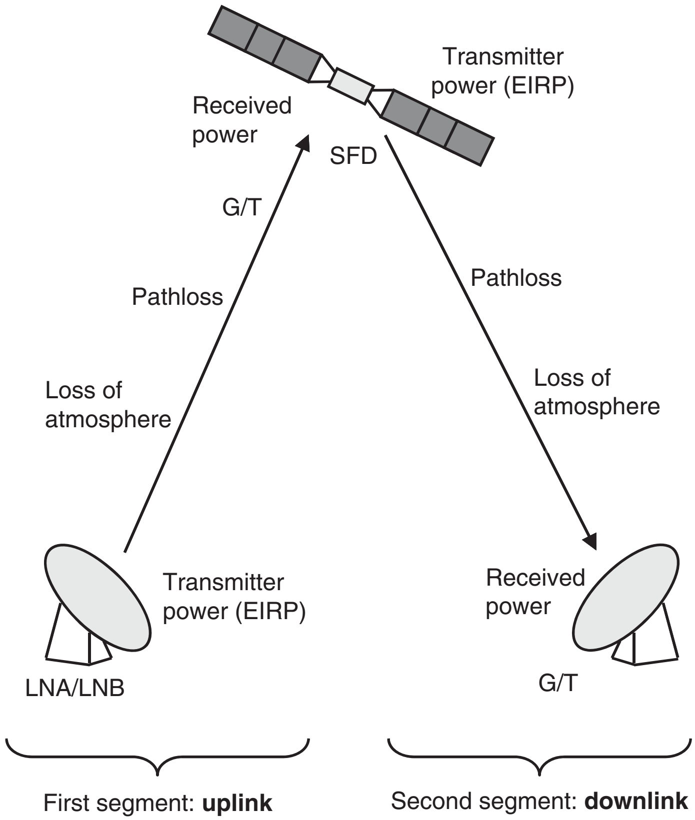

Figure 16.13 *High-level link budget for the satellite communications.*

the sum of gains and losses along the signal path, and the characteristics of the receiver and its antenna. The idea of the same link budget is valid regardless of the earth station types. Figure 16.13 presents the main idea of the link budget.

In addition to the main components of the link budget presented in Figure 16.13, there are elements contributing to the complete gain as well as to the loss. The gain in the uplink includes the output power of the transmitter that is fed into the antenna cable and the antenna gain compared to isotropic radiator. In the receiving and, the receiver antenna gain increases further the power level. The components for the loss include the earth station's antenna feeder and connector losses, the radio frequency path loss, additional loss due to the atmospheric conditions, as well as the feeder and connector losses in the receives side of the satellite. For the downlink, the principle is the same, now satellite being the transmitting end and earth station the receiving side.

The path loss depends on the frequency, with the higher frequencies attenuating more than lower ones. One of the additional limiting factors is also the general interference at the same frequency or nearby spectrum which means that in order to receive the contents with adequate quality level, the sum of the gain components should be respectively higher. Also noise from other radio frequency sources lowers the performance both from other satellites and earth stations. Furthermore, there are unpredictable elements like rain and snow which further contributes to the total attenuation of the uplink and downlink.

The main components for the link budget are thus EIRP, G/T, SFD (saturated flux density) in the satellite side, the antenna gain of the earth station, frequency and interferences.

The base for the antenna gain G (dBi) can be calculated by Equation (16.10):

$$G = \eta \left(\frac{\pi d}{\lambda}\right)^2 = \eta \left(\frac{\pi df}{c}\right)^2. \tag{16.10}$$

In this equation, c is the speed of light (3x10^8 m/s), f is the frequency (Hz), n is the efficiency of the antenna (%), and d is the diameter of the antenna (m). Furthermore, the antenna beam width Θ (degrees) which dictates the useful beam above 3 dB attenuation compared to the maximum level in the main beam direction can be calculated by Equation (16.11):

$$\Theta = 70 \cdot \frac{c}{df}. \tag{16.11}$$

The effectively radiating power level compared to the isotropic antenna P_{EIRP} is the power level of the transmitting antenna which is result of the actual output power of the transmitter P_t and the antenna gain G_t minus the sum of the losses L_{cc} between the power connector of the transmitter and antenna connector, that is, the sum of the antenna cables, jumpers and connectors. The equation is:

$$P_{EIRP} = G_t + P_t - L_{cc}. \tag{16.12}$$

The received power P_r, or signal power level in the receiver end is:

$$P_r = P_{EIRP} - L_p + G_r. \tag{16.13}$$

In this equation, P_{EIRP} is the transmitted power level obtained previously, L_p is the path loss of the link, and G_r is the receiving antenna's gain. Furthermore, the path loss can be estimated theoretically by applying Equation (16.14):

$$L_p = \left(\frac{4\pi D}{\lambda}\right)^2. \tag{16.14}$$

In this equation, D is the slant range (m).

For the noise level estimate, Figure 16.13 shows the relevant components that contribute to the value. The first one is the thermal noise Pn which is a result of the random characteristics of the behavior of electronics equipment. The equation for thermal noise power is:

$$P_n = KTB. \tag{16.15}$$

In this equation, K is constant with a value of −228.6 dBJ/K, T is the temperature (K), and B is the received noise bandwidth (Hz).

At a more detailed level, the effective temperature T_e of the equipment is

$$T_e = T_a + \frac{T_b}{G_a}. \tag{16.16}$$

In this equation, T_a is the temperature of the LNA, T_b is the temperature of the D/C, and G_a is the gain of the LNA. Now, the noise temperature can be calculated by the following equation:

$$T_s = \frac{T_{ant}}{L_c} + \left(1 - \frac{1}{L_c}\right) \cdot T_c. \tag{16.17}$$

In this equation, T_{ant} is the temperature of the antenna element, L_c is the antenna cable loss, and T_c is the antenna cable temperature. Now, the complete contribution of temperatures, effective temperature, is:

$$T_{eff} = T_e + T_s. \tag{16.18}$$

It can be seen that as the LNA is the first element in the transmission path, its contribution is major to the effective temperature value. It also can be noted that lower the LNAs noise figure lower also the temperature is.

Once the temperature is known, it is now possible to estimate the G/T value, that is, the gain to noise temperature of the system (dB/K). It is specific for each system, and can be obtained in this form:

$$G/T = G - 10\log\left(T_{eff}\right). \tag{16.19}$$

Now we are ready to estimate the level of the link in terms of the carrier per noise level. There are thus 2 parts in this analysis, uplink and downlink, with C/N for uplink and downlink, respectively. For the uplink, the value is:

$$\left(\frac{C}{N}\right)_{UL} = \left(P_{EIRP}\right)_{UL} - \left(L_p\right)_{UL} + \left(G/T\right)_{sat} - K - B. \tag{16.20}$$

$$\left(\frac{C}{N}\right)_{DL} = \left(P_{EIRP}\right)_{DL} - \left(L_p\right)_{DL} + \left(G/T\right)_{dev} - K - B. \tag{16.21}$$

Finally, we can calculate the bit energy (Eb) compared to the noise power density (No) that is commonly used for evaluating the performance in terms of bit error rate at the input of the receiver. The respective equation is:

$$\frac{E_b}{N_0} = \left(\frac{C}{N}\right)_T + B_n - R_i. \tag{16.22}$$

In this equation, C refers to carrier signal level, N to noise level, B_n to noise bandwidth (Hz), and R_i to the information rate (b/s). The bit error rate is dimensioned as an average value under typical conditions, but it should be noted that the variations in the climate and other phenomena like attenuation due to the rain alter

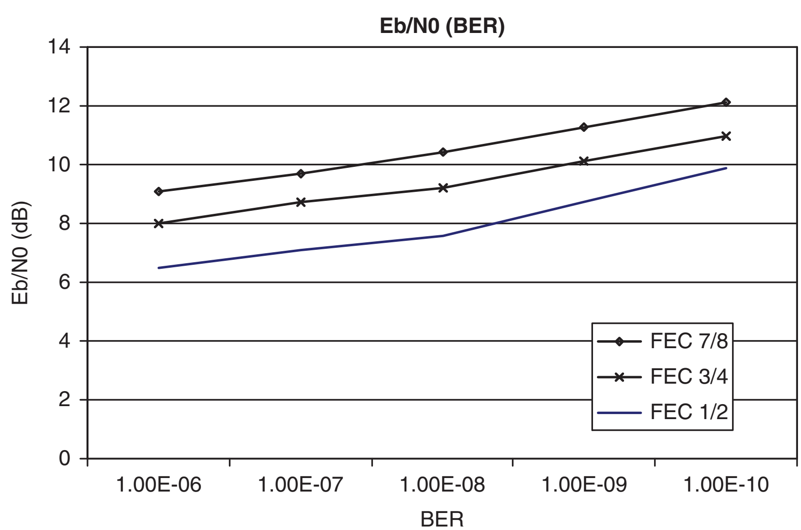

Figure 16.14 *Typical Eb/N0 values as a function of BER, for a set of FEC rates.*

the bit error value. The effect can be seen in practice by varying coverage area of the spot beam as a function of frequency. For the link budget dimensioning purposes, a rough estimate for the rain margin in C-band and Ku-band may be 99.96 and 99.60%, respectively.

The acceptable level of bit error rate, and thus Eb/No depends on the application in question. For the voice call, the value of about 10E-3 can be acceptable whilst for data connections, the raw bit error rate should typically be considerably better. For the latter, the error coding schemes and FEC mechanisms are typically applied in order to lower the effect of varying propagation conditions and interference levels of the radio interface. Figure 16.14 presents typical values of Eb/No in the satellite environment as a function of FEC.

If possible, the link budget should also take into account any uncertainty that may be estimated to affect in the final Eb/No. This category of "other items" may include, but not limited to, estimated margin due to the adjacent satellite interface (ASI) and interference margin due to the other terrestrial and satellite systems.

References

1. Satellite Signals. Explanation of satellite antenna polar mount. http://www.satsig.net/polmount.htm (last accessed 24 August 2014).
2. Initiation of Telstar: http://www.nasa.gov (last accessed 24 August 2014).
3. Space debris: http://www.tethers.com/OrbitalDebris.html (last accessed 2 August 2014).
4. Space debris challenges: http://www.space.com/16518-space-junk.html (last accessed 2 August 2014).
5. Space debris solutions: http://space.epfl.ch/page-39422-en.html (last accessed 2 August 2014.
6. ICO satellite system: http://www.spaceandtech.com/ last accessed 4 July 2013).
7. ICO name changes: http://www.reuters.com/article/2011/06/16/idUS219758+16-Jun-2011+BW20110616 (last accessed 2 August 2014).
8. Inmarsat: http://www.inmarsat.com/ (last accessed 2 August 2014).
9. Thuraya: http://www.thuraya.com/ (last accessed 2 August 2014).
10. Odyssey: http://mars.jpl.nasa.gov/odyssey/ (last accessed 2 August 2014).
11. SkyBridge status: http://www.itu.int/newsarchive/press/WRC97/SkyBridge.html (last accessed 2 August 2014).
12. VSAT solution via Telstar: http://www.satsig.net/africa/q-kon/q-kon.htm (last accessed 2 August 2014).
13. Commercial example of Iridium: http://www.allroadcommunications.com/shortburst/ (last accessed 2 August 2014).
14. Example of Iridium in railway services: http://iap.esa.int/projects/transport/iriss (last accessed 2 August 2014).
15. Jeffrey White. "Iridium Satellite Aviation – We are everywhere you want to fly."
16. ACeS satellite system: http://en.wikipedia.org/wiki/ACeS (last accessed 24 August 2014).
17. ACeS information: http://space.skyrocket.de/doc_sdat/garuda-1.htm (last accessed 2 August 2014).
18. Globalstar: http://globalstar.com (last accessed 2 August 2014).
19. Skyterra: http://www.skyterra.net/ (last accessed 2 August 2014).
20. Satellite radio: http://www.spacefoundation.org/programs/space-technology-hall-fame/inducted-technologies/satellite-radio-technology (last accessed 2 August 2014).
21. Globalstar: http://globalstar.com/en/doubletime2/?gclid=COv0gZ-KnLgCFSZp7Aod1hYANQ (last accessed 2 August 2014).
22. Globalstar: http://www.cs.ucla.edu/classes/winter02/cs216/l1/slides/P24_survey2.pdf (last accessed 2 August 2014).
23. Wireless Link Budget. http://www.rfcafe.com/references/electrical/link-budget.htm (last accessed 24 August 2014).
24. Satellite link budget tutorial: http://www.satcoms.org.uk/satellite/forum/satellite-communication-tutorials_forums_cat3.html?title=satellite-tutorials-vsat-training (last accessed 2 August 2014).
25. Satellite link budget examples: http://complextoreal.com/wp-content/uploads/2013/01/linkbud.pdf (last accessed 2 August 2014).
26. Satellite link budget calculations: http://www.satcoms.org.uk/satellite/forum/topic769.html (last accessed 2 August 2014).
27. Satellite link budget calculator: http://www.satcoms.org.uk/satellite/satellite-link-budget-calculator.asp?title=satellite-link-budget-calculator (last accessed 2 August 2014).

28. Satellite propagation path loss: http://www.radio-electronics.com/info/propagation/path-loss/free-space-formula-equation.php (last accessed 2 August 2014).
29. Satellite link budget tutorial: http://www.rfcafe.com/references/articles/Satellite%20Comm%20Lectures/Satellite-Comms-Link-Budget.pdf (INACCESSIBLE).
30. Horn antenna example. http://www.futurity.org/science-technology/cheaper-lighter-satellite-antennas/ (last accessed 2 August 2014).
31. WRC: http://www.itu.int/ITU-R/index.asp?category=conferences&rlink=wrc&lang=en(last accessed 2 August 2014).
32. FCC: http://www.fcc.gov (last accessed 2 August 2014).
33. IFRB: http://www.itu.int/en/history/Pages/IFRB.aspx (last accessed 2 August 2014).
34. ITU: http://www.itu.int (last accessed 24 August 2014).
35. Intersputnik: http://www.intersputnik.com/ (last accessed 2 August 2014).
36. NTIA: http://www.ntia.doc.gov/about (last accessed 2 August 2014).
37. ITU-T: http://www.itu.int/en/ITU-T/about/Pages/default.aspx (last accessed 2 August 2014).
38. OFCOM: http://www.ofcom.org.uk/ (last accessed 2 August 2014).
39. NAB: http://www.nab.org/ (last accessed 2 August 2014).
40. ITSO: http://www.itso.int/ (last accessed 2 August 2014).
41. INTELSAT: http://www.intelsat.com/about-us/ (last accessed 2 August 2014).
42. NASA: http://www.nasa.gov (last accessed 2 August 2014).
43. JAXA: http://www.jaxa.jp/index_e.html (last accessed 2 August 2014).
44. NCTA: http://www.ncta.com/ (last accessed 2 August 2014).
45. Advanced Communications Technology (ACTS) in Orbit. http://www.nasa.gov/centers/glenn/multimedia/artgallery/art_feature_006_ACTS.html.
46. ISO: www.iso.org (last accessed 2 August 2014).
47. EBU: http://www.ebu.ch (last accessed 2 August 2014).
48. EUTELSAT: http://www.eutelsat.com (last accessed 2 August 2014).
49. TELSTAR history: http://www.corp.att.com/attlabs/reputation/timeline/62trans.html (last accessed 2 August 2014).
50. Globalstar satellite system: http://www.globalstar.com/en/ (last accessed 2 August 2014).
51. Orbcomm satellite system: http://www.orbcomm.com/ (last accessed 2 August 2014).
52. Mars communications: http://mars.jpl.nasa.gov/odyssey/mission/spacecraft/parts/telecommunications/ (last accessed 2 August 2014).
53. SkyBridge satellite system: http://www.itu.int/newsarchive/press/WRC97/SkyBridge.html (last accessed 2 August 2014).
54. Skybridge description. ITU News 8/97.
55. Iridium satellite system: http://www.iridium.com/default.aspx (last accessed 2 August 2014).
56. Jeffrey White. Iridium Satellite – We are everywhere you want to fly.
57. Iridium Short Burst Data: http://www.iridium.com/products/iridium9601.aspx (last accessed 2 August 2014).
58. IRISS: http://artes-apps.esa.int/projects/iriss (last accessed 2 August 2014).
59. IRISS Architecture: http://iap.esa.int/projects/transport/iriss (last accessed 2 August 2014).

17

Satellite Systems: Location Services and Telemetry

Jyrki T. J. Penttinen

17.1 General

Satellite can refer to the manmade equipment placed at some of the useful orbits of the object like Earth or Mars. Satellite can also be any other special piece that is located at the orbit. As an example, Mars is orbited by two very small satellites, Phobos and Deimos. It is assumed that these satellites are based on captured asteroids with dark, carbon-rich rock [1].

In the case of terminology of, for example, spaceflights, a satellite is an object which has been placed into orbit by human endeavor. These satellites are also called artificial satellites. The other variant, natural satellite may refer to, for example, Moon.

The first artificial satellite made by human beings was the Sputnik 1, launched by the Soviet Union in 1957. After that, the number of satellites orbiting the Earth has risen to several thousands. These artificial satellites represent the international cooperation of over 50 countries. Despite the high number of artificial satellites, only a few hundred of them are actually functional and operational. The rest are unused, either complete or fragmented satellites that are considered merely space debris.

Along with the development of satellite techniques, the importance of location based services has increased rapidly since the initial deployment of GPS back in 1990s. Along with the development of mobile communications networks, an increasing set of applications utilizes location based services by combining the location information of satellite and cellular systems. Not only location based services as such, but also high-precision clock information is very useful for many telecommunications solutions. As an example, the synchronization of telecommunication networks can rely on the time stamp of GPS [2, 3].

The most widely utilized commercial satellite navigation systems are American GPS, Russian GLONASS and European GALILEO. Also new programs have been appearing into the market, like Chinese Compass

The Telecommunications Handbook: Engineering Guidelines for Fixed, Mobile and Satellite Systems, First Edition.
Edited by Jyrki T. J. Penttinen.

and Japanese QZSS. All the satellite navigation systems require complete or close to line of sight between sufficient amount of satellites and receiver in order to function, whereas the cellular systems can provide a rough estimate for the location even with blocking obstacles in the radio path, but with reduced accuracy of location.

The satellite systems can also be utilized in many other tasks, including measurements of the weather and other events of the Earth. Also space investigation has an important role, which can be done via satellites for exploring the structure and phenomena of deep space. In general, the delivery of measurement data, that is, telemetry, is a major task of satellites.

17.2 GPS

17.2.1 Background

American solution for the location services is GPS (global positioning system). It is the first system deployed that covers global markets, since 1990s. The basic idea of GPS is to provide location and time for the receivers around globe regardless weather conditions. The 3D-location can be obtained whenever there are at least 4 satellites visible for the receiver, whereas 2D-location can be estimated already with 3 satellites.

Originally meant for the needs of the defense forces of the USA, the system was constructed in such a way that only military receivers were able to obtain the most accurate information whereas additional interferences were created for the publicly available receivers to provide with lower accuracy. This functionality was called selective availability. In the initial phase of GPS service, the accuracy of the commercial devices was thus around $\pm$ 100 meters, and military receivers could obtain some meters of accuracy. There was soon a solution provided for this reduced accuracy via differential stations that received the GPS signals at the same time as the portable devices. As the differential station knows its own location physically, it could then send correction information for the GPS receivers nearby in the field.

Eventually Department of Defense (DoD) of the USA decided to remove the lower accuracy for the commercial devices which has made the differential stations unnecessary. This has also created GPS as a useful tool for nonmilitary environments. One example of this is the utilization of GPS in the aviation because it is independent on the weather conditions. This gives possibility to navigate without visibility physically, and to use autopilot for the flights. It is also a useful tool for emergency field, and as an example, GPS combined with cellular systems own location based services, the emergency calls can be located sufficiently accurately by mobile operators. Obviously, GPS is still a major tool also for the military operations.

Due to the accurate atom clock signal of GPS, the system is used increasingly also in the nonlocation environments. Especially telecommunications networks may utilize GPS for the synchronization of the transmission over large areas.

GPS was developed in order to clearly enhance 0revious navigation systems, which were similar type of ground-based radio-navigation systems like LORAN and the Decca Navigator, and satellite-based Transit and Omega Navigation System. In the military environment, other alternatives were based, for example, on the gyroscope which gradually loses its accuracy after initial calibration.

The American Department of Defense has been the owner of GPS from the initial planning and development. The first fully operational stage of GPS, with 24 satellites and 3 spare satellites, was reached in 1994.

GPS satellites have a certain lifetime, and thus need to be replaced after their useful period. Over time, GPS have also been developed further. This has resulted in the implementation of the next generation of GPS III satellites. Also the rest of the system including the control system is being modernized, leading to the Next Generation Operational Control System (OCX). As a continuum to GPS, DoD is obligated to operate GPS by maintaining a Standard Positioning Service that is available on a continuous, worldwide basis.

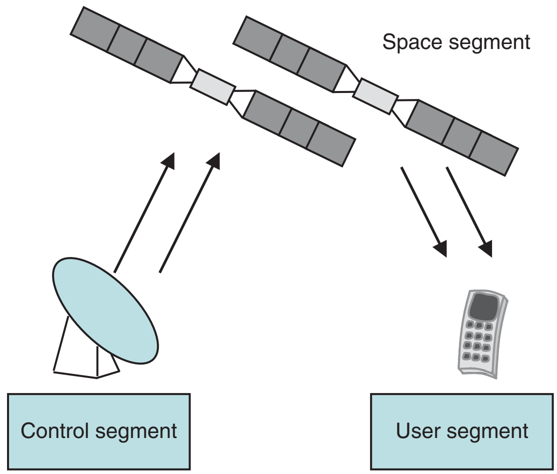

Figure 17.1 *The segments of GPS.*

17.2.2 System Architecture

The Global Positioning System is based on the GPS satellites that constantly send messages from their orbit to Earth. The GPS receivers calculate their position by timing these signals. The messages of the satellites contain the time stamp when the message is transmitted and the satellite position at that specific time. When the GPS receiver gets these signals from different satellites, it can determine the transit time of each radio path, which makes it possible to calculate the distance to each satellite in 2D domain. The combination of these distances, together with the knowledge of the location of each satellite, provides the basis for the calculation of the position of the receiver in 2D domain. The navigation devices typically are able to show the position on graphical map with latitude and longitude, and in the case of a minimum of 4 satellites, also the elevation, that is, the 3D position. Furthermore, typical GPS devices can calculate the speed of the device which is useful in, for example, car road navigation applications.

Due to the limitations of the accuracy of GPS device's own internal clock, 4 satellites must be seen by the device in order to accurately determine the 3D position. It is also possible to combine other technologies to enhance the accuracy in case only 3 satellites are visible.

The architecture of GPS contains three segments: space segment (SS), control segment (CS), and user segment (US). The US Air Force operates the space and control segments as shown in Figure 17.1.

17.2.2.1 Space Segment

The space segment contains 24–32 satellites which are located to Medium Earth Orbit (MEO). The control segment contains master and alternate master control station and host which include dedicated and shared ground antennas and monitor stations. Finally, the user segment contains US and allied military users of the secure GPS Precise Positioning Service as well as commercial, civil, and scientific users of the Standard Positioning Service.

The space segment (SS) includes the orbiting GPS satellites, that is, Space Vehicles (SV), in about 20 200 km of distance from the Earth's surface. The GPS constellation originally has had 24 SVs. First, they were located at three circular-type orbits, and then the constellation has had six orbital planes with four satellites located to each as presented in Figure 17.2.

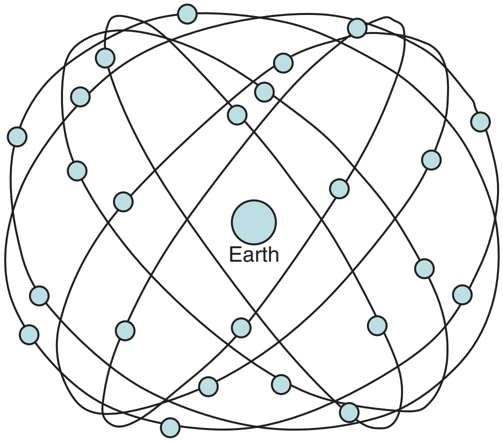

Figure 17.2 *GPS constellation consists of 24 satellites in 6 separate orbit planes, that is, 4 satellites per orbit. The altitude of GPS orbits is 20 200 km from Earth's surface, and the inclination is 55 degrees.*

The orbits are fixed with respect to distant stars, that is, the fixing is not done based on Earth. Nevertheless, these six orbit planes are inclined about 55° relative to Earth's equator. The period for a single orbit is 11 hours and 58 minutes.

There have been additional SVs deployed in the GPS orbits, which has also caused a change from the uniform to nonuniform arrangement. Since 2008, a total of 31 SVs are active, and two spare SVs. These additional SVs provide redundancy for the measurements. Also the visibility of the SVs has been enhanced, resulting in a minimum number of about 9 simultaneous satellites with line-of-sight.

17.2.2.2 Control Segment

The control segment of the GPS system includes the following components: MCS (Master Control Station), Alternate MCS, 4 ground antennas, and 6 monitor stations.

The flight paths of the SVs are tracked by US Air Force monitoring stations and shared NGA monitoring stations in different locations over Earth. Based on this information, each GPS satellite is regularly updated via dedicated or shared ground antennas in order to synchronize the atomic clocks of the SVs, and to adjust the ephemeris of the internal orbital model of the SVs.

The Operation Control Segment (OCS) provides operational capability with the main aim of maintaining the GPS system operational and performing according to the GPS specifications. It will eventually be replaced by Next Generation GPS Operation Control System (OCX) when possible. The latter is able to continue the functionality to manage GPS legacy satellites, with additional functionality to manage the new GPS III (block IIIA) satellites. The GPS OCX program was estimated to support the GPS IIIA 2014 but is delayed.

17.2.2.3 User Segment

The user segment refers to the end-users. It includes the US and allied military users capable of utilizing the secure GPS Precise Positioning Service, as well as all the civil, commercial and scientific users of the Standard Positioning Service of GPS.

The main components of GPS receiver include antenna (either internal or external), receiver and clock that are based typically on a crystal oscillator. In addition, the user interface typically includes a display

Table 17.1 The GPS bands and respective frequencies

Frequency (MHz)	GPS band	Information
157.42	L1	Codes for coarse acquisition C/A, encrypted precision P(Y), civilian L1C and military M.
1227.60	L2	Codes for P(Y), civilian L2C, and military M offered as of Block IIR-M satellites.
1381.05	L3	Dedicated to NUDET (nuclear detonation) activities.
1379.913	L4	Ionosphere correction purposes.
1176.45	L5	Dedicated to SoL (civilian safety of life) activities.

with graphical capabilities. The receiver has a practical limit for handling the signal from various SVs. In the beginning of the service, the limit was only 4 or 5 simultaneous satellites whereas modern versions can handle some 10–20 SVs.

17.2.3 Frequencies

The radio interface of GPS system is based on CDMA spread-spectrum technique. The low bit rate data transmission is encoded with a high-rate Pseudo-Random Noise (PRN) sequence. It is unique for each satellite. GPS receivers know the PRN codes of different satellites which provide the possibility to reconstruct the transmitted data. The related P code rate is 10.23 Mchips/s and is used only for the military environment whereas the respective C/A code rate is 1.023 Mchips/s for the civilian use. GPS satellites operate in two separate frequencies, that is, 1.575 42 GHz, which is referred to L1 band, and 1.2276 GHz, which is referred to L2 band.

It should be noted that the carrier in L1 band is modulated by C/A and P codes, while the L2 carrier is only modulated by the P code. The P code can be encrypted as a P(Y) code. It is available in military equipment when the respective decryption key is known. Both the C/A and P(Y) codes impart the precise time-of-day to the user.

There are also other bands related to the GPS system dedicated for special use. The L3 carrier functions in 1.381 05 GHz, and is meant for the US Nuclear Detonation Detection System (USNDS) to investigate nuclear detonations in the Earth's atmosphere and near space. The L4 carrier at 1.379 913 GHz is a candidate for additional ionospheric correction. The L5 carrier at 1.176 45 GHz is used in the modernized version of GPS in order to provide less interfered frequency.

Table 17.1 summarizes the GPS bands, frequencies and services.

17.2.4 Functionality

GPS is based on the simultaneous broadcasting of signals via two separate frequencies of all the satellites in such a way that the L1 signal is transmitted at 1.575 42 GHz and L2 signal at 1.2276 GHz. The system encodes the low-bit rate CDMA data with a high-rate pseudo-random (PRN) sequence. The latter is unique for each satellite. The receiver knows the PRN code of each satellite which provides the possibility to reconstruct the transmitted data.

The C/A code results in the transmitted data rate of 1.023 x 106 chips per second, and the P code results in 10.23 x 106 chips per second. The C/A code has been deployed for civilian use whilst P code is dedicated to military use. Thanks to the internal satellite reference of 10.229 999 995 43 MHz, the relativistic effects can be compensated. The L1 carrier is modulated by C/A and P codes, whilst only the P code is used for L2

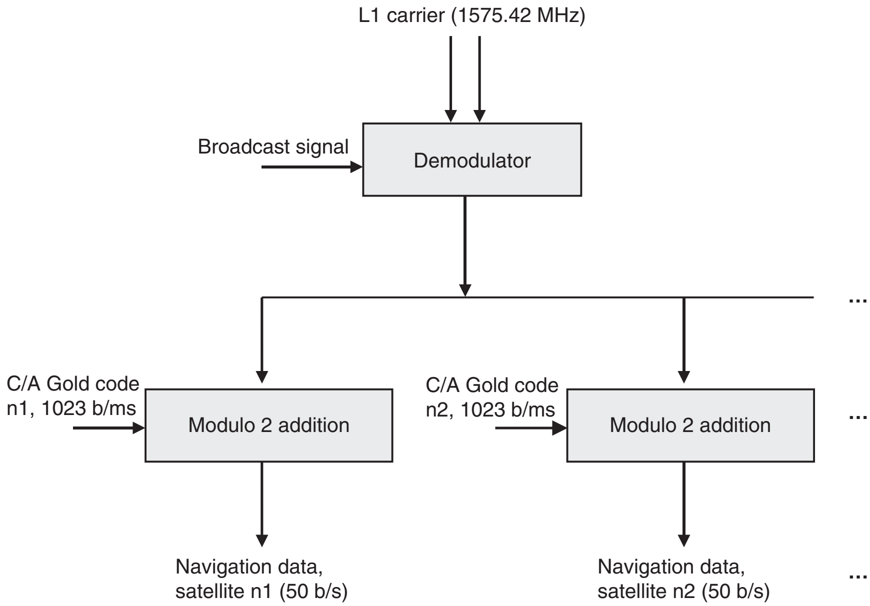

Figure 17.3 *The use of the C/A Gold code in the data processing of the GPS receiver.*

carrier. Furthermore, it is possible to encrypt the P code as a P(Y) code, resulting in its being available solely for military use that has the respective decryption key.

The GPS signals are all modulated to fixed L1 carrier frequency. The separation of the received signals after demodulation is done via Gold code in such a way that it is a unique data stream for each satellite. Thus, the receiver is able to distinguish between the different satellites after the demodulation via the addition of Gold codes.

In the initial phase, if the GPS receiver has not previously acquired satellites or if it has been in off-mode longer time, the equipment lacks almanac information and must enter the search mode. During this time period, the receiver searches for the satellite signals until it can lock on to one of the available satellites. After the lock, the receiver obtains the almanac information along with the information about the satellite constellation, that is, the orbital data or Ephemeris is embedded in the satellite data message. In this way, the receiver gets information about the useful satellites at that time. The individual satellite signals are observed by correlating the C/A code (Figure 17.3), and the receiver is thus able to calculate the position of the satellite in (x, y, z) axis at given time. Once the almanac information is received to the memory of the receiver, the next initiations are faster, as the receiver listens directly to the PRN IDs (Pseudo Random Noise) per each satellite which range from 1 to 32.

Now, by processing the navigation message, the receiver gets the information of the transmission time stamp as well as the satellite position at the respective moment.

17.3 GALILEO

17.3.1 General

The success of GPS has resulted in a highly attractive market for many applications and solutions based on location services. GPS was the first one for the consumer markets and it has reached a leading position.

Even if it was first meant for military use, it provides worldwide service for all the consumers. It is also an important base for, for example, synchronization of many telecommunications systems. The system is also being developed further, which increases usability and accuracy. Nevertheless, the military aspect has stayed an important focus of GPS, and the system is thus operated completely under DoD of the USA.

It can be argued that due to centralization of the control of GPS, there have risen various parallel plans for the satellite navigation systems.

17.3.2 European Variant

Europe is developing GALILEO which is a similar system with GPS, that is, a global navigation system. It has been brainstormed in 1990s, and validated and developed in 2002–2005. The original time schedule has been challenging, the aim having been to get GALILEO in use already around 2008. After successful in-orbit validation in 2011–12, the initial operational capability can be estimated to take place around mid-decade.

The European Union, via the commission and member states, is financing GALILEO both in development phase as well as in the launching and operating of the satellites. GALILEO has an important international role as also many non-EU members have been financing the program. The cost of the project has been over 1 billion euros in the development phase of 2002–2005, and the deployment costs are over 2 billion euros. The yearly operating cost has been estimated to 220 million euros.

Although GALILEO is parallel system with GPS, these systems can be utilized in cooperation. This is possible because the systems are compatible, meaning that in the outage areas of one system, there could be visibility to the satellites of the other system, which increases the accuracy of navigation.

In addition to the European Commission and ESA (European Space Agency), there have been representatives from the private sector for the development and financing of the system. The important nodes of the system are GALILEO Joint Undertaking and the party that utilizes GALILEO.

As an immediate benefit of GALILEO, when deployed, it outperforms the initial version of GPS in its accuracy of navigation. The planned coverage area of GPS is also better.

17.3.2.1 EGNOS

EGNOS (European Geostationary Navigation Overlay Service) is a first step in Europe towards satellite navigation era by its own means. It is a joint effort of the European Commission and ESA, and refers in practice to the joint utilization of the existing satellite navigations systems, which are GPS and GLONASS. As a concrete outcome of this effort, it enhances the accuracy of location information.

EGNOS covers in practice all European countries, and the coverage area can be further enhanced by utilizing two Inmarsat satellites and the own satellite of ESA, Artemis.

EGNOS has been under application testing since 2000, and its functionality is enhanced gradually in such a way that it will be integrated into GALILEO as one part of the system.

17.3.2.2 GALILEO Joint Undertaking

The practical management of the development of GALILEO is GALILEO Joint Undertaking. It is meant to be a temporal setup during the development and deployment phases of GALILEO. The founders of the Joint Undertaking are European Commission and ESA, and its headquarters is located at Brussels, Belgium.

The task of the Joint Undertaking is to observe that GALILEO is advancing as planned, and to coordinate the financing of the project. The most important tasks of it are the coordination of the integration of EGNOS and Galileo systems, the deployment and testing of GALILEO, the launching of the first satellites, facilitation of the research development, coordination of the financing and creation of the business cases.

Table 17.2 The service classes of Galileo

Service class	Description
OS	A set of publicly available signals.
SoL	Technically more advanced as OS, providing, for example, warnings to the users in case there are deviations in the preset criteria. This class includes management of the quality of service.
CS	This class provides two additional signals with enhanced accuracy of the location as well as enhanced data rates. This class also provides small-scale broadcast messages, and includes management of the quality of service.
PRS	This class provides location information and clock for closed user group in such a way that the availability of the service is as good as possible. This class includes the utilization of 2 protected PRS signal.
SAR	This class includes the delivery of alarm messages that can be received from the triggering transmitters. This class is also an extension of international COSPAS-SARSAT finding and rescue service system.

17.3.2.3 Basic Services

The services of Galileo have been divided into open services (OS) and closed services. OS are available for the public without any activation or usage fee. The closed services can be accessed only by limited groups. The closed services include applications that require precise accuracy. These are SoL (safety of life service) and SAR (search and rescue), CS (commercial services) and PRS (public regulated services). Table 17.2 summarizes the services.

The public location services have an accuracy of about 4 meters in horizontal plane and 8 meters in vertical plane when two frequencies are utilized. The respective accuracy is 15–35 meters when via a frequency. As a comparison, the accuracy of GPS is 22 meters in the horizontal plane and 35 meters in the vertical plane. The closed, commercial Galileo location services may offer accuracy of 0.1–1 meters, from which the 0.1 meter accuracy can be achieved locally whereas the 1 meter range is the case for global, closed services.

The accurate clock of GALILEO is meant to benefit various parties, including industry, commercial environment and in general different countries and states. The benefit of GALILEO is similar to the GPS, to provide advanced and accurate means to navigate in all areas of the Earth, in sea, land and air. Some examples of the benefits are optimization of land transportation and safe navigation of airplanes.

An important additional feature of GALILEO is enhancement of personal safety. GALILEO includes functionality that can locate the GALILEO receiver via the return channel. This is beneficial for large-scale crisis and accidents, as well as personal safety of people. In general, the aim of GALILEO is to provide fluent cooperation between the GALILEO and mobile communications systems.

GALILEO thus provides the basic principles of GPS, including provision of service without a separate usage fee.

17.3.2.4 Technology

The global component of GALILEO includes 27 operative and 3 spare satellites (Figure 17.4). The satellites are located to MEO, that is, about 24 000 km height from the surface of Earth. In the deployment phase of GALILEO, 2–8 satellites have been possible to launch simultaneously. Taking into account the cooperation with GPS, the total amount of the satellites of these two systems is thus nearly 60.

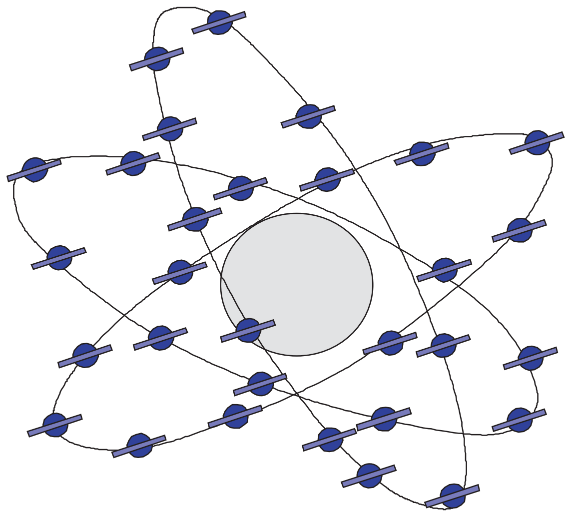

Figure 17.4 The constellation of GALILEO satellites. The system is based on three orbits, each containing a total of 10 satellites (9 operational and one spare).

The GALILEO system also contains earth stations which communicate with and control the satellites. The communication contains messages for evaluating the accuracy of the positioning as well as for fault management.

Via the earth stations, it is possible to achieve the most accurate positioning values by utilizing the local component. In this solution, the earth stations send messages for enhancing accuracy via the existing location based system, or via some other communications system or method.

17.3.2.5 Network Architecture

Galileo consists of satellite and user blocks. In addition, there is a control block in a same manner as in GPS. The height of the satellites is 23 616 km in three separate orbits. Each orbit contains 9 operational satellites and one spare satellite, summing up the total amount of satellites to 30. In case an operational satellite is damaged, the spare satellite can be moved into the respective location. This procedure lasts about 6 hours, providing sufficiently good visibility and accuracy for terminals also during service breakdown.

The tilting of the GALILEO orbit compared to the Equator is 56 degrees, which is optimized for usage in the European continent.

The earth stations of GALILEO operate within Europe. There are two control stations (GCC, ground control center). The tasks of these stations include control, management and correction of the state of satellite constellation. It is also capable of synchronizing the atomic clocks in cooperation between satellites, controlling of tasks and provision of services, and monitoring of satellites in general.

There is also a set of GALILEO sensor stations (GSS) located around globe. They observe the global quality of the signal in space of GALILEO (SIS). This information is forwarded to the earth stations via a separate GALILEO communications network (GCN). This is one of the main differences between GALILEO and other positioning systems.

The earth stations handle the information, for example, the correction of the clock reference. The processed information is then sent back to the satellites via separate frequency bands of GALILEO uplink station (GUS). These stations also receive data, like telemetry and control information, sent by the satellites. For

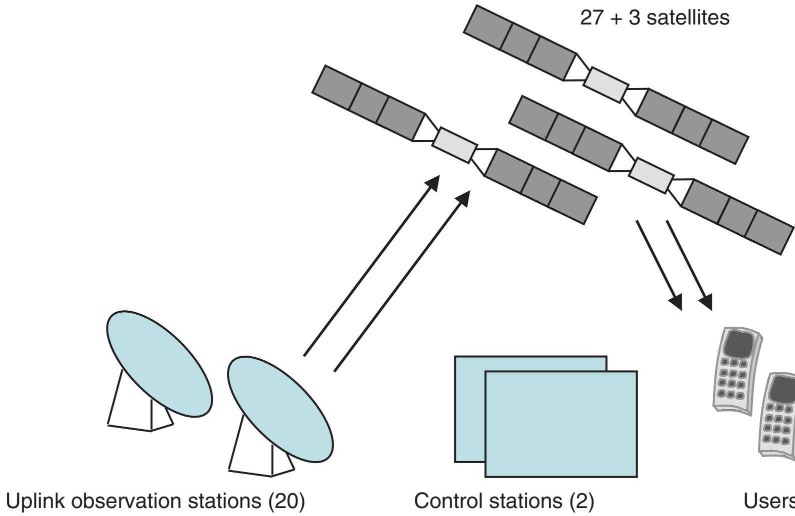

Figure 17.5 The high level architecture of GALILEO contains a satellite block including 30 satellites, and earth observation stations, control stations and user equipment.

this transmission, 5 S bands and 10 C bands are utilized. Figure 17.5 presents the high level architecture of GALILEO.

Although GALILEO can cooperate with GPS and GLONASS, it has been built as independent from the other systems as possible. The difference can be observed from the GALILEO terrestrial reference frame (GTRF), which is similar to the system utilized by GPS coordinate system, WGS-84. The synchronization of GALILEO is based on the international atomic time which provides with the accuracy of 50 nano seconds.

The technical principle of GALILEO is straightforward. Each GALILEO satellite maintains very accurate clock. Each satellite broadcasts its own signal which includes the time stamp of the beginning of the transmission.

The receiver that is located at a certain spot on the Earth, for example, an integrated GALILEO receiver and a mobile communications station, contains accurate information about the orbits of each satellite. When the receiver receives the personalized signal of a certain satellite, the receiver identifies the satellite and calculates the time difference of the time stamps of the received time and transmitted time stamp, and gets the accurate distance between the satellite and receiver. When a minimum of 4 such distance information can be obtained from different satellites, the 3D position of the receiver on Earth can be obtained.

17.3.2.6 Terminals

The marketing aspects direct strongly the adaptation of the usage of the system. This has been taken into account in the basic principles of GALILEO, and thus the set of terminal models should include economically attractive devices. In addition to the economical aspects, the set of GALILEO devices also includes a large variety of special models for different user segments, including devices with joint functionalities with, for example, mobile communications networks.

There will thus be technically many different type of devices, both standalone GALILEO terminals as well as combined devices. The small and integrated devices may include support for the 2G/3G mobile communications, with added value derived from enhanced emergency call functionalities. There will also be special models for aviation and terrestrial vehicles.

17.3.2.7 Positioning Accuracy Depends on the Clock

GALILEO utilizes high-precision atomic clocks. These clocks have been developed in Switzerland, at Observatoire de Neuchatelissa and Temex Neuchatel Time. The daily accuracy of GALILEO clocks is in the range of one per few billions of seconds. Based on this precision, the accuracy of the basic positioning may be in the range of 45 cm at ground level, and 10 cm when a correction to information is applied.

Within each GALILEO satellite, there are 2 separate clocks. One is an atomic clock based on rubidium. The other is based on hydrogen. These clocks differ from each other as for the technology, but both have the same basic principle. When the atoms are forced to move from one energy tier to another, the system radiates in the microwave band in a very precise, constant frequency. The radiating frequency of the rubidium clock is around 6 GHz, and the hydrogen clock is around 1.4 GHz in frequency. As a result, these constant frequencies provide a highly accurate reference clock.

In addition to the forming of the positioning information, these clocks can also be utilized as a frequency for the adjusting of the own clock of the GALILEO devices. As the clocks of the satellites gradually lose accuracy, the earth stations can set the correct time via more accurate cesium-based atomic clocks that would not fit physically on board the satellites. The reference maintained by the earth stations is called Galileo System Time.

17.3.2.8 Frequencies and Modulation

As decided at an international WRC 2003 meeting of ITU-R (World Radio Conference) GALILEO has dedicated frequencies of 1164–1215 MHz (channels E5a/L5, E5, E5b), 1260–1300 MHz (channel E6) and 1559–1593 MHz (channels E2/L1/E1).

The common modulation method with GPS and GALILEO is BOC. It provides a fluent joint functioning of GPS and GALILEO without disturbing each other. It should be noted that the frequency band of GALILEO is slightly wider than GPS, which facilitates the un-interferenced functionality of GALILEO.

BOC takes into account the multipath propagated radio signals in a more efficient way than in previous systems, which enhances positioning accuracy. The modulation of GALILEO depends on the band and service class. The modulation is thus Alt-BOC(15,10) for OS/SOL in E5a and E5b bands, BPSK(5) for CS and BOC(10,5) for PRS in E6 band, and BOC(1,1) for OS/SOL and BOC(15,2.5) for PRS in L1 band.

17.3.2.9 Future

GALILEO provides an alternative for positioning systems, which functions in a global environment but which has been optimized for European use. The basic idea of GALILEO is to provide an independent system that can also combine the techniques of other positioning systems like GPS, GLONASS and methods of mobile communications networks. As a result of this, the availability and quality of service are better than in a single, standalone system.

GALILEO was originally planned to reach its fully operational stage around 2008, but due to the delays in the program, according to the current information, it is planned to be operational possibly around 2015.

17.3.2.10 ESA

ESA (European Space Agency) is a multinational space entity which takes care of long-term, peaceful space politics. ESA consists of 15 member states, from which almost all are also part of the European Union.

ESA has long roots from the development and execution of space programs, including the launching of telecommunications and weather observation satellites. Thus, ESA is also an important member in the development of GALILEO and its early version, EGNOS.

17.4 Positioning Systems: Other Initiatives

In addition to GPS and GALILEO, there is also the Russian, Indian regional navigational satellite system, Japanese satellite system and Chinese navigation systems.

GLONASS is the Russian Global Navigation Satellite System. It is a similar initiative to GPS. In its initial phase, it was used by the Russian military. Eventually, it was opened for public use in 2007 and is thus fully operational worldwide.

BeiDou is a regional system of the People's Republic of China. It is limited to Asia and the West Pacific. Known also as COMPASS, the system is planned to be operational by 2020.

The Japanese variant, QZSS, is a regional system covering Asia and Oceania.

India is also planning to deploy a regional navigation system, IRNSS The plan is to cover India and the Indian Ocean.

17.4.1 GLONASS

GLONASS (Globalnaya Navigatsionnaya Sputnikovaya Sistema), that is, Global Navigation Satellite System, is a Russian satellite navigation system It is operated for the Russian government by the Russian Aerospace Defense Forces. Like GALILEO, also GLONASS is meant for an alternative satellite navigation system for GPS, and it also complements it. Whilst waiting for the fully operational stage of GALILEO, GLONASS is the only parallel system for GPS with a fully global coverage, and with comparable precision.

The Russian satellite navigation system has undergone various stages. The coverage of the latest version of GLONASS was complete within the territory of Russia by 2010. The global coverage was achieved in 2011, with the complete set of 24 satellites in GLONASS orbits.

Like GPS, GLONASS is also meant for both military use as well as for civilians. The GLONASS satellites operate in middle circular orbit at 19 100 km altitude from Earth. The inclination is of 64.8 degrees, which is higher than for GPS and GALILEO. For this reason, GLONASS is especially useful in the high latitudes of both north and south.

GLONASS satellites have a period of 11 hours and 15 minutes. The GLONASS constellation consists of three orbits, each containing 8 satellites. The complete constellation contains a total of 24 satellites, although the Russian territory can be covered already with 18 satellites. As is the case for GPS and GALILEO, GLONASS also requires the reception of at least 4 satellites simultaneously in order to determine the 3D position on Earth.

There are two types of signals in GLONASS: SP (standard precision) and HP (obfuscated high precision) signal. The encoding and modulation is comparable with GPS, being DSSS and BPSK (binary phase-shift keying) modulation. It should be noted that each one of the satellites transmits the same code as their SP signal, but the identification of the satellite is possible due to the slight difference in the transmitting frequency of each one, which is based on 15 frequency division multiple access (FDMA).

As in the case of the primary band of GPS, also the frequency of GLONASS is in L1 band. The center frequency of GLONASS is 1602 MHz + $n \times 0.5625$ MHz. The term n refers to the frequency channel number with a range of $n = -7, -6, \ldots +5, +6$. Please note that before, the range has been $n = 0, \ldots +13$. The radiated isotropic transmitting power (EIRP) of GLONASS is in the range of 25 ... 27 dBW. The transmission is directional with the angle of a cone of 38°. The polarization is right-hand circular polarization. Please note that the number of the frequencies is not the same as the total number of satellites (24), because the frequencies can be reused on the opposite side of the planet. There are thus satellite pairs in the GLONASS constellation using the same frequency.

In the other GLONASS band, L2, the HP signal is broadcasted together with the SP signal in phase quadrature. The signals thus share the same carrier, although the SP signal utilizes bandwidth of ten times higher. The center frequency of L2 transmission is in 1246 MHz + n×0.4375 MHz. The range of the n in this equation is the same as for L1, that is, $n = -7, -6, \ldots +5, +6$.

The vertical accuracy of the first generation of GLONASS is 5–10 meters by the utilization of SP signal together with GPS. The evolved versions enhance the accuracy.

The coordinate system of GLONASS is based on PZ-90.

17.4.2 BeiDou/COMPASS

The Chinese satellite navigation system is called the BeiDou Navigation System, that is, BeiDou, or Compass Navigation Satellite System. Similarly as GALILEO is Europe's intention to build up an independent satellite navigation system, this is China's response to satellite navigation technologies.

The Compass system is being developed gradually. The first version is called the BeiDou Satellite Navigation Experimental System, that is, BeiDou-1. It contains only three satellites and is thus limited in 0overage and applications. Nevertheless, the first generation Compass has been commercially available in and close to China since 2000.

The evolved version of Compass, BeiDou-2, will be representing the second generation of Compass. It will be substantially more complete, consisting of 35 satellites. It is being deployed gradually in such a way that it could serve the Asia-Pacific region. The global coverage is planned to be a reality by 2020.

17.4.3 QZSS

The Japanese response to the satellite navigation systems is the Quasi-Zenith Satellite System (QZSS). The planned coverage area would cover Japan, meaning that it will be a highly local system. According to estimations, QZSS will have 4-satellite constellation in 2018 so that 3 satellites are visible always at the Asia-Oceania regions. QZSS can be used in an integrated with GPS which increases the number of useful satellites to at least eight.

In addition to the actual navigation service, the aim of the QZSS system is to support mobile applications and to provide communications-based services like audio, video and data. The accuracy of the location based service is expected to be limited via QZSS, and for the more accurate service, a joint operation with other systems is needed, for example, with the geostationary satellites of Multifunctional Transport Satellite (MTSAT) of Japan. Nevertheless this latter one is still under development.

The planned constellation of QZSS is based on a periodic Highly Elliptical Orbit (HEO). The period of QZSS would thus be over 12 hours per day, and the elevation would be over 70 degrees. This is the reason for the term quasi-Zenith.

17.4.4 IRNSS

India is also developing a satellite navigation system under the name of Indian Regional Navigational Satellite System (IRNSS), organized by the Indian Space Research Organisation. According to the plan, the IRNSS will provide a Standard Positioning Service which is open for normal civilian use. In addition, there would be an encrypted Restricted Service for military utilization.

17.5 Space Research

The first artificial satellite, Sputnik I of the former USSR on October 4, 1957 can be noted as a milestone for space research from outside the Earth's surface. Ever since, space research has been expanding as new satellites have been launched, the next one having been Explorer I of the USA soon after Sputnik I, on January 31, 1958. The era of space probes had thus begun, and by the end of the 1960s, the USA and USSR had already launched about 50 space probes for investigating the Moon [4]. The early probes were relatively simple as they were planned to pass close to the Moon or even crash into the Moon, before the technology developed and soft landings were possible. These early investigations of the Moon paved the way for preparing the first manned Moon mission. In the later phase, there has been systematic mapping of the lunar surface by the USA, and China has also been active in space probes.

17.6 Weather and Meteorological Satellites

Weather satellites serve for observing Earth's climate. Weather satellites may be polar orbiting or geostationary. The former covers the entire Earth asynchronously and the latter is serving more localized areas spot. The focus of weather satellites is to observe clod systems, but they can be also used for observing, for example, the development of ash originated from active volcanoes and the status of the ozone layer, as well as the cycles of the severe weather conditions caused by the El Niño weather system.

There are also meteorological satellites capable of observing deeper phenomena than weather satellites, including forest fires, status of air pollution, snow and ice, ocean currents, and development of vegetation, among other visual aspects. A practical example of meteorological satellites is the European ENVISAT.

The information collected via weather satellites is typically based on international cooperation in order to provide a continuous service, the main participating countries being the US, Europe, India, China, Russia, and Japan.

Weather and meteorological satellites are typically based on visible and infrared light. A typical range of the visible spectrum is 0.6 μm – 1.6 μm which is useful for cloud cover observation in daytime, as well as storms, snow, ice, fire, smoke, smog and dust among other visible objects. It should also be noted that the status of wind is possible to observe via the development of cloud movement.

Infrared variant, on the other hand, is typically in range of 3.9 μm–7.3 μm for vapor, and 8.7 μm–13.4 μm for thermal imaging. The thermal imaging is used for determining the height and types of clouds, the temperature of water and land, and in general to observe oceanic characteristics. Infrared satellites are useful for investigation of development of severe weather conditions via eye pattern, one technique being Dvorak. It is utilized to estimate tropical cyclone intensity from satellite pictures, and it is based on a pattern recognition decision tree. More detailed, by utilizing satellite picture of a tropical cyclone, the matching of the image can be done with a number of possible pattern types like curved band pattern, shear pattern, eye pattern, central dense overcast pattern, embedded center pattern and central cold cover pattern. As an example, infrared satellite imagery can be used for investigating eye patterns of severe weather system development via the difference between the temperatures of warm eye and the surrounding cold cloud parts. In this case, it can be reasoned that the larger the difference, the more severe the tropical cyclone will be [5].

17.6.1 Geostationary Satellites

Geostationary weather satellites are orbiting at the level of the Equator with altitude of 35 880 km, or 22 300 miles. Thanks to the type of orbit, the satellites will not move compared to the physical location of Earth's surface. The benefit of this solution is that these satellites are capable of constantly informing the status of the complete hemisphere below, by utilizing both visible-light and infrared sensors. The practical use case for these satellites is the delivery of daily weather forecast pictures for local television news and newspapers.

Various geostationary meteorological satellites operate at the international level. As an example, the United States has GOES-12, GOES-13 and GOES-15 satellites. Russia utilizes today modernized weather satellite Elektro-L 1. Japan has MTSAT-1R, Europe uses Meteosat-6, 7, 8 and 9, and India has INSAT.

17.6.2 Polar Orbiting Satellites

Polar orbiting weather satellites are located typically at an altitude of 850 km or 530 miles. Their path is designed to go via North and South Poles. Polar satellites are located to sun-synchronous orbits, that is, the satellites are able to monitor any location on Earth by viewing the spots twice per day with the same light conditions, as a result of the near-constant local solar time. As the polar orbiting satellites are considerably closer to Earth, they are able to provide higher quality imaging compared to geostationary satellites.

The United States has polar orbiting meteorological satellites NOAA 15–18, with spare satellites in standby mode, whilst polar orbiting satellites elsewhere are, for example, Metop-A (Europe), Meteor and RESURS (Russia), and FY-1D and FY-3A (China).

17.7 Military Systems

Military satellites form a part of the biggest nation's arsenal for gathering intelligence. They are also used for secured and closed communications, and in some cases, also as a military weapon. It should be noted that a civilian satellites can also carry military transponders, and military satellites can contain civilian equipment.

As an example of the US military satellites, Milstar (Military Strategic and Tactical Relay) is a set of communications satellites in geostationary orbit. They provide closed and interference immune global communications system for the US Armed Forces. There is currently a set of five operational satellites.

Milstar has the following segments:

- Space segment which includes five satellites
- Ground terminals and users
- Control and command stations.

The Milstar satellites have a designed useful lifetime of 10 years and are intended to be replaced by AEHF satellites (Advanced Extremely High Frequency program).

There are Block I and Block II satellites which provide communications of 75–1200 b/s. In addition, Block II can manage data rates of 4.8 kb/s–1.544 Mb/s. The uplink of the satellites is in the Q band and downlink in the Ka band.

Milstar satellites function as a switchboard for directing traffic between Earth terminals. In addition to communications between satellites and Earth terminals, the satellites are also capable of communicating between other Milstar satellites via cross links, thus minimizing the need for ground-controlled switching. The modes of communication include encrypted voice, data, teletype and facsimile communications.

The new US military satellite system, AEHF, is equally a set of communications satellites operated by the United States Air Force Air Force Space Command, for the joint use of the United States, the Great Britain, Canada and the Netherlands. The system is designed to include six geostationary satellites in such a way that they gradually replace the previous Milstar satellites.

AEHF operates at 44 GHz Uplink (EHF band) and 20 GHz Downlink (SHF band). As previously, AEHF has links between satellites and Earth terminals, as well as between satellites via cross links. The system provides

Table 17.3 Satellite orbits

Orbit	Definition and observations
Areocentric	This is an orbit around Mars. The objects in this orbit include moons and artificial satellites.
Geocentric	This is a basic orbit around Earth, that is, the center of the orbit is represented by Earth. An example of an object on geocentric orbit is the Moon. Currently, the geocentric orbit is heavily occupied by artificial satellites, the number being over 2400.
Heliocentric	As the name suggests, this orbit is formed around the Sun. The objects around the Sun include the planets, comets and asteroids. This orbit contains also artificial satellites and every kind of space debris.

resistance against jamming and eavesdropping, and has, for example, frequency-hopping. Furthermore, the satellites have phased array antennas for dynamic electrical beam forming to further minimize the jamming and to minimize the transmission out from the useful link.

AEHF will operate with the same Milstar data-rate classes of 75–2400 b/s and 4.8 kb/s–1.544 Mb/s. As a novelty, AEHF also has higher data rates of up to 8.192 Mb/s. As is the case of Milstar, AEHF will also be operated by the 4th Space Operations Squadron.

17.7.1 Orbits

Table 17.3 summarizes the definitions for the overall satellite orbits.

Table 17.4 summarizes the Geocentric orbit types as a function of the altitude.

There is also a Geosynchronous transfer orbit which refers to an elliptic orbit with the perigee located at the altitude of LEO, and the apogee at the altitude of a geosynchronous orbit. Another variant is Geostationary transfer orbit which refers to an elliptic orbit with the perigee located at the altitude of LEO, and the apogee at the altitude of a geostationary orbit.

Another form of classification of the orbits is based on the synchronization. A synchronous orbit refers to an orbit with the satellite having an orbital period which is equal to the average rotational period of the piece that the satellite orbits, and that has the same direction of rotation as that piece. In the case of the Earth, the period is 23 hours and 56 minutes with 4.091 seconds.

Semisynchronous orbit (SSO) is an orbit with an altitude of about 20 200 km, that is, 12 600 miles. In this case, the orbital period is equal to a half of the average rotational period. In the case of the Earth, this makes the period about 12 hours long.

Geosynchronous orbit (GSO) refers to an orbit with an altitude of about 35 786 km, that is, 22 236 miles.

Geostationary orbit (GEO), that is, Clarke orbit, refers to a geosynchronous orbit with zero-inclination. This means that the satellite looks like a fixed point for an observer on the Earth.

Table 17.4 The orbit types of the Earth

Orbit	Definition and observations
Low Earth Orbit (LEO)	The altitude of the orbit is in the range of 0–2000 km, that is, 0–1240 miles.
Medium Earth orbit (MEO)	The altitude of the orbit is in the range of 2000 km, that is, 1200 miles, up to almost to geosynchronous orbit at 35 786 km, that is, 22 236 miles. This orbit is also referred to as intermediate circular orbit.
High Earth orbit (HEO)	The altitude of the orbit is exactly at the geosynchronous orbit with 35 786 km, that is, 22 236 miles, from the Earth.

References

1. Smithsonian National Air and Space Museum. Space satellites: http://airandspace.si.edu/etp/mars/MARSdata.html (last accessed 3 August 2014).
2. Los Angeles Air Force Base. Factsheets: GPS Advanced Control Segment (OCX). Losangeles.af.mil. 2011-10-25. http://www.losangeles.af.mil/ (last accessed 3 August 2014).
3. Penttinen, J. (2006) *Tietoliikennetekniikka (Telecommunications Technology)*, WSOY, Helsinki,
4. Infoplease. Satellites and Probes of Infoplease: http://www.infoplease.com/encyclopedia/science/space-exploration-satellites-probes.html (last accessed 3 August 2014).
5. NOAA. Dvorak technique: http://www.aoml.noaa.gov/hrd/tcfaq/H1.html (last accessed 3 August 2014). Finland.

18

Other and Special Networks

Pertti Virtanen and Jyrki T. J. Penttinen

18.1 IS-95

This chapter describes the overall architecture, core and radio systems, and interfaces of North American cellular system IS-95 which belongs to the 2G mobile communications. The evolved 3G variant CDMA2000 is discussed in chapter 18.2, and chapter 18.3 summarizes TETRA system.

18.1.1 General

Interim Standard 95 (IS-95) is a CDMA-based digital cellular system which has been developed by Qualcomm. The system is also known as cdmaOne®, or TIA-EIA-95. IS-95 is one of the second generation cellular systems, that is, it represents the same development phase in the cellular communications field as GSM [1,2].

IS-95 is spread widely in North America, and it is also supported in selected other areas outside the USA. The first specification set for the system was called IS-95A, and an enhanced specification IS-95B was released later. In fact, the later specification also has the commercial name cdmaOne®. The system is meant for voice and data services. The data transfer rate for IS-95A is up to 14.4 kbps, and for IS-95B up to 115 kbps.

18.1.2 Standards

IS-95 has been developed by TIA (Telecommunications Industry Association), and more specifically, in T1P1 committee within ATIS (Alliance for Telecommunications Industry Solutions). IS-95, that is, Interim Standard no. 95, defines the radio interface, whilst IS-41 forms the basis for the core network. IS-95 is prominently North American digital, second generation cellular system with CDMA as a basis for the radio interface which was developed by Qualcomm. Since the initial introduction of the system as a TIA-published standard back in 1993, the system has evolved under standard names IS-95A and IS-95B.

The Telecommunications Handbook: Engineering Guidelines for Fixed, Mobile and Satellite Systems, First Edition.
Edited by Jyrki T. J. Penttinen.

The idea of direct sequence spread spectrum (DSSS) has already been known and utilized by the military environment for a long time prior to its adaptation to the commercial mobile communications systems since 1960s. It has provided means to hide signal efficiently into the radio spectrum below the thermal noise level which makes the detection and jamming of the signal much more challenging than was possible for the previous systems. Qualcomm adopted the idea of DSSS into the multiple access of cellular environment in the 1980s.

As a result of the efforts of Qualcomm, and later other parties like AT&T and Motorola, a standard for the mobile system, IS-95A, was published in 1995 under the Cellular Telecommunications Industry Association (CTIA) and the Telecommunications Industry Association (TIA). The first commercial network was launched three years later by Hutchison Telecom. This development work also resulted in the foundation of CDMA Development Group (CDG) which includes representatives from network operators and device manufacturers. The aim of the group is to promote CDMA as well as to further work on the CDMA technology and standards, along with the currently most active 3GPP2.

The basic CDMA system, that is, IS-95A and IS-95B, later evolved towards the 3G system. The resulting 3G system is called commercially as cdma2000. This system includes several variants, that is, cdma2000 1x, cdma2000 1x EV-DO (Evolution Data Only/Data Optimized). There has also been a theoretical variant called cdma2000 1x EV-DV (Evolution Data and Voice).

18.1.3 CDMA Principles

CDMA (Code Division Multiple Access) is in fact both radio interface definition as well as access method for cellular systems. CDMA is based on the separation of users by different codes dedicated for each user within the same area. To complement the whole network system, the core part is basically based on the TDMA principles in a similar manner as in, for example, GSM system. There are thus considerable similarities in the practical solutions of the upper layers to the air interface, including RRM (Radio Resource Management) and MM (Mobility Management).

18.1.3.1 Functioning

The basic idea of DSSS is to multiply the data stream with another one that is using a considerably higher data rate. This spreading code makes the frequency bandwidth widened for the actual data transmission. The original data stream can be detected when the very same spreading code is applied in the reconstruction of the data. As the spreading codes and data stream of a certain user is seen as background noise for the outsiders, it is thus possible to utilize the same, wide frequency brand for several users in such a way that they are not disturbed by the others. This phenomenon can be compared with a cocktail party where participants form different language areas (comparable to scrambling codes). The unknown languages are interpreted as background noise for the others, and even if the noise level rises, the ones that understand common language can interpret the respective message. The correlation of the codes for scrambling and unscrambling provides with the possibility to apply multiple access in the system.

Via the spread spectrum techniques like DSSS, CDMA thus utilizes the same, relatively wide frequency band for sharing the data transmission for all the users currently located in the area. The essential parts for making the CDMA system work are the transmit power control and error correction codes. There are also other functionalities which increase the performance of the system, like the RAKE receiver for combining multipath propagated signal components of the same signal, variable data rate for data transmission and voice service, and soft handoff mechanism. The complete set of different functionalities provides with possibility to lower the requirements of carrier per interference level that would be required without these additional functionalities, which in turn enhances the spectral efficiency.

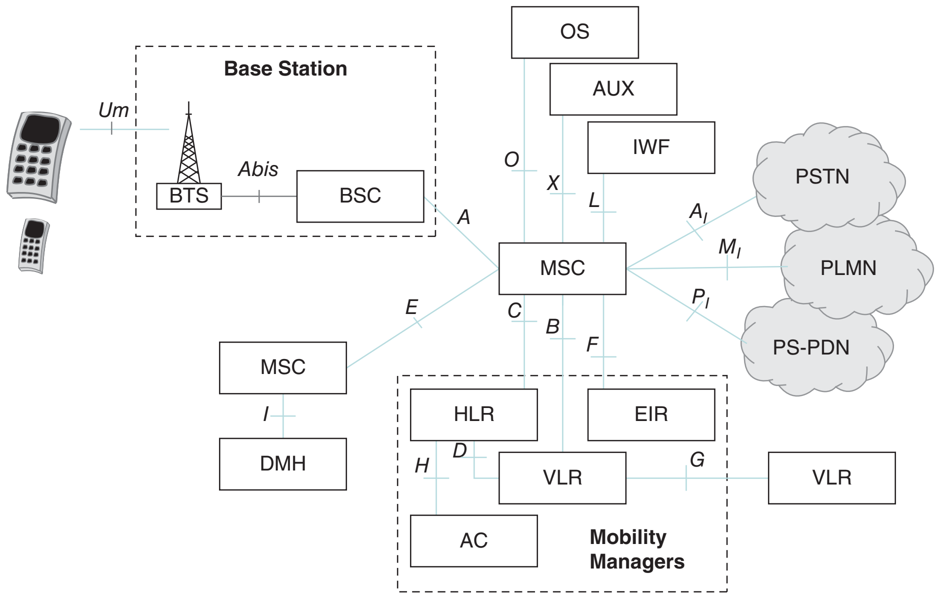

Figure 18.1 *The network architecture for IS-95 defined by TR45/46.*

18.1.3.2 Radio Interface

The CDMA forward link refers to the radio transmission from the Base Station (BS) to the terminal. The respective term in GSM is downlink. In IS-95, the forward link refers to a bandwidth of a single carrier which is 1.25 MHz wide.

IS-95 utilizes QPSK for the transmission of spreading codes in case of the forward channels.

Each IS-95 carrier contains logical channels. As everything is transmitted and received within the same frequency band of the carrier, these channels are separated from each others by variety of spreading codes.

When the signal is processed, it goes through several steps including coding itself, interleaving, and so on, which results in the symbols being transmitted over the radio interface. In order to do this, the symbols are modulated with Walsh Codes.

18.1.4 Network Architecture

The core network, including radio and fixed parts if North American cellular systems, is defined by TIA Committee TR-45 as a reference architecture. Among other aspects, TR-45 contains service standards as well as definitions for the interoperability and performance. These standards define the requirements for terminal equipment, radio base stations, and so on. The division is such that TR-45.3 includes TDMA technology whilst TR-45.5 defines the CDMA systems.

Figure 18.1 presents the IS-95 system architecture based on the reference model of 45/46.

As can be observed from Figure 18.1, the network elements and the principle of the core architecture have many similarities with the GSM system. Nevertheless, they are different. The main elements of the IS-95 core network are:

- Base Station is the element taking care of the physical connection of Mobile Station and the rest of the network. It is thus acting in radio interface via Um, and on the other side, with BSC via Abis interface.

BSC together with BSC form the radio network part of the system architecture. BTS collects information from the radio interface, that is, the quality of the connections with user equipment.

- Base Station Controller is the "brain" of the radio network, and thus makes the decisions, for example, for soft and hard handovers based on the information BTS is providing. BSC maintains a list of the active and idle terminals that are currently in the area.
- Mobile Services Switching-Center (MSC).
- Home Location Center (HLC).
- Visitor Location Register (VLR).
- Authentication Center (AC).
- The Data Message Handler (DMH) collects information about the connections for the charging purposes.
- Terminals.

18.2 CDMA2000

18.2.1 General

CDMA2000, which is also known as IMT Multi Carrier (IMT-MC), belongs to the third generation cellular systems together, for example, with UMTS. CDMA2000 standards consist of CDMA2000 1X, CDMA2000 EV-DO Rev. 0, CDMA2000 EV-DO Rev. A, and CDMA2000 EV-DO Rev. B. [3] These variants comply with the ITU IMT-2000 requirements, and furthermore, CDMA2000 is backward-compatible with the North American 2G system, IS-95. In the United States, CDMA2000 is a registered trademark of the Telecommunications Industry Association (TIA-USA). [4]

CDMA2000 1X (IS-2000) is also known as 1x and 1xRTT (one times Radio Transmission Technology). It forms the nucleus of CDMA2000 radio interface. The meaning of the term 1x is that the system uses the same radio frequency bandwidth as IS-95, which is the original CDMA value of 1.25 MHz for both forward and reverse radio channels.

CDMA 1X adds 64 traffic channels to the forward link compared to the original IS-95. These channels are orthogonal to the previous 64 channels of IS-95.

CDMA 1X system provides supports packet data speeds theoretically up to 153 kbit/s. The practical data rates are typically in the range of 60–100 kbit/s, for example, for [5].

IMT-2000 also made changes to the data link layer for greater use of data services, including medium and link access control protocols and QoS. The IS-95 data link layer only provided "best efforts delivery" for data and circuit switched channel for voice (i.e., a voice frame once every 20 ms).

CDMA2000 1xEV-DO (Evolution-Data Optimized), often abbreviated as EV-DO or EV, is a telecommunications standard for the wireless transmission of data through radio signals, typically for broadband Internet access. It uses multiplexing techniques including code division multiple access (CDMA) as well as time division multiple access (TDMA) to maximize both individual user's throughput and the overall system throughput. It is standardized by 3rd Generation Partnership Project 2 (3GPP2) as part of the CDMA2000 family of standards and has been adopted by many mobile phone service providers around the world – particularly those previously employing CDMA networks. It is also used on the Globalstar satellite phone network. Globalstar GSP 1700 satphone in one of the examples supporting EVDO [6, 7].

The CDMA Development Group states that, as of November 2009, there were 308 operators in 116 countries offering CDMA2000 1X and 1xEV-DO service [8–10].

CDMA2000 is defined by 3GPP2, which is the North American counterpart for the European 3GPP. The general trend is that CDMA would not have such a wide international support as GSM has achieved. Furthermore, the customer base is moving towards newer technologies like LTE and LTE-Advanced. Thus,

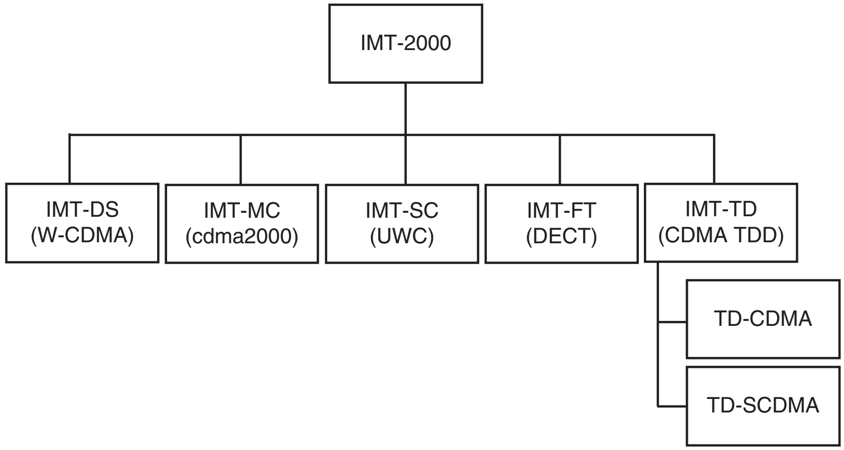

Figure 18.2 *IMT-2000 is divided into several 3G solutions.*

one of the logical network evolution paths would be to include GSM support in IS-95 devices for the roaming purposes, and for supporting also national USA GSM frequencies.

In general, the business environment indicates that the 2G systems – as they are already relatively old and thus less spectral efficient than the next cellular system generations – will be fading out gradually, and the released spectrum would be utilized for 3G and especially for 4G technologies.

There are various evolution paths towards the wider support of next generations. The transition would not be sudden, as there are many IS-95 users without terminal's support of other technologies. Nevertheless, as the penetration of the multimode devices increase, meaning that the IS-95 can be utilized in a parallel way with 3G/LTE/4G by prioritizing the newer technologies, the frequency re-farming would be one of the most logical solutions in the transition phase. The first CDMA2000 networks were deployed in South Korea during 2000.

The radio interface of CDMA2000 system differs from, for example, UMTS. Nevertheless, the core network has synergies between different systems. As an example, TIA packet core network was specified by TIA TR45.6., and the work was based strongly on IETF (Internet Engineering Task Force) solutions.

CDMA2000 belongs to ITU (International Telecommunication Union) IMT-2000 (international mobile telecommunications) as one of the 3G standards as seen from Figure 18.2. CDMA2000 is based on CDMA IS-95 system and is in practice its evolution path. The other entity specifying CDMA2000 is 3GPP2 which assures that the core solution is compatible with 3GPP systems.

18.3 TETRA

TETRA (terrestrial trunked radio) is a radio system meant for professional, global use. TETRA is an open standard developed by the European Telecommunications Standards Institute (ETSI). As a basis for the standardization, there has been TETRA MoU (Memorandum of Understanding) established in 1994, and currently MoU includes multitude of representing organizations from equipment manufacturers, operators, service providers, as well as other interested parties.

The main purpose of the TETRA standard was to define a series of open interfaces, as well as services and facilities, in sufficient detail to enable independent manufacturers to develop infrastructure and terminal products that would fully interoperate with each other as well as meet the needs of traditional PMR (Private Mobile Radio) user organizations.

The technology solutions chosen to meet user requirements contained in the TETRA standards have been, and continue to be, developed primarily by well known and respected manufacturers who have been serving the PMR market with products and services for several decades. This combined knowledge ensures that optimum technology solutions are chosen to meet user requirements.

Although the prime responsibility of ETSI is to develop standards for Europe, many of its standards are also adopted worldwide, as evidenced by the uptake of GSM, the first wireless technology standard to be developed by ETSI. Similarly, TETRA has already been deployed in many regions and nations outside Europe, resulting in TETRA becoming a truly global standard.

There is no doubt that a proprietary technology solution can be brought to market in less time than a solution conforming to a recognized open standard. However, large user organizations, especially those in the public sector, have recognized that some proprietary solutions can meet their needs but the “tie in” to a single supplier can have significant disadvantages. Even though there are some disadvantages, the main advantages and benefits of adopting an open standard are:

- Economies of scale provided by a large harmonized market served by several independent manufacturers and suppliers competing for the same business resulting in competitively priced solutions
- Second source security if existing suppliers exit the market
- Evolution (instead of revolution) of the technology standard ensuring longevity and good return on investment for both users and suppliers
- Choice of manufacturers for new products keeping prices down
- Greater choice of products for specialized applications
- Greater responsiveness to future needs by existing suppliers because of competition.

Because there are several independent manufacturers of both TETRA network infrastructure and radio terminals all the benefits of standardization listed also apply to the TETRA market. This planned evolution of TETRA can be appreciated when considering that traditional PMR user organizations will always require private PMR networks because public networks cannot adequately provide the required RF coverage, Grade of Service (GoS) during busy periods and high levels of reliability. Besides these basic needs, public networks will not be able to provide the specialized voice services such as wide area fast call-set up all informed nets (group calls), Direct Mode Operation (DMO) and high levels of secure encryption for voice and data. Figure 18.3 shows an example of TETRA communications systems in the military environment.

In summary, TETRA will evolve in a similar way to GSM, which evolved from providing a basic V+D “one to one” telephony service (via GSM II+, GPRS, EDGE, etc.) to UMTS/3G supporting powerful multimedia applications and High Speed Data. Also, the focus and technology solution for Next Generation Networks (NGN) will primarily be for public networks.

Taking these previous factors into consideration and the fact that analog MPT 1327 trunking networks are still being deployed across the world more than 28 years after the technology was first developed, TETRA networks are expected to be available for at least another 25 years, thereby ensuring a very good return on investment for user organizations as well as manufacturers and suppliers.

18.3.1 TETRA I

The original TETRA standard first envisaged in ETSI was known as the TETRA Voice plus Data (V+D) standard. Because of the need to further evolve and enhance TETRA, the original V+D standard is now

Figure 18.3 TETRA is a closed radio system for government entities, military service, police, and so on. Its main benefits are the good security level as well as possibility to share the resources amongst the connected entities.

known as TETRA Release I. An overview of the network elements covered in the TETRA standard is shown in Figure 18.4.

The elements and interfaces of Figure 18.4 are the following:

- Um, the air interface (AI) between the network and radio terminal (MS, Mobile Station).
- Ud, the radio interface between two terminals communicating directly with each others (DMO, Direct Mode Operation).
- PEI, Peripheral Equipment Interface, for example, for connecting computer.
- SwMI, Switching and Management Infrastructure, for example, exchanges, base stations, and so on.
- PDN, Packet Data Network.
- PSTN, Public Switched Telecommunications Network (including PABX).
- ISDN, Integrated Services Digital Network.
- IPI, IP Interworking, that is, IP interface between TETRA networks via public telephony network. Please note that this option is not deployed in practice.
- ISI, Inter System Interface, that is, the connectivity between TETRA networks without gateway network.
- Gi, packet data interface towards public packet data networks.

The abbreviation SwMI is used to classify all of the equipment and subsystems that comprise a TETRA network, including base stations. Even though some ETSI Technical Committee (TC) TETRA members felt

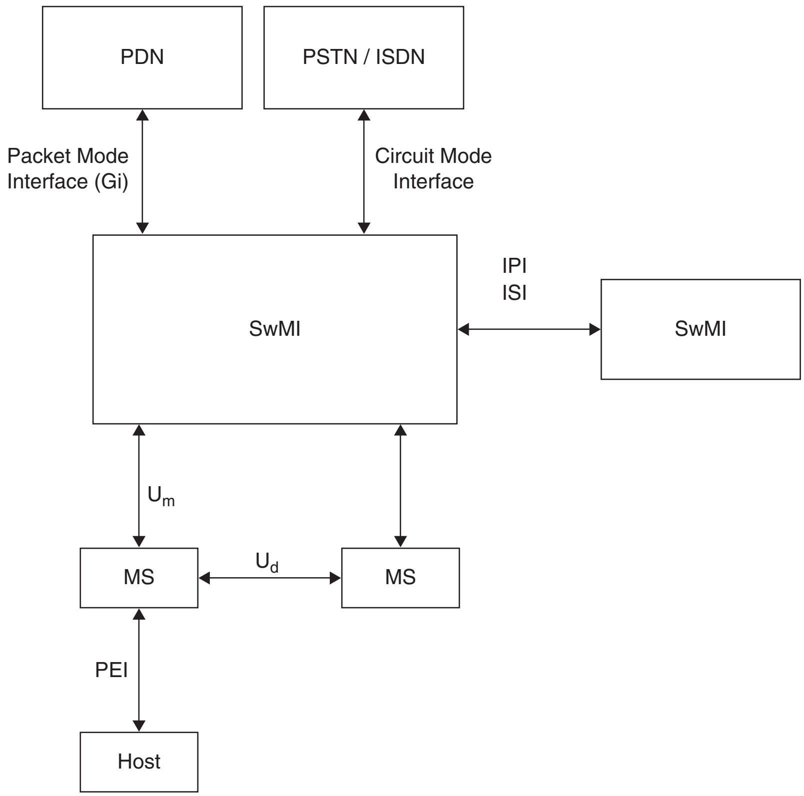

Figure 18.4 *The interfaces of TETRA Release 1.*

that a standard base station interface would be useful (as provided in GSM) it was decided that owing to the way in which different manufacturers configure their networks for optimum performance and design flexibility, it would be impractical to implement.

It was also agreed, for the same reasons as the base station interface, that everything contained inside the SwMI would not be standardized, thereby allowing TETRA infrastructure manufacturers' flexibility in design, and the ability to differentiate their portfolio offerings, when in competition with other TETRA manufacturers. This practical approach also meant that new technologies in the areas of transmission and networking could be used without having to go through a long standardization process.

18.3.1.1 Air Interfaces (1 and 2)

The most important (and complex) interfaces are considered to be the "air interfaces" between the base station and radio terminals (1) and the Direct Mode Operation (DMO) interface (2). DMO is a facility that allows terminals to operate in local radio nets independent of the main TETRA network infrastructure.

18.3.1.2 Peripheral Equipment Interface (4)

This interface standardizes the connection of the radio terminal to an external device, and supports data transmission between applications resident in the device and the connected TETRA radio terminal. The

PEI also supports certain elements of control within the radio terminal from the external device and/or application.

18.3.1.3 Remote Dispatcher Interface (5)

This interface was originally intended to allow connection to remote wire line dispatcher consoles like those located in major control rooms. Nevertheless, work on this interface was dropped in ETSI TC TETRA as the complexity to provide a universal interface without degrading performance was impractical. This was because the PMR industry had specialist manufacturers of control room equipment, the majority of which differed in the way they interfaced to PMR networks. Similarly, the TETRA network architecture of manufacturers also differed adding to the complexity of providing a universal interface. For these reasons only TETRA manufacturer specific interface specifications are available to support the many voice and data applications requiring access to TETRA infrastructures.

18.3.1.4 PSTN / ISDN / PABX (6)

This standardized interface enables TETRA to interface with the PSTN, the ISDN and/or a PABX.

18.3.1.5 Intersystem Interface (7)

This standardized Intersystem Interface (ISI) allows infrastructures supplied by different TETRA manufacturers to interoperate with each other allowing interoperability between two or more networks. There are two methods of interconnection in the standard, one covering information transfer using circuit mode and the other using packet mode.

18.3.1.6 Network Management Interface (8)

Like the local dispatcher interface, it was recognized during standardization activities that a common network management interface was impractical. Fortunately, this early standardization was not wasted as it was later turned into a comprehensive guide to assist users in defining network management requirements.

18.3.1.7 Others

Besides these network element standards, the many services and facilities available on TETRA are also standardized, the most significant of these being:

- Advanced and fast group call services – clear and encrypted
- Individual calls – clear and encrypted
- Short Data Services – clear and encrypted
- Packet Data Services – clear and encrypted.

18.3.2 TETRA II

DTETRA Release I (Voice + Data) already provides a very comprehensive portfolio of services and facilities but as time progresses there is a need to evolve and enhance all technologies to better satisfy user requirements, future proof investments and ensure longevity. Like GSM moving to GPRS, EDGE and UMTS/3G, TETRA also needs to evolve to satisfy increasing user demand for new services and facilities as well as gleaning the benefits of new technology.

As early as 1999, interest groups comprising both users and manufacturers within Technical Committee (TC) TETRA and the TETRA Association identified the need to enhance TETRA in several areas. Although the initial number of areas identified was very comprehensive, significant events in the telecommunications industry, combined with changing market needs, resulted in the following services and facilities being standardized at the end of 2005 as part of TETRA Release II:

- Trunked Mode Operation (TMO) Range Extension
- Adaptive Multiple Rate (AMR) Voice Codec
- Mixed Excitation Liner Predictive, enhanced (MELPe) Voice Codec
- TETRA Enhanced Data Service (TEDS).

18.3.2.1 Trunked Mode Operation (TMO) Range Extension

The ability for TETRA to operate beyond the 58 km range limit (a function of TETRAs TDMA structure) was required by certain user organizations to allow efficient Air-Ground-Air (AGA) communications whilst operating on the main TMO network. By modifying uplink and downlink bursts, as well as guard times, the TMO range of TETRA is extended up to 83 km for AGA applications. (Note: DMO has no TDMA structure range limitation as synchronization takes place in DMO at the start of each transmission.)

18.3.2.2 Adaptive Multiple Rate (AMR) Voice Codec

The AMR codec, operating in the 4.75 kbits/s only mode, has been chosen for possible future applications in TETRA. However, completion of the Air Interface Standard to accommodate the AMR codec is suspended in TC TETRA until sufficient market need is identified.

18.3.2.3 Mixed Excitation Liner Predictive, Enhanced (MELPe) Voice Codec

The STANAG 4591 (MELPe codec), to use its correct NATO reference, has been standardized by NATO for its own military communication applications because of its low bit rate (2400 bit/s), immunity to high background noise and acceptable voice quality performance. Because of TETRAs suitability for certain military communication applications TC TETRA carried out a technical feasibility study to see if could be supported on TETRA. The results of this study indicated potential benefits such as

- Interworking with government systems (no tandem operation)
- Suppression of background noise
- Improved RF Coverage using spare bits available for extra FEC
- Simultaneous V+D using spare bits available for data.

However, the way the MELPe codec needs to be implemented in TETRA increases “end to end” voice delay, which needs to be balanced against its possible benefits. Completion of the TETRA standard to accommodate the MELPe codec will be dependent on the outcome of cost/benefit comparisons with the existing TETRA codec, which will be carried out in TC TETRA.

Table 18.1 Packet data throughput of TETRA in DL. TEDS RF channel bandwidths and data rates. All presented rates are a result of 4 timeslots

	Channel type / kHz			
Modulation	25	50	100	150
Pi/4 DQPSK	15.6			
Pi/8 D8PSK	24.3			
4-QAM	11	27	58	90
16-QAM	22	54	116	179
64-QAM	33	80	175	269
64-QAM	44	107	233	359
64-QAM	66	160	349	538

18.3.2.4 TETRA Enhanced Data Service (TEDS)

TEDS is a new TETRA High Speed Data (HSD) service using different RF channel bandwidths and data rates for flexible use of PMR frequency bands. TEDS is fully compatibility with TETRA Release 1 and allows for ease of migration. It has been optimized for efficient use of PMR frequency bands and designed for all TETRA market segment applications. The RF channel bandwidths supported in TEDS are:

- 25 kHz
- 50 kHz
- 100 kHz
- 150 kHz.

The modulation schemes supported in TEDS are:

- pi/4 DQPSK (for common TETRA V+D and TEDS control channel)
- pi/8 D8PSK (for early migration requiring modest increase in speed)
- 4 QAM (for efficient links at edge of coverage)
- 16 QAM (for moderate speeds)
- 64 QAM (for high speeds).

Table 18.1 presents a matrix with different RF channel bandwidths and data rates supported in TEDS.

With adaptive selection of modulation schemes, RF channel bandwidths and coding according to propagation conditions, user bit rates in the region of 10 to 500 kbits/s can be expected. For ease of evolution and migration from TETRA Release 1 reuse of the TETRA protocol stack and TDMA structure have been maximized. TEDS also allows up to 8 multimedia applications and QoS negotiation for real-time class data applications, such as voice and video and telemetry, with the QoS attributes negotiated being; throughput, delay, priority and reliability. Support for sectored cells is also provided enabling the use of existing TETRA Release 1 Base Sites for TEDS without the need for additional sites. Even though TEDS is capable of providing High Speed Data in 150 kHz RF channels, the current limitation caused by insufficient RF spectrum to support the growth of TETRA will probably limit early deployments to 50 kHz RF channel assignments only.

Now that the TETRA Release II standards are sufficiently complete for product development purposes, actual product availability will be dependent on the different manufacturer's development plans.

18.3.3 Security

The area of TETRA security is extensive as it needs to provide different levels of security ranging from what is acceptable on commercial networks to what is acceptable on a national public safety network. The security mechanisms in the standard are covered through Authentication, Air Interface Encryption (AIE) and End to End encryption. The threats to Confidentiality, Authenticity, Integrity, Availability as well as Accountability are covered with those three mechanisms.

The standard based services are constantly being expanded by a subgroup of the Association – Security and Fraud Prevention Group (SFPG).

Mutual Authentication is a service required to ensure that a TETRA system can control access to it and for a radio terminal to check if a network can be trusted. In TETRA, as in most other secure systems, authentication is the basis for much of overall network security and can also be used to ensure validated billing in public access systems, and can provide the foundation for a secure distribution channel for sensitive information such as other encryption keys. The mutual authentication security mechanisms protect both Voice and Data services.

The TETRA standard supports four AIE TETRA Encryption Algorithms (TEAs), these being TEA1, TEA2, TEA3 and TEA 4. There are differences in the intended use and the exportability of equipment containing these algorithms. For example, TEA2 is intended for use by public safety users in Schengen and related European countries only; the others have wider applications ranging from general commercial use to public safety use in regions where TEA2 is not used. The main benefit of over the air encryption is that it protects all signaling and identities as well as user speech and data. This provides an excellent level of protection from traffic analysis as well as from eavesdropping. The encryption system is closely bound to the TETRA signaling protocols and the algorithms can (if desired) be implemented as software within radio terminals and base station equipment, instead of using encryption modules, which could consume space and increase cost.

The TETRA standard also supports End to End encryption using a variety of encryption algorithms as deemed necessary by national security organizations. The TETRA Association Security and Fraud Prevention Group has extended the work carried out in the TETRA standard to define a general framework for the incorporation of End to End encryption. Recommended sample solutions have also been provided for the International Data Encryption Algorithm (IDEA) algorithm (IPR owned by Ascom) and the newer Advanced Encryption Standard (AES) algorithm (IPR-free), which benefits from a larger cryptographic algorithm block size. Custom and indigenous algorithms are also possible with End to End encryption, although these are not recommended for air interface encryption due to their need for integration in signaling protocols and availability of standard compliant terminals.

Besides these core security capabilities, TETRA can also support a wide range of security management capabilities such as those used to control, manage and operate the individual security mechanisms in a network. The most important of these is Encryption Key management, which is fully integrated in TETRA standard functions. Even though security functions are integrated in a network this does not automatically imply that a network is fully secure. However, what is normally achieved is that the security risks are "condensed," that is they are concentrated to specific elements in the network, which can be adequately controlled.

18.3.4 Benefits

The core technologies used in the TETRA standard, such as Digital, Trunking and Time Division Multiple Access (TDMA) also provide a number of inherent advantages and benefits as follows:

18.3.4.1 Digital System

Nowadays, practically everything electronic uses digital technology and wireless communications are no exception. Even though analog FM PMR communications will remain a viable option for

Table 18.2 Conventional PMR problems solved by trunking

Conventional PMR problem	Trunking solution
Contention	All call requests are handled on the control channel for immediate call processing or in order of queue priority if the system is busy.
Manual Switching of Channels	Automatic cell handover takes away the need for manual channel selection.
Inefficient Channel Utilization	The automatic and dynamic assignment of a small number of communication channels shared amongst a relatively large number of users ensures an equal grade of service for all radio users on the system.
Lack of Privacy	The dynamic and random allocation of channels makes it more difficult for a casual eavesdropper to monitor conversations.
Radio User Abuse	Abuse is minimized as the identity of all radio users and the time and duration of messages are known and can therefore be easily traced to the abuser.

several years, digital radio provides relative advantages and disadvantages in the important performance areas of:

- Voice Quality
- RF Coverage
- Nonvoice Services
- Security
- Cost.

18.3.4.2 Trunking

Trunking techniques have been used for many years in switched telephone networks. The first trunked mobile radio communication systems were deployed as early as the 1970s in North America with proprietary signaling protocols and shortly afterwards in Europe using analog MPT1327 technology. The main benefit of trunking is normally seen as spectrum efficiency, or more radio users per RF channel compared with a conventional radio channel for a given Grade of Service (GoS), brought about by the automatic and dynamic assignment of a small number of communication channels shared amongst a relatively large number of users.

Because trunking systems support more radio users than conventional systems, national administrations actively support the deployment of trunking systems as this helps reduce pressure on meeting PMR spectrum demands. However, from a radio user's operational point of view, spectrum efficiency does not really mean anything. What users want is to solve all the operational problems associated with conventional PMR, yet still retain the simplicity of conventional open channel "all informed net" operation. The fundamental element of trunking that solves these conventional PMR problems is the use of a control channel. Table 18.2 lists the operational problems of conventional PMR and also lists how the use of trunking solves these problems.

It is important to note that the operational simplicity of conventional PMR "all informed net" talk group communications is still retained by employing fast call setup "Push To Talk" (PTT) operation on radio terminals.

18.3.4.3 Additional Services and Facilities

As the control channel acts as a signaling communications link between the Trunking Controller and all mobile radio terminals operating on the system, the Trunking Controller knows the status of the system at any moment in time as well as its historic usage, which is stored in its memory. For example, the Trunking Controller knows:

- The individual and group identity of all radio units registered on the system
- The individual identity and time radio units registered on the system
- The individual identity and time radio units deregistered from the system
- The individual and group identity, time and duration of all messages.

With additional intelligence in both the radio terminals and the trunking controller the advantages and benefits of trunking can be increased. For example, the length of the control channel signaling messages can be increased by a set amount to accommodate a variety of new services and facilities. Also, the trunking controller can be programmed to handle calls in a variety of ways as required by the operator of the system.

18.3.4.4 Time Division Multiple Access (TDMA)

A four time slot TDMA technology was adopted in TETRA as it offered the optimum solution to balance the cost of equipment with that of supporting the services and facilities required by user organizations for a medium to high capacity network providing single site local RF coverage and/or multiple site wide area RF coverage.

RF Spectrum efficiency is a combination of three main factors being the occupied bandwidth per communication channel, the frequency reuse factor determined by the Carrier to Interference protection ratio C/I in dB's and the trunking technology used. As previously mentioned TETRA utilizes the latest in trunking technology. Also, the TDMA technology used in TETRA provides 4 independent communications channels in a 25 kHz RF bandwidth Channel, making it twice as efficient in occupied bandwidth terms as a traditional 12.5 kHz RF bandwidth FDMA channel. Although FDMA technologies tend to have a better C/I performance than TDMA TETRA, the overall spectrum efficiency advantage lies with TETRA, especially for medium to high capacity networks.

Because of using TDMA technology, the cost and equipment space at base station sites can be significantly reduced compared with traditional FDMA technology trunking solutions. Another advantage of TDMA technology is that it enables new services and facilities to be supported with minimum cost. Some examples are: Short Data Service and packet data transfer.

18.3.4.5 Higher Data Rates

The "laws of physics" limit the maximum data rate in a given RF channel bandwidth. For the same modulation scheme applies principle: the wider the channel bandwidth the higher the data rate. Because TDMA uses wider channels than FDMA, the combined data rate on a single RF carrier is greater.

18.3.4.6 Improved Data Throughput in Poor RF Signal Conditions

The net data rate in TDMA is better than FDMA in poor RF propagation conditions. This is because Automatic Repeat Requests (ARQs) are required when received data is corrupted as a result of RF fading. As TDMA terminal devices effectively operate in full duplex ARQs can be sent efficiently after each time slot transmission instead of waiting until the end of each voice transmission, as is usually the case with FDMA.

18.3.4.7 Bandwidth on Demand

In TDMA any number of time slots up to the maximum limit of the technology being employed can be combined to increase data throughput as required for specific applications.

18.3.4.8 Concurrent Voice and Data

Because of the TDMA time slot structure it is possible to assign one time slot to support voice and the next time slot to support data in a two slot transmission from radio terminals. This capability effectively allows a single radio terminal to concurrently transmit or receive voice and data at the same time.

18.3.4.9 Full Duplex Voice Communications

TDMA technology inherently supports full duplex communications. Although full duplex voice communications can be supported on FDMA systems the need for duplex operation requires RF screening between the transmitter and receiver and also a duplexer to allow single antenna working. Because of this, duplex FDMA radio terminals are usually bulkier and more costly to produce than TDMA terminals, which do not need RF screening or antenna duplexers.

18.3.5 Key Services

In developing the TETRA standard to meet the needs of traditional PMR user organizations, numerous services and facilities have been provided. Details of all the TETRA services and facilities can be found in the "About TETRA" Section under TETRA Release 1 and TETRA Release 2. However, in this section it is considered appropriate to list some of the Key Services and Facilities, which clearly differentiate TETRA from other wireless technologies.

The Key Voice Services and Facilities are the following:

- Group Call (commonly called "all in formed net" and "talk group call")
- Preemptive Priority Call (Emergency Call)
- Call Retention
- Priority Call
- Busy Queuing
- Direct Mode Operation (DMO)
- Dynamic Group Number Assignment (DGNA)
- Ambience Listening
- Call Authorized by Dispatcher
- Area Selection
- Late Entry
- Voice Encryption.

18.3.5.1 Group Call

This is probably the most basic voice service in TETRA but yet the most complex to support effectively and efficiently. This is because group calls need to:

Use simple "Push To Talk" operation to provide fast call setup group communications

Be operated and managed in particular ways to optimize network loading, some examples being the following:

- Operate in simplex
- Operate on a “preferred” site for optimum network loading
- Have a defined area of operation (Area selection)
- Have a very reliable call-set up signaling protocol to ensure all users in a group are connected together when a call is first initiated (call acknowledgment signaling is impractical for group calls)
- Have priority mechanisms to ensure that specified users in a wide area group call (spanning multiple base station sites) are connected together when a network is busy.

It is this complexity needed to support group calls that makes public cellular networks unsuitable, simply because they were originally designed to support “One to One” calls, unlike TETRA which was primarily designed to support group calls at the outset.

18.3.5.2 Preemptive Priority Call

This call service, of which the highest priority is the emergency call, provides the highest uplink priority and highest priority access to network resources. If a network is busy, the lowest priority communication is dropped to handle the emergency call. Unlike 911, 112 or 999 initiated public network emergency calls (which can also be supported on TETRA) the TETRA emergency call can be initiated by using a dedicated switch located on the terminal. Activating the emergency call automatically alerts the affiliated control room dispatcher and other terminal users in that persons talk group.

18.3.5.3 Call Retention

This service protects selected radio terminal users from being forced off the network as a result of preemptive calls (emergency calls) during busy periods. When emergency calls are supported in a network, it is essential that only a small number of radio terminal users are provided with this facility as the objective of retaining important calls during busy periods could be lost.

18.3.5.4 Priority Call

During network busy periods, that service allows access to network resources in order of user terminals call priority status. As there are 16 levels of priority in TETRA, this service is very useful in providing different Grade of Service (GoS) levels (and tariff structures) during busy periods. For example, front line officers would be provided with the highest priority levels in a Public Safety network to maintain the highest level of service access whilst routine users would be provided with lower priority levels.

18.3.5.5 Busy Queuing

In TETRA a queue is provided in the trunking controller during network busy periods to store and handle calls on a First In First Out (FIFO) basis in order of user priority level. The advantage is that a user only has to initiate a call request once, knowing that even in busy periods the call will be automatically established once a traffic channel becomes free, thus reducing user stress and frustration when contending with other users on a busy network.

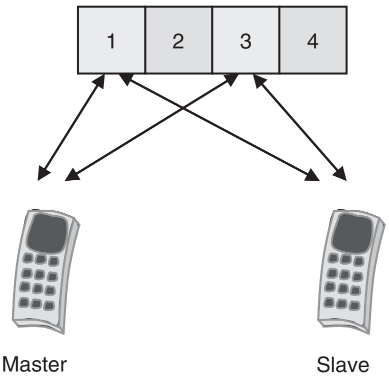

Figure 18.5 In the direct mode of TETRA, there are normally 2 timeslots in use per direction which provide the communications for a single connection per frequency. Nevertheless, the frequency efficient mode provides two calls per frequency.

18.3.5.6 Direct Mode Operation (DMO)

Direct Mode Operation (DMO) provides the ability for TETRA radio terminals to communicate directly with each independent of the TETRA network infrastructure. DMO is not new and has been a facility mandated and used by many traditional PMR user organizations for several decades. The primary requirement for DMO has been brought about by the need to balance the RF Coverage, Grade of Service (GoS) and Reliability of a network with that of the network's overall cost. The requirement for DMO makes the use of public cellular networks unsuitable.

There is also a so-called frequency efficient mode defined for TETRA which provides means for serving two duplex calls per frequency as shown in Figure 18.5. In this mode, only one timeslot is reserved per direction. As can be seen in Figure, the master station has reserved timeslot 1 and the salve uses timeslot 3, which leaves the timeslots 2 and 4 for another frequency efficient direct call.

18.3.5.7 Dynamic Group Number Assignment (DGNA)

This service allows the creation of unique Groups of users to handle different communication needs and may also be used to group participants in an ongoing call. This service is considered by many public safety organizations to be extremely useful in setting up a common talk group for incident communications. For example, selected users from the Police, Fire and Ambulance could be brought together to manage a major emergency where close coordination between the three emergency services is required. Similarly, DGNA is also considered useful for managing incidents by other user organizations such as Utilities and Transportation.

18.3.5.8 Ambience Listening

A dispatcher may place a radio terminal into Ambience Listening mode without any indication being provided to the radio terminal user. This remote controlled action allows the dispatcher to listen to background noises and conversations within range of the radio terminal's microphone. This is an important service to utilize for those persons transporting important, valuable and/or sensitive material that could be "hijack" targets. Similarly, this is a useful service to have implemented in public service vehicles where a driver's health and safety could be at risk.

The number of user applications for the Ambience Listening service are numerous and in many cases application specific. However, it is important to note that many users feel that this service invades a person's privacy and for this reason only those users who need Ambience Listening as part of their work duties should be provided with this service.

18.3.5.9 Call Authorized by Dispatcher

This service allows the dispatcher to verify call requests before calls are allowed to proceed. This is a useful service to utilize when radio user discipline needs to be maintained. This service also reduces the amount of radio traffic on a network as only essential work related calls are permitted. However, the frequent need for all informed net group communications between terminal users and the time delay experienced in authorizing calls can make this service unacceptable for some user organizations.

18.3.5.10 Area Selection

Area Selection defines areas of operation for users and can be chosen on a "call by call" basis. This service basically simulates the ability for a dispatcher to select different base stations to make a call as was possible in conventional networks. This service also helps to improve network loading and overall spectrum efficiency by restricting the area of operation for selected group calls.

18.3.5.11 Late Entry

This service provides continuous call in progress updates to allow latecomers to join a communication channel. This is not a service but an air interface feature that allows a trunked radio terminal to behave in a similar way to conventional PMR terminals. For example, if a user turns on their TETRA terminal the control channel will automatically divert the user's terminal to a talk group call, if a call is already in progress. Similarly, if the user's terminal has been outside radio coverage, for example in a tunnel, the control channel will also divert the user's terminal to a talk group call assuming a call is already in progress.

18.3.5.12 Voice Encryption

The TETRA standard supports a number of over the air TETRA Encryption Algorithms (TEAs), the differences being the types of users who are permitted to use them. The main benefit of over the air encryption is that it can be implemented as software within radio terminals and base station equipment, instead of using encryption modules, which consume space and increase cost. The TETRA standard also supports "end to end" encryption using a variety of other encryption algorithms as deemed necessary by national security organizations.

18.3.6 Functionality

TETRA and GSM have quite a few similarities although they are different systems. Thus, the functionality of TETRA and GSM cannot be compared with each others as for applications and technical setup.

TETRA is based on TDMA (time division multiple access) similarly than, for example, in the case of GSM. TETRA frequencies have been divided into four timeslots which are repeated periodically. As a comparison, GSM has 8 full rate timeslots per frequency. Figure 18.6 clarifies the TETRA frame structure.

The professional point of view is seen in the push-to-tangent which is utilized in the half-duplex mode. The frequency division of TETRA is 25 kHz (compared to 200 kHz in GSM). The modulation technique of TETRA is DQPSK (differential quarternary phase shift keying). The data rate required for the encoding of

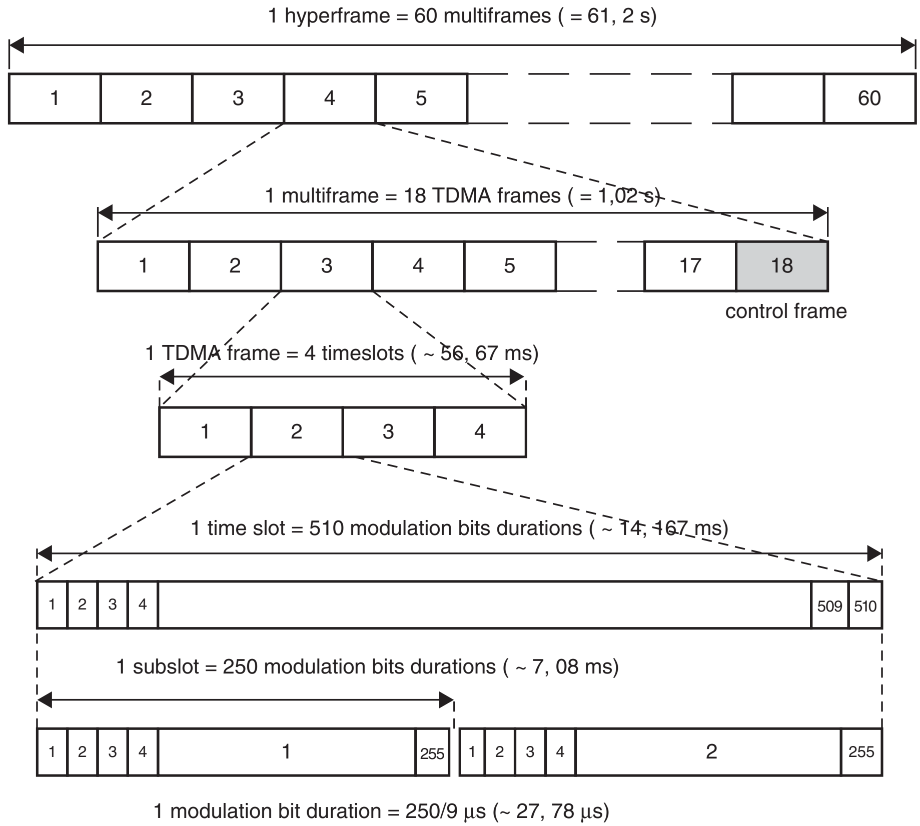

Figure 18.6 *The frame structure of TETRA.*

voice transmission, with the error coding included, is 7.2 kb/s. It should be noted that the voice call can be selected as half-duplex or full-duplex mode.

The elemental radio resource of TETRA is a timeslot which has duration of 14.167 ms. The information is transmitted during that time at a modulation rate of 36 kb/s, corresponding to 510 bits (255 symbols) including guard and ramping times.

The hyperframe represents the highest level of the frame hierarchy. A single hyperframe is divided into 60 multiframes, and it lasts 61.2 s. A single multiframe is divided into 18 frames with duration of 1.02 s. The last, eighteenth frame in the multiframe is a control frame.

A single frame is further divided into 4 timeslots and its duration is 56.67 ms. A single timeslot duration is 14.167 ms corresponding to 255 symbols. It should be noted that the uplink timeslots may be divided further into 2 sub-timeslots.

The TETRA channel types are relatively simple as it only has two types of physical channels:

- Traffic Physical channel (TP) which transports mainly traffic channels.
- Control Physical channel (CP) which transports solely the control channel. One CP channel is defined as the MCCH, the others are called Extended Control Channel (ECCH). The Radio Frequency (RF) carrier containing the MCCH is called the main carrier.

An extensive description of the TETRA functionality can be found in [1].

References

1. ETSI. Definitions of ETSI for TETRA: http://www.etsi.org/technologies-clusters/technologies/tetra (last accessed 3 August 2014).
2. Source: Wireless Intelligence 2008 / http://comm.ae/world-gsm-subscriber-statistics-end-q407/ (last accessed 3 August 2014).
3. ETSI EN 300 392-2 V2.3.2 (2001–03). Terrestrial Trunked Radio (TETRA);Voice plus Data (V+D); Part 2: Air Interface (AI).
4. Pahlavan, K. and Krishnamurti, P. (2001) *Principles of Wireless Networks*. Prentice Hall PTR Upper Saddle River, NJ, USA.
5. Engadget Mobile. Globalstar. http://www.engadgetmobile.com/2006/11/09/globalstar-gsp-1700-satphone-also-loaded-with-ev-do/ (last accessed 3 August 2014).
6. CDG. CDMA2000 explanation, 17 November 2009: http://www.cdg.org (last accessed 3 August 2014).
7. Mobile Engadget. Globalstar. http://mobile.engadget.com/2006/11/09/globalstar-gsp-1700-satphone-also-loaded-with-ev-do/ (last accessed 3 August 2014).
8. Comm.ae. http://comm.ae/world-gsm-subscriber-statistics-end-q407/ (last accessed 3 August 2014).
9. CDG. http://www.cdg.org/ (last accessed 3 August 2014).
10. ITU, CDMA and CDMA 2000 statistics. http://www.itu.int/ITU-D/ict/newslog/CDMA+Subscriber+Base+At+475+Million+Worldwide.aspx (last accessed 5 October 2014).

19

Security Aspects of Telecommunications: 3GPP Mobile Networks

Jyrki T. J. Penttinen

19.1 Introduction

The fraud related to mobile devices can include unauthorized use, manipulation or tampering the equipment needed for communications. The form of fraud depends on the technologies and can be estimated to develop to a more advanced degree along with the general advances of the technologies. As an example, during the first generation of the mobile communications systems, the cloning of subscription identities was one of the most used forms of illegal activity. Some concrete issues were the cloning of the actual subscription number in order to utilize telecommunications services whilst the respective costs were generated to the original legal owner of the subscription. As technologies have been advanced much faster than the legislation, there has been a transition time that has caused important updates to the laws against modern fraud, as can be interpreted in references [1–3].

Not only has legislation been developing, but preventing technologies have taken major leaps to protect users and operators.

19.2 Basic Principles of Protection

Mobile communications is one type of radio communications, and therefore the early cases of frauds related to first and second generation mobile communications were complicated as for the interpretation of typically very outdated radio emission laws. As an example, in the USA, the Telecommunications Act of 1996 had been the first major overhaul of telecommunications laws in the latest decades. The main goal of the new law is to liberate the communications business by allowing fair competition in any market.

The Telecommunications Handbook: Engineering Guidelines for Fixed, Mobile and Satellite Systems, First Edition.
Edited by Jyrki T. J. Penttinen.

In the relatively early phase of the mobile communications, cloning of first generation cellular phones represented the major part of the cellular telecommunications fraud. Legislation was updated in many places due to this, to criminalize the use, possession, manufacture or sale of cloning hardware or software. As the following mobile system generations already took into account the lessons learned from the cloning efforts, at the moment the focus has changed and the primary fraud type is related to subscriptions.

Subscriber fraud refers to signing up for service with a false customer information, for example, by showing a stolen or modified identification. This is relatively tricky as for the personal efforts to correct the situation, if someone has utilized the funds of the victim. It also may take a relatively long time before the victim realizes that there have been fraudulent acts. In any doubts, the recommendation is to contact carrier customer service to discuss the fraud.

For the cloning efforts of the cellular device, the hardware has individual electronic serial number (ESN) and telephone number (MIN) in case of CDMA system, and IMEI (International Mobile Equipment Identity) in case of 3GPP devices. A cloned mobile device refers to the equipment that has been reprogrammed and informs the network other hardware ID. In order to clone a mobile device and to use it in such a way that another person receives the bill requires typically old generation systems like AMPS and CDMA that are based on the HW, not on the encryption of the SIM card.

The LTE/SAE network is based on IP which means that the same threats are valid as in any other packet networks. As for the security processes, the main aim of the LTE/SAE operator is to reduce the opportunities for the misuse of the network [4].

Since the early days of 3GPP 3G system, security has been identified as an essential part of the whole service. The first, Release 99 specifications included thus 19 new specifications by SA3 working group, including the main definitions found in TS 33.102 (3G Security – Security Architecture). Ever since, 3GPP has produced advanced specifications for the security, taking into account heavier also the IP domain as the mobile networks are developing towards IMS and all-IP concepts.

3GPP SA3 has created new specifications for the LTE/SAE protection under TS 33.401 (Security Architecture of SAE) and TS 33.402 (Security of SAE with non-3GPP access). The LTE system provides confidentiality and integrity protection for signaling between LTE-UE and MME. The confidentiality protection refers to the ciphering of the signaling messages. The integrity protection, in turn, assures that the signaling message contents are not altered during the transmission.

All the LTE traffic is secured by using Packet Data Convergence Protocol (PDCP) in the radio interface. In the control plane, PDCP provides both encryption and integrity protection for the RRC signaling messages that are delivered within the PDCP packet payload. In the user plane, PDCP performs encryption of the user data without the integrity protection. It should be noted that the protection of the internal LTE/SAE interfaces like S1 is left as optional.

19.3 GSM Security

When the first phase of GSM was under specification, the security aspects were quite different from the current world. Many of the modern threats we face today were still not experienced by the operators or users, like mobile phone viruses and DoD (Denial of Service) attack attempts. At that time, it was sufficient to comply with the general requirement selected as a base for the specifications, that is, the GSM system needed to have as good as, or better security level than that of fixed networks had. It was thus straightforward to comply with this requirement.

The essential security aspects of GSM are the subscriber authentication and authorization, radio interface encryption for both signaling and communications, and the utilization of temporary identity during communications.

19.3.1 SIM

The subscriber identity module (SIM) of GSM is a smart card. Physically speaking, SIM cards come in several sizes, dictated by the form factor (FF). The first form factor FF1 refers to the credit-card-sized card that was planned to be used in the first phase deployments back in the early 1990s. In practice, the form factor FF2 initiated the GSM era, with approximately 2.5 × 1.5 cm of size. After that, there have been smaller cards deployed, that is, FF3 (micro card) and FF4 (nano card). The SIM card contains permanent records of the subscriber's identity, and authentication algorithms A3 and A8. In contrast, the A5 algorithm, of which there are currently three versions available (A5/1 for good protection, A5/2 for slightly lower protection, and new variant A5/3), is stored directly in the mobile terminal. Also the adapted A5-algorithm taken into use with GPRS is already stored in the terminal, which means that the user does not need to switch SIM cards in order to use GPRS. The user can store changing data such as phone numbers connected with alphanumeric information and short messages on the SIM card. The subscriber can also define blocks to calls to certain numbers and store them on the SIM card.

The subscriber can, in principle, use any GSM terminal using the SIM card. Thus a subscriber to an European GSM 900 or GSM 1800 network can use his/her own SIM card and a suitable mobile phone to use a GSM 1900 network in the USA, as long as the operator in question has signed a roaming agreement with the subscriber's home network operator.

The benefits of the SIM card are for example, the safety functions, interoperability with in principle any GSM phone and network, and the flexible adding and removing of services. As the SIM card is basically a regular smart card, it may be possible in the future to combine functions unrelated to the GSM network such as a bankcard to it. Thus it can be called a multifunctional or multi card.

The SIM card of the GSM system is one example of contact smart cards; other examples include prepaid calling cards, bankcards and access control cards. Such smart cards have been specified in ISO series 7816 standards, whereas basic smart cards can be found in ISO series 14443 standards.

The user can specify a personal identity code (PIN-code), so that the terminal always asks for the code before activating to enable a connection when power is switched on. An exception is the European emergency no. 112, US emergency no. 911, and other national variants, which are possible to dial without the PIN code. It is also possible to call emergency number with a mobile phone even without a SIM card.

The PIN code consists of a maximum of four digits. If the PIN code function has been activated and an incorrect PIN code is entered three times, the SIM card asks for the personal unblocking key (PUK). If an incorrect PUK code is entered ten times, the SIM card locks itself so that the user can no longer activate it.

The same physical card is applied in UMTS and LTE devices, although the security functions have been further developed. For M2M communications, there is also non-removable MFF2 SIM.

19.3.2 Authentication and Authorization

GSM system may check the user's right to utilize the network in the initiation of the call, location area update, and in the termination of the call. This happens with the help of AuC (Authentication Center). Even though AuC is a separate logical functionality, it is typically integrated into HLR because they do have the communication link in any case.

Each subscriber is given with a subscriber-specific key Ki which is stored in the user's SIM as well as in the AuC. This happens in the provision phase when the subscriber is created, and the subscriber profile is activated in the HLR of the home network of the subscriber.

Figure 19.1 shows the situation when the user initiates the call attempt.

AuC calculates in advance so-called triplet which contains parameter values for RAND (Random number), SRES (signed response) and Kc (temporal key). This triplet is stored into the VLR to which the user is

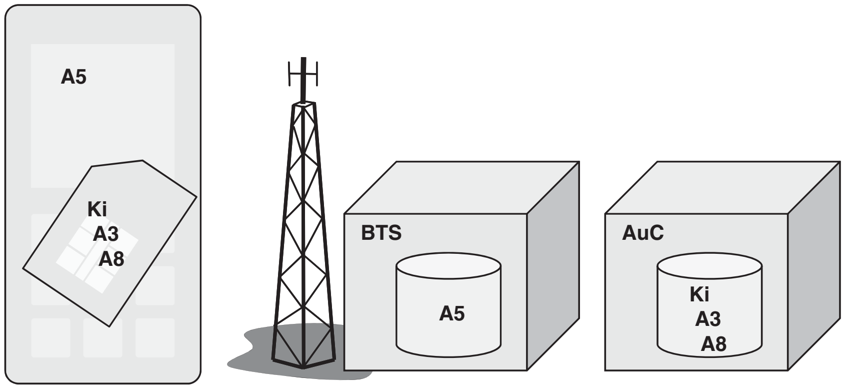

Figure 19.1 The subscriber-specific Ki, as well as the A3 and A8 algorithms are stored in SIM and AuC. The A5 algorithm has been stored, in turn, in the HW of the user device and in the base transceiver equipment.

currently subscribed, as shown in Figure 19.2. It should be noted that the actual user-specific key Ki is never transferred between the network elements or over the radio interface.

As shown in Figure 19.3, the authentication happens in such a way that the RAND generated by AuC. The value range of the RAND is 0.2^{128}-1. This value is sent as a part of the initial signaling via SDCCH (Stand Alone Dedicated Control Channel) to the SIM.

In the next step, based on the received RAND and the Ki key and A3 algorithm stored in the SIM, the SIM calculates the SRES. Based on the same information, AuC has also calculated the SRES already in advanced. Now, the user device sends the SRES value to corresponding VLR via SDCCH, and VLR compares the SRES values it received from SIM and AuC. If the SRES values are different, the call attempt is terminated. SRES may be incorrect if the user has an incorrect A3 or Ki key. The network reasons thus that the call attempt is not authorized.

A3 algorithm is operator-specific and resides in SIM and AuC. Logically, it is recommendable to use as secure algorithm as realistically possible in order to assure that the Ki key cannot be calculated via RAND and SRES values. A3 may be public, but typically it is intended to keep classified.

The length of RAND is 128 bits and SRES has always 32 bits. The length of the Ki key is operator-specific.

19.3.3 Encryption of the Radio Interface

As soon as the network has concluded that the user is authorized, that is, the SRES is correct, the signaling and all the communications will be encrypted. This happens in two phases. First, a temporal key is generated, and secondly the key is utilized for the actual radio interface encryption.

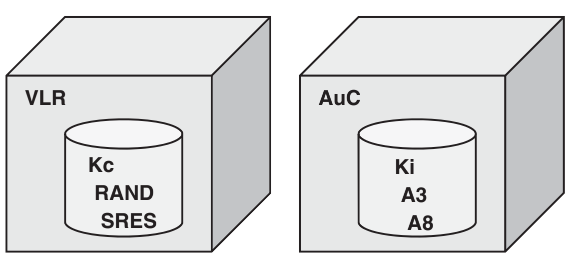

Figure 19.2 By utilizing Ki, A3 and A8, the AuC calculates the triplet, that is, values for the Kc key, RAND and SRES. The triplet is stored into VLR.

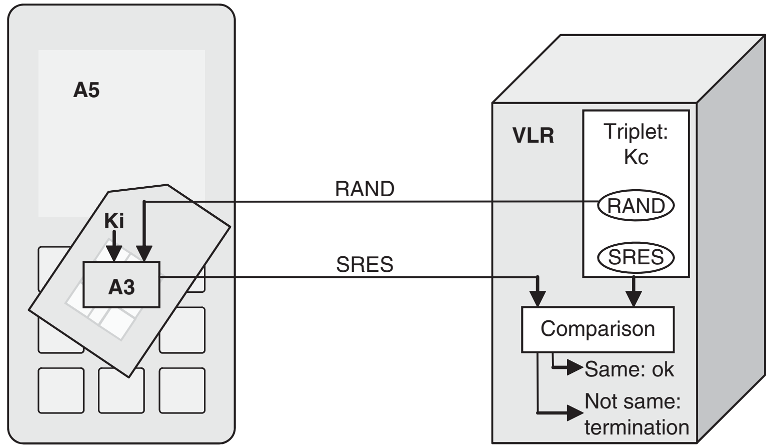

Figure 19.3 *The authentication and authorization happens with A3, RAND and Ki.*

SIM and AuC calculate the connection-specific key Kc by utilizing the same RAND value as described previously, and by applying the A8 algorithm that is stored in SIM and AuC. Figure 19.4 clarifies the procedure. The length of the Kc is always 64 bits in GSM. If the key is shorter, zero-bits are added until the full 64 bit length is reached.

Kc is calculated separately for each connection. The A8 algorithm is operator-specific, and the same principles for its recommended complexity apply as for the A3. As the A3 and A8 are both operator-specific, they can be combined. In that case, the algorithm is called A38. It is of high importance to assure the combined algorithm is not too simple and that the Kc can be protected.

As soon as the Kc is calculated, the actual encryption of the radio interface starts taking place as shown in Figure 19.5. As from this point, all the communications between the user equipment and base transceiver station is scrambled, including the signaling (e.g., location area updates and handovers), voice and data calls and short/multimedia messaging.

For the scrambling, the same connection-specific Kc and A5 algorithm are used. The algorithm takes as an input the Kc key and the super frame number COUNT of the GSM radio interface timeslot structure, and the result is fed into logical XOR (Exclusive Or) operation together with the single burst information of 114 bits.

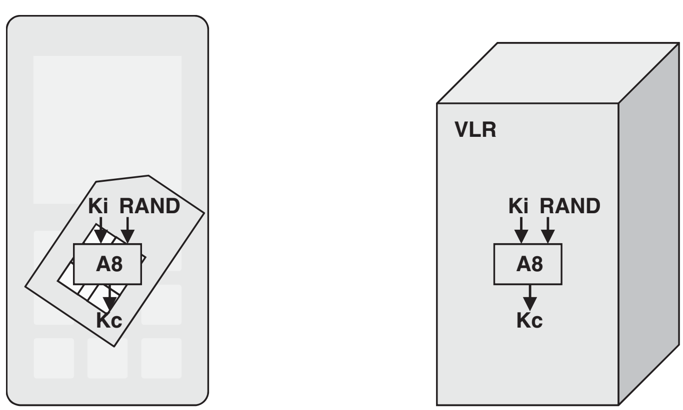

Figure 19.4 *Kc is calculated with A8 algorithm, based on Ki of SIM and RAND received from AuC/VLR.*

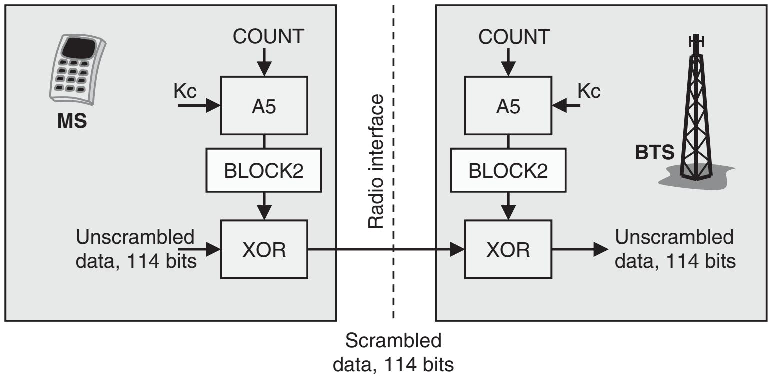

Figure 19.5 *The encryption of the GSM radio interface takes place via A5 algorithm.*

The superframe cycle is about 3.5 hours, and the length of the COUNT is 22 bits. The block that A5 produces is called BLOCK1 or BLOCK2. BLOCK1 is utilized for the encryption and BLOCK2 is utilized for the unscrambling as presented in Figure 19.5.

Because the COUNT gets different value for each block, the scrambling of each burst's 114 information bit are different from each others. In case the connection takes longer that the repetition cycle of the superframe, the COUNT values will be repeated accordingly. The values for Kc and RAND are calculated from scratch each time new connection is established.

After all, the efficiency of the A5 algorithm does not depend on the nonpublicity as each operator has the description of it. In any case, GSM Association has decided to maintain it confidential, so it has not been published officially.

According to the specification, there can be seven different A5 algorithms. At the moment, there are 3 algorithms produced, A5/1, A5/2 and A5/3. The original A5/1 is designed as a strong one whereas the A5/2 is lighter, the latter being easier to decrypt by governments and outsiders. For the case where no encryption is used, A5/0 is applied. There is currently also A5/4 algorithm for 128 bit cases defined as of Release 9.

In the beginning of the connection, the user equipment informs the network about the wanted version of A5 algorithm. In case AuC does not have the common algorithm, and the network does not support nonencrypted calls, the call attempt is not successful. If the network supports nonencrypted connections, the call is established now without encryption. In this case, user will notice a symbol, for example, open lock on the display of the device to indicate that the communications is not encrypted. If common algorithms are present, the network decides the version.

19.3.4 Encryption of IMSI

The IMSI (International Mobile Subscriber Identity) is always priority identifier within GSM network. Nevertheless, the revealing of it in radio interface is typically avoided. Instead, TMSI (Temporary Mobile Subscriber Identity) is used to replace the IMSI. This gives additional value for the user identity protection.

Nevertheless, the support of TMSI is not mandatory, so each operator decides the strategy for its use. If it is supported, the user equipment gets a new value each time it registers to a new location area (LA). The TMSI is sent over the radio interface scrambled via SDCCH.

19.3.5 Other GSM Security Aspects

Additional protection for eavesdropping is achieved via slow frequency hopping which has the interval of a single burst. Another aspect that increases the challenges in the systematic user-specific eavesdropping or getting location information is the relatively small cell size in urban areas. Furthermore, the cells are typically sectorized, and power control as well as discontinuous transmission gives extra challenges for the monitoring.

19.3.6 Potential Security Weaknesses of GSM

As the GSM system specifications were developed already in late 1980s and the first commercial networks were launched as of 1991, the security threats were not seen a significant danger. Thus, the enhanced security for the already existing fixed network was sufficient. This means that the default assumption was that within the networks, the security was sufficiently well guaranteed.

Nevertheless, potential attack methods have been identified during this time. Even if the GSM authentication and authorization is still in highly secure level, there is a potential threat that someone creates a false base transceiver station (BTS) nearby the user. If the network ID is selected identical to the user's home operator, and the BTS signal level is sufficiently higher than the real ones, a fraudulent BTS may capture the calls in the initial phase of the call in such a way that the encryption is forced off – as it is the network that decides the selection of the algorithm, the user may only have indicator on display showing that the encoding is not utilized. If the user continues, the actual delivery of the call is straightforward to connect to the originally intended B-subscriber, and the fraudulent BTS provider can follow all the communications for that connection.

One aspect of GSM is that the encryption of the radio interface is valid only between the MS and BTS. The encryption equipment is located at BTS site, so the communication may be eavesdropped, for example, via radio links within the GSM network.

The addition of GPRS in the beginning of 2000s was at the same time a significant step towards all IP concept that helped to overcome with the "old-fashioned" circuit switched (CS) connections. Nevertheless, the CS data communications were well protected because of the closed environment, and point-to-point connections. GPRS brought the openness to the cellular networks with connections exposed to the public Internet. That also started to require the planning of prevention methods against the security aspects which were already familiar to the fixed Internet access users. As an example, firewalls between the GPRS core and the outside world were needed with ever-increasing methods for analyzing suspicious utilization. Not only the eavesdropping of the data connections, but also other aspects were noted to be at least equally important, including the protection against DoD attacks and economical frauds. As an example of a simple method for DoD is to send a very large set of data connection initiation requests from Internet to GPRS network by selecting random mobile subscriptions as receivers. Prior to the PDP context activation, network needs to signal with HLR to find the location of the receiving party (current VLR). As the user might be nonexisting, the signaling results in failed connection. Nevertheless, by repeating these false call requests, Internet-user might cause overload of HLR signaling and prevent the delivery of other traffic. Thus, this type of activity needs to be analyzed in real time by the GPRS operator in order to block the attempts prior to the extra signaling.

Other potential GMS security aspects relate to data integrity, which is not provided yet in GSM networks. Also the altering of IMEI code might be possible, although the specifications were intended to require good protection for it. The protection for these potential security threat aspects have been enhanced in UMTS specification.

19.4 UMTS Security

When UMTS was specified, there were already experiences available from GSM networks that helped in the further development of the protection. The main aspects in UMTS protection are that it is based on GSM principles, that UMTS takes into account the noted GSM weaknesses, and that it adds further security for 3G services.

Concretely, the subscriber authentication of UMTS was enhanced in such a way that it includes also mutual authentication between system and user. This avoids the potential utilization of false base stations which could be utilized in capturing the user communications. There have also been enhancements in protocols and algorithms. In 3G, the related terminology is User Authentication, whereas GSM utilizes term Subscriber Authentication.

UMTS also enhanced the radio interface encryption by introducing longer keys, publicly verifiable algorithms, and multiple radio interfaces between the terminal and base station. Also the functionality of USIM was enhanced [5].

The UMTS authentication and key agreement protocol, UMTS AKA, are applied in the UMTS networks. Furthermore, 3GPP defines EAP-AKA for the authentication of the users that connect via WLAN, and EAP-AKA' to authenticate users that connect via trusted non-3GPP access networks to the LTE core, that is, EPS.

The UMTS specifications define three entities related to the authentication: Home Environment (HE), that is, the home network, Serving Network (SN), and the USIM (UICC). In the beginning of the call attempt, the SN assures the user's identity is correct via challenge-response procedure as is the case in GSM. In UMTS, the User Equipment also makes sure that the SN is authorized by HE to execute this procedure. This is called mutual authentication.

UMTS authentication is based on permanent key K which is stored to the HE database as well as to the USIM of the user. For the encryption and integrity check, additional temporal keys are generated. As soon as the user is identified, the actual authentication takes place via AKA procedure. For that, authentication vector is generated by AuC and stored to the respective VLR/SGSN. In the beginning of the authentication procedure, the VLR (in case of CS call) or SGSN (for PS calls) sends user authentication request to UE, with two parameter values for RAND and AUTN. USIM of the UICC receives these parameters. USIM already has the permanent UMTS key K which is used to calculate the authentication vectors already prepared also by AuC. This happens via multiple algorithms, and the outcome is the assurance if AuC actually generated the AUTN. Now, if the AuC was noted correct, a RES value is sent back to VLR/SGSN which, in turn, compares it with the one (XRES, Expected Response) generated at AuC. If the RES and XRES are the same, the authentication phase is passed.

Along with the authentication procedure, the temporal CK (key for radio access network encryption) and IK (integrity protection) values are also calculated by USIM which transfers them to the mobile equipment for the actual encryption. On the other hand, CK and IK are transferred from AuC to VLR/SGSN. As the encryption and integrity procedure is initiated, these values are transferred to the RNC via RANAP (RAN Application Protocol) message, Security Mode Command. RNC thus takes care of the actual encryption in UMTS.

Prior to the ciphering, the UE and RNC agree the version of the encryption algorithm. The algorithm set is enhances and there is more variety in UMTS compared to GSM. The encryption is then done in MAC (Medium Access Control) or RLC (Radio Link Control) layer. Each PDU (Protocol Data Unit) increases a relatively small counter. This counter is CFN (Connection Frame Number) in MAC layer, and RLC-SN (RLC Sequence Number) in RLC layer. A counter with larger value range is also applied, HFN (Hyperframe Number). The combination of all these counters is COUNT-C.

To complement the encryption, some additional inputs are needed: BEARER (radio bearer identity), DIRECTION (indication of the UL or DL encryption) and LENGTH (indicates the size of data to be encrypted).

To complement UMTS security, integrity protection mechanism is also applied in RRC layer with IK integrity key. The RRC message together with DIRECTION/1, IK/128, COUNT-1/32 and random number FRESH/32 (the number indicating the bit size for each parameter) is fed into one-way function f9.

19.5 LTE Security

19.5.1 Security Process

The development of the security processes contains many items. The aim of all the security measures is to prevent in advance the possible attacks by shielding all the relevant LTE/SAE interfaces and elements in such a way that outsiders have minimum possibilities to advance in the fraudulent activities.

The security design of LTE/SAE thus includes feature development according to the best knowledge about the current and future methods for the attacks together with their technical and business impacts to the network. For instance, security threats like a denial-of-service attack can slow down, or in worst case, paralyze big part of network and cause limited availability of services, which results in loss of revenue and increases the chances of customer churn. One possibility to develop up-to-date measures against these security threats is to create a security process.

The first step in the security planning is to identify the security threats. Based on the security risk analysis of this phase, the LTE/SAE system is designed and updated accordingly in order to create all the possible countermeasures against the identified security risks. This leads to the list of security requirements, and to the specification of the security architecture layout in the system level.

The next step is to take into account threats in the software level, securing the code as much as possible in the software development processes.

At the end of the security design process, a comprehensive security testing is needed, by taking into account the imaginable attack types in the normal operations of the network, as well as in the unstable conditions that are created either intentionally or accidentally.

This example of the security process is logically an iterative activity for the participating parties. This means that as the technologies develop and new methods and ideas for the system attacks arise, they should be identified and taken into account at the earliest possible moment to be included in the security planning process so that the network can be updated accordingly to work against the new threat types. As one part of the new security threat identification, a network fraud monitoring process should be implemented. This provides information about the possible new security threats at time in order to be taken into account in the security process.

In addition to the security process, it is recommendable to execute a security audits in the operator networks. This is an important task as there are huge amount of different combinations of network elements with their different versions and security levels in the end-to-end chain of the mobile networks. Both hardware and software can be audited in cooperation with the network vendor and operator. If vulnerabilities are detected, the issues can be corrected at time by enhancing updated security threat countermeasures.

Figure 19.6 shows high-level items in the LTE/SAE security environment.

19.5.2 Network Attack Types in LTE/SAE

The LTE/SAE architecture has some special characteristics that should be taken into account in the enhanced security planning. The important fact is that LTE/SAE is based on flat architecture, which means that all radio access protocols terminate to the eNodeB elements. Furthermore, the IP protocols are also visible in eNodeB.

The challenges arise from the architectural realization of LTE/SAE, for example, because it is possible to place eNodeB-elements in more accessible location to potential hacker attacks. Furthermore, the LTE/SAE network interworks with legacy and non-3GPP networks that might open unpredictable security holes even if their normal mode of operation would not open these issues. There are also new business environments with networks whose trustfulness is not necessarily known.

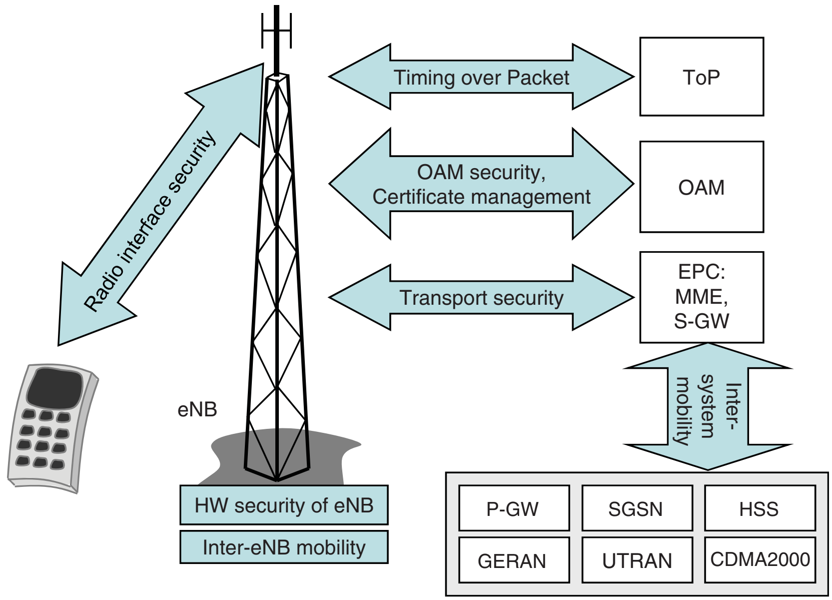

Figure 19.6 *The LTE/SAE security chain includes various high-level aspects.*

Comparing the purely IP based LTE/SAE architecture with the previous 2G/3G principles, it can be noted that LTE/SAE requires extended authentication and key agreements in order to cope with the modern IT attacks. This means that the key hierarchy as well as the interworking security are inevitably more complex that earlier. It also means that the eNodeB element has additional security functionality compared to the previous 2G base transceiver stations and 3G node B elements.

The identification of the potential network attack types in LTE/SAE environment is one of the most essential preventive tasks. As the Home eNB concept basically means that the customer has possibility to try to access physically the hardware and software of the element, it is one of the most potential roots for the fraudulent activities. Some possibilities might be the following:

- Cloning of HeNB credentials.
- Physical attacks on HeNB, for example, in the form of tampering.
- Configuration attacks on HeNB, for example, fraudulent software updates.
- Protocol attacks on HeNB, for example, man-in-the-middle attacks.
- Attacks against the core network, for example, denial of service.
- Attacks against user data and identity privacy, for example, by eavesdropping.
- Attacks against radio resources and management.

19.5.3 Preparation for the Attacks

More detailed list of the LTE/SAE security related items to be taken into account in the security process include the following:

– Air-link security (U-plane and C-plane security). This includes the definition and description of ciphering algorithm for the U-plane and C-plane, definition and description of integrity protection algorithm for C-plane, and description of the access stratum security signaling (including key distribution).

- Transport Security. This item includes the definition and description of ciphering and integrity algorithms for transport network, and description of the transport security signaling (including key distribution).
- Certificate Management. This item includes the definition of public key and key management concepts.
- OAM Security (M-plane security). This item includes the management plane security.
- Timing over Packet (ToP). This item includes the Synchronization Plane security for IEEE v2 packets for frequency and time/phase synchronization.
- eNB Requirement. This item includes the definition of secure environment, requirement definition for eNB according to 3GPP TS 33.401, and secure key and file storage.
- Intra LTE and Inter System Mobility. This item includes the definition of security aspects in handover case (including key distribution).

It should be noted that different planes differentiate the traffic types, and this should be taken into account in the security planning. The planes in the LTE/SAE environment are: U-plane for the delivery of the user data, C-plane for the delivery of the control data, M-plane for the delivery of the management data and S-plane for the frequency and time/phase synchronization information. Figures 19.7–19.10 identify the security related aspects of these planes.

As will be described further in the next sections, IPSec is the 3GPP standardized solution for security on several LTE interfaces. These are S1-MME and X2 Control Plane as well as S1 and X2 User Plane. Security for the Management Plane is not standardized but the use of IPSec or transport security is also suggested. In addition, usage of IPSec in combination with certificates makes it very difficult for any unauthorized person to gain access to the core network or eavesdrop the traffic between the eNBs and the core network. In this way integrity and confidentiality of data can be ensured.

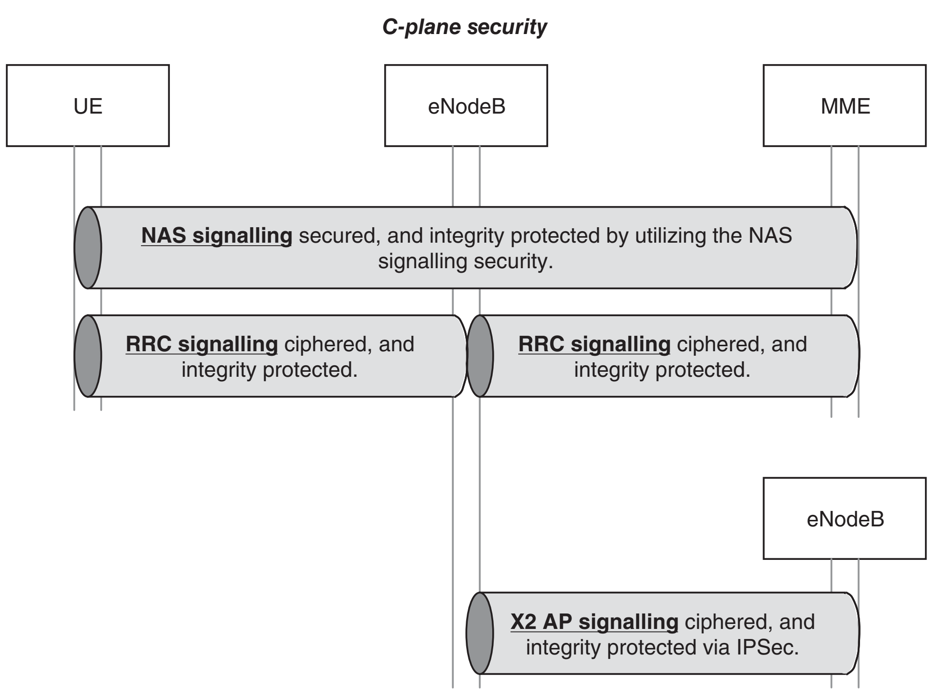

Figure 19.7 The C-plane security principle of LTE/SAE.

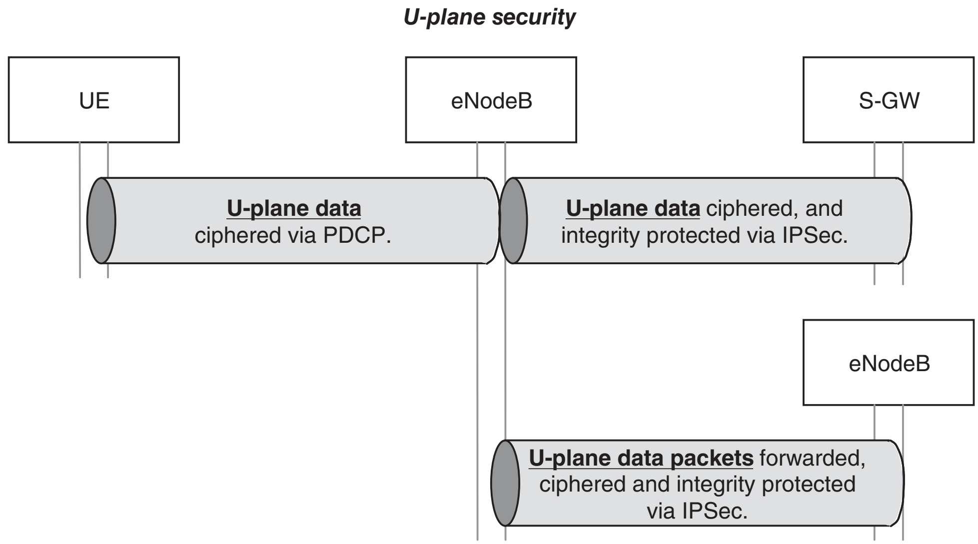

Figure 19.8 *The U-plane security principle of LTE/SAE.*

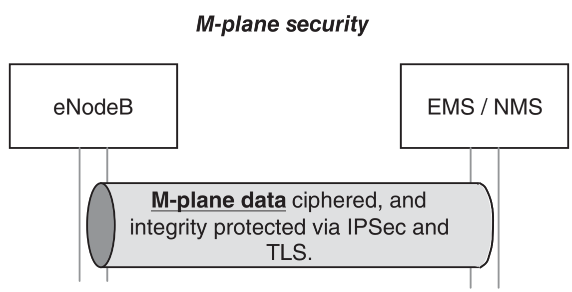

Figure 19.9 *The M-plane security principle of LTE/SAE.*

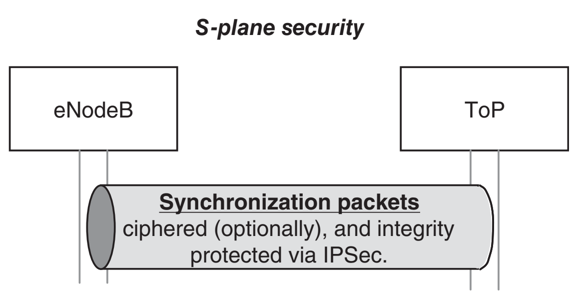

Figure 19.10 *The S-plane security principle of LTE/SAE.*

19.5.4 Certificates

19.5.4.1 X.509 Certificates and PKI

Digital certificates are used to authenticate communication peers and to encrypt sensitive data. They are essentially for Transport Layer Security and for IPSec support. X.509 certificates contain a public key that is signed by a commonly trusted party. Via this method, the receiving end trusts in the correctness of the public key, as long as a trusted party has confirmed the matching identity by using its digital signature as part of the certificate. This idea of the trusting parties and certificates forms a trust chain.

The essential problem of trust chains is to get the keys delivered into the place at the stage when the security does not yet exist due to the absence of keys. The most secure way would be to install the needed keys locally on site. This solution is feasible for a low number of sites, or for a new network which is going to be commissioned in any case. Nevertheless, for a large network such as LTE, this is extremely challenging because the certificates have a limited lifetime and they need to be replaced from time to time.

In order to cope this challenge, the Certificate Management Protocol (CMP) is a functional option that, for example, Nokia Networks (NSN) utilizes. This standardized protocol provides the capability to retrieve, update and revoke certificates automatically from a central server. Initial authentication (when operator certificates are not yet in place) is done based by vendor certificates (which are installed in the factory) which are trusted by the operator certification authority. As a result, the eNB Public Key Infrastructure can be introduced as a flat hierarchy, having one root CA in the operators' network.

The eNB supports NSN secure device identity where a vendor certificate is installed to the eNB in the factory. The vendor certificate is used to identify eNB in the operator CA and to receive operator certificate for the eNB. This functionality is used as a part of SON LTE BTS Auto Connectivity feature and supports the automated enrolment of operator certificates for base stations according to 3GPP TS 33.310.

Figure 19.11 shows an example of the interaction of the operator and vendor. The process begins in the factory, where the public and private key pair is created, and vendor's certificate is created and signed. From the factory, a vendor's device certificate is stored to the eNodeB, and also delivered to the Factory Certification

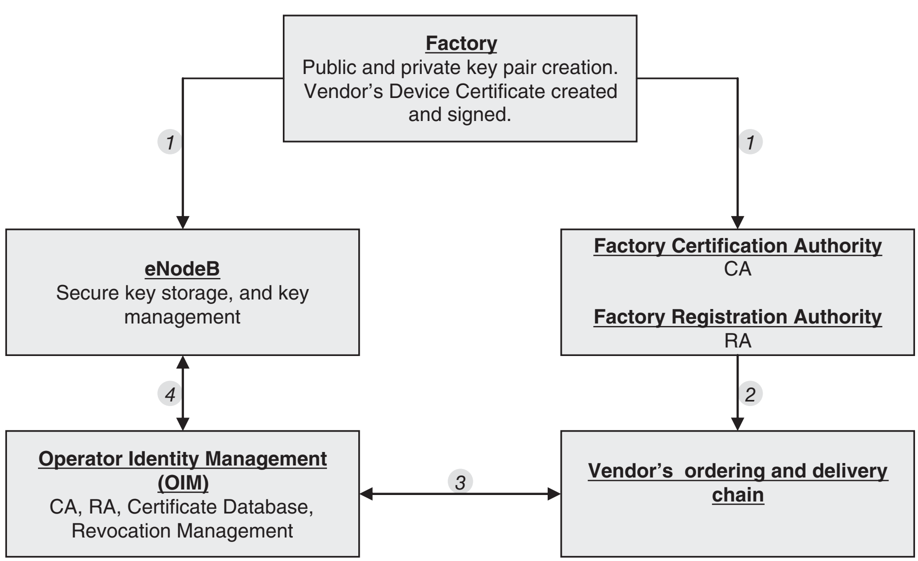

Figure 19.11 *The principle of the vendor certificates. This process can be utilized in realistic LTE/SAE network environment.*

Authority (CA) and Factory Registration Authority (RA) as shown in delivery chain (1) of the figure. Next, the product information, consisting of the module serial numbers and vendor's root CA certificate, is placed to the vendor's ordering and delivery chain. At this point, the eNodeB is shipped to the operator. Following the process, the vendor's device serial numbers and vendor's root CA certificate is delivered to the operator (3), that is, to the operator identity management (IDM). Next, the IDM creates an operator node certificate to assure the authenticity of the eNodeB (4). At this point, the operator node certificate substitutes the previous vendor's node certificate. It is now possible to make sure that the equipment is genuine by observing the serial number and vendor's device certification.

19.5.5 LTE Transport Security

19.5.5.1 IPSec and IKE

The eNB follows the rules established by the Network Domain Security / IP Security (NDS/IPSec) architecture defined in 3GPP TS 33.210. The 3GPP TS 33.210 specification introduces Security Gateways (SEG) on the borders of security domains to handle the NDS/IPSec traffic between the same. All NDS/IPSec traffic passes through a SEG before entering or leaving a security domain. The SEGs are responsible for enforcing security policies for the interworking between networks. In this role they also provide the IPSec functions.

It is possible to implement the SEG functionality in dedicated hardware (external SEG) or to integrate it in existing nodes (internal SEG). From the eNB point of view it is not relevant if SEGs are external to the peer-entities or integrated, that is, it is not visible for the eNB. At eNB side the IPSec function is integrated into the eNB. Therefore, the eNB represents a security domain of its own and can act as SEG according to 3GPP TS 33.210.

The following logical interfaces can be protected by IPSec means:

- S1_U: User data transport (U-plane) between eNB and S-GW (GTP-U tunneling)
- S1_MME: Signaling transport (C-plane) between eNB and MME (S1AP protocol)
- X2_U: User data transport (U-plane) between eNB nodes during handover (GTP-U tunneling)
- X2_C: Signaling transport (C-plane) between eNB nodes (X2AP protocol)
- O&M i/f: Transport of O&M data (M-plane) between eNB and O&M System
- ToP i/f: Transport of ToP synchronization data (S-plane) between eNB and ToP Master.

Figure 19.12 shows the eNB protocol stacks with the embedded IPSec layer.

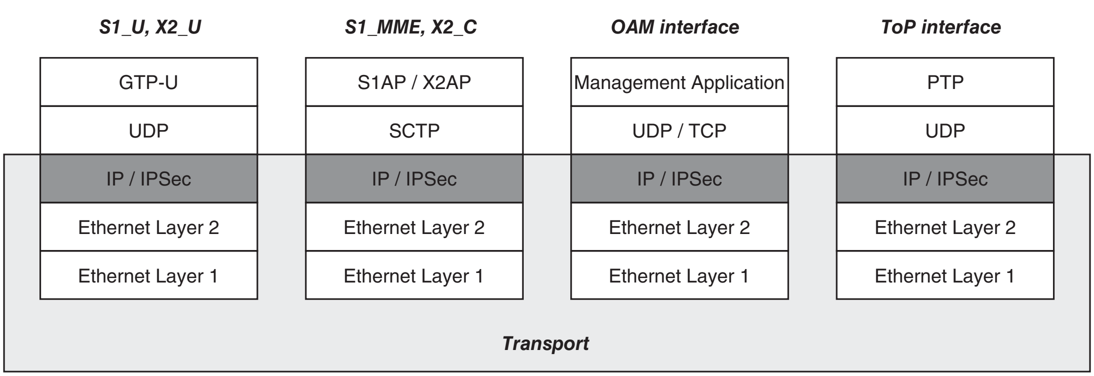

Figure 19.12 *eNB Protocol Stacks with embedded IPSec Layer.*

19.5.6 Traffic Filtering

19.5.6.1 Firewall

It is possible to have the eNB element to support a firewall function with the following key capabilities:

- Ingress IP packet filtering
- Ingress rate limiting
- Egress rate limiting
- DoS countermeasures.

19.5.6.2 Filtering for Site Support Equipment

The eNB might provide access to site support equipment (e.g. battery backup units etc.) via additional Ethernet interfaces. Typically, this type of IP based equipment does not provide own IP packet filter or firewall. Thus, the site support equipment would be directly accessible if there is no packet filter at eNB side. Therefore, the eNB provides an IP packet filter service that protects the site support equipment from harmful network traffic, but also protects the network from unintended traffic from this interface.

19.5.7 Radio Interface Security

19.5.7.1 Access Stratum Protection (as Protection)

The following section provides information about the key hierarchy and key derivation function. The key derivation is relevant for normal operation, that is, call establish and further in case of a handover (i.e. inter eNB handover).

19.5.7.2 Key Hierarchy

Figure 19.13 gives an overview about the LTE overall key hierarchy concept. It shows the EPS key hierarchy valid for steady state, that is, if no handover happens. Nodes are represented by frames, and keys are

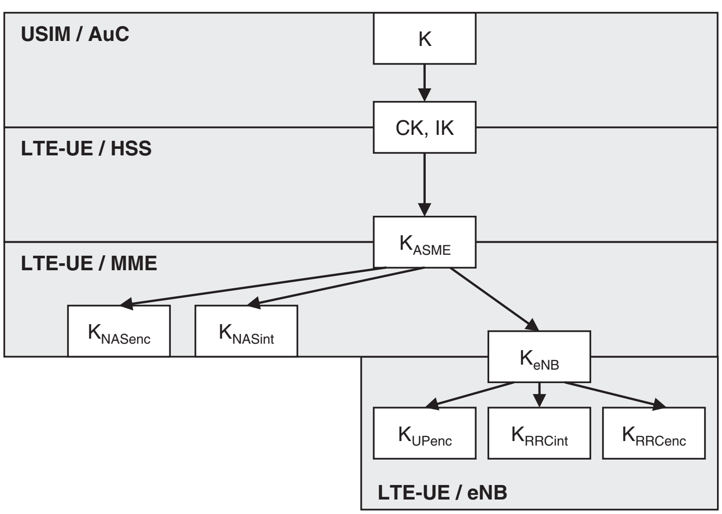

Figure 19.13 *LTE Key Hierarchy Concept.*

represented by boxes. An arrow represents a key derivation function. If a key will be derived at one node and transmitted to another, then its corresponding box is located on the border of the frames which corresponds to the involved nodes.

The part of the key hierarchy which is known to eNB is called the eNB key hierarchy. It comprises all AS keys, that is, K_{eNB}, K_{UPenc}, K_{RRCint}, and K_{RRCenc}. There is one AS base-key K_{eNB} and three AS derived-keys K_{UPenc}, K_{RRCint}, and K_{RRCenc}. The AS derived-keys are used for: UP encryption, RRC integrity protection, and RRC encryption.

The key "K" is the only permanent key. All other keys will be derived on demand using a key derivation function. A key derivation is controlled by a key derivation procedure.

The existence of a key depends on the state in the following way:

- K exists always.
- The NAS keys CK, IK, K_{ASME}, K_{NASenc} and K_{NASint} exists while EMM-REGISTERED.
- The AS keys K_{eNB}, K_{UPenc}, K_{RRCint}, and K_{RRCenc} exists only in RRC-CONNECTED.

19.5.7.3 Key Derivation Functions

Key derivation works with a KDF (Key Derivation Function), which in case of EPS is a cryptographic hashes function, with Ky = KDF(Kx, S), which calculates a hash with key Kx from string S. The hash value becomes the derived key Ky. The following applies:

- Kx is a superior key, that is, a key which is located in the hierarchies on the higher level (besides handover, where it may be at the same level).
- Ky is the derived inferior key.
- The string S is a concatenation of several substrings, which may be classified as follows: (1) Bound: String representation of parameters to which the key Ky shall be bound to. Usually, these parameters describe a part of the environment, for example, a cell identifier, and the key Ky is only valid while these parameters do not change. (2) Fresh: String representation of parameters which shall ensure different (or "fresh") keys Ky also if all other parameters are unchanged. Usually, these parameters are unique for each instant of calculation, for example, a random number.

A cryptographic hashes function provides a result of a fixed size and it is not reversible, that is, it is not feasible (at current state of the art) to derive an unknown parameters if the result and the other parameters are known. In particular, it is not feasible to get Kx, even if Ky and the string S is known.

19.5.7.4 Key Establishment Procedures

There are three basic key establishment procedures:

- Authentication and Key Agreement (AKA) establishes CK, IK, and KASME on the one hand in USIM and UE, and on the other hand in AuC, HSS, and MME. AKA is a NAS procedure and does not have any prerequisite besides the permanent key K. Please note that MME is the Access Security Management Entity (ASME) for the EPS.
- NAS Security Mode Command (NAS SMC) establishes the NAS keys K_{NASenc} and K_{NASint} which are needed for NAS message encryption and integrity protection. NAS SMC is a NAS procedure and needs a valid KASME as prerequisite. In addition, NAS SMC activates the NAS security.

– AS Security Mode Command (AS SMC) establishes the AS keys K_{UPenc}, K_{RRCint}, and K_{RRCenc} which are needed for UP encryption and RRC integrity protection and encryption. AS SMC is an AS procedure and needs a valid K_{eNB} as prerequisite. In addition, AS SMC activates the AS security.

The establishment procedure for the key K_{eNB} depends on case:

– At change to RCC-CONNECTED, K_{eNB} will be derived in MME and transmitted to eNB by the S1AP: INITIAL CONTEXT SETUP REQUEST message.
– At active intra-LTE mobility, K_{eNB} will be derived by a procedure which is shared by source eNB and target eNB.

In case of intra-LTE handover, the key hierarchy differs temporarily, because the K_{eNB} for target eNB will be derived from the K_{eNB} of source eNB.

19.5.7.5 Key Handling in Handover

Figure 19.14 shows the key handling in handover. The boxes represent keys and the arrows KDF. All keys of the same row are derived in a single chain of KDFs starting from an initial K_{eNB} or NH (Next Hop parameter). These chains are called forward chains in short.

At initial context setup, an initial K_{eNB} will be derived from KASME at MME. This starts the first forward chain (NCC = 0). The initial K_{eNB} will be transmitted to the eNB and becomes its K_{eNB}.

At handover, a transport key $K_{eNB}{}^*$ and finally a fresh target K_{eNB} will be derived. Because the derivation uses a cryptographic hashes function, it is not feasible to derive the source K_{eNB} from the fresh target K_{eNB}. Therefore a target eNB cannot expose the security of a source eNB. This is called (perfect) backward security.

However, if derivation happens on one forward chain, that is, if $K_{eNB}{}^*$ is derived from the source K_{eNB}, then the target eNB keys are no secret for the source eNB, because their derivation is known. This is true in a recursive manner. Therefore, all keys of the same forward chain which are located right hand of any K_{eNB} are no secret to this key owner. There is no forward security.

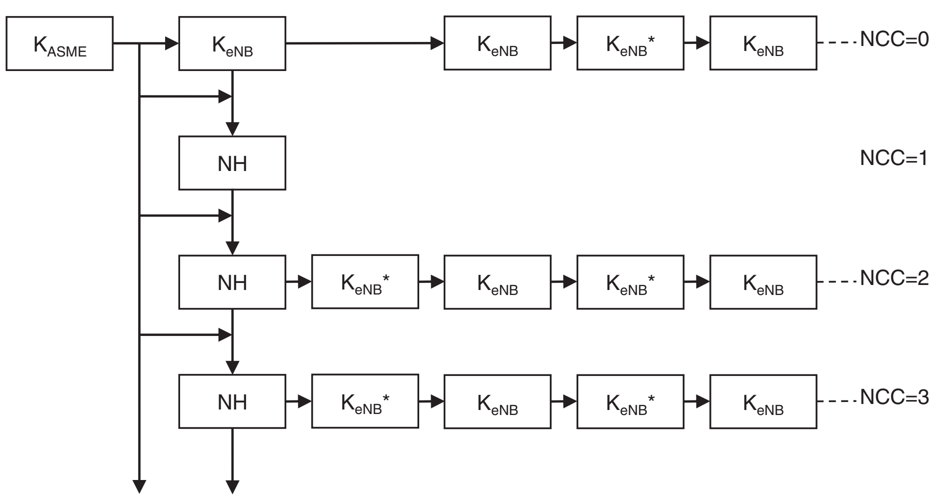

Figure 19.14 *Key Handling in Handover.*

In order to get forward security, the current forward chain has to be terminated and a new one started. This is done by deriving $K_{eNB}{}^{*}$ from a parameter NH which was derived from KASME and not only a source K_{eNB} (vertical chain of yellow boxes). In case of S1 handover, NH is transported by S1AP: HANDOVER REQUEST and applies for this handover. Therefore, forward security is reached by this handover. This is called forward security after one hop, which may also be called perfect forward security. In case of X2 handover, NH is transported by S1AP: PATH SWITCH ACKNOWLEDGEMENT and cannot apply for this handover (because the new keys are already determined at this point in time) but the next one. Therefore, forward security is reached at next handover. This is called forward security after two hops.

If NCC reaches its maximum value of 7, it will wrap around at next increment.

Note: The forward chain 1 will be skipped (by 3GPP standardization) first time (at first round of NCC after initial context setup). At each later rounds (after NCC wrapped around) forward chain 1 will usually not be skipped.

19.5.7.6 Security Handling of RRC Connection Reestablishment

If a UE decides to try RRC connection reestablishment, following steps regarding security happens:

1. The UE transmits on SRB0 a RRC CONNECTION REESTABLISHMENT REQUEST message to the cell it selected for requesting RRC connection reestablishment. This cell shall be called requested cell, and its controlling eNB requested eNB. The message contains a UE-Identity which tells the last serving cell from UE point of view, that is, the cell in which the trigger for RRC reconnection occurred, to the eNB. This cell shall be called serving cell and its controlling eNB serving eNB. Please beware that the serving cell is defined from UE point of view. Usually the point of view for UE or network does not matter, because UE and network are synchronized, but a broken RRC connection may cause different assumptions for UE and eNB about serving cell. At RRC connection reestablishment, the network will adapt its serving cell assumption to UE. The UE-Identity is called related to the eNB which controls the reported physical (=serving) cell. Please note that this SFS uses the term "requested" for RRC connection reestablishment, and "target" for handover. However, other documents (including some 3GPP specifications) do not differ in this way and use just "target."
2. The RRC CONNECTION REESTABLISHMENT REQUEST message cannot be authenticated by PDCP MAC-I integrity protection, because it transmits across SRB0. Instead, the requested eNB checks the authentication of the received UE-Identity by comparing a received authentication code, which is contained in the UE-Identity IE and is called shortMAC-I, with an authentication code calculated by the network. Each cell which shall be enabled to authenticate a RRC reestablishment request applies a dedicated shortMAC-I, because this code is bound to a cell. The following cases may happen:
   - *If the serving cell is controlled by the same eNB as the requested cell*, the authentication on network side is an internal matter of the requested eNB and may happen on demand.
   - If *the serving cell is controlled by a different eNB as the requested cell*, authentication on the network side is matter between two eNBs: (A) The calculation of the network side authentication code happens on the serving eNB, because it requires the RRC integrity protection key (KRRCint) of the serving cell. (B) The comparison of the authentication codes happens on the requested eNB, because it got the shortMAC-I from the UE. The calculation of a set of shortMAC-I and the delivery to another eNB happens at course of a handover preparation. Background: An RRC connection reestablishment is only possible to an eNB which knows some UE context information, this is a prerequisite not only, but also, motivated by security issues. Therefore, a requested eNB must either contain the serving cell (above bullet) or already be prepared for a handover from the serving cell, that is, contain a handover target cell (this bullet) for the UE. Otherwise, the RRC connection reestablishment request will fail.

The requested cell is controlled by the same eNB as the serving cell, if the UE-Identity is related to the requested eNB, that is, if the physical cell identity reported by the UE-Identity belongs to the requested eNB.

If the RRC connection reestablishment request is accepted, both UE and eNB will refresh their AS key hierarchy in the same way as for handover from the serving cell to the requested cell, however always keeping the security algorithms of the serving cell. The possible cases are the same as for authentication:

– If the serving cell is controlled by the same eNB as the requested cell, the reestablishment procedure is on network side an eNB internal matter. In particular, the key refresh happens like for intra eNB HO or, if the requested cell is equal to the serving cell, by intra cell AS security maintenance.
– If the serving cell is controlled by a different eNB as the requested cell, the reestablishment procedure is on network side a matter between two eNBs. In particular, the key refresh happens like for inter eNB HO, according to the HO type which prepared the target cell (which exists, see related description for shortMAC-I). (A) In case of X2 HO, the source eNB needs to calculate a dedicated KeNB* for each cell which shall support a reestablishment and to signal it to target eNB at course of handover preparation. This is similar to the shortMAC-I provision described above. (B) In case of S1 HO, the target eNB derives the fresh KeNB from a NH parameter received from MME. Because the NH is independent from cell, nothing special for reestablishment support is needed (however, there is still the limitation by cell dedicated shortMAC-Is).

In any case, the UE needs to know the NCC parameter for key refresh. It will be signaled by the RRC CONNECTION REESTABLISHMENT message. This message is unprotected, because it transmits across SRB0. Please note: In case the X2 interface is not protected by IPSec then the X2AP messages including the keys are transferred in plaintext. SRB1 (and all later by RRC connection reconfiguration procedure added bearers) will apply the fresh keys immediately.

Please note that in the relationship between RRC connection reestablishment and handover, from UE point of view the RRC connection reestablishment procedure is independent from the handover: The UE will always send its request to a selected cell and, if it was accepted, refreshes its AS keys in exactly the same way according to the received NCC parameter. From the network point of view the RRC connection reestablishment procedure depends on handover:

– If no handover is prepared, a RRC connection reestablishment can only succeed to the serving (this time from network point of view) eNB, because no other eNB knows the UE context. Also, the UE and network must share their assumptions about the serving eNB (but the serving cells may differ).
– If a handover is prepared, a RRC connection reestablishment can succeed to both the source and to the target eNB, because they know the UE context.
– The handover source eNB may always expect to be addressed as serving eNB, there is no difference (besides names) to the "no handover" case above. In contrast, the handover target eNB may be addressed either as serving eNB or not. If it is addressed as serving eNB, the UE regards the HO as completed. Otherwise, the UE regards the HO either as uncompleted or did not become aware of it at all, and will report the source cell as the serving cell.

19.6 LTE/SAE Service Security: Case Example

19.6.1 General

LTE/SAE is changing mobile communications towards all IP type of environment. This also has a potential increase in the fraudulent activities. The motivations of such activities might be, for example, financial, destructive or even political. The motivation can simply also be related to a way of proofing the hacker skills.

The modern information technology combined with the advanced mobile technologies brings changes and new aspects that might increase the vulnerability for the intentional attacks by opening new possibilities for the fraudulent intentions.

As an example, the base station equipment has traditionally been well protected physically. The radio and transport equipment have been shielded inside of the site construction which is not accessible for others than authorized personnel. In the future, this type of equipment may be located increasingly in public places and even homes.

On the other hand, the methods for the attacks evolve along the advanced tools which are available easier via the Internet distribution mechanisms. Furthermore, these activities might include increasingly sophisticated attacks performed by the IT professionals.

19.6.2 IPSec

For LTE, IPSec with PKI is utilized as a standardized security solution. PKI is applied to authenticate network elements and authorize network access whilst IPSec provides integrity and confidentiality on the transport route for the control and user plane. The IPSec concept is based on the certificate server which is the registration authority within the operator's infrastructure. It takes care of the certificates which provide secured IPSec routes between the elements via the migrated SecGW, as shown in Figure 19.15.

Security Gateways with high availability such as Juniper or Cisco are examples of scalable platforms terminating the IPSec traffic from the eNodeBs and covering fast growing performance figures for the next years. Insta Certifier is a Public Key Infrastructure (PKI) platform for issuing and managing digital certificates

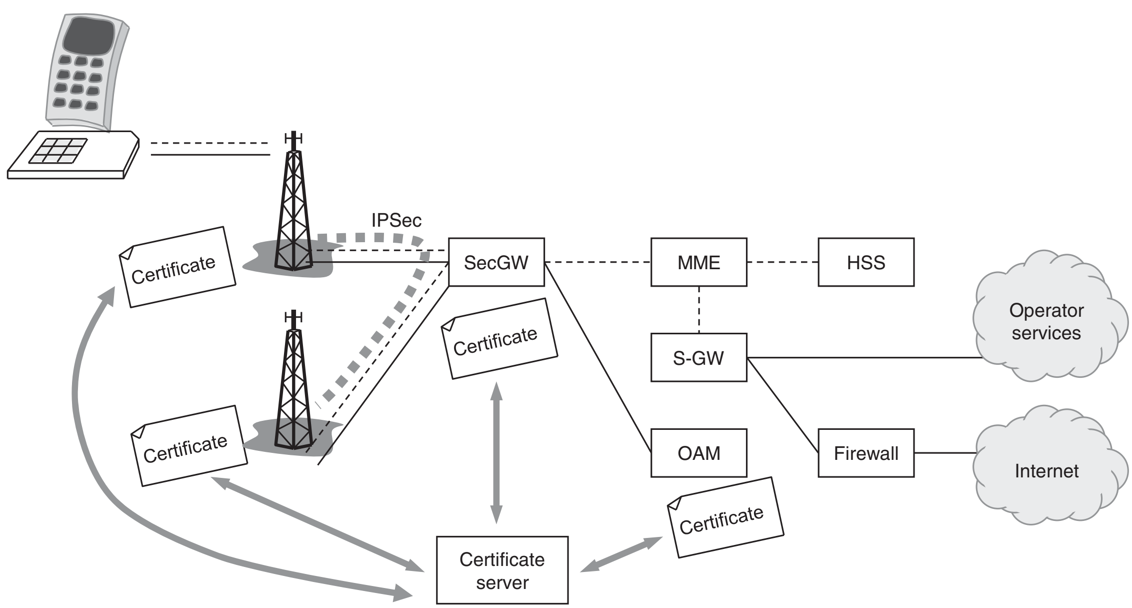

Figure 19.15 The architecture of the combined IPSec and PKI. The light dotted line shows signaling, and solid line shows user plane data flow. The heavier dotted line shows the IPSec tunnel. The communication between SecGW and OAM can be done via TLS/HTTPS.

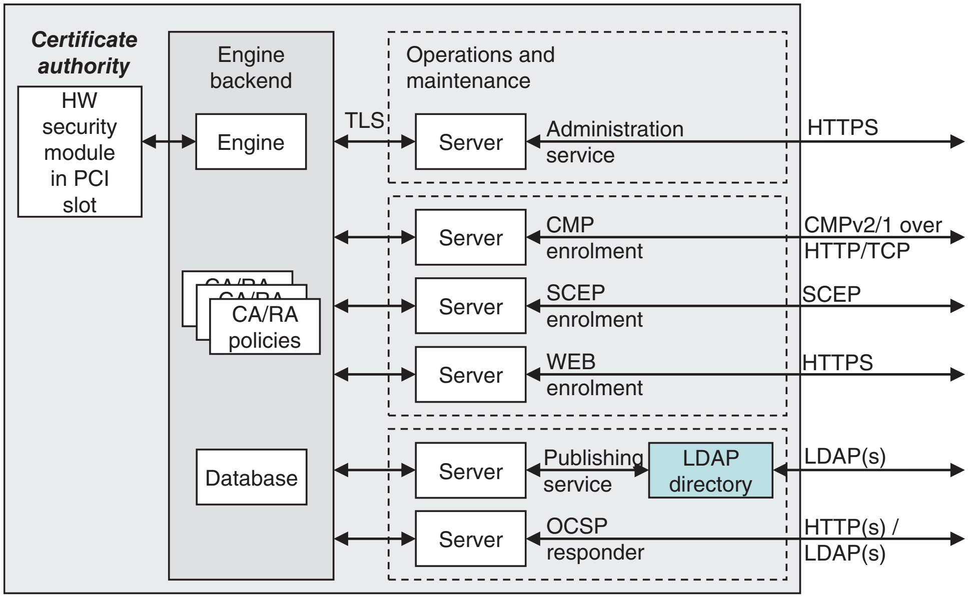

Figure 19.16 *The PKI design with the architecture and interfaces.*

both for users and machines. It provides Certification Authority (CA) and Registration Authority (RA) functionality, that is, manageability for very large PKI environments by introducing centralized management of authentication keys with support for scalable revocation and key updates.

The device identity concept is valid for different entities. LTE/SAE network elements possess identity and are able to authenticate peer and provide their own identity. In addition, there are two kinds of authorities in the security solution. First one is the Factory Registration Authority, which requests vendor certificates. Centralized vendor-wide CA issues and keeps the certificates in a database. The second one is Operator's Certificate Authority which recognizes the vendor's certificates and authorizes requests. This authority issues and manages operator certificates for the network elements.

Figure 19.16 shows the SW architecture and interfaces for the PKI solution.

19.6.3 IPSec Processing and Security Gateway

The LTE standard requires IPSec capability at the eNB level. As an example, Nokia Networks has integrated IPSec and firewall at the Flexi LTE eNB. It includes authentication with the X.509 certificates. On the other hand, the integrated firewall is more complete than a mere address or port filter, and thus the rules are integrated in the eNB with other management, providing automatic configuration in most of the cases.

In general, the LTE standard requires IPSec capability at the eNodeB elements. The support of the functionality is mandatory, but for the trusted networks (considered reliable by the operator), the actual utilization of IPSec is left optional. In practice, the eNodeB can include both IPSec and firewall, with authentication via the X.509 certificates. The combination of IPSec and firewall provides the possibility to integrate the firewall rules with other network management. This means that there is little or no manual configuration needed.

In the following scenario, illustrated in Figure 19.17, the SecGW is complementary placed alongside the Aggregation Router (AR) with both interfaces (incoming and outgoing) directly connected to the AR.

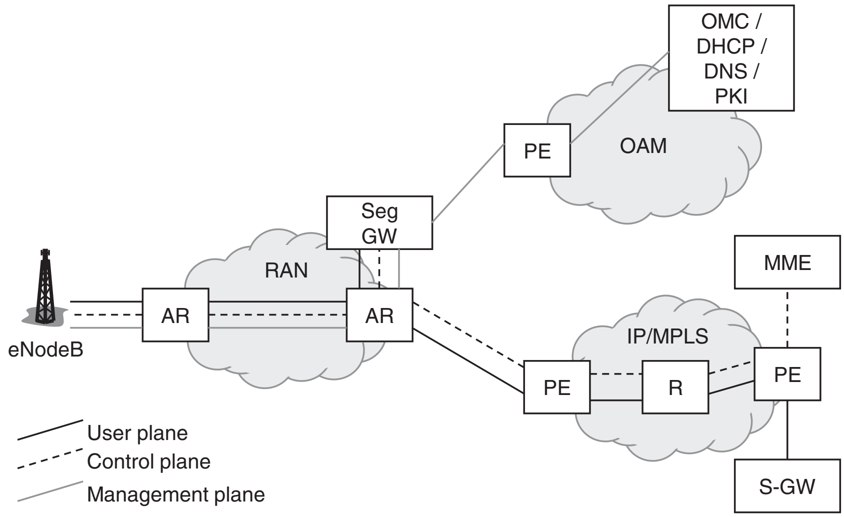

Figure 19.17 *An integration example, with GW attached to the access router.*

The advantage of this scenario is that only minimal changes on the existing network have to be done. Furthermore the scenario allows for aggregation of all interfaces between AR and SecGW into one logical link. Once aggregated, the different kinds of traffic can be separated on Layer 2 by defining corresponding VLAN interfaces on the SecGW. This setup provides a higher level of flexibility and resilience against single link outages.

The SecGW could provide several options to achieve traffic separation:

- **Virtual routers** allow for separating routing domains into separate logical entities with routing table is logically separated. Each virtual routing entity is handling its directly connected networks and static or dynamic routes.
- **VLANs** are used to separate the traffic within physical links. All security concepts work with physical and also logical sub interfaces (tagged traffic on trunk links).
- Definition of dedicated **security zones** as shown in Figure 19.18. These zones are used to logically separate network areas and allow for a more granular traffic filtering and traffic control by defining access control and filter policies.

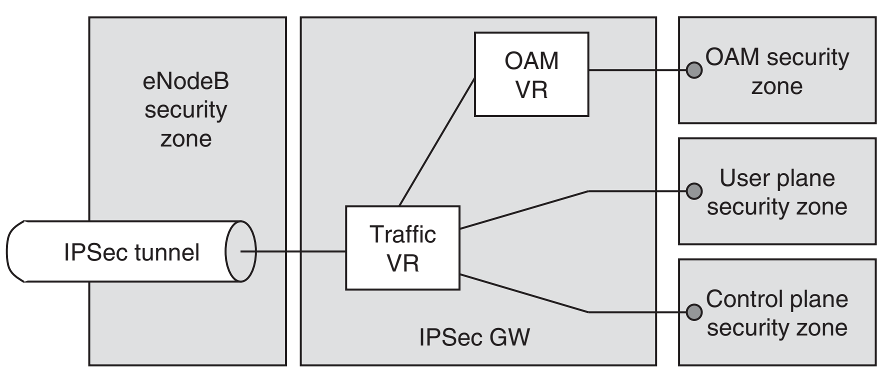

Figure 19.18 *The security zone principle.*

The VPN design of the segGW can be based either on the single tunnel setup, or multiple tunnel setup. For the *single tunnel setup*, the traffic of all planes is encrypted with the same encryption set, that is, either a single IKE-SA or a single IP-SEC SA. The setup can be based on the dedicated tunnel interface per eNB, which means that each eNB have its own tunnel interface on the SecGW. The setup can also be based on the shared tunnel interface, which means that all the eNBs shares one tunnel interface on the SecGW.

For the *multiple tunnel setup*, the traffic for each plane is encrypted with different encryption sets (1 IKE-SA / 3 IP-SEC SA). The multiple tunnel setup can be based on the dedicated tunnel interfaces per eNB, or on the shared tunnel interfaces.

19.6.4 Single Tunnel with Dedicated Tunnel Interfaces

The benefit of this solution is that it provides a persistent tunnel interface per eNB. All routes to one eNB points to the same interface, which makes the design easier. The third benefit is the small amount of security associations (SA). As a drawback, the solution requires a large amount of tunnel interfaces.

19.6.5 Single Tunnel with Shared Tunnel Interfaces

The benefits of this design are that only one tunnel interface is required per chassis. The eNB inner routes can be aggregated on SecGW. As in previous case, only small amount of security associations is needed. As a drawback, the scalability of this design is linked to the IP address concept.

19.6.6 Multiple Tunnels with Dedicated Tunnel Interfaces

The benefit of this solution is related to the dedicated tunnel interfaces per plane per eNB. As drawbacks, three tunnel interfaces per eNB are required. Other issue is the larger amount of security associations. Due to the drawbacks, this is the least feasible solution.

19.6.7 Multiple Tunnels with Shared Tunnel Interfaces

The benefit of this solution is that only one tunnel interface per plane per chassis is needed. In addition, the eNB inner routes can be aggregated on SecGW. As drawbacks, a larger amount of security associations are required. Also, additional IP network for VPN next-hop table is required if the eNB inner routes cannot be aggregated per plane. Furthermore, the scalability is limited by the IP address concept.

19.6.8 Summary

As the solution of multiple tunnels with dedicated tunnel interfaces requires a large amount of tunnel interfaces, it is not recommendable. From the remaining three solution types, the single tunnel design reduces the number of the security associations significantly, which requires alignment between the eNB vendors. As the shared tunnel design provides reduced amount of the tunnel interfaces per chassis, it is a good solution.

19.7 Authentication and Authorization

In the authentication and key agreement process, the HSS generates authentication data and provides it to MME element for the processing. There is a challenge-response authentication and key agreement procedure applied between the MME element and LTE-UE.

The confidentiality and integrity of the signaling is assured via the Radio Resource Control (RRC) signaling between LTE-UE and LTE (E-UTRAN). On the other hand, there is Non Access Stratum (NAS) signaling between LTE-UE and MME. It should be noted that in the S1 interface signaling, the protection is not LTE-UE – specific, and that it is optional to apply the protection in S1.

In the user plane confidentiality, the S1-U protection is not LTE-UE specific. There are enhanced network domain security mechanisms applied that are based on IPSec. Here, the protection is optional. The integrity is not protected in S1-U, for example, due to performance impact, and limited protection for the application layer.

Let's examine the signaling flow for the authentication procedure, which can be seen in Figure 19.19. In the initial phase of the authentication, the MME element evokes the procedure in such a way that it sends the International Mobile Subscriber Identity (IMSI) as well as the serving network's identity (SN ID) to the HSS of the home network of the subscriber in question (2). In case MME does not yet have information about the IMSI code at this stage, it is requested first from the LTE-UE (1). It should be noted that IMSI is delivered

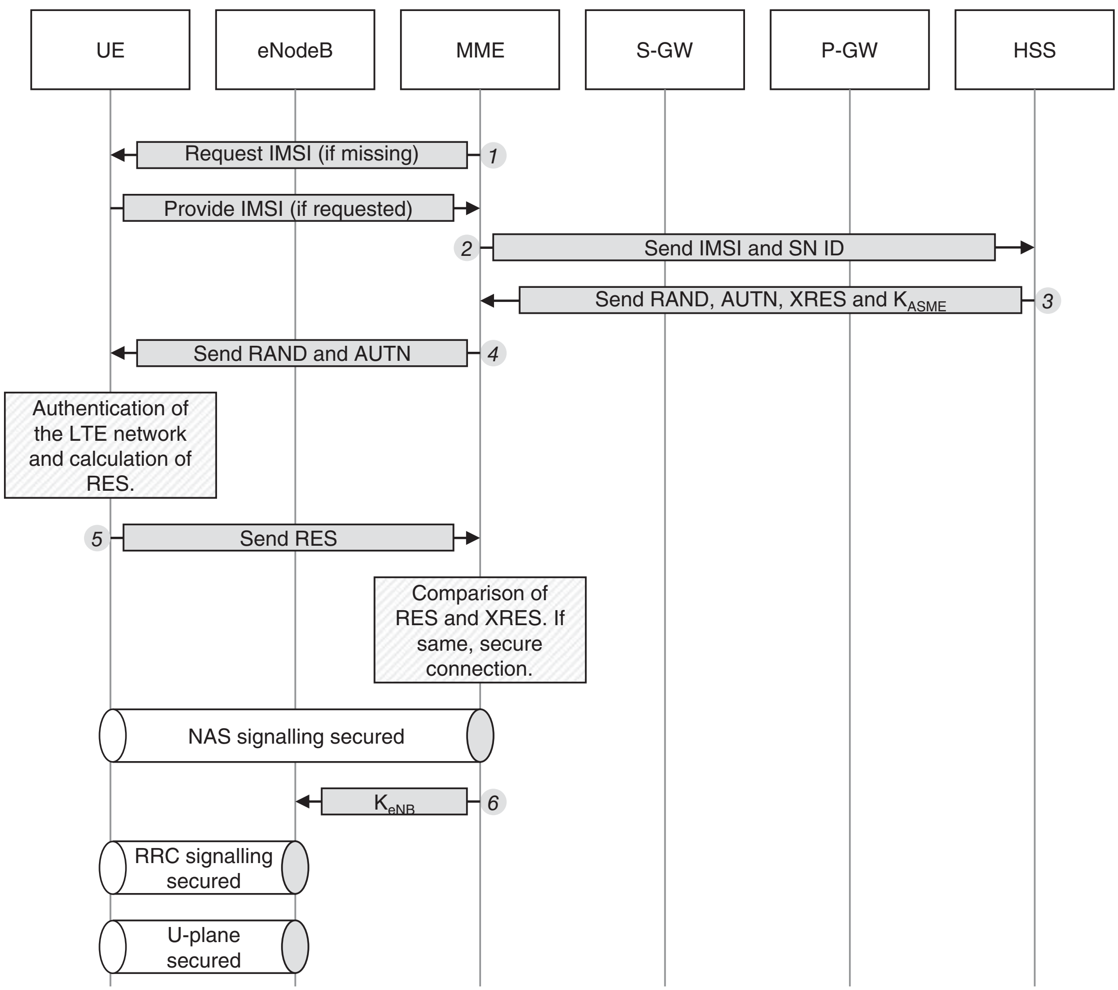

Figure 19.19 *The mutual authentication procedure of LTE.*

over the radio interface in a text format, which means that this procedure should be done only when no other options are available.

As a result to the MME user authentication request to HSS, the latter responds with an EPS authentication vector, which contains RAND (random challenge number), AUTN (authentication token), XRES (expected response), and the ASME key (3).

When MME has received this information, it sends the RAND number and AUTN to the LTE-UE (4). Now, the LTE-UE processes this information in order to authenticate the network (according to the mutual authentication concept). Based on the received information and its own key, LTE-UE also calculates a response (RES) and sends it back to MME (5).

Both LTE-UE and HSS contains the same algorithm which is utilized for the calculation of the response with the same inputs. The RES of LTE-UE, and the expected response XRES calculated previously by HSS are now analyzed by MME. This means that MME compares the RES and XRES values. If they are the same, the LTE-UE is thus authenticated correctly, and the NAS signaling will be secured. The eNodeB key, that is, K_{eNB} is now calculated and delivered to eNodeB in order to secure the radio interface for all signaling and data transmission (6).

19.8 Customer Data Safety

The normal procedures applied in the 2G and 3G environments for the physical protection of the customer subscription data, charging record data and other confidential information are utilized also in the LTE/SAE environment. Also fraud prevention and monitoring that are applied in the previous generations are applicable also in LTE/SAE.

19.9 Lawful Interception

The Lawful Interception (LI) in general has been designed for authorized and official access to the private communications. Via the LI, the mobile/fixed network operator and service provider can collect the delivered traffic and/or identification information of the connections for the purposes of the further analysis by the law enforcement officials with intercepted communications of private individuals or organizations. This possibility has been already for long time in the mobile communications networks. As an example, it was designed as a part of the possibly deployed element, LIG (Legal Interception Gateway) also in the first Release 97 of the basic GPRS solution, which can mirror the traffic delivered via the GPRS nodes.

Although it is possible to store technically basically all the details and contents of the data flows, the legal interception process can only be applied in accordance with the national and regional laws and technical regulations.

Specifically in the case of the LTE/SAE, 3GPP Evolved Packet System provides only IP based services. This logically means that EPS can be utilized for the interception of IP layer's Content of Communication (CC) data flows. The LTE voice connections are considered in this sense as IP data flows via VoIP solutions. If fallback-type of functionality is applied during the LTE voice call in order to continue the voice call as circuit switched version, the respective 2G/3G network contains also the LI. In addition to the user plane interception, the LI solution of EPS can also generate Intercept Related Information (IRI) records in the control plane messages which identifies the called parties, the location of LTE terminal and other call related information.

The functional architecture of the EPS lawful interception is comparable to the functional architecture of the packet switched domain of 3G networks of 3GPP. Figure19.20–19.22 show the standard defined

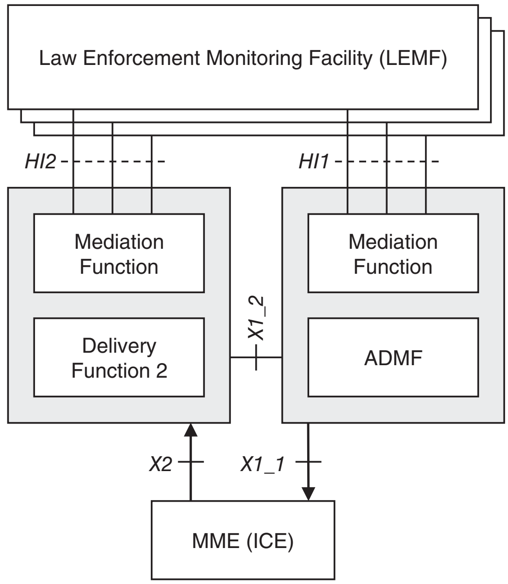

Figure 19.20 *The configuration for the MME intercept.*

configurations for the MME, HSS, S-GW and PDN-GW, respectively, for the EPS lawful interception. The key identities for the interception are IMSI (International Mobile Station Identity), MSISDN (Mobile Station ISDN number) and IMEI (International Mobile Equipment Identity).

As can be seen in the figures above, the MME element handles solely control plane and HSS only handles signaling. The interception of Content of Communication is thus applicable only via the S-GW and

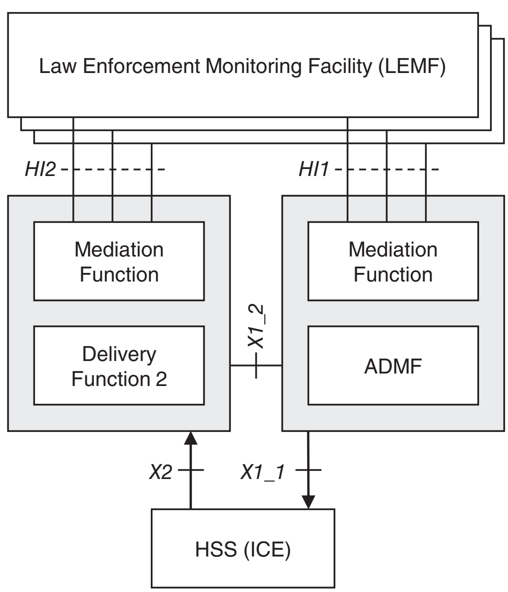

Figure 19.21 *The configuration for the HSS intercept.*

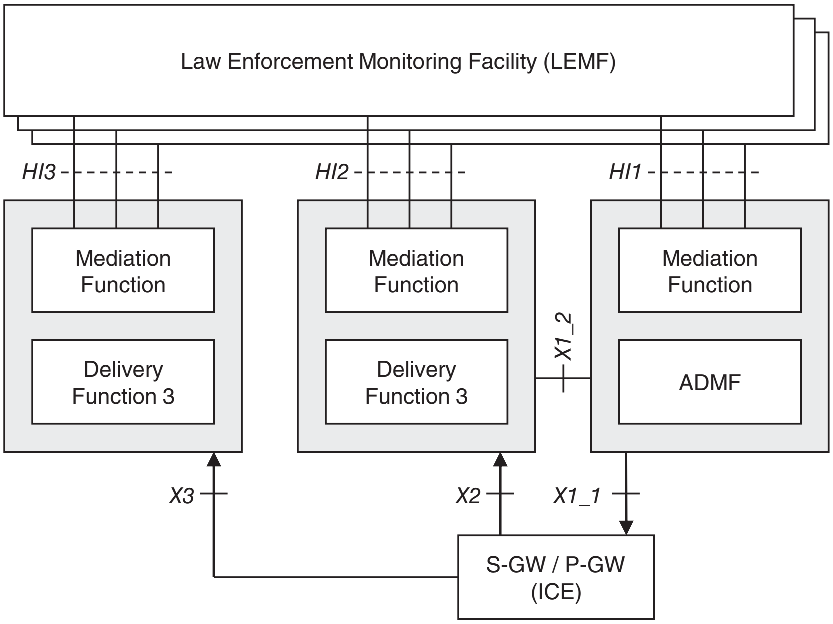

Figure 19.22 *The configuration for the S-GW and P-GW intercept.*

PDN-GW elements. In the figures, the Administration Function (ADMF) is functionality that interfaces with the Law Enforcement Monitoring Facilities (LEMF) of the Law Enforcement Agencies (LEA) that request the interception. The ADMF functionality has been designed in such a way that it has a direct interface with the network elements that are intercepted whilst it keeps the interception related activities of each LEA separated from each others. This is done in such a way that ADMF, together with the delivery functions of the intercepted information, is hided from the Intercepting Control Element (ICE), even in the case of various simultaneous activations on behalf of separate LEAs related to the same subscription.

The physical ICE of the LTE/SAE network is connected to ADMF via an X1_1 interface. This interface delivers all the interception related information from each ICE. Each ICE carries out the interception, that is, the activation, deactivation, interrogation and invocation procedures independently. On the other side of the ADMF, an HI1 interface is defined towards the requester of the lawful interception. For the communication between independent delivery functions and LEA, there are HI2 and HI3 interfaces defined. The task of the delivery functions is to distribute the Intercept Related Information and Content of Communication (CC) to the relevant Law Enforcement Agencies.

Some examples when Legal Interception is activated are the following:

– There is a change in the location information of the subscriber.
– A terminating or originating short message transfer is being initiated by the target.
– A terminating or originating circuit switched call is being initiated by the target.
– A terminating or originating packet data service is initiated by the target.

The content of communications (CoC) can be intercepted via the Legal Interception concept, intercepted from the media plane entities. In addition, various identities related to the intercepted communications can be stored. Some examples of the intercept related information (IRI) that can be intercepted from the subscribers are the following:

– MSISDN (Mobile Sunscriber ISDN Number).
– IMSI (International Mobile Station Identity).
– ME id (Mobile Equipment Identifier).
– Event type.
– Event time and date.
– NE Id (Network Element Identifier).
– Location.

References

1. Federal Communications Commission (FCC). Examples of fraud cases. http://www.fcc.gov/encyclopedia/consumer-publications-library#T-fraud (last accessed 3 August 2014).
2. Federal Communications Commission (FCC). Examples of fraud cases for mobile phones. http://www.fcc.gov/guides/cell-phone-fraud (last accessed 3 August 2014.
3. Federal Communications Commission (FCC). Telecommunications Act of 1996. http://transition.fcc.gov/telecom.html (last accessed 3 August 2014).
4. Penttinen, J. (2011) *The LTE/SAE Deployment Handbook*, John Wiley & Sons, Ltd.
5. Forsberg, D., Horn, G., Moeller, W. and Niemi, V. (2010) *LTE Security*, John Wiley & Sons, Ltd, p. 284.

20
Planning of 2G Networks

Jyrki T. J. Penttinen

20.1 General Planning Guidelines for Fixed Networks

20.1.1 General

The dimensioning of the telecommunications networks is one of the principal tasks of operators in order to guarantee adequate quality level for the end-users. The dimensioning is not a task to be done only once but high quality networks require constant optimization and tuning of parameters.

The main division of network planning can be done between fixed and mobile communications. Many of the tasks are actually common in both wired environment and radio interface as some of the most important high-level principles are to offer sufficient capacity with acceptable QoS (Quality of Service), or GoS (Grade of Service). The terminology of the planning items and parameters may vary depending on the network in question, but the basic principles are often similar.

Some of the common topics to be taken into account in network dimensioning are the following: Estimated need for offered capacity. At the most detailed level, this item may be planned by taking into account daily, weekly, yearly and seasonal variations in the capacity demand. As a basic rule, networks should not be overdimensioned as the business model is impacted greatly due to too high-quality deployment. The fact is that the peak-hour consumes far more resources than off-peak hours, and in average there is considerable waste of capacity. Certain blocking level should thus be accepted by the operator as well as users.

The end-to-end chain of transmission consists of many different elements, interfaces, subnetworks, so it is important to identify potential bottlenecks and minimize impacts so that all interfaces are in balance as for throughput and service levels. Essential ways to plan the network and subnetworks is to understand the theoretical performance, and to verify performance regularly via network operation and management systems as well as via practical field tests by taking representative samples in varying network conditions, topologies, and so on. It is important also to understand the network utilization per user profile types as, for example, the behavior of voice call peak-hour might be very different from data peak hours. Network traffic is typically a mix of various profiles which are weighted as a function of time and geographical area. This means, for

The Telecommunications Handbook: Engineering Guidelines for Fixed, Mobile and Satellite Systems, First Edition.
Edited by Jyrki T. J. Penttinen.

example, that business use profile is weighted at city centers during working hours whilst residential use profiles are strengthened in suburban areas during the evening.

One useful way to estimate traffic profiles and their impacts on network performance are simulations. These might be based on best-guess parameter settings, or on previous experiences of traffic behavior. The outcome of these simulations is of the utmost benefit for correct network parameter tuning as well as for selection of optimal set of functionalities on the network. It should be noted that functionalities might have interdependencies in such a way that the final optimal case is either a result of combined functionality, or sometimes only one functionality without combination results in best performance. As this is extremely hard to estimate without seeing the behavior either in practice or by simulating, it is good method to investigate the effect via simulations or smaller-scale trial or pilot networks prior to heavier investments. The technoeconomical optimization becomes a more and more important task in all telecommunications networks. This chapter presents the effect of the parameters of the network dimensioning by showing that even minor-looking items might have major impact on the total cost per achieved capacity and service level, which means that it is a good idea to investigate the impact analysis as carefully as possible prior to actual network deployment. At the same time, this chapter presents some typical simulation principles and case results as examples for emphasizing the benefits of this task.

One of the modern methods currently under active investigation and development are the self-optimizing networks (SON). This area covers many technologies, functionalities and methods, and can be considered as an umbrella that combines various known and future techniques in order to make the network adjust automatically the parameters for optimal performance, capacity or as a function of some other variable. SON has been developed strongly to 2G, 3G and 4G networks which eases the network planning and optimization task.

This chapter also gives practical information about the experience of nominal and detailed network planning for wireless environments as well as for rollout strategies. Especially for mobile communications networks, there are some extremely important tasks like site hunting, preparation of site, signaling and communication line preparation, road preparation, tower construction and so on that are not directly related to technical and theoretical planning but are very important to take into account in the options of the plans. As an example, quite often it is not possible to obtain optimal location of towers due to restrictions that are hard to estimate in advance, like special local rules for antenna installations and so on. The contents of this chapter are based on the established radio planning theories and practices presented in [1]–[30], including GSM and GPPRS RF performance evaluations of ETSI standardization, practical measurements and theoretical simulations.

20.1.2 Planning of the Networks

The main goal of the dimensioning of fixed telecommunications networks is related to the offered capacity. This item includes preparation for the number of available channels for communications as well as reserve of subscription numbers. On the other hand, network planning should take into account the placement of network elements as well as alternative routing of connections in case of service breakdowns. This is critical for voice traffic that is handled via old-fashioned circuit switched technologies like PCM/TDMA, as well as in any modern network which is based on certain maximum amount of available capacity. The issue is equally important in the packet switched environment, although by the nature of the technology, the rerouting of the traffic is more automated via, for example, the router infrastructure of IP subnetworks. The IP network optimization is thus more related to dimensioning of sufficient capacity and assurance of correct network element functionality.

One of the issues in circuit switched telecommunications infrastructure is the interference level of wired connections like paired cobber lines. The issue increases as a function of distance as parallel lines tend to cause inducted interferences which might trigger problems both in older analog connections as well as in the digital environment. The role of fiber optics increases rapidly in global level which lowers this problem.

If the core network transmission has components based on radio links, for example, via satellites or repeaters, it is equally important to plan the core network based on the same rules as in the RF interface of mobile communications networks, to avoid, for example, cochannel interferences. Furthermore, the service breakdown rate must be dimensioned well enough to comply with all requirements of the operator, regulator and in general the global rules for minimum quality levels for the mean service outage times.

For cellular networks, the radio network includes, in addition to the capacity and quality of service common to the fixed networks, also radio coverage, frequency and radio parameter planning, which can further be divided into outdoor and indoor planning. Typically, the outdoor environment is divided into clusters for coverage planning, like dense urban, urban, suburban, rural and open areas. Each of these area types behaves differently as for radio propagation and thus the achieved cell ranges. Regardless of the technology, it is possible to deploy the radio network by utilizing a single layer approach, or a multilayer strategy. The latter may consist of umbrella cells covering large areas but with only basic capacity whilst the underlying cells can be utilized more locally to take care of the higher capacity. The multilayer structure can consist, for example, of two or three layers in practice, which can also be offered via different technologies. As an example, the basic capacity can be offered via macro cells of GSM 900 and GSM 1800 for large coverage areas for voice calls, and when more capacity is needed for wireless Internet access, it can be offered via smaller high-frequency 3G micro cells, and highest data rate users can be served very locally via pico cells of LTE. It should be noted that the interradio technology for the access (RAT) can be utilized for the seamless handovers more fluently than before, and thus, for example, handover between mobile communications networks and Wi-Fi is possible where functionality is activated and Wi-Fi hotspot is available. Figure 20.1 clarifies the idea of multilayer networks.

Figure 20.2 shows the essential steps of the 2G radio network planning process. The plan can be divided into a nominal plan that gives a rough estimate for the offered capacity in a given coverage area, and into detailed level planning which further clarifies the final behavior of offered traffic as terms of, for example, data throughput and quality level as a function of the number of users that each have certain user profile (a mix of voice and varying data services and SMS and MMS traffic). The basic procedure can be initiated via the capacity plan which follows by the coverage plan. Capacity and coverage are basically tight together regardless of technology, that is, when more capacity is offered, it can be done via loosening the code rate which in turn reduces the useful cell radius.

In case of data networks, there are many special features that need to be taken into account in the radio and core network planning. As an example, the IP numbering plan for subnetworks must be done by estimating

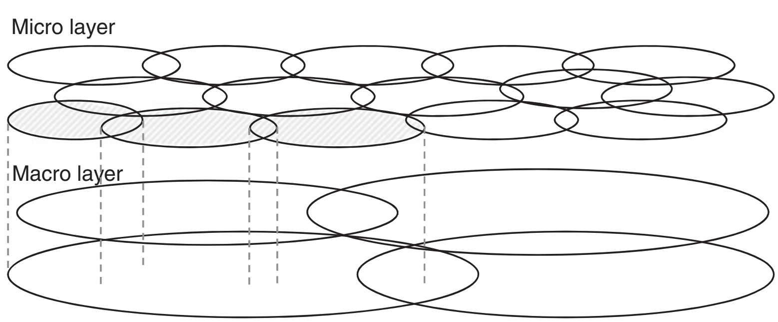

Figure 20.1 *The idea of multilayer network planning. The micro layer offers capacity in the densest parts whilst macro layer provides large coverage.*

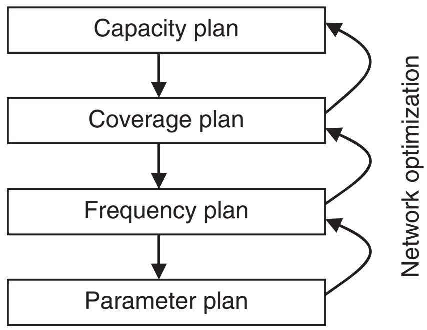

Figure 20.2 *The main steps of the GSM radio network planning.*

the current and future usage. Also the physical aspects of the LAN networks, to mention another example, must be taken into account as each LAN variant has limitations to, for example, segment length of the cables.

In general, it is highly recommendable to take into account the future development of the customer base as well as possible via practical estimates. This is to guarantee the best possible service level along with future network infrastructure investments, and the optimization between the new purchases of the equipment. In the most accurate and detailed estimates, some essential tools for the planning are thus the forecasts of population growth and economical categories of households.

20.2 Capacity Planning

Both fixed and radio networks can deliver certain maximum amount of connections for end-users. Even in networks based on varying capacity along with simultaneous users, there is a practical limit and there are thus no networks that could offer service without any blocking whatsoever. Operators must thus decide the level of offered capacity that is optimal for both end-users (sufficient average quality level for the connections, and reasonable outage probability) and for operators (costs of investments, and reasonable return of investment).

In capacity planning, the need for physical network equipment must be estimated for each geographical location of network access points. Depending on network type, this means that the estimate for the number of routers, bridges, mobile services switching centers, base station controllers, base stations and so on must be known and analyzed for the capacity offer. The essential sources of information for complying with the requirements area also found in the specifications of the systems and equipment, as well as the quality criteria of the regulators and operators. In the radio links and mobile communications network cells, the radio propagation models serve as a good estimation tool for achieved coverage with a planned capacity offering. The topology of the surrounding area must thus be estimated as accurately as reasonable, for example, via 3D models of the geography in the most accurate cases. As the more advanced and accurately modeled maps are more expensive than basic 2D-maps, optimal network planning also takes into account this expense and the task of the operator is thus to decide how accurate predictions are needed in each cluster. As a rule of thumb, the densest urban areas are worth designing with 3D models whilst the planning of suburban areas typically would not benefit considerably from the advanced methods.

The high-quality network planning also includes analysis for future expansions. It is wise to do the planning several years beforehand although the exact subscription numbers would be hard to estimate. The plan would include the estimate for the need of additional functionalities, elements and capacity resources. A good tool for the plan is the user forecasts which convert more accurate during the time.

In case of the circuit switched networks like PCM and TDM based technologies that deliver voice service the most important planning item is to estimate the traffic and thus the amount of the offered channels. This item requires understanding about the average reservation time per channel. On the other hand, the load of the network and thus the service level must be understood. The load of the channels varies greatly along the hour of the day, and there are typically many profiles found along the week, year, and season. There are also occasional special events in many locations that may require special attention from the operator side. The solution for these type of days is to plan in advance additional capacity for the location by, for example, providing additional cells under existing base stations, or additional complete base stations that are vehicle mounted with additional transmission capacity via fixed networks (fiber optics nearby) or via radio links or satellite links.

In the normal case, the offered load is typically dimensioned by estimating the busy hour of the location. The definition of the busy hour varies slightly, as it can be, for example, a single, most loaded hour of the day, or it can be a combination of four nonconsecutive quarters of hour. In any case, the task is to find an estimated moment of the day when the network is busiest.

In each case, the most practical tool for the capacity dimensioning has traditionally been Erlang B which is meant for the systems without queuing. For the systems capable of queuing (e.g., during 10 seconds in the call attempt), there is an enhanced variant, Erlang C. Nevertheless, the basic Erlang B is widely utilized in mobile communications systems as well as in fixed voice networks to estimate the final capacity of the network when a certain blocking probability is designed.

Let's take an example of GSM radio interface. It can be assumed that GSM networks are dimensioned according to 2–3% blocking probability during the peak hour for voice calls. The cells can contain several tens of percents of free time slots for GPRS usage, depending on the total number of TRXs (transceiver units) even during these peak periods. This result is directly derived from the nature of Erlang B formula; in order to offer the service with a certain blocking probability, there has to be, on average, a certain amount of free channels. In order to achieve relatively fluent service, the blocking cannot exceed too many percentage points.

When the GSM network is dimensioned using a certain blocking probability, the transferred traffic can be calculated using Erlang B as shown in Formula 20.1 and Figure 20.3 The Erlang B formula can be considered as a starting point for estimating the traffic and calculating the required channels and the number of TRXs. The formula gives the blocking probability both in the time domain as well as for the number of attempted calls. In other words, the formula gives the probability of the situation, where all channels are occupied:

$$B(N) = \frac{\dfrac{A^N}{N!}}{\displaystyle\sum_{n=0}^{N} \frac{A^n}{n!}}, \tag{20.1}$$

where N is the total number of traffic channels, and A is the product of the average call density and the average time of occupation, that is, A is the offered load. The formula can also be modified to a recursive form, more suitable for computer calculations:

$$\begin{aligned} B(0) &= 1 \\ B(N) &= \frac{AB(N-1)}{N + AB(N-1)}. \end{aligned} \tag{20.2}$$

The basic rule of thumb for the dimensioning of the GSM network is call blocking during the peak hour. Blocking can thus be easily measured from existing networks by dividing the number of blocked calls by

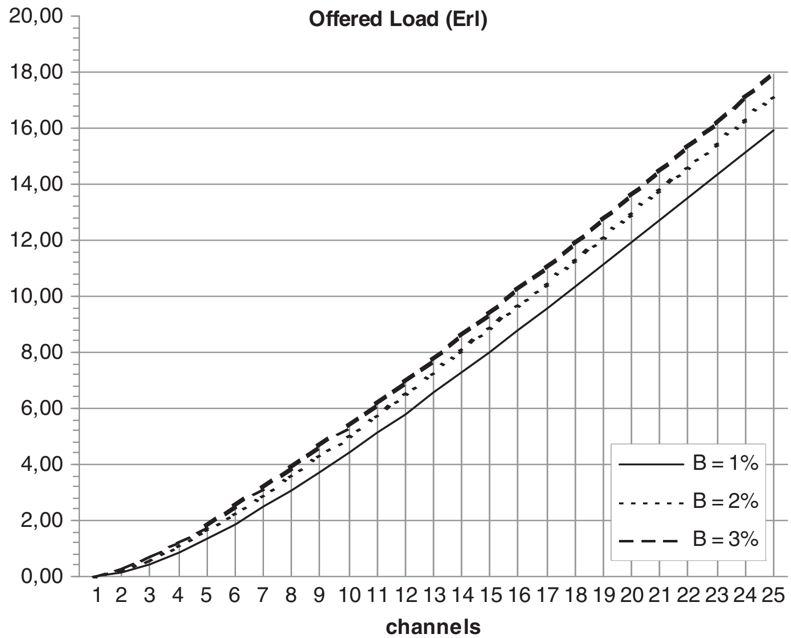

Figure 20.3 *Offered traffic using the Erlang B formula when blocking probability of 1, 2 and 3% are investigated.*

the total number of attempted calls. The normal dimensioning value might be for example, 2–3% during the heaviest period of the network. The served traffic $\bar{x}$ can be calculated with the formula:

$$\bar{x} = A(1 - B), \tag{20.3}$$

and the blocked traffic can thus be calculated as follows:

$$m = A - \bar{x} = AB. \tag{20.4}$$

Figure 20.3 presents a trunking gain effect of Erlang B, that is, the more the available capacity is reserved, the more efficiently it can be utilized which decreases the average blocking probability. It can be noted that Erlang B reveals the trunking effect that also happens in practice. It refers to higher rise of useful capacity along with a higher amount of offered channels, that is, the higher the number of channels, the more efficiently the capacity can be utilized.

The signaling traffic and its dimensioning depend on the complete picture, including user signaling as well as internal network signaling. As an example, in areas where heavy location area updates are present, the internal signaling load also increases accordingly. In these areas, average dimensioning guideline might not be enough but additional signaling capacity is needed. The need for the capacity increase must be verified case-by-case by network statistics and field measurements.

Along with the packet switching GPRS, new aspects are also needed to be taken into account for the traditional circuit switched dimensioning principles. In general, GPRS utilizes by default the “leftover” timeslots which do not interfere directly with voice call capacity, except if the GPRS reserves also part of the timeslots in a fixed manner (dedicated GPRS TSLs) which guarantees better minimum capacity for the customers.

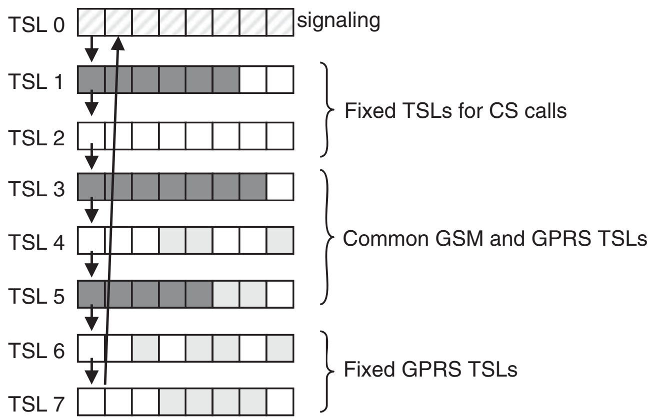

Figure 20.4 *GPRS may have fixed timeslots in order to guarantee minimum capacity to be multiplexed amongst the simultaneous users. Dark TSLs represent voice traffic whilst light grey TSLs indicate the presence of GPRS users.*

There is also the possibility to define maximum number of TSLs for GPRS which means that the available capacity from the circuit switched voice and data calls is changing dynamically for the GPRS users within this region. Figures 20.4 and 20.5 present examples that clarify functionality.

The rest of the capacity is always reserved thus for circuit switched calls. Depending on the manufacturer's algorithms, there is also the possibility to define the whole TRX for only GPRS users in a fixed manner.

It is thus an important optimization task to reserve sufficiently capacity for GPRS users but without affecting too much the voice user blocking probability.

20.3 Coverage Planning

20.3.1 Link Budget

The aim of radio network planning is to provide with sufficient coverage area for the end-users in such a way that the planned capacity can be offered according to certain average quality criteria. The task is one of the most important optimization activities of the operator because "too good" coverage wastes resources, and might turn the business case negative as each base station has an initial cost impact (CAPEX) as well as longer term costs (OPEX).

Figure 20.5 *An example of the GPRS throughput measurement in the radio interface. In this case, the drop of the throughput can be seen clearly when the voice call activity increases, and the multislots for GPRS user drops accordingly from 3 to 2 TSLs.*

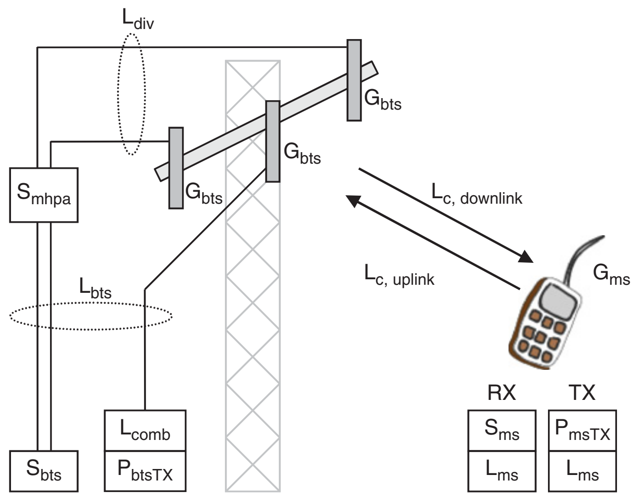

Figure 20.6 *The principle of the GSM radio link budget of GSM. The uplink and downlink are balanced by utilizing the parameters seen in the figure.*

Typically the limiting factor of the link budget is the uplink direction due to the low power level of the device. It is thus important to take care of the correct balance of the uplink and downlink so that there will be no undesired situations where device can receive the base station signaling much farther than it is actually able to transmit to the network.

The GSM radio link budget parameters as presented in Figure 20.6 include P_{btsTX} which is the base station transmitter power in the antenna connector whilst L_{comb} is the loss of filters, L_{bts} is the loss of antenna feeders, and G_{bts} is the antenna gain of the base station antenna. In the reception side of the terminal, G_{ms} is the antenna gain of the device, L_{ms} is the cabling loss of the device, S_{ms} is the sensitivity of the device, P_{msTX} is the outpout power of the device, G_{div} is the antenna diversity gain of the cell antennas, S_{bts} is the sensitivity of the cell, and S_{mhpa} is the potential mast head amplifier's (MHA) gain. It should be noted that the MHA also increases the noise level accordingly, so the practical maximum gain that is used with this element is merely to compensate the cable loss value.

Table 20.1 presents an example of typical GSM 900 link budget for both uplink and downlink in three different environments. The task is thus to balance the uplink and downlink directions.

It should be noted that in the more detailed link budget estimate as presented in Table 20.2, other aspects may also be taken into account such as body loss, fading margin,[1] slow fading margin,[2] possible indoor building loss (which can be on average 20−30 dB if no indoor BTS or repeater is available) and shadowing margin in the cases when the 1st Fresnel zone includes obstacles (causing typically 0−6 dB loss).

The planning value for the received power level at the street level is thus $P_{btstxpeak} - L_{c\_max}$, that is, in this example, the city macro cell has (45–145) dB = −100 dBm, micro cell of the city has (43–137) dB = −94 dB, and rural cell has (51–145) dB = −94 dB. In the case of indoor planning, also the average indoor attenuation value should be added into these values.

The next step for the actual estimate of the coverage areas is to apply radio propagation models. Some typical and useful models are Okumura-Hata (for over 1 km cell radius) or Walfisch-Ikegami (for micro cells).

---

[1] The overall loss due to Rayleigh fading; in GSM, Rayleigh fading effect has been taken into account in the planning values.

[2] Normal distributed confidence level change 50% → 75 %, which corresponds to the success rate of the calls per location of 90%.

Table 20.1 *Example of GSM 900 link budget*

Element	Link budget	Parameter	City, macro cell	City, micro cell	Rural
	Downlink				
BTS	RF power before combiner	$+P_{btsTX}$, dBm	41	39	41
	Combiner	$-L_{comb}$, dB	4	4	4
	Cable and connector loss	$-L_{bts}$, dB	3	3	3
	Antenna gain	$+G_{bts}$, dBi	11	11	17
	= EIRP of the BTS	$P_{btsTXpeak}$, dBm	45	43	51
MS	Sensitivity	$-S_{ms}$, dBm	−102	−102	−102
	Antenna gain	$+G_{ms}$, dBi	0	0	0
	Cable and connector loss	$-L_{ms}$, dB	0	0	0
	Radio path loss	$L_{c\_downlink}$, dB	147	145	153
	Uplink				
MS	Input power	$+P_{msTX}$, dBm	33	33	33
	Cable and connector loss	$-L_{ms}$, dB	0	0	0
	Antenna gain	$+G_{ms}$, dBi	0	0	0
BTS	Sensitivity	$-S_{bts}$, dBm	−104	−104	−104
	Antenna gain	$+G_{bts}$, dBi	11	11	17
	Cable and connector loss	$-L_{bts}$, dB	3	3	3
	Diversity gain	$+G_{div}$, dB	2	0	2
	Mast Head Amplifier gain	$+S_{mhpa}$, dB	0	0	0
	Radio path loss	L_{c_uplink}, dB	147	145	153

20.3.2 Radio Wave Propagation Models

In the simplest case, the theoretical cell coverage area can be estimated via geometry:

$$S = kR^2, \tag{20.5}$$

where R is the radius of the cell, and k depends on cell configuration as presented in Figure 20.7. In practice, major part pf the cell configurations fall into these three main categories, that is, 1, 2 or 3-sector cells.

In the practical radio network planning, more detailed and realistic models are needed because the error margin of such a rough estimate might cause extensive costs due to the overdimensioning of the capacity, or on the other hand, customer complaints due to too optimistic network dimensioning.

Table 20.2 *Some of the more detailed link budget items*

Link budget, additinal parameters	Unit	City, macro cell	City, micro cell	Rural
Body loss	dB	3	3	3
Fading margin	dB	0	0	0
Slow fading margin	dB	5	5	5
Indoor attenuation	dB	0	0	0
Shadowing margin	dB	0	0	0
Maximum path loss margin	dB	139	137	145

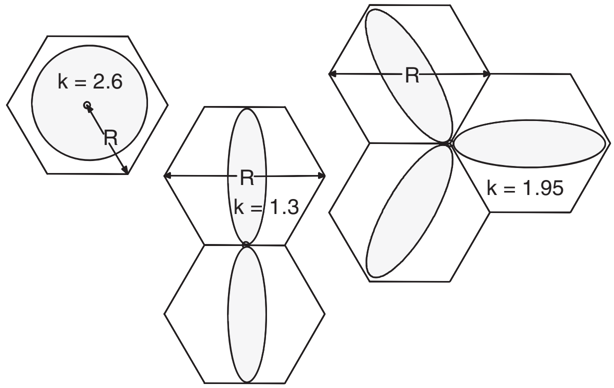

Figure 20.7 *The dependency of the k parameter value as a function of cell configuration.*

One of the most widely utilized radio propagation estimate models is empirically developed Okumura-Hata which is parameterized into different main environments of the practical radio networks. The base for the formula is:

$$L(dB) = 69.55 + 26.16\,\mathrm{lg}(f) - 13.82\,\mathrm{lg}(h_{BS}) - a(h_{MS}) + [44.9 - 6.55\,\mathrm{lg}(h_{BS})]\,\mathrm{lg}(d), \tag{20.6}$$

where f is frequency in MHz (150–1500 MHz), h_{BS} is the effective height of base station antenna in meters (30–200 m) and d is the distance for the successful connections (1–20 km). $a(h_{MS})$ is a correction factor which depends on the environment type. In the *medium and small city* the parameter comes as:

$$a(h_{MS})_{M-S} = [1.1\ \mathrm{lg}(f) - 0.7]\,h_{MS} - \left[1.56\ \mathrm{lg}(f) - 0.8\right], \tag{20.7}$$

where h_{MS} is the height of the mobile device antenna in meters (1–10 m). In *large city,* the correction factor is, for frequencies between 150 MHz and 1500 MHz:

$$\begin{array}{ll} a(h_{MS})_{LC1} = 8.29\left[\mathrm{lg}(1.54h_{MS})\right]^2 - 1.10, & \mathrm{f} \leq 200\ \mathrm{MHz} \\ a(h_m)_{LC2} = 3.2\left[\mathrm{lg}(11.75h_{MS})\right]^2 - 4.97, & \mathrm{f} \geq 400\ \mathrm{MHz} \end{array}. \tag{20.8}$$

In *suburban*, the correction factor is:

$$a(h_{MS})_{SU} = a(h_{MS})_{M-S} + 2\left[\mathrm{lg}\left(\frac{f}{28}\right)\right]^2 + 5.4. \tag{20.9}$$

And finally in *open area* the factor is:

$$a(h_{MS})_{OA} = a(h_{MS})_{M-S} + 4.78\left[\mathrm{lg}(f)\right]^2 - 18.33\,\mathrm{lg}(f) + 40.94. \tag{20.10}$$

Okumura-Hata is suitable, for example, for GSM 900 networks, For the systems that are based on small cell sizes, as is often the case for high band GSM 1800/1900 and 2.1 HGz systems, there are more suitable models like COST 231Hata (1807 MHz). Basically, Okumura-Hata is not accurate for cases where the cell radius is less than 1 km. COST 231 Walfisch-Ikegami is designed thus for these environments. The model assumes that the base station antenna is above average rooftop height, yet below maximum height.

In 900 MHz frequency range, the path loss with typical assumptions and in LOS (Line of Sight) is:

$$L(dB) = 132.8 + 38\,\log(d/km) \tag{20.11}$$

Table 20.3 Some examples of GSM cell range

	City, macro cell	City, micro cell	Rural
Maximum path loss/dB	139	137	145
Propagation model	O-H	W-I	O-H
Cell range/km	2.3	1.3	21.8

and in 1800 MHz frequency:

$$L(dB) = 142.9 + 38\log(d/km)\,\text{medium size city} \tag{20.12}$$

$$L(dB) = 145.3 + 38\log(d/km)\,\text{large city.} \tag{20.13}$$

Table 20.3 shows typical examples of the link budget values based on the previously utilized cases. The outcome of this example is the distance between base station and device antenna per area type. Okumura-Hata (O-H) has been used for the large cells and Walfisch-Ikegami (W-I) for city cells.

In addition to the previously mentioned, there is a variety of other models, from which part is purely theoretical and others, more typically, a combination of theories and empirical test results. One example is dual-ray model which is based on the combination of direct ray accompanied by reflected main ray. The most accurate models use 3D-map as a base for detailed calculations as well as reference measurements from the area in order to calibrate the final outcome.

20.4 Frequency Planning

20.4.1 C/I Ratio

The goal of frequency planning is to increase spectral efficiency in such a way that the same channel interference is kept in a controlled level. As an example, GSM tolerates same channel interference value of C/I_c (Carrier per Interference) of 9 dB. In practice, the planning value needs to be higher than this in order to keep the uncertainty level sufficiently low. For the first neighboring channel protection it is typically sufficient that the neighboring channel is utilized already in the nearest neighboring base station. The second neighboring channel, that is, with one channel between, can be used even in the same base station.

Also GPRS packet channels have the defined maximum tolerable same channel interference limits which depend on the surrounding topology and propagation conditions. In practice, the strongest protected GPRS coding scheme CS-1 functions approximately with the same conditions as GSM voice channels, providing with 9 kb/s data rate per timeslot. The least protected basic GPRS coding scheme CS-4 offers 21.4 kb/s per timeslot in reduced coverage area, and higher order channel coding schemes further reduces the coverage area but offering higher data rates. As an example, CS-4 requires, depending on the situation, minimum 20 dB signal-to interference ratio.

The practical and simple, yet highly useful tool for the nominal frequency planning purposes is hexagonal model. The formula is the following:

$$\frac{C}{I} = \frac{R^{-n}}{\sum_{k=1}^{K_0} D_k^{-n}} \approx \frac{R^{-n}}{6D^{-n}} = \frac{Q^n}{6}, \tag{20.14}$$

where C is the carrier, I is interference, R is the cell range and D the distance between frequency reuse patterns.

The outcome of the hexagonal model is the theoretical C/I value in the cell edge assuming that the attenuation of the radio propagation is purely exponential. The value $K_0 = 6$ indicates the nearest same channel neighboring base station number. As the distance in the second tier of same channel base stations is already relatively far away, the total effect is so low (in range of maximum of 1 dB) that it is not needed to be taken into account. The term Q is the reduction factor of the same channel interference, that is, the higher the value of Q, lower same channel interference is achieved.

The term n is the attenuation factor which depends on the environment. Typical value is in range of 2–5. The attenuation factor can be calculated in the following way:

$$L = L_0 + 10n\log(d), \tag{20.15}$$

where L is the propagation loss (dB), L_0 is the attenuation of the initial point, and d is the distance. In practice, $n = 2$ refers to the free propagation environment (LOS), whilst $n = 5$ means heavily attenuated environment.

The repetition factor of the hexagonal model (Figure 20.8) is K. As the distance is $D = R\sqrt{3K}$, we can obtain the signal-to-interference ratio:

$$\frac{C}{I} = \frac{Q^n}{6} = \frac{\left(\sqrt{3K}\right)^n}{6}. \tag{20.16}$$

Signal-to-interference ratio can also be calculated according to the worst scenario which means that D is $D - R$, and:

$$\frac{C}{I} = \frac{\left(\sqrt{3K} - 1\right)^n}{6}. \tag{20.17}$$

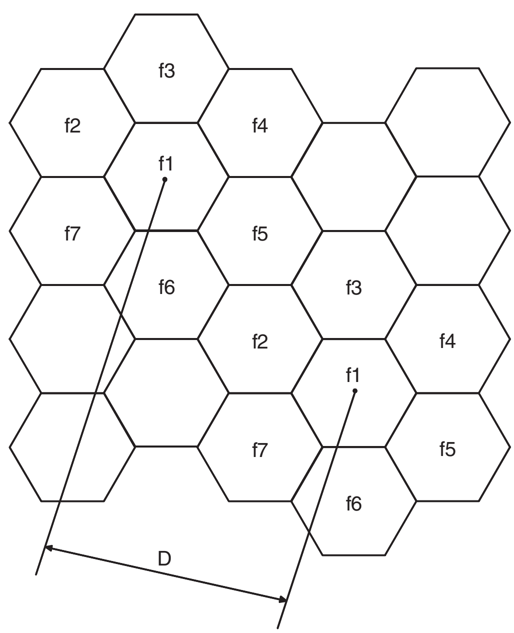

Figure 20.8 *The principle of hexagonal model. D is the distance between two nearest same channel base stations.*

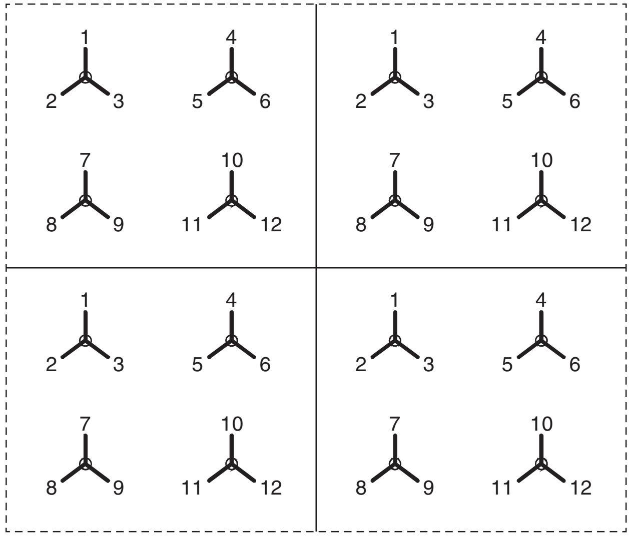

Figure 20.9 *The principle of 4/12 repetition figure. The configuration consists of 4 physical site locations, and 3 sectors (frequencies) in each site.*

This model indicates that, for example, the repetition factor of $K = 7$ and value of $n = 4$ yields C/I ratio in the worst case of $27.4 = 10\log(27.4) = 14$ dB, which in theory is sufficient for guaranteeing sufficient protection for the same channel interference levels. In practice $K = 7$ might be too low in some situations because this model does not take into account the log-normal fading, which can be 6–8 dB, depending on the environment.

Sectorization as presented in Figure 20.9 enhances cellular networks C/I ratio and increases thus the offered capacity. By observing the hexagonal model we can note that 120 degree sectors (with three antenna directions with respectively 3 cells per base station) lowers the number of interfering neighboring cells from 6 to 2. In the further sectorization to a somewhat theoretical 60 degree resolution, there would only be a single interfering cell present.

Same channel interference can occur if the device's transmitter causes too strong signal level in the receiving end (base station or another device). As the interference of the base station is more constant that that of devices the interference analysis is typically done only for downlink.

One practical and widely utilized method for lowering the uncontrolled same channel interferences is downtilting of base station antennas as shown in Figure 20.10. The downtilting can also be done in a reversed way, that is, the antennas are tilted to face slightly upwards as shown in Figure 20.11. Combining the negative mechanical downtilt with the electrical downtilt, the result may be "cleaner" than in pure mechanical downtilt as in the latter case the back lobes are inclined upwards, which in turn increases interferences in the back side of the antenna.

The examples presented previously are applicable for the FDMA and TDMA cellular systems. As an example, the CDMA network tolerates the same channel transmissions because the wanted contents are picked up from the pseudo noise of all the transmissions by correlating the correct code.

20.5 Parameter Planning

It is possible to fine tune the performance and capacity of TDMA/FDMA networks via an already relatively small set of parameters. The most important parameters are related to handover. Also parameters related to

Figure 20.10 *Mechanical downtilt of base station antennas. The downtilt can also be handled electrically which provides possibility to tune the angle remotely.*

Figure 20.11 *Reversed downtilting. This example is from Prague where this type of solution can often be observed.*

power level, frequency hopping and cell selection / reselection are essential to adjust correctly the link budget per environment type. The advanced networks may have hundreds of parameters so before adjusting certain less utilized parameter values it is important to understand also the interdependency of the parameters.

20.6 Network Measurements

The telecom network measurements are essential in all the phases of the network deployment and operation, in order to maintain the planned quality of service level, and to prepare for the moment when future network expansions or frequency reorganization is needed. The load of the networks is one of the most important points to be measured. The measurement can be done via various manners, including radio interface measurements, network signaling and traffic monitoring, upper protocol layer monitoring and via short term and longer periods of network statistics.

The telecom network measurements are based on, for example, network and protocol analyzers which can explore the correctness of the messages, which in turn indicates the bit error rates. Radio interface requires quality monitoring.

Also via the own network elements reveal plenty of useful data for observing the quality, load per service, erroneous packet distribution, and other in-depth information. Typically this kind of measurements is done only by infrastructure owners.

There is variety of equipment available for radio interface measurements. Some examples are Nemo Technologies, Agilent, HP, Comarco, Rohde & Schwarz and Anritsu. Typically, the equipment contains GPS which maps location data with the measurement results for the graphical postprocessing and easier analysis. Figure 20.12 shows an example of such measurement equipment.

Figure 20.12 *Comarco radio network measurement equipment connected to basic GSM test phone. The respective scripts are used for the testing, including data and voice call initiation, maintaining and terminating according to wanted time intervals.*

20.7 Effects of Data Services on GSM Planning

GPRS provides a maximum of 8 simultaneous and dynamically variable timeslots for a single user. The function is called the multislot technique, and it is similar in principle to nontransparent HSCSD. Even if the actual value for the number of multislot channels were smaller than the maximum defined value, the average level of interference due to both terminals and base stations can be assumed to grow compared to traditional GSM networks.

The initiation of the GPRS data call is very fast due to the VRRA MAC (Variable Rate Reservation Access, Medium Access Protocol). In practice, the average initiation time for the data call establishment might be some hundreds of milliseconds. For this reason, GPRS can almost immediately utilize the free time slots of the TRX. The initial signaling of a normal voice or circuit switched data call typically lasts several seconds. During that time period, any established GPRS call will most probably end because of the "bursty" nature of the packet switched service. Thus, the circuit switched call will not experience considerably higher blocking due to GPRS.

The multislot technique influences the selection of the traffic model. In the traditional GSM networks, voice calls can be modeled using the familiar Erlang B formula, as calls are considered to function without queuing. The Erlang B formula will no longer be valid after the introduction of the packet switched and multislot techniques.

20.7.1 Capacity Planning

20.7.1.1 GPRS Traffic

The operator can define a certain minimum value for the number of fixed GPRS timeslots as shown in Figure 20.13, taking into account the options offered by the equipment manufacturer. Thus the operator can provide a minimum level GPRS service also during the voice traffic peak hour and other traffic load peaks, as circuit switched calls can never enter the fixed GPRS timeslots. On the other hand, the possibility also exists to define a maximum value for the number of GPRS-capable timeslots. This means that the GPRS and circuit switched connections can share the same resources between the minimum and maximum GPRS-capable timeslots, prioritizing circuit switched calls. The rest of the timeslots are reserved only for circuit switched traffic. Figure 12.13 clarifies the idea.

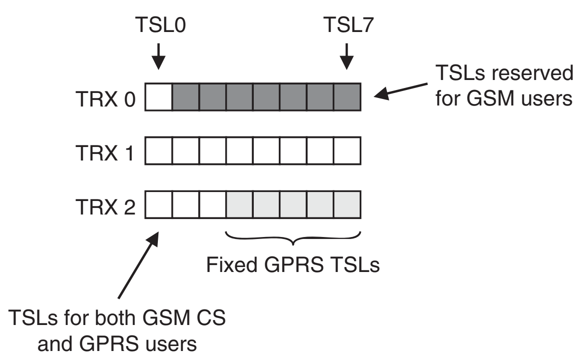

Figure 20.13 *Fixed timeslots can be defined for GPRS users, preventing the circuit switched users from entering those resources. This solution provides a basic minimum GPRS service level.*

The specifications do not define nor restrict the number or existence of the fixed GPRS timeslots, so the equipment manufacturers can freely select their own solutions. Thus an operator might not be able to select freely the minimum or maximum number of GPRS timeslots. If the fixed timeslots can be selected freely, nothing prevents even defining TRXs solely for GPRS use.

The fixed GPRS timeslots decrease the number of available timeslots for the circuit switched calls. Because the Erlang B formula gives a so-called trunking gain, which depends on the offered load, the increase of fixed GPRS timeslots will affect normal GSM traffic, and its already dimensioned blocking probability will decrease more than linearly. In other words, the more efficiently the voice calls utilize the resources, the less capacity will be left for the GPRS connections. The issue is interesting especially when optimizing the GSM network for voice usage using for example, queuing and directed retry functions.

Let's examine a case where a cell has only one TRX. We can assume that one of the timeslots must be reserved for BCCH signaling, which gives us a total of 7 traffic channels for GSM users. Calculating the delivered traffic with Erlang B, assuming 2% blocking probability during the voice traffic peak hour, we receive an average value of 2.94 Erl for the offered load. If we now define two fixed timeslots for GPRS users, 5 timeslots are left for circuit switched users with an offered load of 1.66 Erl. We can also calculate that we could achieve 0.22 Erl offered load with the two resting timeslots, removed from the voice users. We can further calculate that the loss caused by the trunking effect is $(2.94 - (0.22 + 1.66))$ Erl = 1.06 Erl, that is, we would cause the performance of the network to decrease. However, the Erlang B model cannot be used as such for GPRS traffic calculations due to the "bursty" nature of GPRS traffic, and the ability to multiplex various users per timeslot in congested situation (which lowers the data rate per user, respectively). Nevertheless, this simple calculation gives an idea of the effects of GPRS fixed time slots on circuit switched connections.

M/M/N The effect of the GSM users must be taken into account when modeling GPRS traffic. The GPRS network can be considered a data network with variable offered resources. The resources depend on the behavior of the GSM users. This is an interesting problem to examine, since for example, normal local area networks do not change size "on the go" like GPRS.

When utilizing fixed timeslots for GPRS, it could be possible to use, for example, an M/M/N type of model. This means, that while there are an infinite number of queuing resources, only N channels exist for the service. The M/M/N model functions on a "first come, first served" basis. Even if GPRS can serve several simultaneous users within the same timeslot, practical limits exist for the usage of the network, as the system inevitably becomes saturated at some point. ETSI GPRS specification 03.60 defines a maximum of 6–8 users per single GPRS timeslot uplink, depending on channel configurations. This is due to the limitations of the USF (Upper State Flag) parameter.

This standard traffic model does not apply, however, to "bursty" TCP/IP type GPRS traffic. The model assumes a Poisson distribution for the arrival time of calls, and an exponential distribution for the duration of the call. For this reason, the ETSI SMG 2 GPRS AdHoc used better suited models for initial GPRS performance simulations: the FUNET model and the so-called packet train model.

The FUNET Model Theoretical GPRS network studies can be performed by investigating real fixed packet data networks. In this case, it might be hard to estimate the error margin due to fundamental differences between fixed and mobile networks. The utilized application will have a strong effect on the models, as well as on the billing of the services. For example, a fixed ADSL line with flat rate charging is ideal for heavy data stream downloading, even for (near) real time video, within the practical limits of the network data rates. This might not be the case for the GPRS network, where data rates are usually much more limited by the radio interface, and the price of the transferred bits might be charged separately. These are the main reasons why studies of traffic behavior of users in fixed data networks might be quite misleading when adopting the results directly to the mobile environment.

Packet data is “bursty” in nature. GPRS is designed for the packet environment, and for this reason the service occupies the radio resources only when needed. The behavior of GPRS is thus totally different from the circuit switched connections, which reserve network resources until the end of the connection regardless of the real usage of the network. A circuit switched data user is likely to read and write email in offline mode in order to save resources (and, of course, expenses). With GPRS, real time email usage becomes a more logical solution.

The FUNET model has been developed by collecting information on the usage of the Finnish research and university data network, called FUNET. The probability density function in its normal form is:

$$Cauchy(0.8, 1) = f(x) = \frac{1}{\pi\left[1 + (x - 0.8)^2\right]}, \tag{20.18}$$

which is suitable for a maximum of 10 kilo byte packets, that is, the function is defined within the range $0 \leq x \leq 10$ kilo bytes.

The term x represents the size of the packet in kilo bytes. Figure 20.14 is a graphical presentation of the probability density function.

The Packet Train Model Another formula suitable for theoretical GPRS traffic studies is called the Packet Train model, which can be presented in cumulative form:

$$F(X) = P(x \leq M) = 1 - e^{-\frac{x}{170}}, \tag{20.19}$$

where x is the size of the packet in bytes. The function is defined for the range of $0 \leq x \leq 1000$ bytes as presented in Figure 20.15.

The model is suitable for a maximum of 1000 byte-sized packets, that is, for smaller packets than the FUNET is designed for. The model is based on “trains” with certain amounts of cars, representing the data

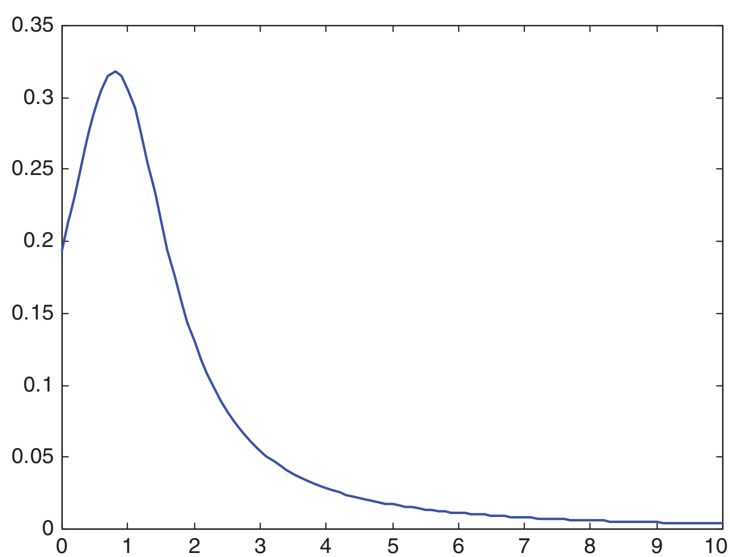

Figure 20.14 *The probability density function (PDF) of the FUNET model. The size of the packet can vary between 0 and 10 kB per second.*

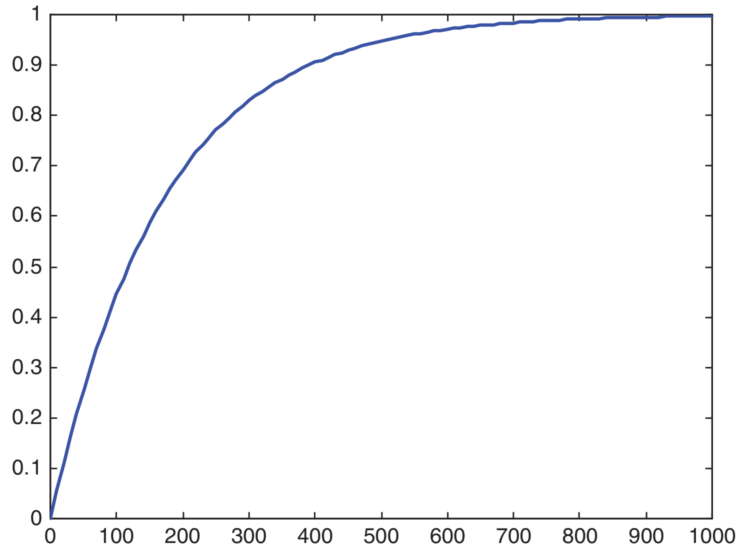

Figure 20.15 *The cumulative distribution of the packet train model. The size of the packets can vary between 0 and 1000 bytes.*

bursts, where the packets of the message follow each other at an almost infinite speed. The arrival time for the messages (trains) is modeled utilizing Poisson distribution. The model expects the average arrival time of the messages to be 170 bytes. Figure 20.15 represents the cumulative distribution function (CDF) of the Packet Train model.

Traffic Profile Measurements Existing GSM networks can be used as a basis for estimating the traffic profiles of new data services. The weight of the future traffic can be estimated separately for voice and data connections, due to the different profiles of each service. The final profiles can naturally be influenced – at least to some degree – by the choices made in billing. Thus if the network operator wants to have most of the data services utilized during the evening time, the balance between voice and data services can be maintained by setting a lower price for data connections in the evenings.

Current traffic profiles can be measured by the accuracy of a single TRX. Figure 20.16 shows typical examples of combined data and voice traffic in rural and urban areas in Finland, measured from the real GSM network. The values of the figure have been normalized in order to allow a comparison of the results.

In both cases, the GSM traffic peak hour seems to hit between 3 and 4 p.m., although the absolute value of the traffic transfer is more than double in the urban area compared to the rural area. The distribution of the rural traffic seems to be more equally spread, whereas the urban traffic decreases rapidly after the peak hour. Using these figures, the operator can estimate the distribution of home and work areas, and thus construct the cells in the right places for the right moments. In the future, automatic capacity load and beam forming procedures could be used to assign capacity for those areas where it is really needed. Nowadays capacity can be added occasionally to for example, areas with lots of summer cottages during the summer time. The “cold” TRXs of the cell are defined functional, and the frequency plan is changed “on the go.” It is also possible to physically add more TRXs by for example, by bringing a vehicle-mounted BTS to the area.

Data traffic cannot necessarily be studied from the networks in much detail. Because the lines to and from the MSC are defined either for voice or data traffic, the data profiles can be examined by the accuracy of

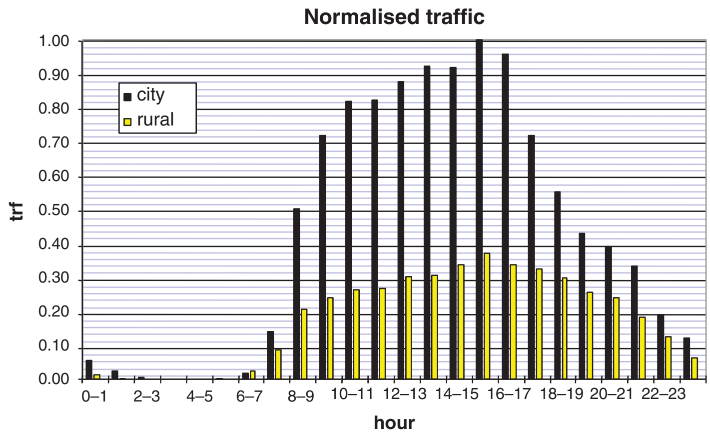

Figure 20.16 *Typical GSM traffic profiles in urban and rural areas.*

MSC site. Figure 20.17 shows an example of a data traffic profile, averaged over a one-month time period. The above-mentioned BTSs are a part of this MSC.

In the example case, voice and data traffic profiles are differentiated by a distinct data peak hour at approximately 10 p.m. The other significant data peak takes place between 11 a.m. and 2 p.m. When transfer directions are compared, it can be seen that MT calls represent only 7% of total data traffic in this specific case. The profile can, of course, vary heavily depending on the area and timing. In any case, this example highlights some of the basic phenomena that should perhaps be taken into account when planning GPRS services.

Data traffic was typically low in the beginning of the GPRS services, approximately 0.5–2% compared to voice traffic. Along with the development of the packet data-capable terminals and new applications, the importance of data traffic has grown significantly. The breaking point for the data growth curve can be found at the time when first smart devices with GPRS were launched. A possible danger is that increased data usage, including GPRS, causes interference to normal voice traffic if networks are not maintained and updated correctly according to data traffic growth.

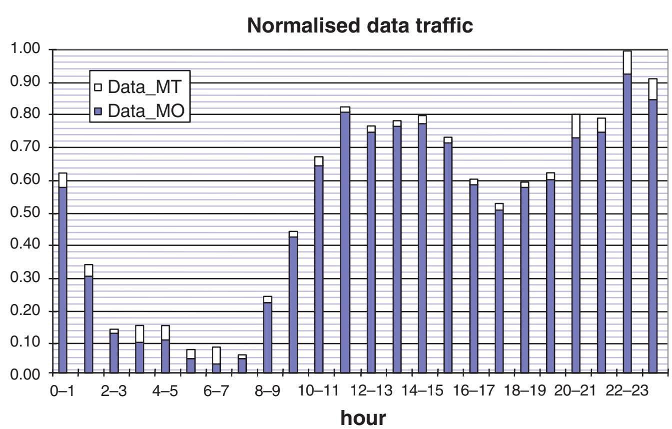

Figure 20.17 *An example of 24-hour GSM data traffic distribution. The mobile originated and mobile terminated directions are shown separately.*

Simulation Studies The reference [1] contains MAC layer simulation results achieved at the University of Aachen, Germany. The results contain estimates for frame transfer delay (FTD), data throughput (S), and blocking rate (B). The FTD stands for the time difference between the sent and received LLC frame. S stands for the amount of successfully transferred data over the air interface. B gives the probability for the failure of the RACH burst sent by the MS.

The simulations are focused on the radio interface. The utilized traffic models were the FUNET model, the packet train model and the Mobitex model. The latter was optional in simulations performed in ETSI SMG 2 GPRS AdHoc group, but was not used by it. The Aachen simulations used the channel coding scheme CS-3, which gives a code rate of about 3/4.

With the FUNET model, approximately 70% throughput compared to the maximum theoretical value was achieved in the simulations. In comparison, a value of approximately 48% was derived when using the packet train model. The below-theoretical values are explained by the fact that in a heavily loaded network with burst collisions, the channel resolution algorithm that takes care of the burst allocation for the following periods, becomes unstable. The low throughput value of the packet train model is a result of the smaller packet sizes, which cause a higher probability for collisions in the situation where an entry to the service is attempted. The blocking rate also depends heavily on the number of Random Access bursts. The maximum transfer delay for the FUNET model is about 2.5 seconds, whereas the delay of the packet train model grows significantly whenever load increases. When interpreting these simulation results, it should be kept in mind that it is likely that CS-3 will not be in use in GPRS networks due to the limitations of the *Abis* interface.

The Italian CSELT has also been active in simulations in the GPRS field, and has for instance carried out RLC/MAC layer simulations during the ETSI SMG 2 GPRS AdHoc work. The reference [2] discusses studies which were based on the simulations carried out on one GPRS time slot, utilizing packet train model. As a result, the throughput, frame access delay and frame transfer delay can be seen as a function of *C/I* with all defined GPRS channel coding schemes. As can be expected, CS-1 provides the most efficient throughput with low *C/I* values. The performance of CS-2 and CS-3 is pretty similar, while CS-4 is suitable only for high *C/I* values. The results indicate that GPRS throughput could be around 22–67% of the theoretical maximum values of 9.05–21.4 kb/s of GPRS traffic, in a single time slot, when the *C/I* value is between 10 and 30 dB.

20.7.1.2 Calculations about Offered GPRS Capacity

The planning of future GSM networks is increasingly an optimization task. Voice, circuit switched data, basic GPRS and highly advanced EDGE2 and Dual Carrier users could be utilizing the same network resources. This is because none of the services can actually act as a substitute for the other services within the near future. Unlike for circuit switched connections, GPRS is intended for variable packet data speeds. The nontransparent HSCSD when supported might partially behave like GPRS; it can utilize dynamic allocation of the resources with upgrading and downgrading functions, which is quite close to the packet service [3, 4].

We can assume that the normal voice peak hour blocking probability is 2%. In this case, using a single TRX cell, one time slot is reserved for signaling purposes, and thus the offered load is 2.94 Erl and delivered traffic is $\bar{x} = (1 - B)A = 2.88$ Erl, according to the Erlang B formula. This means, that the available *average* number of time slots for GPRS would be $7 - 2.88 = 4.12$.

On the other hand, the 2% blocking probability value of the GSM network could be thought to mean that GPRS could be served 98% of the time, at least on one time slot. Similarly, GPRS traffic is totally blocked on average only 72 seconds per hour, due to GSM traffic. As we might imagine, this 72-second period is spread more or less uniformly over the one-hour period, which means that the actual blocking of GPRS is quite insignificant. This applies to reasonable GPRS traffic loads; whenever GPRS traffic is heavy, GPRS users will affect each other's service level (data throughput) negatively, until a proportion of the GPRS users must be rejected due to the lack of resources.

If GPRS utilizes the resources only when they are available GPRS has no effect on the circuit switched traffic load on the traffic channels. Instead, the signaling related to the GPRS call initiation on the BCCH/CCCH time slots could affect GSM signaling, depending on the GPRS channel configuration. Depending on these contradictions, the need might arise to arrange a proper GPRS signaling channel to balance the GSM and GPRS load. A real danger is that the initial signaling of GPRS could become a system bottleneck for both GMS and GPRS users.

Based on the earlier aspects, the average offered GPRS time slots of the cell can be calculated using the formula presented in Refs. [3,4]:

$$N_{GPRS} = N_{tot} - (1 - B)A(N_{tot}), \tag{20.20}$$

where

$$N_{tot} = 8N_{TRX} - N_{sign}, \tag{20.21}$$

and $A(N_{tot})$ is the average number of offered voice channels with a blocking probability B, based on the Erlang B formula. N_{tot} is the physical number of time slots of the cell, N_{TRX} is the number of TRXs in that cell, and N_{sign} is the number of physical circuit and packet switched signaling channels of the cell. Please note that a single GPRS user can utilize multislot channels up to a maximum of 8 time slots, and all of these channels must be allocated on the same TRX.

It can be assumed in these calculations, that one signaling channel exists for one TRX cell, two signaling channels for two TRX cells, and three signaling channels for three and four TRX cells.

The fixed GPRS TSLs reduce the number of available TSLs for GSM users which reduces the trunking gain of the original circuit switched connections. The previous formula can be modified to support the fixed GPRS time slots as follows:

$$N_{GPRS} = \min\left\{ N_{tot} - (1 - B)\,A\left(N_{tot} - \sum_{n=1}^{N_{GPRS\_TRX}} N_{fix-\min,n} \right), \sum_{n=1}^{N_{GPRS\_TRX}} N_{\max\_TRX,n} \right\}, \tag{20.22}$$

where N_{GPRS_TRX} is the number of GPRS TRXs in the cell n, and N_{fix_min} is the number of the fixed GPRS timeslots on TRX n. N_{max_TRX} is the maximum number of time slots that GPRS can utilize in each TRX (this can be defined individually for each TRX). This formula expects the time slots to be filled in an "intelligent" manner, that is, so that there will be no blocking of GPRS, presuming there are free time slots in any of the TRXs. Each vendor defines in practice how the algorithms for allocating GPRS timeslots will function. Basically the system could try to find a block of consecutive GPRS capable timeslots, which optimizes the data throughput for the users. As a method for finding the timeslots efficiently, the voice traffic could be forced to one end of the TRXs, and GPRS in the other end. In any case, the user can have the multislot channels only within the same TRX.

It might be a bit misleading to think about GPRS capacity as time slots. This is because there could be several simultaneous users per single GPRS time slot. The other way of thinking could be to calculate the transferred bits within a certain offered capacity, that is, a kind of "bandwidth". In the familiar case of one TRX with 2% blocking probability, 4.12 timeslots are occupied on average 2% of the time. In an ideal case, using the fastest data transfer coding scheme with 21.4 kb/s, the total data rate on average is thus $4.12 * 21.4$ kb/s = 88.2 kb/s. Using the channel coding scheme CS-1, the average data rate would be $4.12 * 9.05$ kb/s = 37.6 kb/s.

The principle of the GPRS coverage map is presented in Figure 20.18. Each coverage area depends on the *C/I* ratio, which corresponds to the channel coding schemes CS-1–CS-4. The corresponding coverage areas, that is, surfaces (or duration of the usage) are A_{cs1}–A_{cs4}. The traffic distribution within the cell is considered uniform over the whole area or time.

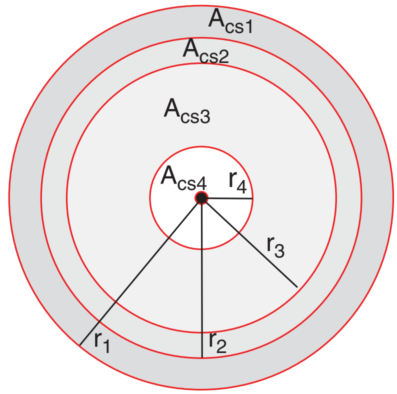

Figure 20.18 *The coverage areas achieved with the channel coding schemes CS-1–CS-4. The corresponding surfaces are A_{cs1}–A_{cs4}.*

Based on the surfaces and the corresponding data rates, we can now obtain a term, which indicates the amount of transferred bits per time period in a certain surface. The data rate naturally depends on the protocol layer, and is different, for example, for LLC, RLC, and user planes. In these calculations we utilize the user data rate derived from the specifications. The parameters p_1–p_4 can be used to express the *usability* of the corresponding channel coding schemes. This can be done by time or by coverage area, compared to the voice service, where A_p is the reference area or time. The average number of the TRXs within the area is T, and the GPRS capable TRX number is T_G. The usability of the GPRS timeslots (*throughput*) is marked as k_G (0–1), and the number of GPRS users per single cell is marked as u_G. The total amount of GPRS cells within one square kilometer is s.

The ideal throughput delivered by a single cell is the sum of the throughputs of each channel coding scheme affecting the corresponding surface. The surfaces or time values are thus:

$$A_{cs4} = p_4 A_p \tag{20.23}$$

and

$$A_{cs(n)} = p_n A_p - A_{cs(n+1)}, \text{ when } n = 3.1. \tag{20.24}$$

The delivered amount of data of the cell is thus:

$$v_{cell}(A_p) = \sum_{n=1}^{4} v_n p_n = \sum_{n=1}^{4} \frac{v_n}{A_{cs(n)}}, \tag{20.25}$$

where v_n is the maximum data throughput of the corresponding channel coding scheme (kb/s). If the number of the cells within one square kilometer is s and the average number of GPRS capable TRXs is T_G, the total maximum (ideal) throughput during the voice traffic busy hour would be:

$$v/km^2 = sT_G v_{cell}(A_p). \tag{20.26}$$

When the number of active GPRS users within the area is u_G, the capacity per user is thus:

$$v/user/km^2 = \frac{sT_G v_{cell}(A_p)}{u_G}. \tag{20.27}$$

Table 20.4 *Some case examples of the allocation of GPRS and circuit switched timeslots on the cell. The maximum number of GPRS capable timeslots, as well as fixed GPRS timeslots is shown in corresponding column for each case*

TRXs	Sign. chs	Traffic chs	GPRS fixed	Case 1 max	GPRS fixed	Case 2 max	GPRS fixed	Case 3 max
1	1	7	0	7	1	7	1	7
2	2	14	0	14	2	14	1	14
3	3	21	0	21	3	21	1	21
4	3	29	0	29	4	29	1	29
5	4	36	0	36	5	36	1	36
6	4	44	0	44	6	44	1	44

Based on the calculations above, we can now create a term G_{d_s}, which indicates the delivered maximum GPRS data compared to the circuit switched data transfer of 9.6 kb/s/timeslot.

$$G_{d\_s} = \frac{v_{cell}(A_p)}{9.6kb/s \cdot A(N)}, \tag{20.28}$$

where $A(N)$ is the offered capacity based on the Erlang B formula, with N traffic channels in each cell. The realistic fine-tuning must be done using a constant k_G, which indicates the real GPRS throughput. The value must be studied, and can be obtained from real networks or from the simulation results.

20.7.1.3 Case Examples

Some possible channel combinations of GPRS and circuit switched traffic exist and are shown in Table 20.4 . Case 1 represents the situation where no fixed GPRS timeslots have been allocated. In this case, in theory, the GPRS service has no effect on normal GSM traffic, as the amount of TRXs is not increased or timeslots are removed from GSM users. The assumption is that GPRS traffic does not significantly increase the signaling load. Case 2 is similar to case 1, but the difference is the allocation of one fixed GPRS time slot per each TRX. In case 3, only one fixed GPRS time slot exists. It should be noted that each GPRS time slot may deliver traffic up to the maximum multiplexed users (e.g., 6–8 in uplink) and with any coding scheme that is possible for the connection. This analysis considers thus the higher level capacity based on overall time slots.

The physical signaling channel configuration is achieved in these calculations according to Figure 20.19. This study is used to figure out the available resources for GPRS, that is, the average amount of timeslots left over from circuit switched calls during the peak hour. The number of available time slots could be called the available GPRS "bandwidth," so as not to confuse it with the normal GSM traffic within the time slots. The bandwidth is available for all GPRS users, and can be shared up to the limits of the simultaneous time slot usage. Please note, that a single user cannot necessarily use all of the available bandwidth only for him/herself, depending on the equipment limitations (terminal and network support of the multislot classes, and the fact that the multislots must be allocated on the same TRX).

Figure 20.19 shows typical channel configurations as a function of the amount of TRX elements. The signaling channels are marked with a grey color (B=BCCH, S=SDCCH) and traffic channels as white. In these calculations, the GPRS TRXs can be selected from the lowest TRX (TRX1) up to the highest (TRX6), or vice versa. The selection order slightly affects capacity results in some cases, since in for example, the 5th or 6th TRX case, TRX1 contains 6 traffic channels whereas TRX5 or TRX6 contains 8 GPRS time slots.

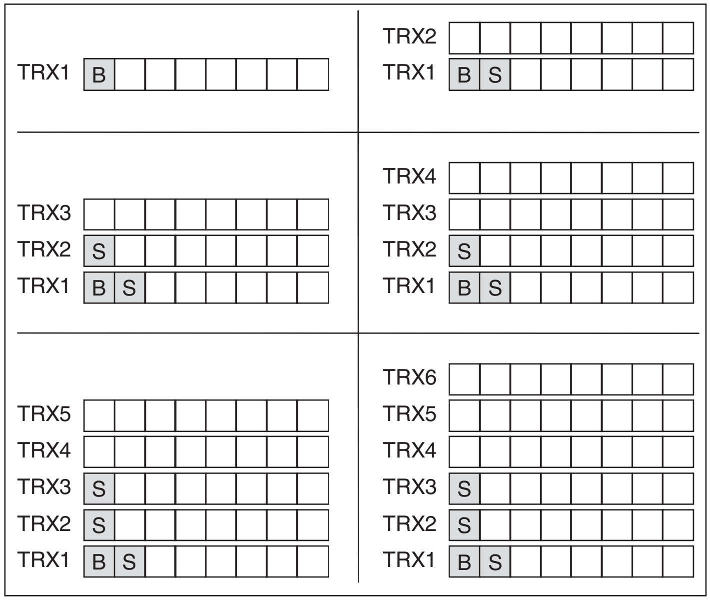

Figure 20.19 *The channel configuration in the case examples for TRX nos. 1–6.*

Case 1: No Fixed GPRS Timeslots Based on the Formula 20.22., no calculations are made in Table 20.4 for the number of available timeslots for GPRS use with different cell structures (TRX numbers). The calculations take into account the combinations of the signaling and traffic channels as shown in Figure 20.19. It is also assumed that the cell's GPRS TRXs are selected beginning from the lowest TRX number, and that the BCCH time slot is always on the first TRX. Thus, for example, in the case of one GPRS TRX, the maximum number of GPRS time slots is 7, because the BCCH is situated on the time slot no. 0.

The results of Table 20.5 show that the use of more than two GPRS TRXs does not give any additional average capacity for GPRS during the voice peak hour with 2% blocking probability. During off-peak traffic the situation is, of course, different, and the more offered capacity GPRS has, the more efficiently the GPRS users can use it for packet traffic.

Table 20.5 *The theoretical maximum number of GPRS time slots as calculated by the Formula 20.22. The assumption for the circuit switched traffic blocking is 2%, and the example is case 1. The GPRS TRXs are chosen in order of TRX nos. 1 -> 6*

GPRS-TRXs	TRXs (tot)					
Total	1	2	3	4	5	6
1	4.12	5.96	6.0	6.0	6.0	6.0
2	—	5.96	7.28	8.42	9.25	10.0
3	—	—	7.28	8.42	9.25	10.0
4	—	—	—	8.42	9.25	10.0
5	—	—	—	—	9.25	10.0
6	—	—	—	—	—	10.0

Table 20.6 *The theoretical capacity for the GPRS service with a circuit switched traffic blocking rate of 2% (case 1), when GPRS TRXs are selected by the TRX order of 6->1*

GPRS-TRXs	TRXs (tot)					
Total	1	2	3	4	5	6
1	4.12	5.96	7.28	8.0	8.0	8.0
2	—	5.96	7.28	8.42	9.25	10.0
3	—	—	7.28	8.42	9.25	10.0
4	—	—	—	8.42	9.25	10.0
5	—	—	—	—	9.25	10.0
6	—	—	—	—	—	10.0

The example below shows a kind of worst case scenario for available GPRS capacity where the multi TRX cell always allocates GPRS users in the BCCH frequency before filling up the other TRXs. This means, that BCCH time slot 0 affects the capacity of GPRS, with possible other signaling time slots. Thus only 6–7 available time slots exist for GPRS use on the BCCH frequency. However, it is possible to allocate the GPRS time slots to any of the TRXs, depending on the manufacturers' solutions. Thus, in the case of a two-TRX cell, the other TRX (non-BCCH TRX) offers the full 8 time slot capacity for the GPRS service. It might be interesting to recalculate the latter example assuming that the GPRS TRXs are selected in reverse order, that is, from the highest TRX number to the BCCH TRX. The result is shown in Table 20.6.

The difference between Tables 20.5 and 20.6 is that the reverse order of TRX allocation for the GPRS is seen in the case of one TRX configuration. The ideal GPRS data rates of both Tables 20.5 and 20.6 with different TRX combinations are shown in Table 20.7. The maximum data rate (i.e. the available average GPRS bandwidth for all users in total) is thus between 37.3 and 214 kb/s.

The data rates of Table 20.7 are calculated assuming that GPRS time slots can be utilized ideally with a 100% load with all GPRS channel coding schemes, and that the circuit switched connections are established mainly on other TRXs than GPRS.

If the possibility exists to allocate priority blocks for both circuit switched and packet switched traffic, there are more means to allocate multislot connections to GPRS. In effect, each GPRS user might have a more efficient connection type in this solution. These algorithms are being developed, and each one is dependent on the equipment manufacturer. The specifications do not define these algorithms, which leads to different solutions in different networks. The calculations presented in this book do not take into account the effect of the algorithms.

If the GPRS time slots are allocated on the BCCH-TRX, there will be no changes in the *C/I* value due to the extra GPRS traffic load in the downlink direction. This is not the case in BCCH uplink or other frequencies uplink and downlink, but the average *C/I* value is lowered by the GPRS traffic. In these cases the network can be assumed to have been constructed to take into account the worst case, that is, 100% load in each frequency, and thus the variation of the *C/I* value is not relevant. Chapter 20.10 presents simulations on the actual effect of the changes of the *C/I* distribution by GPRS load. The subject is very interesting, as increased GPRS traffic seems to cause a situation similar to UMTS cell breathing.

Case 2: One Fixed GPRS TS in Each TRX The results of case 2 are shown in Table 20.8. In this case, each GPRS TRX has one fixed GPRS time slot, whereas the maximum number of available GPRS TSs equals the total number of traffic time slots in each TRX.

Utilizing fixed GPRS time slots causes one problem: the blocking rate of circuit switched traffic rises. The phenomenon can be calculated using the Erlang B formula. The values should be calculated by examining

Table 20.7 *The ideal maximum data rates (kb/s) of the GPRS cells, when using the service among the other circuit switched services, with 2% blocking. The assumption is that GPRS can utilize all of the free time slots, and furthermore, there are no fixed GPRS time slots in use. The result is shown with both the GPRS TRX order 1->6 and 6->1, if they differ from each other. The result is the cell "GPRS bandwidth," that is, the total capacity available for all GPRS subscribers*

GPRS-TRXs Class	TRXs	Total number of TRXs 1	2	3	4	5	6
CS-1	1	37.3	53.9	54.3/65.9	54.3/72.4	54.3/72.4	54.3/72.4
	2	—	53.9	65.9	76.2	83.7	90.5
1 TS = 9.05 kb/s	3	—	—	65.9	76.2	83.7	90.5
	4	—	—	—	76.2	83.7	90.5
	5	—	—	—	—	83.7	90.5
	6	—	—	—	—	—	90.5
CS-2	1	55.2	79.9	80.4/97.6	80.4/107.2	80.4/107.2	80.4/107.2
	2	—	79.9	97.6	112.8	124.0	134.0
1 TS = 13.4 kb/s	3	—	—	97.6	112.8	124.0	134.0
	4	—	—	—	112.8	124.0	134.0
	5	—	—	—	—	124.0	134.0
	6	—	—	—	—	—	134.0
CS-3	1	64.3	93.0	93.6/113.6	93.6/124.8	93.6/124.8	93.6/124.8
	2	—	93.0	113.6	131.4	144.3	156.0
1 TS = 15.6 kb/s	3	—	—	113.6	131.4	144.3	156.0
	4	—	—	—	131.4	144.3	156.0
	5	—	—	—	—	144.3	156.0
	6	—	—	—	—	—	156.0
CS-4	1	88.2	127.5	128.4/155,8	128.4/171,2	128.4/171,2	128.4/171,2
	2	—	127.5	155.8	180.2	198.0	214.0
1 TS = 21.4 kb/s	3	—	—	155.8	180.2	198.0	214.0
	4	—	—	—	180.2	198.0	214.0
	5	—	—	—	—	198.0	214.0
	6	—	—	—	—	—	214.0

Table 20.8 *The theoretical GPRS capacity (time slots) in case 2. The results are shown in GPRS TRX order 1->6/6->1, if they differ from each other*

GPRS-TRXs Total	TRXs (total) 1	2	3	4	5	6
1	4.77	6.0/6.75	6.0/8.0	6.0/8.0	6.0/8.0	6.0/8.0
2	—	7.52	8.95	10.1	11.0	11.9
3	—	—	9.73	11.0	11.9	12.7
4	—	—	—	11.9	12.8	13.6
5	—	—	—	—	13.7	14.5
6	—	—	—	—	—	15.4

Table 20.9 *The effect of the fixed GPRS time slots on the growth of the blocking rate of the cell. For example, in a six-TRX cell, when two fixed GPRS time slot are used, 42 time slots remain for circuit switched connections, causing the original 2% blocking rate to grow to 3.3%*

TRXs	Circuit switched time slots/amount		B of circuit switched traffic, when GPRS-TRX number is:					
Total	Original	New value	1	2	3	4	5	6
1	7	6	4.9	—	—	—	—	—
2	14	12–13	3.5	5.7	—	—	—	—
3	21	18–20	3.0	4.4	6.3	—	—	—
4	29	25–28	2.8	3.8	5.1	6.6	—	—
5	36	31–35	2.7	3.5	4.5	5.7	7.1	—
6	44	38–43	2.6	3.3	4.1	5.1	6.2	7.4

the original offered load of circuit switched traffic, and comparing it with the new situation with fewer circuit switched time slots. For example, in the case of one TRX the original number of available time slots is 7 which leads to an offered load $A = 2.94$. When GPRS receives one fixed time slot, six circuit switched time slots remain. For these six time slots, maintaining the value of A as 2.94, the new blocking rate will be $B = 4.9\%$.

As in case 1, it is possible to examine the reversed order of GPRS TRXs (i.e. selecting the TRXs from the highest number towards the lowest one). These results are also shown in Table 20.8, if they differ from the former calculations.

It can be seen from Table 20.9 that if the acceptable circuit switched blocking rate limit should be 3%, the only valid combination is fulfilled when using one fixed GPRS time slot over two TRX cells. The resulting theoretical data rates range between 54.3 and 128.4 kb/s (GPRS-TRX order 1→6) or 72.4–171.2 kb/s (GPRS-TRX order 6→1). The benefit of using fixed GPRS time slots is securing a certain minimum level of GPRS capacity, even during the circuit switched traffic peak hour.

Case 3: One Fixed GPRS TS Case 3 is otherwise the same as Case 2, but in this case there is only one GPRS time slot defined, regardless of the total number of TRXs. GPRS capacity is shown in Table 20.10 and 20.11.

Table 20.10 *The theoretical GPRS capacity (time slots) in case 3, when using the circuit switched traffic blocking rates according to the previous example. The results are shown in GPRS-TRX order 1->6 / 6->1, if they differ from each other*

GPRS-TRXs	TRXs (total)					
Total	1	2	3	4	5	6
1	4.77	6.0/6.75	6.0/8.0	6.0/8.0	6.0/8.0	6.0/8.0
2	—	6.75	8.06	9.20	10.1	10.9
3	—	—	8.06	9.20	10.1	10.9
4	—	—	—	9.20	10.1	10.9
5	—	—	—	—	10.1	10.9
6	—	—	—	—	—	10.9

Table 20.11 *The effect of the fixed GPRS time slots on cell blocking probability, when one fixed GPRS time slot has been defined for the cell*

TRXs Total	Circuit switched time slots/amount		Blocking probability B, when GPRS-TRX number is:					
	Original	New value	1	2	3	4	5	6
1	7	6	4.9	—	—	—	—	—
2	14	13	3.5	3.5	—	—	—	—
3	21	20	3.0	3.0	3.0	—	—	—
4	29	28	2.8	2.8	2.8	2.8	—	—
5	36	35	2.7	2.7	2.7	2.7	2.7	—
6	44	43	2.6	2.6	2.6	2.6	2.6	2.6

The results show, that the 3% blocking probability criteria can be achieved on 3–6 TRX, regardless of the total number of the GPRS-TRXs. The resulting data rates thus vary between 54.3 and 218 kb/s (GPRS-TRX order 1→6) or 72.4–218 kb/s (GPRS-TRX order 6→1).

Case Comparison If the operator does not want to change the already dimensioned circuit switched traffic blocking value, it makes no sense to define fixed GPRS time slots. According to Case 1, the number of GPRS TRXs that remain available for GPRS transmission during the circuit switched peak hour (with 2% blocking probability) is two. Installing more GPRS TRXs does not add value, since the offered GPRS capacity saturates. Within these one or two GPRS TRXs the available GPRS capacity is, depending on the channel coding scheme, 37.3–214 kb/s in total, to be shared with all GPRS users. In this case, the GPRS, as well as circuit switched traffic, experiences the originally dimensioned blocking probability rate.

Case 2 guarantees the continuous availability of a certain minimum capacity for GPRS, but at the same time increases the original dimensioned value of circuit switched traffic blocking probability. If no changes are made to the dimensioning (adding TRXs etc.), Case 2 results in offered GPRS traffic rates of 43.2–330 kb/s with a guaranteed minimum GPRS capacity $N_{TRX\_LKM} \cdot 9.05$ kb/s – $N_{TRX\_LKM} \cdot 21.4$ kb/s. In addition to adding more capacity to the cell, few other solutions exist where the original 2% blocking probability limit would be maintained. The other solution is to accept a new, changed blocking probability value, which might be for example, 3%. This value can be achieved in more than two TRX cells, when no more than one fixed GPRS time slot is used. In this case, the offered GPRS capacity varies between 43.2 and 171.2 kb/s, and the guaranteed value for GPRS traffic is always between 9.05 and 21.4 kb/s, depending on the used channel coding schemes.

Case 3 is a combination of Cases 1 and 2. It guarantees a minimum GPRS capacity of 9.05–21.4 kb/s, and the offered capacity varies between 43.2 and 233 kb/s with a circuit switched traffic blocking probability of 3%. A graphical presentation is shown in Figure 20.20 as a conclusion of the investigated cases. The results show the total offered GPRS capacity for all users; the maximum GPRS rate for each user depends on the equipment, the supported multislot classes and so on.

If we want to maintain the already dimensioned GSM traffic blocking probability, the logical solution is to define at least two GPRS capable TRXs into the cell, with no fixed GPRS time slots. During the normal circuit switched peak hour, there will not be much benefit received from using more than two GPRS TRXs, but they are of no harm either. Of course, during the circuit switched off-peak hour the GPRS users can have more GPRS capacity. GPRS could, however, cause different traffic peaks than voice traffic, so it might be wise to define GPRS in all of the TRXs.

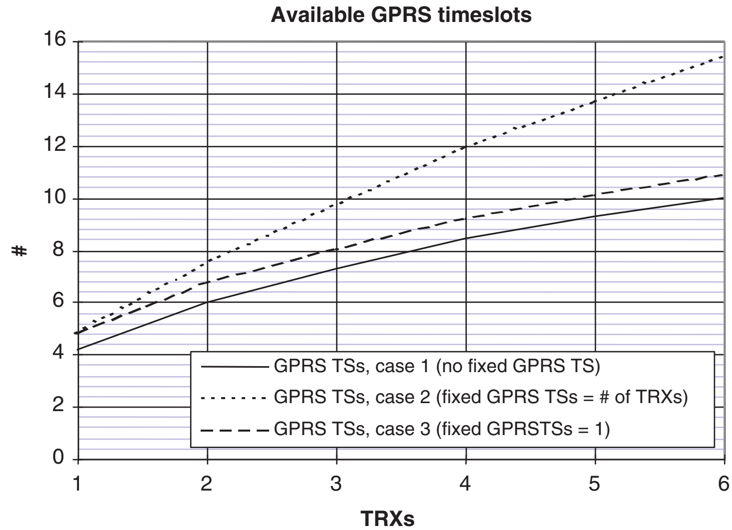

Figure 20.20 *Available GPRS timeslots in different cases during the voice traffic busy hour (with 2% blocking). In these cases, all the TRXs contain the GPRS ability.*

If the growth of the blocking probability from 2% up to 3% is accepted, it is possible to define one fixed GPRS time slot in the cells containing three or more TRXs.

When considering the interference, the downlink C/I value does not change in BCCH frequency due to the growth of the GPRS traffic. In BCCH uplink and in both directions in any other frequency, the GPRS traffic affects the C/I value. This can be seen by the growth of the outage probability, that is, increasing GPPRS traffic will slow down or, in the worst case, even cut the other GPRS calls, as well as the voice or circuit switched data calls. In other words, the typical 90% location probability for a successful call is reduced. If the network is properly dimensioned according to the worst case, the increased GPRS traffic will not affect the situation.

20.7.1.4 Distribution of Traffic by Cell

The above deductions give us the average number of GPRS channels available during the peak voice traffic hour. The distribution of circuit and packet switched traffic within the cell during the peak hour can be studied using the distribution function for circuit switched traffic.

For example in the case of one TRX, the average offered load is 2.94 timeslots out of the total 7 time slots. The average usage rate for the channels is thus $\bar{x} = 2.94 \cdot (1 - 0.02)$ Erl $= 2.88$ Erl. It can be assumed that the Poisson distribution offers a suitable theoretical basis for the study of the usage of the time slots.

When the number of time slots N of the Erlang B formula approaches infinity, according to Ref. [5] traffic is distributed as follows:

$$P_n = \frac{\frac{A^n}{n!}}{\sum_{i=0}^{N} \frac{A^i}{i!}} \xrightarrow[N \to \infty]{} \frac{\frac{A^n}{n!}}{e^A} = \frac{A^n}{n!} e^{-A}. \tag{20.29}$$

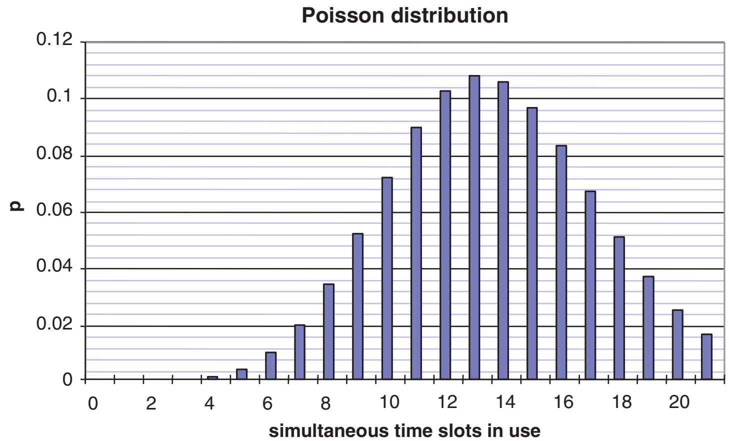

Figure 20.21 *The distribution of the number of time slots calculated according to the Poisson distribution in a cell with three TRXs, during peak hour and a 2% blocking probability.*

Based on the study, the offered traffic also follows the Erlang B-distribution. Thus, with high values of N, the formula for the Poisson distribution offers a sufficiently accurate probability for the usage of the timeslots (usage rate):

$$f(x) = p\,[X = x] = \frac{e^{-k}k^{x}}{x!}, \tag{20.30}$$

where $p(X = x)$ is the probability, that x timeslots are being used simultaneously. In this case, k represents the average offered traffic $\underline{x}$ given by the Erlang B-formula.

Figure 20.21 shows the values of $f(x)$ received in the case of three TRXs. The GPRS channel usage rate can be directly inferred from the picture as the negation of the values.

It can be seen from Figure 20.21 that when all 21 channels of the circuit switched service are in use, GPRS is completely blocked. (This occurs 2% of the time; as the Poisson distribution is not accurate with definite values of N, a slight error exists in Figure 20.21 in the probability of all channels being used.) On the other hand, GPRS can use all 21 time slots, when the number of circuit switched time slots is zero.

A cumulative probability function $f(n_{GPRS})$ for GPRS can further be inferred from the Poisson distribution to indicate the probability, that there are 0–n timeslots available for GPRS [3]:

$$\begin{cases} TS_{CS} = 0, 1, \ldots n \\ TS_{GPRS} = n - TS_{CS} \\ f(n_{GPRS}) = p\left[X \leq TS_{GPRS}\right] = 1 - \sum\limits_{z=0}^{n} f(n - TS_{CS}) \end{cases}, \tag{20.31}$$

where TS_{CS} is the number of simultaneous circuit switched time slots, TS_{GPRS} is the number of simultaneous packet switched time slots (i.e. a negation of TS_{CS}), and n is the maximum number of available traffic time slots.

20.7.1.5 The Effect of the Load

The previous study focuses on the maximum GPRS capacity available with different TRX combinations. This capacity is, however, not the same as actual capacity available for use, as the maximum data throughput

Table 20.12 *Examples of GPRS data throughput (kb/s) with some channel availability values*

	Loading % of the GPRS TRXs							
	20%		40%		70%		100%	
Case	min	max	min	max	min	max	min	max
1	7.46	42.8	14.9	85.6	26.1	150	37.3	214
2	8.64	66	17.3	132	30.2	231	43.2	330
3	8.64	46.6	17.3	93.2	30.2	163	43.2	233

saturates as the GPRS load increases. The phenomenon resembles the saturation of the Slotted Aloha-type of packet data network. The difference compared to Slotted Aloha is that in the GPRS network, the offered capacity depends on the number of GSM users, and thus changes over time. Just how efficiently the available capacity can be utilized depends on the channel coding algorithm and the number of users. As GPRS terminals send their channel allocation requests uplink as a RACH burst, GPRS traffic may in the worst case negatively affect the success of normal calls when sharing a physical BCCH time slot. If GPRS utilizes a separate, dedicated PBCCH time slot, collisions with voice service requests cannot occur over the RACH uplink.

According to Ref. [1] the maximum throughput of GPRS is 48–70% depending on the model used. In addition [2], states that according to simulation results using the packet train model, throughput is 22–67% with C/I levels 10–30 dB. The actual usability of GPRS channels can be deduced with the help of these results.

According to simulation results studied in Ref. [3] actual GPRS throughput is likely to range between 40 and 70%. Of the examples shown in Table 20.12, case 1 seems the likely alternative for an "all-round" service in the startup phase of GPRS services, as it does not affect the dimensioning of the radio network. The average data rates over the entire offered GPRS band (note: not per user, but for all users) achieved range between 14.9 and 150 kb/s during peak voice traffic hour, with GPRS being blocked 2% of the time.

Solutions such as having a cell defined with one fixed GPRS time slot as described in Case 3 can be considered for areas with a high GPRS load. Thus maximum data rate is increased by 10% compared to the solution, where no time slots are fixed for GPRS. This solution guarantees a certain minimum GPRS service while the probability of blocking for the voice service is slightly increased. Also entirely GPRS-specific cells can be considered, if equipment manufacturers support such a solution. This would, however, require changes to the capacity and frequency plans in the cell in question.

20.7.2 Coverage Area Planning

20.7.2.1 Channel Coding

The coverage areas of the GPRS and HSCSD services depend on the channel coding in use. Channel coding can be made lighter closer to the base station, or it can be left out completely with the GPRS service. This improves user data rates while BER is left unchanged. Thus it is theoretically possible to always reach a 22.8 kb/s data rate per traffic channel. The lightest GPRS channel coding enables a 21.4 kb/s rate, whereas HSCSD can utilize rates of 9.6 and 14.4 kb/s.

20.7.2.2 Simulation Results

Channel coding influences the data rate and the achieved coverage area. As the channel coding is optimized for each connection, data rates are in practice higher close to the base station, and lower on the edges of the coverage area. CS-1 is the same as the channel coding used on the SDCCH channel of basic GSM; classes CS-2 and CS-3 utilize puncturing; and class CS-4 does not use coding, that is, FEC is not in use.

Table 20.13 Parameters related to basic GPRS channel coding classes. In addition to the user's data, the data rate contains information on the RLC frame

Class	Code rate	Load	Data rate (kb/s)
CS-1	1/2	181	9.05
CS-2	~2/3	268	13.4
CS-3	~3/4	312	15.6
CS-4	1	428	21.4

All methods result in 456 bit blocks, as in basic GSM. Thus the number of bits per burst and the formation of the bursts for the radio interface are similar to GSM. CS-1 is always used for GPRS signaling regardless of the C/I ratio of the connection. The USF (Uplink State Flag) controls the allocation of the channels, and BCS (Block Check Sum) is used for error correction.

Table 20.13 lists parameters relating to different channel coding classes. The data rate with no channel coding is close to maximum transfer rate, whereas the data rate with the most efficient channel coding is slightly smaller than with basic GSM.

The results of the packet train model simulations in source [2] (throughput as a function of C/I, with 5% blocking probability) give us the maximum throughput of the four channel coding classes as a function of C/I, assuming that GPRS uses ideal link adaptation. The maximum throughput is shown in Figure 20.22.

What is noteworthy in the results of this simulation is that even with good C/I values, the actual data transfer rate of the RLC/MAC layer is only 1.1–1.8 kilobytes per second with a 5% blocking probability, which compares with data rates 8.8–14.4 kb/s. This is probably due to the nature of the packet train model, where the average size of the packets is relatively small. As the number of users increases, collisions occur with the channel allocation requests, leading to delays. It can also be seen from the figure that CS-4 begins to have an effect when C/I is 20 dB or better. CS-3 is in use in the 10–20 dB range, and CS-2 for approx. 7–10 dB. When the C/I ratio is even weaker, CS-1 is used. It is not obvious from the background information given on the simulations whether the header information and so on contained in the RLC frames has been taken into account when calculating the data rate.

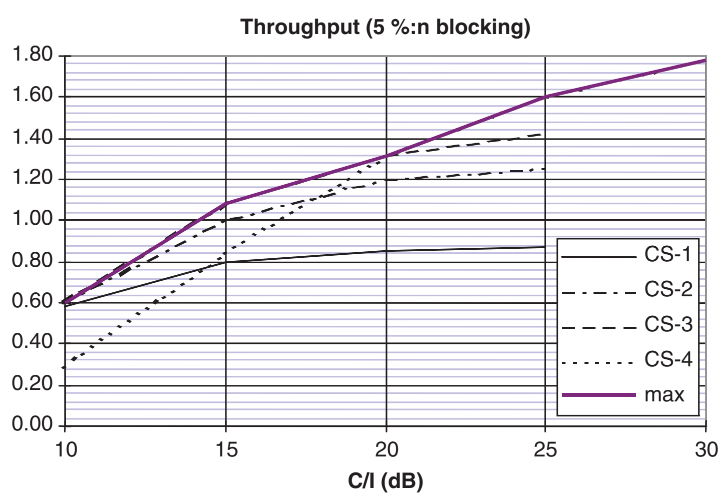

Figure 20.22 *GPRS data throughput (kilobytes per second) as a function of C/I as interpreted from source [2]. Figure interpreted from the original work of Paolo Zanini.*

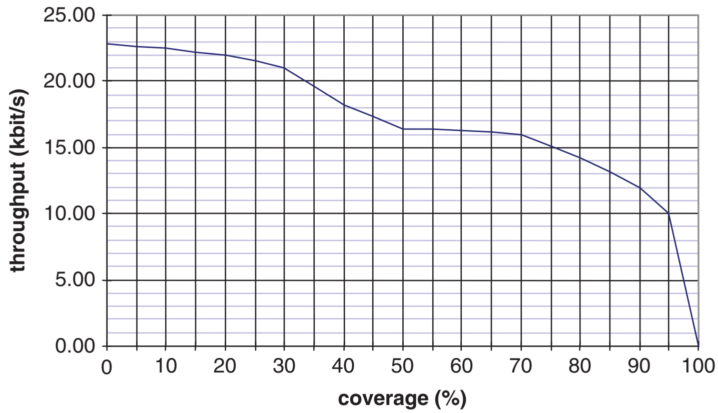

Figure 20.23 GMSK throughput as a function of the coverage area (95% location probability) in an environment where coverage area is constrained; interpreted from sources [6, 7]. Figure interpreted from the drafts of ETSI GSM.

Sources [6, 7] contain results of simulations on the performance of both EDGE and GPRS. The sources include information about the data throughput vs. coverage area (Figure 20.23, a numerical estimate from the original source).

The data rate experienced by the GPRS user is directly dependent on the C/I value, or in the case of a noise limited network, on the E_b/N_0 value. Assuming that GPRS can utilize ideal link adaptation, basic GPRS can reach data rates per TSL, for example, as shown in Figure 20.23, depending on C/I-values. If a network were defined for example, for a 18 dB C/I-relationship, the maximum data rate per each individual time slot would be 1.3 kilobytes or 10.4 kb/s, according to Figure 20.23. The model gives a relatively pessimistic overview of the situation, as it assumes that the packets are small in size.

The coverage areas corresponding to the GPRS channel coding classes in an interference-free environment can be inferred from Figure 20.23. The surface areas corresponding to the C/I ratios are listed in Table 20.14 (C/I is interpreted to mean the same as the signal-to-noise relationship in this context).

20.7.2.3 Link Budget Calculations

Voice Service

The effects of the different channel coding classes on the coverage area can be estimated with the help of a link budget, if it can be assumed that modifications can be taken into account for example, by adjusting the sensitivity of the MS and the BTS. Tables 20.15 and 20.16 show an example of a typical GSM link budget for urban and rural environments for GSM 900-system voice services.

Table 20.14 The coverage areas corresponding to GPRS classes as interpreted from sources [6, 7] as compared to GSM voice traffic

GPRS class	Data rate (kb/s)	Coverage (% from the voice service area)
CS-1	9.05	98
CS-2	13.4	83
CS-3	15.6	70
CS-4	21.4	25

Source: Data by courtesy of ETSI.

Table 20.15 Examples of link budgeting for GSM 900 voice traffic

		Parameter	City, macro cell	City, micro cell	Rural, macro cell
Downlink					
BTS	RF power (bef. combiner)	+ Pbts_tx, dBm	41	39	41
	combiner	−Lcomb, dB	4	4	4
	cable and connector loss	−Lbts, dB	3	3	3
	antenna gain	+Gbts, dBi	11	11	17
	= radiating power	Pbts_txpeak, dBm	45	43	51
MS	sensitivity	−Sms, dBm	−102	−102	−102
	antenna gain	+Gms, dBi	0	0	0
	cable and connector loss	−Lms, dB	0	0	0
	path loss	Lc_downlink, dB	147	145	153
Uplink					
MS	input power	+Pms_tx, dBm	33	33	33
	cable and connector loss	−Lms, dB	0	0	0
	antenna gain	+Gms, dBi	0	0	0
BTS	RX sensitivity	−Sbts, dBm	−104	−104	−104
	antenna gain	+Gbts, dBi	11	11	17
	cable and connector loss	−Lbts, dB	3	3	3
	diversity gain	+Gdiv, dB	2	0	2
	mast head amplifier	+Smhpa, dB	0	0	0
	path loss	Lc_uplink, dB	147	145	153

Table 20.16 Fading margins for the examples in the link budget, taking into account the values of the parameters in the table

Unit		City, macro cell	City, micro cell	Rural, macro cell
body loss	dB	3	3	3
fading margin	dB	0	0	0
slow fading margin	dB	5	5	5
indoor attenuation	dB	0	0	0
shadowing margin	dB	0	0	0
maximum path loss	dB	**139**	**137**	**145**

Different fading margins – body fading, fading margin,[3] slow fading margin,[4] possible indoor fading (on average 20–30 dB) and shadowing margin (approx. 0–6 dB; only in cases where blocks exist in the 1st degree Fresnel ellipse) – are added to the acquired radio path fading margin.

The planned value for the street-level received power level is Pbts_txpeak –Lc_max. Thus, in the example cases, the power level of an urban mega cell is (45–145) dB = −100 dBm, an urban micro cell (43–137) dB = −94 dB, and for a rural cell (51–145) dB = −94 dB.

[3]Total fading effect caused by Rayleigh-fading has been taken into account in GSM systems.

[4]The reliability range of the normal distribution shifts from 50%->75%, which corresponds with a call success location probability of 90%.

Table 20.17 The coverage areas of the cells of the example

		City, macro cell	City, micro cell	Rural, macro cell
prediction model		O-H	W-I	O-H
radius of the cell	km	2.3	1.3	21.8

The cell coverage areas can be calculated using the Okumura-Hata (radius over 1 km) or the Walfisch-Ikegami (micro cell). The values for coverage areas of the example are shown in Table 20.17.

Data Transfer

The shortening of the cell's radius can be further estimated using the Okumura-Hata and the Walfisch-Ikegami, when additional fading caused by packet data transfer is added to the values of the previous example. Figures 20.24–20.28 show examples of cell radii calculated using the Okumura-Hata and the Walfisch-Ikegami models for certain radio signal fading values. Table 20.18 shows calculations on the percentage reductions of cell coverage areas with certain additional fading values. What the actual fading margin is depends mostly on the C/I value required by GPRS channel coding classes compared with the 9 dB-value specified for the GSM voice service.

Both the Okumura-Hata and the Walfisch-Ikegami are logarithmic functions. Both can thus be used to form general-purpose tables over the permitted area to indicate the percentage change in the radius as a function of the change in advancement path loss. Figures 20.27 and 20.28 show the change in advancement fading as a function of the change in the cell area. This can be used to study the change in coverage area depending on the C/I values required by GPRS channel coding classes, compared with GSM voice traffic.

Descriptions of GPRS BER- and BLER-simulations can be found in for example, sources [2, 7], for different channel coding classes. According to Ref. [8], packet traffic channels have a BLER-requirement of 10%, while USF (Upper State Flag) requires 1%, and PRACH channels 15%. Listed in Table 20.19 are the C/I values with a 10% BLER-value as interpreted from Ref. [2] (physical layer simulations, block erasure rate

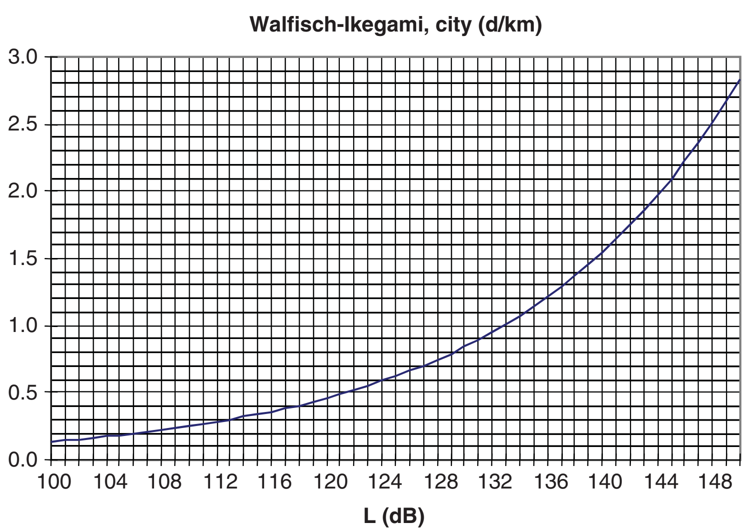

Figure 20.24 *Urban cell sizes calculated using the Walfisch-Ikegami advancement formula for certain fading margins (f = 900 MHz).*

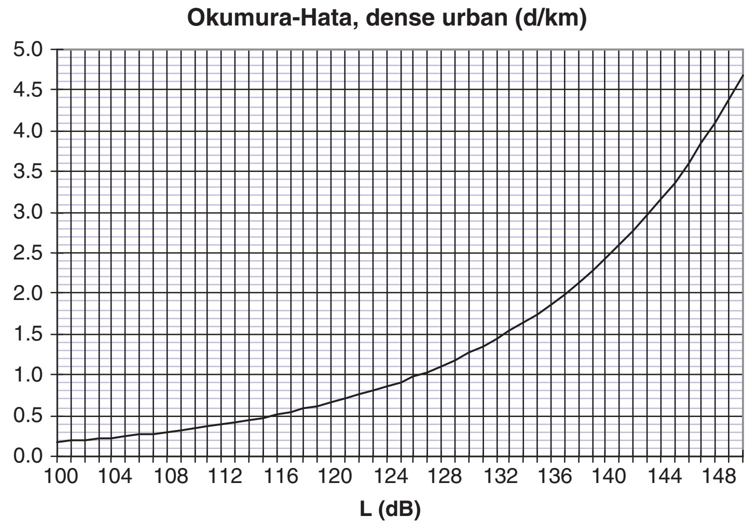

Figure 20.25 *Urban macro cell sizes calculated using the Okumura-Hata advancement formula for certain fading margins (f = 900 MHz, hbs = 30 m, hms = 1.5 m).*

vs. C/I, TU3 no FH / TU50 ideal FH). The C/I values have been compared with the C/I-value requirements set in the specifications [8], with and without frequency hopping. In addition, the table contains the results from Ref. [7], which have been achieved using the TU3 model with ideal frequency hopping. These results represent the situation only in a noise limited environment.

It can be seen from Table 20.19 that according to the simulation results in Refs. [2, 7], the requirements set by the specifications are met with all channel coding classes with an average margin of 2 dB. With channel

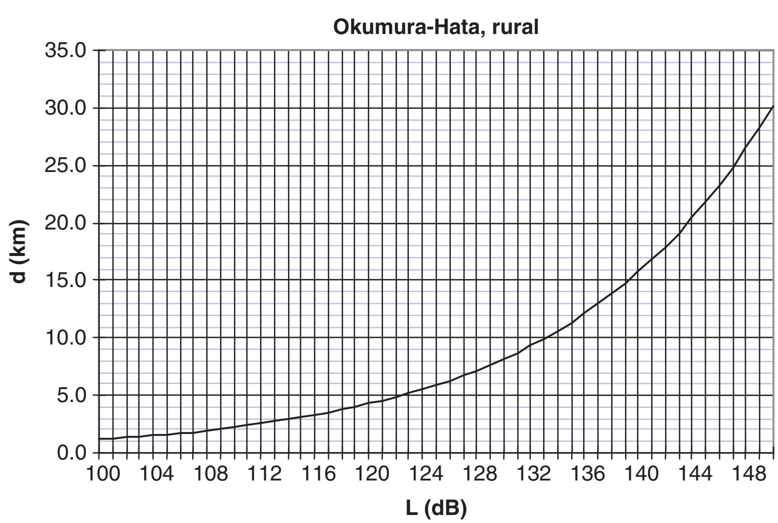

Figure 20.26 *Rural cell sizes calculated using the Okumura-Hata advancement formula for certain fading margins (f = 900 MHz, hbs = 30 m, hms = 1.5 m).*

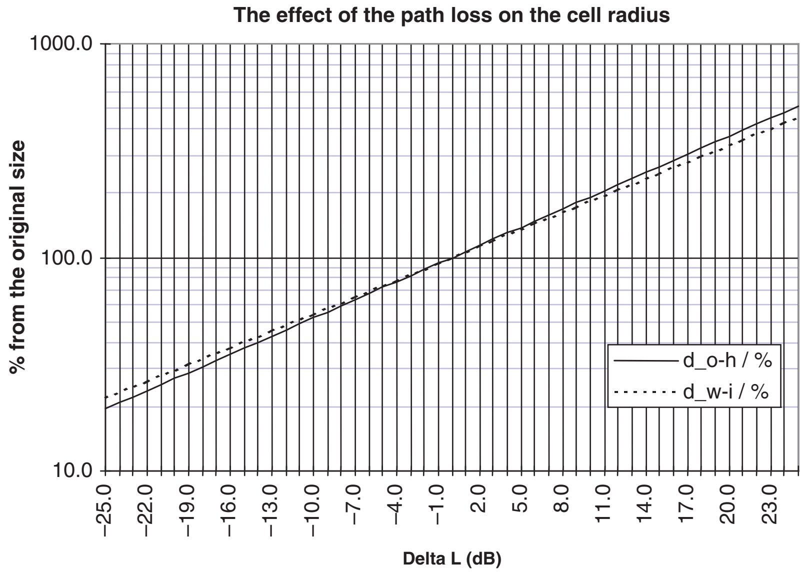

Figure 20.27 *The effect of the change in advancement fading ($\Delta L = \pm 25$ dB) on the cell radius.*

coding classes CS-2 and CS-3 the results given by Refs. [2, 6] differ by approximately 1 dB, but otherwise these two simulations have given quite similar results. Table 20.20 compares the requirements for C/I-values with the 9 dB value set by the specifications for GSM voice traffic.

The coverage areas corresponding to each GPRS channel coding class are compared with GSM coverage areas (with and without frequency hopping) in Table 20.21. Table 20.22 shows the size of each service area compared to speech traffic calculated using the Okumura-Hata (O-H) for large cells and the Walfisch-Ikegami

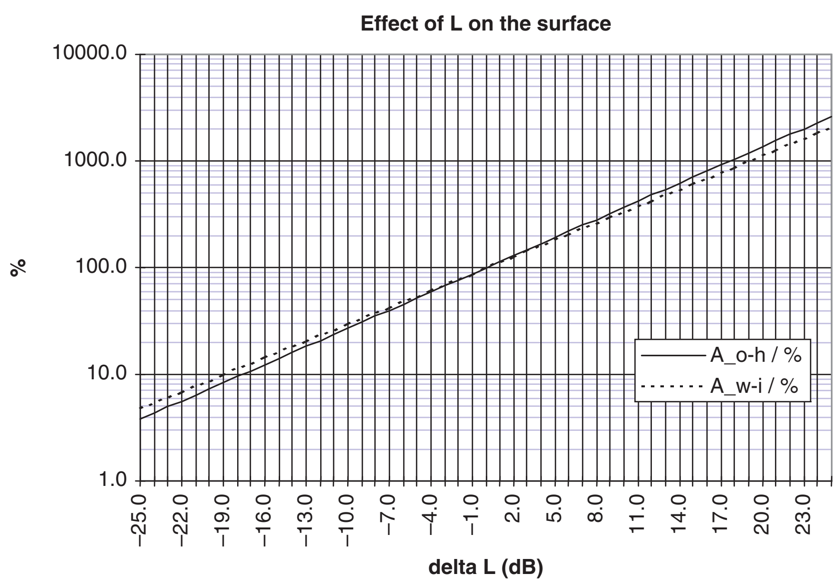

Figure 20.28 *The effect of the change in advancement fading ($\Delta L = \pm 25$ dB) on the cell area.*

Table 20.18 An example of the shortening of the cell radius and the decrease in the cell coverage area for certain fading margin values, as compared to voice traffic

	Radius/surface compared to the voice service (%)	
Additional loss (dB)	Okumura-Hata	Walfisch-Ikegami
0.2	98.7/97.4	98.8/97.6
0.5	96.8/93.7	97.0/94.1
1	93.7/87.7	94.1/88.6
2	87.7/77.0	88.6/78.5
4	77.0/59.3	78.5/61.6
6	67.6/45.6	69.5/48.3
8	59.3/35.1	61.6/37.9
10	52.0/27.1	54.6/29.8
15	37.5/14.1	40.3/16.2
20	27.1/7.3	29.8/8.9

Table 20.19 A comparison of the C/I values for packet traffic channels on the 900 MHz frequency range given by sources [2, 7, 8]. The BLER-value is 10%

	CS-1			CS-2			CS-3			CS-4		
	[8]	[2]	[7]	[8]	[2]	[7]	[8]	[2]	[7]	[8]	[2]	[7]
noFH	13	10.8	—	15	12.8	—	16	13.7	—	19	17.2	—
FH	9	7.1	7.2	13	11.5	10.4	15	13.6	12.4	23	20.8	20.5

Source: Data by courtesy of ETSI and Paulo Zanini.

(W-I) for micro cells. Table 20.22 describes the change in the coverage area with the limit values set by the specifications, and based on the results of CSELT simulations. The C/I-values have been calculated based on the assumption that the level of the interfering sum signal is constant, and that only the change in the serving cell's propagation path is significant.

Sensitivity requirements have been specified for GPRS traffic channels for situations TU3 (typical urban, 3 km/h) without frequency hopping, and TU50 (typical urban, 50 km/h) with frequency hopping. Examples of GSM and GPRS coverage areas can be calculated using the link budget and an applicable model for advancement. Table 20.22 utilizes the above link calculations to show examples of fading margins for traffic in different environments, and balanced between uplink and downlink.

Table 20.20 The values in Table 12.16 compared with the 9 dB C/I requirement of GSM.

	CS-1			CS-2			CS-3			CS-4		
	[8]	[2]	[2]	[8]	[2]	[7]	[8]	[2]	[7]	[8]	[2]	[7]
no FH	+4	+1.8	—	+6	+3.8	—	+7	+4.7	—	+10	+8.2	—
FH	0	−1.9	−1.8	+4	+2.5	−1.4	+6	+4.6	+3.4	+14	+12	+12

Table 20.21 *Service area size (radius r/area A) (%) compared to GSM voice traffic (f = 900 MHz). The new sensitivity limits set in the specifications for the channel coding classes have not been considered. O-H = Okumura-Hata, W-I = Walfisch-Ikegami*

	CS-1		CS-2		CS-3		CS-4	
Model	[2]	[8]	[2]	[8]	[2]	[8]	[2]	[8]
O-H, no FH	89/79	77/59	78/61	68/46	74/54	63/40	59/34	52/27
OH, FH	113/128	100/100	85/72	77/59	74/55	68/46	46/21	40/16
W-I, no FH	90/80	78/62	79/63	70/48	75/57	65/43	61/37	55/30
W-I, FH	112/126	100/100	86/74	78 62	76/57	70/48	49/24	43/18

Table 20.22 *The fading margins in the case examples corresponding to GSM and GPRS channel coding classes. The table does not take into account the sensitivity limits set in the specifications for the channel coding classes. For example, in the case of CS-1 w/out frequency hopping, there is an additional fading of +1.8 dB for the propagation path (based on Ref. [2]). Thus the fading changes from the 139 dB set for GSM to 137.2 dB*

			CS-1		CS-2		CS-3		CS-4	
Environment	Model	GSM	[2]	[8]	[2]	[8]	[2]	[8]	[2]	[8]
city	TU3, no FH	139	137.2	135	135.2	133	134.3	132	130.2	129
macro cell	TU50, FH	142	140.9	139	136.5	135	134.4	133	127.2	125
city	TU3, no FH	137	135.2	133	133.2	131	132.3	130	128.8	127
micro cell	TU50, FH	140	138.9	137	134.5	133	132.4	131	125.2	123
rural	TU3, no FH	145	143.2	141	141.2	139	140.3	138	136.8	135
macro cell	TU50, FH	148	146.9	145	142.5	141	140.4	139	133.2	131

In this study, normal GSM frequency hopping is expected to expand the fading margin by 3 dB. If the changes in GPRS sensitivity values are not taken into account, the result is the values shown in Table 20.23.

Both the results from the CSELT simulations [2] and the limits set by the specifications [8] have been used as a basis for the reference values for the C/I-ratio. The values in the link budget have been compared to the C/I requirement of 9 dB of GSM, and have been summed up in the link budget sensitivity value. The 10% BLER-value set by the specifications has been used.

Table 20.23 *The fading margins in the case example corresponding to GSM and GPRS channel coding classes, when the new sensitivity limits set in the specifications for the channel coding classes have been taken into account*

			CS-1		CS-2		CS-3		CS-4	
Environment	Model	GSM	[2]	[8]	[2]	[8]	[2]	[8]	[2]	[8]
city	TU3, no FH	139	139.2	137	137.2	135	136.3	134	129.8	128
macro cell	TU50, FH	142	142.9	141	135.5	134	131.4	130	115.2	113
city	TU3, no FH	137	137.2	135	135.2	133	134.3	132	127.8	126
micro cell	TU50, FH	140	140.9	139	133.5	132	129.4	128	113.2	111
rural	TU3, no FH	145	145.2	143	143.2	141	142.3	140	135.8	134
macro cell	TU50, FH	148	148.9	147	141.5	140	137.4	136	121.2	119

Table 20.24 The change in the GPRS propagation path loss L as compared to the GSM signal. The GSM L with frequency hopping is expected to be 3 dB greater than without frequency hopping. The new sensitivity limits set in the specifications for the channel coding classes have been taken into account

		GSM	CS-1		CS-2		CS-3		CS-4	
Environment	Model	L	[2]	[8]	[2]	[8]	[2]	[8]	[2]	[8]
city	TU3, no FH	139	+0.2	−2	−1.8	−4	−2.7	−5	−9.2	−11
macro cell	TU50, FH	142	+0.9	−1	−6.5	−8	−10.6	−12	−26.8	−29
city	TU3, no FH	137	+0.2	−2	−1.8	−4	−2.7	−5	−9.2	−11
micro cell	TU50, FH	140	+0.9	−1	−6.5	−8	−10.6	−12	−26.8	−29
rural	TU3, no FH	145	+0.2	−2	−1.8	−4	−2.7	−5	−9.2	−11
macro cell	TU50, FH	148	+0.9	−1	−6.5	−8	−10.6	−12	−26.8	−29

Table 20.25 The GPRS channel coding schemes and the corresponding cell radii and surfaces compared to the basic GSM voice service coverage areas. The calculations have been carried out by using Okumura-Hata model with GSM C/I reference value of 9 dB

	CS-1		CS-2		CS-3		CS-4	
	[2]	[8]	[2]	[8]	[2]	[8]	[2]	[8]
radii/%	101	88	89	77	84	72	55	49
A/%	103	77	79	59	70	52	30	24

The cell radii and areas corresponding to the different channel coding classes shown in Table 20.24 are achieved by analyzing the situation without frequency hopping shown in Table 20.23. The values are based on the Okumura-Hata model. The results of the Walfisch-Ikugami used for urban micro cells do not differ significantly from these results. Compared to the specifications, the values describe a pessimistic situation, while the CSELT simulation results [2] relate to an optimistic situation.

Now, based on the values of the Table 20.25, utilizing the formulas 20.23 and 20.24, the weight parameter values can be calculated as shown in Table 20.26. The comparison has been made based on the specification limits and the CSELT simulation values.

Formula 20.25 can be used to calculate the theoretical GPRS traffic throughput of a single time slot of an individual cell. Based on Ref. [2], the optimistic value can now be formed:

$$v_{cell}(A_p)_{Zan} = (9.05 \cdot 0.24 + 13.4 \cdot 0.09 + 15.6 \cdot 0.40 + 21.4 \cdot 0.30) kb/s = 16.0 kb/s \tag{20.32}$$

Table 20.26 The weight multipliers p1–p4 as inferred from Table 20.25

weight	Source [2]	Source [8]
P_1	0.24	0.18
P_2	0.09	0.07
P_3	0.40	0.28
P_4	0.30	0.24

Source: Data by courtesy of ETSI and Paulo Zanini.

and the optimistic value, according to the limit values set by specification [8] are:

$$v_{cell}(A_p)_{0505} = (9.05 \cdot 0.18 + 13.4 \cdot 0.07 + 15.6 \cdot 0.28 + 21.4 \cdot 0.24)kb/s = 12.1kb/s. \qquad (20.33)$$

In the first phase of the GPRS service, only channel coding classes CS-1 and CS-2 were used, mainly due to the limitations of the *Abis*-interface. In this case, GPRS traffic throughput can be calculated by replacing the CS-3 and CS-4 values in Formula 20.25 with the bit rates of CS-2, as follows:

$$v_{cell\_Ph1}(A_p)_{Zan} = (9.05 \cdot 0.24 + 13.4 \cdot (0.09 + 0.40 + 0.30))kb/s = 12.8kb/s \qquad (20.34)$$

and

$$v_{cell\_Ph1}(A_p)_{0505} = (9.05 \cdot 0.18 + 13.4 \cdot (0.07 + 0.28 + 0.24))kb/s = 9.54kb/s. \qquad (20.35)$$

20.7.2.4 Network Interference Levels

The limit C/I-value of the GSM network can be dimensioned separately for the BCCH-frequencies and other frequencies. Because signal is sent downlink on the BCCH-frequency at a maximum defined power level regardless of the actual usage of the time slots, the BCCH causes higher interference levels than other frequencies. Thus it makes sense to use a loose reuse pattern (e.g. $K \geq 12$–16). The planned value 12–15 dB can thus be used for the C/I of the BCCH-frequencies.

For other frequencies, the C/I value can be a few dB weaker, for example, 9–12 dB with a reuse pattern K>=9–12. An even smaller reuse pattern can be used for micro cells. The C/I of the downlink BCCH in a way represents the worst case or an interference level with a 100% load on the cells. The C/I value of other cells can be a planned value for example, based on the average 40–80% cell load of the above calculation.

The conclusion can be drawn from the example that if GPRS is used with the time slots of the BCCH-TRX, it does not cause changes to the planned C/I-levels of the network nor does it influence the capacity of the circuit switched connections. GPRS capacity is a probability function, where, with a 2% probability over a one-hour time period, GPRS service is blocked in total for $3600 \times 0.02 = 72$ seconds, when all time slots are reserved for circuit switched connections.

The average interference level of the network can be considered to vary between the normal and the 100% load situations when GPRS is used on frequencies other than the BCCH in the downlink direction. The C/I-relationship of the basic GSM network can be planned so that also the non-BCCH-TRXs are assumed to carry a constant load. In this way, the interference in the network can be predicted according to the worst case scenario. This means that the planned 9 dB C/I-value is reached on the edge of the non-BCCH-TRXs' coverage area when the respective time slots of both the serving and the interfering TRXs on the same channel are in use.

20.7.3 Frequency Planning

The simultaneous use of multiple time slots causes momentary load peaks in the network. In regular network dimensioning, the average load is 40–80% of the maximum possible load, depending on the number of TRXs. GPRS can at times utilize the entire remaining capacity and thus add to the level of interference of the network.

Transparent HSCSD data transfer, where a certain data rate and, in effect, a certain minimum number of time slots, is guaranteed for each user, can be quite problematic from the point of view of other voice and data users. For example, a user of four time slots leaves the TRX only three free time slots. The trunking effect based on the Erlang B-formula is relatively weaker on these three time slots, and can cause capacity problems.

Table 20.27 *The C/I values on the edge of the cell with 90% location probability, with certain recurrence patterns for cells with one, two, and three sectors*

	C/I		
K	360 deg	120 deg	60 deg
1	−6.0	−1.3	1.7
3	2.3	7.1	10.1
4	4.5	9.3	12.3
7	8.8	13.5	16.5
9	10.7	15.4	18.4
12	12.9	17.6	20.6
13	13.5	18.2	21.2
16	15.0	19.8	22.8
19	16.3	21.1	24.1

The use of all GPRS time slots is dynamic, which means that a circuit switched connection can be offered when required (that is, if the circuit switched connections do not use up the entire capacity). If fixed or dynamic GPRS time slots are defined for the TRX, usage peaks can occur, and the average level of interference increases. Thus GPRS may lead to a need to expand the used recurrence pattern. Alternatively, the network can be dimensioned according to the worst case scenario, to assume that all TRXs have a full load constantly. Thus the actual use of the time slots does not affect the planned values of the recurrence patterns.

Table 20.27 contains C/I-values on the outer edge of the cell. These values are the results of calculations, where a C/I-relationship is calculated for symmetric reuse patterns using the hexagonal principle and assuming that the fading margin caused by the fading of the log-normal to be approximately 6.6 dB with a 10% interference blocking probability criteria [9]. It has further been assumed that the antenna sectors are ideal, and that a constant amplification exists for the entire sector area.

The actual network may not be built using the hexagonal principle. Instead, frequency planning may even be "random, but in a controlled fashion." The network can also contain lots of areas with coverage overlaps. The recurrence pattern for frequencies on the same channel can also be used to estimate the situation with such networks. The recurrence pattern of such a network can be concluded when the number of cells, frequencies, and TRXs on the same channel in a given area is known.

In real networks, BCCH-frequencies are often placed loosely in the terrain so that the C/I-value would be approx. 12–15 dB which corresponds to the reuse pattern multiplier K = 12–16 for cells with an omnidirectional beam. Other TRXs are often allowed an approximately 9–12 dB C/I-relationship, which means that the 7-reuse pattern can be used without sectoring. The reuse pattern can be further "tightened" from the values presented above in the case of sectored cells.

Table 20.28 lists the sizes of the recurrence patterns required by different GPRS channel coding classes. The values have been calculated assuming a BLER-value of 10%, and the advancement multiplier n has a value of 3.5. The values in the table are the C/I-values defined by the specifications and the results of the CSELT simulations, without frequency hopping. The results have been compared with the 9 dB limit value defined by the GSM specifications, and a more realistic 14 dB-value.

The values in Table 20.28 represent a situation with average advancement circumstances. The advancement multiplier[5] n has a strong effect on the theoretical number of recurrence patterns. Thus, the C/I-value is greater

[5]An improvement of one point in n typically leads to a C/I-value improvement of about 6 . . . 9 dB.

Table 20.28 Recurrence patterns required by different GPRS channel coding classes

Channel coding	C/I-limit (dB)		K 360 deg	K 120 deg	K 60 deg
GSM	Spec limit	9.0	9	4	3
	Practical limit	14.0	16	9	7
CS-1	[2]	10.8	9	7	4
	[8]	13.0	12	7	7
CS-2	[2]	12.8	12	7	7
	[8]	15.0	16	12	7
CS-3	[2]	13.7	13	7	7
	[8]	16.0	19	12	7
CS-4	[2]	17.2	>19	13	9
	[8]	19.0	>19	16	12

in a populated urban area, and the recurrence pattern is smaller than the values shown in the table. Respectively, in a rural, less populated area with open spaces both the N and the C/I-values are smaller, and there are more recurrence patterns. In reality, the C/I-value also depends on the load on the network.

In any case, Table 20.28 indicates the relationship between the GPRS channel coding classes and the recurrence patterns required for the channel coding classes to function. According to the table, the channel coding class CS-4 for fastest data transfer would require two- or three times the multiplier for the recurrence pattern of basic GSM, in order to function in the entire cell area.

20.7.4 Parameter Planning

20.7.4.1 Channel Switching

GPRS channel switch is more dynamic with GPRS than with HSCSD. In fact, the GPRS service does not have a similar channel switch as the GSM system. Instead, the cell reselection functionality is used to switch from one cell to another. Both GPRS cell switching and routing area switching cause signaling, the amount of which depends on for example, whether the location is within the area of one or two SGSNs. Thus the parameters related to cell selection and reselection should be chosen carefully.

The optimal selection of routing areas is important for signaling. When GPRS is first implemented, the network should be constructed so that the GPRS routing areas (RA) are the same as the GSM location areas (LA). When sufficient information has been collected on the nature of signaling, the routing areas can be divided into smaller sections within the location areas in order to balance the signaling traffic.

20.7.4.2 Frequency Hopping

The results of simulations described in Ref. [2] can be collected to form a table with the throughput values for the RLC/MAC-layer as a function of the C/I, both with ideal frequency hopping (TU50) and without it (TU3). These situations represent the extreme cases when studying the impact of fading phenomena. The results have been collected by changing what was graphically represented into a numerical format, which adds a small error factor into the studies presented here. Figure 20.29 shows the simulation results for channel coding class CS-1 for the values in question.

Frequency hopping improves the throughput of CS-1 by approximately 1–3 dB, up to the point where throughput is saturated. On the other hand, frequency hopping can improve data transfer by approximately

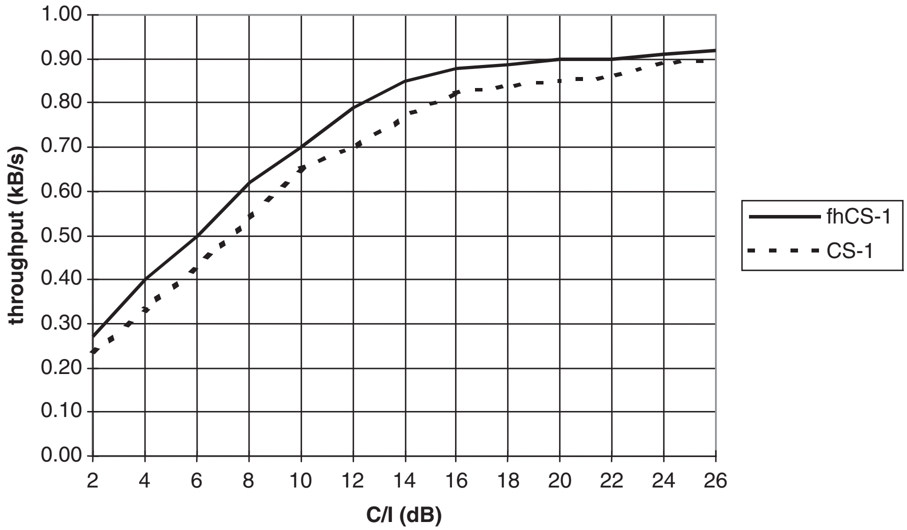

Figure 20.29 *A comparison of the results of simulations in Ref. [2] for CS-1, with and without frequency hopping. Throughput is given as kilobytes per second. Figure interpreted from the original work of Paolo Zanini.*

0.1 kilobytes per second with a certain C/I-value. In comparison, according to the results of the CSELT simulations, frequency hopping does not improve performance with other channel coding classes. Saturation simulations have shown that the situation is almost the opposite with CS-4. A partial explanation may be that the interleaving depth for GPRS traffic is only 4 compared to a value of 8 for voice traffic and a maximum of 19 for data traffic.

The importance of the interference diversity gained via frequency hopping reduces as the load increases. It can be assumed that the benefits from frequency hopping may be eroded completely as GPRS usage increases. However, there is no harm from frequency hopping in principle. Also, even if interference diversity were to decrease, frequency hopping would remain useful due to its help in reducing the impact of frequency diversity or Rayleigh-fading.

20.7.4.3 Power Control

GPRS power control has been defined for both uplink and downlink. Uplink power control is a mix of open loop- and closed loop-procedures. Downlink, the MS measures signal strength and the channel quality. The *open loop*- procedure is used in the initiation phase of packet data transfer, and it is based on the signal strength received by the MS. The network controls the *closed loop*-procedure, which is comparable to GSM power control. The GPRS-MS power control can be represented with the formula

$$P_{CH} = \min\left(\Gamma_{CH} - \alpha C, P_{\max}\right), \tag{20.36}$$

where Γ_{CH} is the parameter sent by the network to the MS at the time of connection, α has a value of either 0 (closed loop) or 1 (open loop), C is the field strength (power) received by the MS, and $P_{\max}$ is the maximum MS transmission power in the cell in question. The measurement results of the MS are transferred on the PDTCH (which contains the PACCH) if in transfer mode, otherwise on the PCCCH. The interference level is measured from all 8 idle time slots on the frequency in question.

It is in principle recommended that GPRS power control be used for both uplink and downlink in order to reduce the interference level of the network.

20.7.4.4 Cell Selection

GPRS cell selection criteria depend on the use of the PBCCH (packet BCCH). If PBCCH has not been defined, and GPRS used the GSM BCCH instead, the C31 (signal strength threshold criterion) and C32 (cell ranking) criteria can be used. The final strategy for cell selection will probably be developed only after some real experience has been gained.

20.7.4.5 Discontinuous Reception

The DRX-functionality of the MS has been specified for GPRS to be used in the standby- and ready-modes. The achieved benefits are basically the same as in the case of basic GSM. The most significant benefit is the extended battery time for the user. The DRX has no direct impact on network planning.

20.7.4.6 Channel Configurations

According to the specifications, GPRS broadcast channel signaling can take place either on the GSM BCCH-time slot or on a dedicated PBCCH channel. In the first alternative, the GPRS naturally takes up some of the resources from basic GSM signaling. Also, when terminals signal the network or reply to a PCH-channel signal on the RACH channel, the signaling of the circuit switched users and the GPRS users can collide. This means, that the initial signaling of GPRS and basic GSM can interfere with each other. This is not likely to be a problem with low GPRS loads, but as capacity shortages appear, it is recommended to use a dedicated PBCCH channel for GPRS. If GPRS uses the BCCH, the traditional 51-frame structure is used, and GPRS signaling reserves its own blocks.

The fixed PCCCH, where the PBCCH is located, takes up one physical time slot. Thus the situation can capacity-wise be compared to the earlier described case 3 keeping in mind that no minimum capacity is yet guaranteed for GPRS traffic without a fixed GPRS traffic time slot. Using the PBCCH prevents the circuit switched and GPRS users from colliding with each other.

If GPRS signaling uses the BCCH, a maximum of eight users can be multiplexed uplink in a single GPRS time slot. The connections are controlled using the Uplink State Flag (USF), which has been defined to be three bits in size and located in the header field of the packets. When PBCCH is used, one USF mode has to be defined in the signaling, which means that there can be a maximum of seven GPRS users per time slot uplink.

20.8 Other Planning Considerations

20.8.1 Multilayer Networks

A real GSM consists of different types of cells; the purely homogeneous hexagonal model is far from reality. A looser recurrence pattern is typically planned for the BCCH channels than for the other frequencies, as the BCCH-TRX has continuous downlink traffic on all its time slots.

The different-size recurrence patterns of one network can also be utilized in a so-called multilayer network, achieving theoretical capacity increases of 50–100%. The multilayer network is based on the controlled increase of the interference level of the network, combined with the interference diversity achieved with frequency hopping. The functioning of a multilayer network in HSCSD and GPRA networks could potentially be an interesting topic for research.

20.8.2 Variation of the Load

Cell-based networks typically experience both short- and long-term variance in loads. The same load peaks can be seen in the load profile on working days. The profiles for suburban cells differ from the downtown business area profiles both during weekday evenings and other freetime periods. During popular holiday seasons some seasons may experience overload situations, as the people on holiday travel to their summer cottages. It is possible to add temporary capacity to overloaded regions by using transportable base station equipment for instance. In such situations, the interference caused by cochannel interferences must be eliminated.

The profile of data calls may differ from the profile of regular calls. The general observation can be made that the peak hour for data calls takes place later than the regular peak hour. Roaming subscribers also seem to make more data calls than the home network subscribers. Both observations can be explained with the GSM data user profile: users such as business travelers typically use GSM data to check their email in the evenings while on a business trip.

It seems likely that at least in the first phase of the services, it is mainly business users who will use HSCSD and GPRS for mobile data transfer. The traffic profile will probably resemble the current GSM data profile. Also, the data-intensive profile of roaming subscribers must be taken into account from the beginning of the GSM and HSCSD service.

One potentially significant factor is the use of the queuing- and Directed Entry-functions in the GSM network for GPRS and HSCSD services, simultaneously. These functions can be used to significantly increase the usability of the voice traffic while maintaining the call blocking probability level constant. The enhanced use of the voice traffic capacity means less resources for GPRS and HSCSD traffic, respectively.

20.8.3 Vegetation and Weather Conditions

In addition to shifts in the overall load on the network, the different seasons can also influence the advancement behavior of radio signals. Humidity, especially during rainy seasons in the spring and autumn, can cause additional fading to the radio signal. Forests account for approximately 70% of the surface area of Finland, which makes forests a significant source of fading. The typical fading effect caused by forests can be estimated using for example, a formula, which has been presented in the CCIR report 236-5 (a less accurate version of the formula can be found in Ref. [10]. The formula is as follows:

$$L(dB) = 0.187(f^{0.284})(d^{0.588}), \tag{20.37}$$

where f equals the frequency in MHz, and d is the depth of the forest area in meters. The formula is valid for frequencies 0.20–90 GHz. Figure 20.30 shows examples of the fading effect caused by forests for frequencies 900, 1800, and 1900 MHz calculated using the formula.

It can be seen from the picture, that the fading effect of the forest increases relatively more with higher frequencies as a function of the depth of the forest.

In the winter when water freezes and causes changes to the cell structure in trees the fading effect is typically 0–2 dB smaller. However, the fading effect of the forest does not in principle change radio network planning from the point of view of new data services.

Exceptional weather conditions, such as the inversion of the layers of the atmosphere, can also influence the so-called "radio weather". The signal may become channeled and advance several times the normal distance. Exceptional weather conditions have not proven a problem with GSM frequencies, so there is no need to prepare for changes in "radio weather" when planning for new data services.

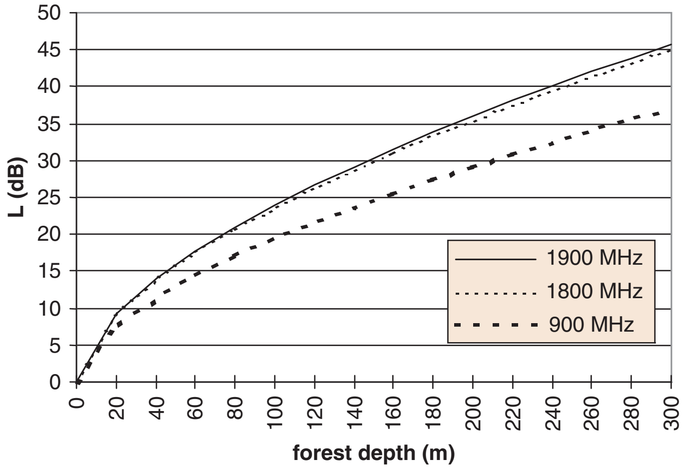

Figure 20.30 *The path loss effect caused by forests.*

20.8.4 Quality of Service Levels

Figure 20.31 shows an example of the QoS-level negotiated at the beginning of a GPRS connection. A default offered QoS-level has been defined in the subscriber's HLR/GR-field. In this example, a 40 kb/s maximum rate (other parameters include average rate and delay) has been defined as the default QoS level. The subscriber in the example wants to use a more modest rate of 25 kb/s, but the network can only offer 15 kb/s due to the load on the network. The network confirms, whether the user accepts the lower-than-requested rate or not. If the user accepts, the transfer is begun with the 15 kb/s rate. Later, when capacity is freed, the network can offer a faster rate, up to the negotiated maximum value.

The QoS of the HSCSD service is related to the number of available time slots. Also the 9.6 kb/s and the 14.4 kb/s rates per time slot can be used with the HSCSD service. Thus the HSCSD QoS means that the network guarantees a minimum of one fixed 9.6 kb/s time slot. In contrast, delays may occur with GPRS where no time slot is guaranteed for the users.

The changes in QoS influence network planning (mainly capacity planning), if QoS requires more than one time slot per user.

20.8.5 Transmission

The most critical part of GSM transmission from the GPRS point of view is the *Abis*-interface. The current *Abis*-interface consists in most cases of the submultiplexed 16 kb/s blocks of the 64 kb/s time slots of the

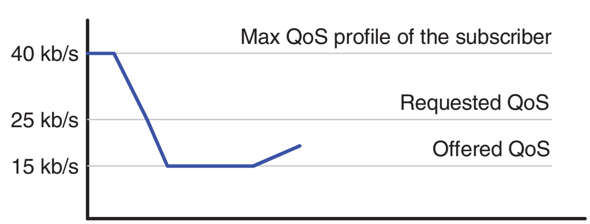

Figure 20.31 *An example of the QoS levels negotiated at the beginning and during the connection.*

PCM. Thus a transfer rate faster than 16 kb/s on the radio interface influences the way the transmission is defined. In principle, the entire 64 kb/s block can be specified for such a connection, but it is a waste of resources. Technically 32 kb/s blocks can also be specified, depending on the equipment.

This means that in practice, GPRS channel coding classes CS-3 and CS-4 cause changes to current transmission. The problem can be solved by redefining the PCM blocks and by adding transmission capacity, if necessary. In any case, the traditional TRAU will be modified, as the GSM network must be able to separate the circuit switched connections from the GPRS packet data. The PCU (packet control unit) separates and transfers the packets between the GSM and GPRS networks. The PCU is typically located with the BSC, but in principle, nothing prevents from locating the PCU with the base station. In such a case, all GPRS channel coding classes can be used, in principle.

The features of the GPRS trunk network elements also influence the transmission performance. Depending on equipment manufacturer, the SGSN and the GGSN can have different capacities for transmission. The typical SGSN and GGSN capacity is tens of megabits per second. As GPRS traffic increases, it is essential that the trunk network has sufficient capacity so that the trunk network will not become a bottleneck for transmission. In addition to the network, each individual element must be capable of handling and storing GPRS traffic without facing capacity problems, even after network upgrades. For example, sufficient capacity must be ensured for the HLR fields in connection with other network-related upgrades.

In addition to the above issues, the transmission must be secured. It depends on the equipment manufacturer, whether the elements have been secured in for example, the same way as the existing centers and BSCs, or if parallel elements are used.

20.8.6 The Effect of the Applications

The data rate, flawlessness, and delays are not equally critical issues for all applications. Table 20.29 is a preliminary list of typical applications, which can be used with different GSM data transfer methods [3].

Table 20.29 *The usability of different data transfer techniques (with respect to sufficient data rate) for some applications rated by importance 1–3. HSCSD can contain 14.4 kb/s and V.42bis-services. On the other hand, circuit switched services can utilize ISDN to make the establishment of the connection faster*

Application	9.6 kb/s	14.4 kb/s	28.8 kb/s	V.42bis	HSCSD (T) < 64 kb/s	HSCSD (NT) <64 kb/s	GPRS >64 kb/s
Email, offline	2	2	3	3	3	3	3
Email, online	1	2	2	2	2	2	3
Text file transfer	1	2	3	3	3	3	3
Bin file transfer	1	2	2	–	3	3	3
Telnet	3	3	3	3	3	3	3
Internet browsing							
-interactive	1	1	2	2	3	3	3
-one-way	1	2	2	2	3	3	3
Location	2	2	2	—	3	3	3
Video							
-surveillance, slow	2	3	3	—	3	3	3
-video conference	1	1	2	—	3	2	2
audio	1	2	3	—	3	3	2
audio/video	1	1	2	—	3	2	2

It seems difficult to define a most suitable application for each data transfer method, as new techniques give birth to new ideas. An example is the short messaging service, which today functions as a platform for a variety of value added services.

The transfer of video picture can potentially become a widespread technique with mobile terminals, especially when more efficient compression algorithms are developed. Slow surveillance video picture has been used relatively successfully even with NMT phones, so the transfer of surveillance video clips is also well suited to newer and faster techniques. Video conferencing requires a much faster data rate than 100 kb/s, and a high quality video conference might require up to several megabits per second. Also, it has been estimated that a delay of over 200 ms while transferring video significantly deteriorates the quality. The delay can, however, be balanced using application-level buffers.

In principle, GPRS suits all kinds of applications except delay-critical data transfer, which is required for example, for real-time video conferencing.

20.8.7 The Usability of GPRS Data

Typical GPRS applications focus on messaging including email, mainly used by business people. The variety of terminals have gradually improved and the terminals got cheaper, which brought GPRS into the reach of a larger audience. The range of applications will also expand, and an increasing number of capacity-intensive applications will enter the markets.

As also the basic GSM network continues to be relevant as the number of subscribers will not lower fast, it may make sense to estimate the percentage of GPRS users out of all GSM users. The concept of a GPRS user is more complex than that of a regular GSM user, as GPRS can be used for example as a solution for surveillance over a soda vending machine. GPRS is not as directly related to a physical person as GSM, as it can also be used for entirely machine-to-machine communication.

If a sufficient number of TRXs are available, GPRS capacity can in theory be offered without changes to the network. Estimates of the percentage of the total capacity used by GPRS are listed in Table 20.30. However, it is a bit problematic to estimate the capacity required by GPRS as reserving time slots during a certain period of time. A more accurate measure would be the total amount of data to be transferred during a single time period in a certain area.

One alternative for studying the problematic subscriber number issue is to calculate the maximum GPRS throughput, which can be guaranteed with the existing network. It can be assumed that there are 100 TRX / km^2 and more than 3 TRXs per cell in the downtown area of a large city. Similarly, it is assumed that in the area outside the downtown area there are 60 TRX / km^2 and 3 TRX / cell, and 25 TRX / km^2 and 2.5 TRX / cell in the suburbs. It is further assumed, that in rural areas, there are 1.5 TRX / cell, and the distance between masts is approximately 10 km.

Table 20.30 GPRS share of the cell's capacity as a percentage of time

TRXs	Traffic TSs total	Delivered traffic (Erl)	GPRS (% of the time)
1	7	2.88	58.8
2	14	8.04	42.6
3	21	13.7	34.7
4	29	20.6	29.0
5	36	26.8	25.7
6	44	34.0	22.7

Table 20.31 *The average number of GPRS time slots available per cell during peak hours for certain TRX numbers in downtown areas, downtown vicinity, suburbs, and rural areas*

				GPRS-TSs/cell		
Environment	cells/km^2	TRXs/km^2	TRXs/cell	case 1	case 2	case 3
Dense urban	28	100	≈3.5	7.5	10.9	8.3
Urban	20	60	≈3	7.0	9.5	7.8
Suburban	10	25	≈2.5	6.4	8.5	7.2
Rural	0.01	0.02	≈1.5	4.9	6.1	5.7

Based on the studies described earlier, we can assume GPRS throughput to range between 40 and 70%, depending on the application. According to the results presented previously, the theoretical maximum transfer rate of a single GPRS cell with 100% throughput is 12.4–16.5 kb/s per time slot, taking into account the cell's entire functioning area (when all GPRS channel coding classes are in use). Furthermore, with the additional information on how many GPRS time slots can on average be used during the peak voice hour, we get an estimate of GPRS transfer rate in different environments. These results are shown in Table 20.31. It has been assumed that the GPRS functionality has been specified for all TRXs in each cell.

As mentioned earlier, case 1 (GPRS-TRXs with no time slots fixed for GPRS) does not impact on capacity planning, and is thus the recommended alternative for the first phase. Case 3, where one time slot is fixed for GPRS, affects the voice traffic blocking probability so that the new probability for a cell with three TRXs is 3% or less. If the increase in the blocking probability is tolerated and accepted, case 3 is a suitable alternative for downtown area cells, but not necessarily recommended for suburbs and rural areas.

Table 20.32 can be used to estimate GPRS transfer rates by cell in different environments using the 40 and 70% throughput values shown in Table 20.12. Ref [2] shown in these Tables stands for the theoretical maximum rate of 16.5 kb/s, and [8] stands for the basic 12.4 kb/s rate of a time slot with 100% throughput. The values in the table have been calculated for case 1, where the B = 2% voice traffic blocking probability remains unchanged. The values have also been calculated for case 3 in the cases where the blocking probability $B \leq 3\%$.

Table 20.33 has been used to calculate GPRS traffic by square kilometer in different environments.

Table 20.32 *GPRS traffic transmission rates by cell. The Ref. "[2]" represents the 16.5 kb/s data rate, and Ref. "[8]" represents 12.4 kb/s*

		GPRS data throughput of cell (kbit/s)			
		Voice traffic B = 2%		Voice traffic B ≤ 3%	
Environment	Throughput	[2]	[8]	[2]	[8]
Dense urban	40%	48.0	36.3	53.1	40.2
	70%	84.0	63.5	93.0	70.3
Urban	40%	44.8	33.9	49.9	37.8
	70%	78.4	59.3	87.4	66.1
Suburban	40%	41.0	31.0	—	—
	70%	71.7	54.2	—	—
Rural	40%	31.4	23.7	—	—
	70%	54.9	41.5	—	—

Source: Data by courtesy of ETSI and Paulo Zanini.

Table 20.33 *The theoretical GPRS traffic throughput per km², when all GPRS channel coding classes are used*

		GPRS data throughput (kbit/s/km²)			
		Voice traffic B = 2%		Voice traffic B ≤ 3%	
Environment	Throughput	[2]	[8]	[2]	[8]
Dense urban	40%	1344	1016	1487	1125
	70%	2352	1779	2603	1968
Urban	40%	896	678	998	755
	70%	1568	1186	1747	1321
Suburban	40%	410	310	—	—
	70%	717	542	—	—
Rural	40%	0.31	0.24	—	—
	70%	0.55	0.42	—	—

Table 20.33 states the available GPRS capacity during the voice traffic peak hour. The table can be used to estimate the number of GPRS users assuming a certain average capacity requirement per user. The estimate is bound to be quite vague at this point, as the capacity requirement depends on the applications used.

One estimation method is to make assumptions on light, medium, and heavy GPRS usage during peak hours. "Light" usage would consist of telnet-type of traffic, email without large attachments, requiring less than 100 kbits of capacity. "Medium" applications require the transfer of small files and slow surveillance video, requiring less than 500 kbits of capacity. Videoconferencing, and the browsing of Internet pages containing pictures would be classified as "heavy" usage, when more than 500 kbits would be transferred per hour.

The amount of GPRS users could be estimated with the help of the applications. Each user requires a certain share of the capacity during the peak hour. Tables 20.34 and 20.35 contain some examples and the potential number of users in different environments.

As only channel coding classes CS-1 and CS-2 was available for first-phase GPRS, the "worst case scenario" for transfer rates is also evaluated. Table 20.36 contains calculations on the maximum number of users assuming an average transfer rate of 9.54 kb/s per GPRS time slot.

The values in the above tables describe situations, where the capacity of the network is strained during peak hours. This means that in principle each GPRS user has one hour for data transfer. In practice such a situation is quite impossible, which means that the values in the tables are, at best, intelligent "guesstimates."

The tables also give indication of the theoretical maximum number of GPRS users per GPRS cell. In practice, the number of users a single GPRS-TRX can serve is limited due to the channel allocation algorithm

Table 20.34 *The maximum number of GPRS users per square kilometer, compared to the theoretical transfer rate of 16.5 kb/s per time slot ([2]) with 2% voice traffic blocking probability*

			Number of GPRS users/km²			
Application	kbits/user/busy hour	Through-put/%	Dense urban	Urban	Suburban	Rural
Light	100	40	48384	32256	14746	11
		70	84672	56448	25805	20
Mid	500	40	9677	6451	2949	2
		70	16934	11290	5161	4
Heavy	4000	40	1210	806	369	0.3
		70	2117	1411	645	0.5

Source: Data by courtesy of Paulo Zanini.

Table 20.35 *The maximum number of GPRS users per square kilometer, compared to the theoretical transfer rate of 12.4 kb/s per time slot ([8]) with 2% voice traffic blocking probability*

			Number of GPRS users/km^2			
Application	kbits/user/busy hour	Through-put/%	Dense urban	Urban	Suburban	Rural
Light	100	40	36 590	24 394	11 151	9
		70	64 033	42 689	19 515	15
Mid	500	40	7318	4879	2230	2
		70	12807	8538	3903	3
Heavy	4000	40	915	610	279	0.2
		70	1601	1067	488	0.4

Source: Data by courtesy of ETSI.

and the functioning of the USF. According to the estimates presented in Ref. [3], the number of users is ranging between 20 and 40 per basic GPRS cell. Source [3] concludes that the throughput for the data traffic of one user grows faster than linearly as a function of the simultaneous GPRS time slots.

When comparing the numbers of GPRS users to the users of basic circuit switched voice services, it can be assumed that the duration of an average call is 2 min / user. This means, that in theory, one time slot can serve 30 users per hour. Table 20.37 contains the numbers of time slots available for voice services during peak hours, assuming a 2% blocking probability. The offered traffic can be calculated using the Erlang B formula, which gives the number of users for the voice service per square kilometer.

The GPRS applications have a significant effect on the network load and dimensioning. If GPRS were used mainly for the transfer of small amounts of data, the existing dimensioning of the network is sufficient, even if the number of users was one or two decades times the number of basic GSM subscribers. If the amount of data transferred via GPRS were measured in megabits, the network would probably already be saturated when the GPRS users make up a small percentage of all users.

If there is sufficient extra capacity in the network, GPRS users can use multiple time slots simultaneously. The closer the network is to its theoretical point of saturation, the slower are the connections experienced by individual GPRS users. Each operator needs to decide how much spare capacity it wants to keep in the network. If an operator wants to offer good quality service, a significantly larger margin than the theoretical maximum should be added to the estimated GPRS peak hour load.

Table 20.36 *The maximum number of GPRS users per square kilometer, compared to the theoretical transfer rate of 9.54 kb/s per time slot ([8], when only CS-1 and CS-2 are available) with 2% voice traffic blocking probability*

			Number of GPRS users/km^2			
Application	kbits/user/busy hour	Through-put/%	Dense urban	Urban	Suburban	Rural
Light	100	40	28 829	19 224	8784	7
		70	50 501	33 624	15 372	12
Mid	500	40	5766	3845	1757	1
		70	10 100	6725	3074	2
Heavy	4000	40	721	481	220	0.2
		70	1263	841	384	0.3

Source: Data by courtesy of ETSI.

Table 20.37 An estimate on the number of users of voice services in the different regions during peak hour, assuming a call duration of 2 min

Environment	TRXs	GPRS TSs	Voice TSs/cell	Offered load/cell	Offered load/km^2	Users/busy hour/km^2
Dense urban	3.5	7.5	17.5	17.2	482	14 460
Urban	3.0	7.0	14	13.7	274	8220
Suburban	2.5	6.4	11.1	10.9	109	3270
Rural	1.5	4.9	5.6	5.5	0.06	1.8

20.9 GSM/GPRS Measurement and Simulation Techniques

20.9.1 GPRS Measurement Devices

Some GPRS terminals are reliable enough to be integrated into measurement devices. The equipment can typically be used to record and monitor GPRS radio network occurrences and parameter values. The equipment can consist of a data storage unit such as a laptop PC, connected to a commercially available GPRS terminal. The connection is established for example via the public Internet to the server, and data transfer is monitored in both directions. A separate data unit accessed via GPRS can potentially be connected to the Internet network. The closer the data unit is to the GGSN element of the studied network, the more reliable are the achieved measurement results. This is because the potential bottleneck-effect of the Internet network should be eliminated. Thus, data can be transferred from the data unit to the measurement device in a controlled fashion to measure issues such as connection time to the GPRS network and the Internet, the data transfer rate, and the variance in channel coding classes and multislot usage. The measurement equipment can also contain satellite positioning equipment such as the GPS (Global Positioning System), which allows studying the measurement data spatially.

Figures 20.32 and 20.33 show some screens from the GPRS measurement device (Anite) when a separate data unit is used. The examples show situations where the throughput of the LLC- and RLC-layers has been evaluated both uplink and downlink. In this case, the measurement terminal supports 3+1 time slots (DL+UL). The results show the variance in the data rate during a connection; in this example, the rate varied between 0 and 40 kb/s.

20.9.2 The C/I Measured from the Network

The performance of the GPRS and other data services can be estimated for example as the actual throughput of the data as a function of the *C/I* (signal/interference) and the *Eb/N0* (energy-noise-ratio). The results of theoretical simulations can be calibrated to better represent reality by estimating the C/I- and Eb/No-values of actual networks.

The C/I-relationship can be investigated from the network or a part of the network, which is clearly interference limited. A dense urban network is most probably constrained by interference. By measuring the interference distribution of the network (Q-classes) and by using simulation results from e.g. Ref. [11], a connection can be established between the Q and the *C/I*.

It makes the most sense to study the *Eb/N0*-values in a noise limited network, such as a sparse rural network. The *Eb/N0*-values are studied by measuring the field strength levels of the network.

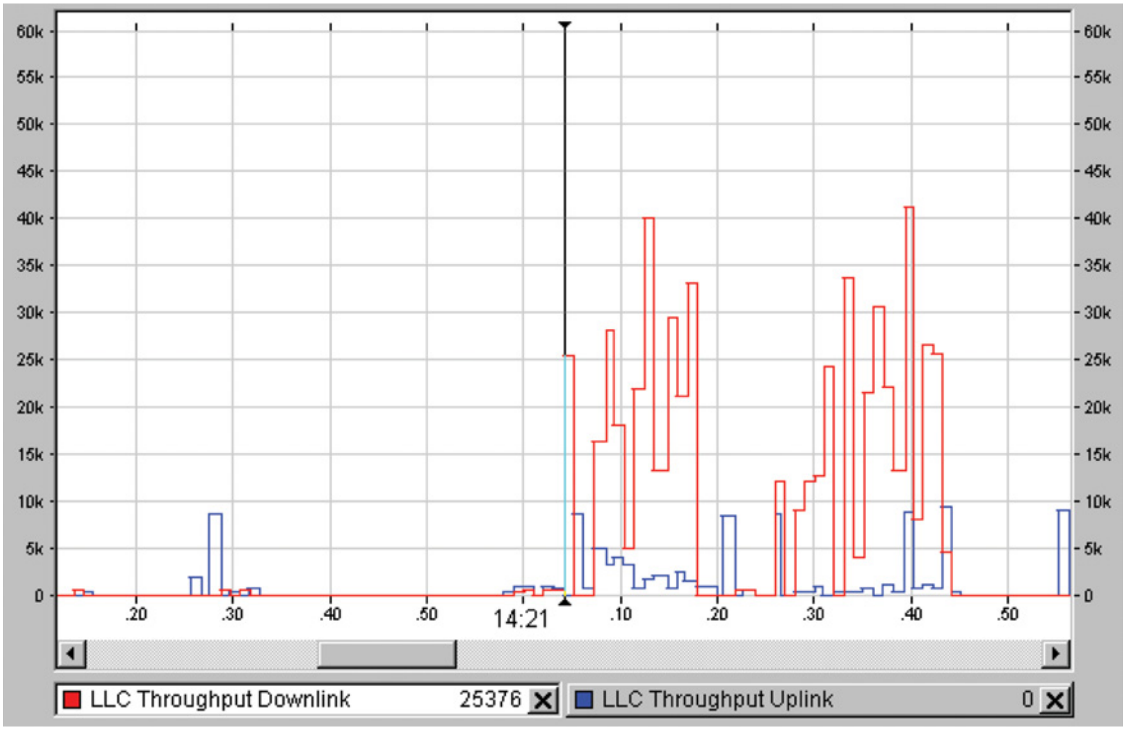

Figure 20.32 *An example of the throughput of the LLC layer of a GPRS connection, measured with GPRS radio equipment (Anite).*

20.9.2.1 RxLevStatistics

Source [3] describes studies done on both rural and urban networks using the "Rx Level Statistics"-tool (found in the operations and maintenance center of the GSM network) in order to obtain field strength and interference levels. The tool can be used to collect the quality level (Q-level) values 0–7 of the traffic channels in a matrix format, and also, the received power level values divided into six regions. This measurement used Rx Level region values 10, 15, 20, 30, 40, and 63, which correspond to values −110 – −100, −99 – −95, −94 – −90, −89 – −80, −79 – −70, and 69 – −47 dBm. The values were collected into a file with 480 ms intervals.

Ten BCCH-TRXs and 8–10 "secondary TRXs" (i.e. TRXs, with no BCCH time slot) from both interference-constrained urban environments and coverage area-constrained rural networks were chosen for the measurements. In both cases, the RxLevel- and Q- values of the traffic channels were monitored for one hour during a peak hour (3–4 p.m.) and at an off-peak hour (6–7 p.m.). All of the studied cells presented her, including urban and rural cases, used uplink power control and uplink-DTX. None of the cells used frequency hopping.

One downside of the Rx Level Statistics-tool is its crude field power scale. When the GSM network measures the power level with a 64-scale, the Rx Level Statistics-tool saves the values in a table with only a 6-scale. The scale can be "expanded" to a 64-scale with a relatively low error margin (see Figure 20.34) by assuming an even spread of the values within each field power region. The examples chosen for the figure are averages of the downlink and uplink TRXs. The assumption of an even spread is naturally not valid in reality, but the field power distribution cannot be studied more accurately with a single round of measurements.

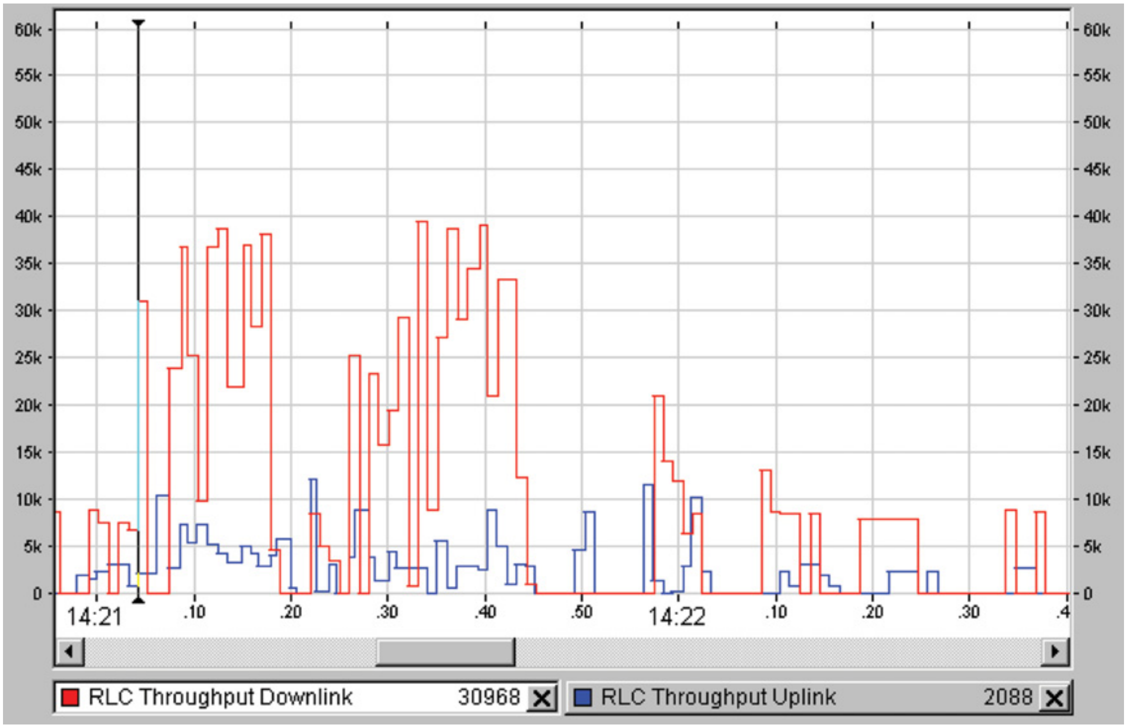

Figure 20.33 *An example of the throughput of the RLC layer of a GPRS connection (Anite, GPRS radio measurement equipment).*

It can be assumed that the bit error relationship, uncovered with the help of the Q-parameter, has the most significant influence on an interference-constrained network. The Q-values can only be saved using a scale with 8 values. In a well-designed network, most of the values collect in Q-class 0.

It can be seen from Figure 20.34 that the urban networks give fewer of the weakest values than do the rural networks. This is due to the abundant coverage overlaps of the urban cells. A low field power level – one which is close to the sensitivity level of the terminal – naturally increases the bit error rate, which means that the principle constraint in rural areas is the signal–noise relationship. The correlation of the field power level and the *Eb/No* is thus evident in rural areas. The cellular network is denser in cities. The planned street level value in cities can be for example, –75 dBm, compared to values of even –85 dBm in rural areas. In the urban areas, it is mostly the *C/I* which correlates with the field power level.

Source [11] describes simulations done on the correlation between the GSM quality level Q and the *C/I*. According to the results, the correlation between the Q-value and the *C/I* is: Q0 = 22 dB, Q1 = 20 dB, Q2 = 18 dB, Q3 = 15 dB, Q4 = 13 dB, Q5 = 9 dB, Q6 = 5 dB, and Q7 = 0 dB. The TU3 model with ideal frequency hopping was used in the simulations. The cumulative %-shares of the Q-classes corresponding to the *C/I*-values (in other words: the percentage of the studied time period when the service level in question or a better quality level has been achieved) are shown in Figure 20.35. The correlation between the *C/I* and the Q-class can probably be represented as a function as shown, but the function must take into account the inaccuracy factor, which depends on for example, the environment.

In the urban and rural environments shown in Figure 20.35, a minimum value from the uplink and downlink TRXs has been chosen to represent the worst case quality level. This is the minimum value achieved in each

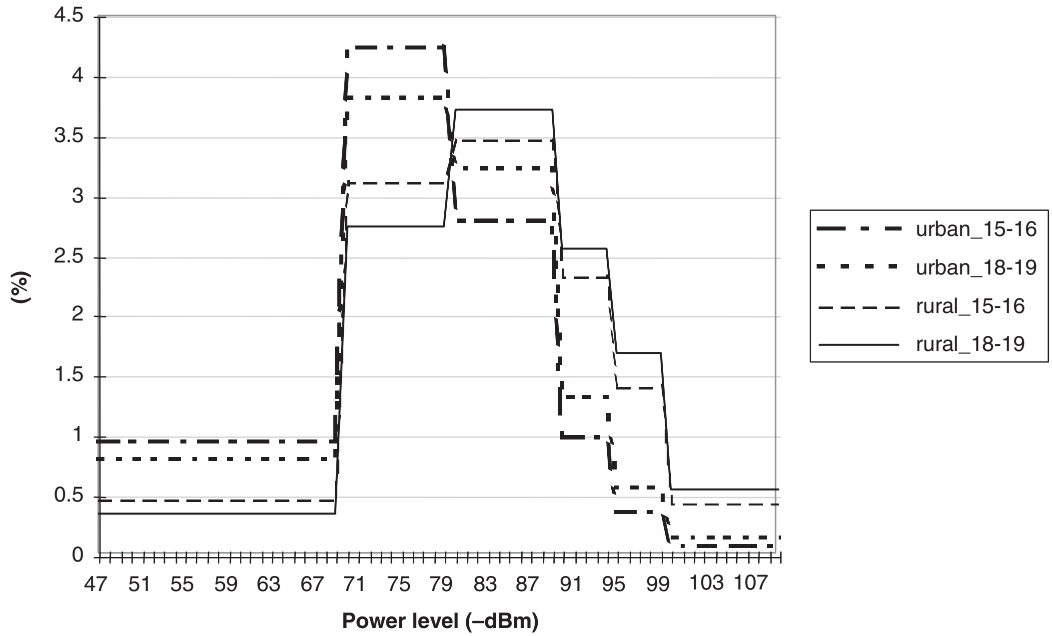

Figure 20.34 An example of the received field power distribution in urban and rural environments.

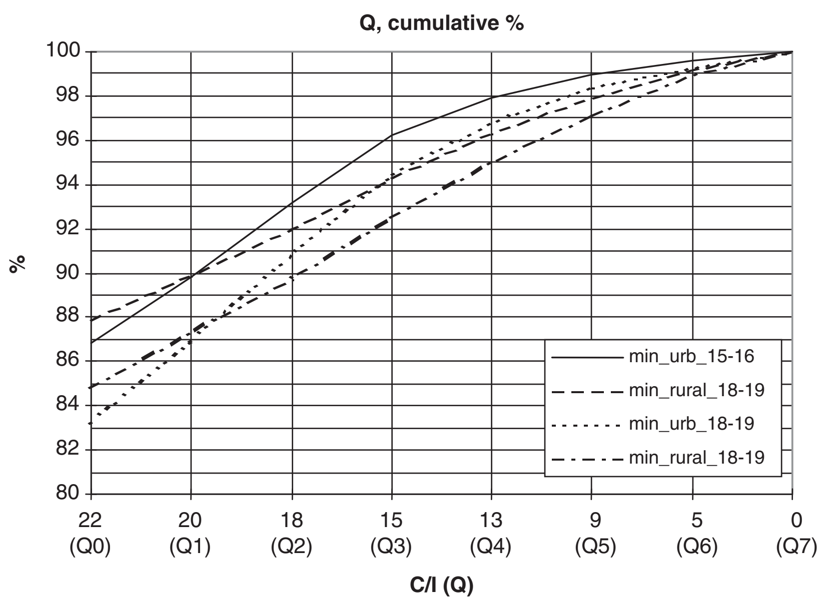

Figure 20.35 The cumulative shares of the minimum quality levels achieved in both urban and rural environments based on the RxLevelStatistics measurements. Note: the C/I scale is nonlinear.

Table 20.38 *A comparison of field equipment and RxLevelStatistics measurement results (received power). The results are shown as percentages. The field equipment results have been achieved using a single GSM phone as the measuring device. RxLevelStatistics-results consist of the measurements of all calls in the studied area*

	Urban		Rural	
RxLev/dBm	Field meas.	RxLevStatistics	Field meas.	RxLevStatistics
−60	27	10	9	5
−60 – −80	65	63	56	45
−81 – −93	8	24	30	39
−94 – −98	0	2	3	7
−99 – -104	0	1	1	3
<−104	0	0	1	1

environment. The *C/I* value in the figure applies also when frequency hopping is used. As the studied cells did not utilize frequency hopping at the time of the measurements, it can be assumed that actual *C/I*-values are a few dB weaker. This means, that for example, the C/I requirement value corresponding to Q-value 4 changes from 13 dB to 15–16 dB.

Some shift in timing between the rural and urban environments is evident in Figure 20.35. However, it can be stated, that the shift is relatively meaningless.

20.9.2.2 Field Measurements

RxLevStatistics can be used to study the average quality level of the network in a relatively reliable statistical manner. However, it makes more sense to use field equipment to monitor the quality level of a single call. Such field equipment can be formed with a test phone and a data storage unit connected to it. The equipment can be used to set up and break up calls at desired intervals, resulting in for example, success rates for calls and channel switches, the duration and area of the service, received power levels (RxLev) and quality level (Q) downlink. Equipment placed in a car was used in simulations described in Ref. [3]. During the measurements, the phone was attached to the car's dashboard. The route consisted of the area of the studied cells. The RxLev-values used in the measurements were chosen using a method somewhat different from the measurements described in the previous chapter in order to get comparable results from the RxLev- and field equipment measurements as presented in Table 20.38.

For comparison purposes, a series of experiments was made where rural and urban areas were studied with both field equipment and the RxLevStatistics equipment. The experiments were made on a working day with normal traffic before peak hours. Both experiments were made at the same time and in the same cells. As the field equipment can only measure downlink, the same direction was studied with the RxLevStatistics equipment. Also the same RxLev-limit values were used in both experiments. The cells did not utilize frequency hopping at the time of the experiments, which means that a weakening of approximately 2–3 dB can be assumed.

20.9.3 Conclusions

The results of the RxLevelStatistics are clearly better than the results of the field experiments. The RxLevel-Statistics measurement is statistically very reliable, as it continuously measures the quality levels of all active connections.

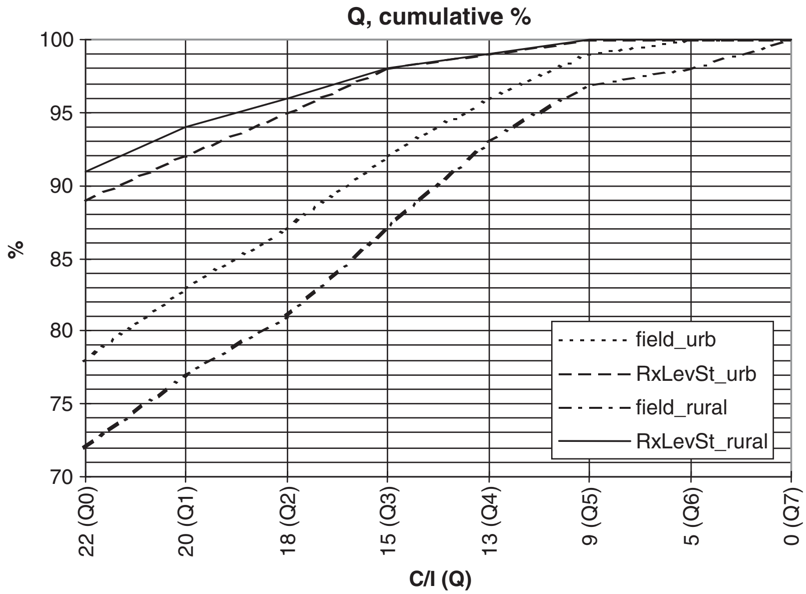

Figure 20.36 *A comparison of the results on studies of the cumulative, %-age share of time of Q achieved using the field equipment and the RxLevStatistics. Note: the C/I-scale is non-linear.*

Field experiments give a pessimistic view of the network quality, as the phone is located on the car's dashboard, causing a very unequal reception radiation pattern. In addition, the measurements are done while the car is constantly moving, and the outer areas of the cells are visited on purpose. The field experiments thus describe the worst case scenario, as cells have, on average, been placed in areas with high traffic. In contrast, RxLevelStatistics can be used to find out, if the cells of the network have been constructed in areas, where the majority of the traffic is.

The result indicates the network's QoS level that the GPRS users experience, for example, in moving environment. Data equipment located in the car is increasing area of interest due to the increased popularity of interactive location based services integrated to the car.

Let's assume that the correlation between the *C/I* and the Q described in Ref. [11] apply so that an additional margin of 2–3 dB must be added to the values, if frequency hopping is not used. This means, that the practical situation can be evaluated with the help of the conclusions of theoretical considerations. Based on Figure 20.36, we can choose a static (corresponding to the results of the RxLevStatistics experiments) and mobile (corresponding to the results of the field equipment) data profile in both urban and rural environments. The functioning areas of the GPRS classes can be "calibrated" to the real network, resulting in the actual data rate experienced by the GPRS users.

It makes sense to study the *C/I* requirement level using a 10% BLER-limit value, as this is the value defined in the specifications. For comparison purposes, the *C/I* values of the specifications and the *C/I* values (10% BLER-value) were achieved as results of the CSELT simulations. The percentage share of time when a certain *C/I*-value is achieved can be seen from Figure 20.35. The results are shown as Table 20.39.

According to the RxLevelStatistics experiments, field power levels required by CS3-4 are achieved more than 93.5% of the time in both urban and rural areas, when no frequency hopping is used. A mobile GPRS user is within the influence area of CS3-4 85.0% of the time in urban environments, and 79.1% of the time in

Table 20.39 *The percentage share of time of each GPRS channel coding class in data connections, based on the quality level distribution of the GSM network. The table also contains the time, when the C/I level of no class is achieved*

				C/I is achieved (% of the time)			
				Static MS (RxLevStatistics)		Moving MS (field measurement)	
Channel coding scheme			*C/I* requirements (dB)	Urban	Rural	Urban	Rural
Outage time	FH	[2]	<7.1	0	0	0.5	2.5
		[8]	<9	0	0	3.0	1.0
	noFH	[2]	<10.8	0.5	0.5	2.4	4.7
		[8]	<13	1.0	1.0	4.1	7.0
CS-1	FH	[2]	7.1	100	100	99.5	97.5
		[8]	9	100	100	97.0	99.0
	noFH	[2]	10.8	99.5	99.5	97.6	95.3
		[8]	13	99.0	99.0	95.9	93.0
CS-2	FH	[2]	11.5	99.3	99.3	97.0	94.4
		[8]	13	99.0	99.0	95.9	93.0
	noFH	[2]	12.8	99.1	99.1	96.0	93.2
		[8]	15	98.0	98.0	92.0	87.2
CS-3	FH	[2]	13.6	98.7	98.7	94.6	91.0
		[8]	15	98.0	98.0	92.0	87.2
	noFH	[2]	13.7	98.6	98.6	94.5	90.8
		[8]	16	97.0	97.3	90.3	85.0
CS-4	FH	[2]	20.8	90.6	92.7	80.9	74.8
		[8]	23	87.5[a]	89.5[a]	76.0[a]	70.0[a]
	noFH	[2]	17.2	95.7	96.5	88.4	82.8
		[8]	19	93.5	95.0	85.0	79.1

[a] Value by extrapolation.

rural environments. Figure 20.37 shows the share of time of each channel coding class in the section of the cell being measured, when the study is made within the limits set by the specifications, and without frequency hopping.

The share of time is not directly proportional to the surface area. The above-described calculations based on real-life situations do, however, show that the measured areas of the network have been constructed taking into account the high traffic areas. As the GPRS service is used in the same places as the existing voice and data services are used, it seems that the measured areas are well suited for GPRS users.

Based on Figure 20.37 is would seem that CS-4 – and thus, the fastest data rate – could be used almost constantly in the measured area of the network, if only it were possible to use CS-4 in actual networks. On the other hand, if it is only possible to use CS-1 and CS-2 in real-life networks, it seems that CS-2 is used almost constantly in the studied area.

As the default BLER-value in the figure is 10%, some re-sending takes place, lowering the actual throughput somewhat. On the other hand, the studied network did not carry the extra load – and additional interference – of GPRS, which in reality weaken the results in Figure 20.37, depending on the number of optimization methods intended for voice.

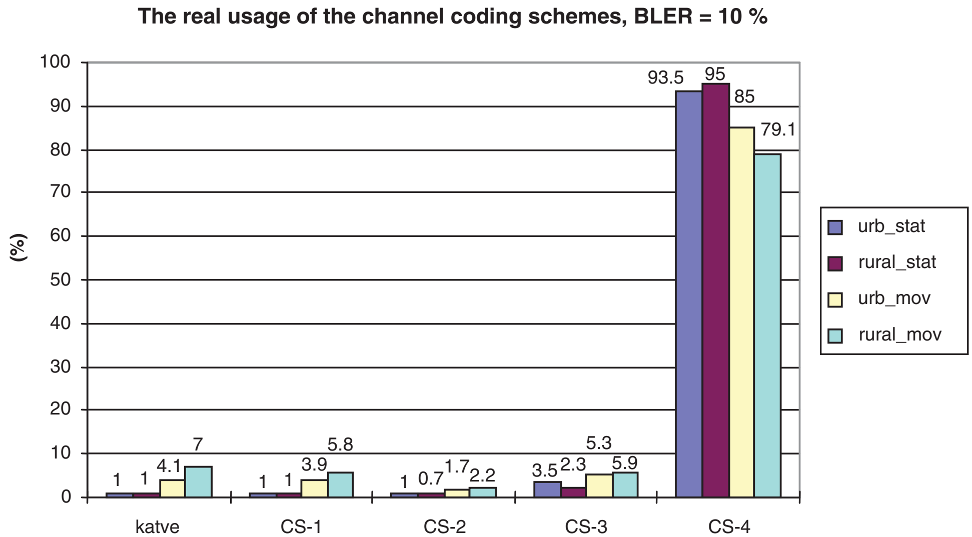

Figure 20.37 *The share of time of each channel coding class in an actual network, when BLER equals 10%. The values are compared to the C/I values set by the specifications, without frequency hopping. The outage is the period of time, when the C/I value of no channel coding class is achieved.*

20.10 Simulations

20.10.1 Interference Level Simulator

Source [3] describes a simulation tool, which can be used to study the downlink direction *C/I*-level behavior according to a GSM network based on the theoretical hexagonal principle with different GPRS loads.

The functioning of the simulator is shown as a flow chart in Figure 20.38. As a result of the simulations, the downlink *C/I*-distribution can be found by varying for example, the TRX levels of the cells, the GPRS load, and the size of the recurrence patterns. The idea behind the simulator is to create traffic, which resembles actual GSM traffic as much as possible to each cell. In these simulations, a full power level was specified for the downlink direction of the BCCH frequencies regardless of the traffic levels. Discontinuous transmission (DTX) was used for the other frequencies. Thus the free time slots on frequencies other than the BCCH frequencies did not cause interference; only the ones used for traffic purposes did. Power level control and frequency hopping were not specified for the simulations.

20.10.2 Simulation Considerations

One problem with interpreting the results of publicized simulations is the way the chosen model and its many parameters influence the results. Traffic models can be developed by studying the data traffic profiles of existing systems. For example, studies made on a company's email or Internet traffic in its fixed network or the GSM network could conclude that GPRS users are expected to behave similarly in a "mobile office environment." However, it should be kept in mind that using the fixed network is very different in nature from using a mobile network, and the results of such studies may not necessarily apply directly to real situations in GPRS networks. In any case, these studies give us some idea of the average length of the data packets and peak times.

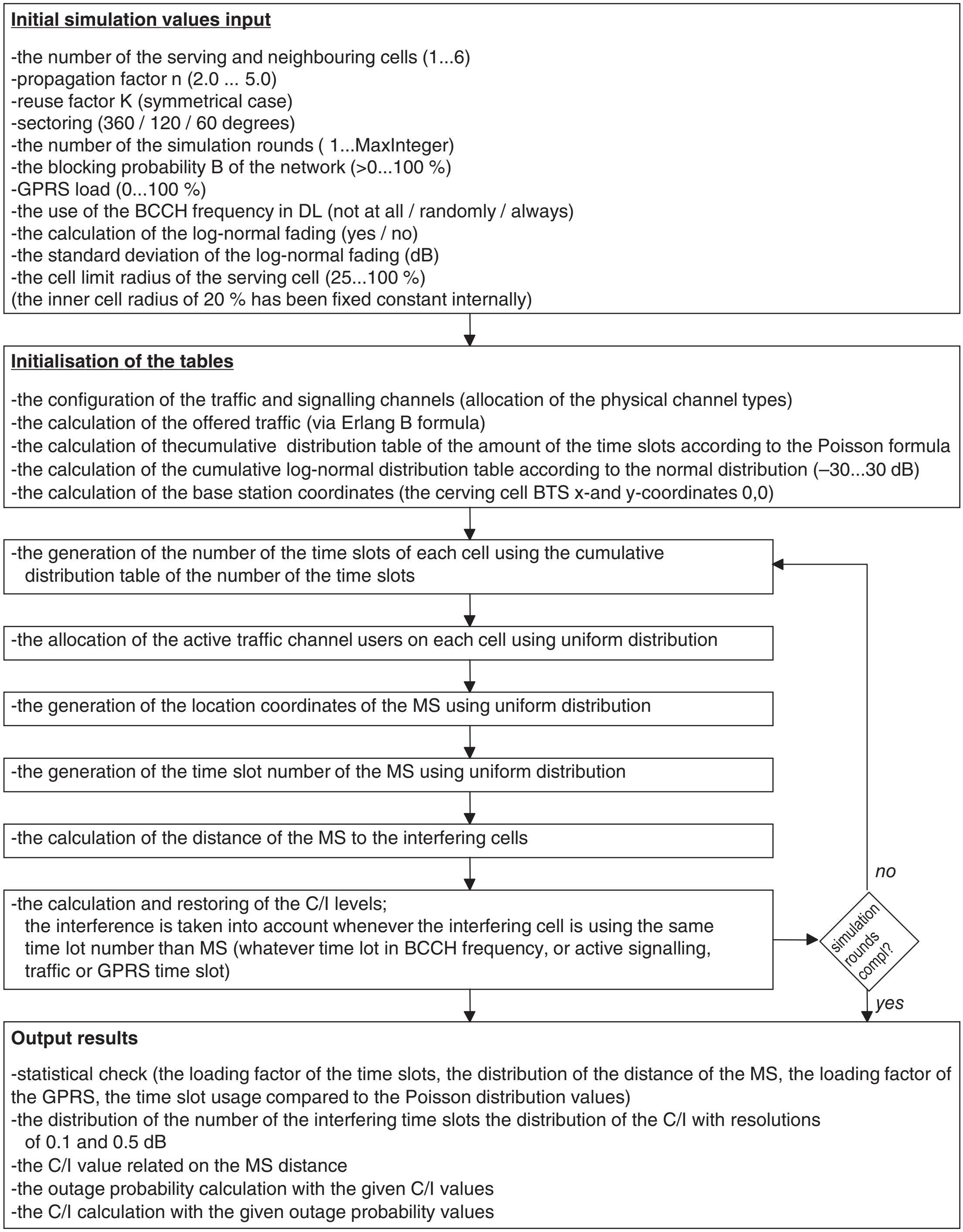

Figure 20.38 *A flow chart of the simulation tool as presented in Ref. [3].*

Other factors influencing the results of simulations are for example, the environment types of the GSM 05.05 –specifications (typical urban TU, rural area RA, etc.), and the use of frequency hopping and power level control. The simulations can be made on different protocol stack levels on the radio interface, or in the entire network. Interesting topics for simulations include throughput, access and data transfer delays, the growth of outage probability as a function of the interference level, and the point of saturation of the system. A common problem is the lack of sufficient background information on the simulation. For example, data throughput simulations may include the implicit assumption that the header field of the RLC-frames is included in the user's data.

Several assumptions can be made to simplify the creation of a simulation environment. For example, it could be decided that the data throughput of the MAC protocol on the RLC/MAC-layer is ideal, which means that throughput $T = (1\text{-}BLER)*R$, where BLER is the block error rate, and R is the channel's bit error rate.

Simulations are always theoretical assumptions made on real situations, and all results should be interpreted with caution.

20.10.3 Simulations: Example

In the example shown here, the network is constructed according to the hexagonal principle, taking into account the six closest interfering base stations. The additional interference caused by base stations farther away is less than 1dB, so they have been excluded from the simulation so that the duration of the simulation is not unnecessarily extended. Figure 20.39 shows the principles of the network to be constructed.

The cells are circular in shape, and they overlap partly so that the intersections are at the corners of the hexagons. The overlapping areas can be considered as areas, where cell reselection takes place.

The following formula can be used to calculate the C/I in a hexagonal model:

$$\frac{C}{I} = \frac{R^{-n}}{\sum_{k=1}^{K_0} D_k^{-n}} \approx \frac{R^{-n}}{6D^{-n}} = \frac{Q^n}{6}, \tag{20.38}$$

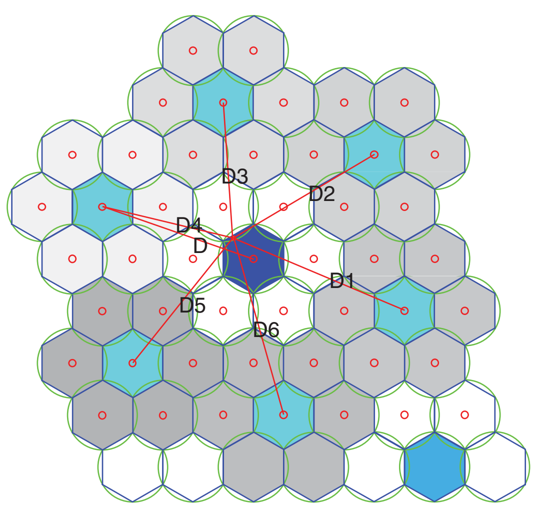

Figure 20.39 *The hexagonal model used in the simulations. The six closest interfering cells are taken into consideration.*

where C is the carrier, I is the interference, R equals cell radius, and D is the distance of the reuse pattern. The theoretical C/I – relationship can be calculated with the model assuming that radio waves advance purely according to the exponential law. The value $Ko = 6$ is due to the number of the closest neighboring cells with the same channel in the hexagonal model; other channels can be excluded. The term Q is the reduction factor of the cochannel interface – the greater is the value of Q, the smaller is the level of the cochannel interference. The term n is the fading multiplier, which depends on the environment. Typical values are 2–5. The fading multiplier can be calculated using the formula

$$L = L_0 + 10n \log(d), \tag{20.39}$$

where L equals the path loss attenuation in dB, Lo is the fading in the starting point, and d is the distance. In practice, $n = 2$ stands for open space, $n = 4$ means complete reflection from the ground, and $n = 5$ stands for an environment which has a strong fading effect.

The reuse model multiplier that is, the number of frequencies can be represented with the term K. Because in a symmetrical hexagonal model $D = R\sqrt{3K}$, C/I equals

$$\frac{C}{I} = \frac{Q^n}{6} = \frac{\left(\sqrt{3K}\right)^n}{6}. \tag{20.40}$$

The signal–interference relationship can also be calculated assuming the worst case scenario, when the distance D of the recurrence model is D-R, and

$$\frac{C}{I} = \frac{\left(\sqrt{3K} - 1\right)^n}{6}. \tag{20.41}$$

The conclusion can be made based on this simple model, that for example, using the recurrence pattern $K = 7$ and value $n = 4$ gives in the worst case a value of 27.4 or 10log(27.4) = 14 dB, which in theory is sufficient as a shield ratio. In practice, $K = 7$ may be slightly too small, as the model does not take into account the normal log-fading.

Only the following values should be used for the multiplier K of a symmetrical recurrence pattern:

$$(k + l)2 - kl, \tag{20.42}$$

where k and l are positive integers. Thus for example, K values 3, 4, 7, 9 and 12 are allowed in recurrence patterns.

When the simulation is initiated, the recurrence pattern multiplier K is defined (symmetrical situations follow formula 20.42, so that $K = 1,3,4,7$ etc.) and the network blocking probability is calculated according to the Erlang B-formula. Each cell can have 1–6 TRXs. In the simulation, the antennas of the cells were defined to have a circular beam.

During the initiation, the offered traffic $\bar{x}$ of each cell is calculated by iterating using the Erlang B formula and the defined blocking probability. After initiation, GSM traffic is generated into each cell according to the Poisson distribution. Figure 20.39 shows an example of the cumulative traffic distribution in a cell with three TRXs with a blocking probability of 2%. On each round of the simulations, traffic is described so that the number of time slots used simultaneously follows the Poisson distribution, but the time slots are chosen from the cells according to a random equal distribution. Thus it is not possible in the simulations to divide GPRS traffic and voice traffic to the different sections of the cells formed by the blocks of time slots. This is a feature which will likely be defined into actual networks to enable multislot usage.

The simulator functions with a "snapshot" principle, that is, a snapshot is taken of the situation in the network on each round of simulations, but the successive snapshots are not time continuous. When the number of

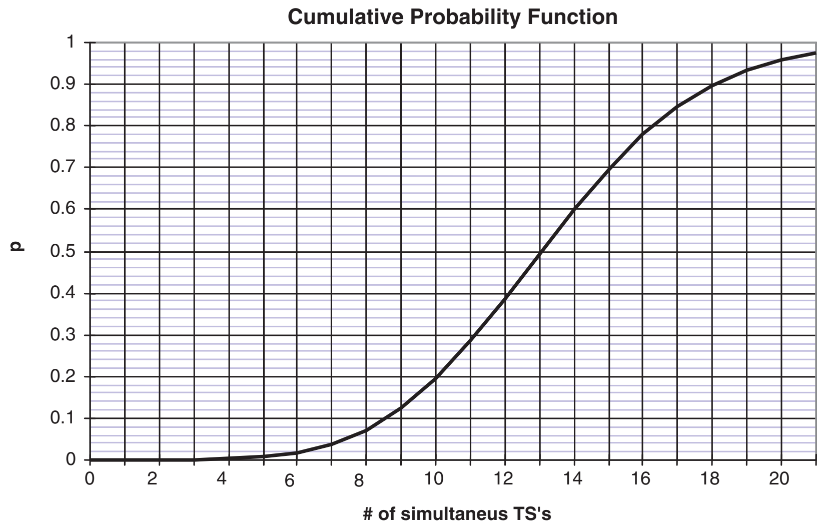

Figure 20.40 *An example of the cumulative probability function of the number of timeslots of a cell with 3 TRXs, calculated using the Poisson distribution and a 2% blocking probability.*

snapshots is sufficient, the distribution of the number of the voice time slots of each cell approaches the Poisson distribution. Figure 20.40 shows an example of the cumulative distribution.

Even if the duration of the traffic has not been modeled, the simulator can achieve a relatively realistic distribution for *C/I* according to the described principle.

In addition to GSM traffic, the GPRS usage ratio (range 0–100%) of the free time slots can be defined for each round of simulations. Each round of simulations describes a static situation, that is, the lengths of the calls or the profile of the GPRS traffic are not considered.

During each round of simulations, the position coordinates of the MS are generated with an equal distribution in the area of the serving base station. The minimum distance of the MS from the serving base station is constricted to 20% of the cell's radius, so that only the relevant *C/I* values are studied. The 20% restriction can be thought to represent a situation, where the antenna of the base station has been placed in a mast, so that the distance between the user and the antenna is at least the difference in height. According to the equal distribution, the 20% restriction means that 4% of the users are located outside of the center of the cell. Thus the impact on the final cumulative *C/I* distribution is insignificant.

When the location coordinates of the MS are known within the serving cell, the *C/I* relationship can be calculated on each round of simulations using a principle modified from formula 20.38, as shown in source [3]:

$$\frac{C}{I} = \frac{-10n\log(r) + L_c}{\sum_{i=1}^{6} h_i\left(-10n\log(D_i) + L_{I,i}\right)}, \tag{20.43}$$

where r is the distance of the MS from the serving base station, and D_i is the distance from a base station on the same channel i. L_c is the variance in signal strength of the serving base station caused by the log-normal fading in decibels, and $L_{I,i}$ is the same variance in the interference signal of a base station on the same channel i. The term h_i is used to represent the interference time slot as follows:

$h_i = 1$, when cell i uses the same time slot
$h_i = 0$, when cell i does not use the same time slot.

The log-normal fading terms L_c and $L_{I,i}$ are generated with a normal distribution, which depends on the standard deviation entered as a starting value. The default value is 6 dB, which is close to the results of measurements done in real GSM networks. The standard deviations in the simulations are independent of each other. This is also usually the situation in real networks, if the cells have been constructed symmetrically according to the hexagonal model. It is not necessary to model the Rayleigh fading, as it is taken into account in the protection ratio requirements of the GSM system. Thus the GSM system is expected to function with an average short-term *C/I*-relationship of 9dB.

The result is the network's *C/I* distribution between 0 and 40 dB. The results can be used to examine the relationship between the *C/I* and the outage probability. The simulator also calculates the *C/I* average distribution as the function of the distance of the MS from the serving cell, and the distribution of the number of simultaneous time slots with interference.

Each simulation consisted of 30 000 rounds. The *C/I* distribution calculated by the simulator was recorded in the dB table between 0 and 40 dB with accuracy of 0.1 dB. On average, 74 samples were generated into each cell of the table, which gives a relatively reliable simulation result considering the reliability requirements of the binomial distribution. Reliability can also be evaluated by comparing the traffic distribution generated by the simulator with theoretical values. After 30 000 rounds, the correlation between the values achieved in the simulations and the theoretical Poisson distribution spread of traffic channels was typically over 0.990. In addition, the usage of the time slots free from voice traffic with a 100% GPRS load was on average 34.7% in the simulated cells with 3 TRXs, which corresponds with the theoretical values.

20.10.4 Results

20.10.4.1 C/I as a Function of the Distance of MS

Figure 20.41 shows the *average C/I* as a function of the MSs distance from the serving cell. The variance in slow fading is omitted due to the use of averages, which means that the result can be interpreted so that 50% of the time, the value of the *C/I* is at least according to the figure.

The studied network was constructed using a 7 recurrence pattern so that each cell used three frequencies. The MS was defined to only use frequencies other than the BCCH frequency.[6] The figure takes into account the situation in a normal GSM network with a 2% blocking probability during peak hour, and the situation when the free time slots of the network are loaded with GPRS traffic with 25, 50, 75, and 100% usage.

It can be seen from the picture that the *average C/I* relationship on the edge of the cell is approximately 15 dB in a normal GSM network (GPRS_0%). GPRS traffic weakens the *C/I* relationship by 0–3 dB, depending on the load.

The previous simulation can be further developed by varying the size of the serving cell. After each simulation, the *C/I* value with 10% outage probability achieved on the edge of the cell can be examined. Figure 20.42 shows the results of simulations, where the radius of the cell has been varied between 25 and 100% on each round (each point represents a value achieved on the edge of the cell). Thus the figure shows

[6]If the MS used only the BCCH-frequency, the C/I-distribution would be exactly the same as in the simulation example with the 100% GPRS load.

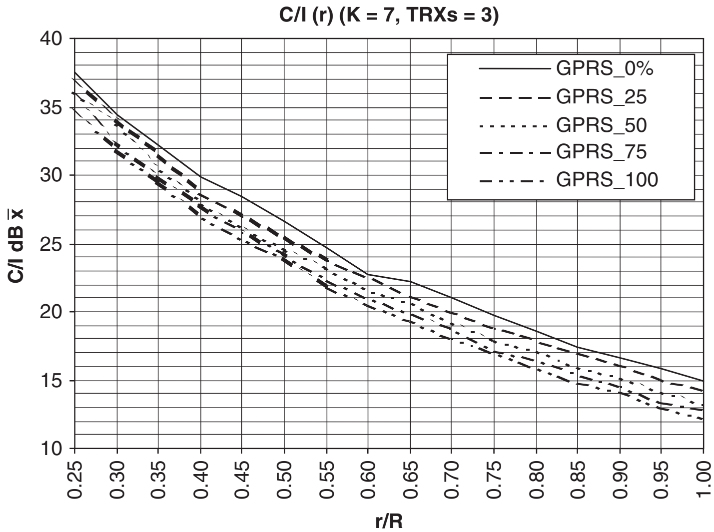

Figure 20.41 *The average C/I vs. the MSs distance r from the serving cell. The cell's radius is R.*

directly the relationship of the *C/I* relationship as a function of the distance from the serving cell, with 10% outage probability.

Figure 20.42 shows, that the simulated network barely meets the GSM specification *C/I*-relationship limit value of 9 dB with 90% location probability during peak hour (B = 2%). Also, the adding of 100% GPRS load weakens the *C/I* value by 2dB when compared to voice traffic peak hours. The usable coverage area for the voice traffic is also reduced to approximately 87% of the original surface area. The GPRS load naturally also affects the usable coverage area of the GPRS service.

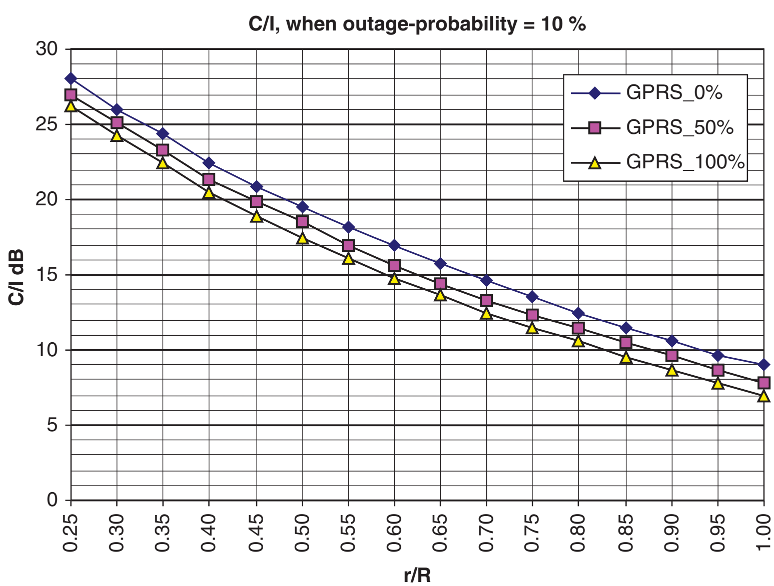

Figure 20.42 *C/I (10% outage probability) as a function of the distance from the cell.*

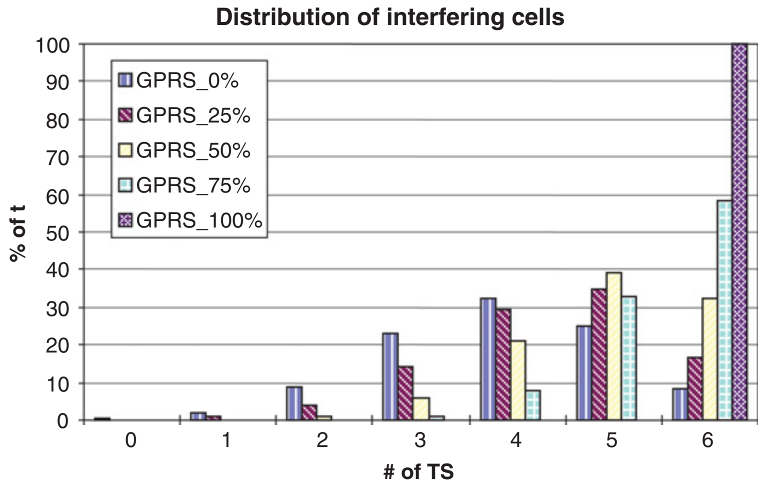

Figure 20.43 *The distribution of the number of interfering time slots with different GPRS loads.*

20.10.4.2 The Distribution of the Number of Interfering Time Slots

The interference level of the network is influenced by the existence of simultaneous interfering time slots. Figure 20.43 shows the distribution of the interfering timeslots with different GPRS loads, which appeared during the simulations.

It can also be seen from Figure 20.43, that in the simulated network without GPRS load (GPRS_0%, i.e. the situation during a normal peak hour with 2% blocking probability) the average number of interfering time slots was 4. GPRS load shifts the average so that the number of simultaneous interfering time slots is 6 with a 100% GPRS load. This would mean that all simultaneous time slots would interfere constantly.

20.10.4.3 The C/I Distribution as a Function of the GPRS Load

Figure 20.44 shows the cumulative *C/I* distribution with different GPRS loads (0.5 dB resolution). The studied cells remain on three frequencies, according to the 7 recurrence pattern.

It can be seen from the picture that the *C/I* distribution weakens as the GPRS load increases. Figure 20.45 shows a more detailed description of the situation where the *C/I* relationship is approximately 9 dB (0.1 dB resolution). 9dB is the limit set by the GSM specifications for the successfulness of GSM calls.

It can be stated by studying the 9 dB value that the simulated GSM network with a 7 recurrence pattern barely functions during a peak hour with normal traffic and a 10% outage probability. 10% is a typical dimensioning principle with GSM networks. When the GPRS traffic load is also added to the network, the *C/I* value is weakened by approximately 0–2 dB (10% outage probability), which means that a share of the voice and GPRS traffic is blocked. The situation can also be studied by examining the 9 dB value, which is met with outage probabilities 10–16%, depending on GPRS load.

20.10.4.4 The Impact of the Recurrence Pattern

The meeting of the 9 dB limit value can be studied as a function of the multiplier *K*. Figure 20.46 shows results of simulations with *K* values 7–19. The simulations were made with cells using three frequencies, and the MS never used the BCCH frequency.

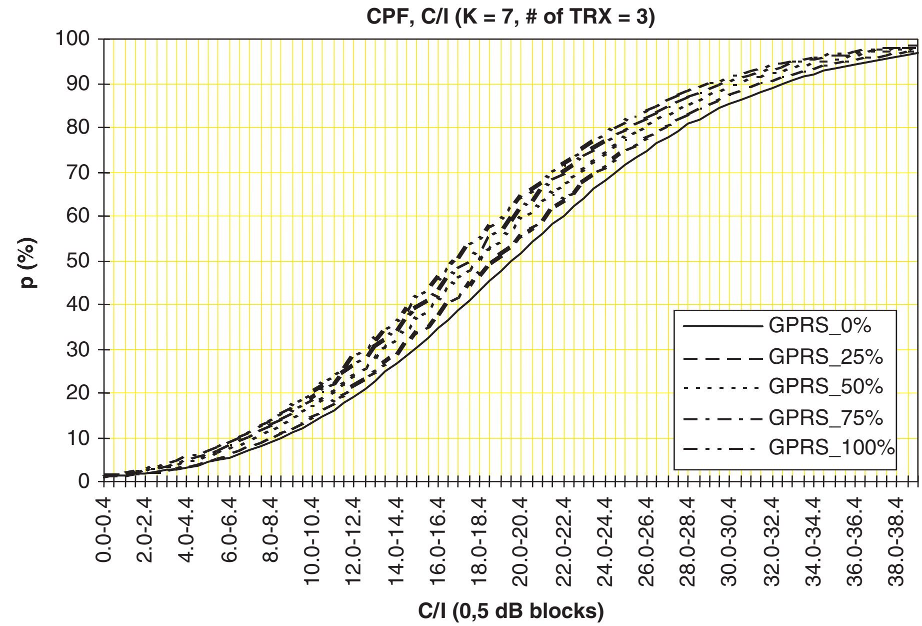

Figure 20.44 The cumulative distribution function of the C/I, when reuse pattern size K = 7 and all cells studied have three frequencies.

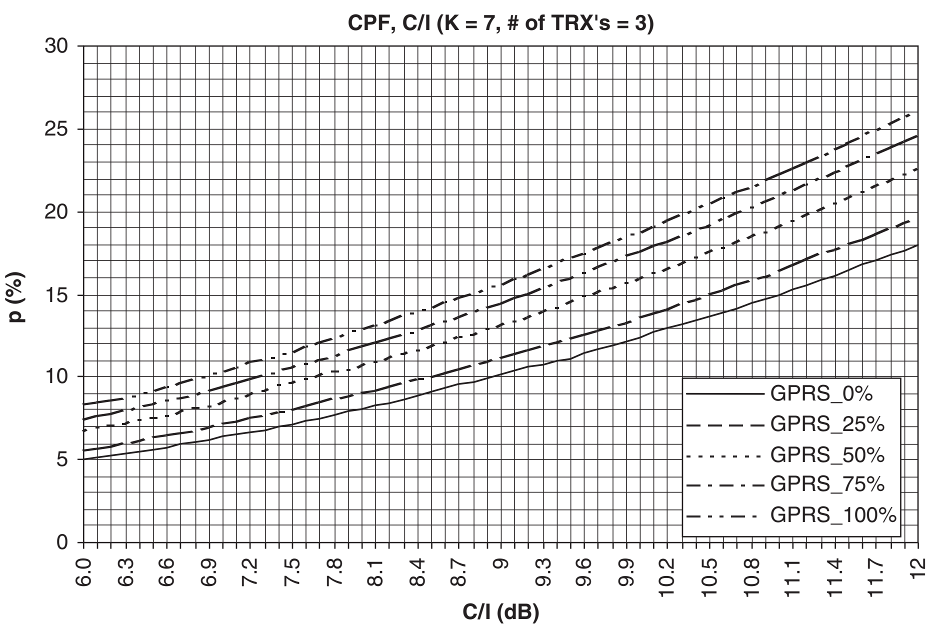

Figure 20.45 The cumulative distribution function of the C/I shown in 12.33- a more detailed view into the 9 dB region.

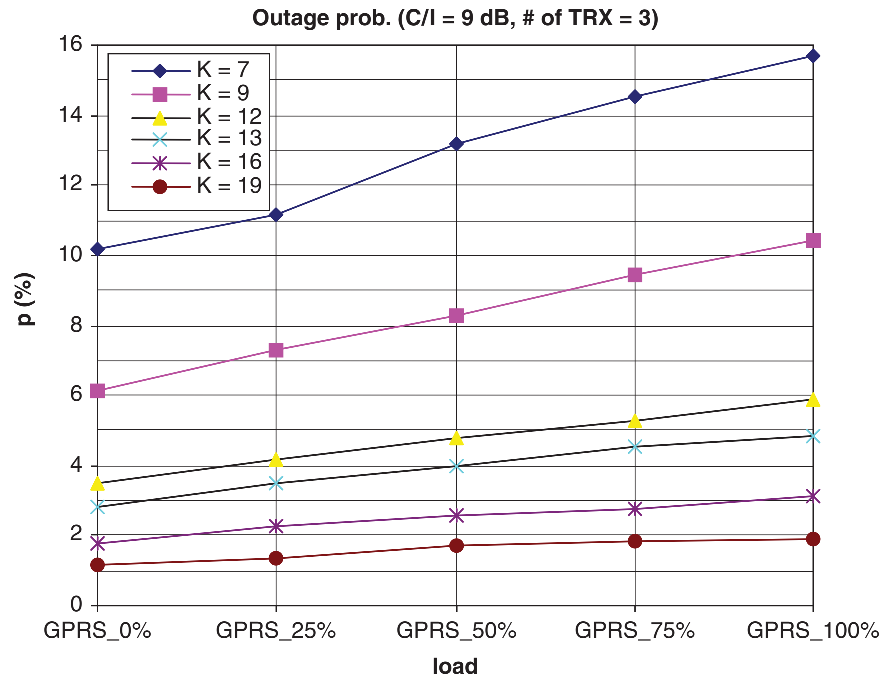

Figure 20.46 *The outage probability achieved with recurrence patterns 7–19, when C/I = 9 dB.*

According to the results of simulations, the 10% outage probability is met with 7 recurrence pattern without GPRS traffic. The 9 recurrence pattern is still met with a 75% GPRS load, but networks with a full load experience outage probabilities of over 10%. Recurrence patterns over 9 seem to keep outage probability clearly below 10% with the 9 dB limit value.

Outage probability can be studied with the *C/I* values required by the different channel coding classes of GPRS. Figures 20.47–20.50 compare outage probability to the *C/I* values achieved in the CSELT simulations, according to which CS-1 requires $C/I = 10.8$ dB, CS-2 12.8 dB, CS-3 13.7 dB, and CS-4 17.2 dB without frequency hopping.

It can be seen from the figures, that CS-1 functions with small GPRS loads with recurrence pattern $K = 9$, and also with heavy GPRS loads, if K has at least a value of 12.

If it is desired that CS-2 functions constantly, a minimum recurrence pattern $K = 13$ is required with light GPRS loads, and $K = 16$ for heavy GPRS loads, respectively. CS-3 requires a recurrence pattern $K = 12$–13 for light and $K = 16$ for heavy GPRS traffic. CS-4 functions with 10% outage probability and light GPRS traffic when $K = 19$.

Figure 20.51 shows the *C/I* values achieved as a result of simulations when $K = 3$–19, and outage probability is 10%.

It can be seen from Figure 20.51 that the *C/I* relationship is reduced due to GPRS load by 0–2 dB with all studied recurrence patterns, depending on the level of the load. This effect is important enough to be taken into account in the precise radio network planning especially in the case of tightly dimensioned reuse patterns, that is, with low amount of overlapping.

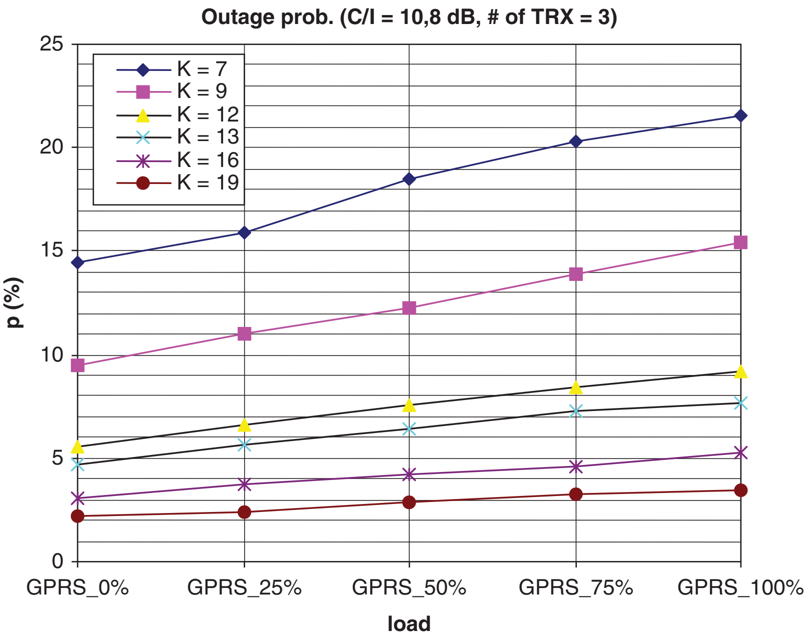

Figure 20.47 *The outage probability achieved with recurrence patterns 7–19 for GPRS channel coding class CS-1.*

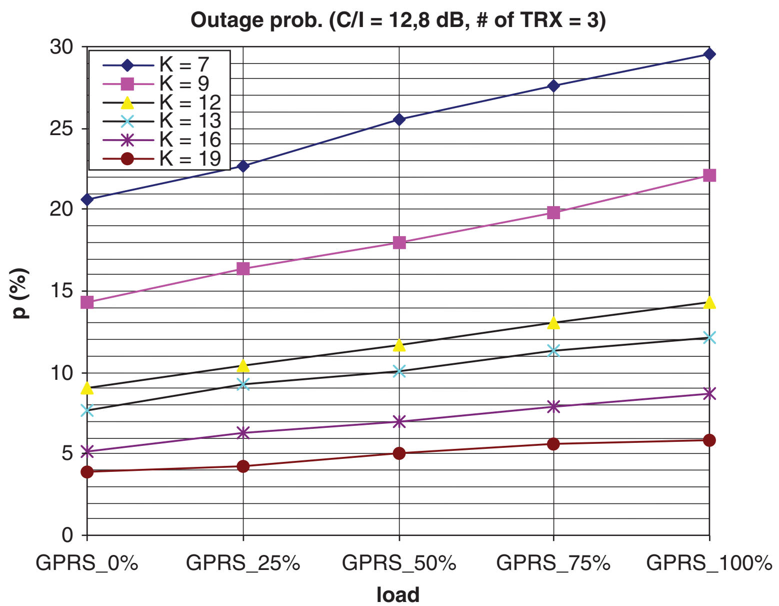

Figure 20.48 *The outage probability achieved with recurrence patterns 7–19 for GPRS channel coding class CS-2.*

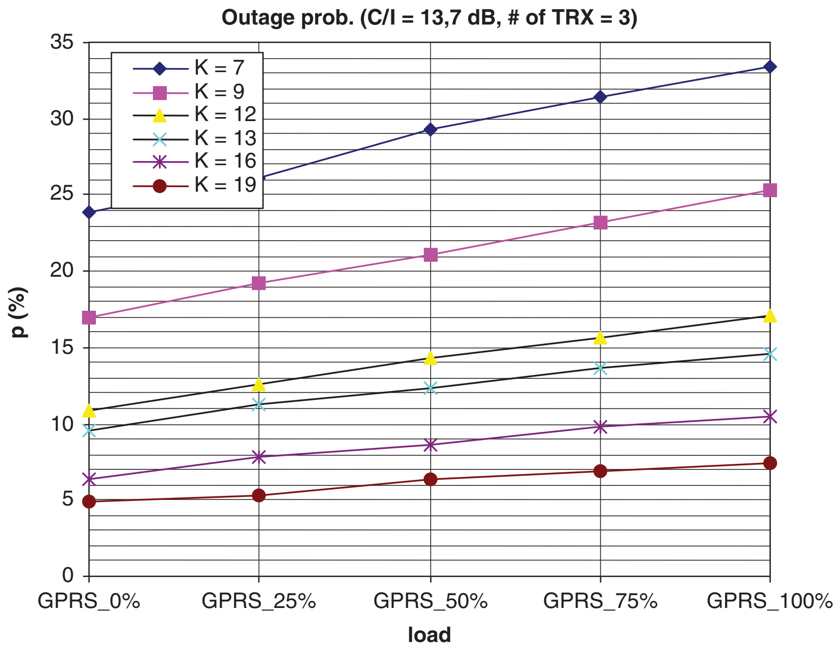

Figure 20.49 The outage probability achieved with recurrence patterns 7–19 for GPRS channel coding class CS-3.

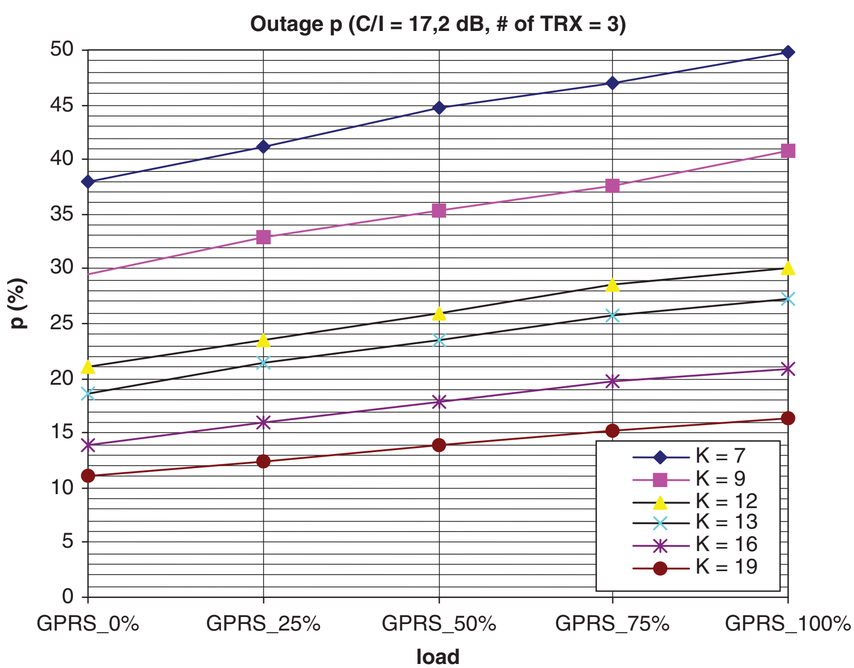

Figure 20.50 The outage probability achieved with recurrence patterns 7–19 for GPRS channel coding class CS-4.

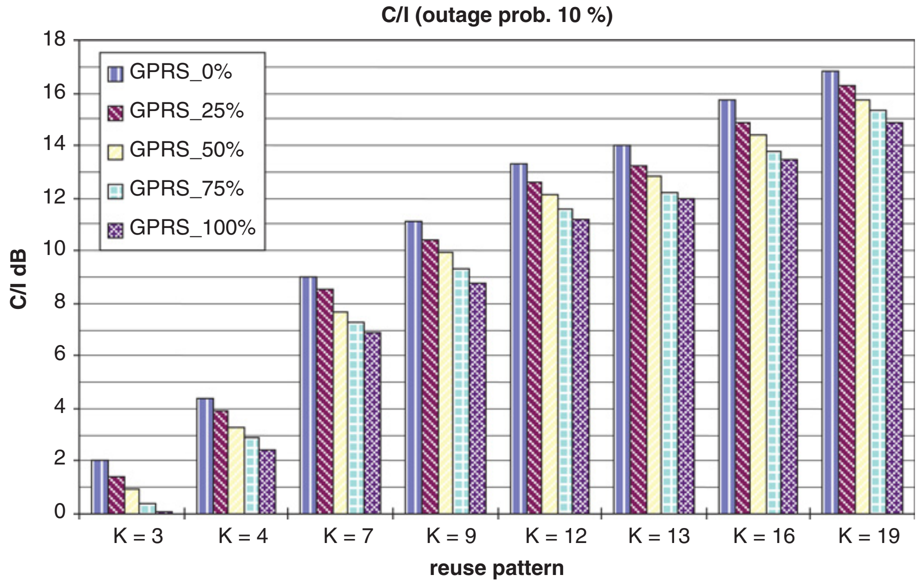

Figure 20.51 *C/I-values achieved with recurrence patterns K = 3–19, and outage probability is 10%.*

References

1. Brasche, G. and Walke, B. (1997) Concepts, Services, and Protocols of the New GSM Phase 2 +General Packet Radio Service. IEEE Communications Magazine, August 1997, pp. 94–104.
2. Zanini, P. (1997) Examining the GPRS radio interface and evaluating GPRS radio performance. GPRS & HSCSD conference. Centre for EuroTelecoms, London 22–24 October 1997.
3. Penttinen, J. (1999) Tulevaisuuden GSM-datapalveluiden vaikutus radioverkkosuunnitteluun (The effects of future GSM data services on the radio network planning). Licentiate thesis, Helsinki University of Technology, p. 123.
4. Penttinen, J. (2000) Effects of GPRS on GSM Radio Network. ICT 2000, International Conference on Telecommunications, Acapulco, México 22–25 May, 2000, pp. 162–6.
5. Halme, S.J. and Rahko, K. (1981) *Tietoliikennetekniikan Perusteet 471*, Otakustantamo, p. 328.
6. EDGE - Evolution of Binary and Quaternary Offset QAM. Tdoc EDGE 007/98. ETSI Joint SMG 1, 2, 3 and 4 EDGE Workshop #1. Helsinki, Finland, 16–17 February 1998.
7. Evaluation of Continuous Phase Modulation. Tdoc EDGE 04/98. ETSI Joint SMG 1, 2, 3 and 4 EDGE Workshop #1. Helsinki, Finland, 16–17 February 1998.
8. ETSI GSM 05.05. GSM/GPRS radio interface.
9. Working Document for the ERO Workshop on Traffic Loading and Efficient Spectrum Use. 4 December 1997, Brussels. 50 p. + annex.
10. Influence of terrain irregularities and vegetation on tropospheric propagation. Recommendations and reports of the CCIR, 1986. Report 236-5. XVIth plenary assembly, Dubrovnik, 1986. International Telecommunication Union, CCIR. pp. 96–101.
11. Terävä, Kimmo. Analysis and implementation of an algorithm for estimating the quality of frequency hopping GSM. Diploma thesis seminar presentation, Helsinki university of technology.
12. Balston, D.M. and Macario, R.C.V. (1993) *Cellular Radio Systems*, Artech House, p. 373.
13. Gibson, J.D. (1996) *The Mobile Communications Handbook*, CRC Press Inc., p. 577.
14. Gilchrist, P. (1996) One stop data shop from GSM's GPRS. Mobile Communications International, March 1996, pp. 62–4.

15. GPRS Conference. 4–5 March 1998, The Waldorf Meridian, London. Centre for EuroTelecoms.
16. Halme, S.J. and Rahko, K. (1987) *Tietoliikennetekniikan Perusteet, 471*. Second revised edition. Otakustantamo.
17. Halme, S.J. (1992) *Televiestintäjärjestelmät 894*. Otakustantamo, p. 402.
18. Hata, M. (1980) Empirical formula for propagation loss in land mobile radio services. *IEEE Transactions on Vehicular Technology*, **VT-29**(3), 317–25.
19. Hess, G.C. (1993) *Land-Mobile Radio System Engineering*, Artech House, Inc., p. 371. ISBN 0-89006-680-9.
20. HSCSD Conference. 3 March 1998, The Waldorf Meridian, London. Centre for EuroTelecoms.
21. Häggman, S.-G. (1999) Frequency allocation to meet interference outage requirements. Radio Network 2000, planning principles for 3rd generation cellular networks. Merito Forum, 5–7 May 1999.
22. Hämäläinen, J. (1996) Design of GSM High Speed Data Services. Thesis for the degree of Doctor of Technology. Tampere University of Technology, Publications 186. Tampere. 104 p. + publications 1–8.
23. Kokko, O. (1994) *Matkapuhelinverkkojen Sisätilasolujen Suunnittelu*, Diplomityö. Teknillinen korkeakoulu, Sähkötekniikan osasto, Espoo, p. 93.
24. Lee, W.C.Y. (1986) Elements of cellular mobile radio systems. *IEEE Transactions on Vehicular Technology*, **VT-35**(2), 48–56.
25. Lee, W.C.Y. (1993) *Mobile Communications Design Fundamentals*, John Wiley & Sons, Inc., USA, p. 372. ISBN 0-471-57446-5.
26. Morgensen, P., Wigard, J. and Fredriksen, F. (1995) *Performance of Slow Frequency Hopping in GSM (link level)*, COST 231 (TD)95-121, Poznan, p. 18.
27. Pirhonen, Riku, Rautava, Tapio and Penttinen, Jyrki (1999) TDMA convergence for packet data services. *IEEE Personal Communications*, **6**(3), 68–73.
28. Influence of terrain irregularities and vegetation on tropospheric propagation. Recommendations and reports of the CCIR, 1986. Report 236-5. XVIth plenary assembly, Dubrovnik, 1986. International Telecommunication Union, CCIR, pp. 96–101.
29. Sreetharan, M. and Kumar, R. (1996) *Cellular Digital Packet Data*, Artech House, Inc., p. 315. ISBN 0-89006-709-0.
30. IEEE 47th Vehicular Technology Conference Proceedings, Volumes 1, 2 and 3. May 4-7, 1997. Hyatt Regency, Phoenix, Arizona, USA. 2247 p.

21

Planning of Advanced 3G Networks

Jyrki T. J. Penttinen

21.1 Introduction

This chapter presents the network planning aspects for the LTE and LTE-Advanced radio interface. Typical methods are presented for estimating the LTE coverage and capacity in the nominal and more advanced phases of the network planning and deployment. As an essential base for the planning, LTE link budget is introduced. Also useful radio propagation models are described for the realistic estimation of the radio path loss.

In general, the LTE and LTE-Advanced radio network planning has many elements that are comparable with the previous mobile communications systems. The common goal of all mobile system radio interface dimensioning is to estimate the coverage and capacity of the services, by adjusting the available parameter values in the planned environment types. Please note that the planning guidelines of the basic UMTS/HSPA are presented in Chapter 12.5 and can also be referred to in more detail in [1].

21.2 Radio Network Planning Process

The radio network planning process can be generalized as presented in Figure 21.1 [2, 3].

The network planning is only one item in the deployment project. The final outcome of the radio plan depends on the given time frame for the rollout, the target quality, offered capacity and quality. Figure 21.2 summarizes the project dependencies. The planning assumptions of LTE can be estimated to be close to the methods and values utilized typically in HSPA, but there are also differences due to the OFDM and SC-FDMA, and higher data speeds. In practice, both HSPA+ as well as LTE radio networks are evolving in such a way that the highest data rates still share many common aspects as for the size of the useful service areas per data rate class.

The focus of the nominal planning phase is to estimate the required number of sites that provide services with sufficiently high quality. Uniform assumptions for the site parameters can be utilized in each cluster

The Telecommunications Handbook: Engineering Guidelines for Fixed, Mobile and Satellite Systems, First Edition.
Edited by Jyrki T. J. Penttinen.

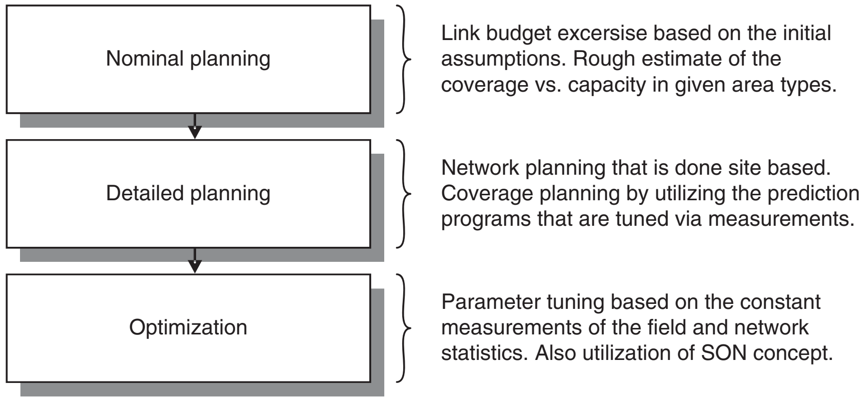

Figure 21.1 *The main items of the LTE radio network planning.*

type. The clusters can be divided, for example, into the dense urban, urban, suburban and rural areas. The first-hand estimate for the required amount of sites per cluster type can be done relatively fast by utilizing a link budget.

The detailed network planning is done site-by-site basis, as the assumptions for the antenna directions, downtilting, power levels and so on varies. A tuned radio propagation models and planning software are utilized at this phase. Digital maps are important as they take into account the topology of the environment, which greatly affects on the local predictions of the coverage areas. The field measurements can be utilized for the tuning of the propagation model parameter settings,

Typically, the optimization is divided into prelaunch optimization and postlaunch optimization. After the actual deployment, the optimization can last until the end of the life cycle of the LTE network. The capacity figures changes during the operation of the network, which requires rearrangements for the plans. The user profiles are investigated constantly at this phase by carrying out regular field tests and by collecting network statistics of the capacity usage, performance figures and possible faults.

The LTE radio network planning depends heavily on the operator's position in the market. For the already existing 2G and/or 3G operators, the reutilization of the base station sites as much as possible is essential in order to keep the deployment costs in optimal level. The most logical strategy in this case is to build LTE as an additional layer. In the beginning of the LTE operation, the LTE coverage is very limited and only hotspots without clear continuous coverage is possible. This means that the services continuity is taken care of via the

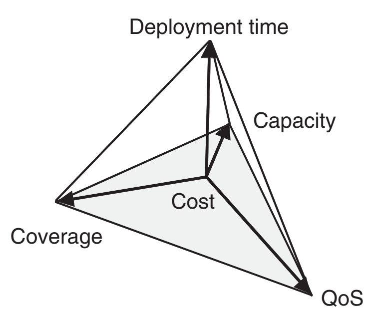

Figure 21.2 *The main items that affect on the total cost of the LTE network.*

existing 2G and 3G networks which act as umbrellas whilst the LTE takes care of the services in the hotspot coverage.

As the LTE network matures, there might be need to construct or rent additional eNodeB sites. At the same time, the already existing 2G and 3G capacity could be lowered gradually, which provides possibility for the radio frequency refarming, that is, the equipment and radio capacity of the previous generations can be lowered whilst LTE takes gradually more bandwidth from the same frequency according to the defined bandwidth definitions of LTE, from 1.4 MHz up to 20 MHz. This requires logically coordination between the radio network planning processes of different systems.

Figure 21.3 gives some ideas about the possible transition of the systems, the 2.1 and 2.6 GHz frequencies being a logical capacity layer solution in the urban environment, and 900 MHz in the rural environment due to its larger coverage areas. The gradual degradation of the earlier systems can happen in such a way

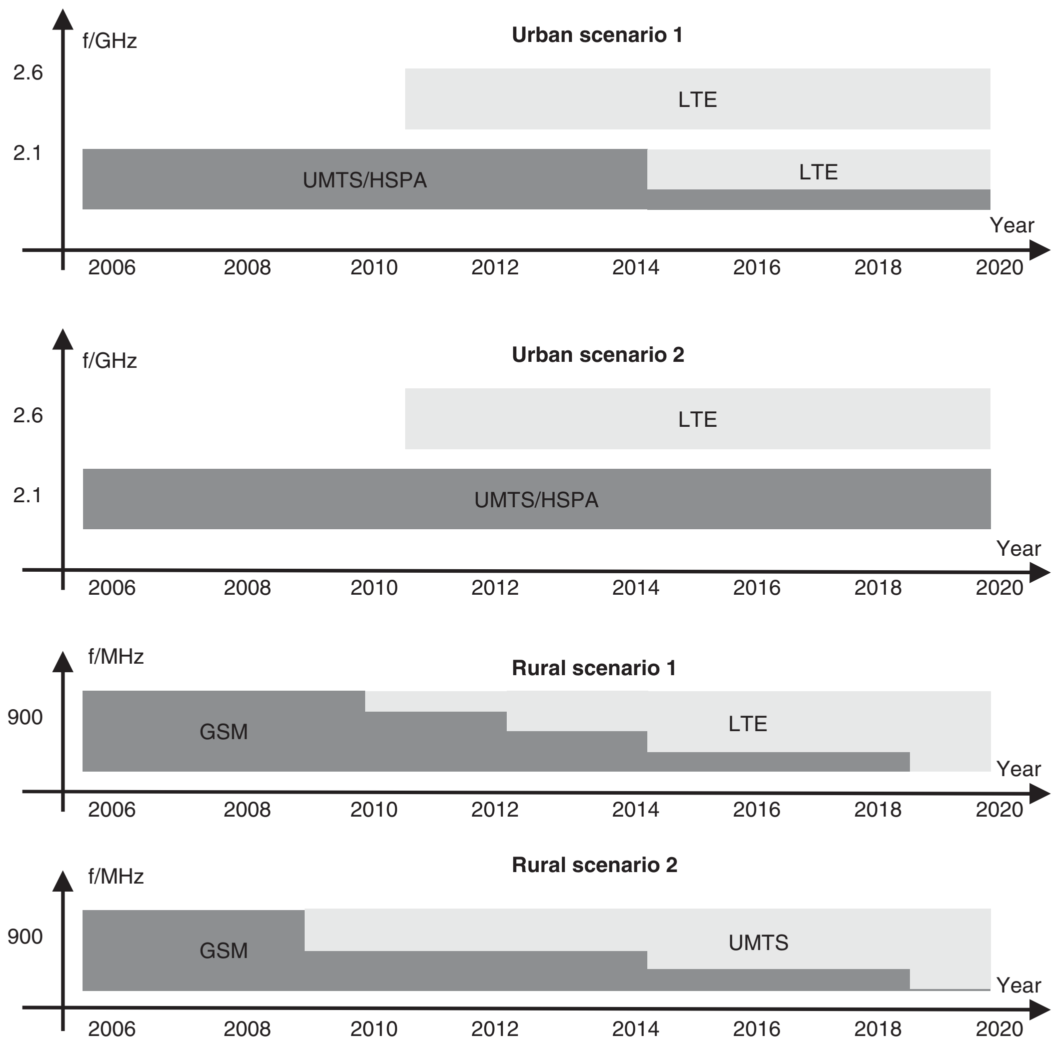

Figure 21.3 *Some possible transition scenarios towards LTE deployment.*

that GSM will be disappearing little by little, and UMTS starts operating in lower capacity, LTE being the parallel system for one of these, until LTE takes over the whole traffic. Another realistic scenario could be even to gradually remove UMTS whilst GSM and LTE stays as parallel systems. This latter case is justified due to the vast amount of GSM capable terminals in the markets which provides the basic voice calls and light multimedia whilst LTE/GSM capable terminals provide the advanced data and part of the voice calls especially in the fall-back situations. The advanced spectral efficiency features like OSC (Orthogonal Sub Channel) and DFCA (Dynamic Frequency and Channel Allocation) of GSM supports this evolution. The more specific examples of the transition scenarios can be investigated by referring Chapter 25.

For the green field operators, the benefit is that the LTE network can be planned in an optimal way from the scratch. The drawback is that there is no existing infrastructure for the sites and respective transmission. The heavy site hunting must thus be executed in an early phase of the planning, by defining the preferred / priority search rings over the planned area. Furthermore, it is essential that there is a constant feedback loop between the planning process and site hunting, as the latter many times do not result in the optimal locations and constant adjustments with the locations, antenna heights and so on must be done. In addition, the transport and core planning must follow at least to some extent the advances of the radio site hunting, before the actual transmission lines or radio links can be constructed physically. Figure 21.4 shows an example of the LTE radio planning process.

21.3 Nominal Network Planning

The main aim of the nominal planning phase of LTE is to estimate the needed number of eNodeBs as accurately as possible, before the actual network deployment begins. Via this estimate, the total cost of the LTE/SAE network will also be estimated by assuming the required criteria for the coverage, capacity and quality of service. Although this phase does not yet produce final network architecture plans or more concrete site distributions, it is essential in order to give a realistic understanding about the capital and operational expenditures. The larger the LTE deployment area is, the more important the accuracy of the eNodeB number and thus the network cost.

21.3.1 Quality of Service

The QoS of the LTE network is better when the blocking of the network is lower. This results in higher average and peak data rates. The balancing of the sufficiently high quality is one of the many optimization tasks of the operator: the higher the average throughput is, the more satisfied the customers are – but with the price of higher deployment costs of the network. It is thus essential to take into account all the relevant technical and nontechnical aspects that affect the average revenue per user (ARPU).

The challenge is to find the optimal point that generates the highest ARPU during the lifetime of the LTE network. Figure 21.5 presents the behavior of the optimal solution.

The dimensioning of the network is done according to the peak hours as well as average load of the network. The load peak generates inevitably lower throughput values per user, or even blocking when the admission control has to limit the entrance of new users. The blocking is designed according to dimensioning criteria, and taking into account the expected future development of the user load as well as occasional load peaks in local, special events and so on. It is important to avoid overdimensioning in order to keep the ARPU close to the ideal level.

The most demanding location for the dimensioning is found at the edge of the cells, where the interference level is highest and thus the SINR (Signal to Interference and Noise Ratio) value is lowest.

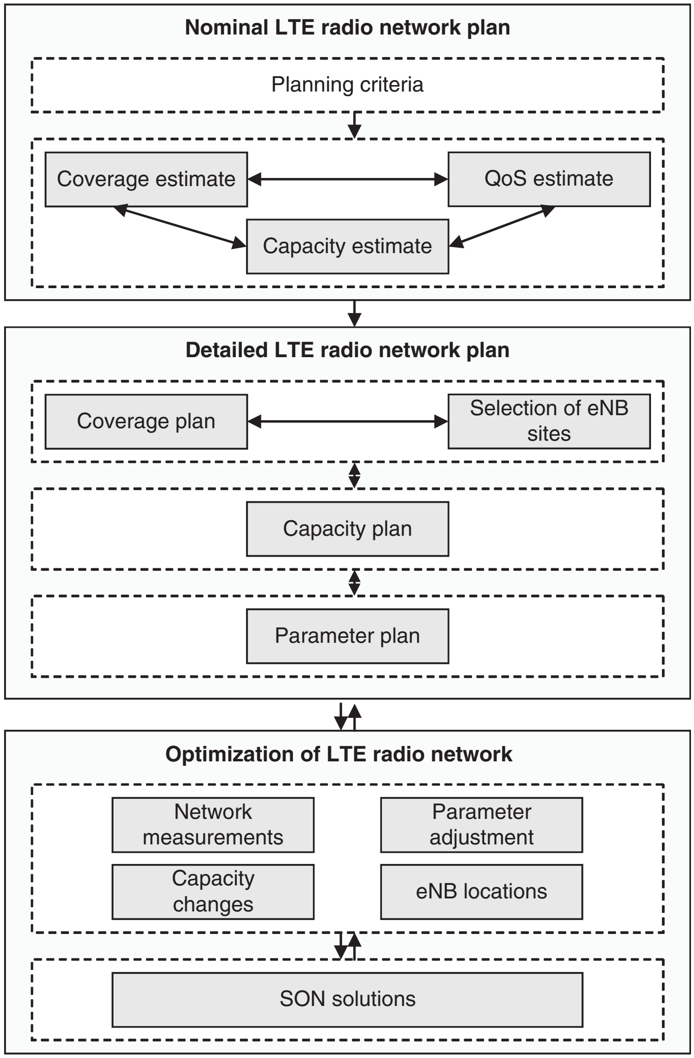

Figure 21.4 *The overall LTE radio network planning process.*

The design and deployment of LTE are quite similar with the earlier mobile communications systems like GSM and UMTS. Especially the HSPA shares many aspects with LTE, both being data solutions of the 3G evolution path.

Similarly to the other general radio network design processes, the radio link budget is the base for the initial planning. It is meant to estimate the maximum allowed functional path loss between the base station and terminal, and it is calculated separately for uplink and downlink. The balancing of the transmission and receiving directions is thus one of the outcomes of the link budget.

The differences between the different transmitting directions are related basically to the different access methods and their requirements for the minimum received power levels for certain data rates. The useful received power level depends on the presence of the useful and interfering signals compared to the noise level, that is, SINR. The network load does have a direct impact on the SINR value because as the amount of

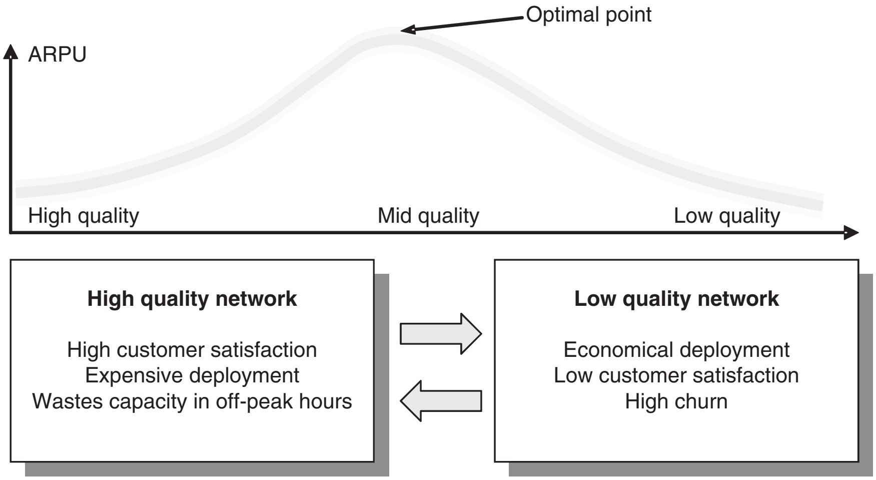

Figure 21.5 *The balancing of the quality.*

the users grows, also the interference share gets higher. A certain SINR value provides accordingly a certain quality of service level providing a respective location and time dependent bit rate and bit error rate, that is, throughput. As the retransmission rate gets higher, the quality of service lowers respectively which can be seen directly by the quality by the customer. The value also depends on the network load, lowering as the capacity utilization gets higher.

The interpretation of the quality of the coverage area depends on the agreed area location probability level. In general, the location variation is considered to follow a log-normal distribution, meaning that the logarithm of the signal level follows a normal or Gaussian distribution. The statistical distribution should be applied in the respective quality level estimations. The mean value means that 50% of the samples are above this value and the other half below. In case of any other percentage for the coverage quality criterion, the relationship between the mean value and standard deviation should be known. The standard deviation value of 5.5 dB can be assumed in the typical suburban type by default unless more accurate values are available. The standard deviation is commonly used as a basis for the mobile communications coverage predictions, informing about the confidence in statistical conclusions.

The relationship between the area location probability and the additional margin that should be taken into account in the link budget can thus be derived from the characteristics of the normal and log-normal distribution, for example, by observing the attenuation points (dB) in cumulative scale that fulfils the required percentage of the area location in the whole area. Furthermore, it can be decided that 90% of area location probability indicates a fair outdoor coverage, whilst 95% is considered as good and 99% provides an excellent quality. Table 21.1 summarizes the mapping of the typical values that can be used in the mobile reception,

Table 21.1 *The area location probability in the site cell edge and over the whole site cell area for the mobile reception when the standard deviation is 5.5 dB*

Area location prob. (minimum coverage target)	Loc probability in site cell edge	Location correction factor	Subjective quality description
90%	70%	7 dB	Fair outdoor
95%	90%	9 dB	Good outdoor, fair indoor
99%	95%	13 dB	Excellent outdoor, good indoor

when the standard deviation is 5.5 dB. In addition to the standard deviation, the criteria vary depending on the environment, that is, on the propagation slope. Slope of 2 (i.e. 20 dB/decade) represents line of sight in free space. The slope of 3.5 (i.e. 35 dB/decade) is used in Table 21.1 representing typical urban environment.

The OFDMA of LTE downlink is especially suitable for the environments with heavy fast-fading of the radio signals, for example, in the dense urban areas where the number of multi path propagated radio components is high. The division of the original information for many lower bit rate subcarriers over a wide frequency band assures that only small set of subcarriers is lost. The spreading of the data (interleaving) over the band, combined with a turbo-coded data, protects fairly well the data as it can be recovered efficiently in the reception side.

The drawback of the OFDMA is the challenge of the SFN network's Peak-to-Average-Power Rate (PAPR) variations, combined with the nonoptimal power efficiency. For this reason, the transmitter of the LTE-UE, that is, the uplink direction, utilizes more energy efficient SC-FDMA (Single Carrier Frequency Division Multiplexing).

The combination of OFDMA and SC-FDMA provides in any case comparable coverage areas with the previous HSPA networks, the difference being higher data rates in LTE.

One of the typical scenarios is that the operator that already has 2G and/or 3G license adds LTE service areas gradually as a part of the complete set of network technologies. This means that especially in the initial phase of the LTE deployment and operation, the proportion of LTE converge is relatively small. The strategy may be to provide LTE coverage only in the most loaded hot spots of the data transmission, providing fast data services and partially load balancing for the voice via VoIP concept. The fallback solution is thus utilized largely for handing over the already established data and voice calls of LTE, providing that the LTE-UE supports the 2G and/or 3G technologies. It can be assumed that the LTE traffic type is heavily data weighted in the beginning with the data dongles (VoIP being possible as typically via Internet and laptop by utilizing the headset), and the VoIP service takes off gradually more as there will be more tradition handset models with the integrated voice call facilitators (microphone and auricular integrated into the LTE-UE).

An important part of the quality of service is the continuum of the LTE service in all the use cases where LTE coverage area is sufficiently good. The functionality of the handovers is thus important. The new X2 interface between the eNB elements optimizes further the success rate of the handovers in LTE compared to the previous techniques, where the direct signaling connection between the base stations was missing. In case of the probable situation where LTE coverage ends but 2G/3G coverage is still available, the connection is changed automatically to 2G/3G networks. This does have an impact on the quality of service level as it is probable that the overall data throughput lowers in this case. For the VoIP connection that is changed to 2G/3G CS call, the impact is not dramatic as soon as the fallback procedure is executed successfully (in fact, the CS voice call quality might be better than LTE VoIP call quality), but the clearest effect will be seen in the high-speed data utilization. The mostly impacted applications are related to the (close-to) the real-time data streaming, when the capacity is lower that in LTE networks. In any case, the continuum of the service is assumingly much more important than the complete breakdown of the call.

21.4 Capacity Planning

The capacity of LTE depends on the services, that is, the required bit rates per user, as well as on the quality of service level. The capacity is related to the utilized bandwidth. The modulation and coding schemes do have a direct relationship with the capacity in such a way that in the cell edge, the offered capacity is lowest due to the highest coding rates and most robust modulation method (QPSK). Nearer the eNodeB, the 16-QAM and 64-QAM can be utilized with lower code rate which facilitates the more efficient utilization of the resources for the actual user data. The modulation and coding scheme of LTE is adaptive which means that the optimal combination is selected. As the interference level of the LTE network increases as a function of the utilization level, also the bit error rate increases which lowers the throughput. This means that there is similar type of "cell breathing" effects in LTE as in W-CDMA as the amount of the calls differ in the area.

As the capacity of LTE depends on various aspects that vary as a function of time and location, one of the possibilities to estimate the capacity behavior is to simulate the most probable use cases, assuming a certain distribution of application profiles. The simulation results can show the distribution of the modulation and coding schemes over the simulated area, taking into account the interfering neighbor cells. This gives information about the expected SINR values and respective throughputs with the given bit error rate, which can be, for example, 10%. The performance depends on the distance between the eNodeBs, antenna heights, radiating power levels, and in general on the surrounding area types and topologies. The most robust modulation of LTE, QPSK, typically requires about 0–3 dB SINR, whilst 16-QAM functions with a minimum of about 7–11 dB, and 64-QAM with 12–15 dB, when the coding rate varies between the highest and lowest possible ones.

The planning of LTE radio network, as well as the process in general, is quite similar with the ones applied in the case of HSPA. Both utilize the same frequency planning principles with the reuse pattern size of 1, which means that the users share the same frequency band for the communications. Also the radio resource management principles are similar.

The LTE network's capacity and coverage thus varies. In order to estimate the effect, a planning tool is required. There are different ways to model the cases. The basic principle is to assume different data usage cases over the whole investigated area in such a way that the average throughput per user, as well as the variations, can be seen via the distribution, for example, in the cumulative presentation. The most challenging part is, in fact, to make the sufficient realistic assumption for the traffic types. For this, the earlier data of UMTS data services can be utilized as a basis, in order to guess the possible share of, for example, the VoIP, WEB, FTP down/uploading, messaging and other traffic types.

The performance indicator can be, for example, throughput per user or spectral efficiency as a function of time, SINR, or other variable. The commonly known Monte-Carlo method is practical and relatively straightforward, when the static snapshots are repeated sufficient amount. The simulation can also be dynamic, although it is fairly more complicated as the mobility of each user has to be modeled. This requires more time and processing power, so the dynamic model is useful merely for the in-depth investigations of radio research management functionality, whereas Monte-Carlo is useful in the general LTE network planning.

Figure 21.6 shows an OFDM snapshot exercise over a theoretical suburban area type when the path loss has been estimated by taking into account a standard deviation of 5.5 dB. In this specific case, Okumura-Hata has been applied.

21.5 Coverage Planning

The LTE coverage area can be estimated in the simplest way via the path loss prediction, although it is much more limited. This method does not reveal the effect of the traffic type distribution on the coverage, but in any case, it gives a sufficiently good first-hand estimate about the general coverage which can be achieved with certain offered capacity case (average throughput).

In the very initial phase of the planning, even a generalized link budget exercise can be executed in order to estimate the rough, average LTE cell range per area type (dense urban, urban and suburban, rural, and open area). By assuming a certain overlapping percentage for the adjacent cells, this exercise provides the initial estimate of the needed number of the eNodeB sites over the planned area.

The coverage planning of LTE radio network is actually a technology dependent. Thus the already existing radio propagation models can be utilized also in LTE planning whenever they are designed for the respective

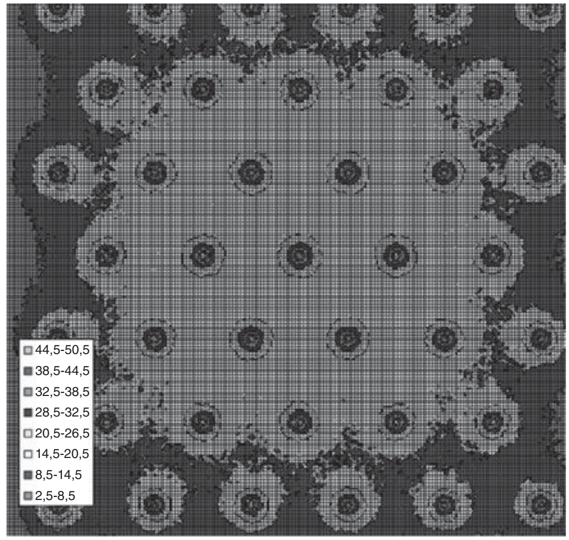

Figure 21.6 *An example of a snap-shot simulation of OFDM in theoretical (flat) sub-urban type. The estimate becomes more realistic when the digital map data is taken into account in the path loss prediction.*

radio frequencies. The radio path loss prediction per area type gives the estimate for the expected radius of the eNodeB. In order to estimate the maximum allowed path loss, a power link budget can be utilized by applying the same principles as in the HSPA radio network planning.

21.5.1 Radio Link Budget

The rough estimate for the initial coverage planning can be obtained by the utilization of the radio link budget. Figure 21.7 shows the general level link budget of LTE. As can be seen, it is quite similar to one of the other mobile communications systems.

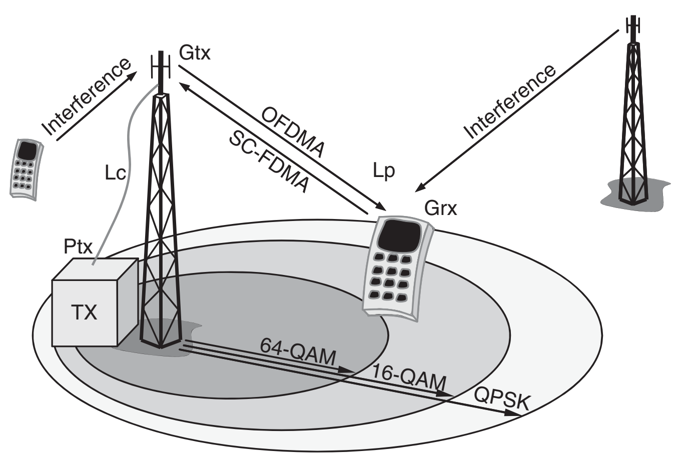

Figure 21.7 *The principle of LTE radio link budget.*

The link budget gives a relatively good overall and average idea about the maximum achievable path loss between the eNB and LTE-UE. This, in turn, provides an estimation of the average cell radius of the sites. The estimate can be obtained for different environment types like dense urban, urban, suburban, rural and open area.

The link budget gives the path loss estimate both for downlink and uplink, which gives possibility to balance the transmission and reception of the data streams according to selected criteria. The balance is important in order to provide a planned quality of service in both directions. In case of LTE, the importance of the balance depends on the application. As an example, for the voice service via VoIP solution, it is important to provide the minimum needed data flow for both directions in order to have a successful call between the A and B subscriber also in the cell edge area In turn, for the downloading of, for example, web pages, the data speed in uplink is not the limiting factor whilst the bit error rate is sufficiently low, so even slow data rates in uplink provides the needed acknowledgements for the fast data download in downlink.

The differences between the downlink and uplink of LTE are related to the different access methods (OFDMA in downlink and SC-FDMA in uplink) and the respective requirement for the minimum functional values. The final useful received power level depends on the balance of the useful carrier signals and the interfering signals, which is compared to the noise level SINR (Signal to Interference and Noise Ratio). A certain SINR value provides a certain quality of service level and respective data rate, which depends on the time and location, especially in the case of multipath propagated radio signals that cause slow and fast fading as a function of time and location. The final quality of the connection is interpreted by the end-users as a useful throughput of the data flow. This takes into account the bit error and respective retransmission, which lowers the useful bit rates even if the erroneous bit rate would be much higher.

The LTE radio link budget is fairly similar with the link budgets utilized in the other mobile technologies, especially in the case of the HSPA. It is quite useful in the nominal planning purposes of the LTE radio network, as it provides a general estimate for the average cell ranges and thus the needed amount of eNBs within the planned area.

Figure 21.7 shows the principle of the LTE radio link budget. The most important factors in the downlink direction are the output power level of the eNB transmitter (P_{tx}), the cable and connector losses (L_c), and the transmitting antenna gain (G_{tx}). It is worth noting that the active antenna systems (AAS) may get more popular along LTE deployments. As the AAS contains the front-end of the transmitter inside the same antenna element shield, the transmitted content can be delivered from eNB up to the AAS by fiber optics which eliminates the cable losses.

The radiating power (EIRP, Effective Isotropic Radiating Power) is distributed via the air interface to LTE-UE, which has certain sensitivity (S), receiver antenna gain (G_{rx}) and noise figure. It should be noted that in case of the small-sized USB stick –type of LTE terminals, the in-build antenna causes merely loss instead of gain. Also external antennas can be utilized with the LTE terminals, but for the link budget purposes the most likely terminal type should be taken into account, which means probably that in-build antenna should be assumed with, for example, 0 dB antenna gain, or in the pessimistic assumptions even as slightly negative.

In the uplink direction, the most important factors are the transmitting power level of the LTE-UE, the antenna gain of the terminal and eNB (which are likely the same as in downlink if the same antenna types are utilized), the cable and connector losses and the sensitivity of the eNB receiver.

The maximum allowed path loss value L is calculated for different modes (modulation schemes) by taking into account the minimum received power level requirement for the modes, and by subtracting the emitted and received powers. The estimate is done separately for the outdoors, and indoors by assuming a certain area-type dependent average building penetration loss.

As a rule of thumb, the QPSK modulation provides largest coverage areas, but at the same time it results in the lowest capacity. The 64-QAM modulation gives highest data rates, but only in limited areas compared to the QPSK case. The third possible LTE modulation, 16-QAM, is a compromise as for the coverage and capacity. Also, the least protected transmission (highest code rates) provides highest data rates but within smaller coverage areas than the heavier protected codes do. In the accurate coverage and capacity estimates, the respective areas can be calculated separately. LTE includes in any case automatic modulation and coding adaptation (MCS, Modulation and Coding Scheme), which provides the optimal combination of the modulation and code rate at all the times. In OFDM, the adaptation could be done even in timeslot basis, but for the practical reasons for keeping the signaling load in optimal level, LTE defines the minimum resolution for the link adaptation to radio block. In practice, this provides in any case enough fast adaptation for any practical situations.

The well-known radio path loss prediction models can be applied as such also for the LTE coverage estimations, when the minimum required power level per investigated mode is known. If the exact practical values are not available, also simulation results can be used as a base for the minimum requirements. In addition, when single frequency network concept is utilized, there is a possibility to add a certain SFN gain value for the link budget. It basically means that the eNBs can be constructed slightly farther away from each others, or alternatively, the reception level is of higher quality in the cell edge areas than it would be without SFN mode. This concept is valid, for example, for the Multimedia Broadcast concept of LTE (MBSFN).

The link budget calculation is typically based on a minimum throughput requirement at the cell edge. This approach provides the cell range calculation in a straightforward way. The order of defining the throughput requirement prior to the link budget calculation, it is possible to estimate the bandwidth and power allocation values for a single user, which mimics the behavior of the realistic scheduling sufficiently well for the link budget calculation purposes.

The LTE link budget can be designed by planning sufficiently deeply the essential parameter values shown in Table 21.2 for downlink, and in Table 21.3 for uplink. The latter assumes 360 kHz bandwidth utilization whereas the downlink assumes a 10 MHz band for the transmission.

The link budget can be planned in the following way, for example, in the downlink direction with the 10 MHz assumption as shown in the example above. The radiating isotropic power (d) can be calculated by taking into account the transmitters output power (a), the antenna feeder and connector loss (b), and transmitter antenna gain (c), so the formula is: $d=a-b+c$.

The minimum received power level (p) of the LTE-UE can be calculated $p = k + l + m - n + o$, utilizing the terminology of the link budgets shown above. The noise figure of the LTE-UE depends on the quality of the model's HW components. The minimum signal-to-noise (or, signal-to-noise and interference, SINR) value j is a result of the simulations. The sensitivity k of the receiver depends on the thermal noise, the own noise figure of the terminal, and SINR in such a way that $k = g + h + j$.

The interference marginal l of the link budget represents the average estimate of the noncoherent interference originated from the neighboring eNodeB elements. The control channel proportion m degrades also slightly the link budget. In the link budget calculations, the effect of the antenna of the LTE terminal can be estimated a 0 dB if no body loss is present near the terminal. In case of the external antenna, the antenna gain increases respectively, but the logical estimate of the average terminal type is a stick model with only its own in-build antenna.

The effect of the data speed can be estimated in a rough level by applying a rule of thumb that assumes a 160 dB path loss in uplink with 64 kb/s data rate. Whenever the bit rate grows, the maximum allowed path loss drops respectively. A simple and practical assumption is that the doubling of the data rate increases the path loss by 3 dB. This can be assumed to work sufficiently well, unless the channel coding or modulation scheme changes.

Figure 21.8 and Figure 21.9 show the theoretical calculation for the effect of the data rates.

Table 21.2 LTE radio link budget example in the downlink direction

Downlink		
Transmitter, eNodeB	**Unit**	**Value**
Transmitter power	W	40.0
Transmitter power (a)	dBm	46.0
Cable and connector loss (b)	dB	2.0
Antenna gain (c)	dBi	11.0
Radiating power (EIRP) (d)	dBm	55.0
Receiver, terminal	**Unit**	**Value**
Temperature (e)	K	290.0
Bandwidth (f)	Hz	10000000.0
Thermal boise	dBW	−134.0
Thermal Boise (g)	dBm	−104.0
Noise figure (h)	dB	7.0
Receiver Boise floor (i)	dBm	−97.0
SINR (j)	dB	−10.0
Receiver sensibility (k)	dBm	−107.0
Interference margin (l)	dB	3.0
Control channels share (m)	dB	1.0
Antenna gain (n)	dBi	0.0
Body loss (o)	dB	0.0
Minimum received power (p)	dBm	−103.0
Maximum allowed path loss, downlink		158.0
Indoor loss		15.0
Maximum path loss for indoors, downlink		143.0

21.5.2 Radio Propagation Models

Some of the widely known path loss prediction models are Okumura-Hata and Walfisch-Ikegami, which are suitable also for the LTE radio network planning as soon as their functional ranges of the antenna height, operating frequency and maximum predictable cell range are taken into account.

The selection of the suitable radio propagation model in the initial phase of the radio network planning depends on the utilized frequency of LTE, the terrain, and the level of accuracy that is wanted. In the early phase of the actual deployment, it is essential to validate the functionality and accuracy of the propagation models, in order to do the respective model tunings for example, as for the cluster types and respective attenuation value. It may be necessary to change the prediction model if the resulting accuracy is still low. One of the limitations is, that the radio network planning tools have only a certain set of propagation models, and the utilized digital map data causes also its own effect on the prediction accuracy. The typical method for the coverage estimate is to estimate the maximum functional path loss for the investigated case (throughput, quality) by defining subregions of, for example, 100 m x 100 m. Within these areas, a sufficient amount of snapshot path loss calculations are done by taking into account the realistic signal variations of the slow and fast fading. If the location probability of this specific subarea is higher than the wanted quality requires, the subregion is selected as functional, otherwise it represents outage of the coverage.

Table 21.3 *The LTE radio link budget in the uplink direction*

Uplink		
Transmitter, terminal	Unit	Value
Transmitter power	W	0.3
Transmitter power (a)	dBm	24.0
Cable and connector loss (b)	dB	0.0
Antenna gain (c)	dBi	0.0
Radiating power (EIRP) (d)	dBm	24.0
Receiver, eNodeB	Unit	Value
Temperature (e)	K	290.0
Bandwidth (f)	Hz	360000.0
Thermal Boise	dBW	−148.4
Thermal Boise (g)	dBm	−118.4
Noise figure (h)	dB	2.0
Receiver Boise floor (i)	dBm	−116.4
SINR (j)	dB	−7.0
Receiver sensibility (k)	dBm	−123.4
Interference margin (l)	dB	2.0
Antenna gain (m)	dBi	11.0
Mast head amplifier (n)	dB	2.0
Cable loss (o)	dB	3.0
Minimum received power (p)	dBm	−131.4
Maximum allowed path loss, uplink		155.4
Smaller of the path losses:		155.4
Indoor loss		15.0
Maximum path loss in indoors, uplink		140.4
Smaller of the path losses in indoors:		140.4

When seeking for the high accuracy, the size of these subregions should be small which results in the longer times with the path loss prediction calculations. In the practice, the typical raster sizes vary between 25 m x 25 m in urban areas to 500 m x 500 m in rural areas.

For the largest coverage area cases, another suitable model for varying environments is ITU-R P.1546. The model is based on predefined curves for the frequency range of 30 MHz to 3000 MHz and for maximum antenna heights of 3000 m from the surrounding ground level. The model is valid for terminal distances of 1 to 1000 km from the base station over the terrestrial and sea levels, or for the combination of these. This model is especially suitable for the large area estimates, extending considerably the functional ranges of the antenna height and cell radius of Okumura-Hata.

If the investigated frequency or antenna height does not coincide with the predefined curves, the correct values can be obtained by interpolating or extrapolating the predefined values according to the annexes of the model and by applying the calculation principles presented in its Annex 5. The case curves represent field strength values for 1 kW effective radiated power level (ERP), and the curves have been produced for the frequencies of 100 MHz, 600 MHz and 2 GHz. The curves are based on the empirical studies about the

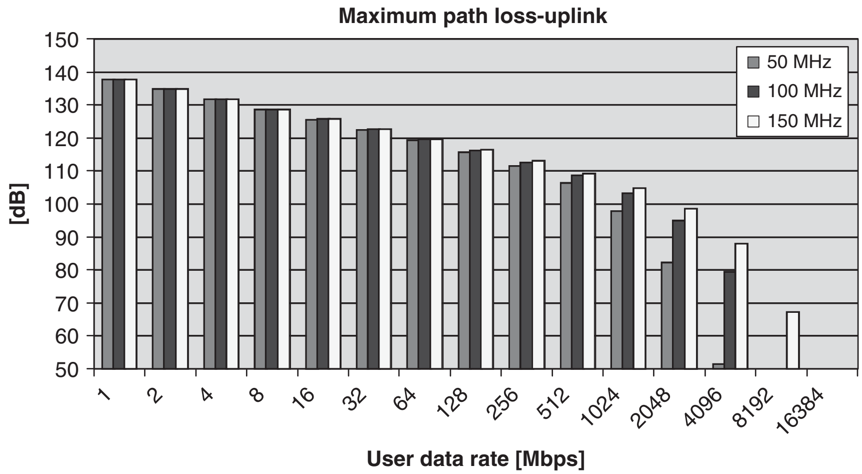

Figure 21.8 *A theoretical estimation for the effect of the data rate on the uplink path loss.*

propagation. In addition to the graphical curve format, the values can be obtained also in a tabulated numerical format.

It can be assumed that the basic and extended versions of Okumura-Hata as well as ITU-R P.1546-3 models provide a sufficiently good first-hand estimate for the LTE coverage areas and respective capacity and quality levels in the initial network planning phase. Nevertheless, there are several other models, including ray-tracing type of estimates for the dense city centers. These models require more detailed digital map data with respective terrain height and cluster attenuations. In the most advanced prediction models, a vector-based 3D map is needed. It logically has a cost effect on the planning but it increases considerably the accuracy of the coverage estimate. It can further be enhanced via local reference measurements by adjusting the model's estimate accordingly. As a cost-efficient compromise, 3D models could be utilized in the advanced phase of the radio network planning in the most important areas.

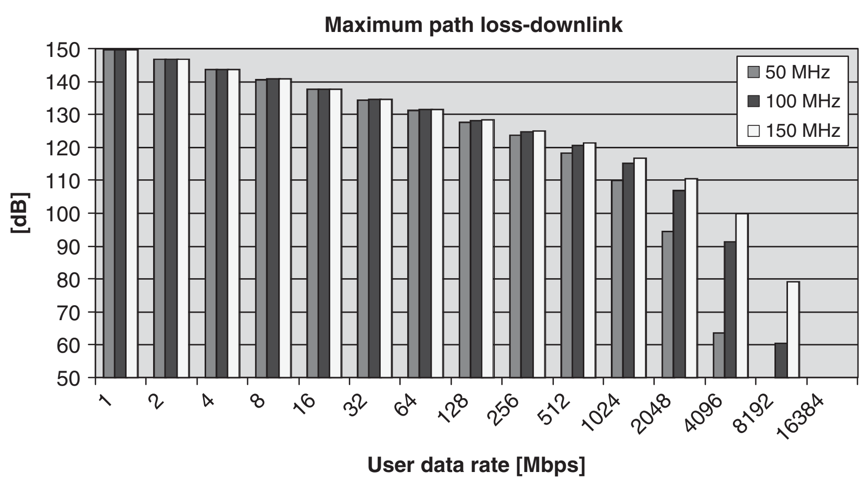

Figure 21.9 *The effect of the data rate on downlink.*

21.5.3 Frequency Planning

The LTE radio network can be constructed in the same band, that is, the reuse pattern size of 1 can be used. As in the case of W-CDMA, it produces a certain level of interferences. The benefit of this solution is that the users can take the full advantage of the high bandwidth with respective peak data rates.

Another solution is to divide the available LTE frequency band to smaller blocks in order to create higher reuse pattern sizes. As an example, if the operator has been granted with a total of 15 MHz for LTE, it can be defined as such for each eNodeB sites which provides the highest possible data rates within this band, although the average data rate suffers at some extend due to the intercell interferences when other users are present. If the 15 MHz block is divided into 5 MHz slices, it gives a possibility to utilize 3 different frequencies per sector. This means that the reuse size of 3 can be applied, and the interferences can effectively be reduced. Nevertheless, the peak data rate per user lowers into third of the previous example die to the bandwidth limitation.

The simulations have shown that despite the increased interference levels, the first case, that is, reuse size of 1, provides higher capacity. Based on this result, a very deep level neighboring plan is not required in the LTE radio network planning. In any case, the overlapping proportions of the network can be optimized by planning carefully the antenna tilting and the balance between the site distances and power levels.

21.5.4 Other Planning Aspects

The tracking area planning is one of the items that impacts to the capacity of the LTE network. If the tracking areas are too small, it results in increased signaling when the LTE-UEs are moving over the tracking area borders. On the other hand, if the tracking areas are too large, the network initiated call and respective paging signaling loads respectively the network. The optimal size of the tracking areas is thus one of the many optimization tasks of the operator in order to find a proper balance for the signaling load.

As some rules of thumb, the tracking areas of LTE could be defined by default the same as the location areas of 2G, and routing areas of GPRS, in case the operator has already previous infrastructure of mobile communications. It can be assumed that the already existing network has had time to mature and a sufficiently good balance of the LA/RA has been found which can be used as such for the LTE TA.

Another guideline is that the tracking area border should not be defined in an area which contains heavily moving LTE-UEs, as is the case in motorways.

21.6 Self-Optimizing Network

Along the deployment of the LTE networks, there will be clearer needs for the Self- Organizing / Optimizing Network concept (SON). The SON concept refers to the complete and automatic optimization of the network via several functions. In fact, it can be claimed that the full spectral efficiency of the LTE network can only be achieved when SON is applied partially or with all possible functionalities. The utilization of SON saves time and resources in the effective response for the changing user load profiles, faults and other dynamically changing phenomena in the network. This, in turn, leads to enhancements in the costs by reducing the expenses both in the deployment (savings of CAPEX, Capital Expenditure) and in the operational phase of the LTE network (savings of OPEX, Operational Expenditure).

SON is a relatively wide concept and has been standardized as a joint effort of 3GPP and NGMN Alliance (Next Generation Mobile Networks Alliance). As a result of this common work, there is already a whole set of SON features specified under the 3GPP Release 8 and 9 documents.

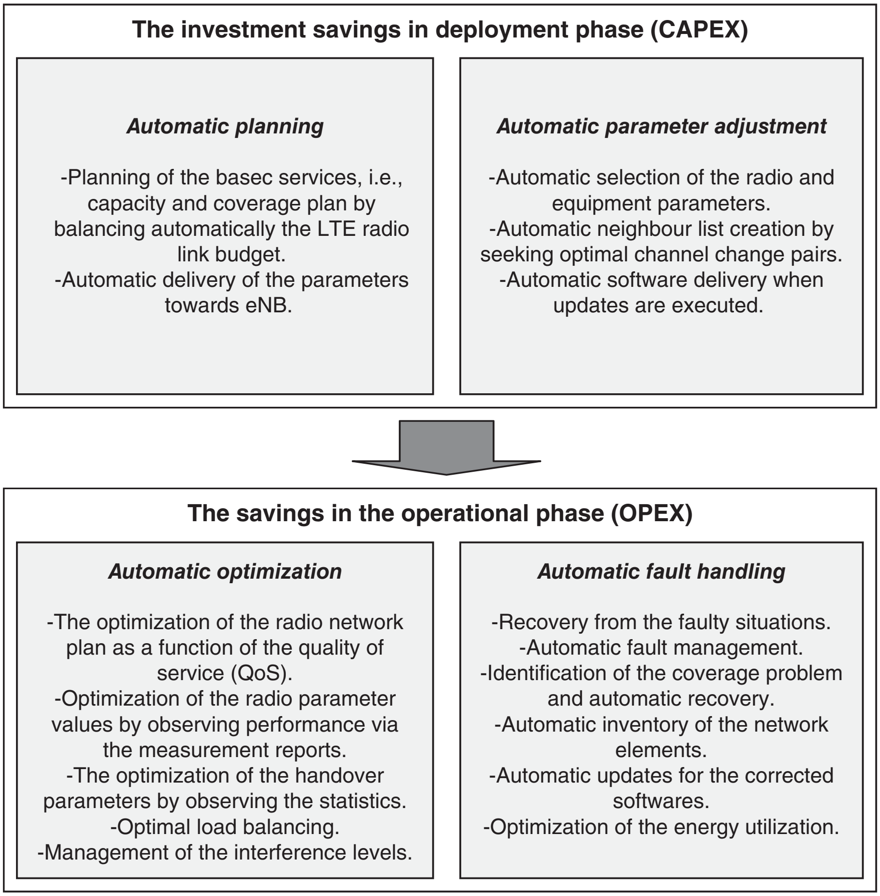

Figure 21.10 *The high-level idea of the self-optimizing network concept.*

The main focus of SON is to minimize the physical work of the technical personnel, that is, to reduce the eNB and element based tuning that traditionally require time and efforts, including the physical travelling to the site, the manual measurements of the faulty elements, software updates, frequency changes and so on. The list of the enhancements is relatively long, and it logically takes time to design and apply these functionalities to the LTE network. It should also be remembered that the highly skilful technical personnel cannot be replaced completely by the utilization of automatic devices and software. Also SON concept itself requires management and tunings, and there is always a part of the trickiest problems that require human intervention to be measured, analyzed and fixed. In any case, the work has been initiated and it can be expected that the self-optimizing functionalities will reduce the costs of the LTE compared to the basic solutions. Figure 21.10 shows the basic idea of the SON for the initial and operational phases of the LTE network.

It can be generalized that the good aim of SON is to provide a certain level of "plug-and-play" functionalities for LTE operators. The strong benefit of the SON concept is that it has been standardized and thus the solutions are global for the LTE system. This provides guaranteed support for the concept for the wide deployment. SON would provide with much faster deployment and parameter tuning of the LTE network as the proportion of the manual work is lower that in the previous mobile network environments. This also means that instead of routine tasks, for example, in the parameter adjustments, the skilful technical personnel can concentrate on more demanding problem solving and on deeper network optimization that inevitably still remains even with the SON concept.

21.7 Parameter Planning

LTE capacity depends on the services and the quality of service level. The capacity is tight to the number of the LTE eNodeB elements and to the bandwidth each is offering. The modulation and coding scheme dynamics results in the highest data rates and thus highest capacity near eNodeBs whereas the cell edge offers minimum capacity. As the bit error rate increases along with the utilization level of the network, it also has an effect on the single coverage areas of eNodeBs.

LTE offers possibilities to combine transmissions in the radio interface as in UMTS, offering soft handover, so the overlapping also provides enhancements to the capacity and quality of service.

The capacity of LTE can be estimated via simulations by varying the number of users, and the capacity each application requires. The simulation results can be ordered as a function of modulation schemes and coding and minimum SINR, as well as on the average data transmission capability of single cells. These results depend on the network deployment, that is, the distance between the eNodeBs, the topology of the area, and distribution of the users with varying data rate needs.

Typical SINR values in LTE are 0–3 dB for QPSK, 7–11 dB for 16-QAM and 12–15 dB for 64-QAM, taking into account the varying coding schemes.

It should be noted that the final data rate offered to the users depend on the complete chain of the interfaces, including the core network as emphasized in Figure 21.11.

The dimensioning of the complete LTE network can be done, for example, in such a way that the transport network supports the combined average data speed of all eNodeB elements attached to that part of the transport network, to assure there is as much average capacity available as possible. Furthermore, the capacity dimensioning of transport network can be done based on the peak data rate of a single eNodeB element which should be possible to deliver as such.

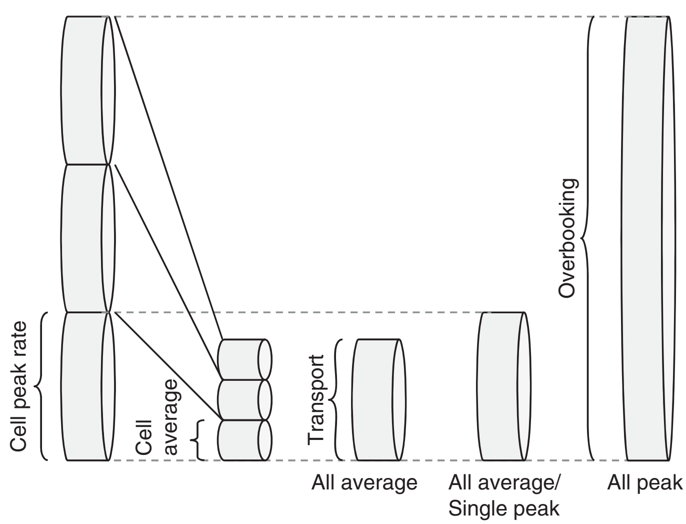

Figure 21.11 *The data rate dependency in LTE network when dimensioning the LTE networks based on the air interface capacity.*

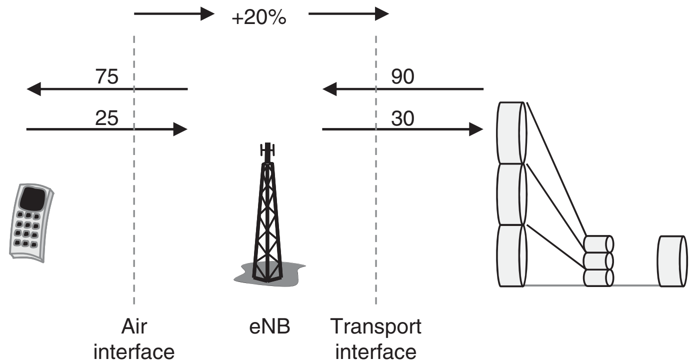

Figure 21.12 An example of the dimensioning principle of the complete LTE network, data rates in Mb/s. This case shows the dimensioning of the network based on all-average / single-peak throughput in 1+1+1 / 10 MHz cell design.

Figure 21.12 shows an example of cell dimensioning the transport to support the aggregated average capacity of all cells, yet supporting still minimum of the peak capacity of a single cell. In the dimensioning, one principle may be to reserve a maximum of 3 times the average rate for the peak rate. In this example, the assumption is 3 cells and 10 MHz bandwith and the following data rates: (1) 17 Mb/s DL net value for physical average rate per cell; (2) 7 Mb/s UL net value for physical average rate per sell; (3) 75 Mb/s DL net value for physical peak rate per cell with 64-QAM and 2x2 MIMO; (4) 25 Mb/s UL net value for physical peak rate per cell with 16-QAM. Please note that the values of 1 Mb/s for M-plane, 300 kb/s for C-plane and 30 ms busts of X2 U-plane are not included in this specific example.

As a general trend, the offered capacity per technology is increasing as indicated in the spectral efficiency analysis of Figure 21.13.

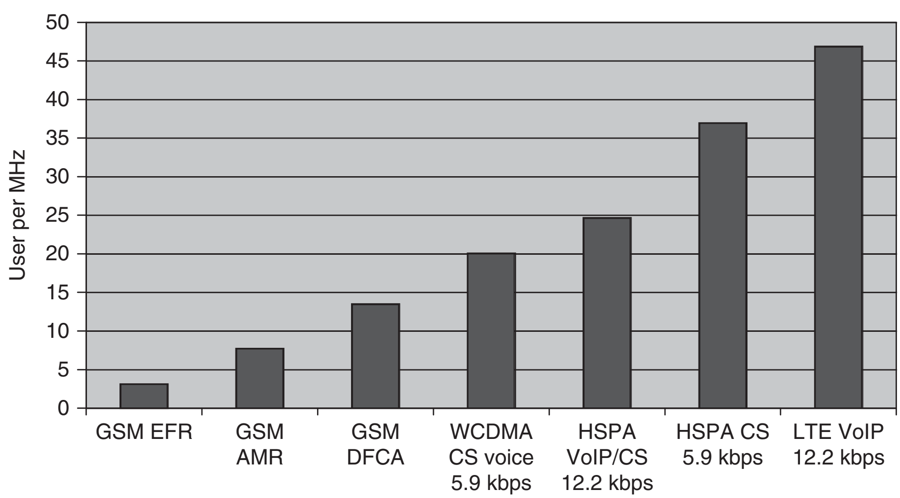

Figure 21.13 The capacity development of the voice calls.

Table 21.4 *The theoretical data rates per bandwidth for QPSK, 16-QAM and 64-QAM*

Resource blocks:			6	15	25	50	100
Subcarriers:			72	180	300	600	1200
			Bandwidth (MHz)				
Modulation	Coding	MIMO	1.4	3.0	5.0	10	20
QPSK	1/2	SISO	0.8	2.1	3.6	7.2	14.3
16-QAM	1/2	SISO	1.7	4.3	7.2	14.3	28.7
16-QAM	3/4	SISO	2.6	6.4	10.7	21.5	43.0
64-QAM	3/4	SISO	3.9	9.7	16.1	32.2	64.5
64-QAM	4/4	SISO	5.1	12.9	21.5	43.0	86.0
64-QAM	3/4	2x2	7.7	19.3	32.2	64.5	129.0
64-QAM	4/4	2x2	10.3	25.8	43.0	86.0	172.0
64-QAM	4/4	4x4	20.6	51.6	86.0	172.0	343.9

The capacity of LTE depends on various parameters. In theory, the offered LTE capacity can be estimated in the following way for the downlink, as for the downlink peak bit rates, when 2x2 MIMO antennas and 64-QAM are applied:

- Control channel overhead is 1 out of 14 = 7.1%.
- Reference symbol overhead is 7.7%.

This results in 172 Mbps in 20 MHz and 86 Mbps in 10 MHz bandwidths. Table 21.4 shows more examples for the QPSK, 16-QAM and 64-QAM.

In the uplink direction, the uplink peak bit rate can be estimated in theory by taking into account the following for the single stream transmission and 16-QAM:

- Control channel 1 resource block varies between 1 and 17%.
- Reference symbol overhead 1 out of 7 resulting in 14.3%.

This results in the data rates of 57 Mbps in 20 MHz and 28 Mbps in 10 MHz bandwidth. More data rate values can be observed in Table 21.5.

As can be seen in these examples, the uplink and downlink rata rates depend on various parameter values. For the operators, it is essential to offer sufficiently wide bandwidth in order to facilitate the data rates. The rest is matter of balancing the data rates per modulation scheme which basically means that there are different areas that offer different data rates depending on the distance between eNodeB elements and mobile devices.

For the operators that can have, for example, 10 MHz or less bandwidth per RF band, there is still one option to be used. In case the operator has two bands, the Carrier Aggregation (CA) can be used to combine data streams, and in theory, the achievable data rate is direct result of the combined rates. So, if the operator has, for example, 10 MHz in LTE band 2 and 10 MHz in LTE band 4, the user is able to experience the same data rates and in a single band with 20 MHz capacity, that is, double speed. It should be noted that the CA is independent from the uplink and downlink, which means that if the operator has only downlink capacity in a certain band, it can be combined via CA with the downlink of other RF band for the increased reception of data.

Table 21.5 *Theoretical uplink data rates for QPSK, 16-QAM and 64-QAM*

Resource blocks:			5	14	24	49	99
Subcarriers:			60	168	288	588	1188
			Bandwidth (MHz)				
Modulation	Coding	MIMO	1.4	3.0	5.0	10	20
QPSK	1/2	SISO	0.7	2.0	3.5	7.1	14.3
16-QAM	1/2	SISO	1.4	4.0	6.9	14.1	28.5
16-QAM	3/4	SISO	2.2	6.0	10.4	21.2	42.8
16-QAM	1/1	SISO	2.9	8.1	13.8	28.2	57.0
64-QAM	3/4	SISO	3.2	9.1	15.6	31.8	64.2
64-QAM	1/1	SISO	4.3	12.1	20.7	42.3	85.5
64-QAM	1/1	V-MIMO	8.6	24.2	41.5	84.7	171.1

Table 21.6 shows still link budgets for downlink with 1 Mb/s data rate. It can be seen that this link budget is similar to an uplink link budget of 64 kbps.

It may also be noted that the link budget for LTE plink with 64 kb/s is at least similar with HSPA 64 kb/s, or with GSM link budget for voice service. The differences are basically for the fast fade margin, macro diversity and interference margin. Table 21.6 shows a case with 40 W base station transmission, 2 dB cable loss, 10 MHz bandwidth. Furthermore, Table 21.7 presents the link budget for 24 dBm device with 2 resource blocks reserved for 64 kb/s data rate.

The downlink link budget is relatively similar with LTE and other OFDM systems like DVB-H. It can be seen that in this specific case, the downlink–uplink link balance is achieved in practice.

Table 21.6 *Downlink link budget for HS-DSCH, when data rate is 1024 kb/s*

Term	TX, NodeB	Value	Info
a	HS-DSCH power (dBm)	46.0	
b	TX antenna gain (dBi)	18.0	
c	Antenna feeder loss (dB)	2.0	
d	EIRP (dBm)	62.0	d=a+b+c
Term	**RX, UE**	**Value**	**Info**
e	UE noise figure (dB)	7.0	
f	Thermal noise (dBm)	−104.5	f=Boltzmann*T*B; T=290K, B=9 MHz
g	Receiver noise floor (dBm)	−97.5	g=e+f
h	SINR (dB)	−10.0	Based on simulations
i	Receiver sensitivity (dBm)	−107.5	i=g+h
j	Interference margin (dB)	4.0	
k	Control channel overhead (dB)	1.0	
l	RX antenna gain (dBi)	0.0	
m	Body loss (dB)	0.0	
n	**Maximum path loss (dB)**	**164.5**	n=d-i-j-k+l-m

Table 21.7 *Link budget for uplink when data rate is 64 kb/s*

Term	TX, UE	Value	Info
a	Max TX power (dBm)	24.0	
b	TX antenna gain (dBi)	0.0	
c	Body loss (dB)	0.0	
d	EIRP (dBm)	24.0	d=a+b+c
Term	**RX, NodeB**	**Value**	**Info**
e	NodeB noise figure (dB)	2.0	
f	Thermal noise (dBm)	−118.4	f=Boltzmann*T*B; T=290K, B=360 kHz
g	Receiver noise floor (dBm)	−116.4	g=e+f
h	SINR (dB)	−7.0	Based on simulations
i	Receiver sensitivity (dBm)	−123.4	i=g+h
j	Interference margin (dB)	1.0	
k	Cable loss (dB)	2.0	
l	RX antenna gain (dBi)	18.0	
m	MHA gain (dB)	2.0	
n	**Maximum path loss (dB)**	**164.4**	n=d-i-j-k+l-m

21.7.1 eNodeB Transmitter Power

For proper designing of the eNodeB power levels in the link budget, realistic limitations need to be known. The maximum useful power level depends on the HW manufacturer and variants.

One typical scenario could be that for FD-LTE, there can be support for 8, 20, 40 and 60 W categories as for the HW is considered, but the final power level is activated via corresponding SW license. It is also important to note if the HW supports single or multiple sectors, that is, in this case, the maximum power level could be either 1 x 60 W for solely one sector, or 60 + 60 + 60 W for a total of 3 sectors. Again, typically these variations are possible to modify via SW license if the respective power amplifiers are included in the HW.

In addition to the actual power level activations, the vendor might offer the possibility to activate certain power levels for certain scenarios in such a way that the power distribution can be varied along with the enhancements of network deployment. The operator can thus select and modify the power levels, for example, by offering via the same HW a low-power micro cell (e.g., 8 W), or high-power rural cell (e.g., 60 W). Table 21.8 shows examples of the power levels.

The SW license activation for the HW features like power levels and number of sectors are beneficial for both equipment vendors as well as for the network operators, as the power levels are already included in the HW, and along with increased traffic, the operator can decide case by case the inclusion of additional capacity without spending the whole CAPEX from the day one. Furthermore, the bandwidth value can also possibly be varied upon the need.

Similar type of configurations is possible to offer for TD-LTE. The Power Amplifier could support, for example, 20 + 20 W via remote radio heads. 3GPP has specified power categories for UE, that is, the broadband transmit power. Table 21.9 shows the specified power classes.

The eNodeB allocates constant power per subcarrier in downlink transmission, which is configurable by the operator as an O&M parameter. The total eNodeB power is shared among all subcarriers, no matter how many of them are used for data allocation. The lower the number of subcarriers assigned to the user, the less power is received at the UE.

Table 21.8 *An example of a possibility for adjusting the transmission power levels and bandwidths of the same eNodeB HW based on the license*

Bandwidth (MHz)	Number of PRB	The resulting eNodeB power (W)
1.4	6	8
3.0	15	20
5.0	25	40
10	50	60
15	75	60+60
20	100	60+60

UE output power is shared between subcarriers assigned to transmission. This means that, when UE operates with smaller amount of subcarriers, it distributes the available power only over the used subcarriers. This results in increased Tx power per subcarrier and, consequently, improves the uplink coverage. Although the uplink direction is controlled by the Power Control algorithm (Open Loop Power Control or Closed Loop Power Control with additional correction factors), it is not relevant for the link budget because the focus is on estimating maximum possible coverage with the UE located at the cell boarder and for this reason the maximum output power of 23 dBm can be assumed.

Power allocation Downlink Power sharing among subcarriers means that if the coverage is limited by Tx power, the planner should increase the output power or decrease the channel bandwidth. Uplink Smaller amount of scheduled resources means higher Tx power per subcarrier. On the other hand, it requires higher modulation and coding scheme (MCS) with better SINR (Signal to Interference and Noise Ratio). Additionally, spreading info bits over smaller amount of resources leads to worse frequency diversity.

21.7.1.1 Cable and Connector Loss

Figure 21.14 shows a realistic example of the cable loss of typical cellular base station site.

The attenuation of each antenna system element depends on its physical properties and can differ for the various dimensioning scenarios (the impact of operating band, cable length, etc.).

There is also a feederless solution possible to apply in the base station antenna systems. This is the most beneficial site construction, enabling very good overall RF performance and leading to a reduced number of sites. The great advantage is that such a site can be deployed with the standard Flexi Triple RF Module mounted close to the antenna system. There are only jumper cables between the Flexi RF Module and antenna connectors. This is the best performing solution from the coverage and capacity point of view because of the small loss of about 0.4.

Table 21.9 *UE TX power classes as defined by 3GPP*

Class	Power (dBm)	Tolerance (dB)
1	30	n/a
2	27	n/a
3	23	+/−2
4	21	n/a

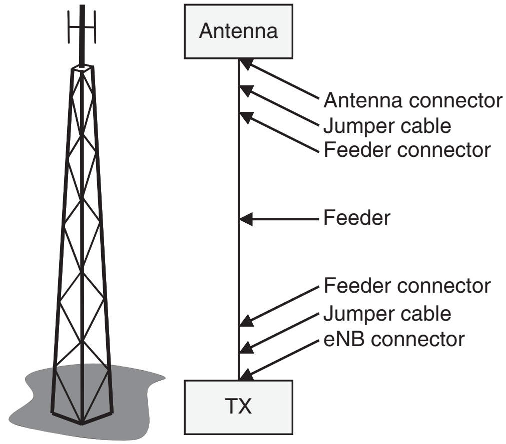

Figure 21.14 *An example of the components that have effect on the cable loss.*

A feeder solution is a standard construction with a long cable between the BTS RF Module and the antenna connector, the attenuation of which depends on its length and the configured operating band. Assuming 15 m cable and jumpers to the antenna/RFM connectors, there is almost 2 dB cable loss in both directions.

Feeder solution with TMA/MHA is third type of solution. If the cable length starts to be the limiting factor for the cell coverage, there might be a need for a Tower Mounted Amplifier (TMA), sometimes called Mast Head Amplifier (MHA). It is possible to compensate UL cable loss with the introduction of TMA. The additional loss must be taken into account for DL direction, though. The TMA is one more element at the antenna line and causes about 0.5 dB loss in DL.

A typical feederless solution offers an optical connection between System Module and RF unit up to 200m. With the introduction of the Distributed Sites feature of selected base station products, it is possible to connect an RF unit located much farther away from the System Module. This is possible with single mode optical transceivers (SFP) installed within Tx/Rx path. This Distributed Sites feature has no explicit impact on coverage dimensioning. Since an optical connection is applied, cable loss shall be considered the same as for the feederless solution. Figure 21.15 compares the solutions described above.

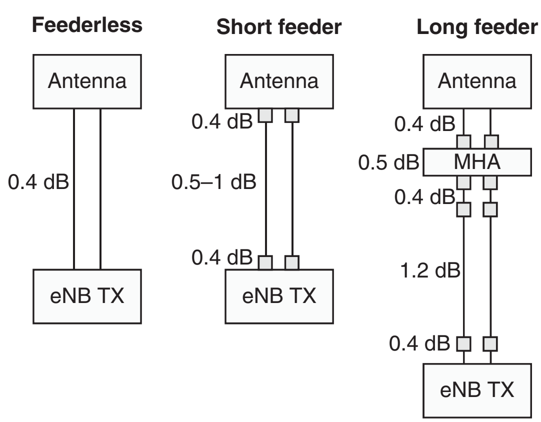

Figure 21.15 *The practical values for elements.*

It is recommended to assume 0.4 dB feeder loss at eNodeB in both directions when considering the feederless site. In this case, there is no need for TMA. Feeder solution is typically characterized by 2 dB feeder loss, compensated in UL to 0 dB if the TMA is introduced (the TMA reduces eNodeB noise figure by 0.2 dB related to Friis formula). Due to the great flexibility of LTE RF parts, it is highly recommended to assume the feederless solution sites for dimensioning unless stated otherwise by general deployment assumptions. It is recommended to assume 0 dB feeder loss at UE for a typical dimensioning case.

Power Design, Case Example 1 (Feederless) Feederless solution offers high capacity and coverage with less sites with better overall RF performance. The total loss between eNb and antenna may be around 0.4 dB in this solution due to the jumper, both for UL and DL.

The DL RF power in a typical case may thus be 43 W – 0.4 dB = 39 W, or 2 x 19.5 W, meaning that the power loss is about 10% in the feederless case.

Power Design, Case Example 2 (Short Feeder) In a relatively short feeder line, that is, for shorter than 15 m of 7/8" coaxial cable, the jumper loss would be 2 x 0.4 dB (eNB and antenna sides), feeder loss 0.5–1 dB (7/8", 2.1 GHz, cable loss may be about 1 dB/15m). In the 43 W case, the antenna connector will receive power of 43 W – 1.3 dB = 32 W (or 2 x 16 W) which corresponds to -25% for 7/8" and 7.5 m cable in DL. In UL, the value may be about 142 dB – 1.3 dB = 140.7 dB which corresponds to -12% for 7/8" and 7.5 m cable.

Power Design, Case Example 3 (Long Feeder) For a long feeder, that is, more than 20 m, the DL RF power in the feeder may be 43 W – 3 dB = 21.5 W or 2 X 10.75 W resulting -44% degradation for 7/8" and 20 m cable. For the UL, the site coverage may be higher than in the feederless case depending on the antenna line quality.

21.7.1.2 Body and Car Loss

Body loss should be considered only when dimensioning is done for a handset (voice service at the cell-edge). On the other hand, there is no need to do so if one talks about CPE (Customer Premises Equipment) at home providing the broadband Internet service with the rooftop/outside-mounted antenna. Dimensioning for two-sector sites located along highways/tracks should be done with additional in-car/in-train loss. The body loss can be assumed to be around 3 dB in typical cases. The in-car loss can be assumed to be around 6 dB.

21.7.1.3 Thermal Noise and Noise Figure

For the link budget calculation there is no need to perform calculations for the exact frequency ranges. In practice, it is sufficiently accurate to assume that only the main center frequencies are considered for link budget calculation. As an example of a case of Band 4 or 10 (1700/2100; AWS) where band separation need to be considered, it is recommended to set two separate link budget scenarios, one for downlink and the other one for uplink and then take the minimum values which limit the cell range.

The bandwidth configuration impacts factors such as overhead ratio and total cell throughput. In case of the basic LTE system, the best network performance (regarding maximum peak data rates and cell throughputs) is achieved by the deployment of 20 MHz bandwidth. One should expect certain performance degradation especially for 1.4 MHz and 3 MHz bandwidth because of worse scheduling gain as well. The LTE specifications define the bandwidths shown in Table 21.10.

The supported bandwidth depends on the vendor solutions, so not necessarily all the bands are supported in the first day. The bandwidth has a direct impact on the maximum offered capacity.

Table 21.10 The LTE bandwidths for Release 8 and 9 solutions

Bandwidth [MHz]	1.4	3	5	10	15	20
No. of PRBs	6	15	25	50	75	100
Subcarrier spacing	15 kHz					

The receiver noise figure depends on the receiver equipment design and represents the additive noise generated by various hardware components. That is why the value should be parameterized in the particular receiver's vendor specification. Also, the receiver noise figure varies depending on the existence of MHA. As an example, the equipment with MHA may have, for example, 2 dB noise figure and 2.2 dB without TMA. The typical UE noise figure can be assumed to be about 7 dB.

The minimum *C/N* ratio that is required for the successful reception depends on the combination of the modulation, code rate and MPE-FEC. Also the radio channel type has a clear effect.

21.7.1.4 Fading Margin

In a cellular network, a propagating radio wave meets many obstacles on the radio path. For indoor environments these obstacles can be floors, walls or furniture. In outdoor environments they can be buildings, cars or just some specific terrain shape. In all cases the radio wave can be reflected, refracted or scattered as presented in Figure 21.16. Hence, the signal received at the receiver is a product of waves passing through paths of different lengths. Multipath propagation is related to the following effects causing signal fading and distortion:

- Shadowing also known as log-normal fading or slow fading.
- Rayleigh fading also known as fast fading.
- Delay spread.

The shadowing margin is the amount by which a received signal level may be reduced without causing system performance to fall below a specified threshold value as shown in the equation below.

$$\text{LNF} = x^{*}\sigma,$$

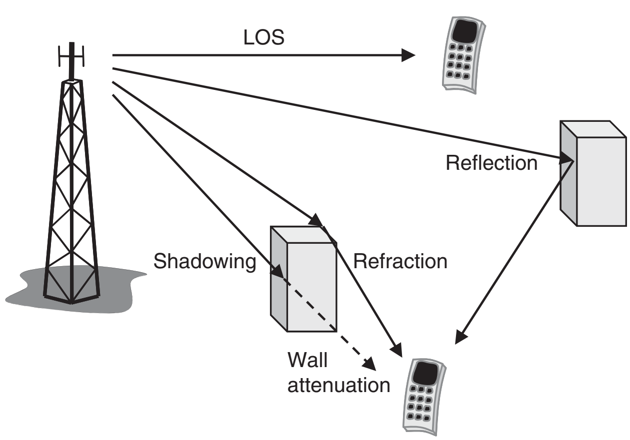

Figure 21.16 *The idea of the reflected RF components.*

Table 21.11 The LNF margin values

Location probability at cell border [%]	Log-normal fading margin [dB]
50	σ x 0
60	σ x 0.25
70	σ x 0.52
75	σ x 0.67
80	σ x 0.84
85	σ x 1.04
90	σ x 1.28
91	σ x 1.34
92	σ x 1.4
93	σ x 1.48
94	σ x 1.55
95	σ x 1.64
96	σ x 1.75
97	σ x 1.89
98	σ x 2.05
99	σ x 2.33

where σ is standard deviation of the slow fading distribution, and x is the argument of the normal cumulative density function F(x) the values of which define the cell edge location probability. The LNF margin values in dependency of the cell edge probability are presented in Table 21.11.

LNF margin is calculated as x * σ, where σ is the standard deviation of the slow fading distribution whereas x is the argument of the normal cumulative density function F(x) the values of which define the cell edge location probability. LNF margins are presented in Table 21.12.

There are two different ways to describe the outage probability. One is the probability related to the cell edge and the other one is related to the cell area. The dependency between these two probability figures is a function of the propagation exponent and the standard deviation. To consider the slow fading impact for a given target cell area probability (usually provided as a design target), Jakes formula must be used. It can be applied only in case of signal reception from one cell. The impact of multicell coverage is considered as the gain against shadowing. This formula gives a relation between the cell edge probability and the corresponding cell area probability. The cell area probability is the probability that an UE receives, within the entire area of the cell, the signal in such a way that at least the "required SINR" condition is fulfilled.

Table 21.12 LNF margins

Location probability	Standard deviation of the signal level		Log normal fading margin (LNF)
In whole cell			On cell edge
95%	86.9%	9 dB	10.1 dB
86.0%		8 dB	8.6 dB
84.9%		7 dB	7.2 dB
78.7%		4 dB	3.2 dB

Table 21.13 Examples of the slow fading margins

Deployment class			Clutter type		
Dense Urban	Urban		Suburban	Rural	
Location probability [%]	Basic	82.6	81.5	81.5	69.8
Mature	84.7	83.7	83.7	73.3	
High end	86.9	86.0	86.0	85.0	
Standard deviation for LNF [dB]	Basic	9	8	8	7
Mature	9	8	8	7	
High end	9	8	8	7	
eNodeB height [m]	Basic	30	30	30	50
Mature	30	30	30	30	
High end	30	30	30	30	
Penetration loss [dB]	Basic	20	15	10	5
Mature	22	17	12	10	
High end	25	20	15	15	

A standard deviation is the deviation from the mean value. This parameter is used for log-normal fading calculation. The standard deviation value is derived from measurements and depends on the clutter type or simulation case. The following deployment classes presented in Table 21.13 have been defined in order to systemize different slow fading related parameters depending on the clutter type.

Please note that the deployment class definition also determines eNodeB height (for propagation model calculation) and the penetration loss. Additionally, penetration loss depends on the carrier frequency. The above listed values refer to 2.1GHz.

There is no need for assuming fast fading margin in the LTE link budget mainly due to very short TTI (1 ms) and fast AMC/scheduling mechanisms preventing falling into the fading gaps. In the presented dimensioning approach the impact of fast fading and user velocity is covered by selection of an appropriate SINR value coming from link level simulation results. For the calculation of shadowing margin, it is recommended to use the mature deployment class.

21.7.1.5 Interference Margin

The "interference margin" is a dB value that represents, as a "summarized value" the safety margin that needs to be taken into account in the overall MAPL in order to compensate interference signals from neighbor cells under traffic load. As a general rule, the higher the traffic load in the neighbor cell, the more subcarriers are used there and, consequently, the higher the interfering signal, especially when frequency reuse 1 is applied (which means that the serving and the neighbor cells use the same subcarriers). The same margin is also included in 3G-specific cell range evaluations. However, the related formula cannot be explicitly applied for LTE. The reason is that the interference characteristic differs between OFDMA-based air interface and WCDMA-based system.

Intracell interference refers to any interference caused by the active terminals within the same cell. The 3G systems suffer from the cell breathing effect because of working with one frequency (frequency reuse = 1) and multiplexing the users based on the orthogonal codes assignment. The OFDMA-based systems outperform the CDMA systems and the mentioned phenomena should not be perceptible. The epexegesis stays in the nature of the OFDM transmission scheme, for which the separate subcarriers are orthogonal. Obviously, it refers to the subcarriers within one particular symbol. The unused ones remain without any meaning for the signal reception whereas the occupied ones cannot interfere with each others because they are orthogonal.

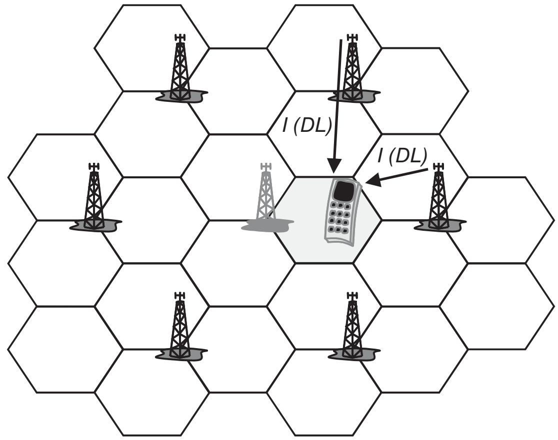

Figure 21.17 Downlink interference.

For this reason the number of multiplexed users in the frequency domain does not influence the interference level between them. Moreover, a single Resource Element is able to carry one user's bits, which is completely different from the CDMA solution where the multiplexed users occupy the same time and frequency resources. Assuming that there is no orthogonality loss inside a cell, no users experience intracell interference. A higher number of active users cause a decrease of available resources but it does not influence the G-factor (wideband C/I). There is no point in considering cell breathing phenomena, which is a well-known disadvantage of WCDMA networks. The eventual conclusion is that intracell interference does not exist and therefore does not require consideration of any additional link budget safety margins during dimensioning.

The situation becomes complicated when the neighboring cells are taken into consideration. The dependencies among the cells served by different base stations must be modeled. The general deployment approach does not foresee any reuse schemes; thus, all sites are allowed to work in the same frequency range (reuse = 1). The proper operation in a single cell is assured by the orthogonal property of FFT. Such a perfect situation is usually reflected in the reality for those terminals which are not located at the cell boundaries, while the cell-edge users suffer from the interference from the neighbor cells.

The aspect of downlink and uplink interference is depicted in Figure 21.17 and Figure 21.18. It is crucial to remember that the orthogonality between subcarriers is preserved only within one single OFDM symbol.

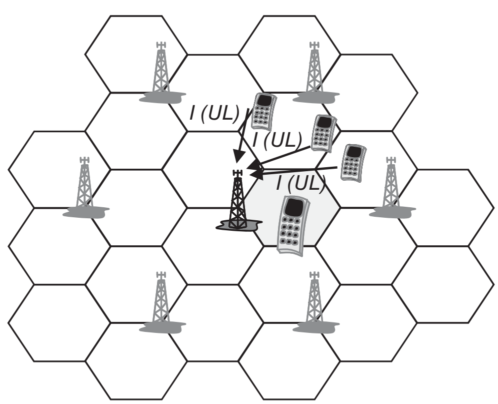

Figure 21.18 Uplink inter-cell interference.

The downlink interference margin can be obtained using analytical methods as presented in the following section, while the uplink interference margin values are usually taken from simulations due to nondeterministic user's distribution.

The higher frequency reuse schemes are possible but the study shows there is no significant gain in the network performance compared to the loss of available resources (peak data rate and capacity decrease). For the purpose of initial coverage dimensioning, one shall assume frequency reuse 1 meaning that all cells operate with the same spectrum.

21.7.2 Calculation of Downlink Interference Margin

For the estimation of the DL interference margin one can assume a constant power per subcarrier, which actually corresponds to the current product specification. The interference margin for LTE is not exactly the same as 3G-specific interference margin. The presented formula expresses the offset between:

- the required SINR value (determined by the selected MCS, the number of allocated PRBs and the scheduling gain in case of DL)
- the C/I at the cell-edge (obtained from a system level simulation for the given cell-edge definition – location probability).

In addition, since the C/I depends on neighbor cell interference, the result must be scaled due to the neighbor cell load. Whilst the location probability remains unchanged, the interference margin effectively depends on the neighbor cell load and the selected MCS. The higher MCS is, the higher margin is assumed. That is why the neighbor cell load of about 70–90% is realistic only for low MCSs. The same network load cannot be achieved for higher MCSs (16QAM / 64QAM). Moreover, this tendency is consistent with the presented cell-edge dimensioning approach. It is unlikely the network operates with the highest MCSs at the cell boundary. On the other hand, the cell load is used for capacity calculation. It determines the intercell interference impact and the average resource utilization in the target cell. Thus, the higher cell load is the higher interference level and the higher resource utilization as well (which can improve the capacity at some point).

The values for uplink interference margin can be obtained using the simulation results. In case of cell load lower than 35% one can use 1 dB by default.

21.7.2.1 Path Loss Estimate

The Transmission power per eNodeB antenna should take into consideration the deployment specific requirements and law regulations in the given country (power spectral density limitations). A general rule could be: low Tx power for small bandwidth, high Tx power for large bandwidth. UE Power Class 3 shall be assumed for dimensioning. It means 23 dBm mobile Tx power in the link budget calculation.

21.7.2.2 Building Loss

The building loss, or building penetration loss, can be estimated at a general level as an average value in different environment types, and it can thus be considered as a fixed radio link budget value for different area types. There are various recommendations for the median value, for example, 11 dB is typical value and a standard deviation value of 6 dB for the building loss. In addition to the building penetration loss, that is, the ratio of the average powers measured outside and inside the building with a fixed transmitter, also the

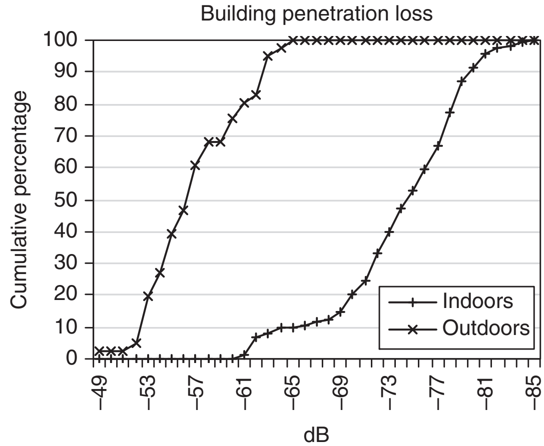

Figure 21.19 *An example of the typical building loss in a format of a cumulative RSSI histogram, which is obtained by the difference of the received power in the indoor and outdoor area. This case represents the deep building type, with frequency band 700 MHz and OFDM system.*

building floor loss may be important to take into account in the detailed network planning. The floor loss can be in certain cases approximately between 30 and 40 dB.

The building loss is frequency dependent. This is logical as the radio wave penetrates into the buildings depending on the conditions. As an example, a typical urban building normally contains metallic supports that might create a Faraday cage. Depending on the wavelength of the signal and the hole-size of the supporting metal the respective attenuation varies.

The average values are used, and the building penetration loss L_P is obtained by comparing the signal level distributions E inside and outside of the building as presented in Figure 21.19:

$$L_P = E_{out\_average} - E_{in\_average} \tag{21.1}$$

21.7.2.3 SFN Gain

One possible item in the radio link budget is the SFN gain, when applicable per system, which can enhance the performance in the overlapping areas, or provide the theoretical possibility to construct the sites further away from each others as the coverage area of each site cell rises. The interpretation of the benefit of SFN varies though.

The challenge of using the SFN item in the link budget is that there is no coherent definition available for the gain. It could be interpreted as the difference between the received power levels in dB, comparing a single stream with a varying number of streams. The value could also be resolved by mapping the $C/(N+I)$ distribution over the whole investigated area.

21.7.2.4 Other Effects

Seasonal conditions might cause low-level yearly fluctuations in the radio propagation due to the variations of the moisture level of vegetation and weather conditions (e.g. via occasional tunneling effects in the ionosphere). As an example, the field measurement results do have certain deviation due to the rain depending on the frequency range. It can be assumed though that in the typical link budget, the effect of the vegetation and rain is minimal for typical cellular and mobile TV low-bands. In case of the higher bands, the effect might be more considerable due to the radio propagation characteristics in higher frequencies. In any case, the seasonal path loss variation can be considered as a minor detail in a practical radio link budget, and due to the challenges in the periodical adjustment of cellular or broadcast type of network, it is not necessary to take into account.

21.7.3 Path Loss Prediction

The next step of the planning is to estimate the high-level cell coverage area in the planned area. The maximum path loss L (dB) is the difference between the effective isotropic radiating transmitter power P_{EIRP} and the required received power in outdoors $P_{\min(out)}$:

$$L = P_{EIRP} - P_{\min(out)}. \tag{21.2}$$

In this formula, P_{EIRP} (dBm) is:

$$P_{EIRP} = P_{TX} - L_{cc} - L_{ps} + G_{TX}. \tag{21.3}$$

The minimum received power level $P_{\min(out)}$ (dBm) is:

$$P_{\min(out)} = P_i + L_{lv} - G_{SFN} + L_{GSM}, \tag{21.4}$$

where P_i is the isotropic received power, L_{lv} is the location variation for a certain area probability, G_{SFN} is the SFN gain, and L_{GSM} is the GSM filter loss due to the isolation of the multiband receiver and GSM transmitter.

The isotropic received power is obtained from:

$$P_i = P_{RX\min} - G_{RX}, \tag{21.5}$$

where $P_{RX\min}$ is the receiver sensitivity, or the minimum power level the receiver requires, and G_{RX} is the receiver's antenna gain. The latter depends on the frequency.

The minimum required receiver's power level $P_{RX\min}$ is obtained by:

$$P_{RX\min} = P_n + (C/N)_{\min}, \tag{21.6}$$

where P_n is the receiver noise input power level, and $(C/N)_{\min}$ is the minimum functional C/N which depends, e.g., on the used modulation and CR. The receiver noise input power P_n (dBW) can be presented further by:

$$P_n = F + 10\ \log(kTB), \tag{21.7}$$

where F (dB) is the receiver's noise figure (component dependent) and $P_t = 10 \cdot \log(kTB)$ is the thermal noise level, k is Boltzmann's constant $1.38 \cdot 10^{-23}$ J/K, T is the temperature in Kelvin (290 K is normally used as an average value) and B is the receiver's noise bandwidth (Hz). As a comparison, in DVB-H, the value for B is 7.6 MHz in case of the 8 MHz variant. As an example, for the 8MHz band, the thermal noise floor is –105.2 dBm. Combined with the terminal's noise figure of 5 dB (this depends on the quality of the receiver's components, and it has frequency dependency), the P_n would be –100.2 dBm.

There is still one important item that should be taken into account when estimating the final maximum allowed path loss. This is the loss that the transmitter filter absorbs from the radiating power. In the general calculations, it can be estimated as 10% of the radiating power level (W). The effect can be taken into account when the transmitter power level is presented in dBm:

$$P_{TX}^{effective}\,[dBm] = 10 \cdot \log\left(\frac{(P_{TX}\,[W] \cdot 0.9)}{1 \cdot 10^{-3}}\right). \tag{21.8}$$

The complete formula for the maximum path loss is thus:

$$\begin{aligned} L(dB) = {} & 10 \cdot \log\left(\frac{(P_{TX}\,[W] \cdot 0.9)}{1 \cdot 10^{-3}}\right)[dBm] - L_{cc}\,[dB] - L_{ps}\,[dB] + G_{TX}\,[dBi] - NF\,[dBm] \\ & - (30 + 10\,\log(kTB)\,[dBm]) - (C/N)_{\min}\,[dB] + (0.013f\,[MHz] - 16.172)\,[dB] - L_{lv}\,[dB] \\ & + G_{SFN}\,[dB] - L_{GSM}\,[dB] \end{aligned} \tag{21.9}$$

21.7.3.1 Propagation Models

The most important radio related task in the nominal as well as in the detailed network planning is to estimate the radio coverage area with the given parameters. There are various models available that are based on the radio propagation theories and experiments. There are also interpolation methods available. The outcome is typically a method that can be applied for the mathematical calculation of the estimated site cell radius. Some of the widely used experimental models in the mobile communications are based on the Okumura-Hata, Cost 231-Hata and ITU-R path loss predictions. This type of models divides the formulation into separate area types, for example, presenting urban, suburban and open areas. Cost 231-Walfisch-Ikegami based model is a slightly different as it tends to quantify the propagation environment.

Typically after the initial presentation of the models, there have been various validation rounds that have confirmed the functionality, or have adjusted the models closer to the reality. As an example, the original Cost 231-Walfisch-Ikegami had a minor error in the initial presentation which was found later. Furthermore, the functionality of the model has been investigated, and the conclusion is that the results of the model are relatively close to the ones obtained from the Okumura-Hata based models especially when the building group height is close to the value of half of the street width.

The original Okumura-Hata path loss prediction model is useful in the approximate coverage estimation of macro cells in many cases especially in the nominal radio network planning phase. As an example, the estimated path loss L (dB) in the large city type can be obtained by:

$$\begin{aligned} L(dB) = {} & 69.55 + 26.16\,\lg(f) - 13.82\,\lg(h_{BS}) - a(h_{MS})_i \\ & + [44.9 - 6.55\,\lg(h_{BS})]\,\lg(d), \end{aligned} \tag{21.10}$$

where h_{BS} is the height of the base station transmitter antenna (in range of 30–200 m), h_{MS} is the height of the receiver (m), and d is the distance between the transmitting and receiving antennas (km). For the frequency range of 400–1500 MHz, the area type factor for the large city is:

$$a(h_{MS})_{LC} = 3.2\left(\log\left(11.75h_{MS}\right)\right)^2 - 4.97. \tag{21.11}$$

The maximum distance d up to 20 km can now be obtained:

$$d = 10^{\left(\frac{L(dB) - [69.55 + 26.16\,\log(f) - 13.82\,\lg(h_{BS}) - a(h_{MS})_i]}{44.9 - 6.55\,\log(h_{BS})}\right)}. \tag{21.12}$$

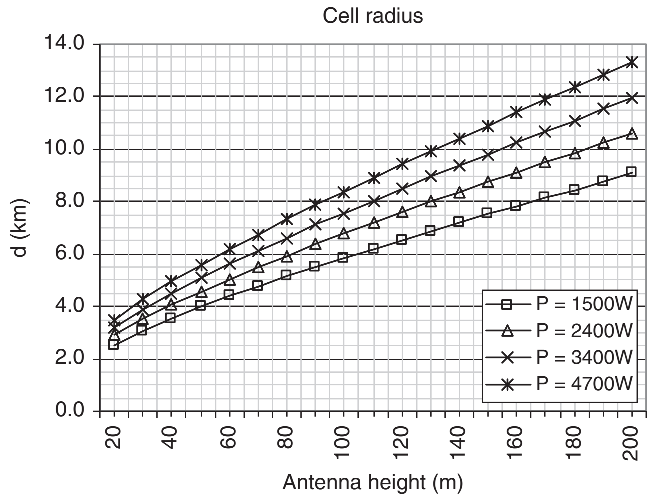

Figure 21.20 *Examples of the high-power site cell radius, when 8 MHz, 16-QAM, CR 1/2 and MPE-FEC 1/2 are applied.*

For the medium-small city type, the correction factor is:

$$a(h_{MS})_{SMC} = (1.1\ \log f - 0.7)\, h_m - (1.56\ \log f - 0.8)\,. \tag{21.13}$$

For the suburban area type, the path loss L of (21.13) is used as a basis with the following:

$$L_{sub-urban} = L - 2\left(\log\left(\frac{f}{28}\right)\right)^2 - 5.4. \tag{21.14}$$

Finally, the loss in the open area can be obtained by applying the following correction:

$$L_{open} = L - 4.78\,(\log f)^2 + 18.33\ \log f - 40.94. \tag{21.15}$$

Figure 21.20 presents the estimated site cell range of the example calculated with the large city correction factor of the Okumura-Hata prediction model and by varying the transmitter antenna height and power levels. As can be noted, the antenna height has a major impact on the site cell radius compared to the transmitter power level.

Another suitable model for practically all mobile communications environments is ITU-R P.1546. The model is based on predefined curves for the frequency range of 30 MHz to 3000 MHz and for maximum antenna heights of 3000 m from the surrounding ground level. The model is valid for terminal distances of 1 to 1000 km from the base station over the terrestrial and sea levels, or for the combination of these.

If the investigated frequency or antenna height does not coincide with the predefined curves, the correct values can be obtained by interpolating or extrapolating the predefined values according to the annexes of ITU-R P.1546 model and by applying the calculation principles presented in its Annex 5. The case curves represent field strength values for 1 kW effective radiated power level (ERP), and the curves have been produced for the frequencies of 100 MHz, 600 MHz and 2 GHz. The curves are based on the empirical studies about the propagation. In addition to the graphical curve format, the values can be obtained also in a tabulated numerical format.

The ITU-R P.1546 method has been evaluated in different sources. Reference [4] has concluded that in the rural area of Australia, P.1546-0 and P.1546-1 provide better overall prediction of the path loss compared to traditional models like Okumura-Hata. The comparison also shows that P.1546-2 on average underestimates the field strength by more than 10 dB in that area type. Nevertheless, it was shown that P.1546-2 improves the standard deviation of the prediction error compared to previous versions of the ITU-R P.1546. This result correlates with [5] which has concluded that the accuracy of theITU- R P.1546 is consistent with the Okumura-Hata model up to about 20 km for urban areas. For rural areas, the predicted field values of the ITU-R P-1546 model differ from the reference solution more than those of the older ITU models do. At the moment, the latest version of the model is ITU-R P.1546-3 [6].

21.7.4 Frequency Planning

Frequency reuse factor is one in LTE. This means that no frequency planning is needed as such for the LTE radio network. As LTE is basically a single frequency network (SFN), the frequency planning is related to the dimensioning of the interfering level of the frequencies that arrives outside of the SFN.

References

1. Holma, H. and Toskala, A. (2011). *LTE for UMTS. Evolution to LTE-Advanced*, 2nd edn, John Wiley & Sons, Ltd, Chichester.
2. Penttinen, J.T.J. (2011) *The LTE/SAE Deployment Handbook*, John Wiley & Sons, Ltd, Chichester.
3. Penttinen, J.T.J. (2009) *The DVB-H Handbook*, John Wiley & Sons, Ltd, Chichester.
4. Östlin, E., Suzuki, H. and Zepernick, H.-J. (2006) Evaluation of the Propagation Model Recommendation ITU-R P-1546 for Mobile Services in Rural Australia. 63rd IEEE Vehicular Technology Conference, Melbourne, Australia, May 2006.
5. Tunc, C.A., Altintas, A. and Ertürk, V.B. (2005) Examination of existent propagation models over large inhomogeneous terrain profiles using fast integral equation solution. *IEEE Transactions on Antennas and Propagation*, **53**(9), 3080–3.
6. Recommendation ITU-R P.1546-3. Method for point-to-area predictions for terrestrial services in the frequency range 30 MHz to 3000 MHz. 2007. 57 pp.

22
Planning of Mobile TV Networks

Jyrki T. J. Penttinen

22.1 Introduction

This chapter presents Mobile TV network planning aspects by generalizing the DVB-H principles, based mainly on source [1]. For more detailed studies, references [2]–[79] present various aspects of mobile TV technologies, standards, theoretical simulations, practical field measurements and performance analysis.

22.2 High-Level Network Dimensioning Process

The dimensioning of the high level DVB-H radio network can be summarized by presenting three main variables, that is, the channel capacity, the coverage area and the quality of the service, which altogether have an effect on the total cost of the network as illustrated in Figure 22.1. The network with poor capacity, coverage and QoS (Quality of Service) has a minimum cost, but the revenue per customer would also be low due to the unsatisfactory service. On the other hand, the highly overlapping and high-capacity network with excellent outdoor and indoor coverage in a large area is technically desirable, but the cost for the building and operating of the network might be too high to recover the expenses as there is a practical limit for user fees.

The task is thus to design a network with sufficiently high quality, and with initial and operating costs that can be recovered in a planned time period, for example, via monthly fees. It can be estimated that the quality of the network is on correct level when the customers are willing to use the service and accept the technical performance of the services as well as the related usage fee. In the complete network design, it is thus essential to take into account both technical issues as well as their costs, and to seek for their balance in order to make sure that the return of investments (ROI) are at an acceptable level. As an example of the cross-relation between the values, by keeping the site number the same, 16-QAM modulation would offer high-capacity with lower coverage, whilst QPSK offers less capacity but in larger area with the same location probability for the coverage.

The Telecommunications Handbook: Engineering Guidelines for Fixed, Mobile and Satellite Systems, First Edition.
Edited by Jyrki T. J. Penttinen.

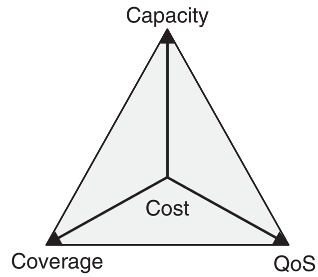

Figure 22.1 *The high-level cross-relations of the most relevant DVB-H radio network planning items.*

In this initial planning phase, there is not yet need for a detailed analysis that take into account the realistic site locations and topological variations, different antenna heights and power levels. Nevertheless, the high-level regulatory limitations for the maximum power should be known. Figure 22.2 summarizes the initial planning steps as presented in Ref. [1].

22.2.1 Capacity Planning

In the initial phase of the DVB-H network planning, the procedure is to decide first the offered capacity of the system. The total capacity in certain DVB-H bandwidth – defined as 5, 6, 7, or 8 MHz – also affects the size of the site coverage area. The dimensioning process is thus iterative, and the aim is to find a balance between the capacity, coverage and cost of the network. This same principle can be generalized over all the mobile TV systems with a certain set of available bandwidths.

The capacity can be varied by adjusting the modulation, guard interval, code rate and channel bandwidth. As an example, the parameter set of QPSK, GI 1/4, code rate 1/2 and channel bandwidth 8 MHz provides a total capacity of 4.98 Mb/s, which can be divided into one or more electronic service guides (ESG) and various audio/video subchannels with about 200–500 kb/s bit stream dedicated for each.

The capacity does not depend on the number of carriers which is indicated by the FFT mode. Nevertheless, the selected FFT affects on the Doppler shift tolerance. As a comparison, the parameter set of 16-QAM, GI 1/32, code rate 7/8 and channel bandwidth of 8 MHz, provides a total capacity of 21.1 Mb/s. It should be noted, though, that the latter parameter set is not practical due to the clearly increased *C/N* requirement which reduces considerably the useful coverage area and makes the reception sensible for the variations in the radio interface.

Table 22.1 shows the reachable capacity in Mb/s per a total DVB-H frequency band as a function of the radio parameter values. This useful bit rate would be reduced accordingly by the direct proportion of MPE-FEC rate when it is present, that is, MPE-FEC rate of 3/4 (which corresponds about 25% of overhead) results in 3/4 out of the original capacity value to be available for the useful bits. In this sense, for example, the combination of CR of 1/2 and MPE-FEC of 1/2 results in the highest protection over the radio transmission but with the lowest capacity.

The first task of the initial capacity planning is thus to select the whole parameter value set that complies with the target capacity value. As an example, if the target capacity value is a minimum of 8 Mb/s, and the bandwidth is 8 MHz, Table 22.1 indicates the compliant parameter values for QPSK that are {GI=1/4, CR=5/6}, {GI=1/4, CR=7/8}, {GI=1/8, CR=3/4}, {GI=1/8, CR=5/6}, {GI=1/8, CR=7/8}, {GI=1/16, CR=3/4}, {GI=1/16, CR=5/6}, {GI=1/16, CR=7/8}, {GI=1/32, CR=2/3}, {GI=1/32, CR=3/4}, {GI=1/32, CR=5/6} and {GI=1/32, CR=7/8}. For the 16-QAM and 64-QAM modulations, all the GI and CR combinations complies with the minimum requirement of 8 Mb/s in this case.

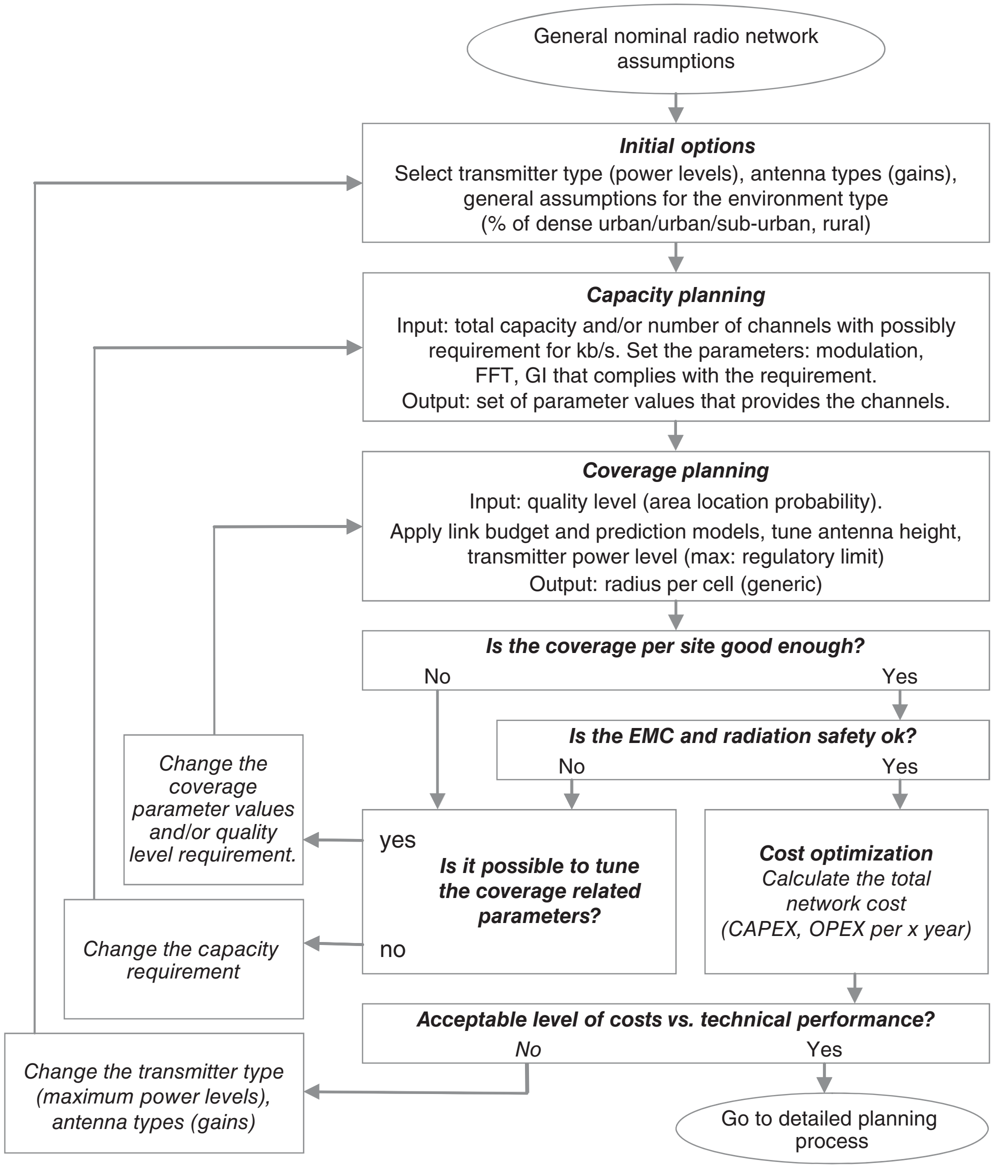

Figure 22.2 *The radio network planning process in the initial phase as presented in Ref. [1]. The process contains high-level estimations of the capacity and coverage by taking into account the economical and regulatory limitations. Figure by Jyrki Penttinen.*

Knowing that the MPE-FEC rate will reduce the final useful data with the direct proportion, the set of the compliant parameters is reduced in the following way when the MPE-FEC is added:

- For MPE-FEC 7/8: {QPSK, GI=1/8, (CR=5/6, 7/8)}, {QPSK, GI=1/16, (CR=5/6, 7/8)}, {QPSK, GI=1/32, (CR=5/6, 7/8)}, {16-QAM, GI=all, CR=all}, {64-QAM, GI=all, CR=all}

Table 22.1 *The summary of total DVB-H bit rates as a function of the parameter values as presented in Ref. [2]. The values for the 5 MHz band can be extrapolated from the others*

		GI=1/4			GI=1/8			GI=1/16			GI=1/32		
Modul.	CR	6 MHz	7 MHz	8 MHz	6 MHz	7 MHz	8 MHz	6 MHz	7 MHz	8 MHz	6 MHz	7 MHz	8 MHz
QPSK	$\frac{1}{2}$	3.73	4.35	4.98	4.14	4.83	5.53	4.39	5.12	5.85	4.52	5.27	6.03
	2/3	4.97	5.80	6.64	5.52	6.45	7.37	5.85	6.83	7.81	6.03	7.03	8.04
	3/4	5.59	6.53	7.46	6.22	7.25	8.29	6.58	7.68	8.78	6.78	7.91	9.05
	5/6	6.22	7.25	8.29	6.91	8.06	9.22	7.31	8.53	9.76	7.54	8.79	10.05
	7/8	6.53	7.62	8.71	7.25	8.46	9.68	7.68	8.96	10.25	7.91	9.23	10.56
16-QAM	1/2	7.46	8.70	9.95	8.29	9.67	11.06	8.78	10.24	11.71	9.04	10.55	12.06
	2/3	9.95	11.61	13.27	11.05	12.90	14.75	11.70	13.66	15.61	12.06	14.07	16.09
	3/4	11.19	13.06	14.93	12.44	14.51	16.59	13.17	15.36	17.56	13.57	15.83	18.10
	5/6	12.44	14.51	16.59	13.82	16.12	18.43	14.63	17.07	19.52	15.08	17.59	20.11
	7/8	13.06	15.24	17.42	14.51	16.93	19.35	15.36	17.93	20.49	15.83	18.47	21.11
64-QAM	1/2	11.19	13.06	14.93	12.44	14.51	16.59	13.17	15.36	17.56	13.57	15.83	18.10
	2/3	14.92	17.41	19.91	16.58	19.35	22.12	17.56	20.49	23.42	18.09	21.11	24.13
	3/4	16.79	19.59	22.39	18.66	21.77	24.88	19.76	23.05	26.35	20.35	23.75	27.14
	5/6	18.66	21.77	24.88	20.73	24.19	27.65	21.95	25.61	29.27	22.62	26.39	30.16
	7/8	19.59	22.86	26.13	21.77	25.40	29.03	23.05	26.89	30.74	23.75	27.71	31.67

Source: Data by courtesy of DVB-H TM.

- For MPE-FEC 5/6: {QPSK, GI=1/8, CR=7/8}, {QPSK, GI=1/16, (CR=5/6, 7/8)}, {QPSK, GI=1/32, (CR=5/6, 7/8)}, {16-QAM, GI=all, CR=all}, {64-QAM, GI=all, CR=all}
- For MPE-FEC 3/4: {16-QAM, GI=1/4, (CR=2/3, 3/4, 5/6, 7/8)}, {16-QAM, (GI=1/8, 1/16, 1/32), CR=all}, {64-QAM, GI=all, CR=all}
- For MPE-FEC 2/3: {16-QAM, GI=1/4, (CR=2/3, 3/4, 5/6, 7/8)}, {16-QAM, GI=1/8, (CR=2/3, 3/4, 5/6, 7/8)},{16-QAM, GI=1/16, (CR=2/3, 3/4, 5/6, 7/8)}, {16-QAM, GI=1/32, CR=all}, {64-QAM, GI=all, CR=all}
- For MPE-FEC 1/2: {16-QAM, GI=1/4, (CR=5/6, 7/8)}, {16-QAM, GI=1/8, (CR=3/4, 5/6, 7/8)}, {16-QAM, GI=1/16, (CR=3/4, 5/6, 7/8)}, {16-QAM, GI=1/32, (CR=2/3, 3/4, 5/6, 7/8)}, {64-QAM, GI=1/4, (CR=2/3, 3/4, 5/6, 7/8)},{64-QAM, (GI=1/8, 1/16, 1/32), CR=all}.

As increased capacity reduces respectively the radio coverage, the task is to find a parameter combination that complies with the original capacity requirement with a possible margin that should be decided beforehand. The division for the margin categories can be decided in such a way that the excess of the capacity of, for example, 0…10% is still acceptable and complies with the target capacity dimensioning. If the excess is, for example, 10…25%, it can be called as slightly overdimensioned capacity, 25…50% can be categorized as clearly overdimensioned, and more than 50% can be considered as heavily overdimensioned.

In this example, the parameter values that comply with the target value of 8.0…8.8 Mb/s are the following:

- For MPE-FEC off: {QPSK, GI=1/4, CR=5/6}, {QPSK, GI=1/4, CR=7/8}, {QPSK, GI=1/8, CR=3/4}, {QPSK, GI=1/16, CR=3/4}, {QPSK, GI=1/32, CR=2/3}
- For MPE-FEC 7/8: {QPSK, GI=1/8, CR=5/6}, {QPSK, GI=1/8, CR=7/8}, {QPSK, GI=1/16, CR=5/6}, {QPSK, GI=1/32, CR=5/6}, {16-QAM, GI=1/4, CR=1/2}
- For MPE-FEC 5/6: {QPSK, GI=1/8, CR=7/8}, {QPSK, GI=1/16, (CR=5/6, 7/8)}, {QPSK, GI=1/32, (CR=5/6, 7/8)}
- For MPE-FEC 3/4: {16-QAM, GI=1/8, CR=1/2}, {QPSK, GI=1/16, CR=1/2}

- For MPE-FEC 2/3: {16-QAM, GI=1/32, CR=1/2}
- For MPE-FEC 1/2: {16-QAM, GI=1/4, (CR=5/6, 7/8)}, {16-QAM, GI=1/8, CR=3/4}, {16-QAM, GI=1/16, CR=3/4}, {16-QAM, GI=1/32, CR=2/3}, {64-QAM, GI=1/8, CR=1/2}, {64-QAM, GI=1/16, CR=1/2}.

When the parameter value candidate short list is selected for the coverage planning, some high-level rules should already be known about the effects of the MPE-FEC, modulation, GI and CR for the final selection of the combination of the parameters. As an example, the optimal performance of MPE-FEC depends on the environment. It functions best when the field strength is low enough, and impulse noise is present. Reference [1] shows that MPE-FEC does have a clear benefit in the extending of the coverage area in the vehicular outdoor channel as the MPE-FEC functionality can provide the same reception quality with a several dB's lower field strength compared to the sole FEC performance.

Nevertheless, when operating within the functional Doppler limits, the importance of MPE-FEC lowers in the low speed pedestrian channel in the city areas where the outdoor field strength is good and the received power levels where MPE-FEC would be most useful are actually rarely present compared to the single site cell. In the indoor pedestrian channel, the negative effect of the occasionally occurring impulse noise lowers even with the lowest MPE-FEC rates. Also, the frame error rate in the low field of the buildings can be enhanced, although it occurs typically only within small areas as the indoor field strength lowers relatively fast in the cell edge region due to strong diffraction attenuation of walls.

The strongest MPE-FEC rates are not recommendable to apply in areas where sufficiently high field strength is found as this would waste capacity but not offering clear performance gains. It has been noted in [35] that the combination of high code rate and low MPE-FEC rate gives better balance between the capacity and coverage compared to the low code rate and high MPE-FEC rate. On the other hand, it is not recommendable to switch off the MPE-FEC as it reduces occasionally appearing impulse noise and helps to extend the useful coverage when the terminal is found in the edge area of the site cell.

QPSK is the most robust of the available DVB-H modulations. It provides largest coverage areas, but with lowest capacity. As Ref. [3] shows, the bit error rate of $1 \cdot 10^{-4}$ for QPSK (4-QAM) requires about 8.2 dB E_b/N_o, whereas the requirement for the 16-QAM is about 12.1 dB and for the 64-QAM about 16.4 dB. This indicates that 16-QAM increases the path loss approximately 4 dB compared to QPSK when applied as such on the radio link budget. On the other hand, 16-QAM provides a double capacity compared to QPSK. 64-QAM further decreases the coverage approximately with an additional 4 dB in theory. 64-QAM has been noted to be very sensitive to the varying radio channel conditions and is thus not recommendable as the primary choice of modulation in large areas.

The QEF point of BER, that is, $2 \cdot 10^{-4}$, is obtained typically with about 7–8 dB stronger C/N values for 16-QAM compared to QPSK. The QEF point in case of the 64-QAM seem to require typically more than 20 dB compared to QPSK, which indicates strong practical challenges for 64-QAM in a mixed radio channel type although the results in this specific case were obtained by collecting the data with a prototype terminal. It should be noted that even if the practical 64-QAM performance might require higher C/N than indicated in theory especially in vehicular channels, there are isolated pedestrian locations where 64-QAM could be used efficiently, for example, in airports and shopping centers, with a clearly separated SFN or MFN.

If the SFN mode is used, smallest GI values provide also smallest functional areas. It means that when using only one or two frequencies and there is a need to cover large areas, GI-values of 1/32 and 1/16 are not recommendable. The largest SFN area can be obtained by using the GI value of 1/4. On the other hand, the small GI values provide more Doppler tolerance which is beneficial in the fast vehicular channel type.

According to the above information, in this specific case, it would be logical to select the following settings as primary generic option for the first iteration of the capacity planning phase: {16-QAM, GI=1/4, CR=1/2, MPE-FEC 7/8}, {16-QAM, GI=1/8, CR=1/2, MPE-FEC 3/4} or {QPSK, GI=1/16, CR=1/2, MPE-FEC 3/4}.

The outcome of this first step of the investigation is a compliant set of DVB-H parameter values for the capacity requirement taking into account the acceptable excess of the capacity. If the forthcoming coverage analysis does not produce a desired plan, the initial capacity target should be changed and the above described process needs to be repeated. If the parameter value set is not considered feasible, the values should be revised until the wanted capacity can be achieved with the required coverage and quality level of the network.

22.2.2 Coverage and QoS Planning

When the capacity requirement and the respective radio parameter value set is known, the next step of the DVB-H network dimensioning is to estimate the coverage of the site cells. The major items for the first hand coverage estimation are: coordinates of the transmitter, radiated power, frequency and antenna pattern [4]. In addition to the antenna height, radiating power and radio path loss in different propagation types, the coverage area size depends on the required quality level of the reception.

The estimation of the radio channel type is important in this phase as it includes the fading profile and has thus effect on the radio link budget and Doppler tolerance limits. In order to get the first estimation of the number of the transmitters, the nominal plan can be carried out by assuming an ideal distribution of sites and the most probable channel type. The practical radio network is always nonideal as for the site locations, so the final plan must be adjusted accordingly, by using nonuniform power levels and antenna heights. The coverage holes, for example, in street canyons and indoors, can be further enhanced by using separate DVB-H gap-fillers.

As there is time and location dependent fluctuation in the received power, the dimensioning is done by estimating the probability for the reception of the sufficiently high-level signal, that is, the task is to design the wanted quality target of the coverage. The margin is presented by the location variation parameter in the link budget. The margin is estimated for the coverage area over the whole site cell.

22.2.2.1 Radio Link Budget

When the coverage criteria are known, the site cell radius can be estimated by applying the radio link budget calculation. As DVB-H is a broadcast system, the radio link budget is calculated only for the downlink direction. For the possible interaction channel, the respective downlink and uplink path losses can be estimated by applying a separate radio link budget of the used system for the interactions (e.g. GSM/GPRS or UMTS). In the normal planning case, though, it can be assumed that the coverage area of the interaction channel is present ideally where also DVB-H is found.

The generic principle of the DVB-H link budget can be seen in Table 22.2. The calculation shows an example of the transmitter output power level of 2400 W, with the quality value of 90% for the area location probability. There are four different cases shown in the table as a function of the modulation and MPE-FEC rate. The SFN gain has assumed as 0 dB in these cases. According to the link budget, the outdoor reception of this specific case results in a successful reception for {QPSK, CR 1/2, MPE-FEC 2/3} when the radio path loss is equal or less than 144.2 dB.

The interpretation of the quality of the coverage area depends on the agreed area location probability level. In general, the location variation is considered to follow a log-normal distribution [5], meaning that the logarithm of the signal level follows a normal or Gaussian distribution. The statistical distribution should be applied in the respective quality level estimations. The mean value means that 50% of the samples are above this value and the other half below. In case of any other percentage for the coverage quality criterion, the relationship between the mean value and standard deviation should be known. The standard deviation of 5.5 dB is normally used in the typical suburban type of DVB-H. The standard deviation is commonly used

Table 22.2 *An example of the DVB-H link budget*

		Case:	1	2	3	4
DVB-H Link Budget		Modulation: CR: MPE-FEC:	QPSK 1/2 1/2	QPSK 1/2 2/3	16-QAM 1/2 1/2	16-QAM 1/2 2/3
General parameters	*Variable*	*Unit*				
Frequency	f	MHz	600.0	600.0	600.0	600.0
Noise floor for 8 MHz bandwidth	P_n	dBm	−105.2	−105.2	−105.2	−105.2
RX noise figure	F	dB	5.0	5.0	5.0	5.0
TX						
Transmitter output power	P_{TX}	W	2400.0	2400.0	2400.0	2400.0
Transmitter output power	P_{TX}	dBm	63.8	63.8	63.8	63.8
Cable and connector loss	L_{cc}	dB	3.0	3.0	3.0	3.0
Power splitter loss	L_{ps}	dB	3.0	3.0	3.0	3.0
Antenna gain	G_{TX}	dBi	13.1	13.1	13.1	13.1
Antenna gain	G_{TX}	dBd	11.0	11.0	11.0	11.0
Eff. Isotropic radiating power	$EIRP$	dBm	70.9	70.9	70.9	70.9
	$EIRP$	W	12309	12309	12309	12309
Eff. Radiating power	ERP	dBm	68.8	68.8	68.8	68.8
	ERP	W	7502.6	7502.6	7502.6	7502.6
RX						
Min C/N for the used mode	$(C/N)_{min}$	dB	8.5	11.5	14.5	17.5
Sensitivity	P_{RXmin}	dBm	−91.7	−88.7	−85.7	−82.7
Antenna gain, isotropic ref	G_{RX}	dBi	−8.4	−8.4	−8.4	−8.4
Antenna gain, 1/2 wavelength dipole	G_{RX}	dBd	−6.2	−6.2	−6.2	−6.2
Isotropic power	P_i	dBm	−83.3	−80.3	−77.3	−74.3
Location variation for 90% area prob	L_{lv}	dB	7.0	7.0	7.0	7.0
SFN gain	G_{SFN}	dB	0.0	0.0	0.0	0.0
MPE-FEC gain	$G_{MPE\text{-}FEC}$	dB	0.0	0.0	0.0	0.0
Building loss	L_b	dB	14.0	14.0	14.0	14.0
GSM filter loss	L_{GSM}	dB	0.0	0.0	0.0	0.0
Min required received power outdoors	$P_{min(out)}$	dBm	−76.3	−73.3	−70.3	−67.3
Min required received power indoors	$P_{min(in)}$	dBm	−62.3	−59.3	−56.3	−53.3
Min required field strength outdoors	$E_{min(out)}$	dBuV/m	56.4	59.4	62.4	65.4
Min required field strength indoors	$E_{min(in)}$	dBuV/m	70.4	73.4	76.4	79.4
Maximum path loss, outdoors	$\boldsymbol{L_{pl(out)}}$	**dB**	**147**	**144**	**141**	**138**
Maximum path loss, indoors	$\boldsymbol{L_{pl(in)}}$	**dB**	**133**	**130**	**127**	**124**

as a basis for the mobile communications coverage predictions, informing about the confidence in statistical conclusions.

The relationship between the area location probability and the additional margin that should be taken into account in the DVB-H link budget (as presented in Table 22.3) can be thus derived from the characteristics of the normal and log-normal distribution, for example, by observing the attenuation points (dB) in cumulative

Table 22.3 The area location probability in the site cell edge and over the whole site cell area for the mobile reception when the standard deviation is 5.5 dB, according to Ref. [2]

Area location prob. (minimum coverage target)	Loc probability in site cell edge	Location correction factor	Subjective quality description
90%	70%	7 dB	Fair outdoor
95%	90%	9 dB	Good outdoor, fair indoor
99%	95%	13 dB	Excellent outdoor, good indoor

Source: Data by courtesy of DVB-H TM.

scale that fulfils the required percentage of the area location in the whole area. Furthermore, it can be decided that 90% of area location probability indicates a fair outdoor coverage, whilst 95% is considered as good and 99% provides an excellent quality. The mapping of the typical values that can be used in the DVB-H planning for mobile reception, when the standard deviation is 5.5 dB, can be seen in Ref. [2]. In addition to the standard deviation, the criteria vary depending on the environment, that is, on the propagation slope. Slope of 2 (i.e. 20 dB/decade) represents line of sight in free space.

Reference [6] presents the correction factors as a function of the reception environment: pedestrian 90% location area probability results in 7.1 dB, pedestrian 95% location 9.0 dB, indoor 90% location 10.4 dB, indoor 95% location 13.3 dB, mobile 90% location 9.0 dB and mobile 99% location 12.8%.

In Ref. [7], the area types have been further divided into different classes, that is, outdoor pedestrian (A), light indoor (B1), deep indoor (B2), mobile rooftop (C) and mobile in-car (D). Reference [7] proposes that for the class A and B, a good coverage quality corresponds to 95%, and acceptable to 70% area location probability whilst the class C and D corresponds to values of 99 and 90%, respectively. It is thus important to clarify the level of the quality in such a way that no misinterpretations may occur in the requirement levels.

It should be noted that in DVB-H, the "cell" coverage refers to the coverage area of one or more sites belonging to a certain SFN. In this work, the term "site cell" refers to the coverage area of a single antenna system of one transmitter. The antenna of the site cell can be omniradiating or a set of directional antennas. The definition of the "cell" can be found in Ref. [2] which describes that in the DVB-H system, the *cell_frequency_link_descriptor* indicates the frequencies that are used for the different cells of the network. The frequencies (and thus cells) are furthermore mapped with Transport Streams. The *cell_list_descriptor* contains the needed information of the coverage area of the cells. Physically, cell is defined as a geographical area covered by the signals that contain one or more transport streams, which can be done with one or more transmitters. In the simulations of this chapter, the hexagonal model with omni-radiating antennas per site cell is used for the coverage estimate.

22.2.2.2 Path Loss

The maximum path loss L (dB) is the difference between the effective isotropic radiating transmitter power P_{EIRP} and the required received power in outdoors $P_{\min(out)}$:

$$L = P_{EIRP} - P_{\min(out)}. \tag{22.1}$$

In this formula, P_{EIRP} (dBm) is:

$$P_{EIRP} = P_{TX} - L_{cc} - L_{ps} + G_{TX}. \tag{22.2}$$

The minimum received power level $P_{\min(out)}$ (dBm) is:

$$P_{\min(out)} = P_i + L_{lv} - G_{SFN} - G_{MPE-FEC} + L_{GSM}, \tag{22.3}$$

where P_i is the isotropic received power, L_{lv} is the location variation for a certain area probability, G_{SFN} is the SFN gain, G_{MPE_FEC} is the MPE-FEC gain, and L_{GSM} is the GSM filter loss due to the isolation of the DVB-H receiver and GSM transmitter.

The isotropic received power is obtained from:

$$P_i = P_{RX\,\min} - G_{RX}, \tag{22.4}$$

where $P_{RX\min}$ is the receiver sensitivity, or the minimum power level the receiver requires, and G_{RX} is the receiver's antenna gain. The latter depends on the frequency. Based on the information of Ref. [2] a linear interpolation can be applied for $\{474\text{ MHz} < f < 858\text{ MHz}\}$ for the antenna gain (which is in fact loss):

$$G_{RX}\,[dBi] = \frac{5f}{384} - \frac{5 \cdot 474}{384} - 10 = 0.013f - 16.172. \tag{22.5}$$

The minimum required receiver's power level $P_{RX\min}$ is obtained by:

$$P_{RX\,\min} = P_n + (C/N)_{\min}, \tag{22.6}$$

where P_n is the receiver noise input power level, and $(C/N)_{\min}$ is the minimum functional C/N which depends on the used modulation, CR and MPE-FEC rate.

The receiver noise input power P_n (dBW) can be presented further by:

$$P_n = F + 10\,\log(kTB), \tag{22.7}$$

where F (dB) is the receiver's noise figure (component dependent) and $P_t = 10{\cdot}\log(kTB)$ is the thermal noise level, k is Boltzmann's constant $1.38{\cdot}10^{-23}$ J/K, T is the temperature in Kelvin (290 K is normally used as an average value) and B is the receiver's noise bandwidth (Hz). In DVB-H, the value for B is 7.6 MHz in case of the 8 MHz variant. As an example, for the 8 MHz band, the thermal noise floor is –105.2 dBm. Combined with the terminal's noise figure of 5 dB (this depends on the quality of the receiver's components, and it has frequency dependency), the P_n would be –100.2 dBm.

There is still one important item that should be taken into account when estimating the final maximum allowed path loss. This is the loss that the transmitter filter absorbs from the radiating power. In the general calculations, it can be estimated as 10% of the radiating power level (W). As an example, Ref. [1] utilized the 10% assumption for the transmitter filter loss. The effect can be taken into account when the transmitter power level is presented in dBm:

$$P_{TX}^{effective}\,[dBm] = 10 \cdot \log\left(\frac{(P_{TX}\,[W] \cdot 0.9)}{1 \cdot 10^{-3}}\right). \tag{22.8}$$

The complete formula for the maximum path loss is thus:

$$\begin{aligned} L(dB) = 10 \cdot \log\left(\frac{(P_{TX}\,[W] \cdot 0.9)}{1 \cdot 10^{-3}}\right)[dBm] - L_{cc}\,[dB] - L_{ps}\,[dB] + G_{TX}\,[dBi] - NF\,[dBm] \\ - (30 + 10\,\log(kTB)\,[dBm]) - (C/N)_{\min}\,[dB] + (0.013f\,[MHz] - 16.172)\,[dB] - L_{lv}\,[dB] \\ + G_{SFN}\,[dB] + G_{MPE-FEC}\,[dB] - L_{GSM}\,[dB]\,. \end{aligned} \tag{22.9}$$

22.2.2.3 Building Loss

The building loss, or building penetration loss, can be estimated in a general level as an average value in different environment types, and it can thus be considered as a fixed radio link budget value for different area types. The DVB-H implementation guideline recommends a median value of 11 dB and a standard deviation

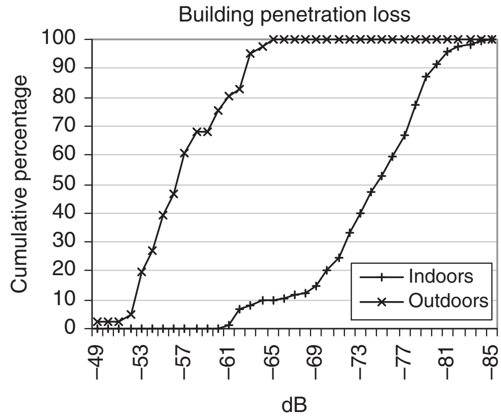

Figure 22.3 *An example of the typical building loss in a format of a cumulative RSSI histogram, which is obtained by the difference of the received power in the indoor and outdoor area. This case represents the deep building type of B2, investigated via field measurements of Ref. [1].*

value of 6 dB to be used for the building loss [2]. In addition to the building penetration loss, that is, the ratio of the average powers measured outside and inside the building with a fixed transmitter, also the building floor loss may be important to take into account in the detailed network planning. Reference [8] has concluded that the floor loss can be in certain cases approximately between 30 and 40 dB.

In Ref. [7], it is stated that the building loss is frequency dependent. This is logical as the radio wave penetrates into the buildings depending on the conditions. As an example, a typical urban building normally contains metallic supports that might create a Faraday cage. Depending on the wavelength of the signal and the hole-size of the supporting metal the respective attenuation varies. Reference [7] includes typical building penetration losses for the general use of the radio link budget. The updated table shows values of 7 dB for class D (in-car) for all the bands of VHF, UHF, L-band and S-band with no standard deviation. For the class B1 (light indoor), the penetration loss is given as 9 dB (with the standard deviation σ_p of 4.5) for VHF, 11 dB ($\sigma_p = 5$ dB) for UHF, 13 dB ($\sigma_p = 5$ dB) for the L-band and 14 dB ($\sigma_p = 5$ dB) for the S-band. For the class B2 (heavy indoor), the respective values are: 15 dB ($\sigma_p = 5$ dB), 17 dB ($\sigma_p = 6$ dB), 19 dB ($\sigma_p = 6$ dB) and 19 dB ($\sigma_p = 6$ dB).

The snapshot measurements as presented in Figure 22.3 carried out during the field tests of Ref. [1] correlate with the above mentioned information. There were two building types investigated, first case (A) being an 8-floor hotel with an open center area representing deep indoor (class B2), and the second (B) being a lighter 1-floor construction (class B1). In both cases, the received power level was stored with one-second interval in a slow-moving pedestrian radio channel type by walking outside of the building on the side where the transmitter antenna was installed, and then by repeating the same measurements in the ground floor inside the center of the building. A UHF frequency of 701 MHz was used in these measurements. The investigated buildings were located in a typical suburban area. It should be noted, that the building loss might vary considerably depending on the environment and building material.

The average RSSI values of indoor and outdoor, that is, the building loss, for the first case A, which represents the class B2, is 16.1 dB with the standard deviation of 5.8 dB in indoors and 3.9 dB in outdoors. The respective value set of Ref. [7] is very close to this result, the difference between the measured one being about 1 dB. When comparing the building loss via the 50-percentile values the result would be about 18 dB. For the second case (B), which represents the class B1, the building loss that is calculated via the differences of the average values, is 14.1 dB, with the standard deviation of 6.1 dB in outdoors and 2.9 dB in indoors.

The difference between the Ref. [7] is now about 3 dB. The respective 50-percentile comparison indicates that the building loss is about 13 dB.

As both cases show differences between the average and respective 50-percentile values, it is important to define which the used method is for solving the building loss. In Ref. [7], the average values are used, and the building penetration loss L_P is obtained by comparing the signal level distributions E inside and outside of the building:

$$L_P = E_{out\_average} - E_{in\_average}. \tag{22.10}$$

The examples presented in this chapter show that sufficiently amount of field tests clarifies the typical building loss values that can be used for the local adjustment of the link budget.

22.2.2.4 Effect of the Antenna Height: Receiver

The mobile environment affects on the coverage area of DVB-H differently compared to DVB-T. The planning assumption of the DVB-H radio link budget is outdoors with a 1.5 meter terminal height, typically in N-LOS in the city area. This causes a penalty for the DVB-H link budget compared to the DVB-T that is based on the fixed rooftop antenna with LOS [9]. Reference [6] indicates that the receiver antenna height loss can be 11 dB for rural area, 16 dB in suburban and 22 dB in urban area in Band IV. For the band V, the respective values are 13, 18 and 24 dB.

22.2.2.5 Effect of the Antenna Height: Transmitter

The height of the DVB-H transmitter site antenna has a key role in the coverage area of the DVB-H site cell. The doubling of radiating power level, that is, adding 3 dB to the radio link budget might raise the radius of the site cell by about 30% in the typical DVB-H antenna heights, meaning that the coverage area would enhance around 70%. This could happen, for example, by changing the 2400 W transmitter model to 4700 W model when using the antenna in 60 m height. On the other hand, if the transmitter antenna height is moved from 60 m to 85 m (40% rise to the height) but using the same 2400 W transmitter, the effect of the coverage area would be the same as doubling the transmitter power. This phenomenon should be considered in the cost optimization of DVB-H.

22.2.2.6 SFN Gain

One possible item in the radio link budget is the SFN gain that can enhance the performance in the overlapping areas, or provide the theoretical possibility to construct the sites further away from each others as the coverage area of each site cell rises. The interpretation of the benefit of SFN varies though.

The implementation guidelines [2] states that there is a potential SFN diversity gain but without specifying more concrete values. The SFN gain in general has been noted as useful in Refs. [10, 11]. On the other hand, [7] recommends that the SFN gain would not be taken into account in the radio link budget. With the parameter sets that results in smaller SFN sizes, as the number of sites grows, part of the sites may start to add interference thus reducing the SFN gain. Depending on the radio parameter set (FFT size and GI), the balance can still be achieved by adjusting the transmitter antenna heights and power levels, but some of the parameter values lead to the highly interfered network.

The challenge of using the SFN item in the link budget is that there is no coherent definition available for the gain. It could be interpreted as the difference between the received power levels in dB, comparing a single stream with a varying number of streams as has been presented in Ref. [12]. The value could also be resolved by mapping the $C/(N+I)$ distribution over the whole investigated area.

22.2.2.7 Minimum C/N

The minimum *C/N* ratio that is required for the successful reception of DVB-H video / audio streams depends on the combination of the modulation, code rate and MPE-FEC. Also the radio channel type has a clear effect. The information about the modulation and code rate dependency can be seen in Ref. [7]. The required carrier level values presented in these references have been used in the presented analysis throughout this chapter. It should be noted that as the values have been published in relatively early stage of DVB-H, they might not be the final ones, though, but as for the accuracy of the results presented in this chapter, it can be assumed that the values are sufficiently close to the reality for different channel types.

22.2.2.8 MPE-FEC

The MPE-FEC functionality has been designed to DVB-H in order to provide additional protection for the radio transmission, which enhances the received signal quality. The MPE-FEC rate can be varied between 0 and 50%. According to Ref. [7] MPE-FEC is suitable for the improvement of the *C/N* performance in mobile radio channels. It also gives additional protection against the impulse noise, and enhances the performance of fast-moving terminals by adding the Doppler shift resistance. The MPE-FEC gain depends on the environment. In general, the closer the radio channel is to the AWGN type, the less gain MPE-FEC offers. On the other hand, the MPE-FEC gain is more notable in the Rayleigh type of fast-fading channel as the OFDM can utilize the separate radio components inside of the SFN area providing this additional gain.

Among various other references, [13] indicates a clear advantage in the use of MPE-FEC. The effect might be in order of several dB. In addition to the streaming services, MPE-FEC is also useful for the file-cast mode of DVB-H. According to Ref. [14], the needed time for the file transfer is considerably reduced, and more content can thus be delivered with the same infrastructure by utilizing MPE-FEC. On the other hand, if the transmission time is kept the same, the area coverage for the reliable reception is enlarged. As noted in Ref. [1], MPE-FEC can move the 5% frame error rate point up to 7 dB in RSSI scale in the single site cell case, indicating that in the best case, the additional error correction can enhance considerably the link budget.

22.2.2.9 Other Effects

The seasonal conditions might cause low-level yearly fluctuations in the radio propagation due to the variations of the moisture level of vegetation and weather conditions (e.g. via occasional tunneling effects in the ionosphere). As an example, Ref. [15] has noted that the field measurement results do have certain deviation due to the rain. It can be assumed though that in the typical link budget, the effect of the vegetation and rain is minimal for DVB-H in VHF/UHF bands. In case of the 1.6 GHz version, the effect might be more considerable due to the radio propagation characteristics in higher frequencies. In any case, the seasonal path loss variation can be considered as a minor detail in a practical radio link budget, and due to the challenges in the periodical adjustment of broadcast type of network, it is not necessary to take into account.

As another possible link budget item, the reception antenna diversity could be utilized to exploit the multipath propagation. According to Ref. [7], this feature may not be implemented on all devices, though, so the diversity effect is not needed to be taken into account in the link budget until the penetration of the terminals containing possible receiver diversity or MIMO type of functionality is sufficiently high.

22.2.3 Propagation Models

The most important radio related task in the nominal as well as in the detailed network planning is to estimate the DVB-H coverage area with the given parameters. There are various models available that are based on the radio propagation theories and experiments. There are also interpolation methods presented [16]. The outcome is typically a method that can be applied for the mathematical calculation of the estimated site cell radius. Some of the widely used experimental models in the mobile communications are based on the Okumura-Hata, Cost 231-Hata and ITU-R path loss predictions. This type of models divides the formulation into separate area types, for example, presenting urban, suburban and open areas. Cost 231-Walfisch-Ikegami based model is a slightly different as it tends to quantify the propagation environment.

Typically after the initial presentation of the models, there have been various validation rounds that have confirmed the functionality, or have adjusted the models closer to the reality. As an example, the original Cost 231-Walfisch-Ikegami had a minor error in the initial presentation which was found later. Furthermore, the functionality of the model has been investigated, for example, in Ref. [17], which concluded that the results of the model are relatively close to the ones obtained from the Okumura-Hata based models especially when the building group height is close to the value of half of the street width.

The original Okumura-Hata path loss prediction model [18] is useful in the approximate coverage estimation of DVB-H in many cases especially in the nominal radio network planning phase. As an example, the estimated path loss L (dB) in the large city type can be obtained by:

$$L(dB) = 69.55 + 26.16\,\lg(f) - 13.82\,\lg(h_{BS}) - a(h_{MS})_i + \left[44.9 - 6.55\,\lg(h_{BS})\right]\lg(d), \tag{22.11}$$

where h_{BS} is the height of the DVB-H transmitter antenna (in range of 30–200 m), h_{MS} is the height of the receiver (m), and d is the distance between the transmitting and receiving antennas (km). For the frequency range of 400–1500 MHz, the area type factor for the large city is:

$$a(h_{MS})_{LC} = 3.2\left(\log\left(11.75h_{MS}\right)\right)^2 - 4.97. \tag{22.12}$$

The maximum distance d up to 20 km can now be obtained:

$$d = 10^{\left(\frac{L(dB)-\left[69.55+26.16\,\log(f)-13.82\,\lg(h_{BS})-a(h_{MS})_i\right]}{44.9-6.55\,\log(h_{BS})}\right)}. \tag{22.13}$$

For the medium-small city type, the correction factor is:

$$a(h_{MS})_{SMC} = (1.1\,\log f - 0.7)\,h_m - (1.56\,\log f - 0.8)\,. \tag{22.14}$$

For the suburban area type, the path loss L of (3-11) is used as a basis with the following:

$$L_{sub-urban} = L - 2\left(\log\left(\frac{f}{28}\right)\right)^2 - 5.4. \tag{22.15}$$

Finally, the loss in the open area can be obtained by applying the following correction:

$$L_{open} = L - 4.78\,(\log f)^2 + 18.33\,\log f - 40.94. \tag{22.16}$$

Source [2] presents the estimated site cell range of the example calculated with the large city correction factor of the Okumura-Hata prediction model and by varying the transmitter antenna height and power levels according to Table 22.2 as presented in Figure 22.4. As may be noted, the antenna height has a major impact on the site cell radius compared to the transmitter power level.

Another suitable model for practically all DVB-H environments is ITU-R P.1546 [19]. The model is based on predefined curves for the frequency range of 30 MHz to 3000 MHz and for maximum antenna heights of 3000 m from the surrounding ground level. The model is valid for terminal distances of 1 to 1000 km from the base station over the terrestrial and sea levels, or for the combination of these.

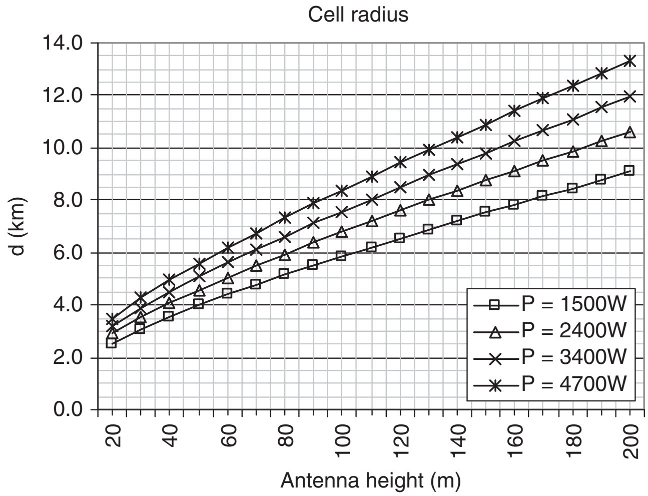

Figure 22.4 *Examples of the DVB-H site cell radius, when 16-QAM, CR 1/2 and MPE-FEC 1/2 are applied. Neither the SFN gain nor MPE-FEC gain are utilized in these calculations.*

If the investigated frequency or antenna height does not coincide with the predefined curves, the correct values can be obtained by interpolating or extrapolating the predefined values according to the annexes of Ref. [19] and by applying the calculation principles presented in its Annex 5. The case curves represent field strength values for 1 kW effective radiated power level (ERP), and the curves have been produced for the frequencies of 100 MHz, 600 MHz and 2 GHz. The curves are based on the empirical studies about the propagation. In addition to the graphical curve format, the values can be obtained also in a tabulated numerical format.

The ITU-R P.1546 method has been evaluated in different sources. Reference [20] has concluded that in the rural area of Australia, P.1546-0 and P.1546-1 provide better overall prediction of the path loss compared to traditional models like Okumura-Hata. The comparison also shows that P.1546-2 on average underestimates the field strength by more than 10 dB in that area type. Nevertheless, it was shown that P.1546-2 improves the standard deviation of the prediction error compared to previous versions of the ITU-R P.1546. This result correlates with Ref. [21] which claims that the accuracy of the ITU-R P.1546 is consistent with the Okumura-Hata model up to about 20 km for urban areas. For rural areas, the predicted field values of the ITU-R P-1546 model differ from the reference solution more than those of the older ITU models do. At the moment, the latest version of the model is ITU-R P.1546-3 [19].

It can be assumed that the basic and extended versions of Okumura-Hata as well as ITU-R P.1546-3 models provide a sufficiently good first-hand estimate for the DVB-H coverage areas and respective capacity and quality levels in the initial network planning phase. These models have been designed for environments with antenna heights and site cell distances that fall into the typical assumptions of DVB-H networks. Reference [6] identifies several other models, including ray-tracing type of estimates for the dense city centers. These models require more detailed digital map data with respective terrain height and cluster attenuations. In the most advanced prediction models, a vector-based 3D map is needed. It logically has a cost effect on the planning but it increases considerably the accuracy of the coverage estimate. It can further be enhanced via local reference measurements by adjusting the model's estimate accordingly. As a cost-efficient compromise, 3D models could be utilized in the advanced phase of the radio network planning in the most important areas.

22.2.4 Safety Distance

A preliminary calculation about the transmitter power levels should be carried out already in the initial radio network planning phase. In this stage, regulatory rules, as well as a rough estimate about EMC and safety zones give a base for estimating the minimum distance between DVB-H antennas and the surrounding population or the antenna systems of other telecommunication systems like GSM and UMTS.

In the initial phase of the radio network planning, it is sufficient to investigate the high level regulatory limits for the nonionizing radiation. A typical maximum allowed value for the DVB-H site might be in order of 50 kW (EIRP), with additional rules to be taken into account, for example, as a function of the antenna height and site type (differentiating the wall-mounted, rooftop and tower-mounted antennas). The international and regional regulation provides sufficient information about the upper limits of the radiation that should be taken into account in the nominal plan. The radiation level might need to be limited further depending on the area type (urban or open), the antenna height, and the frequency. The practical limitations might mean that the highest power class transmitters and high-gain directional antennas can not be used in the implementation and the level should thus be revised case basis.

Detailed safety zone and EMC limit calculations can be carried out when the concrete site locations are known. The allowed antenna distance from the other system antennas or from the installation personnel and the population depends on the type of the installation, that is, the limits vary depending on the rooftop, tower or indoor antenna placement. In a typical broadcast tower case, the main task is to calculate the EMC limitations, that is, the interferences that DVB-H causes to other systems and vice versa, as the antennas are installed sufficiently high in towers by default. In the rooftop and indoor installations, also the safety distance limits for the human exposure should be taken care of with related safety zone marking, for example, for the occasional maintenance visits.

Chapter 24 presents a simple yet practical model for the mobile TV radiation safety distance estimates. When the electrical field is calculated above or below the transmitter antenna, the attenuation factor of the vertical radiation pattern should be taken into account accordingly. The methodology applies in the close distance of the site, and the outcome is to minimize the interference level caused by DVB-H to the other systems nearby, as well as to make sure the human exposure limits are not exceeded. In the back-lobe of the antenna, the method proposes the use of the maximum value (i.e., the minimum possible attenuation value of the radiation pattern) over the whole half-hemisphere. Although the DVB-H frequency usage is regulated, there might also be need to calculate some special cases for the longer distances, like safety zones in the area where sensible space signal reception stations or military bases are present. In these cases, the related safety zone calculation is straightforward and the methodologies of, for example, Ref. [22] can be utilized.

In the site installations, it can be expected that the total exposure increases close to the sites due to the DVB-H. Field measurements presented in Ref. [23] show that in most of the locations around the transmitter site, DVB-H is the dominating source of radiation. The methodology presented in Ref. [1] gives an estimate of the effect, and field tests can be performed in selected locations especially on the rooftop sites in order to fine-tune the calculations.

The outcome of the safety distance calculations in the nominal planning phase helps to reject those radiation levels from the planning assumptions that exceed the regulatory limits and are thus not possible to utilize in the detailed planning, either.

22.2.5 Cost Prediction

In the initial phase of the cost estimation, the strategy is to predict only roughly the capital expense level. This gives an idea about the general amount of costs by varying the most important parameters and by investigating only the main items. The cost effect might be challenging to perform due to the lack of information as stated

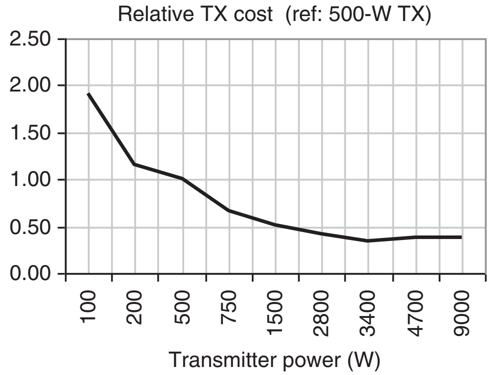

Figure 22.5 *An example of the transmitter cost in terms of a single watt as a function of the total power level.*

in Ref. [24], but especially in the initial phase of the network planning, a high-level estimation can be utilized for the DVB-H specific components if no market data are available. For the rest of the items, for example, for the site construction, antenna, feeders and transmission, typical mobile communications solutions can be used as a basis for the cost estimates.

As an example, the possible DVB-H transmitter list could be limited to models with output power levels of 500–4700 W. The cost of each transmitter includes common parts as well as additional ones. The common parts include the equipment shield and the transmission module. Depending on the model, there is a varying amount of power amplifier units that could fit into the same shield, but higher power levels might require a liquid cooling instead of an air cooling. This means that the price of transmitters does not grow linearly as a function of their power level, but there is a technical as well as marketing related dependency between models and their cost effects on the network planning.

Figure 22.5 shows an example about the transmitter price level behavior, showing the price of a single watt (W) as a function of the total maximum power of the transmitter as studied in Ref. [1]. The price level has been normalized by taking the 500 W-transmitter as a reference.

The next step is to find the other common and variable costs for the individual site setup. The main elements might include the cost of the installation, civil works, and antenna system that includes the antenna elements, cables and other related material.

The study can be done for the initial parameter set, that is, according to the wanted capacity and the estimated average antenna heights. The coverage can be obtained for an individual site based on the suitable radio path loss prediction model. The cost of the network can thus be estimated by multiplying the expenses of a single site and the number of the site cells according to the single site cell radius.

As an example, the initial capacity target could be 5 Mb/s, which can be divided into a program guide (about 300 kb/s) and 10 channels consisting about 450 kb/s each for the combination of audio and video streams. As Table 22.1 indicates, this capacity requirement can be complied with the QPSK modulation, code rate of 1/2 and MPE-FEC rate of 2/3. Assuming that this mode requires C/N of 11.5 dB, and that the environment is urban vehicular with the coverage quality target for the location probability of 70% in the site cell edge and 90% in the site cell area, Table 22.4 can be created by applying the Okumura-Hata prediction model and the hexagonal site cell layout.

After the capacity and coverage analysis, the estimation of the expenses can be done. The cost of the single site can be calculated with the following Formula:

$$C_{site} = C_{common} + C_{\mathrm{vari}able}. \tag{22.17}$$

Table 22.4 *The estimated site cell radius for a set of transmitter antenna heights h_{BS} and transmitter power levels P_{TX} by applying the Okumura-Hata model for the urban area*

	Radius of the site cell (km)			
P_{TX} (W)	h_{BS} = 30 m	h_{BS} = 60 m	h_{BS} = 90 m	h_{BS} = 120 m
500	2.4	3.3	4.0	4.6
750	2.8	3.9	4.8	5.6
1500	3.4	4.8	6.0	7.1
2800	4.0	5.8	7.3	8.7

C_{common} represents the fixed costs and it consists of the site civil and installation work as well as other costs that are constant independently of the power class. $C_{variable}$ consists of the transmitter cost (depending on the power level), feeder cost (which depends on the feeder length and on the type of the feeder which is selected based on the maximum supported power), and on other directly related material that depends on the cable type / length and transmitter power level.

When taking the P_{TX} = 500 W-case as a reference as shown in Figure 22.5, the cost for the transmitters with 500, 750, 1500 and 2800 W is now 1, 0.6, 0.5 and 0.45, respectively. The unit cost per meter of the antenna feeder is estimated in this analysis by comparing it with the normalized cost of 500W transmitter. In a practical case, the costs can logically be expressed in absolute commercial values of each item.

It is now possible to estimate the approximate cost for each site combination in order to build sufficient amount of sites in a certain area. The total cost is thus:

$$C_{tot} = C_{site} \cdot N_A, \tag{22.18}$$

where N_A is the total amount of sites in the area A. As an example, the A could be selected as 100×100 km^2. The radius of the site cell for, for example, 500 W-case and antenna height of 30 m is 2.4 km. The ideal overlapping in the hexagonal model can be taken into account as shown in the network layout calculation of Annex I. Based on this information it is possible to calculate the number of the partially overlapping sites in the investigated area as shown in Table 22.5.

When investigating further the cases, the unit cost of the sites can be obtained per transmitter power level category as shown in Table 22.6. The information represents a snap-shot example of certain transmitter vendor's different models when they are compared with the 500 W-model of this specific provider. It should be noted that the relative comparison depends on each case.

In this analysis, an assumption for the impact of the cable length on the costs can be rejected. As a next step, the normalized cost per transmitter category can be obtained. Table 22.7 shows the cost of the solution as a function of the antenna height in the investigated area. In this analysis, the cost effect of the towers is not considered as the assumption of the cost calculation is to use already existing ones.

Table 22.5 *The total number of sites in the planned area*

P_{TX} (W)	h_{BS} = 30 m	h_{BS} = 60 m	h_{BS} = 90 m	h_{BS} = 120 m
500	1113.1	574.8	379.6	279.5
750	884.2	450.4	294.9	215.7
1500	596.5	296.9	191.5	138.5
2800	418.5	203.9	129.8	92.9

Table 22.6 An example of the unit cost of the sites for a set of power categories

P_{TX} (W)	Normalized cost
500	1.00
750	0.60
1500	0.50
2800	0.45

Table 22.7 The normalized cost for different parameter values in order to cover 100 × 100 km^2

P_{TX} (W)	h_{BS} = 30 m	h_{BS} = 60 m	h_{BS} = 90 m	h_{BS} = 120 m
500	1113.1	574.8	379.6	279.5
750	530.5	270.2	176.9	129.4
1500	298.3	148.4	95.7	69.2
2800	188.3	91.8	58.4	41.8

As can be noted from Table 22.7, by using only the basic parameters, we can note the cost effect of the antenna height compared to the transmitter power category. As can be seen in Table 22.7, the 1500 W transmitter solution with the antenna installed in 120 m comes more attractive than 2800 W transmitter with the antenna installed to 60 m height.

As the CAPEX analysis shows, the strategy for the antenna height and transmitter power category should be done in a sufficiently in-depth level in order to make sure the optimal combinations of the parameter values and respective costs. At the initial planning phase, though, a rough estimation is sufficient in order to understand the cost-effect of different solutions as presented above.

The starting point of the technical parameter designing in the very initial phase of the network planning is the selection of the transmitter type. Table 22.8 summarizes the most important aspects that should be taken into account in the technoeconomic cost optimization when the transmitter strategy is decided.

Table 22.8 The summary of pros and cons of DVB-H transmitter power levels

Transmitter type	Benefits	Drawbacks
Low power	– Low energy consumption. – Economical and easy to install and maintain. – The installation is possible also in rooftop sites due to the smaller safety zones.	– Single site coverage small which requires high number of sites. The obtaining of sites is not straightforward in practice. – High transmission costs in case of terrestrial solution.
High power	– Low number of transmitters which results in the easier site candidate planning and the actual obtaining of sites. – The network can be build up relatively fast. – Provides large coverage areas especially in relatively open area and if the antenna can be installed high.	– Power consumption rises exponentially as the transmitter power is higher. – Liquid cooled transmitter needs additional maintenance. – The outages in coverage if obstacles (especially in urban areas) which requires additional gap fillers. – A single site breakdown causes a large service outage.

The high-level cost estimate is important to include already in the nominal planning phase in order to reject the least feasible parameter values and materials. In the detailed network dimensioning and optimization phase, also the CAPEX analysis should be more in-depth. In that phase, the OPEX should be taken into account, as the importance of the operating costs of the network might rise considerably during the operational years of the network. In that phase, the more thorough investigation of the items that have cost effect should be carried out.

22.3 Detailed Radio Network Design

22.3.1 Identifying the Planning Items

In the detailed network dimensioning, the aim is to find the optimal balance of all the relevant items that have effect on the network performance and costs. Figure 22.6 shows the outcome of the investigations, that is, the main topics to be considered as a basis for the detailed DVB-H network planning phase.

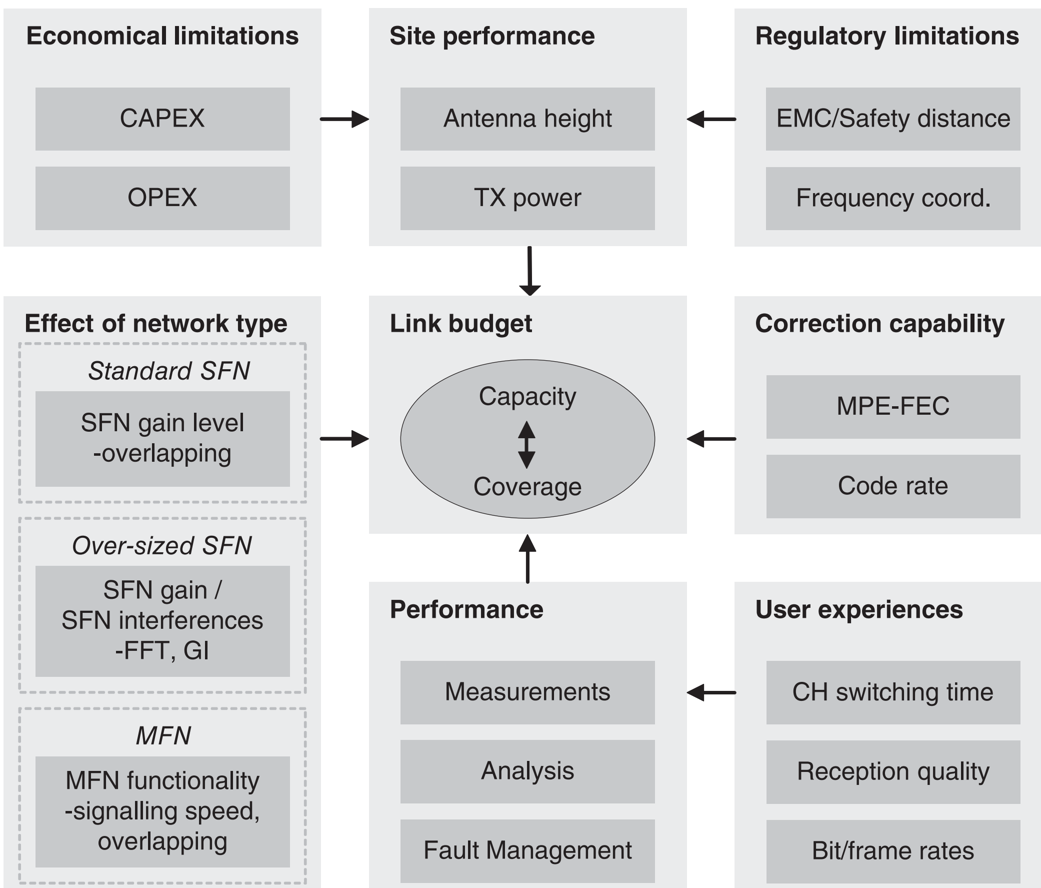

Figure 22.6 *The cross relations of the items affecting on the DVB-H radio dimensioning. In this diagram, the final aim is the balancing of the capacity and coverage by taking into account the restrictions and enhancements related to the technology, commercial and regulatory items [1]. Figure by Jyrki Penttinen.*

The balancing of the radio network performance can be done by investigating and dimensioning the cross-relations of technical, economical and regulatory items in a deep level. The ultimate goal of the detailed DVB-H radio network dimensioning is to find the cost-efficient balance between the coverage, capacity and the quality by taking into account the relevant parameters.

There might be several optimal points for the balance, that is, several different combinations of the parameter values could result in the same optimal solution technologically. Whilst the sufficiently good balance is found, it does not matter what is the theoretical set of solutions. The practical restrictions can determine though the perfect solution. As an example, even if different number of sites could produce the same optimal solution by varying several radio parameter values, the lower amount of sites could be more practical as the obtaining of the site locations is not straightforward task in practice. As an example, it might not be always possible to purchase or rent a site in the technically most suitable location due to the practical reasons that are not always predictable in the network design.

The ideal network dimensioning takes into account the future enhancement needs. It is thus important to make sure that the service areas are not reduced in the later phases. As an example, if the coverage area is done with QPSK providing large operational areas but with low capacity, the preferable aim should be that if the modulation is switched to 16-QAM which provides more capacity but with a lower coverage area per site, there are sufficient amount of gap-fillers and additional main sites operating in the critical areas, so that the customers would not experience a degradation of the services.

The mature network operation phase requires technical optimization, which might be a continuous process even in a stable network. The optimization can be done by carrying out field measurements in selected, most relevant areas of the network. The correct measurement methodology as well as the right interpretation of the measurement data is important in this phase.

22.3.2 Detailed Network Planning Process

Figure 22.7 presents the needed steps of the detailed DVB-H radio network planning process based on the identified topics of Figure 22.6. It is a continuum of the nominal planning phase, and it is meant for adjusting the higher level planning assumptions by identifying the items that may change the coverage, capacity and quality of the radio network compared to the original plan.

In the operational phase of the network, a continuous optimization may take place via the field testing and customer feedback. The optimization includes the performance monitoring as well as the fault management in case the planned coverage changes, for example, due to the faulty antenna that does not trigger alarm for the operations and management system.

This model can be considered as an enhanced version compared to typical processes presented in publicly available sources due to the including of the cost planning and optimization modules.

22.3.3 Capacity Planning

In the detailed phase of the radio network dimensioning, the capacity planning methodology does not change much compared to the one presented in the initial phase. The principle remains the same, but the balancing of the capacity and coverage, that is, the number of the sites and related costs might need several iteration rounds.

The provided capacity does have a direct relation with the cash flow of the operational network. The end-user requires normally several channels to choose from, and the quality of each one should be in acceptable level, taking into account the content type of the channels. The balancing of the number of channels and the bit rate of each one is relatively flexible in DVB-H. There can be several different bit rates defined for different channels, that is, there could be low bit rate channels for audio news type of service, whilst the highest resolution and audio quality might be required, for example, for music video channels. The upper limit for the dimensioning is determined by the bandwidth, which can be divided for several subchannels.

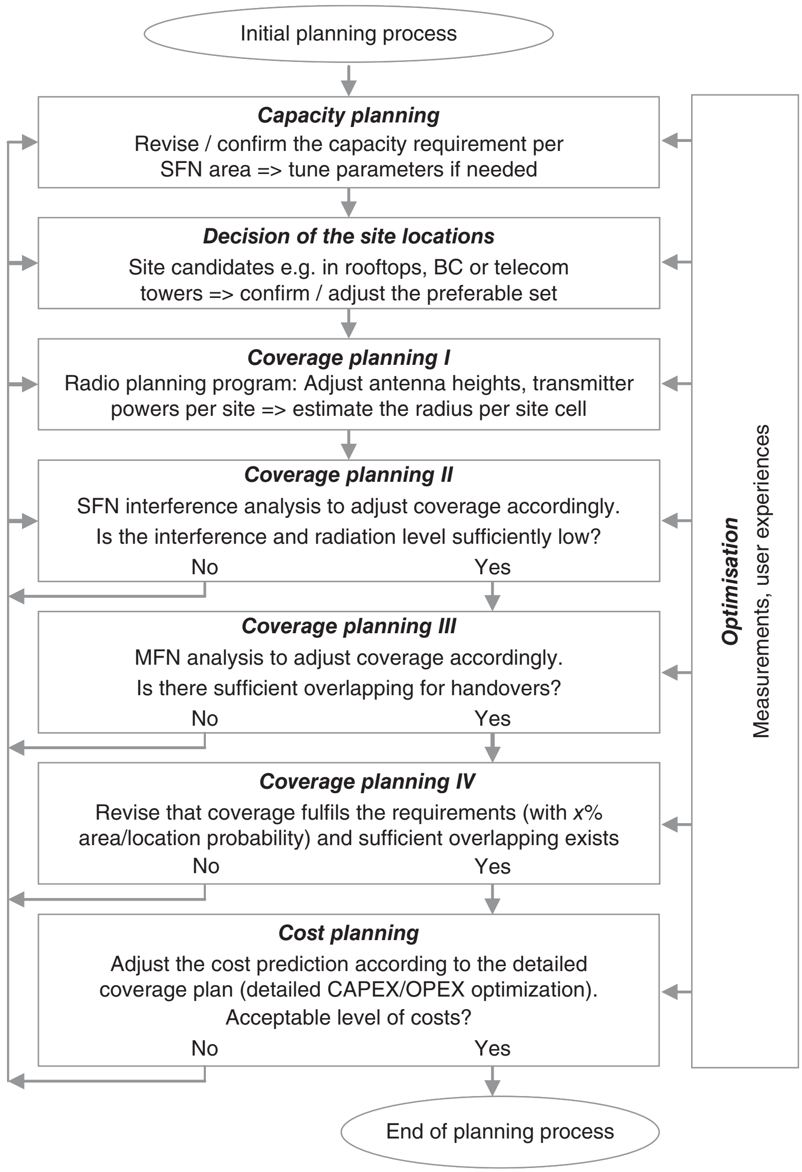

Figure 22.7 *The detailed planning phase contains in-depth analysis of the effects of the performance parameters. The steps can be iterative.*

22.3.4 Coverage Planning

22.3.4.1 Effect of Site Locations

In the initial phase of the coverage planning, a rough estimation of the number of sites is needed. The estimation can be done by carrying out a theoretical estimation of the overall path loss to different environments, for example, urban and dense urban, suburban and rural/open areas. The estimation can be done by filling in the planned areas, for example, by hexagonal type of site cells. Although this approach is theoretical, it gives a first estimation about the needed sites. If the average site antenna height and the transmitter power level are close to the reality, the prediction is sufficiently good for the nominal planning purposes.

In the detailed planning phase, the more accurate site locations should be known. This is the most challenging part of the coverage planning, as the practical sites can rarely be found in the ideal locations according to the uniform type of the initial network layout. In addition, there might be restrictions in the antenna height and transmitter power level usage. As an example, the already existing broadcast or mobile communications towers are normally equipped with other antennas of various systems, as GSM, UMTS, radio, TV and link antennas, which might prevent the additional installations due to the EMC restrictions.

The variation of the practical antenna heights and transmitter power levels affects the optimal CAPEX and OPEX calculations, so the technoeconomical analysis should be done accordingly in order to find the proper balance for the radio parameter values and the network cost.

In order to provide positive user experiences, the coverage areas should be overlapping. In the SFN network, this results in the additional SFN gain, which does have effect on the performance via the link budget enhancement [25]. In case of the MFN, the sufficiently overlapping areas provide the handover functionality without problems.

22.3.4.2 Site Cell Range Predictions

When the final site locations are known, the detailed coverage estimation can be done. The selection of the suitable radio propagation prediction model is important, and in the detailed analysis, a local adjustment of the propagation model parameters is recommendable. The latter can be done by carrying out field measurements and by correlating results with predictions, and correcting the propagation model's parameter set (e.g. by tuning clutter attenuation values).

In the detailed phase of the network coverage planning, the possible SFN interference level should be investigated thoroughly. If the theoretical SFN limits are not exceeded, it is straightforward to assume that there are no interferences present. If reflections are expected from distant obstacles, they increase the effective delay of the radio components and the probability of the occurrence of interferences increases.

In case of interferences, the path loss prediction method, that is, analysis to estimate the carrier levels as a function of the geographical location, should also include the interference analysis. As for the sole coverage within noninterfered SFN area, source [1] shows examples about the commercial planning tool's (NetAct Planner) prediction model usage in urban and dense urban environment. The model of the NetAct Planner is based on the Okumura-Hata [18], which also takes into account the local clutter attenuation factors. As the used map resolution in the presented cases studies is in order of 30 meters, it does not provide the most accurate coverage prediction, for example, inside the street canyons. Nevertheless, the model is sufficiently accurate for planning purposes especially in suburban areas.

NetAct Planner is typically utilized in the 2G and 3G mobile network coverage predictions, and can be adapted to the basic DVB-H coverage prediction by applying the same principles. Various different models can be utilized as a base of the predictions, including the more accurate ray-tracing based 3-dimensional models. If the Okumura-Hata model is utilized, the basic outcome of the tool's propagation model is further enhanced by taking into account additional attenuation values that depend on a separate cluster map of the investigated area. Several different cluster types can be defined, for example, for the water areas and forests with different vegetation densities. The cluster values can be further adjusted by comparing the prediction of different area types with the field test results. If the attenuation values of the clusters are estimated well enough, this method results in more realistic predictions compared to the sole Okumura-Hata model that generalizes the prediction solely over different city and rural environments. Nevertheless, the basic Okumura-Hata model as such indicates the needed number of the sites per area type sufficiently accurately for the nominal planning purposes.

In practice, according to Ref. [6], there might be well over 20 taps in the radio channel. This should be taken into account in the deep level radio network planning and analysis. The recommendation of Ref. [6] is to use at least 12 taps for the respective simulations although the practical terminal deployment would not take the full advantage of all these components. The basic coverage area is studied in Ref. [6].

The balance between the economical aspects and technical solutions depends highly on the wanted quality of the service level. It is thus logical to decide an initial target for the network's coverage and capacity, and to investigate with what parameter combinations the target could be achieved. The detailed network planning might require several iteration rounds depending on the outcome, that is, if the network cost in initial and longer term turns out to be expensive even in the technically optimal point.

22.3.5 Local Measurements

The field measurements are needed for the revisions of the performance level of DVB-H. As DVB-H is a broadcast system without its own uplink channel, the terminals themselves cannot be used directly as reporting devices. The field measurements provide data for performance analysis as well as for fault management.

Other usage of the field measurements is related to the radio propagation prediction model adjustment of DVB-H. The received power level combined with the location information can be utilized in the typical planning programs in order to correlate and correct the model parameter values, for example, the cluster attenuation values. The correction can normally be made either manually or automatically.

When the field measurements are initiated, it is important to calibrate the equipment. Source [1] shows an example of the differences in the channel display when three different DVB-H terminal units are used as measurement equipment. Figure 22.8 shows the further processed cumulative presentation of the differences of the channel displays. As can be seen, the 50%-ile point results in −57.3 dBm, −58.8 dBm and −61.2 dBm for each terminal.

In the field measurements described above, the terminal's already existing field test program displays the received power level based on the interpretation of the gain values of the automatic gain control (AGC). In practice, the minimum step size of the AGC tends to be typically in order of 1 dB. The accuracy of this method is approximately 1…2 dB depending on the received power range. If a power meter is used instead, the accuracy is roughly in the same order. In addition, the quality of the calibration affects on the final estimate

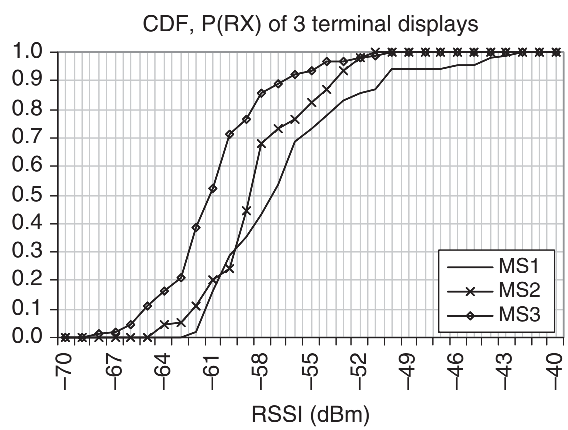

Figure 22.8 Comparison of the RSSI display of 3 different DVB-H terminals used in Ref. [1]. Cumulative distribution of the laboratory measurement with 90 samples per terminal. Figure by Jyrki Penttinen.

of the received power level. The importance of the calibration has been noted also in Ref. [26] which describes the challenge in the interpretation of field tests when high accuracy is required.

In order to minimize the inaccuracy of the channel display, the corresponding calibration can be carried out, for example, by installing a coaxial cable directly to the input of the DVB-H receiver and by connecting a TS generator to it. If instead more advanced field or laboratory equipment like spectrum analyzer is used as in Ref. [8], the accuracy of the averaged samples is logically more reliable. There are also other measurement criteria like error vector that can be utilized in the signal-to-noise ratio calculations as presented in Ref. [27]. Nevertheless, the main idea of the presented field test methodology is based on a portable device which can collect sufficiently accurate measurement data in all the user environments, including indoors, without the need to connect it to the external power supply.

22.3.5.1 Coverage Area

The basic coverage area can be studied by observing the received power level P_{RX} of the site cell around the functional coverage limit as presented in Figure 22.9. The receiving of the radio signal can be done with a scanner using respective bandwidth and averaging settings, or with a terminal capable of showing the radio parameter values as explained in Ref. [1]. In the analysis, it is important to understand the effect of the averaging of the samples during the measurement. If a real audio / video stream can be transmitted, a normal DVB-H receiver can be used as parallel equipment for the subjective quality check. This method would be sufficient in the very initial phase of the planning, in order to understand the parameter value behavior on the quality of the received contents.

One of the aims of the radio coverage measurements is to revise the accuracy of the applied prediction models in the investigated area type. Based on the received power level distribution, the respective model adjustment can be made by adjusting the outcome of the prediction with the offset the field measurements

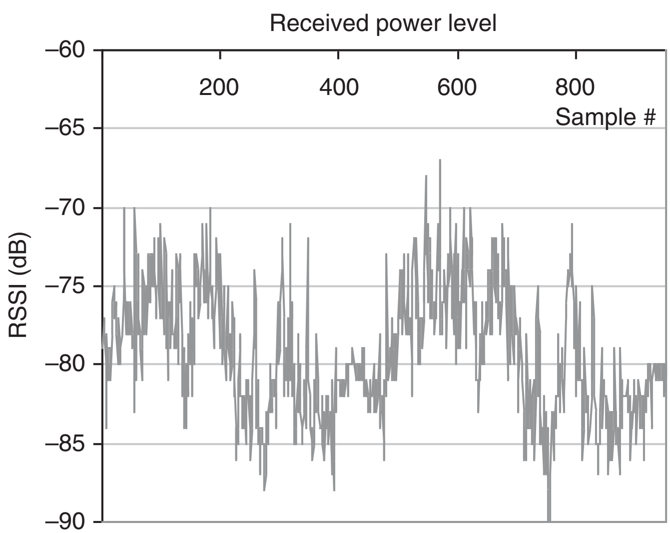

Figure 22.9 *An example of the collected received power levels by moving the receiver close to the functional cell edge. The received power level samples and corresponding geographical locations can be observed by post-processing the measurement data to a map format which indicates the expected coverage limits.*

provide. Depending on the model the area cluster attenuation factors can be modified as it is one of the main items that affect on the accuracy of the predictions.

It should be noted that the accuracy of the received power level measurement depends on the equipment. The DVB-H terminals can be used well for the fast revision of the basic coverage, but the respective accuracy of the analysis suffers typically about 2 dB additional margin due to the lack of calibration of the field strength displays.

For the more in-depth revision of the achieved quality levels as a function of the received power level requires deeper analysis which is presented in the following Chapter 22.3.5.2. The method is based on the data collection from the field and on the respective postprocessing.

22.3.5.2 Error Correction

Source [1] shows methodology for the field measurements and their analysis in order to find the detailed level of information about the performance of the network. For the optimization of the network performance, basic field strength measurements are important to carry out in an early phase of the network deployment, by observing the functional limits via measurement equipment, for example, with a GPS receiver producing the location information for the measurement results.

In this method, the frame error rate has been identified as a useful criterion when the basic performance indicators can be collected from the air interface. This is logical as the end-users experiences the quality level variations in the real-time depending on the correctness of the received frames. The method shows that the quality can be analyzed by arranging the collected FER and MFER occurrences as a function of the received power level. The MPE-FEC gain can be obtained when the results are converted into a cumulative distribution format. As the resolution of the presented equipment is 1 dB for the received power level, the corresponding proportion of error-free frames, frames that can be corrected with MPE-FEC and the ones that remain erroneous can be normalized per each RSSI category. In this way, the breaking point of the observed criterion of MPE-FEC and FEC is possible to obtain. When this method is utilized, it is important to collect a sufficient amount of measurement data for the statistical reliability. According to [2, p. 83] the accuracy of the frame error rate would be sufficient if at least 100 samples are collected. These samples include both errorless and erroneous MPE-FEC frames.

The FER information, that is, frame errors before MPE-FEC specific analysis, is obtained after the Time Slicing process, and the MFER is obtained after the MPE-FEC module. If the data after MPE-FEC is free of errors, the respective data frame is de-encapsulated correctly and the IP output stream can be observed without disturbances. As the erroneous frame indicates directly the user perception of the quality, the frame error rate before and after MPE-FEC are useful indicators for the performance studies. As stated in [2, p. 83], an erroneous frame destroys the service reception for the whole interval between the time-sliced bursts.

According to Ref. [2] the 5% FER / MFER can be considered as a criterion for the acceptable quality level. This point can be studied in various ways. One method is to collect sufficient amount of samples in certain spots, for example, in area of 5 × 5 meters by moving slowly the terminal within the area and thus rasterizing the area. By interpreting in more ample way the statement of the Ref. [2], the respective FER and MFER error rate can now be obtained by applying the following formulas:

$$FER(\%) = 100\frac{F_{err\_bm}}{F_r} \tag{22.19}$$

$$MFER(\%) = 100\frac{F_{err\_am}}{F_r}. \tag{22.20}$$

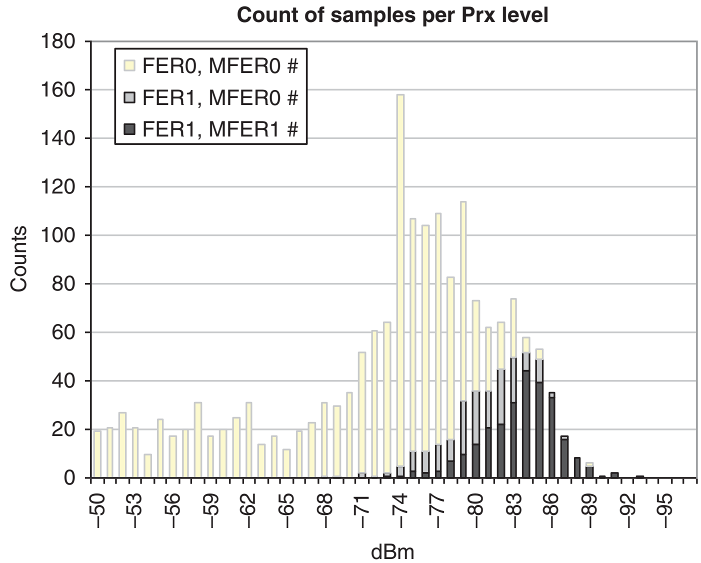

Figure 22.10 *The first step of the proposed methodology arranges the occurred samples as a function of the received power level, showing the amount of error-free instances, occurred instances with frame errors that could be corrected with MPE-FEC, and remaining erroneous frames after MPE-FEC.*

F_{err_bm} represents the erroneous frames before MPE-FEC and F_{err_am} is the number of residual erroneous frames after MPE-FEC. F_r indicates the total number of received frames.

Source [1] presents a methodology that is based on the arranging of the occurred samples as a function of received power levels as shown in Figure 22.10.

The method is based on the normalizing of the occurred events per each RSSI category individually as shown in Figure 22.11. This presentation is valid when other non-RSSI related effecting components are not present. These components might include impulse-noise type of interference which is spread over a wide RSSI scale or the effects that occur after exceeding the Doppler shift limits determined by the used parameter value combination.

The format shows now directly the proportion of the frame errors that could be corrected. It is thus possible to have a closer look on the breaking point of interest, for example, in 5% FER/MFER. Figure 22.11 shows an example of the breaking point. The difference between FER and MFER can be interpreted directly as the level of MPE-FEC gain in dB. This gain can further be interpreted in such a way that the basic coverage area of the site cell is limited to the received power level corresponding to FER 5%, whilst the RSSI level is enhanced by ΔP_{TX} when MPE-FEC is applied.

It is straightforward to analyze the respective performance in terms of received power level from Figure 22.12 which shows an example of the FER5 limit of −74.9 dBm and the MFER5 limit of −78.0 dBm resulting ΔP_{TX}, that is, MPE-FEC gain of 3.1 dB in this specific case.

22.3.5.3 Accuracy of Error Correction Analysis

The applied FER and MFER comparison methodology requires sufficient amount of collected data per each RSSI level especially around the observed breaking point of FER and MFER. The sufficient sample number per RSSI level can be investigated in practice by collecting different amount of samples and by comparing

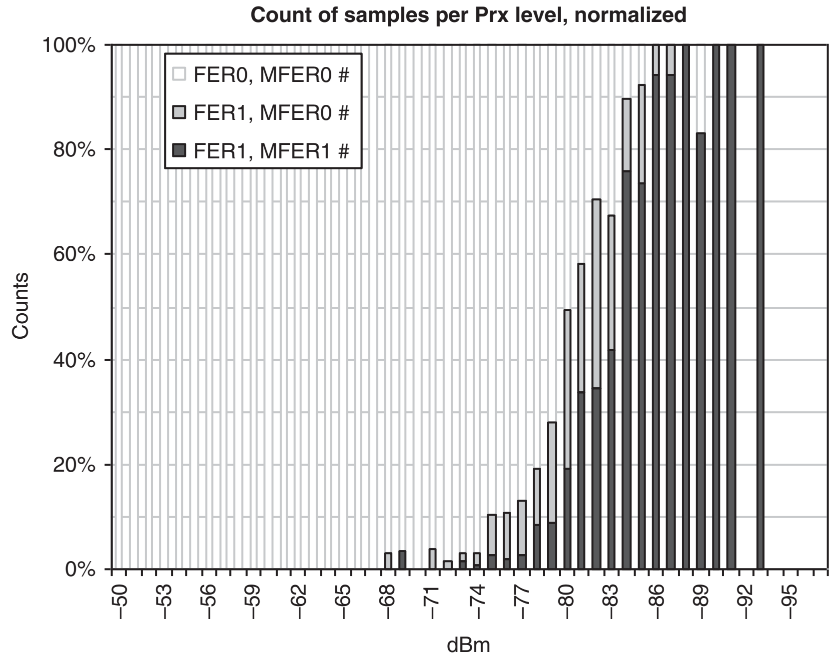

Figure 22.11 The method shows the normalized number of occurred events per each RSSI.

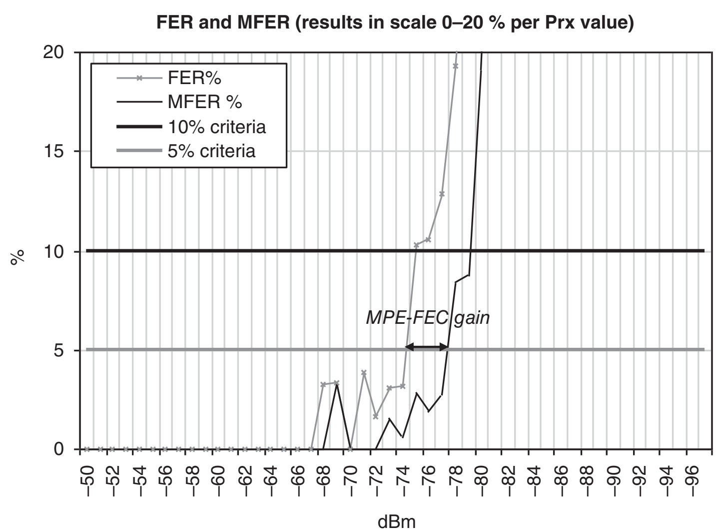

Figure 22.12 An amplified view to the 5% FER / MFER level.

the variations of the results especially in the FER / MFER 5% breaking point. Reference [2] states that in order to provide sufficient accuracy, it is necessary to analyze at least 100 frames. This statement is relatively loose though, although in practice the MFER has a clear breaking point with a small increment of C/N. The statement does not explain more thoroughly about the requirement for the number of related RSSI values.

As a starting point, for the comparison of the FER and MFER graphs, the resolution of the outage-% scale (y) is important. When the number of samples n is known, that is, the number of occurred instances ("no errors," "errors that could be corrected with MPE-FEC," and "remaining errors"), the sampling resolution R in normalized outage %-scale can be obtained via the following Formula:

$$R = \frac{100}{n}\%. \tag{22.21}$$

If the amount of the samples n is 200 per certain RSSI value, the resolution would be 0.5%-units indicating the maximum accuracy that can be obtained with that number of samples.

In order to make a more thorough error estimate, according to the simple random sample from a large set of values – comparable to the polling error estimates as the randomness of the measurement is fulfilled – the maximum error is a single re-expression of the sample size n. Whilst the sampling fraction is less than 5%, the margin of error can be estimated via the simple random sample principle. It assumes that the "population," which refers in this case the entities that can be either erroneous or free-of-errors, is infinite. This assumption can be taken as a basis for the previously described method. As an example, the margin of error at 95% confidence for the simple random sample principle with n samples is:

$$Err\,[95\%] = \frac{0.98}{\sqrt{n}}. \tag{22.22}$$

By applying this principle, an example of 100 samples per RSSI category would produce a margin of error of 9.8% whilst a case with 50 samples produces a value of 13.8%. This estimate can be applied to the values of nonerroneous events, erroneous events before MPE-FEC and erroneous events after MPE-FEC when the total amount of samples is known.

The accuracy of the individual y-axis values of a single curve (FER or MFER) can be estimated via the above mentioned principle. Then, the MPE-FEC gain, which is the difference between the indicated points of FER and MFER curves in 5% line, can be further estimated by observing the extreme values of the accuracy of the single curve. As an example, a measurement result of Ref. [1] can be analyzed. Let's select a case of {16-QAM, CR 2/3, MPE-FEC 2/3, GI 1/4, FFT 4K}, which produced the sample distribution as shown in Figure 22.13.

Table 22.9 summarizes the occurred total amount of samples per RSSI category and shows the respective margin of error for 95% confidence by applying the above mentioned formulas. As can be seen, the accuracy of the RSSI levels of −70 dBm and −71 dBm is not sufficient, but Figure 22.13 indicates that even if the respective number of total samples is low for these RSSI categories, the received power level has been clearly sufficient for error-free functioning. It can be seen from the margin error analysis that the MFER5 can be found between RSSI levels of −76 dBm and −77 dBm when the confidence level of 95% is applied.

Let us observe the RSSI scale of −70…−80 dBm by analyzing individually the error margin for each RSSI category in such a way that the total number of occurred events, that is, "error-free" (FER0, MFER0), "FEC resulting error but MPE-FEC has corrected" (FER1, MFER0) and "error after MPE-FEC" (FER1, MFER1) is taken into account for the error margin calculation.

Figure 22.14 shows the result of the analysis with respective marginal for each respective FER and MFER value. When the FER5 and MFER5 is observed, FER5 results in values of $P_{RXmin} = -71.9$ dBm, $P_{RX} = -72.1$ dBm and $P_{RXmax} = -72.3$ dBm. MFER results in $P_{RXmin} = -76.3$ dBm, $P_{RX} = -76.4$ dBm and $P_{RXmax} = -76.5$ dBm dBm. The respective MPE-FEC gain can thus be informed as $-76.4 - (-72.1)$ dBm $= 4.3$ dB with

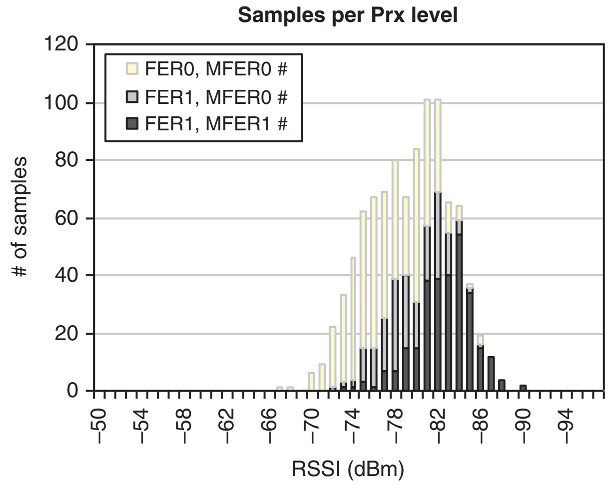

Figure 22.13 *An example of the collected data distribution over the RSSI scale.*

the value range of [4.0, 4.6], taking into account the error margin. This example represents a typical situation for the measurements and the error margin can thus be assumed to reside within 10…15% for individual FER5 and MFER5 curves, and the respective MPE-FEC gain would have typically a maximum of 10% error margin. It should be noted that also the RSSI scale has an error margin that depends on the accuracy of the calibration of the terminal (±2 dB) and on the mapping of the AGC process to the RSSI value (±1 dB).

If the accuracy of the above mentioned measurement methodology should be enhanced, more samples would need to be collected per RSSI category. It is obvious that apart from the impulse noise environment which might produce errors in large RSSI window, a sufficiently good field strength provides with reliable results (nonerroneous samples) even with low amount of samples, whereas erroneous frames are produced relatively often in the lowest field. If the investigation does not concentrate on the impulse noise, it is thus

Table 22.9 *An MPE-FEC error analysis for the example shown above. (* indicates too low sampling rate)*

RSSI (dBm)	# of samples	Margin of error (95%)	Sampling fraction	# of "FER1, MFER 0"	# of "MFER 1" (erroneous samples)	FER and margin of FER (%-units)	MFER and margin of MFER (%-units)
−70	6	40.0	16.7	0	0*	0.0 ± 0.0	0.0 ± 0.0
−71	9	32.7	11.1	0	0*	0.0 ± 0.0	0.0 ± 0.0
−72	22	20.9	4.5	1	0	4.5 ± 0.9	0.0 ± 0.0
−73	33	17.1	3.0	2	1	9.1 ± 1.0	3.0 ± 0.5
−74	46	14.4	2.2	3	1	8.7 ± 0.9	2.2 ± 0.3
−75	62	12.4	1.6	12	3	24.2 ± 2.4	4.8 ± 0.6
−76	67	12.0	1.5	14	1	22.4 ± 2.5	1.5 ± 0.2
−77	69	11.8	1.4	18	7	36.2 ± 3.1	10.1 ± 1.2
−78	80	11.0	1.3	32	7	48.8 ± 4.4	8.8 ± 1.0
−79	67	12.0	1.5	25	15	59.7 ± 4.5	22.4 ± 2.7
−80	84	10.7	1.2	16	15	36.9 ± 2.0	17.9 ± 1.9

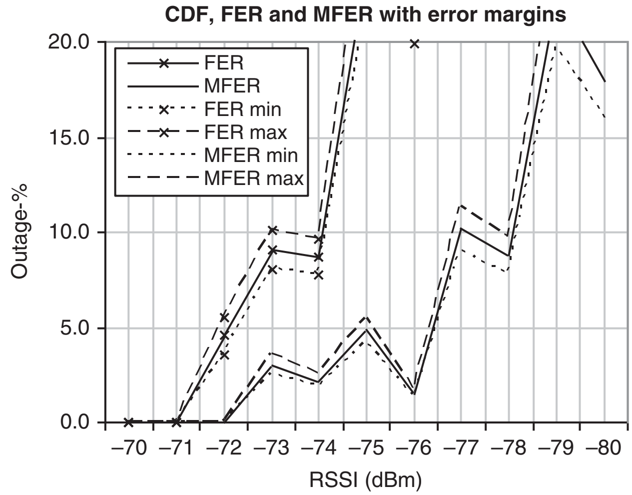

Figure 22.14 *A view to the case result graph showing the FER and MFER curves with the respective error margin that is calculated for the absolute values of the samples for each RSSI category individually.*

sufficient to collect the major part of the samples around the critical RSSI scale of the 5% breaking point of FER and MFER.

On the other hand, the idea of the above-mentioned method is to collect fast snapshot data which means that the task could be carried out in relatively short time period per investigated area. Typically, the storing of each measurement sample (frame) could take about 1 second, and assuming that in sufficiently accurate analysis ±5 RSSI values, that is, about 11 RSSI categories need to be measured, for example, with 300 samples per RSSI category producing margin of error of 5.7%, this would require $11 \cdot 300 \cdot 1\text{ s} = 3300\text{ s}$, that is, 55 minutes would be needed. In typical measurement routines of an operator, this might be in the practical limit of the used work time for a single measurement, although the error margin would provide with a better accuracy.

According to the results obtained in Ref. [1], the analysis described here might produce relatively smooth FER and MFER curves, which results in a straightforward way of comparison of FER5 and MFER5 points and indicates thus the respective MPE-FEC gain with a relatively high confidence level. In some cases, though, the curve behavior is more “aggressive,” producing high peaks along the RSSI scale an making the FER5 and MFER5 interpretation challenging. A part of these peaks can be explained by impulse noise, and some can be occurring due to the lack of samples. Nevertheless, the most important RSSI scale is in the range corresponding the FER5 and MFER5, and normally the phenomena of the correction capability can be seen graphically even if the FER5 and MFER5 breaking points are not implicitly possible to interpret. One way of trying to get the approximate value for the breaking point in this type of cases would be to interpolate regression curves for the FER and MFER graphs. The error margin could be estimated by observing the difference between the regression curve and the original deviating points. Nevertheless, the method would not reflect totally the reality as the errors do occur in nonpredictable way along the RSSI scale in these cases making the reception unstable.

22.3.6 Effect of SFN

The SFN functionality has advantages in the normal operation as it produces gain by combining the signal energy of the multipath propagated separate radio signals. Reference [7], though, has concluded that as SFN gain is not persistent across all locations of the DVB-H network, it is not considered in the radio link budget. In the theoretical case, if the carrier level is considered useful only if the minimum functional level of the investigated parameter settings is achieved, the SFN gain can be obtained only in the overlapping area with *C/N* complying with the minimum functional value. Nevertheless, the *C/N* values just below the functional limit combined with the *C/N* value of other signals below the functional limit might set the sum of these signals above the functional level which enhances the coverage area between the sites. The more carrier components (above the noise limit) there is present in a certain spot, the more probably the sum of the signals exceed the minimum required *C/N* limit.

22.3.6.1 Noninterfered Network

Source [1] presents a variation of the Monte-Carlo simulation method, which can be utilized for estimating the SFN gain as a function of the most relevant radio parameters. The method is based on the investigation of the gain that is achieved in the physical level. The simulations resolve the received power level distribution in a single cell case, which is compared to the combined signals that are produced by n site cells. As the simulations are carried out over the whole investigated area at once instead of individual simulations of the subregions formed by the raster of the area map, the method provides a fast yet sufficiently accurate estimate of the gain, that is, about the increased probability to receive the signal correctly.

The developed method can be compared with the results found in other public references. The practical field test based references are, on the other hand, limited as they are typically based on only few sites. The presented simulation method takes into account all the site cells that contribute either to the useful received carrier power or to the received interfering power within their range of radio propagation.

In the noninterfering case, there are only useful carriers present. The basic principle of the SFN gain calculation is based on the summing of the received power levels $[P_{RX}(C)]_k$ in absolute values (W) as shown in the formula:

$$\left[P_{RX}(C)\right]_{tot}(W) = \sum_{k=1}^{n} \left(\left[P_{RX}(C)\right]_k\right)(W). \tag{22.23}$$

The principle of power summing has been presented in Ref. [28] and it has been used in the received power level estimations, for example, in Ref. [29]. For the power summing, the absolute value is thus first obtained by:

$$\left[P_{RX}(C)\right](W) = \frac{10^{\left(\frac{P_{RX}(dBm)}{10}\right)}}{1000}. \tag{22.24}$$

In order to compare the SFN gain value, the reference level of received power level, at given time and location, is the strongest possible from the sites that can be received above the noise floor:

$$P_{RXref} = \max\left\{\left[P_{RX}(C)\right]_1, \left[P_{RX}(C)\right]_2, \ldots, \left[P_{RX}(C)\right]_n\right\}. \tag{22.25}$$

Now, the SFN gain G_{SFN} is the difference between the summed and reference levels in dBm values:

$$\left[P_{RX}(C)\right](dBm) = 10\log\left(\frac{\left[P_{RX}(C)\right](W)}{1\cdot 10^{-3}}\right) \tag{22.26}$$

$$G_{SFN}(dB) = \left[P_{RX}(C)\right]_{tot}(dBm) - \left[P_{RX}(C)\right]_{ref}(dBm). \tag{22.27}$$

Source [1] assumes that the whole OFDM channels, that is, multipropagated signals, can be summed without taking into account the frequency selectivity within each bandwidth. It is obvious that especially in

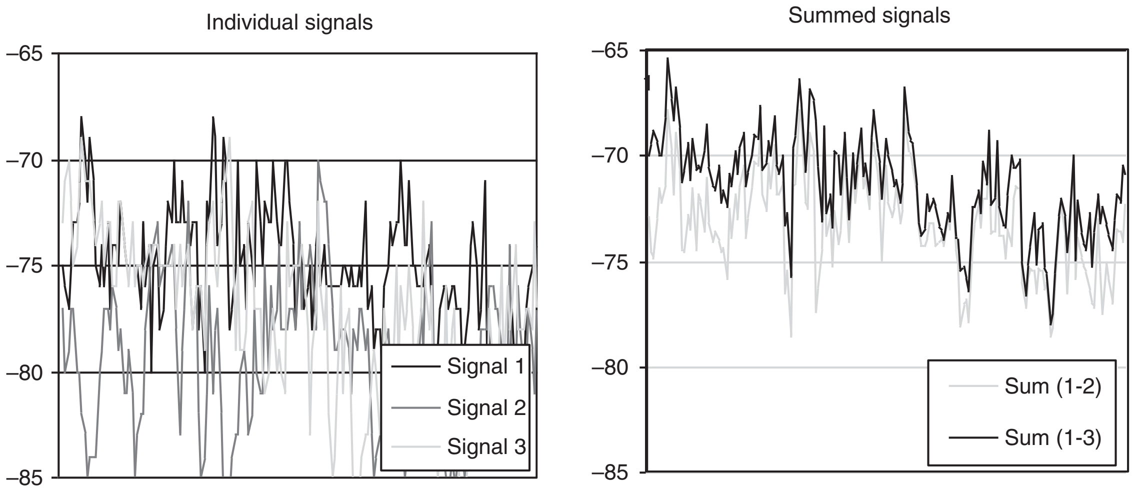

Figure 22.15 *An example of individually received signal levels in the same area, and their total power based on the direct power summing. The same terminal was utilized.*

case of Rayleigh fading, a part of the subcarriers of a single OFDM bandwidth experiences higher attenuation than the others.

As the overall effect is a combination of possibly some destroyed subchannels via the Rayleigh fading, and corrected subchannels via the error recovery properties of MPE-FEC, it can be estimated that the significance of the frequency selectivity is low for the calculation of the SFN gain. The practical observations from the DVB-H trials support this reasoning. Furthermore, the C/N values that are utilized as criteria in the simulations are based on the Ref. [2] which has included the radio channel behavior already in the presented C/I limits.

The principle of the general power level summing for the high-level simulation purposes is shown in Figure 22.15 which presents an example of three noncorrelating individual signals with approximately the same average levels as a function of time and location. These data is derived from the individual measurements of a single site carried out in Ref. [1], and the represent a mix of channel types including the long-term and fast fading. These data are averaged by the terminal to the 1 second resolution, so the highest Rayleigh attenuation peaks can not be shown in this example.

Figure 22.15 presents also the sum of these signals by applying the SFN expression (4-5). As can be noted, the resulting signals, in addition to the increased total received power level, have smaller variations. Figure 22.16 shows the CDF of the individual and combined signal levels. Tables 22.10 and 22.11 summarize this behavior by presenting the numerical values.

The above presented example shows the principle of the summing of power levels. There are also analytical methods developed for the fast fading signal combining, and SFN coverage probability estimation as presented in Refs. [30, 31].

The noninterfered case of the simulation results presented in Ref. [2] and Figure 22.17 can be used for the SFN gain estimations. The suitable principle is based on the filtered coverage area of a single cell, which gives the reference for the cumulative C/N. Then, more cells are added according to the SFN reuse pattern size, and the C/N distribution is simulated for each case. The G_{SFN} can be obtained by comparing the 10% outage probability level (which corresponds to a 90% area location probability within the whole area) of the CDF of each case.

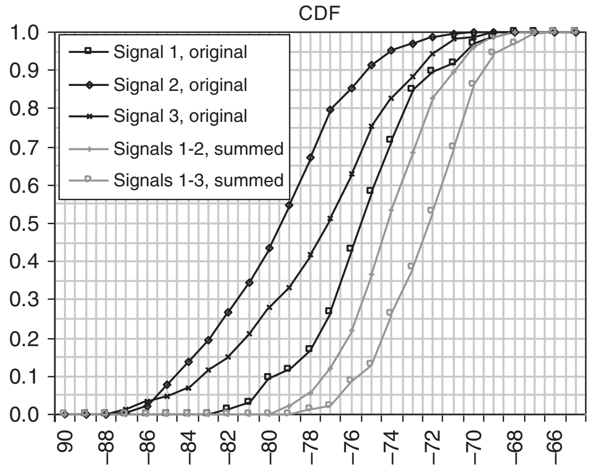

Figure 22.16 *CDF of the 3 individual and summed signals presented in Figure 22.16.*

Table 22.10 *Numerical values of individual signals of the example*

Signal #	Average RSSI (dBm)	Stdev (dB)	50%-ile of CDF (dBm)
1	−75.0	2.9	−75.5
2	−79.2	3.5	−79.5
3	−77.2	3.9	−77.2

Table 22.12 presents comparisons of the SFN gain studies presented in Ref. [1].

According to Table 22.12, the SFN gain value varies largely even with a low number of transmitters (2–3), the minimum gain being 1…1.5 dB and maximum being over 6 dB, depending on the source. The gain obtained from Ref. [1] represents a typical average value range of the other sources thus correlating with major part of the presented references when the number of sites is up to 3.

Unlike the other presented results, Ref. [34] has concluded that the SFN gain would actually be negative in a two-transmitter case due to the significant degradation of the signal quality measured by the modulation

Table 22.11 *The values of the summed signals of the example. G_{SFN} is calculated by comparing the average and 50%-ile level with the strongest individual signal of each case*

Summed signals	Average RSSI (dBm)	Stdev (dB)	50%-ile of CDF (dBm)	G_{SFN}/reference signal #
1, 2	−73.2	2.3	−74.2	+1.8 dB (averages) +1.3 dB (50%-iles) Ref: Signal #1
1, 2, 3	−71.4	2.3	−72.2	+3.6 dB (averages) +3.3 dB (50%-iles) Ref: Signal #1

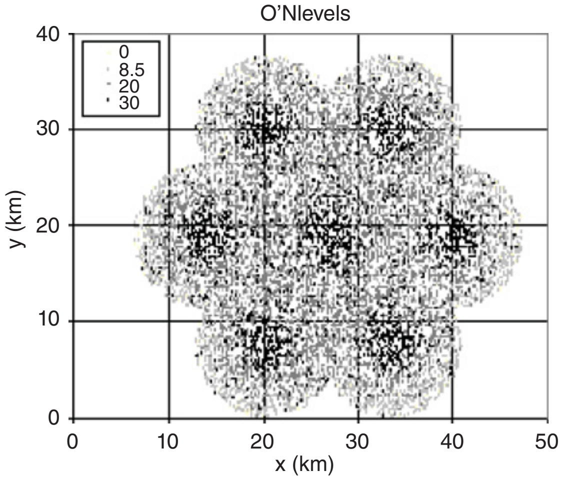

Figure 22.17 *An example of the simulated area with the SFN reuse pattern size K = 7.*

error rate (MER) indicator. This outcome clearly differs from the other references. The source does not present the cause for the effect, but the explanation might be that the presented test setup includes some feedback loop interferences between the input and output of the gap filler. Regardless of the final root cause, this is an important observation and indicates the importance of the correctness of the practical installations.

The reference [6] shows that as the gain margin of the gap filler gets close to zero the ripple increases considerably. At 0 dB gain margin point the system is already unstable. As the gap filler setup is critical in the SFN analysis, the most reliable results might be derived from the test cases where only main transmitters are switched on and off as explained in Ref. [32].

As an overall observation, the field tests that are related to the SFN gain have normally been carried out with only two transmitters. In Ref. [37], a method of adding the electrical field strengths from a total of three main transmitters resulted in a total SFN gain of 2.3 dB. Nevertheless, in a further analysis related to the same network, Ref. [32] has found that the CINR results in only 0.3 dB gain. This indicates that SFN gain should be estimated case-by-case by investigating what is the effect of the interference level, for example, for the MER degradation, not only the sum of electrical field strength.

Figure 22.18 shows a summary of the SFN gain results obtained via the simulations of Publication IV, Figure 12, of source [1]. The originally presented nonuniform SFN reuse pattern size scale has been converted linear and a logarithmic regression curve has been added to the resulting plots. The reference results from Table 22.12 have also been marked in the graph for comparison purposes. As can be observed, the SFN gain values of the found references vary considerably. The main reason for this can probably be explained by the differences in the practical setup, which might not be described in a sufficiently detailed level in the references in order to understand what the size of the coverage area of a single site is compared to the coverage are produced by the gap-filler or secondary transmitter.

Although the results of the source [1] are based on simulations, the setup was made in a controlled way by selecting the overlapping proportions of the site cells according to the hexagonal layout model, that is, there were sufficiently overlapping parts present but without excess. The results are derived always from the filtered cell areas, that is, within the dimensioned cell based on the given area location probability. Furthermore, the simulations are performed for varying amount of transmitters up to 21.

Table 22.12 *Comparison of the SFN gain values via different studies*

Reference	Method	Sites	SFN gain
Publications IV, X of Ref. [1]	Path loss based simulations, hexagonal model, and noninterfering network. Results obtained over the whole area.	1…21	3 sites: 2.8 dB 4 sites: 3.4 dB 7 sites: 4.6 dB 9…21 sites: 4.8–6.0 dB, see Figure 22.18 Note: QPSK, CR=1/2, MPE-FEC=1/2, GI=1/4, FFT=8K. For 16-QAM, the gain is up to 6.4 dB (this maximum occurs with 19 sites).
Ref 1/1: [32]	Field tests, Electrical field measurements and mathematical formula for the gain calculation.	3	2.3 (stdev 1.0…1.2)
Ref 1/2: [33]	Field tests, electrical field measurements with mathematical modelling and CINR estimate.	3	2.3 dB (electric field method) 0.3 dB (CINR method)
Ref 2: [34]	Field tests, main transmitter (TX) and gap filler (GF), Modulation error rate observations.	1 TX + 1 GF	2…3 dB (−52.3 dB → −54.5), but MER decreased about 7…8 dB (20 → 28), resulting negative SFN gain.
Ref 3: [7]	Link budget investigation.	N/A	SFN gain not recommended for radio link budget.
Ref 4: [35]	Field tests.	2	SFN gain was evident, but no exact value presented.
Ref 5: [36]	Indoor field tests, Barcelona, Spain.	2	SFN gain was calculated and treated as a statistical variable. The result was obtained by measuring the average gain in terms of field strength due to the contribution of the gap filler. The variable showed a log-normal behavior with a mean value of 6.1 dB (5.6…6.4 dB) and 6.4 dB of stdev. The range of the values was very wide; study did not thus recommend using SFN gain values in link budget.
Ref 6: [36]	Finnish field tests, outdoor.	2	Case 1: 1.5…3 dB, closer to 1.5 dB Case 2: 3 dB Case 3: 5.4 dB All results are related to the higher field strength.
Ref 7: [36]	Finnish field tests, outdoor and indoor.	2	One transmitter was used in outdoor pedestrian and indoor corridor measurements. The results show better MFER at a certain signal level when using two transmitters. The SFN gain was 1…2 dB (in average 1.5 dB).
Ref 8: [36]	Common outdoor field tests, Barcelona, Spain.	2	The histogram of the accumulated SFN gain values resulted a mean of 6.08 dB and a stdev of 6.36 dB.

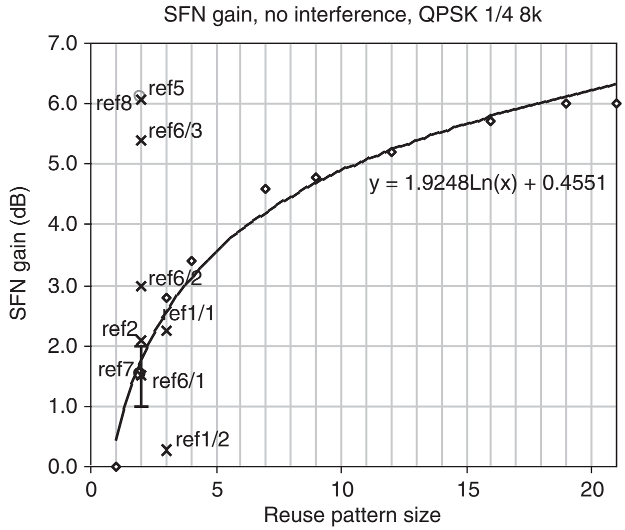

Figure 22.18 *The SFN gain obtained via the simulations of Ref. [1] for the non-interfered network (QPSK, FFT 8K, GI 1/4). The comparable results of the studied references are also presented. Figure by Jyrki Penttinen.*

The presented methodology for the comparison of the *C/N* distributions of a single cell and multiple cell cases creates a logical and controlled environment for the SFN gain estimation. Furthermore, the presented simulation method brings novel aspect for understanding the complete behavior of the SFN gain even in an oversized SFN area where partial interference is present. The model does not explain the practical phenomena of lower MER levels that were observed in Ref. [34], though. The degradation can be assumed to arise from the nonoptimal performance of the terminal or test setup.

22.3.6.2 Interfered Network

In practice, there occur outages for single streams within the planned cell area which can possibly be recovered by the reception of other streams. In this sense, the SFN gain is not limited to the outer border areas of the cell. Furthermore, when the theoretical maximum limit of SFN is exceeded, there might appear interferences as the sites outside the area determined by the maximum allowed guard interval distance converts to interferers instead of sources of useful signals. Source [1] presents the principle of this mechanism via simulations. According to the results, it is possible to find a functional balance between the raised interference level and the compensating SFN gain even in the oversized SFN network.

The simulation method presented in source [1] takes into account the SFN gain in interfered DVB-H networks, that is, when the parameters affecting on the SFN size are selected in such a way that there are sites found outside the theoretical SFN area.

The interferences can first be investigated in a general level by observing the interfered links between the sites that are exceeding the SFN limits. When the number of these potential interference sources is compared with the total number of links between all the sites, the severity of the interference in the investigated area can be noted. According to the results, there exists certain parameter sets which result in interference levels up to 100%. The higher the amount of the potential interfering links is, the more probable the network experiences quality degradation. This indicates clearly the high importance of the correct selection of the parameter values.

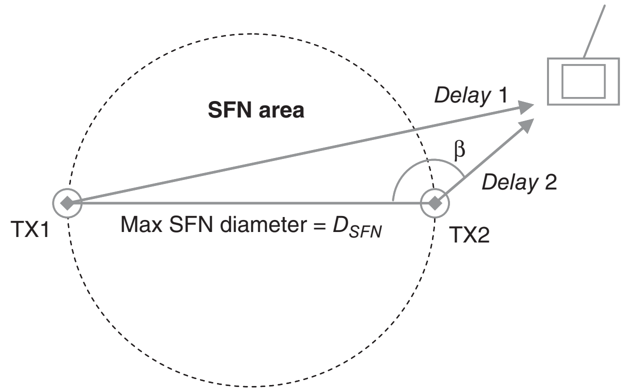

Figure 22.19 *There are no interferences when the sites are within the SFN area even if the receiver drifts outside of the SFN area. The reception within this outer zone is thus possible whenever the minimum required C/N can still be reached.*

There is also a set of parameters that results only partial interference levels. The interference level might not be constant, though, but tends to be switched on and off depending on the location of the terminal in the field. This is due to the fact that the difference between the arriving signals is the one that determinates weather the signal is interfered or not.

When all the sites are within the SFN area, that is, when the distance between of any of the sites is equal or less than the theoretical maximum SFN diameter distance D_{SFN}, there will be no interferences. This is the case also if the receiver is located outside the circle defined by D_{SFN}. This can be shown by investigating the distance between the receiver and the dominating site compared to the distance of the receiver and any of the other sites. The difference between these distances defines the effective distance D_{eff} of the signal. This distance can be investigated by making the receiver drift away of the SFN circle towards any direction as presented in Figure 22.19. The absence of the interferences can be investigated by observing the angle β of Figure 22.19. It can be seen that D_{eff} is maximum when $cos\ \beta = 1$. In that case, $D_{eff} = D_{SFN}$, that is, D_{eff} never exceeds D_{SFN} when no sites are located outside of the SFN area.

When the distance between any of the two sites using the same frequency is longer than the D_{SFN}, there will be interferences in the locations where the $D_{eff} > D_{SFN}$. The interference distribution can be observed by defining two sites in the simulator that was used in Publications III, IV and VII of source [1]. By selecting a parameter set of GI=1/4 and FFT = 2K, the D_{SFN} would be 16.8 km. The simulation of two sites in a 45 $\times$ 45 km^2 area, with *x*/*y*-coordinates of (14.9, 22.5) and (30.1, 22.5) shows that there are no interferences present at any point as the intersite distance is at the maximum allowed value of 16.8 km. The site antenna height was then set to 80 m, EIRP to +70 dBm, and the first site was moved to the coordinate (10.0, 22.5), that is, the intersite distance was extended to 20.1 km. The standard deviation of the long-term fading was maintained in 5.5 dB and the area location probability as 70%. When the maximum D_{SFN} is exceeded by 4.9 km, the interference is present in areas where $D_{eff} > D_{SFN}$ as can be seen via the simulation result shown in Figure 22.20.

Even if the interference starts to be present in those spots that $D_{eff} > D_{SFN}$, the reception is destructed, that is, useless only when the received $C/(N+I)$ is less than the minimum required C/N for the used parameter set. For this reason, we can introduce a term "destructive interference," which can be defined as an interference level at given time and location that lowers the otherwise useful C/N ratio in such a way that the reception gets below the minimum C/N requirement of the investigated parameter set.

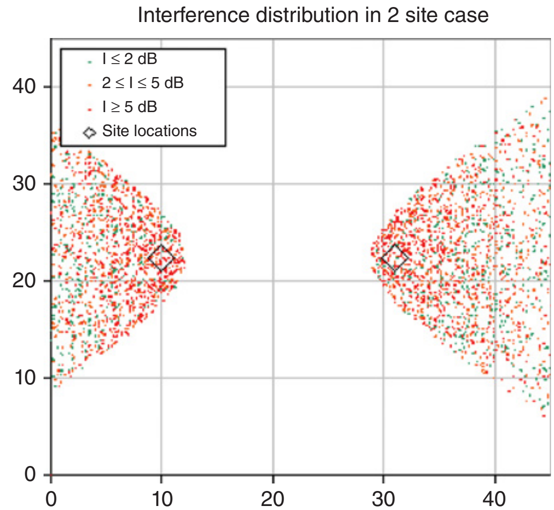

Figure 22.20 *When the distance of two sites exceed the SFN limit, there occur interferences in those areas where $D_{eff} > D_{SFN}$. Figure shows the interference zone that applies for $D_{eff} > D_{SFN}$ everywhere with the I-component greater than the noise floor. In this case, P_{TX} is +70 dBm and the site antenna height is 80 m. The total area size is 45 km × 45 km [1]. Figure by Jyrki Penttinen.*

By continuing the analysis of the previously presented simulation, Figure 22.21 shows the geographical presentation of the distribution of the destructive interferences within the investigated area where $D_{eff} > D_{SFN}$.

Nevertheless, when observing the minimum *C/N* ratio of 8.5 dB in this specific case, the effect of these destructive interference instances is not significant because their percentage of the total events is relatively low, so the remaining *C*/(*N*+*I*) ratio provide still almost as large geographical cell coverage area as the situation would be without interferences. This effect can be observed in Figure 22.22. When the power level rises to +80 dBm, the interference zone starts to be affected more as can be seen in Figure 22.23.

The geographical analysis presented above can be studied in PDF and CDF formats of *C*/*N*, *I*/*N* and *C/*(*N*+*I*) over the whole geographical area of 45 × 45 km^2 for the comparison of different parameter values. Figure 22.24 and Figure 22.25 shows the PDF and CDF, respectively. Figure 22.25 visualizes the previous geographical observation that shows still acceptable interferences in case of +70 dBm. This indicates that it is possible to construct a larger SFN cell than the theoretical D_{SFN} dictates, depending on the radio propagation conditions, site antenna height and transmitter power level.

In order to understand in more detail the impact of the parameter values on the level of interferences, a set of simulations was carried out in Publication III of source [1]. A fixed area of 100 × 100 km^2 was used as a basis for the investigations. The area was filled with partially overlapping cells according to the hexagonal model. As the number of noncorrelating interfering components increases from the above presented 2-site case, the power summing of the received interference components I_k can be applied:

$$\left[P_{RX}(I)\right]_{tot} = \sum_{k=1}^{n} \left(\left[P_{RX}(I)\right]_k\right). \tag{22.28}$$

The *C*/*I* ratio of each simulation round in investigated location is thus:

$$C/_{I}(dB) = \left[P_{RX}(C)\right]_{tot}(dBm) - \left[P_{RX}(I)\right]_{tot}(dBm). \tag{22.29}$$

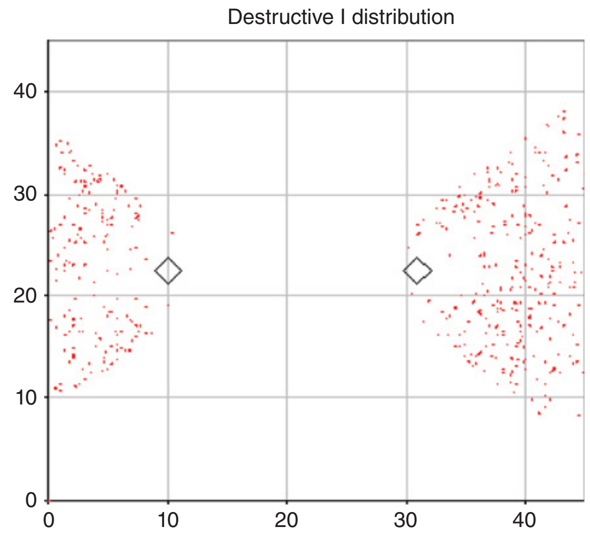

Figure 22.21 The distribution of the destructive interference. It can be seen that the useful field strength is sufficiently high close to the nearest site cell to avoid the destructive interferences, but otherwise this type of interference may occur anywhere within the interference zone [1]. Figure by Jyrki Penttinen.

The results show that by varying the antenna height and relevant parameters affecting on the SFN size, the *C/I* distribution over the whole area clearly varies. The results also indicate that when the parameter values that cause minimal interference levels in the oversized SFN network are utilized together with maximum site antenna height and power level, it is possible to construct a large DVB-H network with only single frequency by accepting a certain outage. The extension of the SFN size is possible especially with GI 1/4 and FFT mode

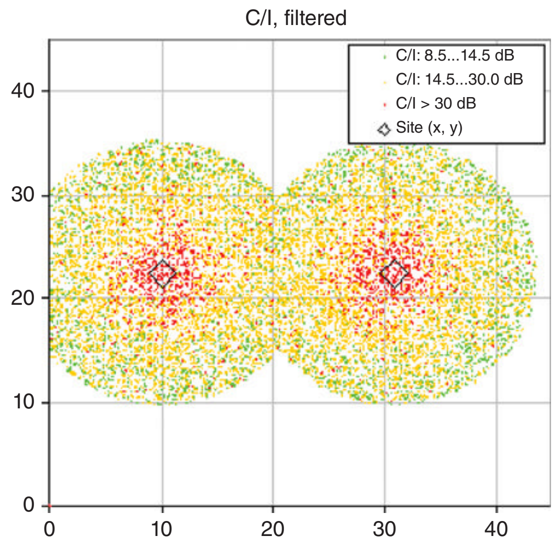

Figure 22.22 The geographical distribution of C/I when P_{TX} = +70 dBm [1]. Figure by Jyrki Penttinen.

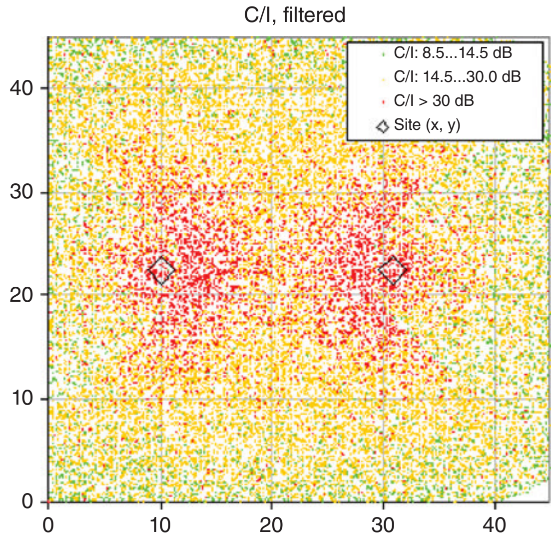

Figure 22.23 *The effect of the interference can be seen when the power level is set to +80 dBm in this specific case. The reduced C/I-level can be seen clearly behind the sites within the interference zone on the left and right hand sides [1]. Figure by Jyrki Penttinen.*

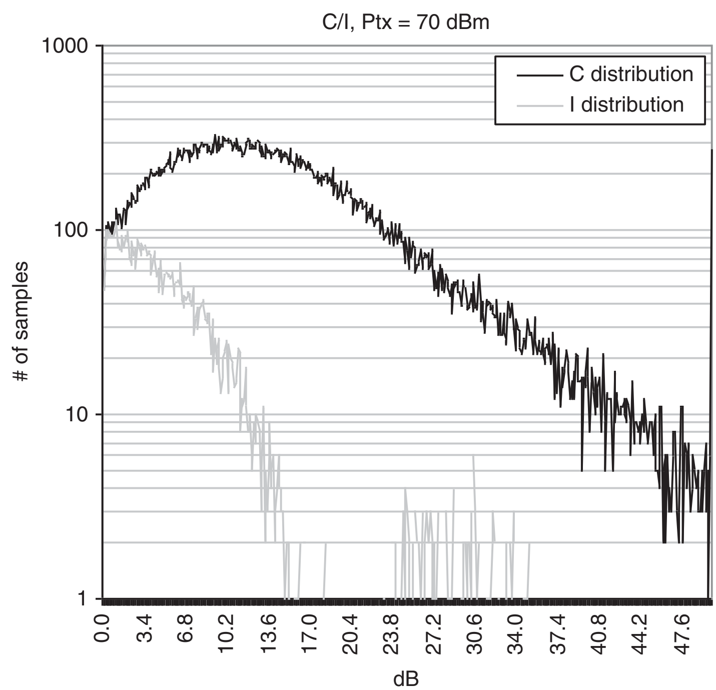

Figure 22.24 *PDF of the 2-site simulations (P_{TX} = +70 dBm).*

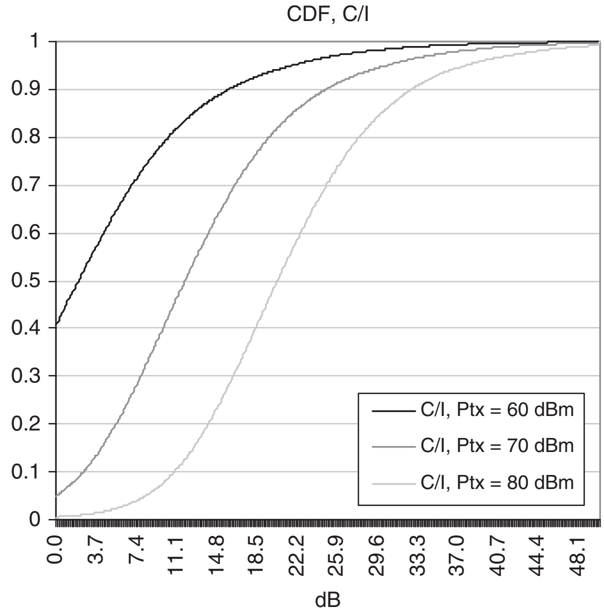

Figure 22.25 *CDF of the 2-site simulations for P_{TX} of +70 dBm. Also comparison results with P_{TX} of +60 dBm and +80 dBm sites are included.*

8K. On the other hand, when applying too demanding parameter values, the network quality might collapse. The optimization chapter describes in detail the respective simulation principle.

22.3.6.3 Conclusion

As source [1] concludes, the SFN gain over the whole investigated network area depends on the GI and FFT modes and the interference level within the SFN area. In the noninterfered SFN network, using GI 1/4 and FFT 8K in a uniform site cell layout with ideally overlapping areas according to the hexagonal model, a maximum of 6 dB of SFN gain can be expected in the best case in average within the whole investigated area. This result clarifies the SFN gain behavior. It can thus be assumed that this result is more realistic than the information of [7, p. 15] that recommends the use of 0 dB as SFN gain value in DVB-H radio link budget.

On the other hand, the definition of Ref. [7] is not very accurate. Instead, it informs that SFN gain is "the reduction of number of transmitters to cover a given area using synchronized transmitters compared to the number of independent MFN transmitters." Reference [7] concludes that "depending on the network topology, this transmitter savings may be carefully translated into an average increase of coverage. However, the SFN gain is not persistent across all points of the network, so it will not be considered in the link budget." It can be interpreted that this definition assumes that the benefit of SFN could be obtained basically in the site cell edges and that the overall gain is challenging to estimate. Publication IV of Ref. [1] shows, though, that the total gain is possible to estimate over the whole investigated area to be utilized as a generalized radio link budget value.

Based on the simulations of source [1], SFN gain of 0–6 dB could be applied to the DVB-H radio link budget depending on the parameter settings and the total number of the transmitter sites per SFN area. For the local adjustments of the SFN value of the link budget, a respective simulation method can be applied by varying the GI, FFT, code rate, power level and antenna height settings. In a realistic network, the SFN gain could be taken into account when the site locations and other parameters are selected. The SFN gain can be simulated over the planned area, and its effect can be taken into account both in noninterfered and oversized SFN cell by reducing the transmitter power levels and/or lowering the assumption of the antenna height, which gives more freedom in the optimal deployment if there are problems in using the values for the power or antenna height that the original plan indicates. If the SFN gain can be simulated in early phase of the first deployment plan, it can influence the optimal site selections.

The method presented in this chapter takes into account the combination of the gain and interference within the whole investigated network, not only non-interfered environment and in limited locations as the other found publications typically show. The case dependent SFN gain value could thus be investigated by applying the proposed simulation method.

22.4 Radiation Limitations

The installation of the DVB-H sites must comply with the local and international regulations of the EMC limits and the safety zones related to the limitations of exposure of the public to nonionizing radiation of the signals. On the other hand, the installation of DVB-H antennas should be designed in such a way that the other systems located nearby will not cause interferences to the DVB-H transmission, and vice versa.

The general limitations of exposure should be investigated via the national and international regulations. European Union dictates the limits for the member countries. As an example, the regulation in Finland covers the frequency band 0–300 GHz [38]. The format and values of the limits varies though depending on the regulator. As another example, FCC informs the limitations for cellular and broadcast environment in USA as a function of the antenna height and installation type as indicated in Ref. [39]. The allowed limit in 700 MHz band is maximally 50 kW ERP, but decreases when the antenna height is lower.

When initiating the planning process, the general maximum power limit should be taken into account. When the plan advances, also the EMC and safety zones should be estimated as accurately as the planning phase allows. When the final plan is undertaken, the site specific safety zones should be estimated.

Chapter 24 describes the principles for the dimensioning of the radiation levels and how the safety zone can be taken into account in rooftop and tower installations of DVB-H. The calculations are based on the formulas shown in Ref. [40]. The respective level should also be taken into account in the CAPEX/OPEX optimization as the maximum allowed power level may change the optimal transmitter power depending on the case. In a standard scenario, the DVB-H site consists of directional antennas either in towers or rooftops. The radiation pattern should be taken into account accordingly in vertical and horizontal layers. The radiation safety distance calculations should be carried out for different sides of the antenna.

As a summary, the regulatory limits dictates the maximum radiating power $P_{TX,reg}$ and EMC / radiation safety analysis results in the maximum radiating power in site-basis $P_{TX,safety}$, the optimal radiating power $P_{TX,opt}$ that is obtained via the CAPEX / OPEX optimization can thus move based on the formula:

$$P_{TX} = \min\left\{P_{TX,opt}, P_{TX,reg}, P_{TX,safety}\right\} \tag{22.30}$$

The selection of the final radiating power level P_{TX}, might thus require changes of either the antenna heights, transmitter power levels or both, reducing the site cell coverage. The respective analysis should be taken into account in general level in the nominal radio network planning (the correct selection of the uniform power levels) and in the detailed radio network planning (by applying the analysis in individual site basis). The final selection of P_{TX} might also require various iteration rounds in order to find the optimal point of technical and economical parameters.

The directional antenna has been studied in the radiation related Chapter 24 because this provides a practical methodology and useful examples of the expected outcome for the installation of the antenna elements, with safety distance estimations below and on the sides of the antennas. The outcome gives a realistic estimation of the radiation for the antenna installations in rooftops, in roof edge parts, where the omnidirectional element would not be a logical solution. Nevertheless, the method is also valid for the omniradiating elements.

22.5 Cost Prediction and Optimization

In complete network planning, the cost effect needs to be taken into account in every phase. In the initial phase of the network planning, a rough estimation of the average site cost as well as the total cost of the area of interest is sufficient. The total channel capacity requirement depends on the number of the channels and their respective quality level. The demand for services and thus additional capacity can be expected to grow as a function of time as stated in Ref. [41].

The detailed plan should already contain more in-depth analysis of the network costs as for the CAPEX and OPEX. In this phase, it is possible to take into account different solutions, including the transmitter type, transmission, preferable and available antenna heights, antenna types, site solutions in general and so on. It is important to balance the costs with the technical solutions, that is, the hardware, software and related technical quality, coverage and capacity of the network. The cost effect is such an important item that it should be taken into account as preferable in an iterative way, that is, in order to seek the optimal point of technical and commercial balance, possibly several planning scenarios should be carried out.

It can be assumed that uniform parameter values may be applied for the site costs in the initial planning, because the final site locations or site specific radio parameters are not necessarily confirmed yet. In the later phase, during the detailed planning, site-specific assumptions should be utilized instead for the more in-depth cost optimization.

The case examples of the DVB-H deployment and operation cost optimization are studied by using the hexagonal network layout. In practice, the sites do have variable antenna heights and power levels as well as nonuniform locations. The presented case examples are shown for the uniform comparison in order to select the most feasible parameter combination from the relative analysis, but the methodology can also be applied in the real environment with site specific parameter values. In that case, the model should be applied by selecting the already existing sites – that can be from both broadcast and mobile networks – and by identifying a set of additional site candidates with a "best effort" assumption for their probable, individual parameter values. As the site locations are fixed in that case in a nonuniform way, the coverage analysis should be carried out by adjusting the power levels, antenna heights, and respective costs of each site separately iteratively way until each case produces the wanted coverage quality level.

The whole chain of the DVB-H network, including the IPDC part (core) and radio networks, includes many business parties [42]. As each one of the delivery chain representatives has their costs due to the investments, they might face the cost optimization issues. It is thus essential to take into account the cost effect of the technical solutions of the DVB-H network deployment as deeply as possible as there might be several parameter values producing a zero-point in the derivative of the cost as a function of the selected parameter values.

22.5.1 Cost Optimization in Noninterfered Network

Source [1] shows a methodology for the seeking for an optimal parameter set for the OPEX and the CAPEX of the network. It is based on the adjustment of the essential variables, that is, antenna heights and transmitter power levels, in order to vary the size of the single site cells of the network assuming there are no interferences found in the planned area, that is, either MFN or noninterfered SFN is used.

The variables for the complete technoeconomic network optimization include the total cost of the site, radius of the site cell and radiating power level vs. the respective cost of the equipment. There are various sources of information as well as tools in order to find the main aspects in the physical environment, including the geographical distribution of the population and economical levels, data bases of the site locations (existing towers, the height of the antennas), digital maps with the area types and respective propagation models and so on. In detailed network optimization, unlike in the nominal plan, uniform power levels would not provide the best performance due to the practical differences of the areas and site solutions, so the power levels and antenna heights should be planned on a site-by-site basis in this phase.

In the initial phase of the planning, the area type can be assumed as uniform, or that there are only few different area types each one being uniform as a cluster class. As an example, the area inside of ring road of a large city could be assumed as dense city area type, whilst outside the ring road, a suburban cluster could be applied.

The first step of the cost estimation is to identify all relevant items that do have cost effect on the network implementation as presented in Figure 22.26. Then, the CAPEX and OPEX of a single site can be estimated, with a respective radio coverage area in that cluster type. In order to have a reliable estimate of the cost effect of the whole network, a sufficiently large area can be selected and filled in with the sites according to the coverage area of each transmitter. In the nominal planning phase, a hexagonal model can be utilized whilst in the detailed planning, site-by-site adjustment and real map data should be applied. Then, by varying

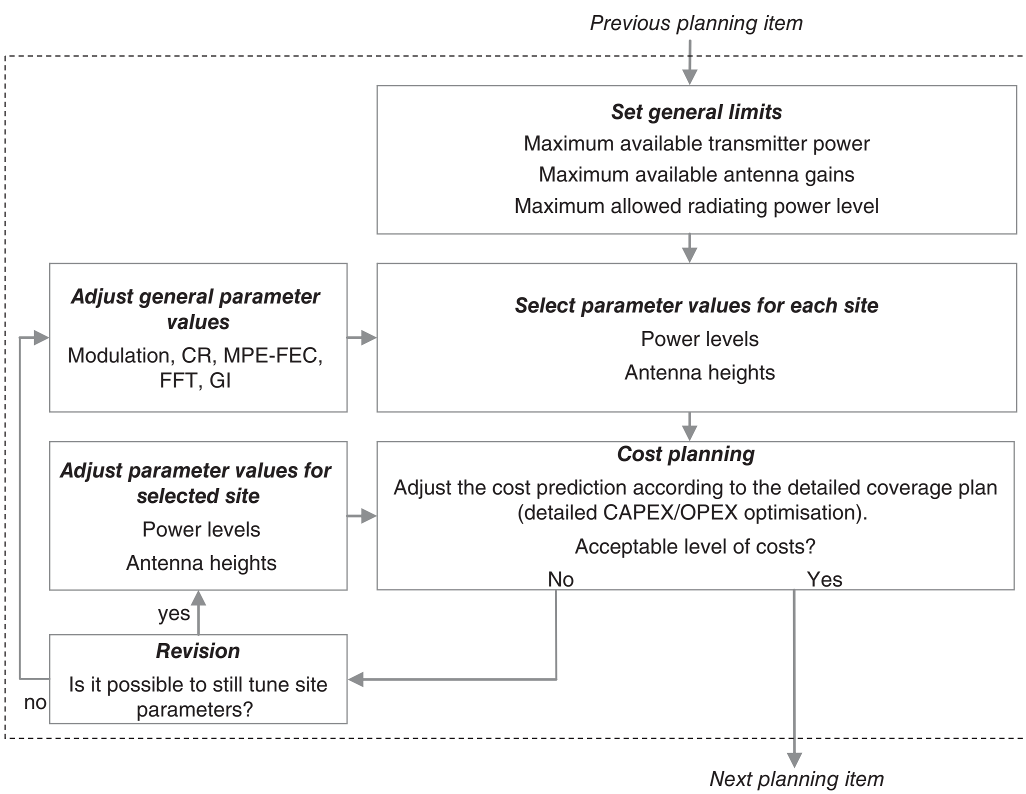

Figure 22.26 *Cost optimization process in noninterfered DVB-H network.*

the essential parameters like antenna heights and gains, transmitter power levels and so on, this process can be repeated. The combined CAPEX and OPEX curves now indicate what would be optimal solution as a function of time.

The CAPEX represents the initial one-time cost of the site. It should contain cost information about the investigated transmitter type, antenna system with related jumpers, connectors, feeders and power splitters, other material for the mounting and so on, the work for site acquisition, legal and technical preparation, and finally the work for actual installation and commissioning of the site. The OPEX indicates the costs that are generated during the time after the initial installation of the site. The most important long-term cost items are related to transmission type, maintenance of equipment, possible tower and site rental agreements, and electrical power consumption.

The analysis can be done by applying an estimate for the transmitter costs as a function of the power level. The cost information was obtained in a snapshot way from the typical commercial transmitter models and the values were normalized by taking the cost of the 500 W-transmitter as a reference. It is obvious that this utilized cost information can vary notably depending on the case, including the effect of possible discounts for the purchase of high amount of transmitters. Nevertheless, it indicates the cost behavior depending on the variations of the power level. As a result, it is possible to create a power vs. cost-dependency graph, that is, what is the cost of single watt as a function of the transmitter type.

Next, the CAPEX and OPEX per site can be presented in graphical format by normalizing the values. In Publication I of Ref. [1], the transmitter type producing 500 W is selected as a reference. For the CAPEX, the higher power transmitters are logically more expensive, they require more installation work, and the respective feeders are more expensive. As for the OPEX, the transmission and site rent are the major cost items. Assuming that the power consumption is about six times more than the produced output power, the energy consumption can be important for the highest power transmitter models. Source [1] gives a rough estimate of the energy consumption which may represent about 25% of the total operating costs of a single site under typical circumstances.

For the coverage estimation with the Okumura-Hata propagation model, Publication I of Ref. [1] assumes that QPSK is used, area location probability 90%, shadowing margin 5.5 dB, frequency 700 MHz, transmitter antenna gain 13 dBi, receiver antenna gain −7 dBi, CR 1/2, and MPE-FEC 3/4. This results in a received power level requirement of −87 dBm.

For each transmitter power level case, the feeder was selected according to the output power requirements, that is, less output power requires thinner feeder, which is more economical and easier to install, but on the other hand its loss is higher reducing the coverage area. As there is clear interdependency, this item requires additional iteration rounds in order to find the optimal balance between the total cost and performance, that is, the task is to select the feeder that complies with the maximum power requirement and that keeps the total cost on minimum level by balancing the cable type (attenuation) and its cost.

In the analysis presented in source [1], the cables were selected in such a way that power levels up to 3400 W use 1-5/8" feeder (with 1.9 dB attenuation per 100 m) and power classes 4700 W and 9000 W used 3" feeders (with 1.5 dB attenuation per 100 m but resulting about 30% more costly material than the previous one). The cable cost for different cable models can vary several tens of percents. The portion of the cable cost reduces to only some percents when the overall site cost gets higher, that is, when the transmitter power increases.

Even if the hexagonal model used in the large network area analysis is theoretical, the model reflects the reality especially for the comparison of cases as the overlapping areas can be utilized in the practical network for handover purposes of the multi frequency network, or for producing the SFN gain in the single frequency network. The presented method of creating a uniform site distribution for each case with same parameter values and ideally overlapping coverage areas of each site cell is thus functional for the relative investigation of the cost effect. As Publication I of Ref. [1] concluded, for example, the operating cost effect of transmission and electrical power might be considerable and thus a significant OPEX item.

22.5.2 Cost Optimization in Interfered SFN Network

When the oversized SFN is applied, the cost optimization method presented previously can still be used, but a feedback loop should be added to the process as proposed in Figure 22.27.

The SFN simulator method of source Ref. [1] can be utilized in the feedback loop. The iteration level of the cost optimization thus rises, but in the in-depth network planning and optimization it is recommendable to carry out as thorough cost optimization as possible as it might have remarkable effects on the final savings in the network deployment and operating phases. The CAPEX/OPEX analysis can be presented in an analytical format in the following way.

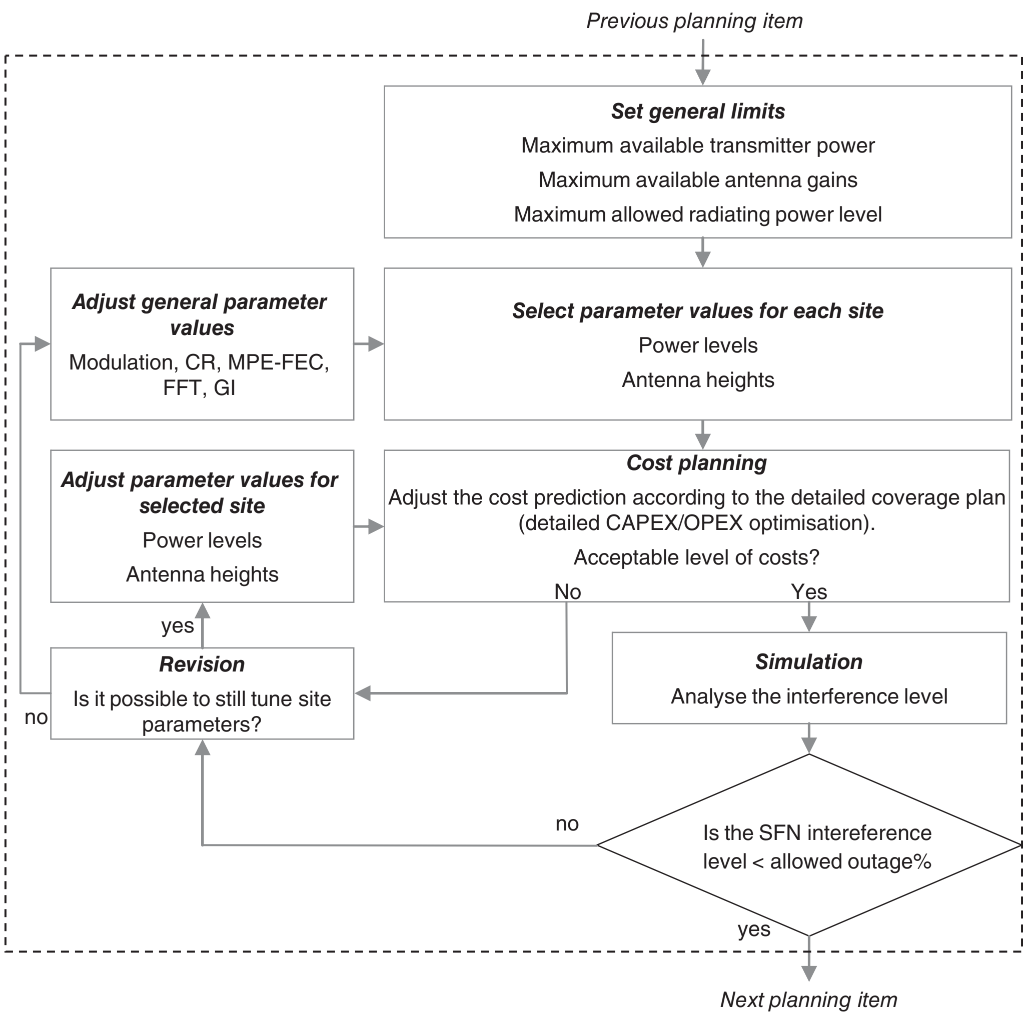

Figure 22.27 *Cost optimization in interfered DVB-H network.*

22.5.2.1 Path Loss

The initial task is to define the basic loss L (dB). The L is obtained from the link budget as presented in Table 22.2. As an example, for the QPSK, CR 1/2 and MPE-FEC 2/3, the required C/N is 11.5 dB according to [2]. With this combination, the maximum allowed path loss in outdoor is 144.2 dB, when the frequency f is 600 MHz, receiver's noise figure F is 5 dB, transmitter output power P_{TX} is 2400 W, cable and connector loss L_{cc} is 3 dB, power splitter loss L_{ps} is 3 dB, transmitter antenna gain G_{TX} is 13.1 dBi, receiver antenna gain –8.4 dBi, and the location variation for 95% area probability is 7.0 dB.

In the cost analysis, the variables shall be the cable and connector loss L_{cc}, the transmitter power P_{TX} and the height of the transmitter antenna h_{BS}. The receiver antenna can be assumed to be fixed to 1.5 meters. Following the principle of the link budget of Table 22.2, the initial L is estimated as shown in the link budget analysis based on Table 22.2.

22.5.2.2 Cell Radius

When the maximum allowed path loss is known for the investigated radio parameter values, the next step of the CAPEX/OPEX analysis is to estimate the single cell radius. A uniform analysis can be utilized in the first phase of the analysis. This gives an overall estimate about the situation by presenting the global values in an ideally overlapping site setup which can be presented by the hexagonal layout model. The drawback of this method is that in practice, it is not possible to build the sites according to the ideal distribution the model suggests. Nevertheless, the model provides a functional estimate about the feasible parameter value ranges as for the antenna heights and transmitter power levels. The method presented in Publication I of Ref. [1] can also be applied in a practical DVB-H network layout, for example, by extending the coverage and interference analysis of Publication VII of Ref. [1] with the CAPEX/OPEX module.

This uniform cost analysis is based on the estimate of a single cell radius. In the simplest format, the free path loss calculation could be applied with a rough estimate of the attenuation factor that represents different area types. Throughout this chapter, the basic Okumura-Hata model [18] or ITU-R p.1546.3 [19] are used as they provide a realistic estimate for the path loss, and they are commonly utilized in the coverage predictions. If the basic Okumura-Hata model is applied for the CAPEX/OPEX analysis, the cell radius d (km) can be estimated as an example in medium and small city environment.

$$d = 10^{\left(\frac{L-[69.55+26.16\log(f)-13.82\lg(h_{BS})-((1.1\log f-0.7)h_m-(1.56\log f-0.8)]}{44.9-6.55\log(h_{BS})}\right)}, \tag{22.31}$$

where L is the outcome of the equation above in dB, f is the frequency (MHz), h_{bs} is the transmitter antenna height (m), and h_m is the mobile antenna height (m).

As soon as the cell radius in known, the next task is to model the cost items in order to understand the expenses of the parameter values. The items can be divided to the initial investments (CAPEX), and to the longer term yearly costs (OPEX).

22.5.2.3 CAPEX Items

The initial network deployment costs include various cost items that impact on the CAPEX. Table 22.13 summarizes the most important variables for a single site.

As can be seen from Table 22.13, there are various relevant items affecting on the CAPEX. The cost variables that are shown in the table have partial interdependencies between the transmitter power levels,

Table 22.13 *The most relevant CAPEX items and relations*

Variable	OPEX item	Observations
C_1^C	Transmitter	The transmitter cost depends on the power level P_{TX} as presented in Figure 22.5, and thus on the cell radius. The complexity of the transmitter has impact on the cost.
C_2^C	Antenna system	The transmitter power level dictates the minimum power level that the antenna elements should support, with a practical margin. The elements can be omnidirectional or directional.
C_3^C	Antenna feeder	The antenna feeder must support the transmitter power level, and a safety margin should be taken into account. The antenna feeder has an impact on the L as the cable and connector losses decrease the cell radius. Higher power requires thicker antenna feeder, which is more expensive, but with also lower losses.
C_4^C	Installation, antenna system	The installation cost depends on the height and weight of the antennas, height of the cables, as well as the length of the cable.
C_5^C	Power splitter	If various directional antennas are used, the power splitter is utilized. The cost depends on the power levels and antenna ports. The power splitter has a direct impact on the cell radius.
C_6^C	Cable brackets	The cable mounting requires brackets. The amount and cost depend on the antenna feeder type and antenna height.
C_7^C	Installation	Personnel's compensation of the work.
C_8^C	Other installation costs	Special fees, for example, due to the helicopter installation and so on.
C_9^C	Site acquisition	The identification of the location and acquiring of the site, including personnel's compensation.
C_{10}^C	Planning, drawings	The preparations of the site, personnel's compensation.
C_{11}^C	Miscellaneous costs	Any other sufficiently significant site related cost.
C_{12}^C	Tower / site building	If the site and/or tower are purchased, this item triggers a one-time cost effect. It should be noted that the building costs might be exponential as a function of the tower height.

antenna height, OFDM parameter values and cell radius. The complete CAPEX for each investigated option is:

$$C_{tot}^C = \sum_{i=1}^{12} C_i^C \tag{22.32}$$

Antenna Feeder Selection When the transmitter power level is selected for the investigation, the antenna feeder type selection depends on the respective maximum possible transmitter power level. Table 22.14 summarizes a set of example values that indicates the relationship between the cable type and power requirements.

As Table 22.14 indicates, the most cost-effective solution (balance of the feeder cost and loss) might be possible to obtain even if the cable type is clearly overdimensioned as for the maximum power because certain cables might increase clearly the cell radius due to the lower loss.

As an example, 1/2″ cable with 100 m antenna height produces about 7.5 dB loss whilst the 7/8″ produces 3.5 dB. This 4 dB difference is significant in the radio link budget, as the same effect can be achieved by, for example, more than doubling the transmitter output power level. It should be noted that the tower installation for a certain antenna height requires typically additional cable length for the equipment room.

Table 22.14 Cable types and main characteristics utilized in the CAPEX/OPEX analysis

Type	Attenuation at 500 MHz / 100 m	P (const) kW / 500 MHz	Attenuation at 700 MHz / 100 m	P (const) kW / 700 MHz
1/2″	6.31	0.98	7.58	0.82
7/8″	2.70	3.38	3.44	2.65
1-5/8″	1.59	6.93	1.91	5.77
2-1/4″	1.31	9.89	1.58	8.21
3″	1.04	18.4	1.46	13.1

By simplifying and assuming that only one piece of cable is utilized per site, the antenna feeder cost item is selected from the list of available types, for example, in the following way in the presented example:

$$C_3^C = \begin{cases} c_{1/2''}/m, P_{TX} < P_{\max}^{1/2''} = 0.82kW - \varepsilon \\ c_{7/8''}/m, P_{TX} < P_{\max}^{7/8''} = 2.65kW - \varepsilon \\ c_{1-5/8''}/m, P_{TX} < P_{\max}^{1-5/8''} = 5.77kW - \varepsilon \\ c_{2-1/4''}/m, P_{TX} < P_{\max}^{2-1/4''} = 8.21kW - \varepsilon \\ c_{3''}/m, P_{TX} < P_{\max}^{3''} = 13.09kW - \varepsilon \end{cases} \cdot \tag{22.33}$$

It can be decided that the additional power safety margin ε is 10% of the maximum supported power level P_{max} of each antenna feeder type. There are no technical restrictions in the use of any of the antenna feeder type whilst the power limits are taken into account. If the effective output power of the transmitter is 2.16 kW (2.4 kW with TX filter loss of 10%), the feeder type of 7/8″ supports 2.65 kW – 0.1·2.65 kW = 2.39 kW, which thus complies with the realistic power of the 2.4 kW transmitter type. In this case, the antenna feeder and related cost C_3^C could thus be selected from $\{c_{7/8,''}\ c_{1\text{-}5/8''}, c_{2\text{-}1/4''}, c_{3''}\}$. The cost c of each antenna feeder should now be known via the available sources. In this analysis, the assumption is that the cost is constant per meter, although in practice, there might be additional discounts due to the volume purchase. On the other hand, it should be noted that the feeder diameter has an effect on the installation cost C_4^C due to the more difficult handling of the material and additional weight of the feeder.

When the transmitter power level is selected, as well as the transmitter antenna height, the cable type or different options for the cable type can be selected, and the respective cable and connector attenuation L_{cc} as well as total cost can be calculated.

The cable is clearly a CAPEX item. It is installed only once and is almost maintenance free, making the related OPEX practically minimal. When estimating the cost effect of different cable types, the balance can be found by evaluating the cost of material and work per cable type (heavier cable requires more work) as well as the effects of different cable losses (thicker cable produces lower losses). It should also be noted that the attenuation is frequency dependent.

Table 22.15 summarizes a case used in source [1] by presenting the site cell radius obtained with different cable types and transmitter power levels. The assumption is to use all the antennas in 100 m tower. This table indicates that the cable types of {1-5/8″, 2-1/4″, 3″} provide similar coverage areas, so the lowest cost type that is supported by each transmitter power level is an obvious selection in order to optimize the cost. Nevertheless, the table shows that there is a clear difference between {1/2″, 7/8″, 1-5/8″}, the thinner cable producing clearly lower coverage areas.

Table 22.15 An example of the site cell radius obtained by varying the cable type and transmitter power levels. N/A is shown if the cable is not suitable for the respective power level

	Cell range (km) for cable type of:				
P_{TX} / W	1/2″	7/8″	1-5/8″	2-1/4″	3″
100	3.7	5.6	6.4	6.6	6.7
200	4.6	6.9	8.0	8.3	8.4
500	6.2	9.2	10.7	11.0	11.1
750	7.0	10.5	12.1	12.5	12.7
1500	N/A	13.0	15.1	15.6	15.7
2800	N/A	N/A	18.4	18.9	19.2
3400	N/A	N/A	19.5	19.9	20.0
4700	N/A	N/A	20.4	20.6	20.7
9000	N/A	N/A	N/A	N/A	22.6

By taking this into account in the cost analysis, that is, by calculating the final cost of the network if thicker cable is utilized in these lower power cases, the outcome could be that the thicker cable, even if the resulting CAPEX is higher than for the thin cables, can provide more cost-efficient OPEX.

22.5.2.4 OPEX Items

Table 22.16 summarizes typical OPEX items of Mobile TV network.

The electrical power consumption P_c of the transmitters can be estimated in a general level as a function of the transmitter output power level, unless there is no more accurate estimate available:

$$P_c = 6 \cdot P_{TX}. \tag{22.34}$$

Table 22.16 The most relevant OPEX items and relations

Variable	OPEX item	Observations
C_1^O	Electricity	The electricity consumption depends mainly on the transmitter type (power level) as well as the power consumption of the related other site elements like modulator, site cooling system, routers and so on.
C_2^O	Maintenance	The maintenance includes principally the transmitter maintenance that depends on the complexity of the transmitter. As an example, air cooled requires less site visits than oil cooled, which in turn also requires oil changes.
C_3^O	Transmission	The transmission can be either terrestrial or via satellite. Both have cost effect, terrestrial triggering site specific costs whilst the costs of a single satellite usage can be divided between all the respective sites.
C_4^O	Tower / site rent	If no own site or tower has been obtained, the rental costs of the equipment shelter and / or antenna system usage are added to the OPEX.
C_5^O	Other expenses	Any other item triggering regular costs like average level of the general site maintenance.

The yearly cost of the transmitter with a power level of P_{TX} is thus:

$$C_1^O = 6 \cdot P_{TX} \cdot c_e \cdot 24h \cdot 365/1000, \tag{22.35}$$

where c_e is the cost of electricity per kWh. Equally, the other OPEX items should be estimated in absolute values in yearly basis, which results in the total OPEX per year:

$$C_{tot/year}^O = \sum_{i=1}^{5} C_i^O. \tag{22.36}$$

The yearly costs might vary over the time, but in this level analysis it is sufficient to estimate an average cost per year.

22.5.2.5 Combined CAPEX/OPEX of the Whole Network

When the CAPEX and OPEX are calculated for all the investigated options as a function of transmitter power levels (transmitter types), antenna heights and OFDM parameters (which results in the balance between the capacity and coverage), the combined network cost can be calculated.

The single cell radius d, or the distance between the site and calculated cell edge with the given area location probability, indicates how many partially overlapping sites can be located within a certain area, which can be the whole planned network or part of that. By taking into account the overlapping share of the hexagonal model (as presented in the appendix of Ref. [1]), the number of the site cells N_{cells} is:

$$N_{cells} = \frac{A_{tot}}{A_{cell}} \cdot 0.827 = \frac{A_{tot}}{\pi d_{cell}^2} \cdot 0.827, \tag{22.37}$$

where A_{tot} is the whole investigated area (km^2) and A_{cell} is the single site cell coverage area (km^2). This information is the key for the CAPEX/OPEX analysis as it provides the possibility to compare the efficiency of the selected parameters in order to minimize the cost when filling the area with the full coverage area.

The total cost of the investigated part of the network is thus:

$$C_{tot} = N_{cells} \cdot \left(C_{tot}^C + C_{tot/year}^O \cdot N_y \right), \tag{22.38}$$

where N_y is the number of the operating years with $\{N_y \in \Re | 0 < N_y < t_{max}\}$, where t_{max} is the maximum operating years of the network. The aim is now to calculate as many options as are logical to investigate, by following the planning process from the beginning. The first task is to fix the maximum offered capacity, which results in a certain limited set of OFDM parameter values for the modulation scheme, CR and MPE-FEC. The analysis can also be limited by selecting realistic general values for the possible maximum antenna heights in the investigated area. Also the regulatory radiation values might be needed to be taken into account when limiting the antenna heights and maximum possible radiating power levels.

The most logical set of parameters can now be tested by calculating the maximum path loss L and respective cell radius d, which gives the cost estimate for each case. The results of source [1] show that there is an optimal point in the combined CAPEX and OPEX curves when varying the transmitter powers. In the presented case analysis of Publication I of Ref. [1], the optimal transmitter power class is found in mid-range models. A further analysis shows that the optimal point varies during the time, higher power transmitters becoming more cost-efficient solution in the longer run.

This outcome of the optimal power level which is not necessarily the highest possible one is an interesting deviation to the typically presented assumption which indicates that the optimization via the reducing of sites by increasing as high masts and as high output powers as possible is assumed to lead directly to the reduction of costs as presented, for example, in Ref. [43].

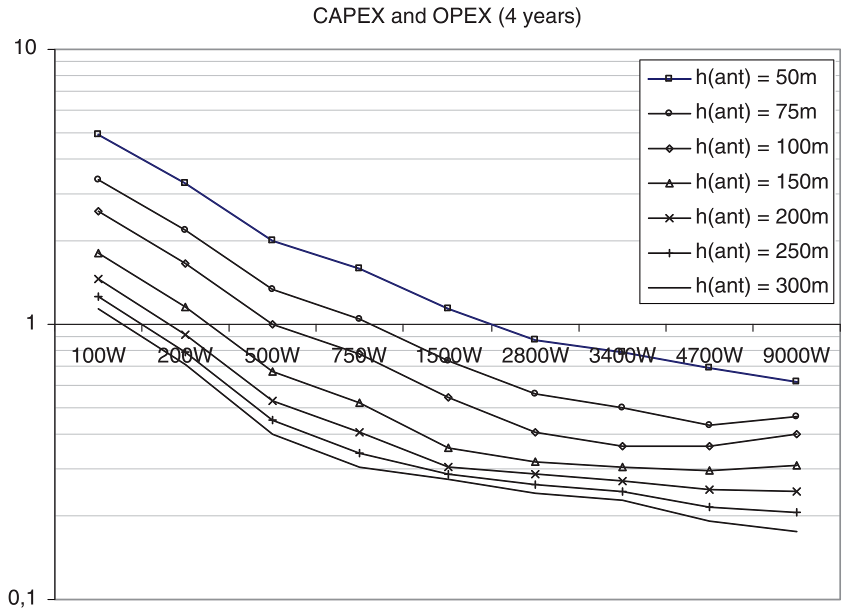

Figure 22.28 *The cost effect of the DVB-H network when the antenna height is varied. This case includes relatively high transmission costs.*

There are many interdependencies between the variables which require various iteration rounds if the optimal set of parameter values are investigated thoroughly. As an example, by keeping the variables the same, but varying the antenna height in a suburban area, Figure 22.28 can be obtained. The presented values are normalized to 500W-transmitter case and for the antenna height h_{ant} of 100 meters. The analysis shows the combined CAPEX and OPEX after 4 years of operation.

Figure 22.28 indicates that the higher the antenna is located the lower the cost of the network is. This is logical outcome because the coverage areas grow as a function of the antenna height. It is interesting to note though, that in this specific case and in a range of about 75…150 meters of antenna height, the optimal transmitter power level is in mid-range models whereas in lowest and highest antenna locations the high-power solution is the optimal in each antenna height category.

For the high antenna cases this is understandable as the number of high-power transmitters is relatively low in order to achieve the same coverage area as with the lower power cases. The explanation for the low antenna height behavior is that the number of the high-power sites in that specific situation is low enough to favor the high-power transmitters even if their relative cost of power consumption is clearly higher.

In the above-mentioned examples, the transmission cost has been assumed to be based on leased lines with relatively high cost (in order of 150 000 euros per year per site). The tower rental and site cost has been assumed to be in order of 20 000 euros per year. If the transmission cost can be reduced considerably, for example, down to 15 000 euros per year per site, the result changes. This is due to the fact that the relative proportion of the electricity consumption gets much higher in the total OPEX. Figure 22.29 shows a related analysis, with all the other parameters being the same as in case of Figure 22.28.

The reference for the normalized *y*-values (costs) has been kept the same in Figure 22.29 as in Figure 22.28, that is, the analysis is based on the 500 W transmitter case and for 100 meters of antenna height. As may be noted, the relative cost of the network is considerably lower with low transmission costs. In this case, the

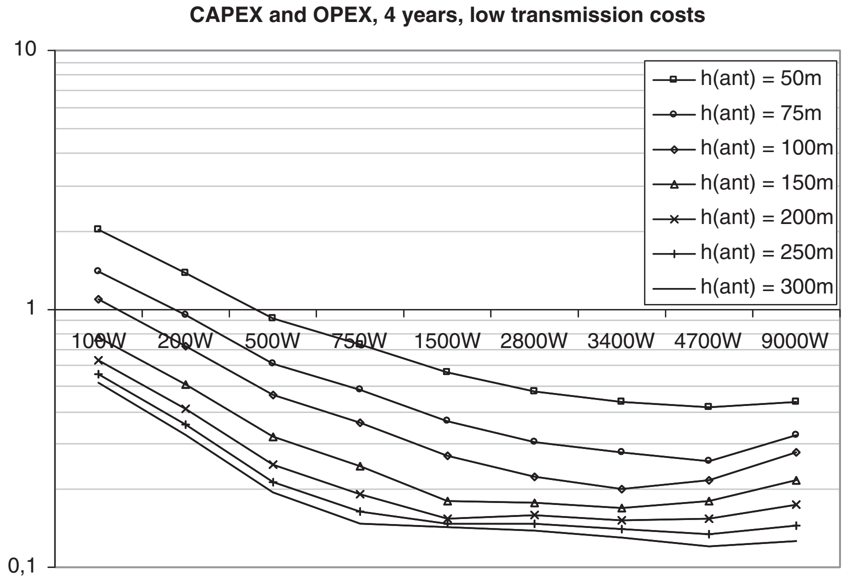

Figure 22.29 *An example of the cost effect as a function of the antenna height when the transmission costs are assumed to be low.*

highest power case of 9000 W results now in about 49% of the OPEX out of the total operating expenses. This effect can be seen in Figure 22.29 as the optimal point per antenna height category is clearly indicating the mid-range power transmitters.

It is worth noting that the above-mentioned analysis results in optimal transmitter power level and antenna height values, which in many cases indicates the mid-level range.

22.5.2.6 Cost Optimization in Varying Sites

The method shown in this chapter for the CAPEX/OPEX optimization is based on the uniform parameter values over the whole investigated network. Nevertheless, the method can also be applied for the varying site configurations. Furthermore, the analysis can be combined to the coverage area study, as well as to the SFN gain and SFN interference balancing study by varying the radio parameters and by calculating the respective costs either in uniform case or site-dependent parameter values.

22.5.2.7 Conclusions

As a conclusion of the cost optimization, it is important to identify all the relevant cost items case-by-case and carry out the related analysis. If the DVB-H transmitter antennas can be located relatively high, it helps in the optimization of the coverage areas. There are limits though, the increased power consumption being one of the critical items in the case of the highest power level transmitters. As Figure 22.28 and Figure 22.29 shows, the antenna height does have a considerable effect on CAPEX and OPEX and the optimal cost point can be found.

On the other hand, the optimal height might depend also on the other aspects like on the fine-tuning of the maximum size of the single frequency network, that is, when sufficiently low EIRP levels are used, the theoretical SFN diameter is not necessarily a limit in practice and the single frequency can be used in large areas. An optimal solution would possibly include part of the antennas installed as high as possible to broadcast towers and others in telecom towers and rooftops due to the easier access.

Furthermore, there might be possibilities to adopt the transmit diversity for the DVB-H networks. It would improve the reception and QoS in the site cell edges and the outage areas within the site cell. The technique provides thus a robust reception with improved QoS, and can reduce the network costs by lowering the transmit power and the number of infrastructure elements as has been noted in Ref. [44].

Reference [43] contains DVB-H specific technoeconomic considerations. Other highly relevant references for the comparison of the presented method and case results are Refs. [45–47]. They present analysis for the optimizing of the network deployment and operation costs, for example, by comparing broadcast and mobile network towers and by combining DVB-H and other cellular systems infrastructure. Unfortunately, none of these considers a DVB-H network deployment scenario, making the final comparison of the results challenging.

As a conclusion of the cost optimization part of this chapter, a sufficiently detailed method was developed to be taken into account in the typical DVB-H deployment. This chapter presents a complete set of CAPEX and OPEX items which were investigated as a function of transmitter antenna height and power levels, taking into account all the relevant deployment aspects of the equipment and planning environment in a more detailed way than was found in other related references.

References

1. Penttinen, J. (2011) The planning and optimisation of DVB-H radio network. Doctorate thesis. Aalto University, 18 February 2011.
2. Digital Video Broadcasting (DVB); DVB-H Implementation Guidelines. ETSI TR 102 377 V1.3.1 (2009-03). 114 pp.
3. Reimers, U. (2005) *DVB – The Family of International Standards for Digital VideoBroadcasting*, 2nd edn, Springler-Verlag, Berlin Heiderberg, 408 p., ISBN 3-540-43545-X.
4. Götz, R. (2002) Supporting Network Planning Tools II. Presentation, LS telcom AG, 2002. 38 pp.
5. van de Beek, J.-J. (1998) Synchronization and Channel Estimation in OFDM Systems. PhD thesis, Lulea University of Technology, Division of Signal Processing, Lulea, Sweden, 158 pp.
6. Milanesio, D. (ed.) (2006) Wing TV. Services to Wireless, Integrated, Nomadic, GPRS, UMTS & TV handheld terminals. D11 – Wing TV Network Issues. May 2006. 140 pp.
7. Link Budgets. Update 02/2009. BMCOFORUM, February 2009. 178 pp.
8. Joseph, W., Tanghe, E., Pareit, D. and Martens, L. (2007) Building penetration measurements for indoor coverage prediction of DVB-H systems. 1-4244-0878-4/07. IEEE 2007, pp. 3005–8.
9. Faria, G., Henriksson, J.A., Stare, E. and Talmola, P. (2006) DVB-H: digital broadcast services to handheld devices. *Proceedings of the IEEE*, **94**(1), 194–209.
10. Single Frequency Network Concepts for cellular OFDM Radio Systems. Wolfgang Zirwas. Siemens AG, München. OFDM workshop, Hamburg, Germany, Sept. 2000. 4 pp.
11. Kim, J.C. and Kim, J.Y. (2006) Single Frequency Network Design of DVB-H (Digital Video Broadcasting – Handheld) System. ICACT2006, February 20–22, 2006, pp. 1595–8.
12. Plets, D., Joseph, W., Verloock, L. *et al.* (2008) Influence of reception condition, MPE-FEC rate and modulation scheme on performance of DVB-H. *IEEE Transactions on Broadcasting*, **54**(3), 590–8.
13. Himmanen, H. (2008) Studies on Channel Models and Channel Characteristics for Mobile Broadcasting. Department of Information Technology, University of Turku, Turcu Centre for Computer Science, Finland, 9 pp.
14. Gómez-Baquero, D. and Bria, A. (2007) Forward Error Correction for File Delivery in DVB-H. Polytechnic University of Valencia, Spain, and Royal Institute of Technology, Sweden. Vehicular Technology Conference VTC 2007, Spring. Dublin, 22–5 April 2007, 5 pp.

15. DVB-H Implementation Guidelines. Draft TR 102 377 V1.2.2 (2006-03). European Broadcasting Union, 108 pp.
16. Backman, W. (2004) Signal Level Interpolation for Coverage Area Prediction. 0-7803-8521-7/04. IEEE 2004, pp. 67–71.
17. Jeong, M. and Lee, B. (2001) Comparison between Path-Loss Prediction Models for Wireless Telecommunication System Design. 0-7803-7070-8/01. IEEE 2001, pp. 186–9.
18. Hata, M. (1980) Empirical formula for propagation loss in land mobile radio services. *IEEE Transactions on Vehicular Technology*, **VT-29**(3), 9 p.
19. Recommendation ITU-R P.1546-3. Method for point-to-area predictions for terrestrial services in the frequency range 30 MHz to 3000 MHz. 2007, 57 pp.
20. Östlin, E., Suzuki, H. and Zepernick, H.-J. (2006) Evaluation of the Propagation Model Recommendation ITU-R P-1546 for Mobile Services in Rural Australia. 63rd IEEE Vehicular Technology Conference, Melbourne, Australia, May 2006.
21. Tunc, C.A., Altintas, A. and Ertürk, V.B. (2005) Examination of existent propagation models over large inhomogeneous terrain profiles using fast integral equation solution. *IEEE Transactions on Antennas and Propagation*, **53**(9), 3080–3.
22. Chubukjian, A., Nappert, H., Mehta, K. and Legris, A. (2000) An example of efficient spectrum management: Army tactical radio operations in broadcasting bands. Industry Canada, Ottawa, Ontario, Canada. 0-7803-6521-6. IEEE, 2000, pp. 15–18.
23. Plets, D., Verloock, L., Joseph, W. and Martens, L. (2009) Influence of DVB-H Network Deployment on Total RF Exposure. Proceedings of the 20th International Zurich Symposium on EMC, Zurich 2009, pp. 329–32.
24. Haddad, K. (2007) DVB-H in Denmark. Technical and Economic aspects. Master's Thesis. Technical University of Denmark (DTU), Center for Information & Communication Technologies (CICT), September 10, 2007, 98 pp.
25. BPN – 066, issue 1.0, guide of SFN Frequency Planning and Network Implementation with regard to T-DAB and DVB-T, EBU – technical department 2005, 124 pp.
26. Aparicio, M. (ed.) Wing TV. Services to Wireless, Integrated, Nomadic, GPRS-UMTS&TV handheld terminals. D6: Common field trials report. November 2006, 86 pp.
27. Fanton, M.D. (2006) Analysis and Measurement of RF System Reflections and DTV Transmission. Electronics Research, Inc. ERI Technical Series, Vol. 6, April 2006, 7 pp.
28. BPN – 066, issue 1.0, guide of SFN Frequency Planning and Network Implementation with regard to T-DAB and DVB-T, EBU – technical department 2005, 124 pp.
29. Silvennoinen, A. (2006) DVB-H Network Optimization under Finnish Conditions (DVBH–lähetysverkon optimointi Suomen olosuhteissa). Written in Finnish. Helsinki University of Technology, Department of Electrical and Communications Engineering. Master's Thesis. Espoo, Finland, 15.5.2006, 111 pp.
30. Ligeti, A. and Zander, J. (1999) Minimal cost coverage planning for single frequency networks. *IEEE Transactions on Broadcasting*, **45**(1), 78–87.
31. Ligeti, A. (1999) Coverage Probability Estimation in Single Frequency Networks in Presence of Correlated Useful and Interfering Components. VTC 1999, pp. 2408–12.
32. Plets, D., Verloock, L., Joseph, W. and Martens, L. (2009) Influence of DVB-H Network Deployment on Total RF Exposure. Proceedings of the 20th International Zurich Symposium on EMC, Zurich 2009, pp. 329–32.
33. Plets, D., Verloock, L., Joseph, W. *et al.* (2009) Weighing the Benefits and Drawbacks of an SFN by Comparing Gain and Interference caused by SFN Operation. IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (Broadband Multimedia 2009), Bilbao, Spain, 13–15 May, 2009, 5 pp.
34. Bóveda, J., Marcos, G., Pérez, J.M. *et al.* (2009) MER degradation in a Broadcast Mobile Network. IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (Broadband Multimedia 2009). Bilbao, Spain, 13–15 May, 2009, 5 pp.
35. Aparicio, M. (ed.) (2006) Wing TV. Services to Wireless, Integrated, Nomadic, GPRS, UMTS & TV handheld terminals. D6: Common field trials report. November 2006, 86 pp.
36. Aparicio, M. (ed.) Wing TV. Services to Wireless, Integrated, Nomadic, GPRS-UMTS&TV handheld terminals. Wing TV Country field report. November 2006, 258 pp.
37. Plets, D., Verloock, L., Joseph, W. *et al.* (2008) New Method to Determine the SFN Gain of a DVB-H Network with Multiple Transmitters. 58th Annual IEEE Broadcast Symposium, 15–18 October 2008, Alexandria, VA, USA, 4 pp.

38. Decree on the limitation of exposure of the public to non-ionizing radiation 294/2002, based on the law on radiation protection 592/1991, 43.
39. Federal Communications Commission. Report and order and further notice of proposed rulemaking (FCC 07-72). Adopted: April 25, 2007. Released: April 27, 2007. 191 pp.
40. Limits of Human Exposure to Radiofrequency Electromagnetic Fields in the Frequency Range from 3 kHz to 399 GHz. Safety Code 6. Environmental Health Directorate, Health Protection Branch. Publication 99-EHD-237. Minister of Public Works and Government Services, Canada 1999, ISBN 0-662-28032-6, 40 pp.
41. Skiöld, D. (2006) An economic analysis of DAB & DVB-H. EBU Technical Review – January 2006, 10 pp.
42. Sattler, C. (2006) Mobile Broadcast Business Models: A State of the Art Study. BMCOFORUM, November 2006, 36 pp.
43. Hoikkanen, A. (2006) Economics of wireless broadcasting over DVB-H networks. Wireless Telecommunications Symposium 2006, pp. 1–5.
44. Zhang, Y., Zhang, C., Cosmas, J. *et al.* (2008) Analysis of DVB-H network coverage with the application of transmit diversity. *IEEE Transactions on Broadcasting*, **54**(3), 568–76.
45. Gómez-Baquero, D. and Bria, A. (2007) Forward Error Correction for File Delivery in DVB-H. Polytechnic University of Valencia, Spain, and Royal Institute of Technology, Sweden. Vehicular Technology Conference VTC 2007, Spring. Dublin, 22–5 April 2007, 5 pp.
46. Johansson, K., Zander, J. and Furuskär, A. (2007) Cost efficient deployment of heterogenerous wireless access networks. IEEE 65th Vehicular Technology Conference 2007, pp. 3200–4.
47. Bria, A. and Gómez-Barquero, D. (2005) Scalability of DVB-H deployment on existing wireless infrastructure. IEEE 16th International Symposium on Personal, Indoor and Mobile Radio Communications 2005, pp. 716–20.
48. FCC, human exposure. http://www.fcc.gov/guides/human-exposure-rf-fields-guidelines-cellular-and-pcs-sites (last accessed 3 August 2014).
49. Bouttevin, T. (ed.) Wing TV. Services to Wireless, Integrated, Nomadic, GPRS-UMTS&TV handheld terminals. D8 – Wing TV Measurement Guidelines & Criteria. Project report, 45 pp.
50. IP datacast over DVB-H : Implementation guidelines for mobility, DVB document A117, DVB organization, July 2007, pp. 18–19.
51. Erceg, V., Greenstein, L.J., Tjandra, S.Y. et al. (1999) An empirically based path loss model for wireless channels in suburban environments. *IEEE Journal on Selected Areas in Communications*, **17**(7): 1205–11.
52. ETSI EN 301 192, V1.4.1 (2004-11). Digital Video Broadcasting (DVB); DVB specification for data broadcasting. European Telecommunications Standards Institute 2004, 78 pp.
53. ETSI, "Digital Video Broadcasting (DVB); Framing Structure, channel coding and modulation for digital terrestrial television." ETSI standard, ETSI EN 300 744, November 2004.
54. ETSI, "Digital Video Broadcasting (DVB); Transmission System for Handheld Terminals." ETSI standard, ETSI EN 302 304, June 2004.
55. ETSI, "Digital Video Broadcasting (DVB); Transmission System for Handheld Terminals (DVB-H)," EN 302 304 v1.1.1, Nov. 2004.
56. ETSI, "IP Datacast over DVB-H; Set of Specifications for Phase 1," TS 102 468, 2006.
57. ETSI, "IP Datacast over DVB-H; PSI/SI," TS 102 470, 2006.
58. ETSI, "Digital Video Broadcasting (DVB); Specification for Data Broadcasting." ETSI standard, ETSI EN 301 192 v1.4.1, November 2004.
59. Faria, G., Henriksson, J.A., Stare, E. and Talmola, P. (2006) *DVB-H: Digital Broadcast Services to Handheld Devices*, IEEE, 16 pp.
60. Gardikis, G., Kokkinis, H. and Kormentzas, G. (2007) Evaluation of the DVB-H data link layer. European Wireless 2007. 1–4 April 2007, Paris, France, 6 pp.
61. Henriksson, J. (2005) DVB-H standard, principles and services. HUT seminar T-111.590. Helsinki, 24.2.2005. Presentation material, 53 pp.
62. Jeong, M. (2001) *Comparison Between Path-Loss Prediction Models for Wireless Telecommunication System Design*, IEEE, 4 pp.
63. Joki, H. and Paavola, J. A Novel Algorithm for Decapsulation and Decoding of DVB-H Link Layer Forward Error Correction. Department of Information Technology, University of Turku, 6 pp.

64. Jokela, T. and Lehtonen, E. Reed-Solomon Decoding Algorithms and Their Complexities at the DVB-H Link-Layer. Turku Centre for Computer Science, Department of Information Technology, University of Turku, 5 pp.
65. Lee, W.C.Y. (1986) Elements of cellular mobile radio system. *IEEE Transactions on Vehicular Technology*, **VT-35**(2), 48–56.
66. Mazenc, J. (2007) Multi protocol Encapsulation – Forward Error Correction (MPE-FEC). ETUD/Insa. Toulouse, France, January 9, 2007. 7 pp.
67. Limits of Human Exposure to Radiofrequency Electromagnetic Fields in the Frequency Range from 3 kHz to 399 GHz. Safety Code 6. Environmental Health Directorate, Health Protection Branch. Publication 99-EHD-237. Minister of Public Works and Government Services, Canada 1999. ISBN 0-662-28032-6, 40 pp.
68. Fanton, M.D. (2006) Analysis of Antenna Beam-tilt and Broadcast Coverage. ERI Technical Series, Vol. 6, April 2006, 3 pp.
69. OMA: "Mobile Broadcast Services," Draft Version 1.0 – 03 Nov 2008, Open Mobile Alliance, OMA-TS-BCAST_Services-V1_0-200801103-D.
70. OMA: "Service Guide for Mobile Broadcast Services," Draft Version 1.0 – 28 Oct 2008, Open Mobile Alliance, OMA-TS-BCAST_Service_Guide-V1_0-20081028-D.
71. Paavola, J., Himmanen, H., Jokela, T. *et al.* (2007) The performance analysis of MPE-FEC decoding methods at the DVB-H link layer for efficient IP packet retrieval. *IEEE Transactions on Broadcasting*, **53**(1), 263–75.
72. Penttinen, J. (2008) Field Measurement and Analysis Method for DVB-H Terminals. The Third International Conference on Digital Telecommunication, ICDT 2008, 29.6.-5.7. 2008 – Bucharest, Romania, 6 pp.
73. Penttinen, J.T.J. (2008) The Simulation of the Interference Levels in Extended DVB-H SFN Areas. The Fourth International Conference on Wireless and Mobile Communications, IEEE 2008, pp. 223–8.
74. Penttinen, J.T.J. (2008) The SFN gain in non-interfered and interfered DVB-H networks. The Fourth International Conference on Wireless and Mobile Communications, IEEE 2008, pp. 294–9.
75. Penttinen, J.T.J. (2009) DVB-H Performance Simulations in Dense Urban Area. The Third International Conference on Digital Society, IEEE 2009, 6 pp.
76. Silvennoinen, A. (2006) DVB-H –lähetysverkon opti-mointi Suomen olosuhteissa (DVB-H Network Optimization under Finnish Conditions). Master's Thesis, Helsinki University of Technology, 15.5.2006, 111 pp.
77. Väre, J. and Puputti, M. (2004) Soft Handover in Terrestrial Broadcast Networks. Proceedings of the 2004 IEEE International Conference on Mobile Data Management, Berkeley, CA, USA, Jan 2004, pp. 236–42.
78. Väre, J., Alamaunu, J., Pekonen, H. and Auranen, T. (2006) Optimization of PSI/SI Transmission in IPDC over DVB-H Networks. Proceedings of the 56th Annual IEEE Broadcast Symposium, Washington, DC, USA, Sept. 2006.
79. Digital Mobile Radio Towards Future Generation Systems. COST 231 Final Report.

23

Planning of Core Networks

Jyrki T. J. Penttinen and Jukka Hongisto

23.1 Introduction

This chapter generalizes the modern core network planning aspects by describing how LTE can be supported by the packet core network of LTE/SAE. This chapter concentrates on the 3GPP release 8 LTE/SAE architecture and Evolved Packet Core, and presents potential migration steps from the existing 2G/3G PS core towards fully compatible Rel-8 EPC.

23.2 General Planning Guidelines for Fixed Networks

The dimensioning of telecommunications networks is one of the principal tasks of operators to guarantee adequate quality level for end-users. The dimensioning is not a task to be done only once but the high quality networks require constant optimization and tuning of parameters.

The main division of the network planning can be done between fixed and mobile communications. Many of the tasks are actually common in both wired environment and radio interface as some of the most important high-level principles are to offer sufficient capacity with acceptable QoS (Quality of Service), or GoS (Grade of Service). The terminology of the planning items and parameters may vary depending on the network in question, but the basic principles are often similar.

Some of the common topics to be taken into account in the network dimensioning are the following. Estimated need for offered capacity. At the most detailed level, this item may be planned by taking into account daily, weekly, yearly and seasonal variations in the capacity demand. As a basic rule, the networks should not be overdimensioned as the business model is impacted greatly due to the too high-quality deployment. The fact is that the peak-hour consumes much more resources than off-the peak hours, and in average there is considerable waste of capacity. Certain blocking level should be thus accepted by the operator as well as users.

The Telecommunications Handbook: Engineering Guidelines for Fixed, Mobile and Satellite Systems, First Edition.
Edited by Jyrki T. J. Penttinen.

The end-to-end chain of the transmission consists of many different elements, interfaces, subnetworks, so it is important to identify the potential bottlenecks and minimize the impacts so that all the interfaces are in balance as for throughput and service levels. Essential ways to plan the network and subnetworks is to understand the theoretical performance, and to verify the performance regularly via network operation and management systems as well as via practical field tests by taking representative samples in varying network conditions, topologies, etc. It is important to understand also the network utilization per user profile types, as, for example, the behavior of voice call peak-hour might be very different form data peak hours. The network traffic typically is a mix of various profiles which are weighted as a function of time and geographical area. This means for example that business use profile is weighted at city centers during working hours whilst residential use profiles are strengthened in suburban areas during the evening.

One useful tool to estimate the traffic profiles and their impacts on the network performance are simulations. These might be based on the best-guess parameter settings, or on previous experiences of the traffic behavior. The outcome of these simulations is of utmost benefit as for the correct network parameter tuning as well as on the selection of optimal set of functionalities on the network. It should be noted that the functionalities might have interdependencies in such a way that the final optimal case is either a result of the combined functionality, or sometimes only one or functionality without combination results in the best performance. As this is extremely hard to estimate without seeing the behavior either in practice or by simulating, it is a good practice to investigate the effect via simulations or smaller-scale trial or pilot networks prior to the heavier investments. The technoeconomical optimization becomes more and more important task in all of the telecommunications networks. This chapter presents the effect of the parameters of the DVB-H radio network dimensioning by showing that even minor-looking items might have major impact on the total cost per achieved capacity and service level, which means that it is a good idea to investigate the impact analysis as carefully as possible prior to the actual network deployment. At the same time, this chapter presents some typical simulation principles and case results as examples for emphasizing the benefits of this task.

One of the modern methods currently under active investigation and development is that of self-optimizing networks (SON). This area covers many technologies, functionalities and methods, and can be considered as an umbrella that combines various known and future techniques in order to make the network adjust automatically the parameters for optimal performance, capacity or as a function of some other variable.

This chapter also gives practical information about the experience about nominal and detailed network planning for wireless environments as well as for roll-out strategies. Especially for the mobile communications networks, there are some extremely important tasks like site hunting, preparation of site (signaling, communication line preparation, road preparation, tower construction etc. that is not directly related to the technical and theoretical planning but is very important to take into account in the options of the plans. As an example, quite often the optimal location of the towers are not possible to obtain due to the restrictions that are hard to estimate in advance, like special local rules for the antenna installations etc.

23.3 Planning of the Networks

The main goal of the dimensioning of the fixed telecommunications networks is related to the offered capacity. This item includes the preparation for the number of the available channels for the communications as well as the reserve of the subscription numbers. On the other hand, the network planning should take into account the placement of the network elements as well as the alternative routing of the connections in the case of service breakdowns. This is critical for the voice traffic that is handled via the old-fashioned circuit switched technologies like PCM/TDMA, as well as in any modern network which is based on certain maximum amount of available capacity. The issue is equally important in the packet switched environment, although by the nature of the technology, the rerouting of the traffic is more automated via, for example, the router infrastructure of IP subnetworks.

One of the issues in the circuit switched telecommunications infrastructure is the interference level of the wired connections like paired cobber lines. The issue increases as a function of the distance as the parallel lines tend to cause inducted interferences which might trigger problems both in older analog connections as well as in the digital environment. The role of fiber optics increases rapidly in global level which lowers this problem.

If the core network transmission has components that are based on radio links, for example, via satellites or repeaters, it is equally important to plan the core network based on the same rules as in the RF interface of the mobile communications networks, to avoid, for example, cochannel interferences. Furthermore, the service breakdown rate must be dimensioned well enough to comply with all the requirements of the operator, regulator and in general the global rules for the minimum quality levels for the mean service outage times.

For the cellular networks, the radio network includes, in addition to the capacity and quality of service common to the fixed networks, also radio coverage, frequency and radio parameter planning, which can further be divided into outdoor and indoor planning. Typically, the outdoor environment is divided into cluster for the coverage planning, like dense urban, urban, suburban, rural and open areas. Each of these area types behave differently as for the radio propagation and thus the achieved cell ranges. Regardless of the technology, it is possible to deploy the radio network by utilizing single layer approach, or multilayer strategy. The latter may consist of umbrella cells covering large areas but with only basic capacity whilst the underlying cells can be utilized more locally to take care of the higher capacity. The multilayer structure can consist, for example, of two or three layers in practice, which can also be offered via different technologies. As an example, the basic capacity can be offered via the macro cells of GSM 900 and GSM 1800 for the large coverage areas for voice calls, and when more capacity is needed for wireless Internet access, it can be offered via smaller high-frequency 3G micro cells, and highest data rate users can be served very locally via pico cells of LTE. It should be noted that the interradio technology for the access (RAT) can be utilized for the seamless handovers more fluently than before, and thus, for example, handover between mobile communications networks and Wi-Fi is possible where functionality is activated and Wi-Fi hotspot is available.

Figure 23.1 shows the essential steps of the radio network planning process. The plan can be divided for the nominal plan that gives a rough estimate for the offered capacity in a given coverage area, and the detailed level planning further clarifies the final behavior of the offered traffic as a terms of, for example, data throughput and quality level as a function of the number of users that each have certain user profile

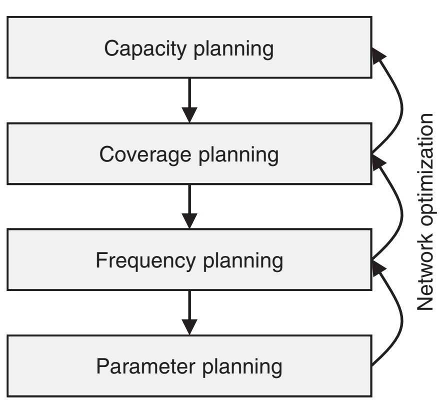

Figure 23.1 *The main steps of the radio network planning.*

(a mix of voice and varying data services and SMS and MMS traffic). The basic procedure can be initiated via the capacity plan which follows by the coverage plan. The capacity and coverage are basically tight together regardless of the technology, that is, when more capacity is offered, it can be done via loosening the code rate which in turn reduces the useful cell radius.

In case of data networks, there are many special features that need to be taken into account in the radio and core network planning. As an example, the IP numbering plan for the subnetworks must be done by estimating the current and future usage. Also the physical aspects of the LAN networks, to mention another example, must be taken into account as each LAN variant has limitations to, for example, segment length of the cables.

In general, it is highly recommendable to take into account the future development of the customer base as well as it is possible via the practical estimates. This is to guarantee the best possible service level along with the future network infrastructure investments, and the optimization between the new purchases of the equipment. In the most accurate and detailed estimates, some essential tools for the planning are thus the forecasts of the population growth and economical categories of the households.

23.4 Capacity Planning

Both fixed and radio networks can deliver certain maximum amount of connections for the end-users. Even in the networks that are based on the varying of the capacity along with the simultaneous users, there is a practical limit and there are thus no networks that could offer service without any blocking whatsoever. The operators must thus decide the level of offered capacity that is optimal for both end-users (sufficient average quality level for the connections, and reasonable outage probability) and for operators (costs of investments, and reasonable return of investment).

In the capacity planning, the need for the physical network equipment must be estimated for each geographical location of the network access points. Depending on the network type, this means that the estimate for the number of routers, bridges, mobile services switching centers, base station controllers, base stations etc. must be known and analyzed for the capacity offer. The essential sources of information for complying with the requirements area also found in the specifications of the systems and equipment, as well as the quality criteria of the regulators and operators. In the radio links and mobile communications network cells, the radio propagation models serve as a good estimation tool for the achieved coverage with a planned capacity offering. The topology of the surrounding area must thus be estimated as accurately as reasonable, for example, via 3D models of the geography in the most accurate cases. As the more advanced and accurate modeled maps are more expensive than basic 2D-maps, the optimal network planning takes into account also this expense and the task of the operator is thus to decide how accurate predictions are needed in each cluster. As a rule of thumb, the densest urban areas are worth designing with 3D models whilst the planning of suburban areas typically would not benefit considerably form the advanced methods.

High-quality network planning also includes analysis for future expansions. It is wise to do this future planning for several years beforehand although the exact subscription numbers would be hard to estimate. The plan would include the estimate for the need of the additional functionalities, elements and capacity resources. A good tool for the plan is user forecasts which convert more accurately during the time.

In case of circuit switched networks like PCM and TDM based technologies that deliver voice service, the most important planning item is to estimate traffic and thus the amount of the offered channels. This item requires understanding about the average reservation time per channel. On the other hand, the load of the network and thus the service level must be understood. The load of the channels varies greatly during the hours of the day, and there are typically many profiles found during the week, year, and season. There are also occasional special events in many locations that may require special attention from the operator side. The solution for these types of days is to plan in advance additional capacity for the location by, for example, providing additional cells under existing base stations, or additional complete base stations that are vehicle mounted with additional transmission capacity via fixed networks (fiber optics nearby) or via radio links or satellite links.

In the normal case, the offered load is typically dimensioned by estimating the busy hour of the location. The definition of the busy hour varies slightly, as it can be, for example, a single, most loaded hour of the day, or it can be a combination of four nonconsecutive quarters of hour. In any case, the task is to find an estimated moment of the day when the network is busiest.

In each case, the most practical tool for capacity dimensioning has traditionally been Erlang B which is meant for the systems without queuing. For the systems capable of queuing (e.g., during 10 seconds in the call attempt), there is an enhanced variant, Erlang C. Nevertheless, the basic Erlang B is widely utilized in mobile communications systems as well as in the fixed voice networks to estimate the final capacity of the network when certain blocking probability is designed.

Let's take an example of GSM radio interface. It can be assumed that GSM networks are dimensioned according to 2–3% blocking probability during the peak hour for voice calls. The cells can contain several tens of percents of free time slots for GPRS usage, depending on the total number of TRXs (transceiver units) even during these peak periods. This result is directly derived from the nature of the Erlang B formula; in order to offer the service with a certain blocking probability, there has to be, on average, a certain amount of free channels. In order to achieve relatively fluent service, the blocking cannot exceed too many percentage points.

When the core network is dimensioned using a certain blocking probability, the transferred traffic can be calculated using Erlang B as shown in the following formula. The Erlang B formula can be considered as a starting point for estimating the traffic and calculating the required channels and the number of TRXs. The formula gives the blocking probability both in the time domain as well as for the number of attempted calls. In other words, the formula gives the probability of the situation, where all channels are occupied:

$$B(N) = \frac{\frac{A^N}{N!}}{\sum_{n=0}^{N} \frac{A^n}{n!}}, \tag{23.1}$$

where N is the total number of traffic channels, and A is the product of the average call density and the average time of occupation, that is, A is the offered load. The formula can also be modified to a recursive form, more suitable for computer calculations:

$$\begin{aligned} B(0) &= 1 \\ B(N) &= \frac{AB(N-1)}{N + AB(N-1)}. \end{aligned} \tag{23.2}$$

The basic rule of thumb for the dimensioning of the GSM network is the call blocking during the peak hour. Blocking can thus be easily measured from existing networks by dividing the number of blocked calls by the total number of attempted calls. The normal dimensioning value might be for example, 2–3% during the heaviest period of the network. The served traffic $\bar{x}$ can be calculated with the formula:

$$\bar{x} = A(1 - B), \tag{23.3}$$

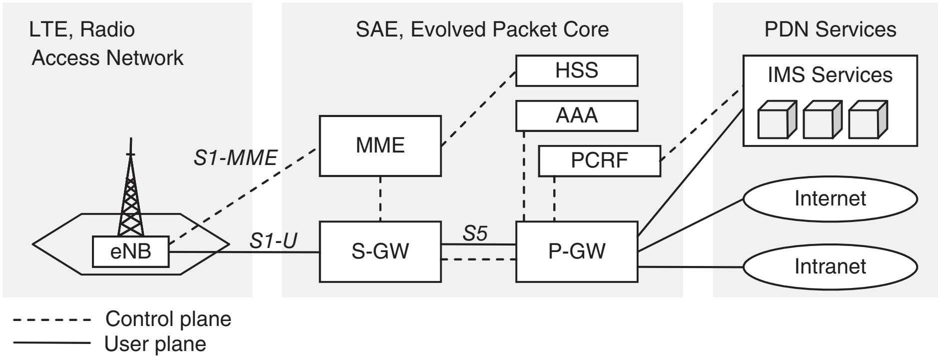

Figure 23.2 *3GPP R8 architecture for LTE/SAE.*

and the blocked traffic can thus be calculated as follows:

$$m = A - \bar{x} = AB. \tag{23.4}$$

23.5 Network Evolution from 2G/3G PS Core to EPC

The following chapters are based mainly on Ref. [1], Chapter 12.

23.5.1 3GPP R8 Requirements for LTE Support in Packet Core Network

3GPP Release 8 defines a new Evolved Packet Core (EPC) for LTE access. The EPC can be used also for other access technologies like GERAN, UTRAN and CDMA2000. Figure 23.2 shows the standard Release 8 architecture with the relevant interfaces between the logical entities.

The Mobility Management Entity (MME) is the equivalent of the SGSN in 2G/3G GPRS networks. In the LTE/SAE network, MME is a pure control plane element. It initiates a direct tunnel between the eNodeB and Serving Gateway in order to deliver the user plane traffic.

According to the 3GPP Release 8, the mobile gateway functionality is divided into Serving Gateway (S-GW) and Packet Data Network Gateway (P-GW) functionalities. These S-GW and P-GW functionalities can be implemented in the same physical node or in two separate entities. They are connected logically via an open S5 interface, which is called as S8 in the case of roaming.

The S-GW and P-GW functionalities are defined as mandatory for the LTE radio network deployment. The LTE subscribers are always connected to the services via P-GW even if 2G/3G radio access would be utilized. 2G/3G subscribers may access existing APNs (and possible new APNs) via GGSN or P-GW as the P-GW element also contains the GGSN functionality.

S-GW terminates the LTE core user plane interface towards the evolved UTRAN. An LTE User Equipment (UE) is assigned to a single Serving Gateway at a given point of time. The Serving Gateway acts as a user plane gateway for the LTE radio network in case of the inter-eNodeB handovers, and for the inter-3GPP mobility (terminating S4 and relaying the traffic between 2G/3G system and PDN-GW).

The Packet Data Network Gateway (PDN-GW) acts as a user plane anchor and terminates the SGi interface towards the service networks. It allocates the IP address for the UE. PDN-GW applies policy enforcement to the subscriber traffic and performs packet filtering on individual user's level (by performing, e.g., a deep

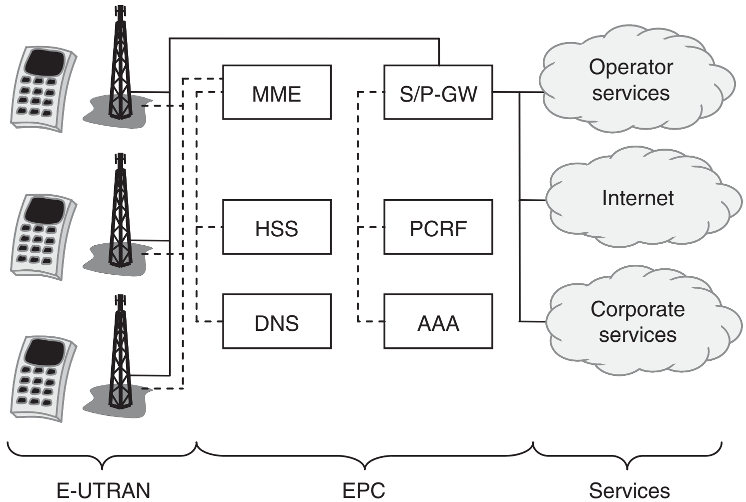

Figure 23.3 EPC for LTE access.

packet inspection). PDN gateway interfaces with the Communications Service Provider online and offline charging systems.

23.5.2 Introducing LTE in Operator Network

Operators who want to start LTE deployment in an early phase will likely begin with a technology trial to test LTE radio and EPC capabilities according to Figure 23.3. The trial network is introduced as an overlay by leaving the existing production machinery intact. When the LTE radio is introduced, it is mandatory to introduce also the MME and S/P-GW capabilities in the core network. In the trial phase, the subscribers are offered with the mobility within the LTE radio network. The Internet access and selected operator services can be included in the set of services of the trial network.

23.6 Entering Commercial Phase: Support for Multimode LTE/3G/2G Terminals with Pre-Release 8 SGSN

When the subscriber moves outside of the LTE network coverage area, the service continuity is a necessary requirement for the operators in order to be able to commercially launch feasible LTE networks. This requires availability of multimode terminals that are capable of supporting 2G, 3G and LTE. Also the integration of 2G/3G core network with the Evolved Packet Core is needed. The integration is required to allow handovers between the LTE and 2G/3G access networks. The gateway acts as an anchor point for all the subscriber sessions. In other words, the same gateway element serves the subscriber session when the subscriber moves between the LTE and 2G/3G networks. This means that even if the subscriber initiated a session in an area where only 2G/3G coverage is available, the session would be served by the P-GW. In practice, this integration means the provisioning of the connectivity between the 2G/3G SGSN and EPC.

As a simplest option, the SGSN can be connected to PDN-GW via the Gn interface which is used for the LTE/3G/2G terminals in order to provide the IP connectivity, and Gn interface between MME and SGSN for

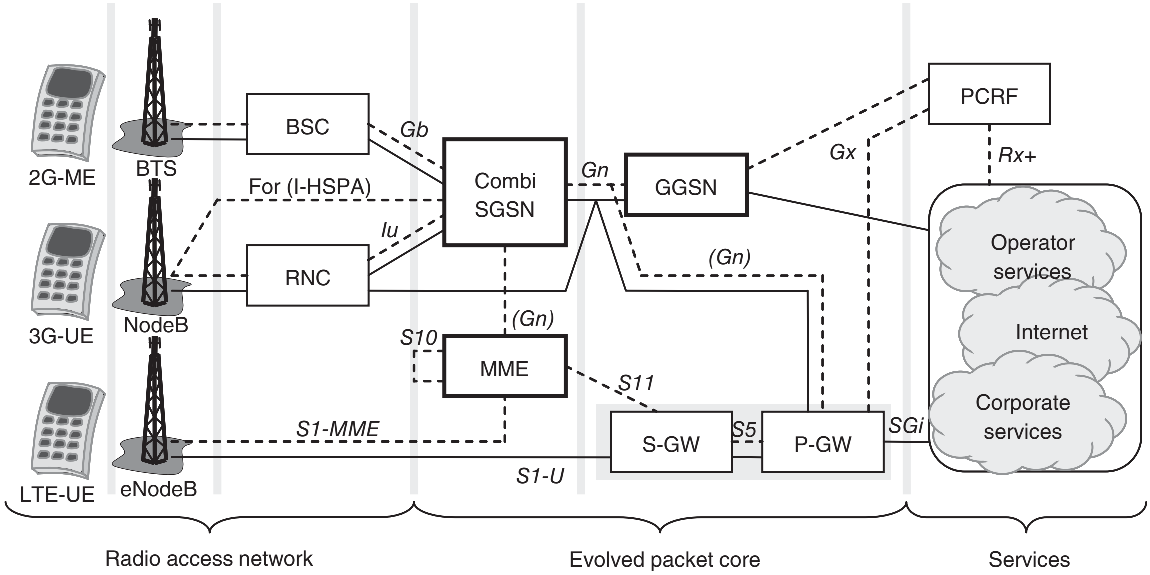

Figure 23.4 *LTE/3G/2G interworking with Pre Release 8 SGSN. The control plane interface from NodeB to SGSN is for Internet-HSPA (I-HSPA).*

the mobility between the LTE and 2G/3G networks. If the Direct Tunnel of the 3GPP Release 7 is used, the user plane traffic goes directly between UTRAN and P-GW.

In this phase, the operator can use GGSN for the 2G/3G traffic of all the terminals that are not LTE capable. If the operator uses the same APN definitions for both 2G/3G and LTE services, then SGSN has to support PDN GW selection based on, for example, IMEI or the terminal capability. Figure 23.4 presents an example of this phase.

23.6.1 Support for Multimode LTE/3G/2G Terminals with Release 8 Network

The deployment of the Release 8 SGSN allows the operator to introduce so-called Common Core where also 2G and 3G accesses are linked to S-GW. The Release 8 QoS model, which means that the network controls the QoS level, is applied also for the 2G/3G traffic.

When SGSN is upgraded to 3GPP Release 8 level, it will have a new S4 interface towards S-GW, and an S3 interface towards MME. In this phase, S-GW is used also for the 2G/3G traffic. Furthermore, it is possible to utilize the Release 8 Direct Tunnel (S12 interface) between UTRAN and S-GW. The already existing GGSN of the operator can still be used for the 2G and 3G subscribers. Figure 23.5 presents an example of this phase.

At this point, it is also relevant to have more optimized interworking with integrated SGSN/MME implementation with following options:

– 3G SGSN and MME for control plane only.
– Combined SGSN/MME for all 3GPP accesses.

From the S-GW/P-GW point of view, it does not matter which option is used because all the traffic goes anyway through S-GW/P-GW, and the intersystem mobility needs to be signaled to the gateway for a proper bearer, policy and charging control purpose.

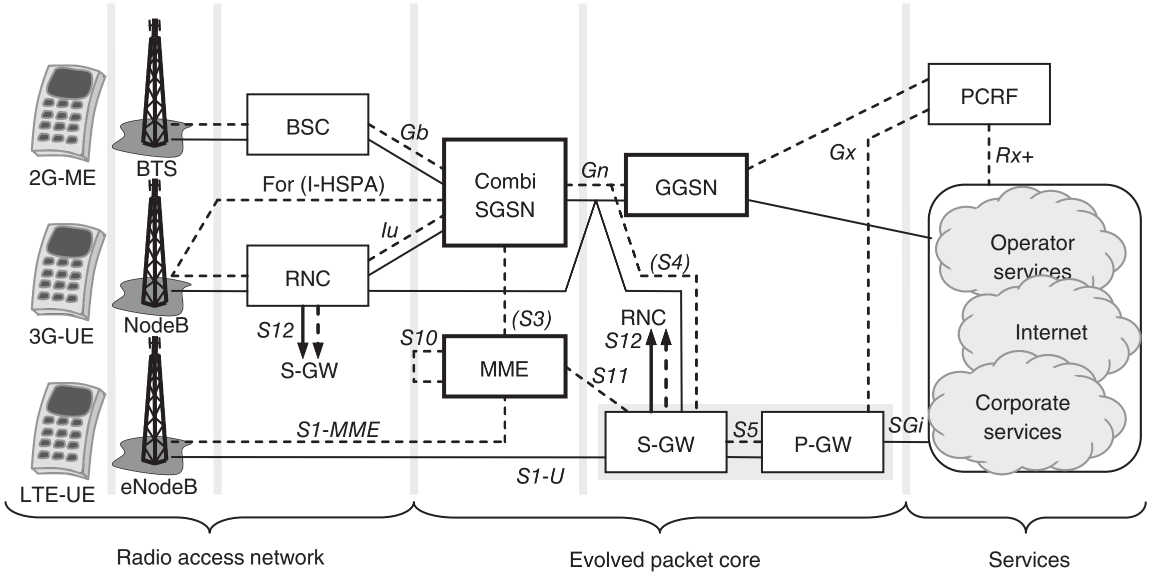

Figure 23.5 *LTE/3G/2G interworking with Release-8 SGSN.*

23.6.2 Optimal Solution for 2G/3G SGSN and MME from Architecture Point of View

The 3GPP Release 8 architecture mandates the User Plane and Control Plane separation. This is also being increasingly applied in the current 3G networks with the Direct Tunnel functionality. In the 2G Packet Core network, the User Plane and Control Plane functions are so tightly coupled that the separation is not feasible. Based on these facts, the optimal solution from the architecture point of view is to apply separate physical nodes for 3G SGSN/MME and for 2G SGSN. The combining of all these three functions in the same physical element is technically possible but it means that the same physical element has to support two very different architectures.

In the presented optimal solution scenario, the assumption is that the 2G traffic growth is moderate compared to the 3G/LTE traffic evolution, and that the operator has 2G SGSN elements that can continue to deliver the 2G traffic. Most of the 2G transport networks may still use the legacy E1/T1 and Frame Relay interfaces but may not be upgraded to support IP based interfaces.

The flat architecture with minimum number of user plane elements has been one of the basic principles of the 3GPP network architecture evolution. As LTE and 3G share common network architecture, this gives an opportunity to align the 3G SGSN and MME functionalities into a single entity. 3G SGSN as a pure control plane element makes it behave like MME and to offer most benefits as a combined element. On the other hand, S-GW and P-GW handle user plane traffic in both nonroaming and roaming cases. Figure 23.6 presents an example of this phase.

By separating the control plane and user plane, the operator can build a modernized all-IP network that accomplishes future data traffic and mobility growth with the flat architecture. The benefits of this optimized architecture include the following:

- Most of the user plane traffic of 3G/HSPA/LTE coming from the Internet can be optimally routed based on the operator peering points.
- With combined S-GW and P-GW, the operator can decrease CAPEX and OPEX by nearly 50%.

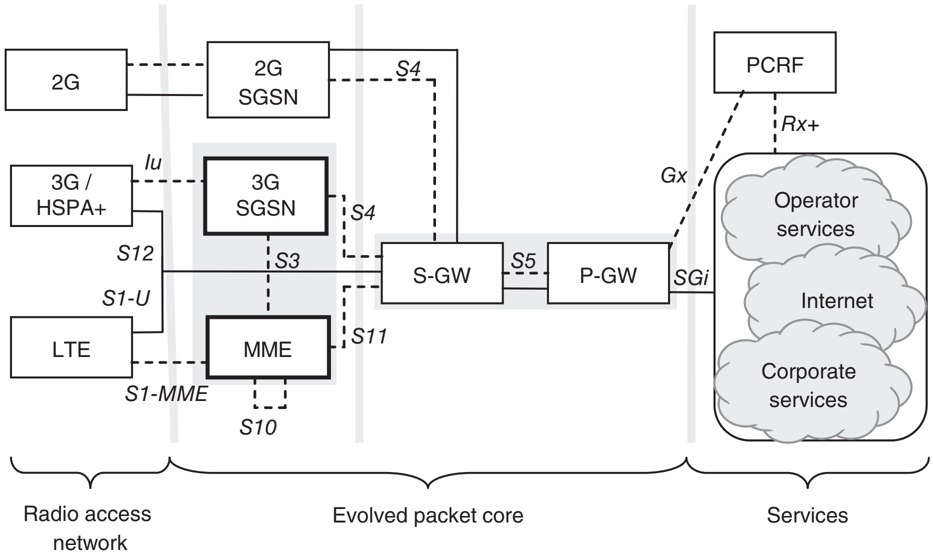

Figure 23.6 Optimized architecture for high speed 3G/HSPA/LTE traffic.

- Combined MME and 3G SGSN allows shared capacity for the LTE/3G subscribers and optimizes the mobility management with the reduction of the signaling traffic when UE moves between LTE and 3G service areas.
- MME and 3G SGSN can support pooling with geographical redundancy. The pooling location can also be close to MSS in order to simplify the connections to the CS domain (SGs and Sv interfaces).
- Release 8 3G Direct tunnel means that SGSN is meant for the control plane only and S-GW and P-GW are acting in the user plane.

23.6.2.1 Combined SGSN/MME for all 3GPP Accesses

It is technically possible to combine all the 2G/3G SGSN and LTE/MME functionalities into a single network element. This can, in fact, be a preferred solution in case the operator wants to upgrade the existing SGSN elements to the new hardware or wants to deploy a minimum number of network elements in the network.

There are two alternative paths leading to this combined scenario:

- The MME functionality is introduced on top of the flat-architecture-optimized network element. The 2G and 3G subscribers are migrated to this element as these accesses are supported. This is the preferred option as the 3G and LTE traffics are driving the mobile data evolution and the solutions are to be optimized according to the needs of these accesses.
- The MME functionality is introduced as a SW upgrade to the existing SGSN. This can be considered as an option in case the operator has unused capacity in the existing SGSN elements.

When a single element includes all the SGSN and MME functionalities, it makes subscriber migration from 2G or 3G to LTE easier and reduces also the inter-SGSN-MME signaling. On the other hand, there will be only few LTE subscribers in the beginning of the operational phase of the LTE network, which means that not all the combined SGSN elements will be upgraded to support MME functionality. When the combined

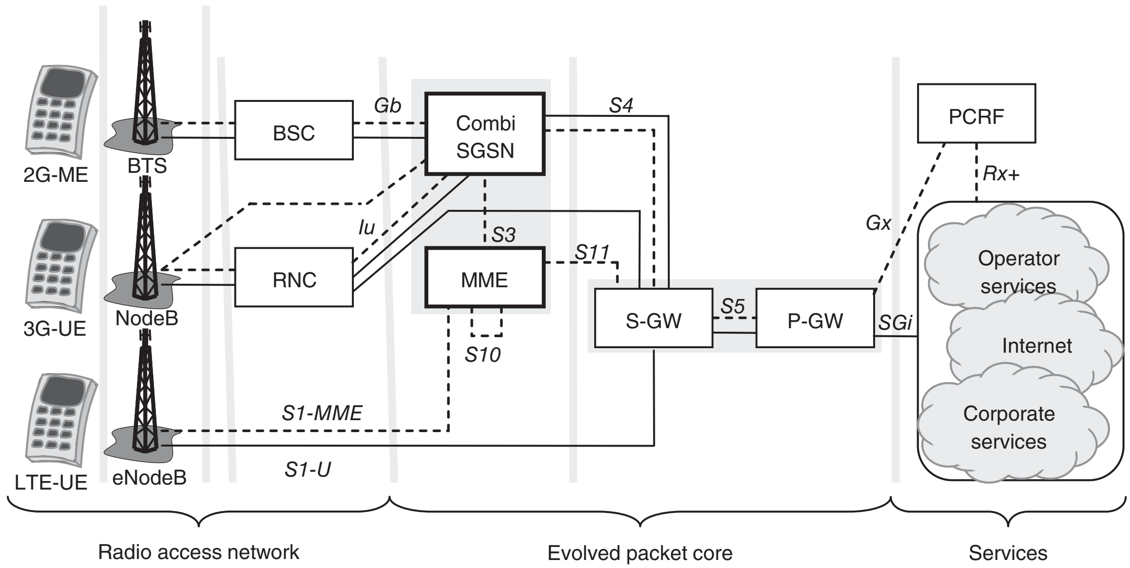

Figure 23.7 *2G/3G SGSN and MME functionalities supported in a single element.*

SGSN-MME scenario is selected, it is preferred that all the interfaces are upgraded to work in the IP based environment, including the Iu and Gb interfaces. Figure 23.7 presents an example of this phase.

23.7 SGSN/MME Evolution

23.7.1 Requirements to MME Functionality in LTE Networks

In the 3GPP Release 8 architecture, the Mobility Management Engine (MME) is a pure control plane element which takes care of the Control Plane traffic handling, session and mobility management, idle mode mobility management and paging.

The flat network architecture, as presented in Figure 23.8 where eNode-B interfaces directly with the core network elements without an aggregating RNC layer, makes all the mobility management events visible for the core network nodes.

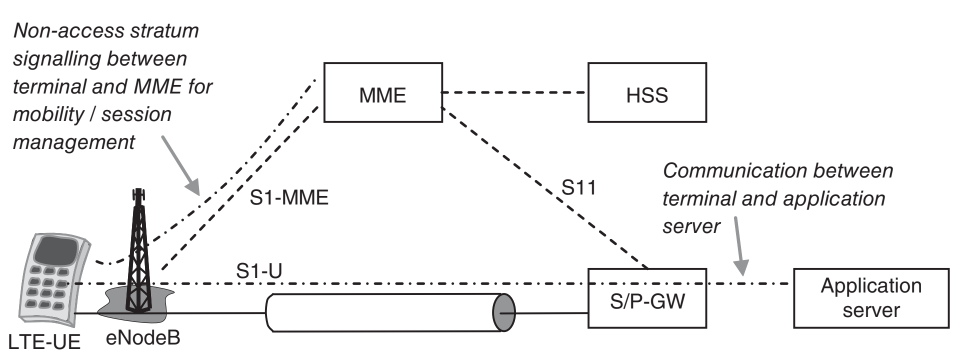

Figure 23.8 *Flat network architecture and bad terminal behavior set challenges to MME.*

The signaling traffic load is already today a more severe bottleneck than the user data throughput in the networks where a large number of smartphones are used. Always active services like Virtual Private Network (VPN), email, chat and machine to machine applications generate continuous keep-alive signaling. Some "badly behaving" smart phone types try to optimize the battery lifetime by switching the state to idle immediately when there is no more packets being sent or received, which results in continuous idle-active signaling load to the network.

23.8 Case Example: Commercial SGSN/MME Offering

23.8.1 Nokia Siemens Networks Flexi Network Server

Nokia Siemens Networks Flexi Network Server (Flexi NS) is a control and transaction machine that implements the 3GPP Release 8 compliant MME functionality. The SGSN functionality can be activated via a SW update.

The Flexi NS has been designed for flat architecture. The ATCA HW platform has been built to cope with the future evolution of LTE/SAE. The SW architecture is optimized to signaling and control plane traffic handling. This allows Flexi NS to implement high signaling capacity in terms of transactions and simultaneous sessions as well as to support high amount of Simultaneously Attached Users (SAU), high bearer and PDP context capacity and 2G throughput capacity (3 shelf configuration).

There are two capacity dimensions where flexible scaling is needed: (1) user plane vs. control plane, and (2) scaling between the access technologies 2G/3G/LTE.

Flexi NS allows dynamic capacity usage to share the node capacity among 2G, 3G and LTE subscribers. A common database is implemented to handle 2G, 3G and LTE subscriber data. This implementation allows flexible allocation of general purpose HW blades between 2G, 3G and LTE, and makes the element straightforward to dimension. Flexi NS implements session resiliency to allow maintaining the subscriber session despite a unit failure. This means that there will be no interruptions in real time services in case of failure.

Flexi NS simplifies connectivity towards the radio network and the Gateway (Serving Gateway, PDN Gateway, GGSN). IP addressing virtualization is applied to hide Flexi NS internal architecture and to show one IP address per MME node towards the radio network and the Gateway.

ATCA hardware provides advanced energy saving options with low traffic power saving mode that allows shutting down selected CPUs when the traffic load is load.

Current DX200 HW of Nokia Siemens Networks SGSN has been designed for a complex interface environment (FR, ATM, E1/T1, IP). SGSN demonstrates carrier grade performance and reliability in live networks. The evolution of this equipment base supports Direct Tunnel, and is straightforward to use in the flat mobile packet core architecture deployments.

23.8.2 Aspects to Consider in SGSN/MME Evolution Planning

When analyzing the possible network evolution scenarios for SGSN/MME solutions there are three key aspects to be considered:

23.8.2.1 Timing of LTE Introduction

Flexi NS provides an overlay solution for operators who are introducing LTE in an early phase. An overlay solution ideally isolates LTE introduction from the existing packet core network allowing easy troubleshooting and upgrades in the technology introduction phase.

23.8.2.2 Network Modernization to all-IP

With transport network being modernized to all-IP and with 3G SGSN functionality becoming available in Flexi NS the operator can start migrating the 3G subscribers from existing SGSNs to Flexi NS.

23.8.2.3 SGSN Investment Utilization

New software releases continue to be provided for the DX200 based SGSN. DX200 SGSN will evolve into a R8 SGSN that interworks with the MME. SGSN/MME interworking is critical especially in the LTE introduction phase as handovers to 2G/3G Packet Core need to be supported.

23.9 Mobile Gateway Evolution

23.9.1 Requirements to Mobile Gateway in Mobile Broadband Networks

Mobile broadband gateway needs to scale up to accommodate the growing traffic volumes. This is, however, not enough but user plane performance is also very important. Flat network architectures, where NSN has demonstrated market leadership already in 3G with Direct Tunnel, will change the connectivity architecture of the EPC gateways making the radio network directly visible to the gateway.

LTE increases the signaling load removing as there is no more concentrating RNC layer between the base stations and the gateway. This will increase the amount of signaling per subscriber. When signaling is combined with additional signaling for AAA, online charging and policy control, the overall processing requirements related to signaling will increase from 2G/3G substantially.

Subscriber density will also be very high due to LTE always-on connectivity, multiplying the signaling by the amount of attached subscribers. All these will impact to the gateway and MME performance profile. Mobile gateway platform needs to provide high performance, scalability and flexibility in all three dimensions. Failure to perform in any of them will lead to suboptimal solutions. Figure 23.9 presents the interdependencies of capacity, throughput and subscriber density.

23.10 Case Example: Commercial GGSN/S-GW/P-GW Offering

23.10.1 Nokia Siemens Networks Flexi Network Gateway

NSN equipment includes the concept of service awareness, combined with online charging capabilities, which allows differentiated charging for operator provided services. With the increasing demand for open Internet

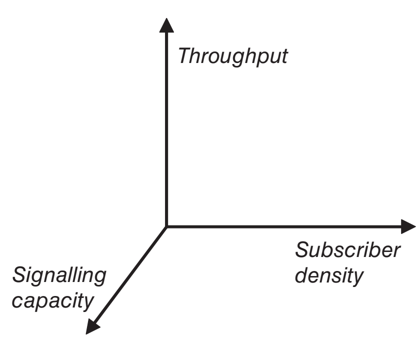

Figure 23.9 *Key mobile gateway scalability dimensions.*

access there are solutions to track proprietary peer-to-peer applications like file sharing and communication services.

Flexi NG is designed to serve as high capacity gateway for mobile broadband networks. It serves 2G/3G, HSPA, I-HSPA and HSPA+ networks as GGSN and LTE networks as S/P-GW.

Flexi NG uses ATCA platform that provides optimal fit with current and future packet core and evolved packet core requirements. NSN EPC based on ATCA provides feasible performance for throughput, signaling and subscriber density. ATCA scales for centralized and decentralized deployments, providing a flexible platform for potentially evolving topologies.

Flexi Network Gateway implements integrated Deep Packet Inspection (DPI) capabilities. It provides recognition for 300+ protocols and applications including peer-to-peer protocol tracking. The protocol and application signatures are constantly updated in Flexi NG. Product capabilities include Service awareness, that is, protocol analysis on L3/L4 and Deep Packet Inspection, that is, analysis on L7/L7+. L7+ analysis includes heuristic analysis that is typically applied to track proprietary protocols like peer-to-peer applications and services.

23.10.2 Aspects to Consider in GGSN/S-GW/P-GW Evolution Planning

Most of the 2G/3G networks still have a very limited number of active PS sessions, typically 10–20% of users have PDP context activated. In LTE there is at least one PS session per subscriber, which means that LTE subscriber growth is directly visible in S-GW and P-GW. Also new services, like VoLTE introduce more PS sessions, which mean that overall active PS session amount is more than 100–200%.

23.11 EPC Network Deployment and Topology Considerations

Network topology means hierarchy of connections between functional units in the network element. 3GPP R8 mandates flat architecture that allows flexible choices in network topology implementation.

IPv4 address exhaustion is expected to materialize in 2012 which puts pressure in transformation from IPv4 to IPv6 addressing. LTE introduction is a natural point to start IPv6 migration in operator networks. The most important change compared to 2G/3G networks is that every attached subscriber requires at least one IP address due to always on connectivity. It is a natural choice to deploy new services like operator VoIP and other IMS services based on IPv6 addressing. More addresses are also needed as the LTE network is fully IP based and every eNode-B can have several addresses.

23.11.1 EPC Topology Options

Packet core is typically deployed in a hierarchical topology that includes National sites (GGSN location), Regional sites (SGSN location), Local sites (RNC location) and Base Station sites.

Current 2G/3G deployment consists of radio network, packet core network including SGSN and GGSN and transport network that provides interconnectivity between the sites. Transport networks are being upgraded to IP based solutions. 3G operators are starting to deploy Direct Tunnel that allows transmission savings by user plane bypassing the SGSN that is routed straight between RNC and GGSN.

3GPP R8 architecture allows modernizing the network topology. Flat architecture becomes mandatory as user plane traffic is routed directly from eNode-Bs to Serving Gateway. Serving Gateway and PDN Gateway can be deployed either at the same site or they can be separated at different sites. OPEX and CAPEX costs can be minimized with combined deployment and this will likely be the most common scenario. MMEs can also

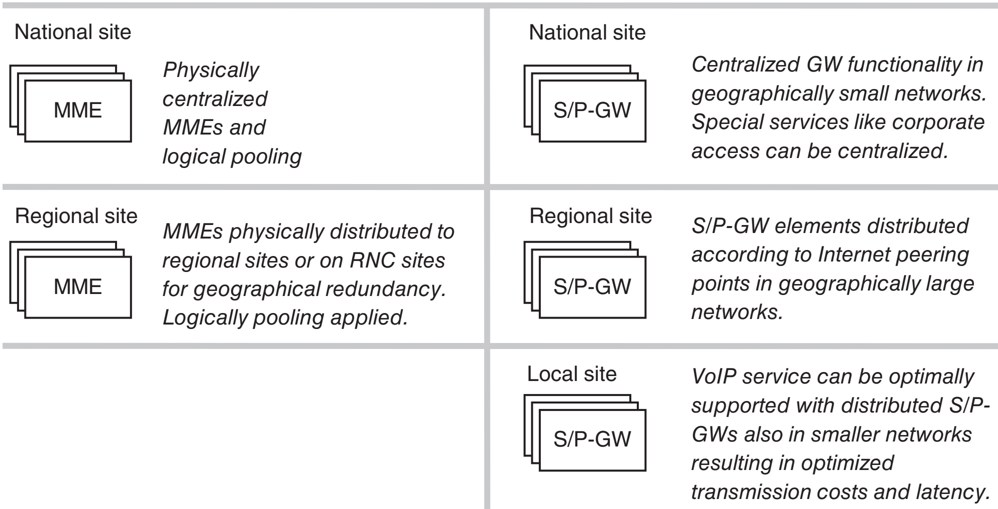

Figure 23.10 *Network topology options.*

be deployed either in centralized or distributed topology. MME pooling can be implemented irrespectively of the selected topology and it is highly recommended for optimal capacity usage.

Figure 23.10 illustrates different network topology and network element location options in a practical LTE deployment.

23.11.2 EPC Topology Evolution

It is likely that in the LTE technology introduction phase the EPC is deployed in a highly centralized manner. LTE will require new terminals and this will limit the amount of LTE subscribers in the service introduction phase. Initial LTE usage can be compared to existing 2G/3G/HSPA usage with centralized GGSN.

When network usage increases and operator expands LTE radio coverage, more Evolved Packet Core elements will be needed and operator can add number of sites for regional redundancy purposes. With MME pooling functionality (S1-flex), the MMEs can be deployed in a very centralized manner in a few sites.

Inter system mobility between LTE and 2G/3G networks may drive SGSN and MME collocation to optimize mobility management and to allow smoother user migration. It needs to be noted that as 2G SGSN continues to handle user plane traffic, highly centralized 2G SGSN is not recommended for latency reasons.

From connectivity perspective MME has Home Subscriber Server (HSS) interface and it is interworking with Circuit Switched domain via SGs/Sv connections to MSC, which may drive harmonization of MME and MSC topology.

The P-GW provides connection to content and service networks including Internet, operator services and corporate connectivity. Routing of high volume traffic needs to be optimized to minimize transmission costs and VoIP traffic needs minimized latency. Because of these reasons it is likely that in the future gateways will be deployed in a more distributed manner according to peering point locations.

The P-GW has interfaces to PCRF, AAA server and charging system, which need to be considered if P-GW is very decentralized. P-GW location can also be driven by service/APN usage for example, corporate VPN or the need to provide connectivity for outbound roamers.

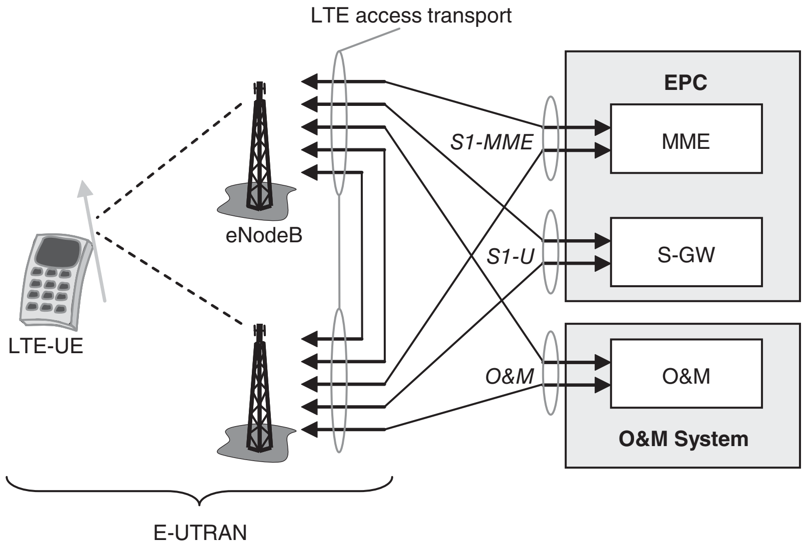

Figure 23.11 *LTE access network reference architecture.*

23.12 LTE Access Dimensioning

The LTE access network provides interconnection between the LTE, that is, Evolved UTRAN (E-UTRAN) and Evolved Packet Core (EPC) domains. The interfaces to be considered in the access network dimensioning are S1_U and S1_MME between eNodeB, and S-GW and MME, respectively. Furthermore, the X2_U and X2_C between the eNodeB elements for the user and control plane signaling, respectively, belong to the access network. Finally, the interface between the eNodeB elements and the Operations and Management element should be taken into account. In other words, the lines that end to eNodeB elements need to be dimensioned within the access network. Figures 23.11 and 23.12 clarify the access network dimensioning points.

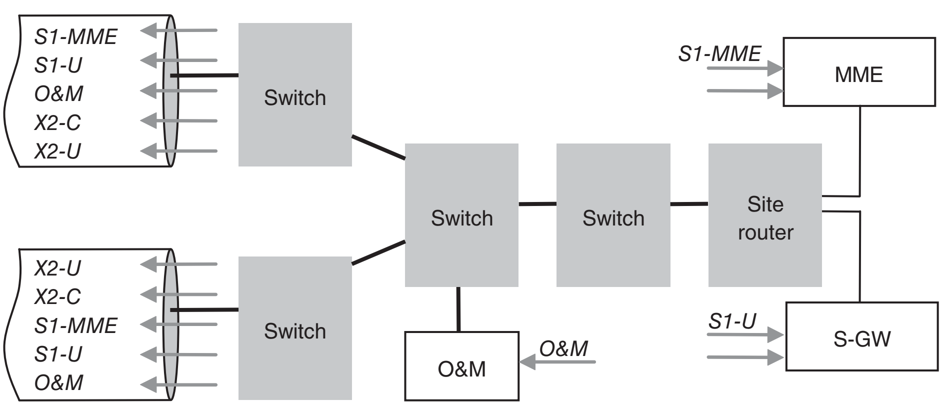

Figure 23.12 *The SAE network interfaces for the dimensioning.*

The access network dimensioning requires the relevant inputs that are the traffic profile, radio network topology and radio interface performance values. The traffic profile gives idea about the user plane traffic per LTE radio cell, handover traffic percentage via the X2 interface, signaling traffic percentage and transport overhead for the user and control planes. The radio network topology and performance indicates the number of cells of the eNodeB, the cell throughput (average and peak values are needed), and radio interface overhead estimate. The result of the dimensioning is the needed capacity per logical interface, that is, for the user plane (S1_U and X2_U), control plane (S1_MME and X2_C), and maintenance plane (OAM interface). The final dimensioning indicates what is the required capacity per whole logical interface (S1 and X2), and the total transport capacity for each eNodeB.

The LTE traffic profile is essential in order to dimension correctly the interfaces. The best estimates can be derived from the operator network statistics of the general data traffic by taking into account the estimated increase of the traffic caused by different LTE applications (VoIP, web browsing, FTP etc.). Also the LTE-UE penetration as a function of time should be taken into account for sufficiently long term.

The estimate can be estimated for the user plane traffic per radio cell in terms of the average data traffic in Mb/s values generated during the LTE traffic busy hour. Note that the busy hour for the voice and data traffic profiles typically occurs in different hours, but in case of LTE only the general peak hour can be considered as the voice service is packet based.

Next, the signaling share for the handovers in the X2_U interface is estimated. The value might oscillate typically between 2 and 3%, but the final estimate should be done case-by-case basis as the network topology does have an effect on the signaling load.

The signaling traffic share in the control plane could be around 1–2% compared to the user plane. This value can be shaped via the laboratory and field tests prior to the actual LTE deployment.

The user plane transport overhead in S1_U and X2_U interfaces depends on the utilization of IPsec in such a way that the total transport protocol headers for GTP-U, UDP, IP and Ethernet is 144 bytes with IPsec, and 78 bytes without it, resulting in 25 or 15% overhead with and without IPsec, respectively.

Finally, the control plane transport overhead for S1_MME and X2_C interfaces can be estimated depending similarly on the use of IPsec. The size of the protocol headers, in this case for SCTP, UDP, IP and Ethernet, is 140 and 74 bytes for IPsec and without IPsec respectively, causing 179 and 95% control plane transport overhead respectively.

As for the radio overhead, the protocol overhead for the PDCP, RLC and MAC is 9 bytes. Depending on the payload packet size, the overhead could be estimated to be around 2% if a mix of packet sizes is assumed.

Based on these assumptions, the user plane, control plane and management plane are dimensioned accordingly. For the user plane, the capacity for the S1_U and X2_U interfaces can be estimated jointly due to the common transport scheduler buffer. It should then be decided if the interface is dimensioned by assuming the aggregated average capacity of all cells, or peak traffics, or a combination of these.

For the control plane dimensioning, that is, for the S1_MME and X2_C, the bandwidth can be estimated, for example, based on the certain percentage from the user plane traffic. This is a straightforward task when the control plane inputs are estimated as shown previously. For the management plane, a simple rule of thumb is to allocate an additional 1 Mb/s at the eNodeB. This capacity includes also the transport overhead.

Reference

1. Penttinen, J.T.J. (2011) *The LTE/SAE Deployment Handbook*, John Wiley & Sons, Ltd, Chichester.

24

EMF – Radiation Safety and Health Aspects

Jouko Rautio and Jyrki T. J. Penttinen

24.1 Introduction

"EMF" – short for Electro Magnetic Fields – is a term used to point to the radiation properties of radio transmitters, the related safety regulation and related health issues (including speculated ones).

The EMF issue also includes the scientific, public and even political debate about the RF radiation and its interaction with living biological tissue – real or speculated. Please note that the "EMF" generally covers radio frequency [10 MHz to 300 GHz], intermediate frequency [300 Hz to 10 MHz], low frequency [up to 300 Hz] and static fields, but the focus in this book limits us only to RF and all further mentions about EMF should thus be understood as RF only, if not otherwise mentioned.

The reader should note, that EMF and EMC (Electromagnetic Compatibility) are, although relative terms, two different items as the EMF refers to radio waves influencing living tissues but EMC refers to radio waves influencing nonliving equipment (usually electronics). Typically the power strength of a radio transmission able to cause a harmful effect to a living tissue is several magnitudes larger than those which can cause interference to nonprotected electronics circuits.

This chapter will consider the scientific background and EMF related organizations, and look at the science in various EMF topics as well as briefly looking at radiation safety in the vicinity of antennas e.g., for the cases presented in Figures 24.1–24.4.

Understanding the scientific principle regarding research results in general and EMF in particular will help the reader to see the many writings in the media and the Internet in rational perspective, and helps discussing about EMF with the customer (man of the street).

However, this chapter is intended to give an overall, basic knowledge about the EMF issue to an ordinary telecommunications engineer who does not in his or her work have to specialize in the area. Space restrictions deny a deeper handling of the issue, so for a more detailed view the reader should refer to the plentiful literature published about the area.

The Telecommunications Handbook: Engineering Guidelines for Fixed, Mobile and Satellite Systems, First Edition.
Edited by Jyrki T. J. Penttinen.

Figure 24.1 *Antennas of mobile networks in a typical mast installation. Notice the tilting of the topmost antennas. Antenna height is enough to cause no RF radiation hazard to general public on the ground, but antenna fitters must take proper actions to avoid overexposure from neighboring antennas. Photo by Jouko Rautio.*

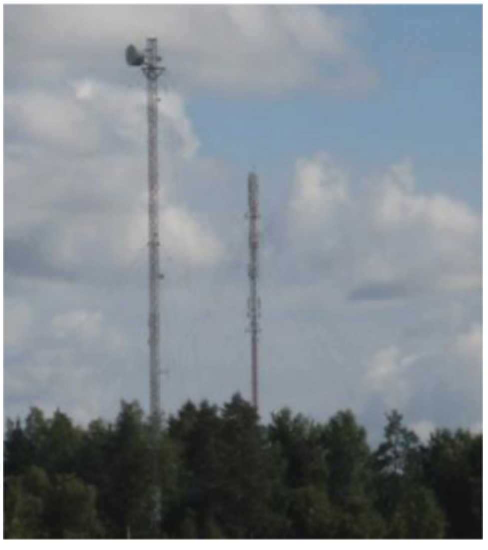

Figure 24.2 *For various reasons (land price, topology etc.) it is not uncommon to have two masts close to each other. When working in a mast one must also pay attention to a possible RF radiation from the other mast, especially when antenna height matches the working height. Photo by Jouko Rautio.*

Figure 24.3 *Installing antennas on the edge of a roof is a preferable way when looking the radiation safety, as we have here. Notice also the VHF-antenna on the short pole. Photo by Jouko Rautio.*

Figure 24.4 *Not an ideal way for antenna installation. Part of the antenna is above the roof level and the radiation can penetrate people working on the roof. Fortunately in a sector antenna the back lobe is only about 1/100 of the main beam. Photo by Jouko Rautio.*

24.2 The EMF Question

Radio waves have been part of the life of the mankind for over 100 years. Various wireless technologies – including sound broadcasting, television services, navigation radars and so on – have helped the life of ordinary people in ways which could have been seen as witchcraft before the invention of radio transmission. One can travel with a boat at mid-Atlantic and make navigation decisions based on weather forecasts delivered by a radio broadcast. One can talk to a friend on the other side of the world while sitting in a moving vehicle, perhaps both in cars. One can cut a foot with an axe deep in a forest, make an emergency call with a mobile phone and be transported to a hospital within half an hour.

Yet as these examples show how the radio technology has benefited safety, well-being and matters of health – even in life and death situations – there are also some who ask the EMF question; do the radio waves possibly have a negative effect or effects to human health? It is a universally acknowledged fact that above a certain level of RF exposure (SAR) to living tissue there will be negative health effects, caused by the rise in tissue temperature (which also are acute; the damage appearing very soon). But some people ask if there are other effects, below the level which cause aforementioned thermal effects? Are there negative health effects, which manifest themselves only after years or decades (like brain cancer)? Is there a cumulative effect? Are some people more (hyper) sensitive to EMF than the majority of the population and who might develop a multitude of symptoms varying from loss of consciousness to a mild itch or nausea?

The EMF Question is widely debated in various arenas, not least on the Internet with many discussion forums and even EMF (or rather anti-EMF) dedicated sites run by activists. Numerous nongovernmental organizations (NGO) are sure that the answer to the EMF Question is an overall lowering of exposure limits, banning cellphones from certain subpopulations or places and wide exclusion zones around “sensitive” locations like schools and kindergartens. The rational view is often overshadowed by the loud message from the so-called “self-appointed experts” who do and promote low-quality research or plain superstition. The general public and even the telecommunications engineers are often unable to understand the true (or lack of) values of these as the science is not a familiar thing, and also they do not necessary know how the current regulation is formulated.

24.3 The Scientific Principle and Process: The Precautionary Principle

An ordinary telecommunications engineer probably has not a good knowledge with the (biological) science and the scientific principles and process, so a short summary about these is welcome. This text will especially concentrate on the research for mobile systems, as the public debate and media attention has focused on this area in recent years. However, the general ideas do apply also for other RF research (and non-RF EMF).

Before we go to the actual laboratory tests it is good first to point to the general ethic rules and culture for the scientists. The work and conclusions (“judgments”) of the scientist must be academically independent. This means that for example those who finance the research cannot say what the result will, or must be, either to limit some aspects out of the study or prevent publishing the results (no matter what they are). Also the researchers should be free of any bias, predetermined views or political goals in their studies; there have been real life examples of how a researcher has falsified a test specimen and/or statistics to prove that mobile phones are a health hazard (e.g. DNA breaks). Also in some studies there are situations in which the researcher must dedicate special attention to the privacy, security and safety of test persons and/or data.

The need for security and control can also be seen as part of the overall need for control of the studies. Every parameter and possibility must be controlled, and even trivial things can interfere. We have a sad real life example, when a cell culture test showed some unexpected results; it transpired that during the night the cleaning lady had opened the windows in the laboratory for fresh air, thus ruining the needed stability in temperature of the Petri dishes.

In EMF science particular the control of dosimetry is very important. Even small changes in the dose of RF exposure or the distribution of the dose can significantly influence the result of the study. This fact is especially important when the question of the existence of possible nonthermal (athermal) effects is probably the main question in EMF issue. Several real life examples show how the uneven RF dose distribution in Petri dishes (especially on the edges) can cause disproportionally big effects on the results, sometimes the error noticed only years after the publication by other scientist. There are also sad examples of Electro hypersensitivity (EHS) provocation studies in which the RF exposure was from a mobile phone connected to a commercial network (although the DTX, the dynamic range 1:1000 of the power control and random changes of the level and so on make the dose estimation impossible).

The starting point for a good scientific study is a hypothesis of supposed mechanism. For example, if you have a theory how exactly the RF field provokes EHS symptoms it is much more exact to arrange dose and exposure periods. If one has no mechanism it is also difficult to assure the scientific commune that the result is something else than an anomaly.

The main principle in science is that the results must be able to be replicated. What this means is that if you have a study with certain arrangements and results any other researcher who does a study with similar arrangements should have same results. There can be some tolerances in this because of the statistical nature of some biological processes but in general the results must be same. The replication is also expected to be done by someone else than the original researcher. It is OK to replicate your own study (if for no other reason, because of typos) but to be really accepted by the science community at least 2–4 successful independent replications by different researchers are expected. In science the “false positive” and “false negative” results (“negative” meaning no find or no effect) are natural phenomena (because of many parameters, statistical variance and biological processes some results differ from the majority and/or the scientific consensus even if no actual mistakes happened in the research).

The need for replication is an especially important thing to notice in EMF science as many nonreplicated studies with alarming results easily find lot of media attention and worry laypeople. Typically, when the replications are made there are no such results to cause alarm but “good news is no news” and do not create many lines or time in the media.

The reader should also remember that in science the zero-result is as good and valuable as any other result. It is in human nature to give more attention to those which are something else than zero, but in science this tendency must be avoided. If we have one result with nonzero result and three replications with zero results the scientists must make a conclusion without weighting a single anomaly too much.

Statistics play an important part in science and especially in EMF topics, not only in epidemiology but also in other studies. One has to have a certain number of samples to have statistical weight and reliability. This works also in the opposite way; if one picks a small enough subpopulation the random effect plays a significant part. A well-known folklore in Finland tells that the radar technicians have many more daughters than sons; when we look at this claim we see that the subpopulation of radar technicians is only a tiny fraction of the male population (less than 100 men) and the tale has existed for only a limited time (there were no radars before World War II), so the statistics in 50 years from now may show that radar technicians have more sons and in the whole 100 year period there has been a 50–50 proportion of genders. Enough test persons (statistical weight) is important in EHS studies, and statistical weight play an important part in many other areas too so when looking for news about EMF research the reader should remember this.

At the end of the scientific process the results of a study must be published in a scientific journal. There is a hierarchy between different journals, as others are considered more prestigious than others. A paper about results will not get published too easily, thanks to the scientific peer-review. This means that preferably at least two or three experienced scientists of the area will go through the paper and check whether the execution of the study and writing of the paper follow the scientific principles and rules, and of course that there are no errors in the study. The best journals are highly selective and the number of accepted papers can be just a small percentage of those offered.

Publishing in a peer-reviewed journal is a very important thing for the acceptance of a study and results in the scientific community and for example many expert organizations recognize or give much more value to only those when they do their evaluation of state of science and possible conclusions (this applies both EMF area and science in general).

But as the peer-review is a strict process and many papers do not reach the pages of journals, in recent years more and more of the unpublished texts have been put available in the Internet. The reader should notice that, for example, a search on the Internet about certain EMF subjects can result in tens of scientific papers, many of which have actually not passed critical standards.

On the EMF debate one can see references to the Precautionary Principle (PP), mostly by those who do not consider the existing scientific data is enough to be sure that the current EMF exposure limits protect the population (especially for long-term effects and/or possible special subpopulations like EHS persons). What exactly is meant with the PP is expressed in many forms by many writers, but the following couple of quotes will be illustrative enough: Rio Declaration (by 172 countries) from 1992 states:

> "In order to protect the environment, the precautionary approach shall be widely applied by States according to their capabilities. Where there are threats of serious or irreversible damage, lack of full scientific certainty shall not be used as a reason for postponing cost-effective measures to prevent environmental degradation."

And the European Commission stated in 2000:

> "The precautionary principle applies where scientific evidence is insufficient, inconclusive or uncertain and preliminary scientific evaluation indicates that there are reasonable grounds for concern that the potentially dangerous effects on the environment, human, animal or plant health may be inconsistent with the high level of protection chosen by the EU."

The PP was formulated mainly the chemical agents in mind. Regarding the PP to EMF issues is based on the claim that we as yet have insufficient amount of research data about the effects of RF radiation, especially long-term. The PP supporters call wide range of EMF regulation changes, like overall lowering of exposure limits, large exclusion zones around base stations or "sensitive areas" (e.g. 500 m to all no matter what power is used), ban of using cellphones for children (allowed age varying from 7 to 16 years) and so on.

However, the reader can note two things about PP regarding EMF; firstly we have already over 2000 peer-reviewed published studies about RF exposure to biological tissue (over 800 of them in the mobile system frequencies) for a period of over 60 years. The scientific consensus based on these is that there is no health hazard within the existing radiation limits although some minor gaps in knowledge exist. The second thing to note is that actually the existing regulation in EU and USA has precautionary aspect included, as the limits are not set to the point of "just below damage" but have a wide safety margin (one does not get burnt even if the SAR value of a mobile phone is 2.1 instead of regulation maximum 2.0 W/kg).

24.4 The Expert Organizations and Regulation

Even with the principles of the science are rather simple and straightforward, the process of evaluating the scientific evidence and putting up recommendations and guidelines for real life people needs expertise. One must be able to value the hundreds of results from studies of varying quality, and on the other hand to be able to see the needs of real life, industrial possibilities and even the views – perhaps prejudices – of the population. Even though there are several countries which for political or historical reasons have their own limit values for EMF regulation, most of the countries today completely or partially follow the recommendations set by the ICNIRP and IEEE.

The International Committee for Non-Ionizing Radiation Protection is an independent nonprofit scientific body, which is not associated with any country, industry or political movement. The members of the main commission (14), The Scientific Expert Group and Project Groups are scientists, who represent themselves, not the institute or country where they have their actual day job. An extra limitation is that the members cannot be employed by any commercial company to further guarantee neutrality from business interests. Most come from universities and national radiation protection agencies. Founded in 1992, the ICNIRP was separated from the International Radiation Protection Association, which since has concentrated to the ionizing radiation. The reader should note that as the ICNIRP deals with nonionizing radiation (NIR) it is not limited to RF only but also to optical area (e.g. laser equipment, ultraviolet tanning devices) and low frequency area (e.g. power lines, trams) and so on. The ICNIRP also works for ultrasound safety. It is headquartered in Germany and thus their net site has the respective land code [1].

The ICNIRP evaluates the current body of published research on EMF (RF and others separately) in irregular interval (years) and builds its opinion about state of science. Without going to details, it can be said that the process will look at each study on how they are made, if they follow good scientific principles and so on and how the results stand compared to the existing scientific consensus, and if there should be changes in recommendations.

The Institute for Electrical and Electronics Engineers (IEEE) has half its members in the USA and develops EMF limits for USA, though those are also used in Canada and some other countries.

The World Health Organization (WHO) is one of the specialized bodies under the United Nations. Mandated to develop health in all issues it also has had interest in EMF. WHO has its own experts and works in cooperation among others ICNIRP and the International Agency for Research of Cancer (IARC).

International Telecommunications Union (ITU) is the oldest of the specialized organizations under the UN. The ITU has focused lot of its development work on the Third World countries and one of the topics in there is also EMF, as many African and Asian-Pacific countries lack expertise and resources in radiation safety issues and regulation.

The European Union (EU) has no mandate to EMF regulation but it is done in national level. However, the EU has made a recommendation about the EMF exposure in 1999 and this has been implemented at a national level in most member countries as a whole or with minor exceptions (in some Eastern European member countries regulation is still based on Soviet decrees).

The European Commission has mandated the Scientific Committee for Emerging and Newly Identified Health Risks (SCENIHR) to look some new technologies, among others EMF. The members of SCENIHR are all scientists from different areas. They have after 1999 done several auditions of new research result to see if changes should be made to the recommendation, but have not seen reason to do any.

It is also worth noting that the EU has the mandate on occupational health including EMF protection in that area. A new directive about workers exposure is expected to come in 2014 but as this is written there are no probably major changes are expected in the RF area.

At a national level there are different things in different countries. Some countries have their own specialist radiation protection agencies, like SSM in Sweden, STUK in Finland, ARPANSA in Australia and BfS in Germany. In some countries the EMF issues fall with the health authorities, like FDA in the USA and HPA in the UK. One example is Lithuania, where there is radiation agency but the EMF license issues and so on fall in practice with the health officials. There are also still quite a lot of countries (especially in Africa and Asia-Pacific) which lack specialized EMF officers but rely on international support on this (e.g. ITU).

In the EMF debate references to the BioIniative can sometimes be seen. To clarify to the reader it is worth mentioning that the BioIniative is not an organization per se nor an official or governmental body but a loose group of persons with various backgrounds (originally about a dozen persons). Some of the members are university scientists who do not agree with the scientific consensus but also others are involved, including environmental consultants and activists. The BI report has provoked criticism from scientists and some health authorities, like those in Netherlands and Australia, who have considered it “unbalanced and unscientific.” However it is widely spread in the Internet and other media as an alternative to the established view and a basis for demanding lower exposure limits.

As exposure limits are set by expert organizations and national regulators, equipment manufacturers can make their products according to the rules. However, there is still some room for choices, and to ease the production process and also to guarantee similarity for consumers, standardization is deemed necessary. This is mainly done by two international organizations, which also do standardization in many other areas (including EMC); CENELEC and IEC.

Comité Européen de Normalisation Électrotechnique (CENELEC; European Committee for Electrotechnical Standardization) is working with national EMF subcommittees in European countries. The International Electrotechnical Commission (IEC) has headquarters in Geneva and develops standards for 84 member countries and 84 Affiliate Countries. In recent years the CENELEC and IEC have increasingly harmonized their standards and probably more so in the future.

Setting standards for EMF includes one for measuring SAR values for mobile phones (of which more later). This is of significant importance, as for example the phone’s position in relation to head and the distance from the tissue influence much the result, and therefore a common standard is needed to ensure that all results from different laboratories are comparable with each other’s. The reader might be interested to hear that it took several years to agree this standard; a testimony of the fact that even minor details have great importance in the measurement process. There are also standards for putting base station equipment into use and for BS EMF measurements.

All the organizations mentioned above work for EMF radiation protection in civilian life. The military have their own bodies for EMF issues and also rules and exposure limit values can differ from those for civilians. In many armed forces they follow the rules set by the North Atlantic Treaty Organization (NATO), including several non-NATO countries. There are many of situations for RF exposure (e.g. high-power radar antenna near warship’s bridge) but EMF issues in military life will not be detailed in this book.

24.5 Some Topics of the EMF Debate

We now can have a short look at some topics in the EMF debate, including some historical perspective how they have evolved. As in the media some sensational news about EMF research easily gains much attention it is good for the reader to have a little background.

24.5.1 Cancer

Probably the best known (speculated) health hazard from the radio transmitters for the great public is cancer. A brain cancer caused by the use of mobile phones has been an established part of public debate since the 1990s. Even popular television shows may have in their dialogue references to the cancer by cellphones. Much of the blame for this widespread myth can be traced to early 1990s, when a couple of court cases in the United States were widely reported in the media, including detailed claims of the plaintiffs accusing mobile phones for their tumors. The judges did not find the evidence for connection between cellphone use and cancer good enough, but that did not stop many lawyers from starting new cases with other cancer victims. Also, there have been lots of speculation about possible causal relationship with phones and cancer in the scientific community and even more in the Internet and other popular media. The reader should also notice that developing better cellular networks with new technologies has paralleled the similar type of development in medical equipment technology. This means that even more cancer cases have been diagnosed including ones that with older type equipment would have not been detected (notice that an undetected tumor can spread unhindered to other organs). The better medical technology has partly created an illusion of the increase in number of tumors, as more undetected ones can be found nowadays. However, even this trend has not followed the exponential curve of mobile phone user number but more like a straight line.

A big problem in finding or not a causal relationship between mobile phone use (or living near a base station) and cancer is the latency period. In medical terminology this means that there is a certain time between the exposure to the causal agent (for example, radioactive radiation or a carcinogenic chemical) and the appearance of the tumor. In the organs in the human head the latency periods of the tumors are typically 10–20 years. What this means in the EMF science is a certain problem, because the mobile phone at first was an elite gadget and only became a device for mass consumption during the late 1990s. Add to this fact the 2–5 years delay until epidemiological data is available to the scientists and we have the problem that our data does not yet cover well the mass usage time, meaning that the statistical weight of the results is not as good yet as one would hope. Brain cancers are relatively rare and thus a big mass of the population is needed to find out if there really are more cases since the spreading of cellphones. However, the data is at least giving more and more light to the problem and the results have been mostly pointing to same conclusion; no connection.

The possible causal relation between EMF exposure (cellphones and base stations) and cancer has been studied in several epidemiologic research programs in many countries. Until now the biggest one was INTERPHONE involving 13 countries, including Japan, Scandinavian and West European countries. The result of the program was “no causal relation.” In the CEFALO program (in four European countries) a possible connection between cellphones and cancer was studied especially in children of 7–19 years of age; again no connection was found.

In May 2011 the International Agency for the Research of Cancer (IARC) classified RF field exposure as 2B “possibly carcinogenic.” This was influenced by the findings of one Swedish research team; although the main body of the research data suggests that there is no connection; also the aforementioned latency period problem was noticed. When looking at this the reader should notice that in the four class system 2B is by far the most numerous of the classes (Class 4 “Not Carcinogenic” includes only one agent) and to help put it into the right perspective is the fact that Class 2B also includes such everyday agents as caffeine (coffee, cola-drinks) and pickled vegetables. Also, within a month the World Health Organization (WHO) issued a statement saying that there is no evidence of RF causing cancer; this was done to clarify the IARC classification to the great public.

24.5.2 Electro Hypersensitivity

Modern society has seen the rise of several environmentally related illnesses; that is a rise to the consciousness of the public at least. Some of them are highly disputed whether the unspecified and varying symptoms actually are (if real) manifestation of a medically proven disease. So-called “Electro hypersensitivity” (EHS) is one of those new diseases, a newcomer. The first widely known cases of what was later called EHS appeared in the latter 1980s, when some office workers connected various different kinds of symptoms (not necessarily the same ones in everyone) to their work near computer screens. As personal computers had just been introduced to office life people had no previous experience about them nor if any occupational health problems appeared with them. While many cases of obscure symptoms could be traced to the fact that sitting hours in front of the computer screen have effects to the shoulders, neck and arm muscles and to the circulation (including that in the head), there were some people who associated their symptoms to the electromagnetic radiation emitted from the screen and processors (even if that is very weak).

As this view spread, more people started to suspect themselves to be hypersensitive to electricity. Of course, some entrepreneur-minded businessmen saw an opportunity to sell devices to limit or completely remove the EMF radiation and the advertisement campaigns raised the public knowledge of the so-called new disease further. At the same time some people started to claim that they had EHS symptoms also from other sources of EMF, such as from electric wires or fluorescent lights. Enter the mobile phones with their base stations in the late 1980s to early 1990s and some people claimed to be hypersensitive to the RF fields from them (the reader should notice, that the EHS patients do not necessarily claim symptoms with all aforementioned devices; for example, one can have symptoms with a cellphone but not with a fluorescent light and vice versa).

At this point the reader may be a little puzzled that 1000+ kW television and sound broadcast transmitters caused no EHS symptoms during their 70 years of operation! Anyway, the rapid spreading of mobile technology paralleled the fast spreading of the Internet, which allowed people to speculate with stories (true or false) about EHS also bypassing the traditional and perhaps more critical media (newspapers, medical journals etc.). As mentioned, the EHS patients complain a wide array of symptoms (e.g. headaches, dizziness, nausea, skin rash etc.), not every symptom for every patient, and so it was easy for people reading about the stories to identify at least some symptoms in themselves, especially those living near base stations.

"The placebo effect" is a well-known phenomenon; one can give a patient some pills and tell them that they are a medicine for the patient's illness and soon the patient is healed, even though the pills were actually just colored sugar. Patients believe that they will become healthy again and they do. But in a similar way it is possible that they think that some chemical or technical agent – perhaps an RF field – does they harm and they get symptoms. This is called "the nocebo effect." The author himself during his career in a mobile operator company faced several times the following chain of events: (1) a contractor lifts up a mast in the middle of a village; (2) a person calls the mobile operator and claims to get EHS symptoms because of the RF radiation from the mast; (3) a second contractor builds the base station equipment in the mast and the transmitter starts to send RF field! Naturally, one would have expected that there must be an RF field before EHS symptoms appear (3 before 2). The delay caused by using two contractors clearly showed the nocebo effect in practice in real life. One cannot see the radio waves but one can imagine them and imagine that they make one ill.

But as these kinds of real life events show that (at least some) EHS patient's symptoms correlate with nocebo effect, there are also tens of scientific studies about electro hypersensitivity. These have been made in major universities in European and American countries as well as in some minor institutes and even by some research groups with ties to pro-EHS NGOs. Typically the provocation studies have involved 10–30 test persons (self-claimed EHS patients) but the biggest ones have had about 200. The vast majority of the studies show that EHS patients as test persons cannot tell the difference between existence and nonexistence of RF exposure, and in some studies they even got more symptoms in the test period when the RF field was not transmitted (but because of the test arrangements they thought it was).

The scientific community is of the opinion that there are no EHS symptoms from RF sources but that they are of psychosomatic origin. The World Health Organization (WHO) has announced about EHS that "the symptoms are real but they have no connection to electromagnetic fields."

24.5.3 The Children's Issue

What is meant by "the Children's Issue" in the EMF debate is the speculation that infants and adolescents are more easily harmed by nonionizing radiation, especially at levels below the current exposure limits. This concentrates mainly on two things; a more easily (RF-) penetrated skull (and other tissues) and the growing nerve system (also child cancer is widely debated). There is still too little research data on both these issues and thus room for various hypotheses. Some scientists say that children have in their heads a skull and other tissues which are more easily penetrated than those in adults. On the other hand there are scientists who say the opposite – that children have actually thicker and less easily RF-penetrated skulls and tissues compared to adults. The thinner-skull party has attracted media coverage by publishing computer generated images on how radiation penetrates child's head; however these have been based on computed models of penetration and absorption in different tissues with many parameters still open to debate.

The growing nerve system is a complex biological structure which not only includes the long nerves connecting different organs but also the spinal cord and the brain. In humans the growing period can take over 20 years. Thus there has been speculation that exposure to EMF (like starting to use cellphone at seven years old or living near a base station) can cause harmful changes in the growth process. Many activist groups have successfully directed a lot of their communication especially to parents, so the worry about the Children's Issue has spread widely. The debate has been highly speculative as there are not many research results about the issue. Fortunately in recent years some studies have been published and more will probably follow in the coming years. For example the University of Kuopio, Finland published their animal study of growing nerve system exposed to RF in 2007; no effects were found. The aforementioned CEFALO about possible connection between child cancer and mobile phone use also had a negative result (published 2011). The Children's Issue debate has been largely based on the lack of studies but little by little more is coming available.

24.5.4 So-Called Funding Bias

In the EMF debate one can sometimes see references to the so-called "funding bias." In science the "false positive" and "false negative" results are a natural phenomenon (because of many parameters, statistical variance and biological processes some results differ from the majority and/or the scientific consensus even if no actual mistakes happened in the research). However, there are claims that in those studies funded by the mobile industry, there are more results which show no effect from exposure to RF fields (a "negative") than all results and especially those with no industry funding at all. There was, for example, one article which claimed that industry funded studies found positive results in 33% of them but nonindustry funded ("public") found positive results in 88% of them! This is called "funding bias" and at first glance seems to point out that the industry is doing something to cover a massive health menace.

But when we take a closer look, things are not so simple. There are a couple of things one should remember. First, the mobile industry can be said to have money – at least more money than volunteer based NGOs – and thus can hire professional experts to evaluate and choose between research proposals, and can give significant funding to best research. Usually there is a firewall between the financier and researcher (third party body) to guarantee the freedom of the scientist. Industry funded research is looked at constantly with a magnifying glass by the public, activists, media and authorities and therefore scientists must follow the rules of ethics and science precisely. Typically these studies are very well-done and published in peer-review science journals.

The second thing to note is that if you remove the industry funded studies from the whole picture of EMF research, all the rest are not actually "public" either. While some truly are 100% government funded some are made by well-wishing but science-ignorant zealot individuals (even lawyers by education!) and quite a lot are studies which have ties to the activist groups (NGOs). These groups cannot be regarded as being neutral, as many of them publicly announce their purpose to get benefits (government money or compensation from the industry) to their members because of EHS.

For example, one Nordic EHS study was done by an emeritus-professor of medicine, who seemed to follow scientific practices and was published in a science journal; however, in a TV clip from the field tests one could see the chairman of the national EHS association as a research assistant (not credited); the results of the study claimed that EHS was a real existing thing! Other studies can be traced to be financially linked with consultants for "EMF insulation for houses," or to manufacturers of shielding devices for cellphones (apparatus which covers the antenna and reduces radiation to the head – as well as to the base station), in other words to those who will increase their business when more people are worried about EMF.

The reader should also notice the fact that many of the best researchers already consider the EMF question answered (= no negative health effects exist below current exposure limits) and have started to study different areas of interest (non-EMF), so that much of the newer EMF research is done by others including the aforementioned ones.

So when we see mentions about "funding bias by the industry" we should be aware that it is more like an illusion and perhaps we can even say that it is a reflection of the funding bias in nonindustry funded research.

24.5.5 Ana-Digi

In the EMF debate it is often claimed that digital systems (e.g. GSM, WLAN) are more dangerous than analog ones (e.g. NMT, AMPS) because of the nature of the modulation. The reader should know that this claim first appeared in the late 1990s or early 2000s, when the main body of the cumulative EMF research knowledge started to point out that mobile systems are not a health hazard (with existing regulation). Because analog systems had existed longer most EMF research on cellphone systems had naturally been done with them. Lack of research data on digital systems opened a window for speculation about the effects of RF modulation. Another factor to give fuel for the claim is that ordinary citizens can actually hear the effect of a digital phone, as they cause typical "r-r-r-r" sound to many electronic devices like car radios (unlike an analog phone). This is naturally a thing which can puzzle a person with no radio engineering background.

However the claim of the different effects because of modulation type does not have any factual basis. Several studies have been made about this. During the last four years there have been no less than three separate independent, international groups of scientists which have looked through the results and have announced that modulation is not a factor for any health hazard.

24.6 SAR

Probably the first technical term involved with the EMF debate which a radio engineer sees is the SAR. This term is used to give a numeric value to the radiation a mobile phone gives to the user, and is in many countries demanded by the regulator to be published for the handset buyers.

When thinking about radiation from radio transmitters, we normally suppose situations where the absorption (or measurement) is done in the far field. This is the "normal" EM field, which is in density inversely proportional to the second power of the distance [1/(r x r), when r is the distance]. However, when we operate extremely close to the absorbing tissue, as is the case when having a mobile phone's antenna only millimeters from the human head, we are actually dealing with the near field. This is a situation where also elements inversely proportional to higher powers and the shapes (dimensions) of the antenna relative to the radiated object have a significant part on the overall strength (for a closer look on the far field and the near field the reader should see other parts of the book). Thus thinking about exposure to tissues in the human head (or other parts of the body very close to the antenna) we must have a term which better describes the strength of the absorption; how much energy is going to the tissue. For this purpose the term Specific Absorption Rate is used, commonly known by the abbreviation SAR.

Local SAR is given by

$$SAR = \frac{\sigma E^2}{\rho}. \tag{24.1}$$

Where we have the electric conductivity, E is the electric field and as divisor we have the mass density.

The Specific Absorption Rate is defined as the amount of energy absorbed by the mass of the biological tissue, and therefore is a very specific term related to the tissue in question (muscle, fat, eye etc.) because of their different conductivities. The term is also dependent on the elements and shapes of the radio transmitter/antenna (in practice the mobile phone), position relative to the tissue (head, hand etc.) and also the distance. These points, however, will not be detailed in this book. The reader should just know that the SAR is a unique, measured value for every model of mobile phone. The measurement position is today an internationally standardized one to make sure every model/brand is measured in a similar way and thus the values are comparable to each other. For obvious reasons the measuring of EM fields inside the head are not done with real ones but using a phantom head (with artificial liquids mimicking electrical properties of head tissue). The SAR values are measured by mobile phone manufacturers themselves, but also by some governmental agencies and consumer NGOs. For example the Finnish Radiation and Nuclear Safety Agency (STUK) have made SAR measurements for randomly selected samples of cellphones since 2003, no SAR values exceeding the regulation limit have been found. Typically also the values found by the agencies fall within +/−15% of the one published by the manufacturer Figure 24.5 shows and example of SAR measurement setup.

In many markets the national (or regional) regulation demands that the SAR value for a mobile phone model is displayed on the package or in the user manual. This is typically the only EMF parameter of the phone available for the customer.

The maximum SAR values of mobile phones are set by national regulations, but typically follow just a couple of international examples. In the European Union the Council made in 1999 their recommendation on EMF, including also the maximum of 2.0 W/kg for SAR for the head. The United States has set the head SAR maximum at 1.6 W/kg. Most EU countries have since implemented the 1999 recommendation to their national regulation. The EU and USA values for SAR have also been adopted in national regulation

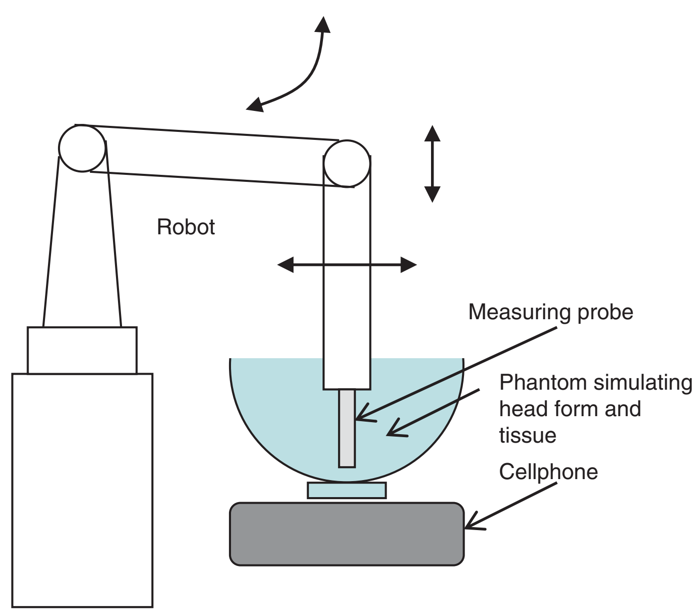

Figure 24.5 *Simplified test arrangement for measuring SAR with a human half-head phantom. The phone is connected to a base station simulator, which guarantees constant radiation level during the measurement. Whole operation takes place in a radio echo-free room. The robot is computer operated, and the liquids inside the phantom can be made to simulate different human tissues.*

in numerous countries in other continents (for example India, Australia etc.) and in practice today cover almost the whole world. The basis for setting the maximum around 1.6–2.0 W/kg comes from numerous EMF studies, which found out the SAR to cause a predefined temperature raise in the tissue, and added a considerable safety factor. In EU the whole body SAR is set in 0.4 W/kg. The reader should note that the limits for EMF exposure (including SAR) guide all situations and equipment, not only those in telecommunications but also for example civilian radars, microwave ovens, RF drying equipment, antitheft devices and so on.

The term SAR should as mentioned earlier be used only in situations very near the antenna, in most (far field) situations one can go with the power density (W/m^2) or field strength (V/m). In practice the SAR is used with the mobile phones, not with for example the base station antennas. Rare examples to use SAR with base stations can be found in some cases of occupational exposure, when mast workers are accidentally near a live antenna (or touching). However, as the antenna size and shape relative to human body is a significant factor in this kind of non-hoped-for case it is strongly recommended to leave a more detailed estimation of the exposure to the specialist radiation safety authorities.

The SAR is in practice the only term which is associated only with EMF issue, as for example the term "power density" is also used in other topics (for example, reception conditions). It is also a clear, numeric value and is therefore popular among the debate of EMF. However it is limited mostly to the thermal effect of the RF exposure and thus relates rather little to the main debate about (speculated) EMF health effects.

24.7 The Safety Distance and Installation

A powerful radio transmitter is a potential nonionizing radiation hazard if one is too close to the antenna. The area around the antenna, which should not be entered, is called the compliance boundary or exclusion zone. The area depends on transmitter power and also on the gain and radiation pattern of the antenna. Also, as the NIR exposure limits are different in different frequencies this is a contributing factor too. Here we will focus on mobile systems. The reader should also note that exposure limits for occupational exposure are higher than those for the general public (e.g. in EU 5 times more) so the safe distance is also different (closer to the antenna).

Please note that "the Exclusion Zone" as a term is also used in the EMF debate for a wider area around base stations based on precautionary ideas, especially by pro-PP activists. The demand for this type of exclusion zone calls for a uniform (randomly picked) distance to every direction from the antenna, inside of which there should be no inhabited buildings in general. More often exclusion zones are demanded around "specific areas," meaning schools, hospitals, kindergarten and so on, inside of which no base stations should be allowed. Most commonly the proposed zone radii are either 300 m or 500 m. As we can see below, for cellular base stations the calculated safe distance based on rational, science-based exposure limits is only some meters, relates to the transmitted power and also depends on the direction (main lobe – side lobes). The PP-type exclusion zone relies more on psychological factors and simplicity. However, in some countries they are included to national or regional regulation because of political pressure. The reader is advised to be cautious when seeing the term Exclusion Zone in a text; whether a PP-type or rational type is the question.

In the far field exposure is expressed either as power density or electric field strength (unit V/m). The power density (S) – or electromagnetic power flux density – depends on the following parameters:

- Distance from the radiating antenna
- Amount of TRXs
- Transmitting power
- Antenna parameters (gain, tilting, radiation pattern)
- Specific attenuation (cables, splitters etc.).

The Power Density (S) towards main beam can be calculated as depicted in:

$$S = \frac{NP * 10^{(G-L)/10}}{4\pi r^2}, \tag{24.2}$$

where

N	=	number of TRXs
P	=	average TX power for each transmitter
G	=	Antenna gain

For calculating other directions than the main beam the antenna gain should be adjusted according to the (horizontal) radiation pattern. When tilting is used this can be seen as utilizing the vertical radiation pattern of the antenna.

Should we insert Maximum Power density values as depicted in Table 24.1 into Equation 24.3, the minimum distance can be calculated as below:

$$r_{\min} = \sqrt{\frac{NP * 10^{(G-L)/10}}{4\pi S_{\max}}}. \tag{24.3}$$

By using GSM 900 and GSM 1800 base stations as examples we can calculate the following safety distances (r_{min}) as presented in Table 24.2.

As can be seen, the minimum safe distances vary at about 2–4 meters for public and about 1–2 meters to occupational exposure. These distances are for the main beam and for the side lobes much smaller. These kinds of figures do not pose a great nuisance to people working near antenna on a roof of a building, but one has to remember that in some countries there can be a big number of transmitters in a pole, and also there can be bigger transmitter power and antenna gain. Also, with the introduction of the LTE and the HDTV the safety distances can grow considerably.

When we have the base station equipment in a mast the antenna is typically so high (10–30 meters) from the ground that the general public has no possibility to enter the compliance boundary. Also, with transmitter power of only 10–30 W EIRP mast workers are safe if they do not climb closer than about 1 m to the antenna (however, there can be other transmitters colocated in the mast). Today it has become more and more common to use a manlift to help in mast work, and with that it is possible to accidentally put the mechanic to the main lobe too close to the antenna. This danger is of course more obvious on masts with numerous antennas, especially if some of those belong to more powerful systems like LTE and TETRA (or broadcasting). It is advisable to turn off all transmitters during the manlift work, or at least below agreed maximum operating height.

In urban areas base station antennas are usually installed on roof (often on a short pole) or outer wall of a building. As the antennas used are directional (with side lobes //1/10–11100 of the main beam) and the material in the roof and the wall causes attenuation the RF field in the room under or behind the antenna is

Table 24.1 *Maximum power density values on some mobile systems frequency bands (ICNIRP based)*

Frequency, MHz	General public, W/m²	Occupational, W/m²
900	4.5	22.5
1800	9.0	45
2100	10.0	50

Table 24.2 Examples of the safety distances for GSM 900 and 1800 MHz

	Amount of TRXs	Pmax (W)	Pmax (dBm)	Gain (dB)	Lmin (dB)	Public exposure limits		Occupational exposure limits	
BTS type	N	P^a/TRX	P^a/TRX	G	L	S_{max} (W/m^2)	r_{min} (m)	S_{max} (W/m^2)	r_{min} (m)
				GSM900					
outdoor	2	31.6	45.0	12	6	4.5	2.1	22.5	0.94
	4	31.6	45.0	12	6	4.5	3.0	22.5	1.33
	6	31.6	45.0	12	6	4.5	3.7	22.5	1.63
indoor	1	12.6	41.0	7	6	4.5	0.5	22.5	0.24
	2	12.6	41.0	7	6	4.5	0.7	22.5	0.33
	4	12.6	41.0	7	6	4.5	1.1	22.5	0.47
				GSM1800					
outdoor	1	5.0	37.0	18	6	9	0.8	45	0.37
	2	5.0	37.0	18	6	9	1.2	45	0.53
	4	5.0	37.0	18	6	9	1.7	45	0.75
indoor	1	5.0	37.0	7	6	9	0.2	45	0.11
	2	5.0	37.0	7	6	9	0.3	45	0.15

[a] TX Power before combiner.

weak and clearly under the maximum allowed exposure limit. Real life measurement by STUK show typically values of 0.03–0.1% of the regulation maximum in these kind of situations.

It is recommended that the antennas are installed so that exposure is very clearly under regulation maximum value. Also, it is advised that the antenna cannot be reached by members of the general public. Especially in wall installations the planner should make sure that the antenna cannot be touched from a window or a balcony. If the antenna is within easy reach for ordinary people there should be a warning sign near it. The reader must also note that there can be different regulatory requirements for warning signs in different countries and should familiarize themselves with the rules in the installation place, especially if the site is in another country than the planning office.

The laptop computers are nowadays connected to wireless networks with WLAN, RLAN and Bluetooth techniques. These use transmitter powers which are weak: typically WLAN 0.1 W, RLAN 0.2 W and Bluetooth 1 mW–0.1 W. The base stations for these networks also have radiating powers considerably lower than those used in cellular base stations; 1–4 W maximum, typically even weaker and installed on roofs or walls. The exposure is clearly under regulatory maximum and need not be discussed in more detail here.

As mentioned at the very beginning of this chapter, the RF area runs from 10 MHz to 300 GHz. Below that band lies the Intermediate frequency area in which the EMF effects have a somewhat different influence mechanism. The telecommunication and broadcast bands of LW, MW and SW will thus not be discussed here. The reader is advised to refer to literature dedicated to them.

On satellite ground stations and radio links the antenna gain and possibly also the transmitter powers are higher. On the other hand, higher antenna gain means that in other directions than the main beam the radiation is relatively even weaker compared with mobile base stations. Radio links with "pencil beams" should be installed on walls or edges of the roof when they are installed on buildings. Figure 24.6 presents an example of the fence for indicating the protection distance.

On FM broadcast transmitters and VHF & UHF Television transmitters we have the opposite situation, as the antennas are typically omnidirectional (on horizontal radiation pattern) and thus have low gain. On the

Figure 24.6 *A fence around the antenna of a satellite ground station is a good idea also from the RF radiation safety perspective. Photo by Jouko Rautio.*

other hand the transmitter power is very high, typically tens to hundreds of kilowatts. Fortunately the broadcast antennas are usually installed on high masts and heights like 200 m from the ground are common. On the ground by the mast these antennas do not pose danger, as the radiation downwards is weak enough because of the "flat" vertical radiation pattern. So the general public is not in danger of overexposure but mast workers can be in danger if they go too near a live antenna. The safe distance to a transmitting FM antenna is 15 m under or above, and to a VHF-UHF antenna 5 meters. There should not be workers inside or outside a transmitting antenna, and one should not touch a radiating one. If needed, the transmitter power should be lowered or the transmitter shut down. Mast elevators are safe places, as the metallic structure functions as a Faraday cage and the EMF cannot penetrate inside. Figure 24.7 shows the principles of the antenna installation locations.

During the last few years HDTV transmitters have appeared on the roof of some buildings. Regarding compliance boundaries and other radiation safety rules they can be treated as mobile base stations – remembering, however, that the transmitted power is higher.

24.8 Summing Up

In this chapter the reader had an elementary look at the EMF issue and the science involved in it, plus a short overview of the radiation safety on the point of safety distances and installation recommendations. The debate around radiation from transmitters, especially from those in mobile phones and their base stations, is mostly about science, speculated health hazards and demanded changes to existing exposure regulation; the actual measurements of RF radiation levels are important but only a small portion of the issue. The tabloids and the Internet can spread at high speed highly alarming news about single studies which are claimed to turn upside down the whole picture of safety in mobile systems. An engineer meets laypeople (relatives, neighbors, customers etc.) in everyday life and it is very beneficial to understand the process with which (EMF) science works and to know how the existing consensus and exposure limits come from a wide base of careful research. Also, the engineer must have some thoughts for the radiation safety when planning installations of the antennas in buildings or masts, or when designing a new model of cellphone. Remember that the EMF is not black magic but a physical phenomenon among others.

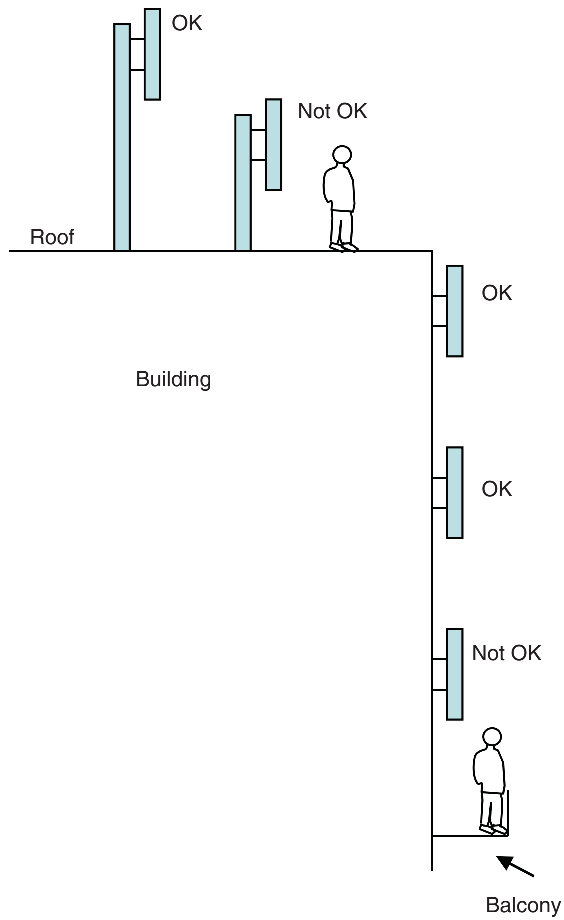

Figure 24.7 Examples of good and not recommended antenna installations.

24.9 High-Power Network Planning

GSM, UMTS and LTE represent cellular networks that are based on the reception and transmission of the radio signals with primarily handheld devices and that are thus close to the human body when they are transmitting to the uplink. The maximum power level of the respective handheld devices is low, GSM representing the highest level of 33 dBm, that is, 2 W in 200 kHz band, whilst the UMTS handheld devices are limited to 30 dBm (1 W) in 5 MHz band and LTE hand held devices have a maximum power level of 20–23 dBm in variable, 1.4–20 MHz band. Equally, the transmission power of the base stations of these systems is low, in range of some tens of watts.

Another mobile network type is mobile TV with various technical and commercial solutions. These networks are meant solely for reception of radio signals for streaming audio and video content in the mobile device. As the coverage areas of the mobile TV networks are large and the infrastructure is separate from the cellular communications networks, the power levels of the transmitters of the mobile TV networks are typically higher, and the antennas are of high gain, which results in clearly higher radiation power levels.

For mobile TV networks, a preliminary calculation about the transmitter power levels should be carried out already in the initial radio network planning phase. At this stage, regulatory rules, as well as a rough estimate about EMC and safety zones give a base for estimating the minimum distance between the transmitter antennas and the surrounding population or the antenna systems of other telecommunication systems like GSM and UMTS.

In the initial phase of radio network planning, it is sufficient to investigate the high level regulatory limits for the nonionizing radiation. A typical maximum allowed value for the mobile TV site might be in the order of 50 kW (EIRP), with additional rules to be taken into account, for example, as a function of the antenna height and site type (differentiating wall-mounted, rooftop and tower-mounted antennas). International and regional regulation provides sufficient information about the upper limits of the radiation that should be taken into account in the nominal plan. The radiation level might need to be limited further, depending on the area type (urban or open), the antenna height, and the frequency. The practical limitations might mean that the highest power class transmitters and high-gain directional antennas can not be used in the implementation and the level should thus be revised case basis.

Detailed safety zone and EMC limit calculations can be carried out when the concrete site locations are known. The allowed antenna distance from the other system antennas or from the installation personnel and the population depends on the type of installation, that is, the limits vary depending on the rooftop, tower or indoor antenna placement. In a typical broadcast tower case, the main task is to calculate EMC limitations, that is, the interferences that the transmitting antennas of the mobile TV network cause to other systems and vice versa, as the antennas are installed sufficiently high in towers by default.

In rooftop and indoor installations, the safety distance limits for human exposure should also be taken care of with related safety zone marking, for example, for occasional maintenance visits.

The following sections present studies about the safety distance calculations of the DVB-H installation in rooftops or towers. A simple yet functional model that is suitable for the DVB-H deployment is presented for safety distance estimates.

When the electrical field is calculated above or below the antenna, the attenuation factor of the vertical radiation pattern should be taken into account accordingly. The methodology applies in the close distance of the site, and the outcome is to minimize the interference level caused by DVB-H to the other systems nearby, as well as to make sure human exposure limits are not exceeded. In the back-lobe of the antenna, the method proposes the use of maximum value (i.e., the minimum possible attenuation value of the radiation pattern) over the whole half-hemisphere.

Although the DVB-H frequency usage is regulated, there might also be a need to calculate some special cases for longer distances, like safety zones in the area where sensible space signal reception stations or military bases are present. In these cases, the related safety zone calculation is straightforward and the methodologies of free space loss can be utilized.

In site installations, it can be expected that total exposure increases close to the sites due to DVB-H. The presented methodology gives an estimate of the values, and field tests can be performed in selected locations, especially on rooftop sites in order to fine-tune the calculations.

The outcome of the safety distance calculations in the nominal planning phase helps to reject those radiation levels from the planning assumptions that exceed the regulatory limits and are thus not possible to utilize in detailed planning, either.

24.9.1 Introduction of DVB-H Interference Estimation

The typical radiating power levels of the DVB-H sites are normally between the values used in mobile and television broadcast networks. It is thus important to take into account accordingly the EMC (Electro Magnetic Compatibility) as well as the human exposure limits in the DVB-H transmission facilities. This chapter presents a practical method for estimating the safety limits of human exposure and EMC in different DVB-H installation environments, based on references [2]–[8]. Furthermore, case studies are presented for the most typical environments [9].

The radiation of the DVB-H transmitter sites can be assumed to be higher than in mobile networks. As there are normally mobile or broadcast systems installed in the same site as DVB-H, it is essential to dimension the safety distances in order to minimize the risk of intersystem interferences. On the other hand, human exposure limits must be calculated for both technical installation personnel and for the public.

The DVB-H transmitter site antennas can be located to the telecom or broadcast tower as well as on the rooftop or indoors of the buildings. The power level should be adjusted accordingly in order to avoid any interferences or safety zone extensions.

24.9.2 Safety Aspects

The DVB-H radio transmission is defined to the frequency range of 470–862 MHz (UHF IV and V) with a channel bandwidth of 5, 6, 7 or 8 MHz. The channels can be divided into several subchannels as the used bit stream is relatively low. DVB-H can also be deployed to VHF III and L bands. The power level of DVB-H depends on transmitter manufacturer solutions. Typically, the power amplifiers can produce around 100–9000 W (output power to the feeder), although the maximum power level might be limited to a few thousand watts in practical solutions.

DVB-H radiation is nonionizing as in the case of mobile network technologies. The radiation does not thus alter the human cell structure as can happen in ionizing systems like X-ray equipment. Nevertheless, a sufficiently high nonionizing radio transmission power can increase the cell temperature. The radiation at a given distance can be estimated with various propagation models. The simplest one giving the maximum values is the far-field attenuation in free space:

$$L = 32.44 + 20\ \log(d) + 20\ \log(f). \tag{24.4}$$

The safety distance of the DVB-H antennas must be assured in the deployment of the system. A simple but practical method can be created based on the fact that the antenna is located to the tower or rooftop, and that the most meaningful area to be investigated is found below the antenna.

The DVB-H antenna solution can be based on omniradiating poles or directional antennas. Figure 24.8 shows an example of the far-field horizontal and vertical radiation patterns of the directional antenna used in DVB-H. As can be observed from the figure, in the practical installation environments, the vertical radiation pattern is the most meaningful when estimating the field strength and safety zones.

In order to obtain safety limits below the DVB-H antenna, a loss analysis with different angles from the antenna can be done as shown in Ref. [9]. Having the vertical antenna pattern, respective coordinate system and linear scale for the antenna radiation attenuation as shown in Figure 24.9, an antenna attenuation table can be created with the scale from 270 to 180 degrees (back lobe) and from 180 to 90 degrees (main beam). 180 degree vertical angle means the point below the antenna. In this specific example, Table 24.3 can be obtained from the respective antenna data. The 90 degree angle represents the main beam of the antenna with 0 dB loss. Please note the difference with the conventional marking of the angles, as normally 0 degree elevation represents the horizon and −90 degrees the point below the antenna.

The values of the table can now be analyzed in two phases. The first task is to obtain the minimum attenuation value for the back and side lobes, that is, in beam angle of 270–180 degrees. In this example it is 20.9 dB. As also the minimum attenuation of the horizontal pattern falls into this value, it can be used as a reference for the complete back side (hemisphere) of the antenna. The vertical angle between 180 and 90 degrees is used for the analysis of the main lobe. The angle is the independent variable used to obtain the respective safety distances.

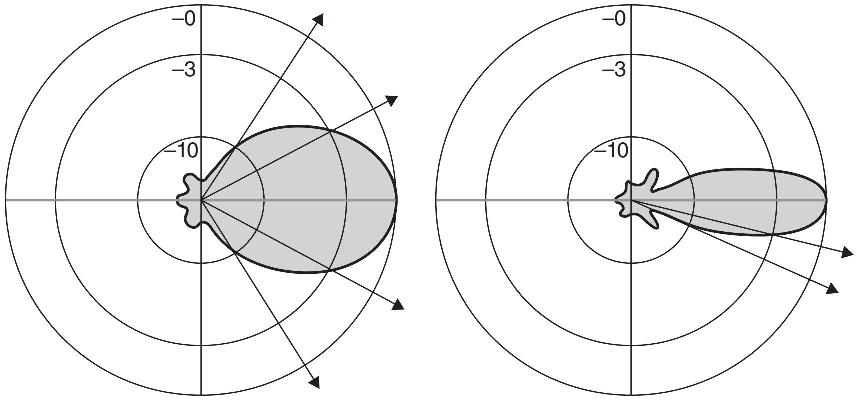

Figure 24.8 Example of the horizontal plane of the antenna radiation pattern (left). The 3 dB attenuation determinates the beam width. In this specific case, the beam width is about 60 degrees. The second pattern (right) shows an example of the vertical plane of the directional antenna radiation. In this case, the beam width in 3 dB attenuation points is about +/−14 degrees, that is, 28 degrees [9]. Figure by Jyrki Penttinen.

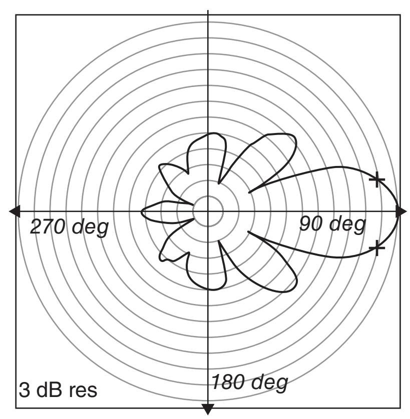

Figure 24.9 An example of the vertical pattern with 3 dB resolution.

Table 24.3 Example of the antenna attenuation values (dB)

Deg	Av	Deg	Av	Deg	Av
270	23.1	210	26.0	150	23.1
260	26.0	200	24.4	140	16.5
250	30.5	190	21.9	130	14.0
240	28.0	180	20.9	120	20.0
230	24.4	170	20.9	110	10.8
220	24.4	160	30.5	100	2.2

The next step is to calculate the safety distance for the radiation power, which is the result of the DVB-H transmitter power, transmitter filter, cable, connector and jumper losses, possible power splitter loss for multiple antenna arrays and the radiation pattern loss in function of the elevation angle. The safety distance depends on the regulatory decisions. As a common type of regulations, the reference [10] can be used in order to obtain the respective formulas.

The immediate area close to the antenna is the near-field region. Most of the electromagnetic energy in this region is stored instead of radiating. The field has considerable variations within this zone, making the field estimation extremely challenging. Further away from the antenna, the reactive near-field decreases and the radiating field becomes predominating as a function of the distance until the far-field zone finally stabilizes the characteristics of the radiation making the calculation of the field strength reliable.

The dimensions of the radiating antenna have impact on the minimum distance where the far-field starts dominating. Assuming the variable D indicates the largest dimension of the antenna and if λ is the wavelength of the observed signal, the following formula can be used for the calculation of the minimum far-field limits in case of large antennas (i.e. if the greatest dimension of the antenna is much more than the wave length), according to Ref. [103]:

$$d_{far-field} = \frac{0.5D^2}{\lambda}. \tag{24.5}$$

For small antennas, the following formula can be used:

$$d_{far-field} = \frac{\lambda}{2\pi}. \tag{24.6}$$

The following calculations are valid for the far-field, for frequency range of 300–150 MHz which falls into the operating frequency range of typical mobile TV broadcast systems like DVB-H. In this specific case, the maximum power density for the general public in the DVB-H frequency range can be obtained via the following formula:

$$W = \frac{f(MHz)}{150}. \tag{24.7}$$

Furthermore, the power density in the far-field region can be obtained via the following equation:

$$W = \frac{EIRP}{4\pi r^2} = \frac{PG}{4\lambda r^2}. \tag{24.8}$$

In the formula, EIRP is the effective isotropic radiated power (W), r is the distance from the radiating antenna (m), P is the power fed to the antenna, and G is the antenna gain compared to the isotropic antenna (10log10[G] in dBi).

It is now possible to estimate the safety distance in front of the DVB-H antenna. Investigating Table 24.3, it is possible to estimate the safety distances also in different sides of the antenna. Let's select the 700 MHz frequency and transmitter power of 1500 W after the filter loss. The additional cable, connector and jumper loss (L) can be estimated to total 3 dB. Using the antenna pattern presented in Figure 24.9 assuming its gain (G) is 13.64 dBi and the physical dimensions are $190 \times 500 \times 1000$ mm, the power density and EIRP values are:

$$W = \frac{700\ MHz}{150} = 4.67\ W/m^2. \tag{24.9}$$

$$EIRP(W) = P \cdot 10^{\left(\frac{G-L}{10}\right)} = 1500\ W \cdot 10^{\left(\frac{13.64-3}{10}\right)} = 17.38\ kW. \tag{24.10}$$

The minimum safety distance within the main beam area is thus:

$$r = \left(\frac{EIRP}{4\pi W}\right)^{0.5} = \left(\frac{17.38\ kW}{4\pi \cdot 4.67\ W/m^2}\right)^{0.5} = 17.2\ m. \tag{24.11}$$

The wavelength of the used frequency is, when c is the speed of light:

$$\lambda = \frac{c}{f} = \frac{300{,}000\ km/s}{700 \cdot 10^6\ Hz} = 0.43\ m. \tag{24.12}$$

Let's investigate the minimum distance for the far-field in order to make sure the calculation is valid. As the extreme dimension of the antenna is more than wavelength (1m), the antenna is considered large, and the far-field distance limit is thus:

$$d_{far-field} = \frac{0.5}{\lambda} = \frac{0.5 \cdot (1\ m)^2}{0.43\ m} = 1.2\ m. \tag{24.13}$$

The calculation is thus valid as the value is within the far-field zone. For the radiation angles of 180… 90 degrees, Table 24.3 can be created.

For the vertical back and side lobes, that is, for the vertical radiation angles of 270… 180 degrees, the common value of 20.9 dB was obtained from Table 24.3. This value corresponds the minimum safety distance of 1.6 meters for the whole range of the above mentioned angles. As can be observed from Tables 24.3 and 24.4, the relatively narrow vertical beam width provides reduced safety distances outside of the main lobe.

Figure 24.10 clarifies the idea of the safety distances in graphical format.

In case of Figure 24.10, using a 2 meter high antenna pole, the safety distance below the antenna is thus achieved for the personnel inside the building. As the antenna is on the edge of the rooftop, the main beam is also secured. In addition, the roof material provides additional attenuation for indoors.

Figure 24.11 shows an extended analysis with the same radio parameters as previously but varying the transmitter power level between 100 and 9000 W.

As can be observed, the lower the power level is and closer to the back side of the antenna, many of the calculated points falls below the minimum far-field distance of 1.2 m. As the calculation is not accurate in those near-field spots, the 1.2 m limit can be considered as a minimum limit for all the cases in near-field.

In case of Figures 24.9 and 24.12 the safety limits on the rooftop should be marked accordingly in case of, for example, the maintenance personnel closing the antenna. The back side of the antenna can be marked as

Table 24.4 *The safety distances r (m) as a function of the observed angle (degrees)*

Degrees	Av	r(min)
180	20.9	1.6
170	20.9	1.6
160	30.5	0.5
150	23.1	1.2
140	16.5	2.6
130	14.0	3.4
120	20.0	1.7
110	10.8	5.0
100	2.2	13.4
90	0.0	17.2

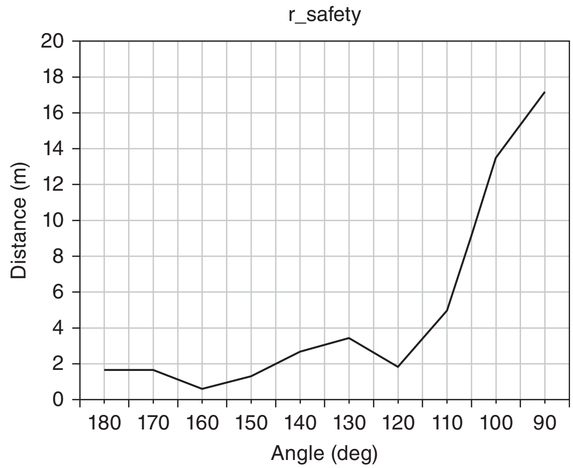

Figure 24.10 *An example of safety distances as a function of angle of the vertical antenna radiation pattern.*

round with the minimum of the calculated safety distance, with preferably an extra margin as the antenna pole can affect the final radiation pattern of the antenna (i.e. the pole might act as a part of the antenna elements).

For the main beam, the horizontal radiation pattern must be taken into account by calculating the safety limits on the sides of the antenna. A mask with sufficient additional margin can be used, for example, by observing the reference attenuation points of 20, 10 and 3 dB and the respective angles. In practice, it is

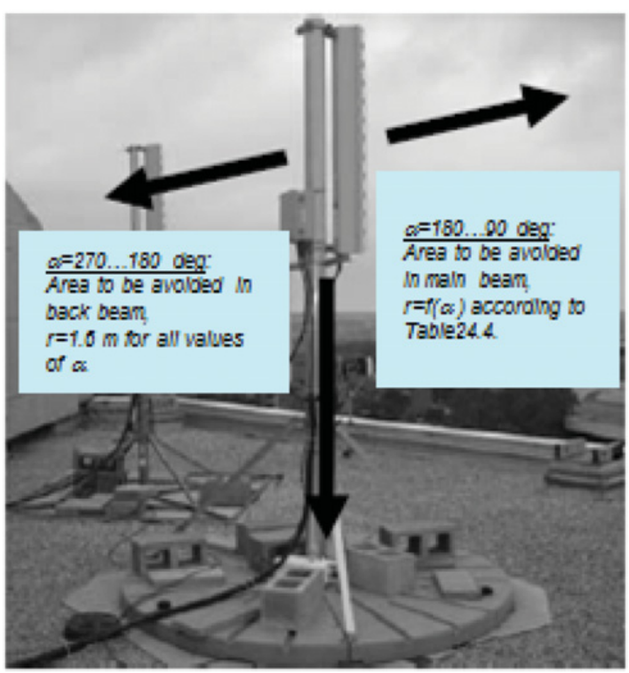

Figure 24.11 *An experimental antenna installation. By taking into account the values presented in the example, the vertical safety distance behind the antenna is 1.6 m. Photo by Jyrki Penttinen.*

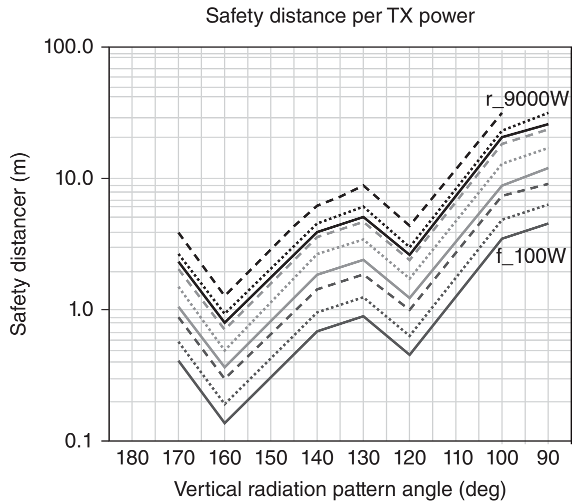

Figure 24.12 The safety distances for a set of TX power levels presented in logarithmic scale and with the parameter setting of the presented example.

important to assure the personnel cannot cross the main beam accidentally as the full EIRP might mean considerable safety distances in front of the beam. It is possible to estimate the safety regions now, e.g., according to Figure 24.13.

24.9.3 EMC Limits

When the DVB-H antenna is installed in the tower, there might be antennas of the other telecommunications systems on top and below the DVB-H antenna. The EMC calculations should thus take into account mainly the upper and lower side lobes of the DVB-H antenna.

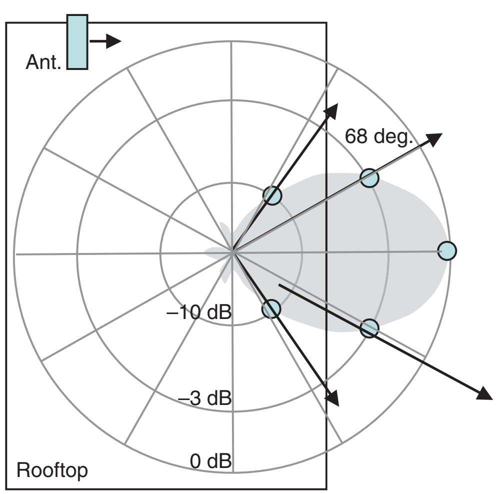

Figure 24.13 An example of rooftop installation. The safety distance on the sides of the antenna should be calculated according to the horizontal radiation pattern of the antenna.

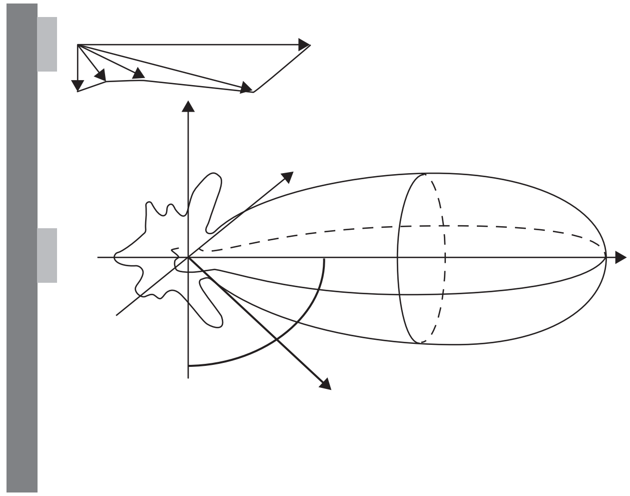

Figure 24.14 *When using directional antenna, the relative field strength above and below the antenna depends on the characteristics of the vertical radiation pattern.*

The analysis methodology presented in the previous chapter can be used also for the EMC calculations. Figure 24.14 shows a principle of the EMC in the towers. The DVB-H antennas create a certain electro magnetic field strength around the antenna, which might be considered as an interference field for the antennas of the other systems nearby.

As in the previous case, the vertical radiation pattern can be used as a basis for the EMC field estimations. In fact, the vertical pattern right above and below the antenna is the most important as the antennas of the other systems are normally in the same vertical line of the tower.

As can be observed from the Figure 24.14, the vertical pattern in this case example is symmetrical above and below the antenna element. The attenuation value around 180 degrees (below the antenna) as well as 0 degrees (above) is thus 20.9 dB as shown in Table 24.3.

In order to calculate the interfering electrical field, the following formula can be used:

$$E_i = \frac{1}{2r}\sqrt{\frac{\eta P}{\pi}}. \tag{24.14}$$

The variable η is the wave impedance in air with the value of 377 ohms, P is the transmitter power (EIRP) including the gains G and attenuations A.

Let us investigate the electrical field for the previously presented case examples with the power levels of 1500 W in 2 meter distance form the main beam. As shown previously, the example results in a total EIRP of 17.83 kW which produces the following field in 2 meters distance from the main beam:

$$E_i = \frac{1}{2r}\sqrt{\frac{\eta GP}{\pi}} = \frac{1}{4}\sqrt{\frac{377 \cdot 17830\ W}{\pi}} \approx 366\ V/m. \tag{24.15}$$

The field strength reduces according to the radio propagation environment. Assuming the worst case, the free propagation loss can be assumed as shown in Equation (24.4).

The antennas of the other systems might be relatively near the DVB-H antennas due to the lack of the space in the tower. The essential question is thus, what is the minimum allowed distance between the antennas? The allowed field strength depends on the immunity level of each system. As an example, the commercial electrical equipment are able to handle interference field strengths of about 2–10 V/m. For the professional electrical equipment as well as for example, mobile network elements, the requirement is considerably higher.

In addition to pure field strength, the frequency is also essential. As an example, the GSM 850 and GSM 900 might suffer from distortion caused by the harmonic components of the DVB-H transmission in the radio spectrum whilst GSM 1800/1900 is safe due to frequency difference and filtering of the equipment.

For the situation with the antennas located on top of each other, the attenuation value of the vertical radiation pattern of the DVB-H should be taken into account. The value for the EMC calculations can be estimated by observing the 180 degree angle of the vertical radiation pattern. In practice, the value might need an additional margin in order to take into account possible variations, for example, for the pattern inequality around 180 degrees, and the possible side-effect of the tower itself which can interfere the radiation pattern. The effect of the radiation pattern can now be extended in the following way:

$$P_{tot}\,[dBm] = P\,[dBm] + G\,[dB] - A_v\,[dB]$$

$$P_{tot}\,[W] = \frac{10^{\frac{P_{tot}[dB]}{10}}}{1000} \tag{24.16}$$

$$E_i = \frac{1}{2r}\sqrt{\frac{\eta P_{tot}\,[W]}{\pi}}.$$

The A_v represents the attenuation (dB) in the observed vertical direction. In this case, the value is 20.9 dB both right above as well as below the antenna as can be seen in Table 24.1 and Figure 24.15. The severity of the interfering field depends on the frequency, that is, how considerable the frequency separation is.

The effect of the main beam should be observed especially when the DVB-H antenna faces the antenna of another system. The EMC safety distance is sufficient in most of the installation cases as there should be a sufficiently open area in front of the antenna, but there might be situations, for example, with antennas located on rooftops in both sides of a street. It is thus important to make sure that the facing antennas (e.g. DVB-H antenna on rooftop and GSM 850 antenna on the other side) are producing sufficiently low field within the respective distance. The field in the main beam might be quite high as can be observed from Figure 24.16, meaning that the mid level and highest power transmitters should be avoided in rooftops.

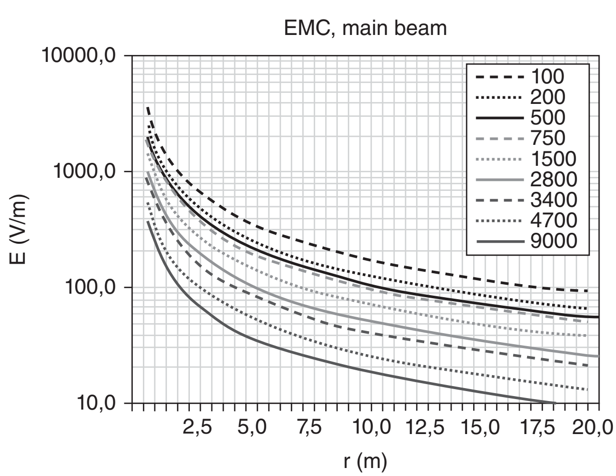

Figure 24.15 *The electrical interference field in the main beam of DVB-H antenna with the transmitter power levels of 100–9000 W (corresponding EIRP of 1.2–104 kW).*

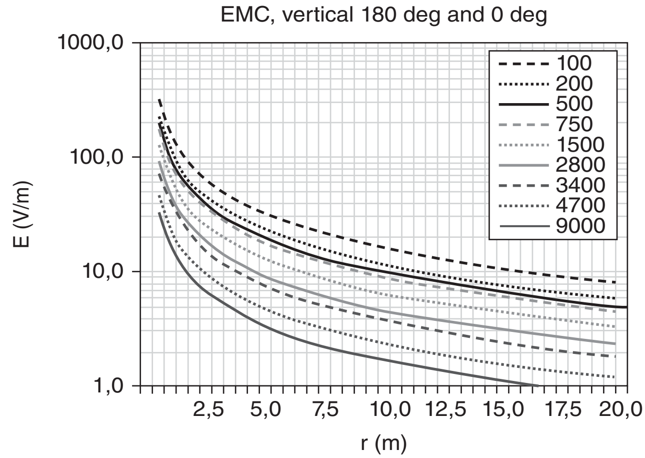

Figure 24.16 *The electrical interference field above and below the typical directional DVB-H antenna with the TX power levels of 100–9000 W (corresponding to EIRP of 9.4–848 W).*

24.9.4 Conclusions

The results of the case study show that the installation of the directional antenna in telecom or broadcast tower, as well as on the rooftop of the building, can be done in a controlled and safe way. The logical way of obtaining safety distance limits is to observe the vertical radiation pattern of the antenna as the antenna is above the population in the tower installations. In rooftop cases, the vertical beam dictates the safety distance which should be analyzed in more detail in all the angles, from the point below the antenna up to the main beam direction. The corresponding height of the antenna is relatively straightforward to design based on this information.

As for the EMC, the vertical beam analysis can be used in the designing of the location of the DVB-H antenna in the telecom or broadcast tower. In multiple directional antenna installation, the final vertical radiation pattern depends on the antenna array and should be calculated or measured taking into account the power splitter loss. The EMC safety distance above and below the antenna depends on the frequency separation and on the maximum field strength that the other systems can support, taking into account the respective vertical beam attenuation of the other systems.

References

1. ICNIRP (International Commission on Non-Ionizing Radiation Protection) www.icnirp.de (last accessed 3 August 2014).
2. DVB-H Implementation Guidelines. Draft TR 102 377 V1.2.2 (2006-03). European Broadcasting Union, 108 pp.
3. Henriksson, J. DVB-H standard, principles and services. HUT seminar T-111.590. Helsinki, 24.2.2005. Presentation material, 53 pp.
4. Editor: Thibault Bouttevin. Wing TV. Services to Wireless, Integrated, Nomadic, GPRS-UMTS&TV handheld terminals. D8 – Wing TV Measurement Guidelines & Criteria. Project report. 45 pp.
5. Faria, G., Henriksson, J.A., Stare, E. and Talmola, P. (2006) DVB-H: Digital Broadcast Services to Handheld Devices. IEEE, 16 pp.
6. Lee, Wi.C.Y. (1986) Elements of cellular mobile radio system. *IEEE Transactions on Vehicular Technology*, **VT-35**(2), 48–56.

7. Fanton, M.D. (2006) Analysis of antenna beam-tilt and broadcast coverage. ERI Technical Series, vol. 6, 3 pp.
8. Fanton, M.D. (2006) Analysis of antenna beam-tilt and broadcast coverage. Electronics Research, Inc. ERI Technical Series, vol. 6, 9 pp.
9. Penttinen, J. (2011) The DVB-H radio network planning and optimisation. Doctoral thesis, Aalto University, School of Electrical Engineering.
10. Minister of Public Works and Government Services. Limits of Human Exposure to Radiofrequency Electromagnetic Fields in the Frequency Range from 3 kHz to 399 GHz. Safety Code 6. Environmental Health Directorate, Health Protection Branch. Publication 99-EHD-237. Minister of Public Works and Government Services, Canada 1999. ISBN 0-662-28032-6, 40 pp.

25

Deployment and Transition of Telecommunication Systems

Michał Maternia

25.1 Introduction

"June 17, 1946 – A driver in St. Louis, Mo., pulled out a handset from under his car's dashboard, placed a phone call and made history. It was the first mobile telephone call." [1] Since this event the wireless telecommunication systems has made a long way. This chapter tries to capture this progress starting from the brief description of the first generation of mobile network deployments through the deployment of GSM, WCDMA, LTE and ending with the outlook on what the future network may look like. The chapter tries to describe the process of network transition and give a short insight into the factors that contributed to this transformation: spectrum aspects and end-user terminals development. It also highlights trends and challenges in modern cellular solutions.

25.2 Why to Deploy Wireless Systems

Over the years wireless cellular solutions were always compared against wireline technologies. First generations of wireless cellular telecommunication products (cf. Figure 25.1) were confronted with fixed telephony products; later when mobiles started to offer broadband connections they were set against wireline Internet access solutions. Over many years those two groups of products were regarded as complementary. Wireless solutions allow mobility handling in combination with cheap and easy access. On the other hand data transfer speed realized via copper is still way above the possibilities of over-the-air devices, not to mention optical fibers. This is the reason why backhaul connections and core parts of wireless telecommunication networks are still in most cases realized via wirelined approach. But in the end it is the mobility and the possibility to experience the undisrupted, always connected state that determined the tremendous growth of wireless devices in recent years.

The Telecommunications Handbook: Engineering Guidelines for Fixed, Mobile and Satellite Systems, First Edition.
Edited by Jyrki T. J. Penttinen.

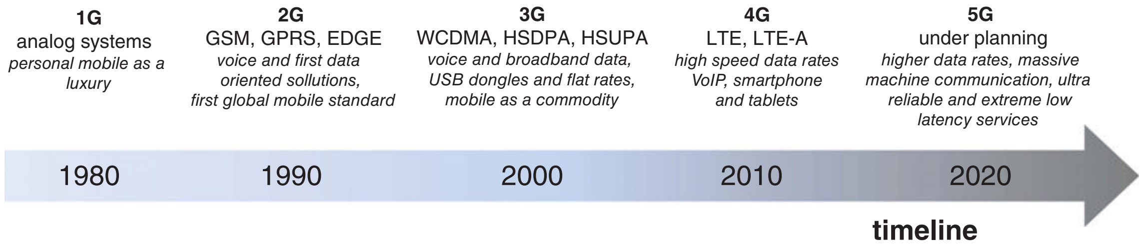

Figure 25.1 *Evolution of a wireless cellular system.*

According to the statistics of the International Telecommunication Union (ITU), the number of mobile-cellular subscriptions already exceeded the number of fixed-telephony subscription, reaching in 2012 more than 6 billion as depicted in Figure 25.2 [2]. Similarly to the situation observed in the first decade of the new millennium when portable laptops sales started to dominate over stationary desktop PCs, recent trends show that tablets, which are even more convenient in transport, have started to take over the market share from laptops. Interestingly cellular networks are completely dominating the telecommunication domain in many underdeveloped or developing countries. We can be tempted to think that in some poor areas we will see a corded telephone rather than a cellular device, but in reality it is the opposite. Today's wireless technologies are so cheap and effective that it is much easier to deploy such solutions on the savannah of Africa or desolated tundra forests in Siberia, than to provide a fixed copper or optical last mile infrastructure in those areas (the so-called "latecomer effect").

On top of proliferation of cellular wireless networks we observe another phenomenon. Since the introduction of 3G High Speed Packet Access (HSPA) the profile of cellular traffic has changed significantly, shifting from plain voice service to heavy data consumptions. Additionally, the penetration of mobile devices and also data traffic volume per single user has drastically increased. This is also reflected in both the software and hardware domains. High definition Youtube or USB dongle wireless modems that mobile operators sell along with their Internet flat rate offers, are just two examples of what has become a commodity for a typical cellular user. This trend is still observed and it is estimated that between 2010 and 2020 the overall traffic

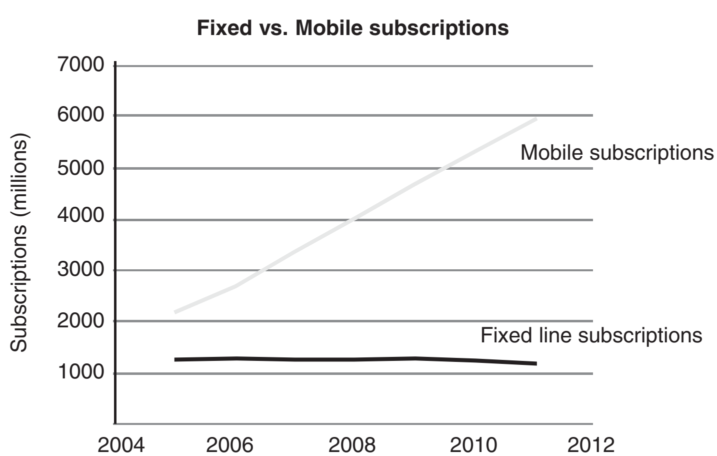

Figure 25.2 *Fixed vs. mobile cellular subscription. Figure interpreted from the data of ITU [2].*

volume will increase by a factor of ∼100–1000x. All this massive traffic growth needs to be handled by wireless network.

Another tendency that can be observed is massive price erosion, reflected not only in the lower prices of end-user devices or land infrastructure, but also in terms of data transfer cost. Until the introduction of HSPA, mobile operators tried to convince their customers to use wireless data transfer by offering very cheap flat rate wireless access. Nowadays, when browsing Internet over a smartphone or tablet is a regular activity for many, it is not only a question of how fast, but also of how cheap this fast broadband service can be provided to end-users.

Also in the infrastructure domain we can witness some interesting trends in the recent years. One observation is the deployment of several different cellular layers in wireless networks. HSPA needs to cooperate with LTE technology, macro cells are complemented with layers of micro, pico and femto stations. This poses new challenges for network operators and telecommunication vendors, for example, how to provide backbone connection to these densely deployed nodes or how to provide a seamless transition between those layers.

All those aforementioned factors contribute to the fact that the cellular ecosystem has became one of the most sophisticated and complicated solutions known to the mankind and efficient network deployment its transition is a continuous and a very challenging process.

25.3 Transition of Telecommunication Systems

Evolution of a telecommunication system is a graduate process. It is usually driven by a growing consumers' demand, but there are multiple other factors that can influence it.

One obvious trigger is of course technological progress. As an example we could give cheap and efficient production of Digital Signal Processors (DSPs) that allowed for incorporating FFT/IFFT based OFDM/SC-FDMA operations which are the key components of 4G radio interface. Another example is the minimization of electronic components that allows for packing huge computational capabilities into modern end-user devices, allowing for example, display of a high data rate video or advanced online gaming, services linked previously with stationary personal computers. A growing number of such high-end personal devices brings more challenges on mobile operators who needs to provide large traffic volumes to end-users with high Quality of Experience (QoE). Interestingly, nowadays wireless telecommunication systems became such a huge part of the IT industry, that it is often a chicken-egg problem where one can't distinguish if it is the technology progress that triggers the telecommunication network transition, or if it is the need for performance enhancement of wireless systems that fuels technology progress.

Another very important phenomenon that influences the development of modern telecommunication networks, are the societal changes that manifest in people wanting to share their daily experience with others, using not only low data rate services like SMSs or simple voice calls, but rather via services that require higher data rates over social network applications for uploading/downloading of photos, songs, videos and so on.

Additional factors influencing the network transition are regulatory issues. One good example from the past is the limit on users' devices radiated power that made the 2G devices smaller and more user friendly (by lowering the energy drain we ended up with smaller battery sizes and eventually with smaller and more convenient phones). Paradoxically one can even state that sometimes it is the regulatory constraints that push the telecommunication solution forward. For instance, limited availability of spectrum for cellular communications forces more advanced and sophisticated network solutions that offer higher spectral efficiency and better utilization of this scarce resource.

It is also worth underlining that wireless evolution is not only the network transition driven by the technical progress. The network evolution process is also propelled by appearances of more and more sophisticated user services and the expansion of existing network. The goal of telecommunication engineers is to provide a smooth transition between the generations of cellular systems, painless from the end-user perspective who should not suffer from challenges and hardships related to that change. Figures 25.1–25.2 summarize the key evolution and statistics of mobile communications systems.

25.4 Network Deployments

In the modern cellular networks we can distinguish several key technologies that emerged one after another: first generation (1G) systems marked by analog methods of data exchange, second generation (2G) characterized by transition to digital radio, third generation (3G) networks that provided broadband connectivity in cellular communication and finally fourth generation (4G) systems, that allow for 1 Gbit of peak download speeds and focus on data exchange. The following section provides a generic overview of the first network deployments and transitions between these generations.

25.4.1 1G Systems

Hexagonal cells system concept for mobile phones, as depicted in Figure 25.3, was first proposed in December 1947, by Bell Labs engineers: Douglas H. Ring and W. Rae Young [3]. However, it took more than 30 years for the first commercial analog and circuit switched cellular networks to appear. In contrary to the later 2G, 3G and 4G systems there was no single 1G standard with the global penetration. Most popular 1G systems was Nordic Mobile Telephony (NMT), Advanced Mobile Phone Systems (AMPS) and Total Access Communications Systems (TACS). Those networks used analog signals to carry data and were primarily intended for voice services.

First 1G network deployments in the early 1980s were dictated by various conditions. The aforementioned standards usually operated at frequencies below 1 GHz (e.g. NMT 450 used carrier frequency around 450 MHz, AMPS operated on 850 MHz etc.) so the cell sizes were large, reaching up to 30 km. The penetration of cellular capable devices was very low and in a typical radio network deployment the idea was to provide sufficient coverage for the voice service, not the capacity. Since user devices were large and bulky, many of them were installed in vehicles or on the ships, hence first base stations were also deployed close to the

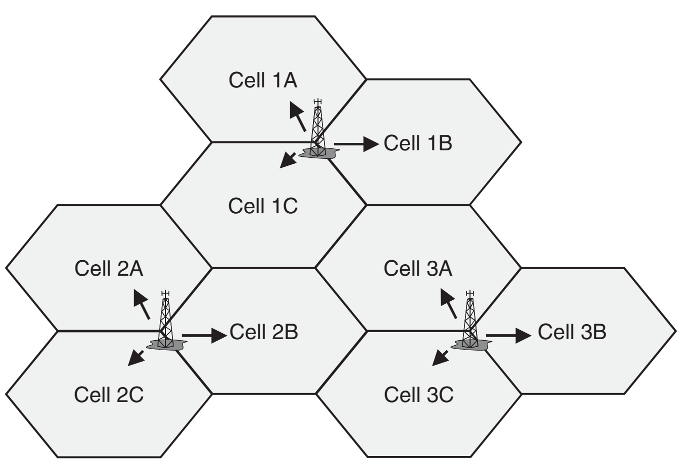

Figure 25.3 *Simplified cellular system concept.*

coast and along the popular roads and large cities. Because analog cellular communication was an emerging technology not foreseen for the mass market, but rather for limited number of customers, there was also no need for dense network deployment. Still some of those 1G standards were quite popular and existed even after the introduction of HSPA technology, for example, some NMT networks in east Europe were shut down in the beginning of 2008, just at the same time when the first LTE trials were ongoing.

25.4.2 2G Systems

What is easily noticeable in the development of the first 2G standards is that in Northern America several different 2G systems emerged. Some examples to mention are Interim Standard 54 (IS-54), IS-136 (also known as Digital AMPS), Integrated Digital Enhanced Networks (iDEN) or Cellular Digital Packet Data (CDPD). On the contrary, in Europe, Global System for Mobile Communications (GSM) was practically the only 2G system and it has gained momentum from the moment of its introduction to the market. This tendency in both geographical regions (Europe and North America) is in total reverse in comparison to the development of 1G networks. In case of the United States practically the only big 1G analog standard that had emerged was AMPS, while Europe experienced development of several different and incompatible 1G technologies. The drawbacks of this approach (problems with roaming, lack of interoperability etc.) were the reason why European Telecommunications Standard Institute (an organization that developed GSM) decided to go for a single solution. In the world of second generations of cellular network solutions GSM achieved a tremendous success and it became the first global cellular standard. There are other 2G standards that served millions of users, but their penetration, in contrary to GSM, was not global. IS-95 (cdmaOne) that was popular in America and Japanese PDC are just two examples to mention, if not stated explicitly, a GSM standard is meant.

The first GSM network was launched on 1 July 1991 by Radiolinja in Finland, a decade after introduction of first 1G networks. Early GSM networks were mostly oriented on providing voice connections, but digital format of the carried signals brought much better voice quality and higher reliability, comparable with the ones that users could experienced with classical plain old telephony systems. This, along with the reduction of size of user terminals and a significant reduction of prices for mobile services contributed to a rapid growth of GSM subscribers base. As for the operators, in couple of years after first 2G deployments, the ones who decided to enter this risky and uncertain business at that time, witnessed a true explosion of wireless cellular voice connections followed by great financial success. This gave the momentum and business justification for the development of the next wireless communication technologies.

Early deployments of GSM networks followed the approach known from 1G. The difference was that in the case of GSM, device manufacturers were obliged to pass the state authorities requirements for the limits on exposure to radio frequency energies. Those Specific Absorption Rate (SAR) limits enforced low Mobile Station power, hence limiting the uplink cell range. This resulted in smaller cell sizes for GSM, despite the fact that the first GSM standard utilized similar carrier frequency ranges as many 1G technologies (around 850 MHz and 900 MHz). Later in the standard development process, GSM was enhanced with a new frequencies set in 1800 MHz and 1900 MHz band that enforced even smaller cell sizes due to higher radio signal attenuation at higher frequencies. On the other hand limitations on total power radiated by the Mobile Station helped to address many health concerns shared by the mobile device consumers. This problem, however, still raises a great deal of controversy despite the fact that even nowadays, when many people can’t imagine living without mobile devices and we are surrounded by a dense wireless network, there is no conclusive scientific evidence on the harmful impact of usage of mobile devices to human body.

Along with the growing popularity of 2G the network operators had to change their approach towards the radio network deployments. If the first networks were usually high power macro sites that aimed at providing a large coverage, in case of GSM deployments, in order to meet the growing demand for wireless voice connections, the operators had to start differentiating the size of their cells. Based on the placement of the

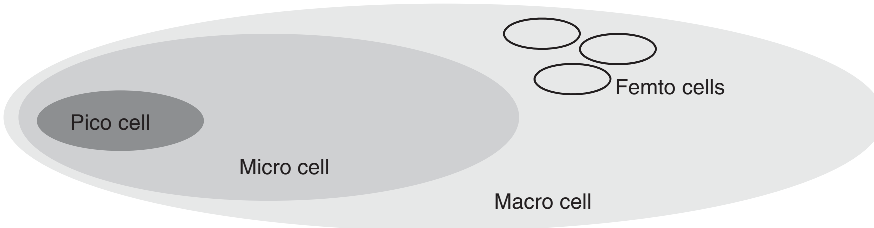

Figure 25.4 *Cell size for macro, micro, pico, femto cells.*

antenna, the output power and their general purpose several cell types can be distinguished as shown in Figure 25.4:

- Macro cells – large cells with radius up to several kilometers and tens of watts of total output power in the serving base station. Their main purpose is to provide coverage over the large (macro) area.
- Micro cell – medium-sized cells with radius of hundreds of meters and total output power limited down to typically 10 watts. Micro cells are usually deployed to provide coverage in quite large areas which are, however, smaller than the ones handled by macro cells but with higher traffic densities, for example, large shopping malls. They provide a compromise between capacity and coverage demands.
- Pico cells – small cells, usually used to provide indoor and outdoor capacity in the area of tens of meters characterized by high traffic volume, for example, train stations or office buildings. The total output power of the base station operating a pico cell typically ranges between 1 and 5 watts, while the base station itself tends to be quite simple, for example, have only omnidirectional antennas.
- Femto cells – smallest cell of all cell types with the smallest output power, usually below 1 watt. Foreseen for residential indoor usage where the backbone connection towards the mobile operator network may not be provided by the operator (e.g. it may be a simple DSL line). The advent of femto devices is rather linked with the 3G HSPA and 4G LTE networks, but it is not limited to any particular technology. In 2G systems femto cells didn't gain much attention; however, it is foreseen that it will play a very important role in future cellular networks.

Nowadays, this basic cell type division, that is depicted in Figure 25.4, is sometimes complemented with additional two types:

- Umbrella cells – term related to the combination of larger cells like macro and several smaller cells like micro or pico, where larger cell is used to provide coverage and mobility aspects between smaller cells.
- Metro cells – a cell type that has a very similar properties as micro cell but in contrast to them metro cells integrate all elements required for the base station operation (including antenna) in one compact device.

Please keep in mind that there is no strict 2G standard with the explicit definition of different cell types. The aforementioned division is rather loose; however, it is very useful in the network planning and dimensioning process. Additional distinction of macro, micro and pico cells is sometimes made based on the basis of where the antennas for the particular cells are placed. Macro sites have their antennas usually mounted much above the average rooftop level, for instance on dedicated masts or on top of tall buildings, while micro sites have their antennas placed usually below rooftop level. Pico and femto cells in this definition are distinguished by their indoor deployment. Also the typical equipment's shape and size reflect the division in different cell types. Equipment for macro cells is rather large and bulky with external antenna systems while the equipment for femto cell can be encapsulated in less than 3 L volume including the integrated antenna.

This variety of different cells sizes and equipment reflects a different approach toward the network deployment that operator had to consider. In contrast to its analog predecessors, 2G networks started to be not only about coverage, but also about capacity. This trend continues today.

From the network deployment point of view one very interesting standard which should be mentioned while discussing 2G is the solution created by the European Rail Traffic Management System known as GSM – Railway or simply GSM-R. It is a technology based on GSM that is used to provide a standardized communication for broadly understood railway applications. Deployment of the GSM-R radio network is quite different from the typical deployment for cellular networks, since the movement of its users follows a very predictable pattern. As a consequence the coverage is also very selective and simply limited to railways areas. Antennas used in GSM-R usually offer high directionality coordinated with the railway positions. This allows for very high distances between the stations, even above tens of kilometers. Additional challenge come from the handovers at tunnel's entrances or exits. To solve the problem of connection losses in high loss tunnels, an extra set of antennas are deployed to provide a sufficient handover region.

25.4.3 2G Evolution from GSM to EDGE

From the point of view of modern, packet-oriented cellular networks, one of the first major enhancements of GSM networks towards efficient data exchange was High Speed Circuit Switched Data (HSCSD) feature developed on top of first GSM based data transfer solution – Circuit Switched Data (CSD). HSCSD was introduced in Release 96 of phase 2+ of GSM and provided a significant improvement in transfer speed offered by 2G. The previous method of data transfer, CSD, utilized a single dedicated radio time slot to carry user's traffic, reaching transfer speed up to 9.6 kbps. As the name of the feature suggests, the data connection was realized in a circuit switched mode similarly to a simple voice connection. HSCSD that used a different error protection coding methods pushed the peak data rates limit up to 14.4 kbps. Additionally, in a later stage of development up to four time slots could be used to carry user's data, which boosted the transfer speed up to 57.6 kbps. However, this maximum speed was available only in a very good radio channel conditions due to error protection overhead.

Despite this improvement the introduction of HSCSD didn't gain massive user attention, even though it didn't require any hardware change, only a software upgrade. The reason behind the lack of interest among users was mostly related to the fact that reservation of four time slots was charged by the operators as equivalent of four voice calls. This approach has limited the number of potential recipients of the service. Along with the introduction of multislot HSCSD a new important step in 2G evolution took place. A new data service called General Packet Radio Service (GPRS) was standardized in the late 1990s. GPRS was a first packet-switched cellular technology oriented on the data transfer that gained a worldwide attention in the cellular community. It added a packet switched functionality on top of circuit switched GSM network and required two additional core network nodes as seen in Figure 25.5:

- Gateway GPRS Support Node (GGSN): the most important component of GPRS networks which works as a router to external packet switched network, Internet or SGNSs and GGSNs in other Public Land Mobile Networks (PLMNs). The main task of GGSN is to encapsulate and deencapsulate the traffic sent to and from SGSN. Additionally it converts the external packet data protocol addresses to the ones used by GSM. It is also responsible for charging, authentication and session management procedures.
- Serving GPRS Support Node (SGSN): its main task is to route the packets to and from the user in the served area. SGSN is used to store the information of the user location and the necessary data to perform authentication and security procedures together with GGSN. The functionality of the SGSN and GGSN may be combined into one physical node and one GGSN is usually connected to a multiple SGSNs.

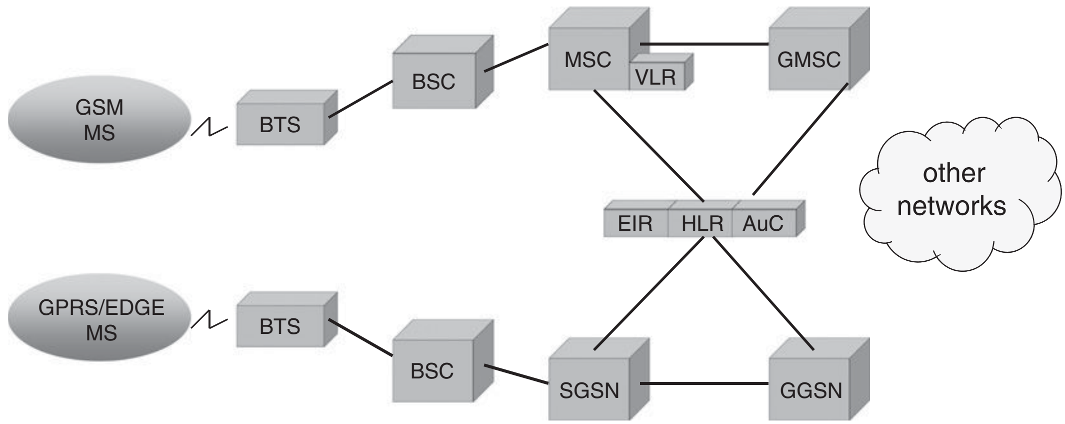

Figure 25.5 *GSM and GPRS/EDGE architecture.*

In order to introduce GPRS functionality in the first 2G networks, apart from the aforementioned changes in the core side, additional hardware upgrade was usually required in the radio network of GSM. A typical upgrade was the hardware enhancement of the Base Station Controller (BSC) with the unit called Packet Controlled Unit that could distinguish packet traffic from voice and provided the interface towards SGSN. From the end-user perspective also new GPRS Mobile Station device was a must in order to use the new packet switched service because old GSM legacy phones didn't have the capability to handle packet switched traffic.

With evolved GPRS features from Release 98 and Release 99, the 2G system could offer services with the theoretical peak data speeds of up to 171.2 kbps with 8 time slots combination. Unfortunately the data rates typically available for users were in most cases much lower and very often not higher than 50 kbps. The key limitation behind that are radio channel limitations and the overhead related to the forward error protection of transfer data, as well as the fact that GPRS is a best effort service, which practically means that it has lower priority than other traffic like voice. In some cases GPRS transfers suffered from higher latencies comparing to HSCSD services, due to the difference in how circuit switched (HSCSD) and packet switched (GPRS) data is prioritized and handled. However, the popularity of the GPRS feature and enhanced speed compared with a simple GSM networks resulted in a situation that in cellular ecosystem and in marketing those solutions are sometimes referred to as 2.5G. The development of new standards enhancing the digital transmission was also in line with the development of new technologies like Wireless Application Protocol (WAP) and Wireless Markup Language (WML) which were one of the first milestones in the Internet usage over wireless mobile devices.

Another important step forward in 2G systems was the introduction of Enhanced Data Rates for GSM Evolution (EDGE) that some people refer to as 2.75G. EDGE is built on top of the existing GPRS network elements and it requires no new infrastructure nodes. Thanks to the new channel coding schemes and new modulation types (8PSK instead of GMSK used in GSM/GPRS) EDGE offered almost threefold increase in available data rates and first revision of the standard allowed the theoretical transfer speed of up to 384 kbps. The introduction of EDGE to the market happened in the same time as the standardization of UMTS 3G technology. To allow smooth transition from 2G based system to 3G ones, the standardization process of EDGE resulted in a new Release 7 upgrades that reduced the latencies down to 100 ms and increased the available transfer rates up to 1.3 Mbps in the Downlink (DL) direction and 653 kbps in the Uplink (UL). Those significant enhancements are known as the EDGE Evolution, EDGE II or Evolved EDGE and are based on the utilization of more than one radio carrier (Dual Carrier feature), lowering the Transmission Time Interval (TTI) duration from 20 ms to 10 ms and higher order modulation [4]. As a low investment system for

GSM operators, EDGE gained a lot of popularity and was introduced in most of the 2G networks that were enhanced with GPRS. In the end EDGE standard was performing with the capacity below the Release 99 3G UMTS, but it was capable of delivering the services considered as 3G.

To complement the evolution of 2G networks it is necessary to mention the evolution of the transport network as well. With the densification of 2G base stations one interesting technology emerged, namely wireless backhauling. With the advent of cheap small stations (e.g. micro or pico nodes) it turned out that the costs of providing a backhaul to the network element are sometimes much higher than the cost of the element itself. It is quite easy to imagine it if one considers the network capacity enhancements in the form of network densification in the center of a big city. Construction works and the main effort needed for the placement of a copper cable or an optical fiber would certainly exceed the price of the network hardware itself. This problem in some cases can be solved with a wireless backhaul that provides a high speed connection between the Base Transceiver Station and a Base Station Controller. Two groups of solutions have gained the biggest popularity:

- Microwave Links: based on the wireless transfer technologies at carrier frequencies from 10 GHz to 30 GHz.
- Free Space Optics (FSO) based on laser optics in the unlicensed spectrum.

The drawback of both is a strong sensitivity against weather conditions (connections can break or limit its capacity in case of rain or fog). Additional drawback is the requirement of the Line Of Sight (LOS) connection between connected devices, which means that there can't be any major obstacle on the propagation path for the wireless backhaul transmission.

To conclude the development of GSM network, the reader should keep in mind that the transition to data oriented service brought additional challenges for the network equipment vendors and to mobile network operators, both on the radio and on the core network side. Despite the fact that GSM transmission was already digital, the main service was still voice transmission. This type of service has rather loose requirements when it comes to errors on the end interface. On the other hand, focusing on data transmission brings much more severe requirements for error tolerance. This means that data transfer needs to be protected against errors to a much greater extent compared with the classical voice service. Error correction schemes also need to be better and more efficient. Additional requirements come from security point of view. Data transmission also requires more complex security mechanisms, especially when dealing with sensitive data, for example, in the case of e-commerce.

25.4.4 3G Systems

At the initial phase of standardization work for the 3rd Generation Partnership Project (3GPP) 3G, the approach of the researchers was to completely separate next generation 3G system from GSM. However, due to the overwhelming success of GSM networks and its evolutions, this approach was changed. This was in contrast to the GSM development which wasn't backward compatible with any of the 1G standards.

One of the first founding milestones for the new generation of 3G cellular mobile technology was the ITU publication on a series of IMT-2000 specific requirements that should be met in order for a telecom system to be classified as 3G compliant. Those initial requirements were as follows:

- Minimum 144 kbps user throughput for high movement speed in vehicular scenario
- Minimum 384 kbps user throughput for medium speed
- Minimum 2 Mbps user throughput for indoor scenario (office use)
- Voice quality comparable to the public switched telephone networks
- Flexible design for introduction of new services.

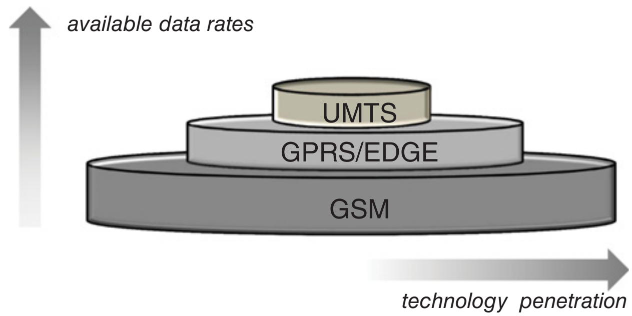

Figure 25.6 *Typical network deployment setup for early 3G rollouts.*

Without a doubt in the first years of 3G's existence, at the beginning of the new millennium, the Universal Mobile Telecommunication Systems (UMTS) networks were growing faster compared with their 2G counterparts a decade earlier. Profits derived from GSM operations fueled 3G deployment and mobile operators were aggressively trying to gain new customers by offering new technology and a totally new user experience. However early 3G network deployments were facing different problems than the ones encountered while GSM kick off. The approach toward the network rollout was also different. In case of UMTS systems the dense network of 2G base stations was already in place, hence the problem of finding new base station locations or antenna masts was less problematic, because in many cases the new base stations and antennas were collocated with the existing GSM elements.

For many 3G operators the business scenario for operating their cellular solutions was the same as in Figure 25.6:

- Provide a GSM coverage in the biggest area possible to provide voice service everywhere where it is needed.
- Complement the GSM technology with EDGE support in areas where there is a high possibility of data service usage (other than voice).
- 3G coverage in hot zone areas, to provide best user experience in limited, but lucrative scenarios.

Such a model allowed operators to gradually introduce the new 3G technology in the existing networks and to reflect operators' commercial priorities and quality of service expectations. EDGE was already a mature and reliable technology with a worldwide penetration. Moreover, EDGE provided similar user data experience as initial WCDMA solutions from Release 99, especially for the indoor users that benefited from better signal propagation in low frequency bands (EDGE) comparing to high frequency bands (WCDMA).

What was one of the most important objectives for the operators was to provide a seamless and undisrupted service change between those two technologies. This also affected the standardization process, where since the beginning of commercial 3G releases, UMTS networks supported seamless transition of some services like voice by means of 3G-2G inter system handovers.

Another different aspect of 3GPP 3G deployment in comparison to 2G rollouts was that WCDMA usually operated on a higher frequency bands than its GSM/EDGE counterparts. Higher frequency band means higher signal attenuation and less coverage. The latter is additionally limited by the fact that uplink coverage in WCDMA is changing depending on the load. Traffic imbalance between UL and DL (users tend to download more than they upload), which we observe in modern networks, was not the case in the first 3G rollouts, since at the beginning of this technology, despite its potential for the better user experience in data transfer domain, the key service and most common application that brought the majority of operators income was still a simple voice connection, where bit transfer in both directions is approximately the same.

The introduction of UMTS technology to the market meant that most of the radio network elements had to be replaced in case of early 3G deployments. However, the fact that usually new 3G stations shared their locations with the existing 2G BTSs had additional advantage that in many cases the backbone network could be reused for carrying 3G traffic and signaling. This advantage is not to be missed since in some cases the costs of backbone connection (installation and maintenance) can exceed the costs of a base stations itself.

Also on the core network side initial 3G setup could reuse part of the existing 2G equipment, usually by adding UMTS specific software and/or hardware to support new interfaces between core and radio network. Network nodes like MSC, SGSN, GGSN, Home Location Register (HLR), Visitor Location Register (VLR), Equipment Identity Register (EIR) and Authentication Center (AUC) are utilized in both standards. The first two elements, however, changed their functionality significantly in UMTS compared with the legacy standard. To facilitate 2G to 3G transition, the functionalities of the key GSM core network element MSC was split between the MSC Server and Media Gateway (MGW), which simply separated signaling handling from circuits or packet commutation. It lead to a better scalability and easier dimensioning of the core network, but in practice also allowed an efficient reuse of the old GSM equipment for UMTS purposes and supported unified core for both 2G and 3G users. Also the GPRS core network nodes could be relatively easily upgraded to 3G core nodes (cf. Figure 25.7).

It is also worth adding that enhancement standards like CAMEL are reusing similar concepts in 3G like Intelligent Network solutions in 2G networks.

25.4.5 Evolution of 3G Networks

3G has transformed the way that people think about using their mobile phones, but not in the way they were told to expect. Commercially available UMTS networks were available since 2001. Early Release 99 solutions provided theoretical maximum data rates of 2 Mbps as dictated by IMT-2000 specifications. In practice however, speed offered by operators was limited to 384 kbps. Since 3G UMTS and 2G family technologies (GSM, GPRS and EDGE) were developed by the same standardization body, 3GPP, many of the existing infrastructure elements, especially in the core network, are common for both generations.

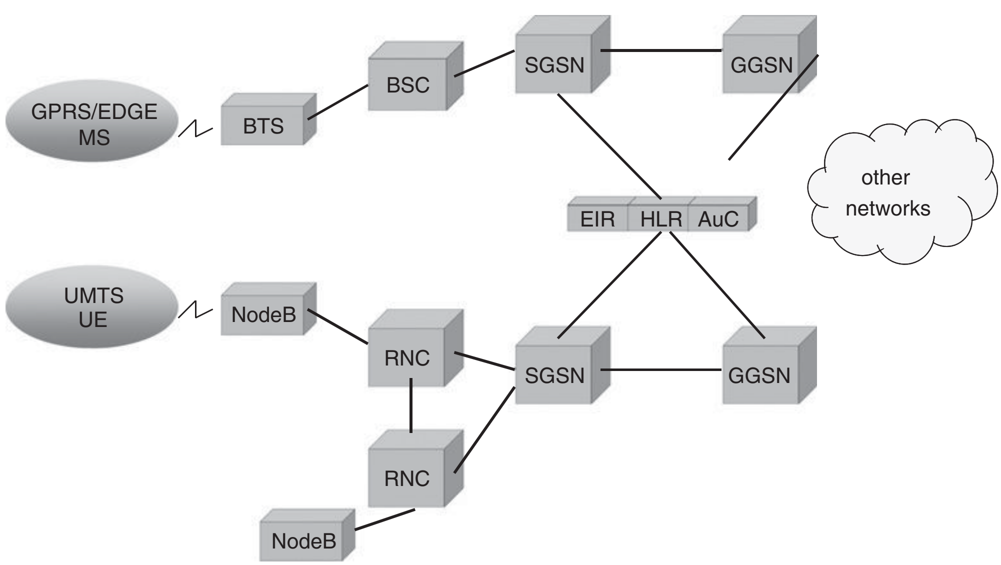

Figure 25.7 *Comparison of 2G and 3G architecture.*

The biggest revolution in 3G networks was the introduction of High Speed Downlink Packet Access (HSDPA) in Release 5 (year 2002) and High Speed Uplink Packet Access (HSUPA) (year 2004) known jointly as HSPA. Those enhancements focused solely on the packed transmission and evolved the UMTS standard to the next level. To highlight their meaning and to differentiate it from the Release 99 UMTS solution sometimes the name 3.5G is used when referring to HSPA (similarly as GPRS enhancement of 2G GSM network is known as 2.5G). Other names that can be found in the initial operator's advertisements of HSPA also highlight this change: UMTS Broadband, 3G+ or Internet Speed. One particular aspect that operators really embraced is that due to the focus on the packet switched domain only, HSPA has proven to be a much simpler standard compared with Release 99 which was designed to support different traffic types like voice, video or data by different means.

The HSPA and its evolution was already described in Section 12, but from the network perspective it is worth underlining that the introduction of many features, especially related to scheduling, was possible because of bolstering the importance of NodeB element in the Radio Network. In pre Release 5 networks the NodeB was a simple device that was used to convert the information conveyed by the transport network to a suitable form for the radio interface and to provide wireless channel measurements. All decisions related to the scheduling, like the choice of transmission format or ACK/NACK handling, were done at the Radio Network Controller (RNC) side. One of the reasons for this particular setup was the conformance with the legacy standard. During the time of GSM hype reliable digital elements that could provide aforementioned functionalities were expensive, and this was also the case in the time of Release 99 standardization. Moving high computational power and capabilities to the multiple base stations was simply not cost effective. After a couple of years, however, the prices of such elements dropped significantly, allowing moving scheduler's decision functionality to NodeB, closer to the air interface itself, hence reducing the reporting delays and allowing for a better adjustment to dynamic channel conditions and further exploitation of the wireless channel's capabilities. This transition allowed boosting downlink download capabilities to the speed of up to 14.4 Mbps at the peak data rate in Release 5 and for the upload speed in Release 6 up to 5.76 Mbps.

This huge upgrade in the cell's capacity had also the negative side effect as in many cases introduction of HSPA triggered the necessity for the backhaul upgrade. Initial 3G network deployments, despite the variety of supported services like video or data, were mainly dimensioned for voice centric user consumption, hence many 3G base station sites were linked with the RNC using simple copper wires based on E1 or T1 circuits, which are able to carry not more than 2 Mbps or 1.5 Mbps of data respectively. In case of moving to HSPA technology, if the traffic growth followed, the backhaul had to be upgraded to multiple E1/T1 connections, or sometimes even replaced with an optical fiber since now the maximum throughput in one 5 MHz sector could be as high as 14.4 Mbps and the base station site usually comprises three sectors, not to mention the ability to have multiple 5 MHz carriers introduced in further HSPA releases. Of course the peak data rates are observed very rarely, but the cases where the backhaul is the bottleneck of the user transmission is usually unacceptable for the operators as it simply means an inefficient network planning.

In Release 7 the process of moving the RNC functionalities to the NodeB has gone even further. The evolution of the HSPA standard, that among others aimed also at latency reduction, resulted in the flattening of the overall network through shifting of most of the RNC functionalities to the NodeB, allowing NodeBs for direct tunneling of user plane to GGSN, without the involvement of RNC and SGSN on a user plane level. The reason for this is that latency in networks is for many modern services and applications as important as peak data rates, especially for services like online gaming or streaming. With flat HSPA architecture and direct tunneling the round-trip time of the packets can be decreased by almost a factor of two [5]. Direct tunneling feature requires changes only in SGSN and doesn't affect existing GGSN platforms.

In contrast to most network enhancements, it doesn't have any impact on the radio interface or user devices, which allows operators to improve their network without the necessity of introduction of new user terminals to the market. The additional value of such a solution, besides simplification of the deployment, is elimination of man work for laborious RNC and SGSN redimensioning in case of NodeB capacity upgrades like addition of new carriers. The other advantage not to miss is that the flat architecture streamlines the evolution of HSPA network towards the next generation of mobile cellular technology, namely LTE. The reader should keep in mind that in case of control plane handling the SGSN node is still needed in a flat HSPA architecture to facilitate the tunneling operations of user plane.

Additionally to the flat architecture Release 7 brought several other enhancements to the HSPA systems. Since they have already been discussed in Chapter 13 we will limit the evolution to mention the most important features of HSDPA:

- Release 7: 64 Quadrature Amplitude Modulation (QAM) or 2 x 2 Mulitple Input Multiple Output (MIMO). The first one required upgrades in the radio part of the NodeB. The latter needs at least two receive antennas at User Equipment (UE) side and two transmit antennas at the NodeB side. In return the peak data rate was uplifted to 21 Mbps (64QAM) or 28 Mbps (2 x 2 MIMO).
- Release 8: Dual Carrier (DC) operations with 64QAM (DC + 64QAM) or 2 x 2 MIMO with 64QAM pushed downlink peak data rates to 42 Mbps. The first feature requires 10 MHz of continuous bandwidth to be available at the operator side. Compared with the 5 MHz bandwidth requirement from pre Release 8 it is the first step towards spectrum utilization similar to the one defined by LTE Release 8 that could accommodate up to 20 MHz of bandwidth.
- Release 9: Dual Carrier operations with 2 x 2MIMO and 64QAM: 84 Mbps
- Release 10: Quad Carrier (QC) + 2 x 2MIMO + 64QAM: 168 Mbps, 20 MHz of bandwidth required
- Release 11: QC + 4 x 4 MIMO + 64QAM or Eight Carriers (8C) + 2 x 2MIMO + 64QAM: 336 Mbps, first feature requires 40 MHz of bandwidth, the second one requires at least four transmit antennas at NodeB side and minimum four receive antennas at UE side for 4 x 4 MIMO operations.

25.4.6 4G Systems Considerations

In March 2008 ITU-R submitted a circular letter inviting proposals for candidate radio interface technologies for the next generation of cellular mobile standard known initially as IMT-Advanced (IMT-A) [6]. At the end of the same year in three ITU-R reports followed [7–9] containing evaluation framework and a set of technical requirements that have to be met for a radio interface technology in order to qualify as an IMT-A radio, and effectively to be considered as a 4G. These two sets of releases officially set the framework of what is an IMT-A radio. In October 2010, LTE-Advanced and Wireless Interoperability for Microwave Access (WiMAX) radio technologies passed the milestone and became officially acknowledged as 4G solutions.

The IMT-A requirements for the qualifying radio technologies to meet certain minimum levels depend on system's peak spectral efficiency, bandwidth flexibility, latency and handover interruption. These key requirements, mostly oriented at enhancing user experience, are shown in Table 25.1.

Table 25.1 *Key IMT-A requirements*

Requirement name	Requirement
Peak spectral efficiency, downlink	$\geq$ 15 bits/s/Hz
Peak spectral efficiency, uplink	$\geq$ 6.75 bits/s/Hz
Bandwidth flexibility	up to 40 MHz
Latency, user plane	$\leq$ 10 ms
Latency, control plane	$\leq$ 100 ms
Handover interruption, intrafrequency	$\leq$ 27.5 ms
Handover interruption, interfrequency	$\leq$ 40 ms
Handover interruption, interband	$\leq$ 60 ms

Table 25.2 *Key environment-specific IMT-A requirements*

Requirement	Unit	InH	UMi	UMa	RMa
Average spectral efficiency, downlink	bits/s/Hz	3.0	2.6	2.2	1.1
Cell edge spectral efficiency, downlink	bits/s/Hz/user	0.1	0.075	0.06	0.04
Average spectral efficiency, uplink	bits/s/Hz	2.25	1.8	1.4	0.7
Cell edge spectral efficiency, uplink	bits/s/Hz/user	0.07	0.05	0.03	0.015
High speed mobility spectral efficiency	bits/s/Hz	1.0	0.75	0.55	0.25
Voice over IP (VoIP) capacity	UEs/MHz	50	40	40	30

Additionally, specific requirements regarding average and cell edge spectral efficiency were defined for a set of diverse deployment environments, namely for: indoor hotspot (InH), urban microcellular (UMi), base coverage urban macro (UMa) and high speed (rural macro, RMa). These environment-specific requirements are gathered in Table 25.2, whereas most important characteristics for each environment are summarized in Table 25.3.

Values presented in Table 25.2 can be obtained through system and link level simulations according to the guidelines provided by IMT-A. To exemplify, cell edge spectral efficiency in the downlink can be obtained using system level simulations where base stations are forming a well known hexagonal urban cell layout and user terminals are dropped uniformly across the simulation area resulting in density of 10 users per sector. Cell edge throughput is evaluated as the throughput value at the fifth percentile of cumulative distributed function of particular user terminal throughput at physical layer, obtained over several simulation realizations (drops), where the user position and other stochastic variables are randomized again and again. Calculated cell edge throughput value is further divided between the amount of spectrum utilized by the base station for the transmission in downlink. The end spectral efficiency value of cell edge users denotes the efficiency of spectrum utilization by users with most challenging radio conditions (usually located at the boarder of the cell, exposed to a large extend to the intercell interferences).

Not all IMT-A requirements need to be evaluated by means of system simulations. Some of them, like downlink or uplink peak spectral efficiency presented in Table 25.1, can be derived out of analytical analysis.

As for the values in Table 25.3 they should be treated only as a guideline for simulation purposes. Despite being close to real life solutions, by no means should they be treated as requirements for realistic realization of the 4G systems.

Table 25.3 *Environment and deployment characteristics for IMT-A*

	InH	UMi	UMa	RMa
Base station antenna height	6 m	10 m	25 m	35 m
Maximum number of antenna elements at base station side	8 Rx 8 Rx	8 Rx 8 Tx	8 Rx 8 Tx	8 Rx 8 Tx
Base station transmit power per 20 MHz band	21 dBm	44 dBm	49 dBm	49 dBm
User terminal antenna system	2 Rx 2 Tx	2 Rx 2 Tx	2 Rx 2 Tx	2 Rx 2 Tx
User terminal maximum transmit power	21 dBm	24 dBm	24 dBm	24 dBm
Minimum distance between user terminal and base station antenna	3 m	10 m	25 m	35 m
Carrier frequency used for evaluation	3.4 GHz	2.5 GHz	2 GHz	800 MHz
Intersite distance	60 m	200 m	500 m	1732 m
User terminal speed	3 kmph	3 kmph	30 kmph	120 kmph

25.4.7 Can HSPA+ Become a 4G System?

The need for HSPA evolution has already been taken seriously by 3GPP, where further enhancements into HSPA are being proposed and standardized. In fact, many of these proposed enhancements improve the technology so substantially that it is relevant to review its performance in the context of IMT-A, as proposed by ITU-R for the so-called 4G systems [10].

To meet the peak spectral efficiency and bandwidth flexibility of IMT-A requirements, advanced upgrades for HSPA are needed. Therefore features such as 4x4 MIMO for downlink, 2x2 MIMO with 64QAM for uplink peak spectral efficiency, and 8-Carrier HSDPA for bandwidth flexibility should be considered. All of them were under development in 3GPP, and were finalized in the end of 2012 and included in Release 11.

Despite the theoretical capability to pass the IMT-A requirements HSPA is not perceived as an IMT-A technology. Additionally the costs of providing 4G like user experience with evolved HSPA systems are higher comparing to LTE systems. As one of the reasons for this we could give higher spectral efficiency of LTE Orthogonal Frequency Division Multiplexing (OFDM) based solutions comparing to HSPA access utilizing WCDMA approach. Frequency selective scheduling allows better utilization of multipath environment and to allocate frequency resources in the form of smaller frequency chunks in a more optimal manner as compared with WCDMA where users share whole 5 MHz band simultaneously. Additionally OFDM based solutions are more suitable for MIMO, due to the fact that they allow for simple frequency domain equalization. Also according to real life measurements, despite the technical possibilities to provide a 4G experience, at the beginning of 2013, download speed in HSPA+ networks were 3x lower than the ones experienced in LTE networks [11].

25.4.8 4G Systems

Only two 4G systems were deployed so far: WiMAX and LTE. Initially those two standards were considered as competing and despite all 3GPP support, many forecasts suggested that WiMAX may dethrone LTE, giving the fact that many big industry players (Intel, Fujitsu, Samsung, AT&T and other) showed great interest in this technology. Additionally WiMAX was commercially available around 2, 3 years ahead of first LTE commercial deployments. In the end however, despite its great potential, in 2012 after a massive LTE rollouts, WiMAX was rather perceived as a supplementary technology used to provide the broadband connectivity by the greenfield operators which didn't have a legacy 3G networks in place. At the same time, in spite of the fact that for most mobile operators it was still 3G which brought most of the operator's revenue, LTE solution was perceived as a natural and inevitable evolutionary step in providing the cellular wireless connections to the end-user. It is worth underlining that LTE standard, as for the state when it was introduced to 3GPP standard family (Release 8 and Release 9), can't be considered as a full 4G compliant, although in mass media and commercial announcement it was advertised as such. Some operators even tried to coin their own phrases like 3.9G or super 3G, but none of those ever gained massive support in the wireless ecosystem. Real 4G networks, as defined by ITU was brought by 3GPP Release 10 standard revision and it is usually referred to as LTE-Advanced. However, throughout this section the term 4G will be jointly used for both LTE and LTE-Advanced.

The initial specifications for LTE, that is the ones of Release 8, were finalized by 3GPP in 2008 and the IMT-A standards were ratified by ITU in November 2010 [12].

The first LTE networks appeared at the end of 2010 while the first LTE-A ones started to be commercially available in 2013. The early deployments of 4G networks were quite different from the deployments of its predecessor. First of all LTE networks from the beginning were data oriented. In the first 3G rollout, despite the announced capabilities for video calls and enhanced speed of data transmission, the lack of terminals that could facilitate broadband experience narrowed down the UMTS services to simple voice connection. On the other hand LTE, from the very first 3GPP release of the 4G standard, was packed oriented. It was standing on the shoulders of the HSPA evolution that set the path to the broadband user experience (devices like Apple's iPhone or Google's Android based user terminals were already in large volumes in the market). Simple voice connections in LTE networks could be realized only via VoIP applications as classical voice service is not part of the LTE or LTE-A standard. Hence to provide this common service a redirection by means of, for example, inter Radio Access Technology (RAT) handovers to the 3G or 2G legacy system is required for the operators who want to provide it. As a consequence of this simple dependency there were two large groups of operators that rolled out early 4G networks.

In the first group we could find classical 3GPP operators for whom the deployment of LTE networks was a natural evolutionary step of their cellular network. The goal that they wanted to achieve was to provide faster wireless connections in hot zone areas, but the transfer speeds were not the only important factor for many of them. Despite the obvious marketing potential of a "new 4G technology" one of the key challenges was to provide wireless broadband by cheaper means compared with HSPA. LTE, though undeniably faster than its complementary HSPA releases, didn't bring such a big quality leap as for instance the introduction of 3G WCDMA solutions after 2G networks. LTE offers unquestionably faster transfer speeds, especially when it comes to utilization of MIMO technology and higher bandwidths, but its advantage also lies in the ability to provide faster data rates by cheaper means compared with HSPA technology. Those operators have initially deployed LTE access nodes in hot zone areas to provide high capacity and fast connections in places where existing HSPA networks were throttling.

The rollout of the first 4G networks was also fueled by micro and pico cells deployment as those solutions were gaining momentum in the timeline of advent for 4G. Micro/pico cells were already known since GSM, but initially in many cases they were perceived as a simple and economical means for the coverage extensions in places where macro stations signal couldn't reach, for example, due to high path losses. By the end of the first decade of the twenty-first century network operators' mindset had already changed and micro and pico cells, together with emerging femto technology, were starting to be seen as the only way to offload macro stations and the easiest means to provide a high capacity over limited areas, especially indoors.

Higher operating frequencies in 2.6 GHz band of the LTE were perfectly fitting this approach, providing higher signal attenuation and effectively better cell's isolation in terms of interferences. Of course macro scenarios utilizing 4G technology were still an attractive option for mobile network operators, especially in 800 MHz band. Another factor that propelled new network rollouts was societal demand. Fast broadband 4G was perfectly fitting new types of communication based on social applications like Facebook or Youtube that usually required high Internet speeds. LTE users that utilized such applications via mobile cellular networks were mostly equipped with tablets or smartphone devices and instead of classic telephony they communicated using Internet based solutions. From the operator's side the financial aspect not to miss was also the ability of 4G to cover a flexible amount of spectrum starting from 1.4 GHz and 3 GHz 1.4 GHz and 3 GHz. This flexibility allowed many operators to efficiently refarm the narrowband spectrum that they already had. In WCDMA/HSPA networks the lowest amount of spectrum with full guard bands that could be utilized was 5 MHz (or little less with limited guard bands and the risk of possible degradation of the service quality on such carriers).

Another large group of first LTE users that could be distinguished were people equipped with LTE capable USB dongles to whom LTE technology was suppose to be a substitution for a fixed broadband connection. This group of users was especially attractive for operators who wanted to provide wireless broadband in rural areas lacking fixed Internet infrastructure. This new group of LTE providers was in many cases consisting of broadband Internet providers who just entered the cellular business, previously focused on providing a fixed line broadband and had no existing cellular network infrastructure.

One could also distinguish a third quite specific group of LTE providers. These were the operators who didn't have an existing WCDMA/HSPA networks in place, but were already in the cellular mobile business. They decided to move directly from 2G to 4G, or they had a Code Division Multiple Access 2000 (CDMA 2000) based networks instead of UMTS.

An important factor that triggered faster rollout of the LTE technology was the so-called traffic avalanche, the phenomenon experienced in wireless networks during the time of the introduction of 4G technology to the market. Most operators were experiencing a massive explosion of data consumption in their networks, hence the rollout of the 4G was in many cases much faster compared with early 2G and 3G networks rollouts. LTE technology, despite its short presence on the market, had also a significant impact on this upsurge of traffic numbers. In 2012 4G connections generated 19 times more traffic on average than non-4G sessions. 4G connections were representing only about 1 percent of mobile connections, but they accounted for 14 percent of the total mobile data traffic [13]. Fast LTE deployment was also supported by the National Regulatory Authorities. For instance in Canada during the auction of 700 MHz band for LTE the government obliged the operators that control two or more of paired spectrum in this band to provide a broadband service for 90% of country population within 5 years' time and for 97% after 7 years [14].

To boost the infrastructure development the 4G spectrum auctions in many countries ended up with quotes much lower compared with similar 3G auctions at the beginning of UMTS introduction to the market. As an example we can give a 4G spectrum auction in United Kingdom, where the rights to operate on frequencies in 800 MHz and 2600 MHz bands were sold to operators for only $3.5 billion, which is ten times lower compared with $35 billion paid for the 3G frequencies in the year 2000 [15]. Such standpoint of National Regulatory Authorities may be considered as a positive for the 4G network development as the saved money could be spent on the infrastructure development.

Also from the 3GPP standardization perspective there were many features supporting LTE rollouts. From the very first LTE Release 8 standard revision, 3GPP has specified several useful SON enabling functions such as Automatic Neighbor Relation (ANR) or Mobility Load Balancing (MLB). ANR facilitate faster and cheaper rollout of LTE networks by reducing manual configuration costs of each eNodeB. This is achieved by the automatized handling of so-called Neighbor Relation Tables (NRT). NRT kept in each eNodeB stores the information of the eNodeBs neighbors' identifiers and this information can be used in a potential handover and mobility management procedures. The concept of such lists were already known in 2G and 3G systems, but it required labor intensive manual input from maintenance teams of those legacy technologies. By using the ANR feature, NRT lists are updated using UE measurements, hence no human action is required which significantly reduces efforts and minimizes overall costs of the new node deployment. The second mentioned feature, MLB, makes use of a cell load and capacity information exchanged between eNodeBs over X2 interface. This particular concept introduces a procedure to negotiate handover setting between different eNodeBs. As a result radio network can achieve better and more balanced resource utilization across different cells, by means of, for example, negotiating handover setting.

As a consequence of these parameter negotiations, some UEs connected to congested cells may be handed over to cells with probably less favorable channel conditions, but with potentially lower load and lower utilization of radio resources. This allows an even load distribution among different cells and results in better overall user experience. Another SON feature which is useful from the network deployment perspective is Mobility Robustness Optimization (MRO). This mobility related mechanism introduced in Release 9 and enhanced with the inter system capability option in Release 10, collects different Key Performance Indicators (KPIs) of network performance relevant to the mobility aspects, such as: number of too early or to late handovers, call drops, ping pong between different cells (situation when user is repeatedly handed over between the same cells), radio link failures and so on. Based on those KPIs, MRO functionality modifies mobility related parameters in the eNodeB in order to improve mentioned KPIs. This feature is especially interesting in its inter Radio Access Technology version for the early LTE rollouts, as the new 4G cells are typically placed in an already existing cellular environment consisting of a overlapping 3G and 2G coverage areas. Manual setting of optimal mobility parameters (like handover thresholds, handover hysteresis values etc.) is usually a very time consuming task which of course can cost a lot of man effort. MRO allows for automatization of this process and additionally provides means to dynamically adapt to changing traffic.

The end result in terms of network deployment, of all aforementioned factors was visible very fast. In the beginning of 2013, only 3 years after the introduction of LTE, there were 62 countries with LTE networks, 19 another had LTE networks scheduled for deployment in 2013 [11]. The country with the most LTE networks was USA with total of 17 LTE running. The revenue of LTE business was also increasing fast, reaching 115% growth in 2012 comparing to the previous years [16].

As a matter of historical heritage and tendencies, first LTE networks rolled out by established operators were mostly Frequency Division Duplexing (FDD) based. On the other hand greenfield players in the wireless business, like the ones described above, geared rather toward LTE Time Division Duplexing (TDD). This particular variant of LTE didn't gain a worldwide attention initially, being perceived by some people as a niche technology (based on the 3G experience with UTRAN TDD whose large-scale deployments were limited to China). However many new operators started to deploy their networks based on LTE TDD. The reason for that is TDD's interesting features that are attractive to newcomers. The most important ones are mainly cheaper spectrum (no need to buy paired spectrum) and cheaper devices (single receiver/transmitter chain for one frequency).

It was also an interesting option for WiMAX operators that already have a TDD spectrum, TDD technical expertise and would like to switch to LTE. Additionally TDD technology is perfect for asymmetric traffic as people tend to download much more than they upload. Adaptation for the asymmetric traffic can be obtained by time domain resource split for UL and DL transmission. FDD technologies can adapt to this phenomenon to a far smaller extent as symmetrical paired spectrum is required regardless of the traffic asymmetry, even at the price of underutilization of the resources. According to the Global LTE-TD Initiative (GTI), an open platform to advocate cooperation among global operators to promote LTE-TDD, in June 2012 there were 52 trial networks using this duplexing (majority of them in Asia or Europe), 10 commercial networks launched, 22 publicly announced TDD LTE contracts and at least 26 operators had a clear TDD LTE commercial deployment plans [17].

25.5 Spectrum Considerations for Network Transition

The air interface is usually the bottleneck of the wireless mobile systems. In the network evolution process newest wireless standards utilizing techniques like OFDM, Coordinated Mulltipoint (CoMP), advanced receivers and Higher Order Modulation combined with low error protective channel coding are slowly reaching the physical limits for the provision of a certain data capacity within a given bandwidth bounded by Shannon's Theorem [18]. Modern standardization processes are shifting the efforts from providing higher peak data rates towards provision of more uniform service offer (i.e. fairness) across the cell and on the interference prevention and mitigation techniques. Bandwidth is a very scarce resource, but on the other hand, with the peak spectral efficiency of modern systems which is close to the theoretical limit for users in favorable radio conditions, one of the most obvious methods for providing higher network capacity, besides not trivial network nodes densification, is the increase in bandwidth which the data transmission may utilize. One should also remember that in contrary to the network densification where gains are not scalable (in interference limited scenarios 10x more base stations do not give 10x higher network capacity but much less), the capacity of the network is fairly proportional to the bandwidth that it may utilize. In addition some frequency diversity gains can be expected on top of this proportional capacity increase while performing in the multicarrier bandwidth [19].

One good example of this available bandwidth extensions that was plugged into the real life, is the concept of multiple carriers operation for HSPA that was introduced in Release 8 in the form of Dual Cell HSDPA and was later followed by the Quad Carrier HSDPA in Release 10 and 8 Carrier HSDPA in Release 11. Dual Cell HSDPA allows operating on 2×5 MHz, Quad Carrier on 4×5 MHz and consequentially 8 Carrier allows for utilization of 8×5 MHz of bandwidth. Also for the uploads HSUPA offers an upgrade from the inherent 5 MHz carrier concept by means of Dual Cell HSUPA from Release 11.

In LTE networks this concept was further refined. While HSPA is only able to operate on bandwidth consisting of multiples of 5 MHz, with LTE standard the operators may utilize their frequency band chunks of the sizes of 1.4 MHz, 3 MHz, 5 MHz, 10 MHz, 15 MHz and 20 MHz. Further evolution in the form of LTE – Advanced standardization allows for the utilization of multicarrier feature that can simultaneously exploit up to 100 MHz of spectrum by the combination of five 20 MHz spectrum parts. It is worth remembering that throughout the standardization process the ability to pool carriers not only in the adjacent frequency bands, but also on non adjacent and separated ones (so-called multiband operations) were added to both HSPA and LTE.

The total share of global spectrum for mobile cellular usage is coordinated by the International Telecommunication Union and the most important decisions about the global spectrum allocations are announced on a World Radio Conferences (WRC). However, the utilization form of the given bandwidth within each country is in many cases regulated by National Regulatory Authorities. This is why the 3GPP standardization body covers a whole range of available frequencies for their technologies. Below the reader can find the possible spectrum options for LTE operations based on Release 11 specifications.

Obtaining the right to operate in one of those frequency bands is not a trivial task. In many countries a license for the exclusive utilization of a given band can only be obtained via so-called spectrum auctions. As already mentioned, exclusive utilization of a specific spectrum over the limited area (e.g. country wise) is usually linked with high capital investment. As an extreme case we can recall a 3G spectrum auction in UK that was carried out in 2000. For the possibility to utilize certain amount of spectrum for 3G transmissions, five different 3G operators paid together the overwhelming amount of £23 billion GBP ($35 billion).

In case of 3G systems most of the operators used the paired spectrum in FDD mode. As for the TDD mode for WCDMA, a standard called UMTS-TDD based on the 1.28 Mcps low cheap rate is available and commercially deployed, mostly in China, where in October 2011 more than 40 million of people were using it [22]. Its 4G counterpart is LTE TDD and it seems to have a potential to reach a far greater number of users and a worldwide penetration. The spectrum for the operation on TDD band is much cheaper than in the case of FDD solutions and TDD systems are far more suitable for asymmetric traffic than its FDD counterparts. Also the ITU regulations foresaw the demand for this kind of solutions and assigned among others a 100 MHz of bandwidth in attractive 2.3 GHz region for TDD operations.

What is also visible from Table 25.4 is that LTE can operate on existing 2G and 3G bands. Especially refarming of existing 2G spectrum can be considered as an enticing solution for the operators giving the declining of 2G traffic in favor of 3G and 4G uptake. For the LTE the 2.6 GHz spectrum allocation is foreseen as especially attractive for the urban areas where the base station deployment is dense in order to provide high capacity over limited areas. The lower bands are also heavily utilized. As an example we could give a case of Germany where first commercial LTE networks were launched in 800 MHz band for the rural broadband services [23]. Also in US the operators are planning to deploy their LTE networks using 700 MHz bands [24]. In Japan 2.1 GHz and 1.5 GHz spectrum is available for the LTE deployments [25].

One of the particular interests of wireless operators, vendors and regulations bodies are digital dividend parts of spectrum released in the process of the transition from analog to digital TV. The example already mentioned is 700 MHz spectrum that was auctioned in the USA for wireless operations.

So far the most popular form of spectrum utilization among operators in cellular ecosystems is individual (licensed) access, which gives exclusive rights to use the certain spectrum at a given time and geographical region (usually country or a state) to a single operator. One of the options for deciding on who has the privilege to use the given spectrum chunk are the spectrum auctions, which have already been mentioned. Completely different approaches are offered by shared spectrum methods. An example of such a solution would be the usage of Wi-Fi technology in Industrial, Scientific and Medical (ISM) radio band.

Table 25.4 *LTE operating bands [20]*

LTE Operating Band	Uplink operating band BS receive UE transmit	Downlink operating band BS transmit UE receive	Duplex mode
	$F_{UL\_low} - F_{UL\_high}$	$F_{DL\_low} - F_{DL\_high}$	
1	1920 MHz–1980 MHz	2110 MHz–2170 MHz	FDD
2	1850 MHz–1910 MHz	1930 MHz–1990 MHz	FDD
3	1710 MHz–1785 MHz	1805 MHz–1880 MHz	FDD
4	1710 MHz–1755 MHz	2110 MHz–2155 MHz	FDD
5	824 MHz–849 MHz	869 MHz–894MHz	FDD
6[1]	830 MHz–840 MHz	875 MHz–885 MHz	FDD
7	2500 MHz–2570 MHz	2620 MHz–2690 MHz	FDD
8	880 MHz–915 MHz	925 MHz–960 MHz	FDD
9	1749.9 MHz–1784.9 MHz	1844.9 MHz–1879.9 MHz	FDD
10	1710 MHz–1770 MHz	2110 MHz–2170 MHz	FDD
11	1427.9 MHz–1447.9 MHz	1475.9 MHz–1495.9 MHz	FDD
12	699 MHz–716 MHz	729 MHz–746 MHz	FDD
13	777 MHz–787 MHz	746 MHz–756 MHz	FDD
14	788 MHz–798 MHz	758 MHz–768 MHz	FDD
15	Reserved	Reserved	FDD
16	Reserved	Reserved	FDD
17	704 MHz–716 MHz	734 MHz–746 MHz	FDD
18	815 MHz–830 MHz	860 MHz–875 MHz	FDD
19	830 MHz–845 MHz	875 MHz–890 MHz	FDD
20	832 MHz–862 MHz	791 MHz–821 MHz	FDD
21	1447.9 MHz–1462.9 MHz	1495.9 MHz–1510.9 MHz	FDD
22	3410 MHz–3490 MHz	3510 MHz–3590 MHz	FDD
23	2000 MHz–2020 MHz	2180 MHz–2200 MHz	FDD
24	1626.5 MHz–1660.5 MHz	1525 MHz–1559 MHz	FDD
25	1850 MHz–1915 MHz	1930 MHz–1995 MHz	FDD
26	814 MHz–849 MHz	859 MHz–894 MHz	FDD
27	807 MHz–824 MHz	852 MHz–869 MHz	FDD
28	703 MHz–748 MHz	758 MHz–803 MHz	FDD
...			
33	1900 MHz–1920 MHz	1900 MHz–1920 MHz	TDD
34	2010 MHz–2025 MHz	2010 MHz–2025 MHz	TDD
35	1850 MHz–1910 MHz	1850 MHz–1910 MHz	TDD
36	1930 MHz–1990 MHz	1930 MHz–1990 MHz	TDD
37	1910 MHz–1930 MHz	1910 MHz–1930 MHz	TDD
38	2570 MHz–2620 MHz	2570 MHz–2620 MHz	TDD
39	1880 MHz–1920 MHz	1880 MHz–1920 MHz	TDD
40	2300 MHz–2400 MHz	2300 MHz–2400 MHz	TDD
41	2496 MHz–2690 MHz	2496 MHz–2690 MHz	TDD
42	3400 MHz–3600 MHz	3400 MHz–3600 MHz	TDD
43	3600 MHz–3800 MHz	3600 MHz–3800 MHz	TDD
44	703 MHz–803 MHz	703 MHz–803 MHz	TDD

Note 1: Band 6 is not applicable.
Source: Data by courtesy of 3GPP.

Table 25.5 Winning bids for the 3G spectrum in UK. Bids given in 10ˆ6 GPB [21]

Licence	Paired spectrum [MHz]	Unpaired spectrum [MHz]	Total spectrum [MHz]	Minimum opening bid	Winner	Winning bid
A	2×15	5	35	5	TJW	4385
B	2×15	0	30	30	Vodafone	5964
C	2×10	5	25	25	BT	4030
D	2×10	5	25	25	One2One	4004
E	2×10	5	25	25	Orange	4095

Source: Data shown in ITU spectrum management white paper.

Despite the fact, that Wi-Fi traffic offloading in cellular systems is in the scope of interest to many operators, those ISM frequency bands in most cases are not considered by cellular operators due to the lack of effective means to provide reasonable radio link quality or guaranteed coverage in this shared medium as it can be utilized by anyone. As an intermediate solution between exclusive and shared spectrum usage, alternative ways can be considered such as Light Licensing or Authorized Shared Access (ASA) also know as Licensed Shared Access (LSA). The background of those solutions is an assumption that a given chunk of spectrum, belonging in a licensed approach exclusively to a single licensee, is not always utilized to a full extent or even not utilized at all in some time periods, so it can be made temporarily available for other interested parties. The first of the aforementioned schemes, Light Licensing, has a broad definition and may be applicable to both individual authorization (exclusive rights of use) or general authorization schemes (no individual rights of use), but is characterized with simplified procedures of acquiring given spectrum part. In many countries only a small fee is necessary to cover the cost of the registration used in Light Licensing approach. This scheme may also allow limiting the number of users who can use a given spectrum chunk and is applied in situations where the given bandwidth future usage may be changed, but for the given moment there is no concern about interference issues. An example where Light Licensing authorization scheme was issued is the 5.8 GHz band in some European countries [26, 27].

The second of the mentioned schemes, Licensed Shared Access, is a typical general authorization scheme. This regulatory concept allows shared use of spectrum under well defined and predictable conditions [28]. In the trivial example one can imagine the original spectrum holder (LSA incumbent), who can provide temporary access to his frequency resources to users subscribed to a friendly operator (LSA licensee), under the specifically pre-agreed individual conditions providing that it doesn't affect the Quality of Service (QoS) of his original subscribers who are utilizing the same bandwidth at the same time, or that the QoS of those users doesn't drop below a certain level. Giving the fact that spectrum is a limited and expensive resource, the possibility of flexible authorization schemes is a very attractive solution for mobile operators. The additional advantage of such an approach is the potential to reduce the interference level in the operator's network. Interference mitigation can be achieved through reliable sharing arrangements based on clear, effective sharing rules and conditions in a band, creating certainty for both incumbent and licensee users [29]. The importance of shared access methods was recognized by the European Commission which issued a full report highlighting the importance of shared spectrum access, assessing and promoting its socioeconomic values [30]. It is also worth mentioning that some parts of the spectrum that are of interest for mobile operators are exclusively reserved by ITU-R for shared access.

So far 3GPP cellular technologies have operated in radio bands ranging from 300 MHz up to 3 GHz, at least in the radio access domain. Recent wireless technologies have also utilized 5 GHz band (e.g. 802.11a, 802.11n, 802.11ac). However, the amount of spectrum below 10 GHz which may be harnessed by the network operators for cellular communication is limited and as a consequence spectrum becomes increasingly crowded. This is the reason why many vendors are starting to consider frequency ranges associated so far with radio astronomy, radars or satellite communication – the millimeter waves (30 GHz – 300 GHz). The usage of this spectrum is more challenging comparing to a typical cellular spectrum: mm waves propagate only in line of sight and are very prone to atmospheric changes, which results in much higher signal absorption during rain periods. They can also suffer from heavy scattering losses from flora (plants, trees etc.). On the other hand, line of sight propagation makes it ideal for wireless backhauling and such solutions are already widely deployed using the microwave links technology. Additionally increased signal attenuation makes it more attractive in interference limited deployment, for example, in the case of very dense networks topologies. If we also consider that mm waves spectrum is largely underutilized and there are large adjacent spectrum chunks available for mobile communication in this region, then it is more than likely that such solutions will come in the future. Some first implementations will come in the form of IEEE 802.11ad (WiGig) that operates on 60 GHz. Millimeter waves technology is also a strong candidate for one of the key pillars of next generation of wireless cellular mobile technology, especially in a local access domain for small cells.

25.6 Terminals Support for the Network Transition

Practically every new upgrade of the Radio Access and the Core Network requires corresponding changes in the implementation of user's terminals. Many such schemes are a proprietary terminal manufacturer's solutions that are only bounded with the end performance requirements specified by standardization bodies like 3GPP, which don't propose a specific technical approach, but rather sets some performance constraints for them.

First mobile radio user devices used in 1G networks were bulky and inconvenient terminals, heavy, large in volume and with a very limited battery work time. Due to the large dimensions they were often mounted in cars or ships. In a portable version they were perceived as a luxury device, usually associated with the business class.

The advent of 2G changed this situation. Limits set upon the total radiated output power emission mobile devices, initially made up to address the health concern issues, had a positive side effect in the form of solutions and implementations that limited battery drain consumption. This, in combination with the digitalization of the radio equipment of user devices, has lead to much smaller device sizes and extended inter charge operability time. Together with the growing popularity of a mobile cellular technology it created a critical mass for cheaper large scale production of the equipment, which became affordable for a wider circle of consumers. An additional factor contributing to the proliferation of 2G technology that should not be missed are the security aspects posed for 2G networks, which are far more advanced and stringent compared with 1G legacy systems. In this sense 2G phones became much more private devices. This tendency is continued up to nowadays and recently we can observe an increased number of forecasts suggesting that in the nearest future mobile cellular devices can even replace our wallets.

Looking backwards in the evolution of mobile devices we can see that the first products were in the majority of cases nothing more than a portable equivalent of plain old phones, that were used for nothing more than making voice calls. This is why the overwhelming success of the Short Message Systems (SMS) in 2G was a huge surprise to both vendors and operators. SMS was one of the first features that differentiated the new mobile cellular devices from corded phones. The inevitable step in the evolution of mobile equipment was data transfer over the air, from and to portable cellular equipment. Initially mobile devices that allowed this over-the-air data exchange via cellular networks did not resemble modern handhelds; typically a separate modem was used to facilitate the transmission. On the infrastructure side one of the first big steps toward the user-friendly wireless cellular data exchange could be associated with 2G Dual Transfer Mode (DTM). This feature introduced in GPRS allowed for a simultaneous voice and data communication, which could be perceived as bridging the gap between the GPRS packet-oriented approach and GSMs circuit switching [31].

To cover for the multiband evolutionary aspects one can also recall the example of first multiband GSM devices. GSM is typically operating in frequency bands of 900 MHz, 1800 MHz (typical European and worldwide band) and 1900 MHz (the frequency band used by the Northern American operators). Without the user device's ability to operate on all of those frequencies the user traveling outside its country's border could face the hardships of SIM roaming. This means that whenever the user travels to a country where their mobile device doesn't support the given frequency band they would have to transfer their SIM card to a new device which does so. Nowadays a dual band (e.g. 900 MHz and 1800 MHz) or triple band (900 MHz, 1800 MHz and 1900 MHz bands) functionality is a default configuration in practically every 2G handheld device.

Another important step in the evolution of terminal devices is so-called multimode. Multimode terminals term refer to terminals supporting different generations of cellular network evolution for example, GSM, WCDMA and LTE. It can also refer to terminals that additionally support a WLAN radio interfaces from the 802.11 standard family. In case of 3G and WLAN technologies one of the first mobiles that allowed these dual network operations (dual mode) was N900iL released by NTT DoCoMo in November 2004 [32]. Using this equipment, users could receive incoming calls via both networks in dual mode or by each network separately. Also features like instant messaging were supported in dual mode.

From the specification point of view the 3GPP standardization body has also put a lot of effort into providing the interworking between new cellular generations. UMTS standard is backward compatible with GSM/GPRS/EDGE and LTE is inherently compatible with both 2G and 3G legacy systems. Of course it doesn't mean that by having a 2G phone we will be able to connect to 3G network or experience LTE data rates. This backward compatibility is realized by means of measurements, roaming and handovers to the legacy system.

In case of measurements in the simplest example the user device is able to measure the pilot signal levels for supported network generation types. By roaming we refer to the ability of seamless transition between supported networks and to make or receive phone calls or other supported services without additional manual user intervention. Finally by handover we mean a transfer of an ongoing connection from one network to another. All of those features serve so-called the Always Best Connected (ABC) approach, which stands for connecting to the most suitable network by means of offered service and networks signal quality.

The look and capabilities of user terminals naturally also reflect the evolution of the telecommunication systems. The first mobile handheld was released by Motorola in 1983 and it was the famous DynaTAC 8000x with the size of $33 \times 4{,}5 \times 9$ cm, weighing almost a kilogram and with a price of almost $4000. In 1999 a Canadian company Research In Motion Ltd introduced a first push email solution called a BlackBerry which allowed instantaneous transmission, and what is more important for the instantaneous reception of sent emails, a first Always On, Always Connected solution. In 2007 Apple released the first iPhone, a very user friendly smartphone that marked the beginning of the domination era for smartphones and tablets in the handheld market. The introduction of the iPhone was a true revolution. Without risk, one can state that it has changed the way in which the people can interact with the Internet beyond using a wireless phone as a replacement for a landline. It has paved the way for other smartphones as well as tablet devices and become a true societal phenomenon.

Nowadays, tablets and smartphones are superseding standard personal computers. In the last quarter of 2012 the shipment of PCs visibly declined compared with the same period the previous year and the reason was the growing customer demand for devices like tablets and smartphones. In 2012 device makers shipped more than 700 million units of smartphones worldwide [33]. At the beginning of 2013 from the total global website traffic 8% was generated by tablets and 7% was generated by smartphones [34]. Despite the fact that smartphones were introduced earlier than tablets, the latter ones are preferred by users for longer surfing in the web. Additionally smartphones are preferred devices for online shopping and ecommerce.

There are several options regarding how cellular mobile devices may develop in future. One possible evolution option may be transparent phones, handhelds which are made out of completely translucent materials. Another game changing technology may be the usage of graphene. This specific, two-dimensional form of carbon is said to be strongest and lightest material known to the mankind and has multiple electrical and optical properties that make it interesting not only to the cellular end-users device manufacturers, but also to the whole electronic industry. It is also worth noticing that in 2013 the European Commission has chosen graphene as one of Europe's first 10-year, 1000 million euro Future Emerging Technology flagships [35]. Another step in evolution may be devices like Google glasses. This augmented reality equipment is capable of enhancing our everyday experience by proactively sensing our surrounding and displaying processed data on the screen embedded in the form of old-style glasses. Finally, wearable devices such as smart watches, may further impact the way we use to interact.

One of the consequences of the advancements on mobile devices market and increased need for personalized and customized equipment is the so-called Bring Your Own Device (BYOD) trend. The reasoning behind this particular tendency was the growing number of employees that actually bring their own mobile equipment (smartphones, tablets etc.) to the workplace and actively use it to facilitate their daily work. This tendency, which also poses some possible challenges to company's safety (by means of security breaches), may bring significant savings in terms of costs for the employees' equipment. It can also increase the efficiency of work as employees are using their own devices, tailored and customized to their needs.

Paradoxically, advances in IT infrastructure and faster data rates on the air interface may also, in some specific device application, lead to a simplification of end-user devices. Increasing potential of cloud technologies, which are foreseen as one of the leading and most promising solutions for future IT applications, may be used to move the computational aspects from the user device to the cloud, which is not constrained (or far less limited) by the power consumptions, that is a typical restriction for wireless mobile devices. This could result in user terminal equipment whose key task is collecting information, transmitting and receiving data from the network connected to cloud solutions and presenting postprocessed information in a user friendly form.

25.7 Evolution of Macro Sites and Deployment of Small Cells

Growing traffic consumption is driving a constant upgrade and evolution of telecommunication networks. One of the key questions which network operators are facing is whether they should focus on improvement of existing macro sites or if they should put their efforts into the deployment of small cells.

There are several arguments behind both alternatives. In favor of macro sites evolution is the fact that macro solutions are a proven and well-known option to provide large area capacities. The macro layer is accessible for most of users as antennas of macro station are usually placed above rooftop level and the total output power of macro stations is in the range of 10–40 watts compared with several watts or milliwatts output power of micro, pico or femto sites. Hence the upgrade of one site improves the performance of a large group of users, contrary to the micro or small cells solution that affects a limited group of users on a smaller area. Also, handling the large group of users by one large central node offers an easy way of resource management coordination. In case of a large number of nodes coordinating smaller groups of UEs each, the signaling overhead related with the coordination information exchange may become excessively large comparing to the useful data volume.

As an argument in favor of small cell solutions one can state that small access nodes are usually deployed in hotspot areas where density of users is larger comparing to the classical macro scenarios. They are also closer to the user, so the energy efficiency aspects can be observed for both user terminals and the infrastructure side. The argument against proliferation of the small cells are the backhaul challenges that operators can encounter while trying to deploy a large number of access nodes. In case of existing macro sites the backhaul infrastructure is already there. Deployment of many network nodes requires also more laborious planning and dimensioning efforts.

To compare the pros and cons of both approaches we can also compare possible upgrade paths using both approaches. Below several most popular ways for macro site improvement are described.

One of the most efficient methods to improve macro sites capabilities is so-called Higher Order Sectorization. Most common macro scenarios assume a three-sector configuration, that is one base station operating on three cells, each one controlled by directional antennas set. Those antennas can be placed in the same location for example, mast or a tower, but are pointed towards different direction. In this case Higher Order Sectorization usually means upgrade to six sector solution. Two ways to achieve this is either by further dividing the horizontal plane using a larger number of antennas/sectors or using the split in the vertical plane by means of solutions like Advanced Antenna Solutions (AAS). Such multisectorization can lead to coverage and capacity gains of:

- Up to 80% more capacity for 6×1 deployment (horizontal plane division) compared to 3×1
- Up to 65% more capacity for 3×2 deployment (vertical plane division) compared to 3×1
- Up to 40% increased coverage [36].

This optimization requires of course some additional investments due to the increased number of antenna elements and increased number of radio processing units in the base station, but it is much cheaper compared with macro site densification that could provide gains in a similar range.

Unfortunately site upgrades by means of Higher Order Sectorization cannot be used in all deployment scenarios. In areas where propagating radio signals can experience large angular spreads due to, for example, reflections from the buildings, Higher Order Sectorization configuration may potentially increase inter sector interference level, hence limiting the overall network performance. In case of 3×2 vertical sectorization, where the main beam of each sector is split into two with different vertical tilts, increased interference is also an issue. However such solutions are really beneficial if users in a sector tend to group, for example, close to the cell center or close to the cell edge.

Another very cost-effective yet simple solution for macro site performance upgrade is tilt optimization. The rationale behind it is that the initial deployments of radio networks usually aim at providing maximum coverage. Fine tuning of the antenna tilt can provide significant capacity enhancement, due to the lower inter site interferences level and thanks to providing higher SINRs in desired areas (by focusing the main beam of the antenna in the vertical plane in the particular place of focus). This feature is especially attractive to SON applications that aim at dynamic optimization of the network. Also solutions for remote electrical tilt steering already exist and are commercially available.

The other way of enhancing macro cell capacity is using the multicarrier features available in 3G and 4G standards. They are available in HSPA since Release 8 and in LTE since Release 10. Additionally the standard allows for multiple carriers in a noncontinuous bandwidth (since Release 9 in HSPA) which creates the possibility to refarm the spectrum of legacy standards like GSM, where penetration of legacy network devices is declining. This feature is of course also available for station types other than macro, but due to the fact that it requires a separate receiver/transmit radio chain and/or additional antennas, the macro layer upgrade require proportionally less investment. The upgrade to the multiple carrier solution in case of small cells usually requires replacement of the entire node hardware as the small cells nodes are usually compact and integrated devices. In the case of additional carriers, operator strategy regarding carrier assignment to macro, micro and pico layers also has to be taken into account to maintain the optimum level of interferences in the system.

On the downside, macro cells, despite their obvious commercial potential (historically the most popular form of cellular deployment with the largest coverage) also have several drawbacks. One can recall here problems with indoor coverage, high costs of redundancy and problems with providing high capacity over a limited area.

The first drawback (problem with indoor coverage) may be partially solved by operating on low frequencies in sub 1 GHz bands. Typically signals operating on low frequencies experience lower attenuation on building construction materials; however, the amount of spectrum available for operators in this frequency region may not be sufficient to provide the necessary broadband speeds. Additionally in case of larger buildings, accumulated attenuation on lower frequency bands can easily become a problem, especially in the UL direction as UE devices have much lower total output power (100 mW–250 mW) comparing to the typical macro stations (10–40 W). This problem is visible in case of the high-rise urban environments, where the coverage of macro layer may be insufficient to provide satisfying transfer speeds on the highest floors of buildings.

The second mentioned macro drawback, cost-related limitations, is the inherent problem of macro solutions. Base stations operating at high output power require rather costly analog components. As macro stations often have to receive very weak signals from distant UEs, their receiver chains' parts have to introduce minimal distortions to received transmission which further increase prices. Since macros provide coverage to a large group of users, their hardware and software components must be very reliable. To illustrate this problem we can exemplify, that damage caused by a small cell node breakdown that provides broadband to a single household, is far less of a problem compared with the outage of a macro base station that provides a service to hundreds or even thousands of cellular users. This is the reason why the equipment which is used for larger cells operations have to be either highly redundant, or characterized by a large Mean Time Between Failure (MTBF) factor, which of course boosts the price higher.

The last-mentioned drawback of the macro base station is being not suitable for providing capacity over limited area. This problem may be also partially solved in modern network. Operators may use highly directional antenna, optimize the tilt or use Higher Order Sectorization. There are also potential features like 3D MIMO that can help the operators to provide the capacity in more specific localizations. Those methods, however, requires additional man effort or hardware/software upgrades which of course increase the price of the macro equipment.

On the other side of the network evolution is the massive deployment of small cells. Small cells are known since 2G, but they started to gain momentum in HSPA and 4G network rollouts. In 2012 the small cell for residential usage (femto cell) market totaled $425 million [37]. As from the beginning of 2013 the installed base of small cells was almost 11 million units and the predictions are that this number will increase eight times, up to the value of 92 million in 2016 as compared to 2013 [38]. Small cells initially installed in customers' households are also foreseen to be deployed in large volumes in places like shopping centers, enterprises or bus stops. Due to their small dimensions they can easily blur into the environment.

This evolution path also has a very strong reasoning behind it, especially when it comes to the provision of indoor capacity or in case of high nonhomogenous user distribution in a cell. Smaller cells are a straightforward way to deliver higher data rates and they are usually deployed much faster comparing to their macro counterparts. They are characterized by low costs, reduced power, size and environmental footprint. However, their deployment, without any self-optimization and self-configuration solution usually needs to be performed with respect to the existing macro network in order to obtain optimal network performance.

Within small cells we could distinguish three groups of products:

- Micro and pico cells: medium and small size cells with the radius ranging from up to hundreds of meters (micro) down to tens of meters (pico) and with power not larger than one watt. What distinguish them from the second group is that their deployment is coordinated and triggered by operators.
- Femto cells: smallest cells of all cell types, foreseen for the residential usage. They look like a typical modem or router and they connect to the core network over a public Internet backhaul. In LTE femtos are called Home Evolved NodeB (HeNB). Their deployment is customer driven and not coordinated by the mobile operator. Femtos are often compared to Wi-Fis and their commercial large scale availability dictates very cheap prices. Their main advantage to the mobile operators must be considered not in creating additional capacity, but rather as an easy way to offload the existing network due to closer proximity to the end-user. An exemplary LTE macro layer would need to spend much more efforts in terms of physical

resource blocks in order to provide similar performance as femto cells deployed close to the user. Femtos can also operate in so-called Closed Subscriber Group (CSG) mode which allows the owner of the Femto to restrict the number of devices that can connect to that node.

- Access points based on IEEE Wi-Fi standards, for example, 802.11g, 802.11n or 802.11ac are usually also considered as a small cell products and play a very important part in offloading regular cellular traffic in many use cases. The detailed description of wireless IEEE standards is beyond the scope of this chapter.

An interesting option to provide an indoor coverage in case of low capacity scenarios is Distributed Antenna Systems (DAS). This solution doesn't fit exactly the macro/small cell division as it is based on splitting the single antenna signal among several other antennas with respectively lower power. In typical applications antennas are distributed in several locations and the end result is solving one of the biggest drawbacks of macro cells solutions, inability to focus resources in limited and distributed hot zone areas. DAS help to overcome this situation by moving antennas closer to users, hence providing higher chances of Line of Sight transmission and limiting propagation losses. As the DAS cells share the resources of the one cell which are distributed over several areas of interest, they become inefficient when larger capacity is required.

Small cells are usually deployed in scenarios where macro layer is already available. They cover blind spots and provide extra capacity enhancements to the existing topology creating, together with the macro layer, a so-called Heterogeneous Network (HetNet) consisting of macro and smaller cells (please note that the HetNet term is also sometimes used to indicate the coexistence of different wireless network technologies like EDGE, HSPA and LTE or to point out that the existing network is built with network components originating from different vendors). This dense heterogeneous network topology, however, can lead to increased intercell interferences level, especially if small cells are operated on the same carrier frequencies as macros. One simple problem to imagine is the situation where due to the power imbalance of macro and small station, user device is connected to the macro based on its downlink pilot signal power strength, but it is much closer to the small cell. During ongoing uplink transmission terminal assigned to a macro station can create strong interferences in the small cell access node because the power of received signal will be inversely proportional to the distance from the terminal. Another example of the problems that can emerge in the real networks is the case of the already mentioned CSG option related to femto cells solutions. This leads to a situation where the user may be close to the femto access point, but is not allowed to connect to it and eventually it may become a source of heavy interference.

Those problems have driven several upgrades to LTE standard like enhanced Intercell Interference Coordination (eICIC). This particular feature is a frequency domain power coordination between network nodes, for example, between high power macro and low power pico. It allows blanking of some subframes of macro layer to reduce the interference level in the small cell layer. Another method which is common in HetNet deployment is Cell Range Expansion (CRE). It applies a certain bias to the cells' pilot signal measurements, which manifest in favoring connection to the desired cells, for example, to move some users from macro cells to picos.

Aside from interference challenges, another aspect that needs to be taken into account while deploying small cells are the legal aspects, for example, when considering the total transmitted output power of a small cell device or in case of law violations when the small cell facilitates law infringement. This, however, is a region-dependent problem and beyond the scope of this book.

The final decisions whether operators are going for the macro upgrade or small cells enhancements are usually scenario-dependent. In case of uniform user distribution and low to moderate overall traffic volume macro enhancements seems to be a more suitable path than network densification using small cells. On the other hand, with high indoor traffic, small cells and Wi-Fi solutions seem to be the most reasonable way to provide high network capacity. The general telecom industry trend is rather gearing toward the latter option. This is why in the future further network densification is very likely to be expected. Recent network solutions proposals seem to be the balance of both of those ways. The macro layer deployed at lower frequencies (e.g. in 800 MHz band) seems to provide umbrella coverage for mobility users while smaller cells at higher frequencies (e.g. 2600 MHz) provide a capacity extension for hotspots and indoor traffic. Definite enhancement to the HetNet are so-called SON features that support the deployment of small cells by automatization of optimization and configuration of new sites, thus limiting capital investments of operators.

The future may also bring a hybrid solution of both macro and small cells. Concepts like Virtual or Phantom Cells are based on the assumption that signaling traffic is carried by the larger cells while user data is handled by small cell layer. In the end the trend nowadays is that cell sizes are becoming smaller and smaller so the hybrid option that decouples the control plane from the data plane gives interesting scalability potential and is definitely an interesting way forward for future cellular systems.

25.8 Beyond 4G Systems: 5G

As a rule of thumb, it can be assumed that since the 1980s every decade brought a new wireless telecommunication breakthrough or new global wireless communication system. The 1980s were marked with the emergence of the first generation of analog systems, a decade later 2G systems appeared with the dominance of GSM technologies, then in the beginning of the new millennium 3G solutions with WCDMA and HSPA technologies became a leader in the group, and finally in the second decade of the new millennium it seems that LTE and 4G LTE-Advanced will take over the position of their 3GPP predecessors as the global leader in the wireless cellular solutions market. According to this pattern it is natural and logical to predict the arrival of a new telecommunications standard around the year 2020, which researchers refer to as 5G systems.

At the moment of this book creation it is impossible to give any detailed information about the shape of this new technology, but there are many predictions about what challenges it will have to cope with. The core challenge of the new system will be for sure handling of massive traffic increase which many engineers refer to as so-called 1000x challenge or 1000x traffic avalanche. 1000x indicates a massive, thousand-fold traffic growth as compared with the year 2010 which is expect to happen in the first half of the 2020s. To reach this kind of a growth a universal solution in the form of the 10x 10x 10x equation is often quoted. This is usually translated as:

- 10 times more base stations
- 10 times more spectrum
- 10 times higher spectral efficiency.

However, it should be treated rather figuratively as an indicator of where exactly we can expect system improvements, that is, network densification, more spectrum for cellular communication and its better utilization. Network densification, as explained in the previous section, is an already ongoing process on top of which one can add possible novel approaches like decoupling of control and user plane, pooling of the baseband resources, Network Function Virtualization or Software Defined Networks (those last methods are further described in the next section). More spectrum for cellular application is also expected from the global (Word Radio Conference) and local (National Regulatory Authorities) level. New spectrum access methods should also address the financial challenges for obtaining new frequencies.

A very interesting option in this context is also the utilization of millimeter waves region (30–300 GHz) for wireless access. As for the efficiency of spectrum utilization the possible methods to achieve this are massive antenna systems, advanced coordinated multipoint transmission schemes and interference management systems. One cannot also exclude the utilization of a variety of context information to improve efficient network resource allocation. By context information we mean here different data that facilitate a better description of past, current or future situation of network devices or network nodes and their mutual interaction.

Contrary to previous breakpoint moments, there is no clear 5G driver like OFDM for LTE standard. More, 4G solutions that were deployed and are operating are already quite mature and well performing technology which means that except future disruptive solutions, all above-mentioned methods to cope with the increased traffic volume are very likely to be incorporated into LTE standard or be a part of LTE technology. Despite that in 2013 there were several projects targeting at 2020 cellular solutions. One can mention here Future Forum, 5GNOW and even first ITU activities toward IMT-2020. To show the challenges that the 2020 systems will have to cope with we can quote the challenges of one of the biggest project in this area, METIS. Those technical challenges are:

- 1000 times higher mobile data volume per area
- 10 times to 100 times higher number of connected devices
- 10 times to 100 times higher typical user data rate
- 10 times longer battery life for low power massive machine communication
- 5 times reduced end-to-end latency.

What is also important is that the project itself is aiming at providing the potential enablers for those objectives at a similar cost and energy consumption as today's network. The future wireless cellular system should also be able to efficiently support Machine Type Communication. This may facilitate the advent of so-called Internet of Things (IoT), a concept of wireless connection between variety of devices, even the ones that we use in everyday life and traditionally don't associate with any "communication functionality" like fridge or temperature sensors. The future 5G system may lay the foundation or become a big part for this kind of solutions. An attractive use case is also a mission critical information exchange in Vehicle-to-Vehicle (V2V) or Vehicle-to-Anything (V2X) communication. Here, 5G solutions could act as enabler for ultra reliable and ultra fast information exchange for traffic safety purposes.

25.9 Challenges and Possibilities

The mobile wireless communications industry is characterized by a constant change. End-users are constantly upgrading their access devices because they utilize more and more advanced and resource consuming features that require higher and higher transfer speeds. Mobile network operators have to constantly upgrade their network in order to provide the necessary capacity and service offer that users demand. Also vendors need to constantly develop their product offer in order to survive in a highly competitive market. The Standard Development Organizations (SDO) like 3GPP are constantly releasing new revisions of wireless standards and new features often become a field of continuous debate for the favorable solutions and Intelectual Property Rights (IPRs).

For the last two years the mobile industry has also been facing an interesting phenomenon where peak data rates are not the ultimate goal of the race. At the beginning of the new millennium network operators tried to convince end-users to utilize wireless connections in a similar manner as they utilized their fixed broadband access or Wi-Fi solutions. Peak data rates are the easiest means to convince consumers that cellular mobile systems are capable of providing similar user experience. Since the advent of HSPA, USB dongles with mobile wireless connectivity and the flat rate have become a commodity followed by the traffic avalanche that has forced the change of mindset.

As mentioned already in the previous section in 2020 rough estimates predict that operators will have to accommodate 100–1000 times larger traffic volume compared with ten years ago. Now the goal is not only about speed, but also about providing large capacity for wireless services by the cheapest means possible and relatively fairly across the cell area. The latter option is especially important from the business point of view since many users, tempted by the advertisement showing peak data rates that are usually only available in very favorable channel conditions in the proximity of the access note, can be easily disappointed if located close to the cell edge, where interference limitations or high signal attenuation result in the end speed far lower than the advertised ones. Besides that, future users are very likely to expect the same kind of user experience everywhere and anytime. This is why many operators consider small cells solutions, where the user is located closer to the base station, as the necessary panacea to provide the promised service. Also features like advanced CoMP, massive MIMO, Advanced Antenna Solutions or virtualization of the network functionalities can help here a lot.

As for the problem of continuous price erosion, the ball is in both the operators' and vendors' courts. The operators have several options to limit their expenses. They can decide to go for the Femto solutions to limit the costs of the backhaul. They can also try to implement different options related to network sharing. Depending on the agreement between operators it may be simple site sharing, which means using the same mast, tower or equipment room for BTS/NodeB colocation. In this particular case all common elements are nonactive, hence another term used for this kind of practice is known as passive sharing. Since site costs constitute ~30% of 3G rollout's CAPEX and OPEX, this is definitely an interesting opportunity for cost reduction.

Having this solution already in place it is relatively easy to go one step further and expand the pool of shared elements, adding there the backhaul network. If the network elements are in the common place then transmission cables and the backhauling infrastructure may become common as well. Another step forward is achieved by so-called Active RAN Sharing, where one or more network elements like base station or RNC is used by different operators. In this case operators are differentiated by the dedicated frequencies on the carrier level. They still have their own PLMN-id's, cell or site level parameters and can offer different services for their customer. Such solution can lead to a reduction in equipment volume in low traffic areas, increased rollout speeds and reduced network and site operating costs. With an even higher level of cooperation, where also the core network elements are shared, for example, MSC, VLR or SGSN, it is possible to reach Roaming Based Sharing, where the customers of one operator roam seamlessly in the 'host' operator's network, plugging gaps in the network coverage of a single provider.

At this point it is also worth mentioning Mobile Virtual Network Operators (MVNOs). MVNO is an operator that doesn't own exclusively any part of the spectrum and its network infrastructure is limited to minimum. This kind of operator has a business agreement with one of the traditional mobile network operators and it uses its network as defined in a business contract in order to provide some differentiating services. In many cases MVNOs have their own intelligent network (IN) infrastructure for value added services and most of those operators limit their services to the low cost prepaid offering. However, other options are also appearing, for example, low cost international phone calls or other new opportunities may also emerge in the future.

Another wireless network trend which aims at limiting capital investments and expenditures is Software Defined Networking (SDN) and Network Functions Virtualization (NFV). Those functionalities intend to enable the capability of supporting different present and future services irrespective of the hardware platform they operate on, decouple control and user plane and for the utilization of the network functionality and services without direct access to network level interface. The related term is so-called a Software Defined Radio (SDR), the ability of hardware to operate on multiple air interfaces using the same hardware platform. To give a practical example, one can imagine a situation where a certain operator wants to switch one of its NodeBs from HSPA to LTE operations having different carrier operating frequencies for both standards. With the SDR concept already in place, neglecting the network management issues, a simple software operation would be needed to make this transition, without any hardware changes. For the operator it means a shorter time to market for the new products and services, but also preservation of hardware investments and reduction of operating costs. The challenges of SDR are definitely the future interfaces and their possible hardware realizations that today's engineer needs to foresee during the development of SDR platforms. Also the cost of a one-time payment for SDR hardware can be higher compared with a single system oriented components due to the additional support of multiple air interface requirements, filters, modulators and so on.

Further extension of SDR concept is Cognitive Radio (CR) where not only is it possible to adjust to a certain air interface using software, but additionally user device is able to interact with the air interfaces and change the transceiver parameters to adapt to this environment. Operating frequency may be changed along with the power setting and other transmission parameters in order to improve spectrum utilization. CR, regardless of the user experience improvements, may also provide an access to emergency services in case of natural disasters like earthquakes or tsunamis.

One of the latest trends in radio network technology is centralized RAN (C-RAN), which utilizes so-called baseband pooling. In today's radio access network most of the base stations follows a pattern that baseband units are colocated with radio frequency units. However there is a growing number of network equipment vendors that offer pooling of baseband units in "hotel" locations that are connected with remote radio frequency units using fast connections with low latency (which is usually based on fiber optics). Such configuration may offer significant savings to operators by OPEX reduction as the baseband parts can be kept and maintained in one central location. Additionally it creates a potential for simplified introduction of advanced coordination techniques like enhanced ICIC, CoMP or load balancing.

References

1. AT&T. [Online]. Available: http://www.corp.att.com/attlabs (last accessed 24 August 2014)
2. ITU. [Online]. Available: http://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx (last accessed 4 August 2014).
3. Das, S.K. (2010) *Mobile Handset Design*, John Wiley & Sons, Inc.
4. 3GPP TS 41.101 v11.0.0 *Technical Specifications and Technical Reports for a GERAN-based 3GPP system (Release 11)*, 2013
5. Nokia Siemens Networks, *The broadband world is flat: Profitable mobile broadband with Internet HSPA*, 2007
6. ITU-R, *Invitation for submission of proposals for candidate radiointerface technologies fot the terrestrial componenets of the radio interface(s) for IMT-Advanced and invitation to participate in their subsequent evaluation*, 2008
7. ITU-R, *M.2133, Requirements, evaluation criteria and submissions templates for the development of IMT-Advances*, 2008
8. ITU-R, *M.2134, Requirement related to technical performance for IMT-Advanced radio interface(s),* 2008.
9. ITU-R, *M.2135, Guidelines for evaluation of radio interface technologies for IMT-Advanced*, 2009
10. Ranta-aho, K. and Maternia, M. (2012) *On Performance of Long Term HSPA Evolution*, *ICC*.
11. OpenSignal, *The State of LTE*, 2013
12. Toscala, A. and Holma, H. (2012) *LTE Advanced: 3GPP Solution for IMT-Advanced*, John Wiley & Sons, Inc.
13. Cisco, *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2012-2017*, 2013
14. rcrwireless, 08 March 2013. [Online]. Available: http://www.rcrwireless.com/article/20130308/spectrum_auction/canada-auction-700-mhz-spectrum-nov-19/ (last accessed 4 August 2014).
15. Ofcom, 20 February 2013. [Online]. Available: http://media.ofcom.org.uk/2013/02/20/ofcom-announces-winners-of-the-4g-mobile-auction/ (last accessed 4 August 2014).
16. Infonetisc Research, "www.infonetics.com," 13 March 2013. [Online]. Available: http://www.infonetics.com/pr/2013/4Q12-Mobile-Infrastructure-Market-Highlights.asp (last accessed 4 August 2014).
17. Global TDD Initiative*, TD-LTE Industry Briefing, August,* 2013
18. Shannon, C.E. (1998) Communication in the presence of noise. *Proceedings of the IEEE*, **86**(2)
19. Maternia, M., Gorzynski, J., and Lachowski, J. (2011) *Multiband and multicarrier operations in HSDPA*, ISWCS.
20. 3GPP TS 36.101 v11.5.0 *User Equipment (UE) radio transmission and reception (Release 11),* 2013
21. ITU, *Spectrum Management for a Converging World: Case Study on the United Kingdom*, [Online]. Available: http://www.itu.int/osg/spu/ni/spectrum/UK-RSM.pdf (last accessed 4 August 2014).

22. Cellular-news, 27 October 2011. [Online]. Available: http://www.cellular-news.com/story/51544.php (last accessed 4 August 2014).
23. GSMAIntelligence, 2 June 2011. [Online]. Available: https://gsmaintelligence.com (last accessed 24 August 2014).
24. FCC, [Online]. Available: http://www.fcc.gov/topic/700-mhz (last accessed 4 August 2014).
25. Ministry of Internal Affairs and Communications, January 2013. [Online]. Available: http://www.tele.soumu.go.jp/e/adm/freq/search/share/plan.htm (last accessed 4 August 2014).
26. Ofcom, *Frequency Band Review for Fixed Wireless Service*, 2011
27. European Commission Committee, *Light Licensing, Licence-Exempt and Commons*, 2013
28. Nokia Siemens Networks, Optimised *Spectrum Use by Authorised Shared Access*, 2012.
29. Blackman, C., Forge, S., and Horovitz, G. (2012) *Final Report for the European Commission: Perspectives on the value of shared spectrum access*.
30. European Commission, *Promoting the shared use of radio spectrum resources in the internal market*, 2012
31. 3GPP TS 43.055 v11.1.0, *Technical Specification Group GSM/EDGE Radio Access Networkl Dual Transfer Mode (DTM); Stage 2 (Release 11)*, 2013.
32. NTT DoCoMo, 15 November 2004. [Online]. Available: http://www.nttdocomo.co.jp/english/info/media_center/pr/2004/001219.html (last accessed 4 August 2014).
33. rcrwireless, 25 January 2013. [Online]. Available: http://www.rcrwireless.com/article/20130125/uncategorized/more-700-million-smartphones-shipped-2012-even-growth-slows/ (last accessed 4 August 2014).
34. Adobe Digital Index, 6 March 2013. [Online]. Available: http://blogs.adobe.com/digitalmarketing/digital-index/tablets-trump-smartphones-in-global-website-traffic/ (last accessed 4 August 2014).
35. Graphine Flagship, 28 January 2013. [Online]. Available: http://www.graphene-flagship.eu (last accessed 24 August 2014)
36. Nokia Siemens Networks, *Deployment Strategies for Heterogeneous Networks*, 2012
37. Infonetics Research, 7 March 2013. [Online]. Available: http://www.infonetics.com/pr/2013/4Q12-Femtocell-Equipment-Market-Highlights.asp (last accessed 4 August 2014).
38. Small Cell Forum, 26 February 2013. [Online]. Available: http://www.smallcellforum.org/newsstory-public-access-small-cell-market-to-hit-us-16-billion-in-2016 (last accessed 4 August 2014).

26

Wireless Network Measurements

Jyrki T. J. Penttinen

26.1 Introduction

The overall physical radio measurements largely share the same principles over cellular and wireless system generations. The same principles can thus be utilized as a base for GSM, UMTS, HSPA+ as well as for LTE measurements, and also for the other systems like CDMA. Nevertheless, each system has their own special characteristics, and they have evolved further from the previous generations. The more advanced performance and respective signal processing also require new abilities from measurement equipment.

26.2 Principles of Radio Interface Measurements

Radio interface measurements are essential in the deployment and operational phases of cellular networks.

The physical radio interface can be investigated by monitoring and storing numerical values about the received power level over the functional bandwidth. In the deeper measurement principles, the modulation scheme and other characteristics need to be observed case-by-case.

26.3 GSM/GPRS

26.3.1 GSM/GPRS Measurement Devices

The GSM/GPRS equipment can typically be used to record and monitor radio network occurrences and parameter values. The equipment can consist of a data storage unit such as a laptop PC, connected to a commercially available terminal. The packet data connection is established, for example, via the public Internet to the server, and data transfer is monitored in both directions. A separate data unit accessed via GPRS can potentially be connected to the Internet network. The closer the data unit is to the GGSN element

The Telecommunications Handbook: Engineering Guidelines for Fixed, Mobile and Satellite Systems, First Edition.
Edited by Jyrki T. J. Penttinen.

of the studied network, the more reliable are the achieved measurement results. This is because the potential bottleneck-effect of the Internet network should be eliminated. Thus, data can be transferred from the data unit to the measurement device in a controlled fashion to measure issues such as connection time to the GPRS network and the Internet, the data transfer rate, and variance in channel coding classes and multislot usage. The measurement equipment can also contain satellite positioning equipment such as the GPS (global positioning system), which allows studying the measurement data spatially.

26.3.2 The C/I Measured from the Network

The performance of the GPRS and other data services can be estimated for example, as the actual throughput of the data as a function of *C/I* (signal/interference) and *Eb/N0* (energy–noise ratio). The results of theoretical simulations can be calibrated to better represent reality by estimating the C/I- and Eb/No-values of actual networks.

The C/I-relationship can be investigated from the network or a part of the network, which is clearly interference limited. A dense urban network is most probably constrained by interference. By measuring the interference distribution of the network (Q-classes) and by using simulation results from, for example, a connection can be established between the Q and the *C/I*.

It makes the most sense to study the *Eb/N0*-values in a noise limited network, such as a sparse rural network. The *Eb/N0*-values are studied by measuring the field strength levels of the network.

26.3.2.1 RxLevStatistics

Source [1] describes studies done on both rural and urban networks using the "Rx Level Statistics"-tool (found in the operations and maintenance center of the GSM network) in order to obtain field strength and interference levels. The tool can be used to collect the quality level (Q-level) values 0–7 of the traffic channels in a matrix format, and also, the received power level values divided into six regions. This measurement used Rx Level region values 10, 15, 20, 30, 40, and 63, which correspond to values −110 – −100, −99 – −95, −94 – −90, −89 – −80, −79 – −70, and −69 – −47 dBm. The values were collected into a file with 480 ms intervals.

Ten BCCH-TRXs and 8–10 "secondary TRXs" (i.e. TRXs, with no BCCH time slot) from both interference-constrained urban environments and coverage area-constrained rural networks were chosen for the measurements. In both cases, the RxLevel- and Q- values of the traffic channels were monitored for one hour during a peak hour (3–4 p.m.) and at an off-peak hour (6–7 p.m.). All of the studied cells, urban and rural, used uplink power control and uplink-DTX. None of the cells used frequency hopping.

One downside of the Rx Level Statistics-tool is its crude field power scale. When the GSM network measures the power level with a 64-scale, the Rx Level Statistics-tool saves the values in a table with only a 6-scale. The scale can be "expanded" to a 64-scale with a relatively low error margin (see Figure 26.1) by assuming an even spread of the values within each field power region. The examples chosen for the figure are averages of the downlink and uplink TRXs. The assumption of an even spread is naturally not valid in reality, but the field power distribution cannot be studied more accurately with a single round of measurements.

It can be assumed that the bit error relationship, uncovered with the help of the Q-parameter, has the most significant influence on an interference-constrained network. The Q-values can only be saved using a scale with 8 values. In a well-designed network, most of the values collect in Q-class 0.

It can be seen from Figures 26.1 and 26.2 that the urban networks give fewer of the weakest values than do the rural networks. This is due to the abundant coverage overlaps of the urban cells. A low field power level – one which is close to the sensitivity level of the terminal – naturally increases the bit error rate, which means that the principle constraint in rural areas is the signal–noise relationship. The correlation of the field power

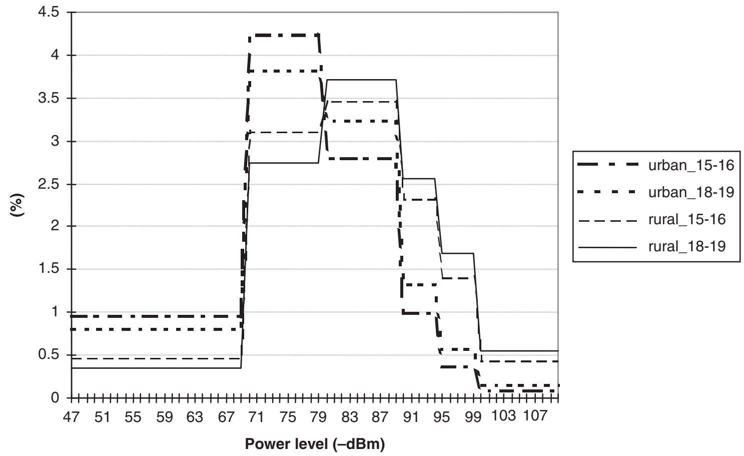

Figure 26.1 *An example of the received field power distribution in urban and rural environments. Figure by courtesy of Nemo Technologies.*

level and the *Eb/No* is thus evident in rural areas. The cellular network is denser in cities. The planned street level value in cities can be, for example, –75 dBm, compared with values of even –85 dBm in rural areas. In the urban areas, it is mostly the *C/I* which correlates with the field power level.

Source [1] describes simulations done on the correlation between the GSM quality level Q and the *C/I*. According to the results, the correlation between the Q-value and the *C/I* is: Q0 = 22 dB, Q1 = 20 dB,

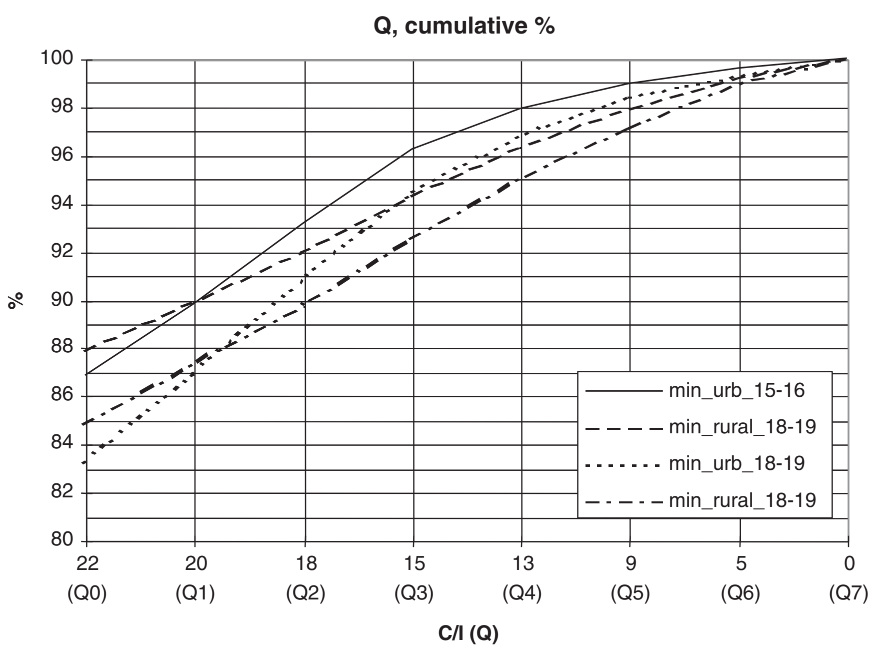

Figure 26.2 *The cumulative shares of the minimum quality levels achieved in both urban and rural environments based on the RxLevelStatistics-measurements. Note: the C/I scale is nonlinear. Figure by courtesy of Nemo Technologies.*

Q2 = 18 dB, Q3 = 15 dB, Q4 = 13 dB, Q5 = 9 dB, Q6 = 5 dB, and Q7 = 0 dB. The TU3 model with ideal frequency hopping was used in the simulations. The cumulative %-shares of the Q-classes corresponding to the *C/I*-values (in other words: the percentage of the studied time period when the service level in question or a better quality level has been achieved) are shown in Figure 26.2. The correlation between the *C/I* and the Q-class can probably be represented as a function as shown, but the function must take into account the inaccuracy factor, which depends on for example, the environment.

In the urban and rural environments shown in Figure 26.2, a minimum value from the uplink and downlink TRXs has been chosen to represent the worst case quality level. This is the minimum value achieved in each environment. The *C/I* value in the figure also applies when frequency hopping is used. As the studied cells did not utilize frequency hopping at the time of the measurements, it can be assumed that actual *C/I*-values are a few dB weaker. This means, that for example, the C/I requirement value corresponding to Q-value 4 changes from 13 dB to 15–16 dB.

Some shift in timing between the rural and urban environments is typically evident. However, it can be stated, that the shift is relatively meaningless.

26.3.2.2 Field Measurements

Operations System statistics, like RxLevStatistics measurement, can be used to study the average quality level of the network in a relatively reliable statistical manner. However, it makes more sense to use field equipment to monitor the quality level of a single call. Such field equipment can be formed with a test phone and a data storage unit connected to it. The equipment can be used to set up and break up calls at desired intervals, resulting in, for example, success rates for calls and channel switches, the duration and area of the service, received power levels (RxLev) and quality level (Q) downlink. Equipment placed in a car was used in simulations described in Ref. [1]. During the measurements, the phone was attached to the car's dashboard. The route consisted of the area of the studied cells. The RxLev-values used in the measurements were chosen using a method somewhat different from the measurements described in the previous chapter in order to get comparable results from the RxLev and field equipment measurements. The results can be seen in Table 26.1.

For comparison purposes, a series of experiments was made where rural and urban areas were studied with both field equipment and the RxLevStatistics equipment. The experiments were made on a working day with normal traffic before peak hours. Both experiments were made at the same time and in the same cells. As the field equipment can only measure downlink, the same direction was studied with the RxLevStatistics

Table 26.1 *A comparison of field equipment and RxLevelStatistics measurement results (received power levels). The results are shown as percentages. The field equipment results have been achieved using a single GSM phone as the measuring device. RxLevelStatistics results consist in the measurements of all calls in the studied area*

	Urban		Rural	
RxLev/dBm	Field meas.	RxLevStatistics	Field meas.	RxLevStatistics
−60	27	10	9	5
−60 – −80	65	63	56	45
−81 – −93	8	24	30	39
−94 – −98	0	2	3	7
−99 – −104	0	1	1	3
<−104	0	0	1	1

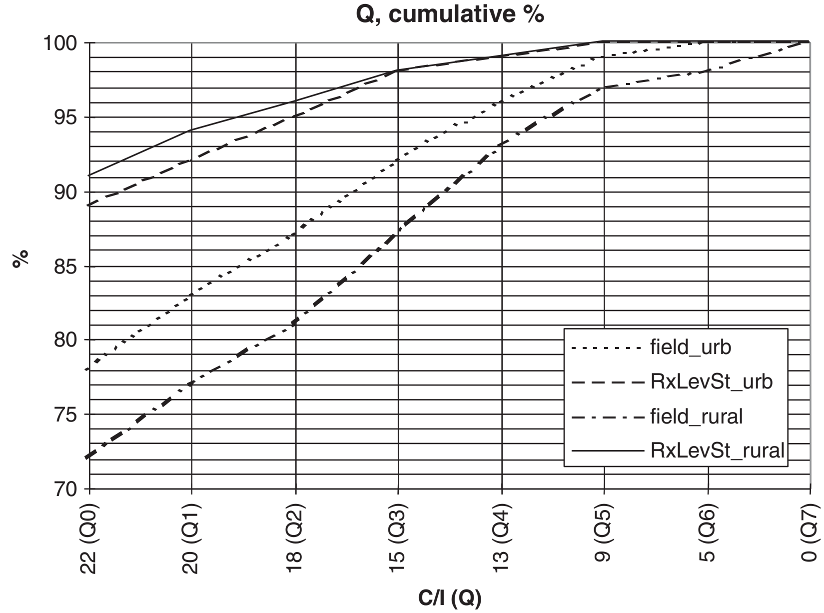

Figure 26.3 *A comparison of the results on studies of the cumulative, %-age share of time of Q achieved using the field equipment and the RxLevStatistics. Note: the C/I-scale is nonlinear.*

equipment. Also the same RxLev limit values were used in both experiments. The cells did not utilize frequency hopping at the time of the experiments, which means that a weakening of approximately 2–3 dB can be assumed.

The results of the RxLevelStatistics, as seen in Figure 26.3, are clearly better than the results of the field experiments. The RxLevelStatistics measurement is statistically extremely reliable, as it continuously measures the quality levels of all active connections.

Field experiments give a pessimistic view of the network quality, as the phone is located on the car's dashboard, causing a very unequal reception radiation pattern. In addition, the measurements are done while the car is constantly moving, and the outer areas of the cells are visited on purpose. The field experiments thus describe the worst case scenario, as cells have, on average, been placed in areas with high traffic. In contrast, RxLevelStatistics can be used to find out, if the cells of the network have been constructed in areas where the majority of the traffic is.

The conclusion that can be seen in Table 26.2 gives an idea of the actual QoS level of the network, which GPRS users are likely to experience in different kinds of environments. Data equipment located in the car is not an impossible idea, even if the use of current applications in cars is not popular. However, applications for positioning cars may emerge with the further development of packet data services.

Let's assume that the correlation between the *C/I* and the Q described in [1] apply so that an additional margin of 2–3 dB must be added to the values, if frequency hopping is not used. This means that the practical situation can be evaluated with the help of conclusions of theoretical considerations. Based on Figure 26.3, we can choose a static (corresponding to the results of the RxLevStatistics experiments) and mobile (corresponding to the results of the field equipment) data profile in both urban and rural environments. The functioning areas of the GPRS-classes can be "calibrated" to the real network, resulting in the actual data rate experienced by the GPRS users.

Table 26.2 *The percentage share of time of each GPRS channel coding class in data connections, based on the quality level distribution of the GSM network. The table also contains the time, when the C/I level of no class is achieved*

Channel coding scheme			*C/I* requirements (dB)	*C/I* is achieved (% of the time)			
				Static MS (RxLevStatistics)		Moving MS (field measurement)	
				Urban	Rural	Urban	Rural
Outage time	FH	[2]	<7.1	0	0	0.5	2.5
		[3]	<9	0	0	3.0	1.0
	noFH	[2]	<10.8	0.5	0.5	2.4	4.7
		[3]	<13	1.0	1.0	4.1	7.0
CS-1	FH	[2]	7.1	100	100	99.5	97.5
		[3]	9	100	100	97.0	99.0
	noFH	[2]	10.8	99.5	99.5	97.6	95.3
		[3]	13	99.0	99.0	95.9	93.0
CS-2	FH	[2]	11.5	99.3	99.3	97.0	94.4
		[3]	13	99.0	99.0	95.9	93.0
	noFH	[2]	12.8	99.1	99.1	96.0	93.2
		[3]	15	98.0	98.0	92.0	87.2
CS-3	FH	[2]	13.6	98.7	98.7	94.6	91.0
		[3]	15	98.0	98.0	92.0	87.2
	noFH	[2]	13.7	98.6	98.6	94.5	90.8
		[3]	16	97.0	97.3	90.3	85.0
CS-4	FH	[2]	20.8	90.6	92.7	80.9	74.8
		[3]	23	87.5[a]	89.5[a]	76.0[a]	70.0[a]
	noFH	[2]	17.2	95.7	96.5	88.4	82.8
		[3]	19	93.5	95.0	85.0	79.1

[a] Value by extrapolation.

It makes sense to study the *C/I*- requirement level using a 10% BLER-limit value, as this is the value defined in the specifications. For comparison purposes, the *C/I*-values of the specifications and the *C/I*- values (10% BLER-value) were achieved as results of the simulations described in [1].

According to the RxLevelStatistics experiments, field power levels required by CS3-4 are achieved more than 93.5% of the time in both urban and rural areas, when no frequency hopping is used. A mobile GPRS user is within the influence area of CS3-4 85.0% of the time in urban environments, and 79.1% of the time in rural environments.

The share of time is not directly proportional to the surface area. The above-described calculations based on real-life situations do, however, show that the measured areas of the network have been constructed taking into account the high traffic areas. As the GPRS service will likely be used in the same places as the existing voice and data services are used, it seems that the measured areas are well suited for GPRS users.

Based on the results seen in Figure 26.4, it would seem that CS-4 – and thus, the fastest data rate – could be used almost constantly in the measured area of the network, if only it were possible to use CS-4 in actual networks. On the other hand, if it is only possible to use CS-1 and CS-2 in real-life networks, it would seem that CS-2 is used almost constantly in the studied area.

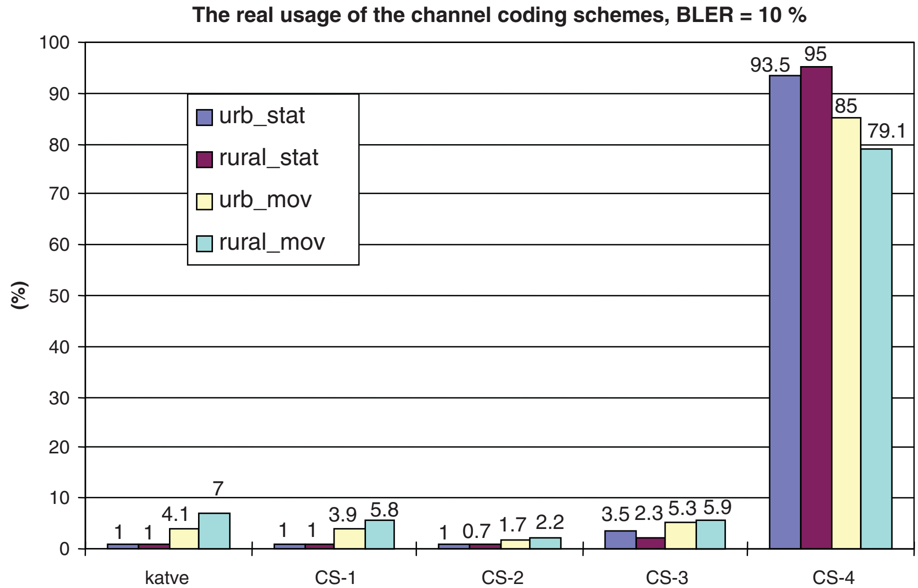

Figure 26.4 *The share of time of each channel coding class in an actual network, when BLER equals 10%. The values are compared to the C/I-values set by the specifications, without frequency hopping. The outage is the period of time, when the C/I-value of no channel coding class is achieved.*

As the default BLER-value in the figure is 10%, some resending takes place, lowering the actual throughput somewhat. On the other hand, the studied network did not carry the extra load – and additional interference – of GPRS, which in reality weaken the results, depending on the number of optimization methods intended for voice.

26.4 LTE

26.4.1 Principle

LTE radio interface is based on OFDM (Orthogonal Frequency Division Multiplexing) in downlink, and on the SC-FDMA (Single-Carrier, Frequency Division Multiple Access) in uplink. This solution provides a scalable bandwidth selection between 1.4 and 20 MHz. In addition, LTE supports both FDD (Frequency Division Duplex) and TDD (Time Division Duplex) modes.

Figure 26.5 shows the principle of the LTE downlink transmission in the time domain. Each resource element corresponds to a single OFDM subcarrier over a period of one OFDM symbol. The channel separation of downlink is defined to 15 kHz in the normal mode of the transmission. In addition, the broadcast mode of LTE can also utilize 7.5 kHz raster size. In uplink, LTE utilizes SC-FDMA with scalable band.

Because LTE functions within a maximum of 20 MHz bandwidth – which is 4 times larger than previously has been the case in the UMTS system – it creates special requirements for measurement equipment. The HW and SW of the measurement equipment should be able to interpret all the received radio parameters and their values, as well as the performance figures of the measured connection. The latter includes, for example, data transmission speed (Mb/s), latency of data transmission (s), jitter, that is, the variations of the latency, bit error rate (%), modulation error rate (%) and frame error rate (%).

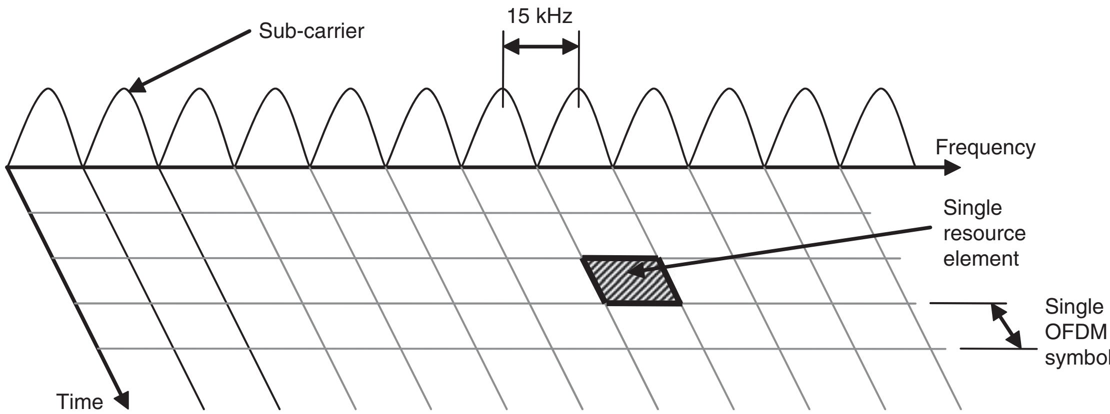

Figure 26.5 *The principle of the OFDM downlink.*

The measurement needs of LTE depend naturally on the development phase of the network. In the initial phase of the system, when the equipment is not yet necessarily available partially or completely, the LTE simulator is a highly useful tool for estimating the network performance under the planned circumstances. They can be utilized for production of performance figures for different traffic scenarios and user profiles, including the combination of different profiles with the varying uplink and downlink data transmission of different applications as FTP upload and download, web browsing, VoIP traffic with varying codec qualities and email traffic. In this way, the functionality and performance can be estimated in advance under the normal and stressed conditions, to make sure the selected capacity and other technical characteristics of the network comply with the estimated traffic.

Also the coverage areas as a function of capacity and quality can be estimated. One of the most useful criteria is the bit error rate as a function of the radio channel type that can be varied between the line-of-sight types of AWGN (Additional White Gaussian Noise) to highly multipropagated environment in the dense city center that produces Rayleigh fading as shown in Figure 26.6. The error margin of the estimated

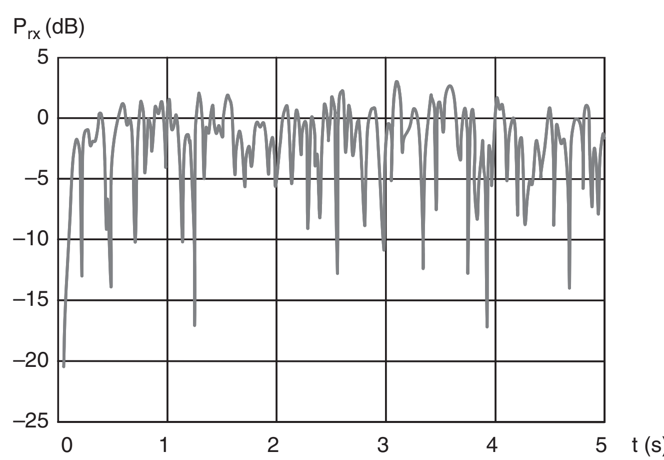

Figure 26.6 *An example of the measurement of the received power level when the Rayleigh fading is present.*

performance depends on accuracy of the simulator's channel models, which can be enhanced via the practical measurements.

At the beginning of the system introduction, in the early phase of the element production, a base station tester is a typical example of the equipment that is needed in the type approval processes. Equally importantly the type approval has to be carried out for the new terminal types. The quality of both hardware and software of the radio equipment needs to be measured in order to avoid problems in the early deployments of the networks. The tester equipment can be integrated into the complete measurement centers that include a large set of tool libraries for the development and execution of the test cases. The tester package typically includes signal generators and RF test equipment meant for the laboratory conditions.

The measurement equipment can be modular, which includes a tester platform for LTE and other mobile technologies, in order to assure the common interworking of the equipment of different systems in accordance with the standards definitions. This type of measurement center may include third party measurement equipments that are integrated to a common control center that takes care of the management of the separate parts of the measurements. This type of solution is useful especially in the terminal testing and pretype approval testing.

In the implementation phase of the LTE, it is still possible to utilize, for example, LTE signal generators and signal analyzers. Both LTE network and terminal interoperability testing between different equipment vendors can be carried out in laboratory conditions by utilizing the signal generators or in the more realistic environment in the indoor and outdoor test networks by utilizing commercial or precommercial terminals.

In the initial phase of new technology as LTE, it can be assumed that there may appear differences in the functionalities of different solutions of different equipment vendors. This is due to the fact that the LTE standards are highly complicated and that it is challenging to create such test specifications and requirements that cover all the essential functionalities of the system. In the case of problems in the interworking, one of the most useful tools is a protocol analyzer. It displays and stores the signaling flow in the wanted level between the network elements. The analysis can be done in real-time or by playing the measurement results in the display of the same analyzer. Also postprocessing can be performed based on stored measurements, either with the same equipment or with a separate, more specific post-processing and analyzer tools.

When the LTE base station infrastructure is about to be integrated and commissioned, the coverage area studies can be carried out with suitable field test equipment. The RF analyzer is one of the most logical selections in the early phase equipments, but as the networks develop, the measurements are most fluent to be carried out with a vehicle mounted outdoor receivers, or with portable testers in the indoors and other locations that are difficult to access with a car. One of the examples is the Nemo Outdoor field tester, which consists of the common platform for the data collection, display and preanalysis, whereas the actual radio measurements are done with commercial LTE terminals and/or third party scanners.

26.4.2 LTE Traffic Simulators

For the core network, that is, EPC, protocol analyzers are useful in network tuning and optimization, as well as in fault management. Nevertheless, if there is no functional network yet available, or the planned features requires further investigation without yet introducing them to the live environment due to potential faults, the performance and functionality of the complicated scenarios can be investigated via the simulators. In case of LTE, which generates considerably more signaling and traffic load compared with previous mobile networks, they can include separate high-performance servers for the creation of realistic loads in large areas.

One example of this type of equipment for the LTE performance evaluation is the NetHawk (EXFO), which has together with Aeroflex developed an integrated test system EAST500. The system includes simulator, measurement and analyzer blocks, which can be utilized for the LTE performance in a large scale before

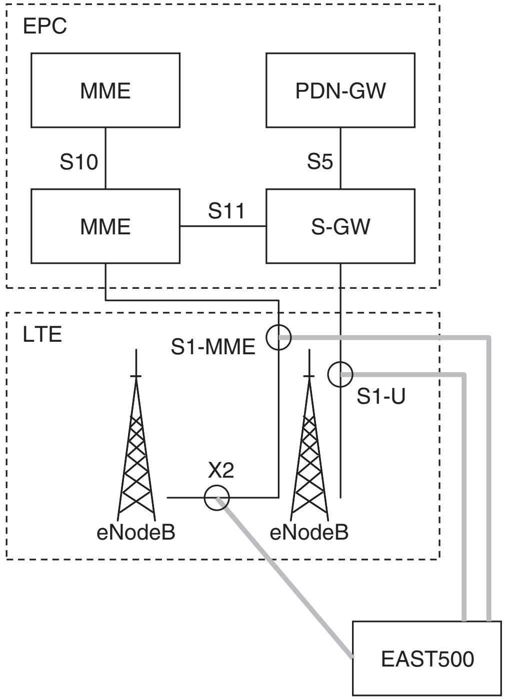

Figure 26.7 *The reference measurement points of LTE load simulations.*

commercial traffic is available. The packet contains the radio interface measurement module by Aeroflex and application tester of NetHawk. The outcome of the investigations gives a realistic estimate about eNodeB performance and capacity behavior in the realistic field conditions and traffic profiles of users. Because eNodeB contains considerably more functionalities compared with UMTS/HSPA NodeB, the measurements and simulation equipment must be updated to support the higher data rates and heavier signaling loads. The data traffic could be estimated to be around ten times more in a typical scenario, which increases the requirements of the analysis accordingly. The equipment thus contains updated IT elements like servers in order to handle the simulated traffic.

Figure 26.7 shows the typical network reference points that can be investigated with the EAST500 equipment. The equipment can be utilized further for the simulating of the user data transmission layer as well as the signaling traffic load via realistic terminal emulation. As the equipment supports all LTE protocol layers, it can be utilized also for the functional testing of eNodeB with varying load, and with different bit error rates that the users sees.

The typical solutions to the fast provisioning of LTE measurements are the SW upgradeable equipment. Some examples of such an earlier equipment type that will have the needed LTE capabilities are the vector analyzer, signal generator, signal emulator and signal analyzer. The updated SW brings the LTE capabilities both in uplink and downlink with the respective LTE definitions. These equipments can be utilized for the measurements of the eNodeB and terminal, for example, in the type approval tests, combined with the baseband signal generator channel emulator and vector signal emulator. The functionality can also include automatic DCI channel coding, channel definitions, power setting and scheduling as well as access of the terminal to the network.

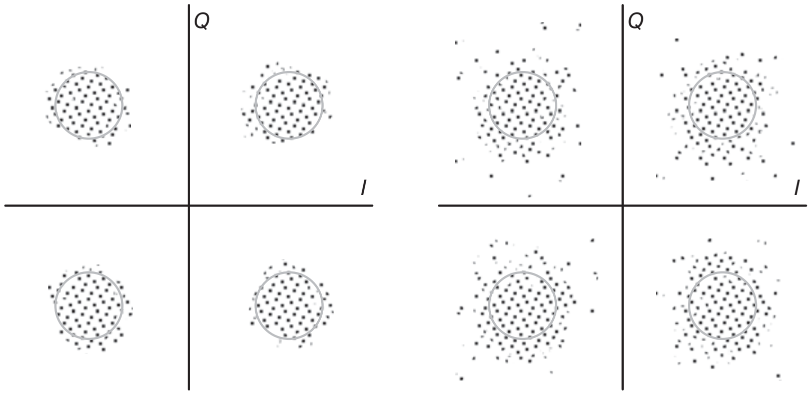

Figure 26.8 An example of the MER levels. The first diagram shows an acceptable MER level, whereas the second diagram might cause already misinterpretations due to the larger variation of the received constellation points.

26.4.3 Typical LTE Measurements

The LTE measurements can be done both in uplink and downlink directions. The typical measurement items may be related to frequency error, transmitted power level, modulation error rate (MER) and error vector magnitude (EVM). MER indicates the accuracy of the modulation, which can be interpreted in the I/Q diagram as shown in Figure 26.8. The EVM value indicates the performance of the demodulator in noisy environments in a slightly different way. The measurement equipment needs to note the position of the symbol obtained from the carrier wave in the output of the demodulator. This result is then compared with the theoretical location of the symbol in the I/Q axis, and the occurred error in the position is shown in percentage scale. Figure 26.9 clarifies the idea of the EVM. The measurements are carried out typically to all of the subcarriers within the selected segment. If the averaging of the results is selected, the maximum and minimum values can also be displayed.

EVM is closely related to the modulation error rate. In practice, it indicates the signal-to-noise ratio of the digitally modulated signal, and it is informed in dB scale. Other essential LTE measurements are:

The constancy of the spectrum indicates the amplitude, difference in the amplitudes, phase and group delay for each subcarrier. This measurement is useful for solving the OFDM specific issues, like intersymbol errors.

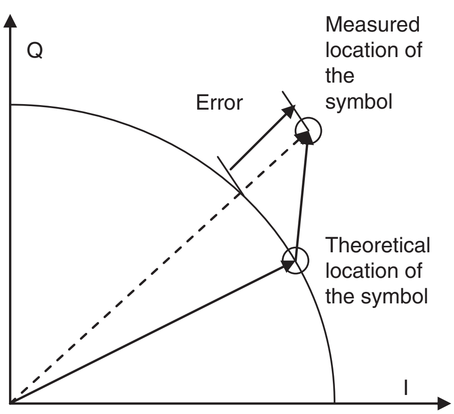

Figure 26.9 The principle of the Error Vector Magnitude (EVM).

The EVM as a function of subcarrier shows graphically the vector errors for each subcarrier for a certain symbol or a group of symbols. The measurement can display at the same time the mean values (RMS) and peak values, which provides the momentary vector error measurements.

The EVM as a function of symbols shows graphically the vector error ratio for each symbol either for a part or for the whole set of subcarriers.

CQI (channel quality information) is one of the most important LTE measurement items. It is derived from the HSPA definitions and indicates the channel quality in a single value by combining different quality related items.

Other useful LTE measurements are also the channels received power level, the utilization level of the band, the leak power of the neighboring channel and the mask of the spectrum source.

Downlink specific LTE measurement set can include the following items: The power vs. resource block shows the power level of each resource block over a single sub frame or over a larger set of subframes. The impact of the power increase for each resource block can thus be visualized by observing the power distributions. It is also possible to observe simultaneously a set of constellation diagrams, which eases in the fault resolutions.

In the uplink direction, the time-based EVM measurement shows individually for each symbol.

As an example, let's investigate the Rohde & Schwarz R&S FSQ-K100 measurement equipment. It can display the following downlink and/or uplink LTE measurement results:

- Power level as a function of time (DL, UL)
- EVM as a function of the carrier (DL)
- EVM as a function of the symbol (DL, UL)
- Frequency error as a function of the symbol (DL)
- EVM as a function of the subframe
- The flatness of the spectrum (DL, UL)
- The group delay of the spectrum (DL, UL)
- The difference of the flatness of the spectrum (DL, UL)
- Constellation diagram (DL, UL)
- Statistics about the CCDF and allocations
- EVM per timeslot (DL)
- Statistics about CCDF (DL, UL), allocations (DL, UL) and signal flow (DL).

26.4.4 Type Approval Measurements

The 3GPP specification 36.214 describes the measurement definitions of the radio interface of E-UTRAN terminals and network, terminal being either in idle or connected mode. The most important examples are the following.

Reference Signal Received Power (RSRP) is calculated as a linear average (in absolute power levels) from those resource elements that are transporting cell specific information. If reception diversity is activated, the received power level is calculated by averaging the power levels of the separate reception paths.

Reference Signal Received Quality (RSRQ) is defined as a ratio of the power level of the resource blocks (RB) and the signal strength of the whole E-UTRA carrier, which is called Carrier Received Signal Strength Indicator.

Other measurements are, for example, the received code power level of UTRA FDD mode, received chip based energy divided by the power density of the band (Ec/No), RSSI and code power of the UTRA TDD mode, downlink power level of the E-UTRA, received interference power level, and thermal noise power level.

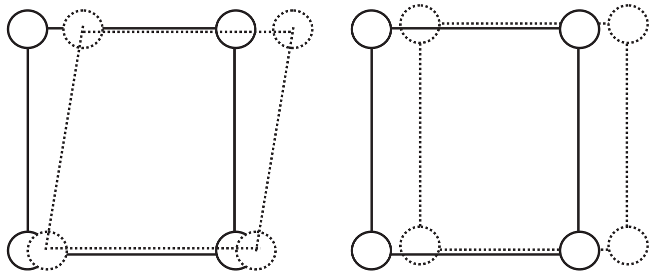

Figure 26.10 Two examples of the possible error behavior in the modulation constellation diagram.

The basic measurements of LTE radio interface are related to the quality of the signals in both uplink and downlink. These measurements include MIMO analysis when applicable. In the basic measurements, the measurement equipment should support in all cases the bandwidth up to 20 MHz, which is the maximum value LTE can utilize. Also all the modulation types should be found in the basic set of measurements, that is, QPSK, 16-QAM and 64-QAM, as is defined in the 3GPP specification of the 36 series. The OFDM basic investigations can be related to the quality of the modulation (e.g. Modulation Error Rate, MER) as well as to the other OFDM functionalities like the size of the transport block, the allocation of the resource block, different protection cycle sizes (CRC) and their combinations.

The quality of the modulation of the transmitter can be investigated at a detailed level, for example, by measuring the EVM value per OFDM carrier, symbol, timeslot and resource block.

26.4.5 Modulation Error Measurements

The interferences and nonoptimal radio conditions in the fast varying signal levels due to, for example, multipath propagation or intersymbol interferences, can change the positions of the occurred samples in the I/Q constellation diagram. Figure 26.10 shows two examples of the erroneous reception. The first diagram shows the twisting of some of the constellation points, and the second one shows the movement of the whole constellation from their theoretical positions. Both can lead to the increased misinterpretation of the values, which leads to the increasing of the bit error rate.

Furthermore, the nonideal equipment may start to twist the constellation as a function of the signal level, that is, further away from the 0-point the sample is, the more it may start finding itself off the constellation point as Figure 26.11 shows.

26.4.6 LTE Performance Simulations

The development of the LTE-based communication system sets major challenges for the testing process. Testing of particular protocol implementations is challenging, as testing often needs to be carried out in both (simulation-based) host and target environments. There is also often a need to simulate the communication towards over- and underlying protocol layers as these may not be available during early phase testing.

As an example, in the CrossNet project, VTT Finland and University of Oulu (CWC) have developed cross-layer testing and simulation framework together with the industrial partners. The developed framework integrates existing industrial tools, EXFO Nethawk EAST simulator, M5 network analyzer and EB Propsim radio channel emulator, to support wrap-around testing of LTE-based telecommunication systems. Figure 26.12 shows the principle of the simulator environment. The setup is valid, for example,

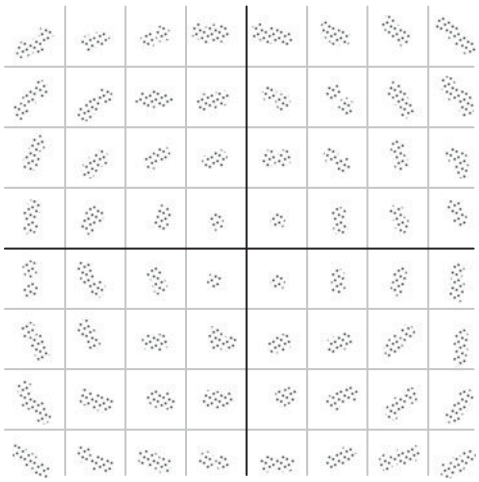

Figure 26.11 An example of the effect of the twisting constellation in the outer part of the diagram, which may be a result of for example, component failure of the receiver.

for the simulations of eNodeB functionality for investigating system performance with a multitude of LTE-UEs.

26.5 LTE Field Measurements

The LTE field measurements indicate a feasible performance of the system in typical radio conditions. The motivation of early technology field trials is to be able to identify inconsistent behavior and ensure that

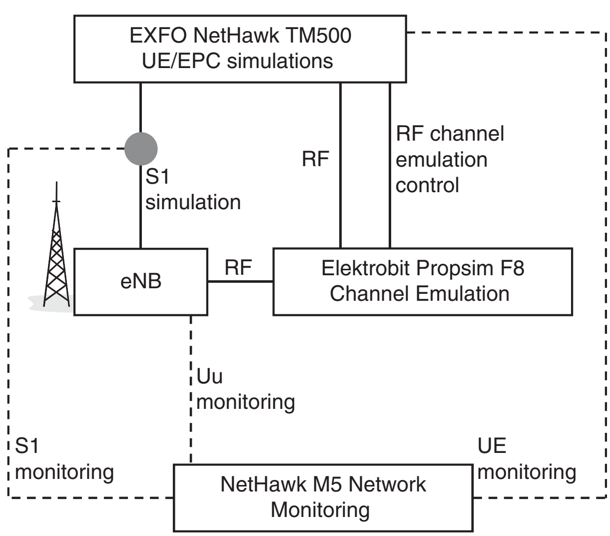

Figure 26.12 CrossNet LTE wrap-around testing environment.

infrastructure and terminal vendors together with the operators identify, implement and test the necessary corrections and optimize end-to-end performance. This assures that the actual commercial network deployments are executed as fluently as possible both as technological issues as well as considering project management.

26.5.1 Typical Field Test Environment

The radio field trial can be located into an area where the EPC, HSS emulator, IMS and applications are also available. In a typical setup, eNBs cover a relatively small area, but sufficiently large in order to provide adjacencies to test mobility scenarios which represent realistic conditions of a live network in terms of interference. The eNB sites can be colocated with existing 2G/3G sites, which normally require the swap of the antenna systems in order to support the LTE band signal on air.

For the operation and maintenance of the eNBs a suitable system is needed., as it is the basis for the parameter, fault and operations handling, as well as for example, the O&M and SON feature testing. A typical LTE network setup can be the following:

- Evolved Packet Core (EPC) – S-GW; P-GW; MME placed in a same site
- HSS emulator for user authentication and authorization
- IMS platform acting as a SIP server for VoIP call support
- Application Server for the TCP/UDP streaming, FTP transfers, latency tests (e.g. ping-based), and HD video streaming.
- Internet access for Internet based scenarios.

In the operator site, the supported elements can logically be located in the equipment room(s) for the purpose of:

- hosting of eNBs and routers; note that the room can also be controlled to manage eNBs of the field remotely by operator's transport network.
- hosting the O&M system equipment that are dedicated to manage LTE eNBs and used for, for example, O&M and SON testing.

In addition, the actual eNBs can be installed on the field both to indoor and outdoor in such a way that they can be activated individually to form partial or complete radio coverage in the test area. In typical trial cases, part of the eNBs can also be utilized as a base for the demos. In the detailed field testing, each eNB can be set up in such a way that it either contributes useful signals, or acts as an interference source in downlink. It should be noted that the interfering eNBs can be configured without S1 connections.

In a typical field testing, static, pedestrian and vehicle radio channel types can be measured and analyzed. This requires a respective plan for the test cases, including the test setup and procedure description, the measurement/signal generator equipment setup and the storing of the performance indicators for postprocessing and analysis. Multiple sites are required in order to confirm the functionality of the handovers. The coverage area can be checked in the cell edge of the area both in noise-limited and interference-limited situation.

One of the important tasks is to select the LTE terminals. In the early state of the LTE system, it can be expected that only data-capable USB stick models are available. In order to make sure that the terminal itself is not the limiting factor, multiple models from various vendors should be considered. This minimizes the effect of the possible nonoptimal behavior of some terminals, and the results thus reflect the network performance.

As the LTE system is capable of delivering considerably faster data connections over the air interface, the physical end-to-end capability of the complete LTE/SAE network should be designed accordingly. The core side should have sufficiently capacity in order to avoid possible bottlenecks in any of the interfaces. If limiting elements or interfaces occur, it should be known well before the actual LTE performance analysis takes place so that misinterpretations of possible lower performance figures can be avoided.

The easiest solution for the remote control of each LTE/SAE element is the centralized O&M system. If it is not available for trial purposes, it is worth considering a setup of local control systems (PC/laptop with element manager software and direct connection to the element) in each site, which can be utilized remotely, for example, via a VPN/CITRIX type of solution. Such remote connections ease the configuration of the network for each test case. It is also worth noting that the possible interferer sites do not require S1 connections in order to have the interference signal on air. This avoids the additional backhaul load and saves capacity. If the eNBs are connected to the backhaul via, for example, over a Gbit connection, a single E1 connection of PDH network or other feasible low-capacity connection is enough for the connectivity of the interfering sites to the controlling test network PC/laptop. E1/Ethernet converters can thus be utilized in the interferer sites.

26.5.2 Test Network Setup

The LTE field test equipment can consist of the drive test tool for generating the data calls and the respective KPIs, scanner to store the received power level, and core protocol analyzer. GPS should be included to the radio measurement setups by default in order to correlate the results geographically. Also additional troubleshooting tools are probably needed. These can be more specific signal and protocol analyzers in the core side, and additional analyzer elements for the test drive tools.

The LTE performance can be measured under varying conditions as for the propagation environment (dense urban, urban, suburban, rural, open) and user profiles. The effect of the latter is quite important as the load figures of the network dictate the single user throughput, for example, in the average and peak load. In order to create the loading in a controlled way, general models can be applied. These models are available, for example, from 3GPP, which has defined a macro case 1 (with 500 m of intersite distance) and case 3 (with 1752 m of intersite distance) by assuming 10 users per cell as a reference.

For test cases where interference is present, the interfering signal can possibly be generated via the eNB itself if the equipment provides this type of test signal generation. As an example, NSN eNB can be set up to send artificial signals on the physical DL channels which can be configured by the so-called PhyTest Interface. As an example, to generate 50% interference load means that half the available active PRBs in PDCCH and PDSCH are used to send dummy data. For the full 100% load, all physical channels are transmitting at all times, and all the PRBs are used.

Figure 26.13 shows the idea of the measurement setup when DL interferences are created intentionally.

The uplink interference is generated by the transmitting power of UEs served by neighboring/other cells to the one considered. The closer the interfering UEs to the cell border the more transmitting power it will use and the lower the path loss to the victim cell which means increased interference to the UEs in the victim cell.

The uplink interference is challenging to simulate or generate in the field as it varies over time and frequency. A common approach in trials is to generate UL interference in the field. The equipment can typically be configured in Layer1 measurements to transmit a pseudo-random sequence on the PUSCH with a determined amount of Tx Power.

The total amount of TX power transmitted by the UE is adjusted according to the path loss of the TM500 location and the victim cell in order to generate the desired noise level at the victim cell. As the TM500 is configured in Layer 1 measurements it does not require to be camping on any interferer cell. This allows the TM500 to be strategically placed anywhere in the victim's cell coverage area, increasing the flexibility of the configuration setup and the amount of interference that can be generated. For field test cases, the location in

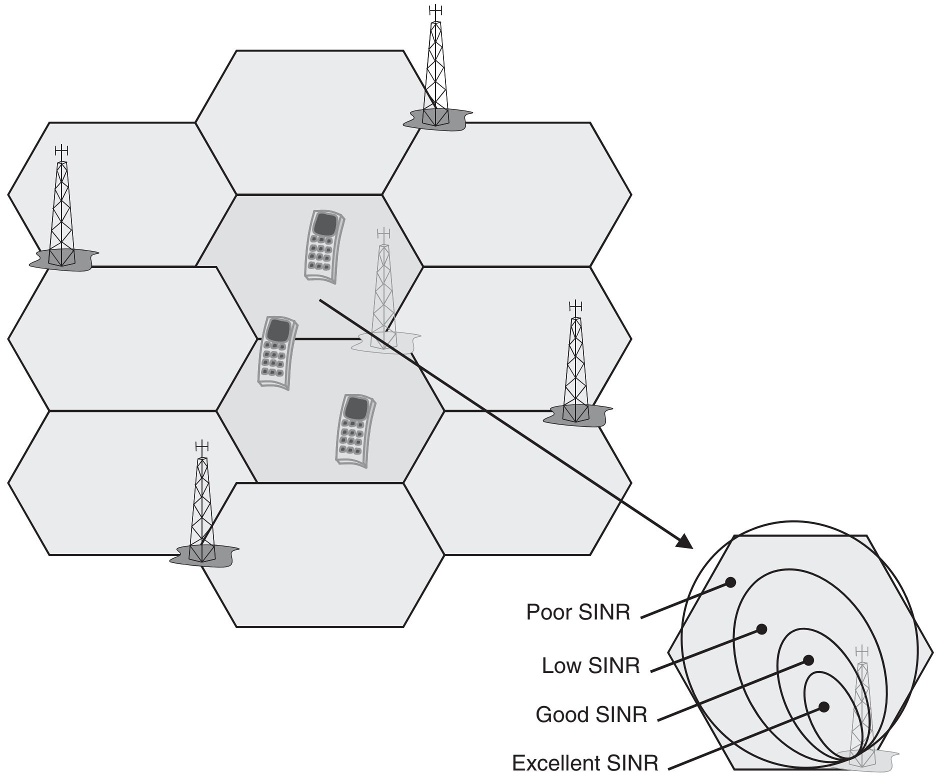

Figure 26.13 *A possible field measurement setup for the interference investigations [4].*

the victim cell and L1 parameters of the TM500 is adjusted in order to generate the desired noise rise required for the UL loaded test cases.

Uplink interference, as shown in Figure 26.14, can also be generated with a Signal Generator by using the transmission patterns obtained previously. This approach is chosen for early LTE field trials as it is a field-proven concept and does not require HW modifications at the eNB site.

The Vector Signal Generator can typically be equipped with an internal baseband generator which allows the generation of LTE digital standards with a sufficiently wide frequency range.

It is quite clear that the uplink interference depends on “Interfering” UE position, that is, the more the interfering set of UEs are located at the cell border, the more transmitting power they will use and also the path-loss to the victim cell becomes lower. Consequently, the produced interference to the victim cell increases.

As a basis for DL performance indicator evaluation, the SINR (Signal to Interference and Noise Ratio) ranges can be fixed, for example, in the following way for 100% DL load:

– Excellent location: SINR $\geq$ 20 dB
– Good SINR (High Location) 12 $\leq$ SINR < 20
– Average SINR (Medium Location): 6 $\leq$ SINR < 12
– Poor but functional: 0 $\leq$ SINR < 6
– Bad SINR (Low Location): $\leq$ 0 dB.

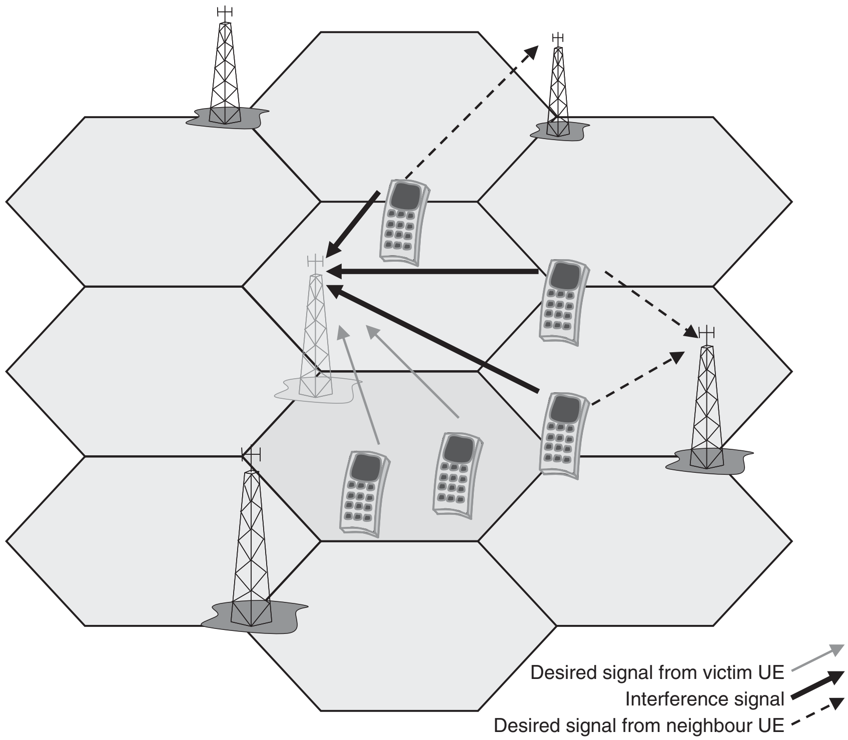

Figure 26.14 *The idea of the UL interference creation for the trial test cases [4].*

For the UL, Load Conditions (%) as a function of IoT Range can be fixed, for example, as following: 50% load => 8 dB, 100% load => 11.5 dB.

In the LTE radio test setup, it is important to decide whether a standalone terminal as such, or a terminal with an external antenna, is utilized. The reasoning for the external antenna mounted, for example, on the rooftop of the vehicle, is that it creates always similar receiving conditions per test drive route regardless of the position of the actual terminal. The coverage area turns out to be larger, though, in this setup compared to the terminal without additional antenna. The benefit of the terminal without external antenna is thus close to the typical conditions of normal users.

Also the characteristics of the terminal are important to take into account. The most limiting factor is the terminal class. Also the support of different MIMO antenna ports has an effect on data throughput.

The successful execution of a trial requires a project plan and a realistic time and resource planning. The main tasks of the trial are the following:

– Initial plan of the scope
– Overall test plan
– Network architecture plan
– Detailed test plan
– Test network deployment
– Test case execution
– Data postprocessing and analysis
– Reporting.

As an outcome of the trial, the functionality of the LTE system and the investigated features are evaluated. Also the performance of the network with different test cases is evaluated.

26.5.3 Test Case Selection

The overall LTE test scope may cover the following study items:

Validation period. This is the first stage test period with the overall aim to make sure the network is correctly constructed and that the equipment is functioning as planned. This stage can include the system validation tests, site configuration and design validation tests. The outcome of this stage is information about the general coverage area, with, for example, single user throughput in downlink and uplink, functioning and performance of the terminal attach and detach procedures, data streaming performance, possibly in static, pedestrian and vehicular radio channel types. The tests can include also downlink and uplink interference, when its significance is altered from noninterfered level up to maximum level.

Extensive testing period. This stage aims to analyze the capacity and throughput sanity testing. It can include the cases above in a more thorough way, including single user throughput in downlink and uplink, deeper data streaming by utilizing TCP and UDP, in static, pedestrian and vehicular environment by altering the vehicular speed and interference levels. This stage can also contain MIMO cases, and varying capacities by altering the number of the LTE terminals. In addition, the following items are some of the examples that can be included in the test set of this stage:

- Scheduler tests via TCP streaming by varying the number of terminals under different channel types and interference levels.
- End-to-end latency tests for the attach procedure latency. The cases can be varied under different channel types, network load levels and interference levels. The straightforward way to execute these tests is the use of PING by varying the packet size (small, medium, large). In order to increase the statistical accuracy, several PINGs can be collected by test case in order to present the latency distribution.
- Mobility and handover procedure performance test, including the measurements of the handover success rate and interruption time for different environment and traffic types (TCP and UDP streaming as well as VoIP calls), and by varying the terminal speed. The handover performance can be investigated for inter and intra cell and site situations. For the voice calls, also Mean Opinion Score (MOS) can be measured with respective field test equipment.
- Radio features tests by varying feature parameter settings, for example, for the Quality of Service related features, or for the uplink closed and open loop power control variants with a certain number of terminals placed in a certain constellation and moving them between the site and coverage border.
- Application performance tests, including, for example, the success rate and setup speed of Internet access, IMS VoIP, FTP, video streaming, HTTP browsing, e-mailing and HTTP download.
- Robustness and availability tests, including, for example, the attach procedure under different interference levels and the TCP streaming by adjusting the TCP window size.
- Operations and maintenance related tests, including functionality of the main functionalities of the tool.
- Self Organizing Network (SON) related tests, including the validation of the correct execution of the commands and the functionality of the SON features.

26.5.4 Items to Assure

In the trial and commercial LTE network deployment, the maximum data speed is much higher in the radio interface compared to any other previous mobile communication networks. It is thus important to take into account all possible bottlenecks of the LTE/SAE network in order to make sure that the end-to-end performance will be at the expected level. It can thus be expected that the data velocity limitations in the transport network can possibly cause problems due to the existence of several equipment, interfaces and settings that not all are under the control of the operator.

As an example, there are experiences about the TCP DL single-user streaming which does not exceed a certain, relatively low value compared with the expected theoretical maximum data speed, even under excellent radio conditions. An analysis of the transport network element distribution along the whole communications route is thus advisable. The problem might be derived from the router data speed limitations related to each single interface. As an example, a data speed limitation of 45 Mb/s can be found in a single Gigabit router, which is sufficient to lower the whole end-to-end data speed. Adjusting this limit to the maximum default value that can be, for example, 200 Mb/s, can remove the problem.

The problem can also be related to the sporadic throughput fluctuations for UDP streaming in the uplink, and for TCP streaming in both directions. The fluctuations can be seen even with relatively simple tools like iPerf, which may show the sudden droppings to some lower data speed level, and slow initiation of the data transmission. This might be caused in nonoptimal performance of the LTE terminal, especially in the initial phase of LTE systems. In fact, there are cases where the LTE terminal occasionally sends a buffer status report "empty" even if the case is not so, which can trigger the problem. Also the TCP window size can cause the problem. In the testing purposes, parallel TCP stream with smallest TCP window size can be utilized in order to reach the throughput stability.

The aim of the trial phase of the LTE is to test the performance of the network, but in some cases this phase also reveals behavior of some of the involved elements that is not directly related to LTE, but needs either parameter adjustments or even complete replacement. Especially, if the core network contains "old-fashioned" solutions, they might work sufficiently well for any previous data rates, but the higher speed of the LTE might reveal the limitations of the hardware and/or software.

References

1. Penttinen, J. (1999) Tulevaisuuden GSM-datapalveluiden vaikutus radioverkkosuunnitteluun (the effects of future GSM data services on the radio network planning). Lic. thesis 28/1999, Helsinki University of Technology, Department of Electrical Engineering, 123 p.
2. Zanini, P. (1997) Examining the GPRS radio interface and evaluating GPRS radio performance. GPRS & HSCSD Conference, Centre for EuroTelecomms, London 22–24 October 1997.
3. GSM TS Specification 05.05, Radio Interface.
4. Penttinen, J.T.J. (2011) *The LTE/SAE Deployment Handbook*, John Wiley & Sons, Ltd.

Index

The Telecommunications Handbook: Engineering Guidelines for Fixed, Mobile and Satellite Systems, First Edition.
Edited by Jyrki T. J. Penttinen.