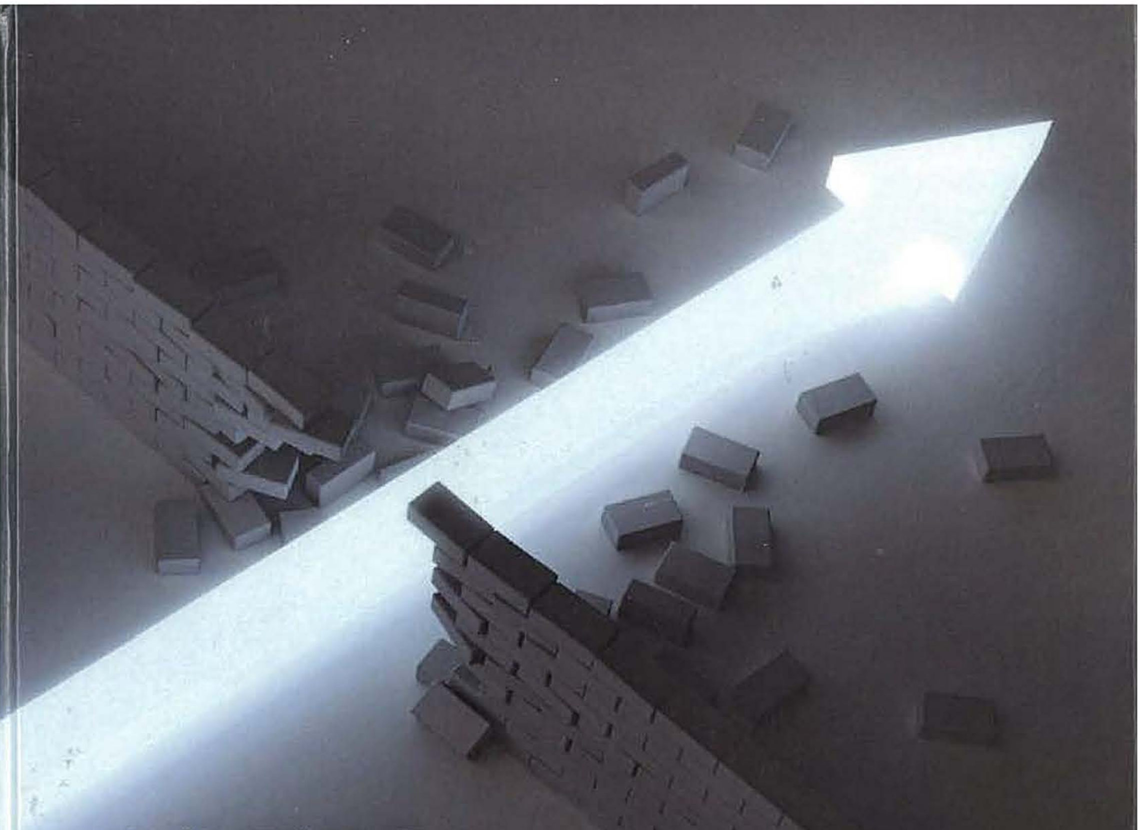


LTE

The UMTS Long Term Evolution

FROM THEORY TO PRACTICE

Edited by: Stefania Sesia • Issam Toufik • Matthew Baker

WILEY

LTE – The UMTS Long Term Evolution

From Theory to Practice

Stefania Sesia
ST-NXP Wireless/ETSI, France

Issam Toufik
ST-NXP Wireless, France

Matthew Baker
Philips Research, UK

WILEY
A John Wiley and Sons, Ltd, Publication

This edition first published 2009
© 2009 John Wiley & Sons Ltd.

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.

The right of the author to be identified as the author of this work has been asserted in accordance with the Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, except as permitted by the UK Copyright, Designs and Patents Act 1988, without the prior permission of the publisher.

Figure on cover: Reproduced by permission of © 3GPP™.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Designations used by companies to distinguish their products are often claimed as trademarks. All brand names and product names used in this book are trade names, service marks, trademarks or registered trademarks of their respective owners. The publisher is not associated with any product or vendor mentioned in this book. This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold on the understanding that the publisher is not engaged in rendering professional services. If professional advice or other expert assistance is required, the services of a competent professional should be sought.

Library of Congress Cataloging-in-Publication Data

Sesia, Stefania.
LTE–the UMTS long term evolution : from theory to practice / Stefania Sesia, Matthew Baker, and Issam Toufik.
p. cm.
Includes bibliographical references and index.
ISBN 978-0-470-69716-0 (cloth)
1. Universal Mobile Telecommunications System. I. Baker, Matthew (Matthew P. J.) II. Toufik, Issam. III. Title.
TK5103.4883.S47 2009
621.3845'6–dc22
2008041823

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

ISBN 9780470697160 (H/B)

FSC
Mixed Sources

List of Acronyms

3GPP 3rd Generation Partnership Project
3GPP2 3rd Generation Partnership Project 2
AC Access Class
ACI Adjacent Channel Interference
ACIR Adjacent Channel Interference Ratio
ACK Acknowledgement
ACLR Adjacent Channel Leakage Ratio
ACS Adjacent Channel Selectivity
ADC Analogue to Digital Converter
ADSL Asymmetric Digital Subscriber Line
AGI Antenna Gain Imbalance
AM Acknowledged Mode
AMC Adaptive Modulation and Coding
AMPS Analogue Mobile Phone System
AMR Adaptive MultiRate
ANR Automatic Neighbour Relation
ANRF Automatic Neighbour Relation Function
AoA Angle-of-Arrival
AoD Angle-of-Departure
APN Access Point Name
APP A-Posteriori Probability
ARFCN Absolute Radio Frequency Channel Number
ARIB Association of Radio Industries and Businesses
ARP Almost Regular Permutation*
ARP Allocation and Retention Priority*
ARQ Automatic Repeat reQuest
AS Access Stratum*
AS Angular Spread*
A-SEM Additional SEM
ATDMA Advanced TDMA
ATIS Alliance for Telecommunications Industry Solutions
AuC Authentication Centre
AWGN Additive White Gaussian Noise
BCC Base station Colour Code
BCH Broadcast CHannel
BCCH Broadcast Control CHannel
BCJR Algorithm named after its inventors, Bahl, Cocke, Jelinek and Raviv
BER Bit Error Rate
BLER BLock Error Rate
BM-SC Broadcast-Multicast Service Centre
BP Belief Propagation
BPRE Bits Per Resource Element
bps bits per second
BPSK Binary Phase Shift Keying
BSIC Base Station Identification Code
BSR Buffer Status Reports
CAZAC Constant Amplitude Zero AutoCorrelation
CB Circular Buffer
CCCH Common Control CHannel
CCE Control Channel Element
CCI Co-Channel Interference
CCO Cell Change Order
CCSA China Communications Standards Association

CDD Cyclic Delay Diversity

CDF Cumulative Distribution Function

CDL Clustered Delay Line

CDM Code Division Multiplex(ed/ing)

CDMA Code Division Multiple Access

C/I Carrier-to-Interference ratio

CF Contention-Free

CFI Control Format Indicator

CFO Carrier Frequency Offset

CINR Carrier-to-Interference-and-Noise Ratio

CIR Channel Impulse Response

CM Cubic Metric

CMHH Constant Modulus HouseHolder

CN Core Network

CODIT UMTS Code DIvision Testbed

COFDM Coded OFDM

CP Cyclic Prefix

CPICH Common PIlot CHannel

CPR Common Phase Rotation

CPT Control PDU Type

CQI Channel Quality Indicator

CRC Cyclic Redundancy Check

C-RNTI Cell Radio Network Temporary Identifier

CS Circuit-Switched

CSG Closed Subscriber Group

CSI Channel State Information

CSIT Channel State Information at the Transmitter

CTF Channel Transfer Function

CVA Circular Viterbi Algorithm

CVQ Channel Vector Quantization

CW Continuous-Wave

DAB Digital Audio Broadcasting

DAC Digital to Analogue Converter

dB deci-Bel

d.c. direct current

DCCH Dedicated Control CHannel

DCFB Direct Channel FeedBack

DCI Downlink Control Information

DFT Discrete Fourier Transform

DFT-S-OFDM DFT-Spread OFDM

Diffserv Differentiated Services

DL DownLink

DL-SCH DownLink Shared CHannel

DMB Digital Mobile Broadcasting

DM RS DeModulation RS

DOA Direction Of Arrival

DPC Dirty-Paper Coding

DRB Data Radio Bearer

DRX Discontinuous Reception

DS-CDMA Direct-Sequence Code Division Multiple Access

DSP Digital Signal Processor

DTCH Dedicated Traffic CHannel

DTX Discontinuous Transmission

DVB-H Digital Video Broadcasting – Handheld

DVB-T Digital Video Broadcasting – Terrestrial

DwPTS Downlink Pilot TimeSlot

ECM EPS Connection Management

EDGE Enhanced Data rates for GSM Evolution

EESM Exponential Effective SINR Mapping

EMM EPS Mobility Management

eNodeB evolved NodeB

EPA Extended Pedestrian A

EPC Evolved Packet Core

EPS Evolved Packet System

ESP Encapsulating Security Payload

ETSI European Telecommunications Standards Institute

ETU Extended Typical Urban

E-UTRA Evolved-UTRA

E-UTRAN Evolved-UTRAN

EVA Extended Vehicular A

EVM Error Vector Magnitude

FACH Forward Access CHannel

FB Frequency Burst

FCCH Frequency Control CHannel
FDD Frequency Division Duplex
FDE Frequency Domain Equalizer
FDM Frequency Division Multiplexing
FDMA Frequency Division Multiple Access
FDSS Frequency Domain Spectral Shaping
FFT Fast Fourier Transform
FI Framing Info
FIR Finite Impulse Response
FMS First Missing SDU
FSTD Frequency Switched Transmit Diversity
FTP File Transfer Protocol
FTTH Fibre-To-The-Home
GBR Guaranteed Bit Rate
GCL Generalized Chirp-Like
GERAN GSM EDGE Radio Access Network
GGSN Gateway GPRS Support Node
GMSK Gaussian Minimum-Shift Keying
GPRS General Packet Radio Service
GPS Global Positioning System
GSM Global System for Mobile communications
GT Guard Time
GTP GPRS Tunnelling Protocol
GTP-U GTP-User plane
HARQ Hybrid Automatic Repeat reQuest
HD-FDD Half-Duplex FDD
HFN Hyper Frame Number
HII High Interference Indicator
HLR Home Location Register
HRPD High Rate Packet Data
HSDPA High Speed Downlink Packet Access
HSPA High Speed Packet Access
HSPA+ High Speed Packet Access Evolution
HSS Home Subscriber Server
HSUPA High Speed Uplink Packet Access
HTTP HyperText Transfer Protocol
ICI InterCarrier Interference
ICIC InterCell Interference Coordination
IDFT Inverse Discrete Fourier Transform
IETF Internet Engineering Task Force
IFDMA Interleaved Frequency Division Multiple Access
IFFT Inverse Fast Fourier Transform
i.i.d. Independently identically distributed
IM Implementation Margin
IMD InterModulation Distortion
IMS IP Multimedia Subsystem
IMSI International Mobile Subscriber Identity
IMT International Mobile Telecommunications
IP Internet Protocol
IR Incremental Redundancy
IRC Interference Rejection Combining
ISD InterSite Distance
ISI InterSymbol Interference
IST-WINNER Information Society Technologies-Wireless world INitiative NEw Radio
ITU International Telecommunication Union
ITU-R ITU Radiocommunication sector
J-TACS Japanese Total Access Communication System
LA Local Area
LB Long Block
LBP Layered Belief Propagation
LBRM Limited Buffer Rate Matching
LCID Logical Channel ID
LDPC Low-Density Parity Check
LI Length Indicator
LLR Log-Likelihood Ratio
LMMSE Linear MMSE
LNA Low Noise Amplifier
LO Local Oscillator
LOS Line-Of-Sight
LS Least Squares
LSF Last Segment Flag
LTE Long-Term Evolution

MA Metropolitan Area

MAC Medium Access Control

MAC-I Message Authentication Code for Integrity

MAN Metropolitan Area Network

MAP Maximum A posteriori Probability

MBMS Multimedia Broadcast/Multicast Service

MBMS GW MBMS GateWay

MBR Maximum Bit Rate

MBSFN Multimedia Broadcast Single Frequency Network

MCCH Multicast Control CHannel

MCE Multicell/Multicast Coordination Entity

MCH Multicast CHannel

MCL Minimum Coupling Loss

MCS Modulation and Coding Scheme

Mcps Megachips per second

MDS Minimum Discernible Signal

MediaFLO Media Forward Link Only

MIB Master Information Block

MIMO Multiple-Input Multiple-Output

MIP Mobile Internet Protocol

MISO Multiple-Input Single-Output

ML Maximum Likelihood

MLD Maximum Likelihood Detector

MME Mobility Management Entity

MMSE Minimum MSE

MO Mobile Originated

M-PSK M-ary Phase-Shift Keying

MQE Minimum Quantization Error

MRC Maximum Ratio Combining

MSAP MCH Subframe Allocation Pattern

MSB Most Significant Bit

MSE Minimum Squared Error

MSR Maximum Sensitivity Reduction

MTCH Multicast Traffic CHannel

MU-MIMO Multi-User MIMO

NACC Network Assisted Cell Change

NACK Negative ACKnowledgement

NACS NonAdjacent Channel Selectivity

NAS Non Access Stratum

NCC Network Colour Code

NCL Neighbour Cell List

NDI New Data Indicator

NF Noise Figure

NGMN Next Generation Mobile Networks

NLMS Normalized Least-Mean-Square

NLOS Non-Line-Of-Sight

NMT Nordic Mobile Telephone

NNSF NAS Node Selection Function

NodeB The base station in WCDMA systems

O&M Operation and Maintenance

OBPD Occupied Bandwidth Power De-rating

OBW Occupied BandWidth

OFDM Orthogonal Frequency Division Multiplexing

OFDMA Orthogonal Frequency Division Multiple Access

OI Overload Indicator

OOB Out-Of-Band

P/S Parallel-to-Serial

PA Power Amplifier

PAN Personal Area Network

PAPR Peak-to-Average Power Ratio

PBCH Physical Broadcast CHannel

PBR Prioritized Bit Rate

PCC Policy Control and Charging

PCCH Paging Control CHannel

P-CCPCH Primary Common Control Physical CHannel

PCEF Policy Control Enforcement Function

PCFICH Physical Control Format Indicator CHannel

PCG Project Coordination Group

PCH Paging CHannel

PCI Physical Cell Identity

P-CPICH Primary Common Pilot CHannel

PCRF Policy Control and charging Rules Function

PDCCH Physical Downlink Control CHannel

PDCP Packet Data Convergence Protocol

PDN Packet Data Network

PDP Power Delay Profile

PDSCH Physical Downlink Shared CHannel

PDU Protocol Data Unit

PF Paging Frame

PFS Proportional Fair Scheduling

P-GW PDN GateWay

PHICH Physical Hybrid ARQ Indicator CHannel

PLL Phase-Locked Loop

PLMN Public Land Mobile Network

P-MCCH Primary MCCH

PMCH Physical Multicast CHannel

PMI Precoding Matrix Indicators

PMIP Proxy MIP

PN Pseudo-Noise

PO Paging Occasion

PRACH Physical Random Access CHannel

PRB Physical Resource Block

P-RNTI Paging RNTI

PS Packet-Switched

P-SCH Primary Synchronization CHannel

PSD Power Spectral Density

PSS Primary Synchronization Signal

PUCCH Physical Uplink Control CHannel

PUSCH Physical Uplink Shared CHannel

PVI Precoding Vector Indicator

QAM Quadrature Amplitude Modulation

QCI QoS Class Identifier

QoS Quality-of-Service

QPP Quadratic Permutation Polynomial

QPSK Quadrature Phase Shift Keying

RA Random Access

RACH Random Access CHannel

RAN Radio Access Network

RAR Random Access Response

RA-RNTI Random Access Radio Network Temporary Identifier

RAT Radio Access Technology

RB Resource Block

RE Resource Element

REG Resource Element Group

RF Radio Frequency

RFC Request For Comments

RI Rank Indicator

RLC Radio Link Control

RLS Recursive Least Squares

RM Rate Matching

RNC Radio Network Controller

RNTI Radio Network Temporary Identifier

RNTP Relative Narrowband Transmit Power

ROHC RObust Header Compression

RoT Rise over Thermal

RPF RePetition Factor

R-PLMN Registered PLMN

RRC Radio Resource Control*

RRC Root-Raised-Cosine*

RRM Radio Resource Management

RS Reference Signal

RSCP Received Signal Code Power

RSRP Reference Signal Received Power

RSRQ Reference Signal Received Quality

RSSI Received Signal Strength Indicator

RTCP Real-time Transport Control Protocol

RTD Round-Trip Delay

RTP Real-time Transport Protocol

RTT Round-Trip Time

RV Redundancy Version

S/P Serial-to-Parallel

S1AP S1 Application Protocol

SAE System Architecture Evolution

SAP Service Access Point

SAW Stop-And-Wait

SB Short Block*
SB Synchronization Burst*
SBP Systematic Bit Puncturing
SC-FDMA Single-Carrier Frequency Division Multiple Access
SCH Synchronization CHannel
SCM Spatial Channel Model
SCME Spatial Channel Model Extension
SCTP Stream Control Transmission Protocol
SDMA Spatial Division Multiple Access
SDO Standards Development Organization
SDU Service Data Unit
SEM Spectrum Emission Mask
SFBC Space-Frequency Block Code
SFDR Spurious-Free Dynamic Range
SFN System Frame Number
SGSN Serving GPRS Support Node
S-GW Serving GateWay
SI System Information
SIB System Information Block
SIC Successive Interference Cancellation
SIMO Single-Input Multiple-Output
SINR Signal-to-Interference plus Noise Ratio
SIP Session Initiation Protocol
SIR Signal-to-Interference Ratio
SI-RNTI System Information Radio Network Temporary Identifier
SISO Single-Input Single-Output*
SISO Soft-Input Soft-Output*
S-MCCH Secondary MCCH
SMS Short Message Service
SN Sequence Number
SNR Signal-to-Noise Ratio
SO Segmentation Offset
SON Self-Optimizing Networks
SPA Sum-Product Algorithm
SPS Semi-Persistent Scheduling
SPS-C-RNTI Semi-Persistent Scheduling C-RNTI
SR Scheduling Request
SRB Signalling Radio Bearer
SRNS Serving Radio Network Subsystem
SRS Sounding Reference Signal
S-SCH Secondary Syncronization CHannel
SSS Secondary Synchronization Signal
STBC Space-Time Block Code
S-TMSI SAE-Temporary Mobile Subscriber Identity
STTD Space-Time Transmit Diversity
SU-MIMO Single-User MIMO
SVD Singular-Value Decomposition
TA Tracking Area
TACS Total Access Communication System
TB Transport Block
TCP Transmission Control Protocol
TDD Time Division Duplex
TDL Tapped Delay Line
TDMA Time Division Multiple Access
TD-SCDMA Time Division Synchronous Code Division Multiple Access
TEID Tunnelling End ID
TF Transport Format
TFT Traffic Flow Template
TM Transparent Mode
TMD Transparent Mode Data
TNL Transport Network Layer
TNMSE Truncated Normalized Mean-Squared Error
TPC Transmitter Power Control
TPD Total Power De-rating
TR Tone Reservation
TSC Training Sequence Code
TSG Technical Specification Group
TTA Telecommunications Technology Association
TTC Telecommunications Technology Committee
TTI Transmission Time Interval

TU Typical Urban

UDP User Datagram Protocol

UE User Equipment

UL UpLink

UL-SCH UpLink Shared CHannel

UM Unacknowledged Mode

UMB Ultra-Mobile Broadband

UMTS Universal Mobile Telecommunications System

UP Unitary Precoding

UpPTS Uplink Pilot TimeSlot

US Uncorrelated-Scattered

USIM Universal Subscriber Identity Module

UTRA Universal Terrestrial Radio Access

UTRAN Universal Terrestrial Radio Access Network

VA Viterbi Algorithm

VCB Virtual Circular Buffer

VCO Voltage-Controlled Oscillator

VoIP Voice-over-IP

VRB Virtual Resource Block

WA Wide Area

WAN Wide Area Network

WCDMA Wideband Code Division Multiple Access

WFT Winograd Fourier Transform

WG Working Group

WiMAX Worldwide interoperability for Microwave Access

WINNER Wireless world INitiative NEw Radio

WLAN Wireless Local Area Network

WPD Waveform Power De-rating

WRC World Radiocommunication Conference

WSS Wide-Sense Stationary

WSSUS Wide-Sense Stationary Uncorrelated Scattering

ZC Zadoff–Chu

ZCZ Zero Correlation Zone

ZF Zero-Forcing

ZFEP Zero-Forcing Equal Power

*This acronym can have different meanings depending on the context. The meaning is clearly indicated in the chapter when used.

2

Network Architecture

Sudeep Palat and Philippe Godin

2.1 Introduction

As mentioned in the preceding chapter, LTE has been designed to support only packet-switched services, in contrast to the circuit-switched model of previous cellular systems. It aims to provide seamless Internet Protocol (IP) connectivity between User Equipment (UE) and the Packet Data Network (PDN), without any disruption to the end users' applications during mobility. While the term 'LTE' encompasses the evolution of the radio access through the Evolved-UTRAN (E-UTRAN), it is accompanied by an evolution of the non-radio aspects under the term 'System Architecture Evolution' (SAE) which includes the Evolved Packet Core (EPC) network. Together LTE and SAE comprise the Evolved Packet System (EPS).

EPS uses the concept of *EPS bearers* to route IP traffic from a gateway in the PDN to the UE. A bearer is an IP packet flow with a defined Quality of Service (QoS) between the gateway and the UE. The E-UTRAN and EPC together set up and release bearers as required by applications.

In this chapter, we present the overall EPS network architecture, giving an overview of the functions provided by the Core Network (CN) and E-UTRAN. The protocol stack across the different interfaces is then explained, along with an overview of the functions provided by the different protocol layers. Section 2.4 outlines the end-to-end bearer path including QoS aspects and provides details of a typical procedure for establishing a bearer. The remainder of the chapter presents the network interfaces in detail, with particular focus on the E-UTRAN interfaces and the procedures used across these interfaces, including those for the support of user mobility. The network elements and interfaces used solely to support broadcast services are covered in Chapter 14.

LTE – The UMTS Long Term Evolution: from Theory to Practice Stefania Sesia, Issam Toufik and Matthew Baker

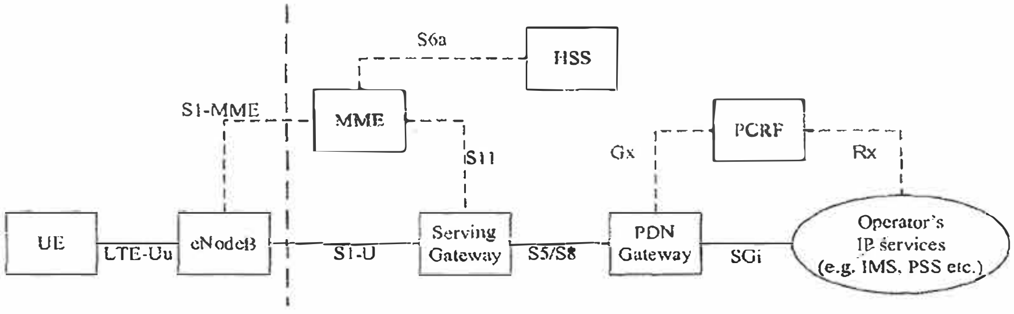


Figure 2.1 The EPS network elements.

2.2 Overall Architectural Overview

EPS provides the user with IP connectivity to a PDN for accessing the Internet, as well as for running services such as Voice over IP (VoIP). An EPS bearer is typically associated with a QoS. Multiple bearers can be established for a user in order to provide different QoS streams or connectivity to different PDNs. For example, a user might be engaged in a voice (VoIP) call while at the same time performing web browsing or File Transfer Protocol (FTP) download. A VoIP bearer would provide the necessary QoS for the voice call, while a best-effort bearer would be suitable for the web browsing or FTP session.

The network must also provide sufficient security and privacy for the user and protection for the network against fraudulent use.

This is achieved by means of several EPS network elements which have different roles. Figure 2.1 shows the overall network architecture including the network elements and the standardized interfaces. At a high level, the network is comprised of the CN (EPC) and the access network (E-UTRAN). While the CN consists of many logical nodes, the access network is made up of essentially just one node, the evolved NodeB (eNodeB), which connects to the UEs. Each of these network elements is inter-connected by means of interfaces which are standardized in order to allow multivendor interoperability. This gives network operators the possibility to source different network elements from different vendors. In fact, network operators may choose in their physical implementations to split or merge these logical network elements depending on commercial considerations. The functional split between the EPC and E-UTRAN is shown in Figure 2.2. The EPC and E-UTRAN network elements are described in more detail below.

2.2.1 The Core Network

The CN (called EPC in SAE) is responsible for the overall control of the UE and establishment of the bearers. The main logical nodes of the EPC are:

- PDN Gateway (P-GW);
- Serving Gateway (S-GW);
- Mobility Management Entity (MME).

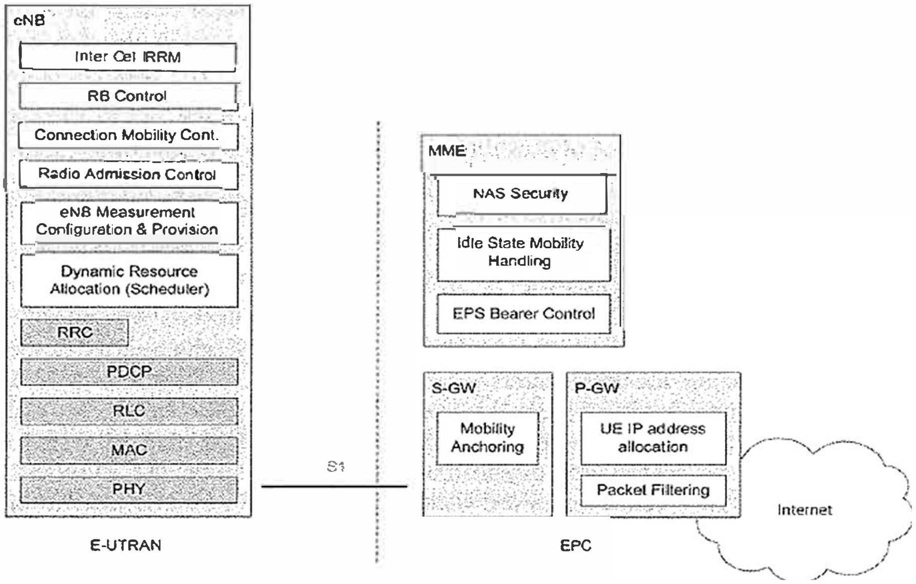


Figure 2.2 Functional split between E-UTRAN and EPC. Reproduced by permission of © 3GPP.

In addition to these nodes, EPC also includes other logical nodes and functions such as the Home Subscriber Server (HSS) and the Policy Control and Charging Rules Function (PCRF). Since the EPS only provides a bearer path of a certain QoS, control of multimedia applications such as VoIP is provided by the IP Multimedia Subsystem (IMS) which is considered to be outside the EPS itself.

The logical CN nodes (specified in [1]) are shown in Figure 2.1 and discussed in more detail in the following.

- **PCRF.** It is responsible for policy control decision-making, as well as for controlling the flow-based charging functionalities in the Policy Control Enforcement Function (PCEF) which resides in the P-GW. The PCRF provides the QoS authorization (QoS class identifier and bitrates) that decides how a certain data flow will be treated in the PCEF and ensures that this is in accordance with the user's subscription profile.
- **Home Location Register (HLR).** The HLR contains users' SAE subscription data such as the EPS-subscribed QoS profile and any access restrictions for roaming (see Section 2.2.3). It also holds information about the PDNs to which the user can connect. This could be in the form of an Access Point Name (APN) (which is a label according to DNS[1] naming conventions describing the access point to the PDN), or a PDN Address (indicating subscribed IP address(es)). In addition the HLR holds dynamic information such as the identity of the MME to which the user is currently attached

---

[1] Domain Name System.

or registered. The HLR may also integrate the Authentication Centre (AuC) which generates the vectors for authentication and security keys.

- **P-GW.** The P-GW is responsible for IP address allocation for the UE, as well as QoS enforcement and flow-based charging according to rules from the PCRF. The P-GW is responsible for the filtering of downlink user IP packets into the different QoS based bearers. This is performed based on Traffic Flow Templates (TFTs) (see Section 2.4). The P-GW performs QoS enforcement for Guaranteed Bit Rate (GBR) bearers. It also serves as the mobility anchor for inter-working with non-3GPP technologies such as CDMA2000 and WiMAX networks (see Section 2.2.4 and Chapter 13 for more information about mobility).
- **S-GW.** All user IP packets are transferred through the S-GW, which serves as the local mobility anchor for the data bearers when the UE moves between eNodeBs. It also retains the information about the bearers when the UE is in idle state (known as ECM-IDLE, see Section 2.2.1.1) and temporarily buffers downlink data while the MME initiates paging of the UE to re-establish the bearers. In addition, the S-GW performs some administrative functions in the visited network such as collecting information for charging (e.g. the volume of data sent to or received from the user), and legal interception. It also serves as the mobility anchor for inter-working with other 3GPP technologies such as GPRS and UMTS (see Section 2.2.4 and Chapter 13 for more information about mobility).
- **MME.** The MME is the control node which processes the signalling between the UE and the CN. The protocols running between the UE and the CN are known as the *Non-Access Stratum* (NAS) protocols.

  The main functions supported by the MME are classified as:

  Functions related to bearer management. This includes the establishment, maintenance and release of the bearers, and is handled by the session management layer in the NAS protocol.

  Functions related to connection management. This includes the establishment of the connection and security between the network and UE, and is handled by the connection or mobility management layer in the NAS protocol layer.

  NAS control procedures are specified in [1] and are discussed in more detail in the following section.

2.2.1.1 Non-Access Stratum (NAS) Procedures

The NAS procedures, especially the connection management procedures, are fundamentally similar to UMTS. The main change from UMTS is that EPS allows concatenation of some procedures to allow faster establishment of the connection and the bearers.

The MME creates a *UE context* when a UE is turned on and attaches to the network. It assigns a unique short temporary identity termed the SAE-Temporary Mobile Subscriber Identity (S-TMSI) to the UE which identifies the UE context in the MME. This UE context holds user subscription information downloaded from the HSS. The local storage of subscription data in the MME allows faster execution of procedures such as bearer establishment since it removes the need to consult the HSS every time. In addition, the UE context also holds dynamic information such as the list of bearers that are established and the terminal capabilities.

To reduce the overhead in the E-UTRAN and processing in the UE, all UE-related information in the access network can be released during long periods of data inactivity. This state is called EPS Connection Management IDLE (ECM-IDLE). The MME retains the UE context and the information about the established bearers during these idle periods.

To allow the network to contact an ECM-IDLE UE, the UE updates the network as to its new location whenever it moves out of its current Tracking Area (TA); this procedure is called a 'Tracking Area Update'. The MME is responsible for keeping track of the user location while the UE is in ECM-IDLE.

When there is a need to deliver downlink data to an ECM-IDLE UE, the MME sends a paging message to all the eNodeBs in its current TA, and the eNodeBs page the UE over the radio interface. On receipt of a paging message, the UE performs a service request procedure which results in moving the UE to ECM-CONNECTED state. UE-related information is thereby created in the E-UTRAN, and the bearers are re-established. The MME is responsible for the re-establishment of the radio bearers and updating the UE context in the eNodeB. This transition between the UE states is called an idle-to-active transition. To speed up the idle-to-active transition and bearer establishment, EPS supports concatenation of the NAS and AS procedures for bearer activation (see also Section 2.4.1). Some inter-relationship between the NAS and AS protocols is intentionally used to allow procedures to run simultaneously rather than sequentially, as in UMTS. For example, the bearer establishment procedure can be executed by the network without waiting for the completion of the security procedure.

Security functions are the responsibility of the MME for both signalling and user data. When a UE attaches with the network, a mutual authentication of the UE and the network is performed between the UE and the MME/HSS. This authentication function also establishes the security keys which are used for encryption of the bearers, as explained in Section 3.2.3.1. The security architecture for SAE is specified in [2].

2.2.2 The Access Network

The Access Network of LTE, E-UTRAN, simply consists of a network of eNodeBs, as illustrated in Figure 2.3. For normal user traffic (as opposed to broadcast), there is no centralized controller in E-UTRAN; hence the E-UTRAN architecture is said to be flat.

The eNodeBs are normally inter-connected with each other by means of an interface known as *X2*, and to the EPC by means of the *S1* interface – more specifically, to the MME by means of the S1-MME interface and to the S-GW by means of the S1-U interface.

The protocols which run between the eNodeBs and the UE are known as the *Access Stratum* (AS) protocols.

The E-UTRAN is responsible for all radio-related functions, which can be summarized briefly as:

- **Radio Resource Management.** This covers all functions related to the radio bearers, such as radio bearer control, radio admission control, radio mobility control, scheduling and dynamic allocation of resources to UEs in both uplink and downlink (see Chapter 13).

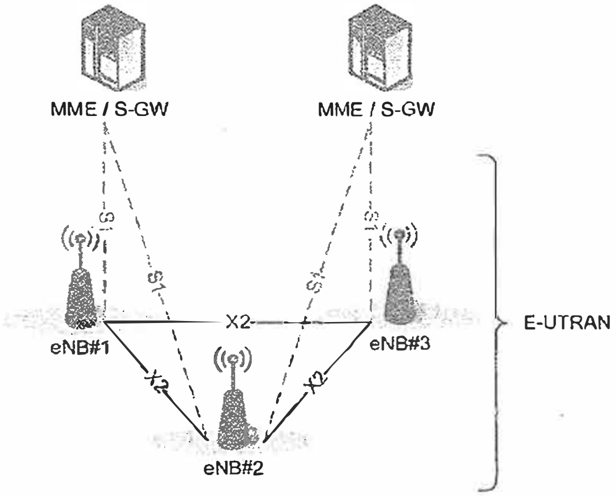


Figure 2.3 Overall E-UTRAN architecture. Reproduced by permission of © 3GPP.

- **Header Compression.** This helps to ensure efficient use of the radio interface by compressing the IP packet headers which could otherwise represent a significant overhead, especially for small packets such as VoIP (see Section 4.2.2).
- **Security.** All data sent over the radio interface is encrypted (see Sections 3.2.3.1 and 4.2.3).
- **Connectivity to the EPC.** This consists of the signalling towards the MME and the bearer path towards the S-GW.

On the network side, all of these functions reside in the eNodeBs, each of which can be responsible for managing multiple cells. Unlike some of the previous second- and third-generation technologies, LTE integrates the radio controller function into the eNodeB. This allows tight interaction between the different protocol layers of the radio access network, thus reducing latency and improving efficiency. Such distributed control eliminates the need for a high-availability, processing-intensive controller, which in turn has the potential to reduce costs and avoid 'single points of failure'. Furthermore, as LTE does not support soft handover there is no need for a centralized data-combining function in the network.

One consequence of the lack of a centralized controller node is that, as the UE moves, the network must transfer all information related to a UE, i.e. the UE context, together with any buffered data, from one eNodeB to another. As discussed in Section 2.3.1.1, mechanisms are therefore needed to avoid data loss during handover. The operation of the X2 interface for this purpose is explained in more detail in Section 2.6.

An important feature of the S1 interface linking the Access Network to the CN is known as *S1-flex*. This is a concept whereby multiple CN nodes (MME/S-GWs) can serve a common geographical area, being connected by a mesh network to the set of eNodeBs in that area (see Section 2.5). An eNodeB may thus be served by multiple MME/S-GWs, as is the case for

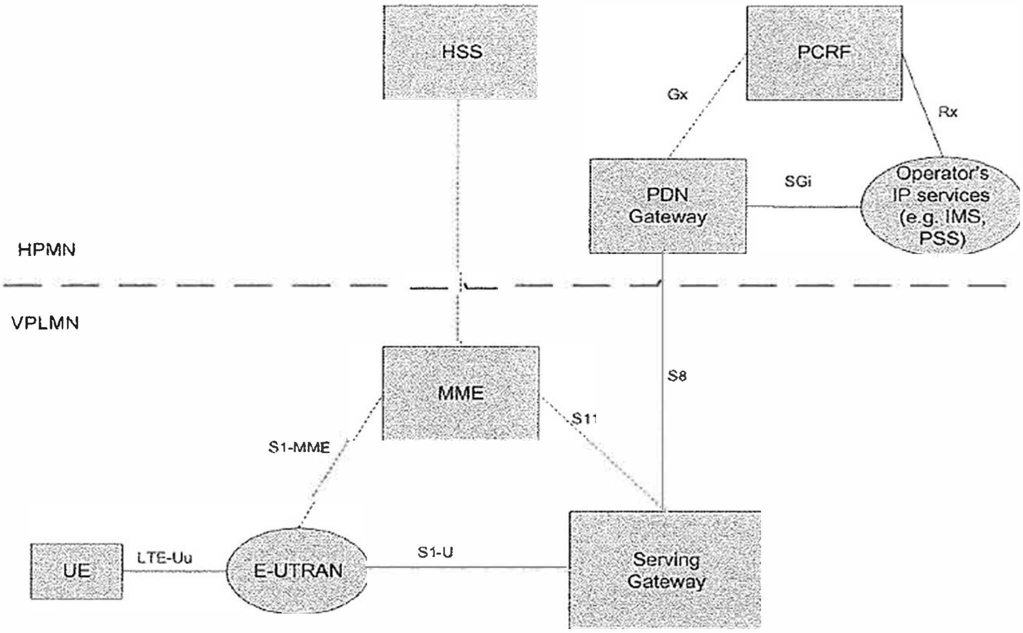


Figure 2.4 Roaming architecture for 3GPP accesses with P-GW in home network.

eNodeB#2 in Figure 2.3. The set of MME/S-GW nodes which serves a common area is called an *MME/S-GW pool*, and the area covered by such a pool of MME/S-GWs is called a *pool area*. This concept allows UEs in the cell(s) controlled by one eNodeB to be shared between multiple CN nodes, thereby providing a possibility for load sharing and also eliminating single points of failure for the CN nodes. The UE context normally remains with the same MME as long as the UE is located within the pool area.

2.2.3 Roaming Architecture

A network run by one operator in one country is known as a Public Land Mobile Network (PLMN). Roaming, where users are allowed to connect to PLMNs other than those to which they are directly subscribed, is a powerful feature for mobile networks, and LTE/SAE is no exception. A roaming user is connected to the E-UTRAN, MME and S-GW of the visited LTE network. However, LTE/SAE allows the P-GW of either the visited or the home network to be used, as shown in Figure 2.4. Using the home network's P-GW allows the user to access the home operator's services even while in a visited network. A P-GW in the visited network allows a 'local breakout' to the Internet in the visited network.

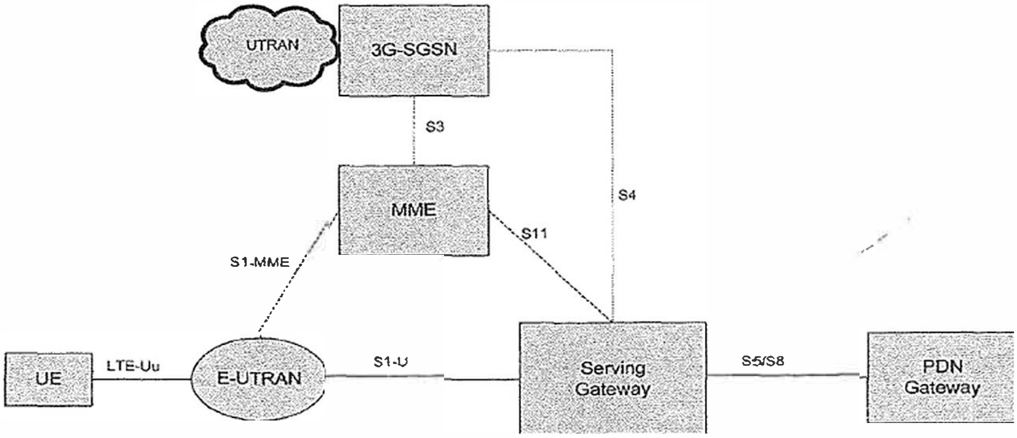


Figure 2.5 Architecture for 3G UMTS interworking.

2.2.4 Inter-Working with other Networks

EPS also supports inter-working and mobility (handover) with networks using other Radio Access Technologies (RATs), notably GSM, UMTS, CDMA2000 and WiMAX. The architecture for inter-working with 2G and 3G GPRS/UMTS networks is shown in Figure 2.5. The S-GW acts as the mobility anchor for inter-working with other 3GPP technologies such as GSM and UMTS, while the P-GW serves as an anchor allowing seamless mobility to non-3GPP networks such as CDMA2000 or WiMAX. The P-GW may also support a Proxy Mobile Internet Protocol (PMIP) based interface. More details of the radio interface procedures for inter-working are specified in [3] and are also covered in Sections 2.5.6.2 and 3.2.4.

2.3 Protocol Architecture

We outline here the radio protocol architecture of E-UTRAN.

2.3.1 User Plane

An IP packet for a UE is encapsulated in an EPC-specific protocol and tunnelled between the P-GW and the eNodeB for transmission to the UE. Different tunnelling protocols are used across different interfaces. A 3GPP-specific tunnelling protocol called the GPRS Tunnelling Protocol (GTP) [4] is used over the core network interfaces, S1 and S5/S8.[2]

The E-UTRAN user plane protocol stack is shown greyed in Figure 2.6, consisting of the PDCP (Packet Data Convergence Protocol), RLC (Radio Link Control) and MAC

---

[2] SAE also provides an option to use PMIP on S5/S8. More details on the MIP-based S5/S8 interface can be found in [3].

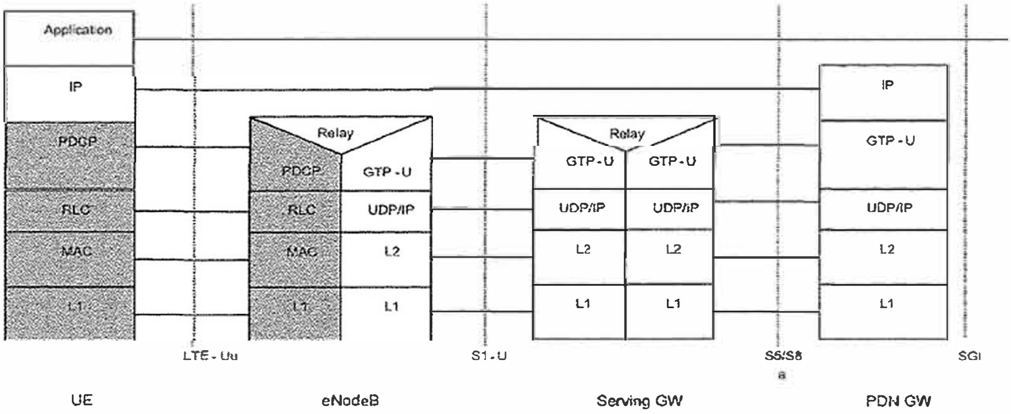


Figure 2.6 The E-UTRAN user plane protocol stack. Reproduced by permission of © 3GPP.

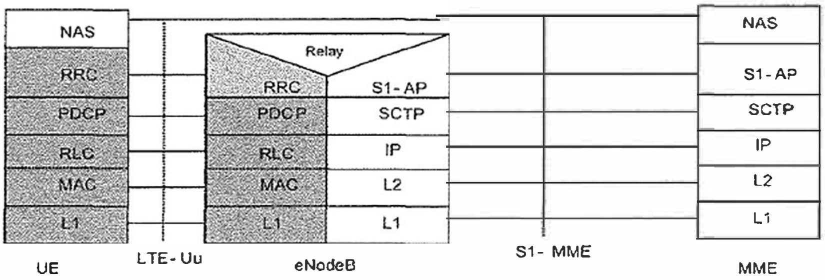


Figure 2.7 Control-plane protocol stack. Reproduced by permission of © 3GPP.

(Medium Access Control) sublayers which are terminated in the eNodeB on the network side. The respective roles of each of these layers are explained in detail in Chapter 4.

2.3.1.1 Data Handling During Handover

In the absence of any centralized controller node, data buffering during handover due to user mobility in the E-UTRAN must be performed in the eNodeB itself. Data protection during handover is a responsibility of the PDCP layer and is explained in detail in Section 4.2.4.

The RLC and MAC layers both start afresh in a new cell after handover.

2.3.2 Control Plane

The protocol stack for the control plane between the UE and MME is shown in Figure 2.7.

The greyed region of the stack indicates the access stratum protocols. The lower layers perform the same functions as for the user plane with the exception that there is no header compression function for control plane.

The RRC protocol is known as 'Layer 3' in the access stratum protocol stack. It is the main controlling function in the access stratum, being responsible for establishing the radio bearers and configuring all the lower layers using RRC signalling between the eNodeB and the UE. These functions are detailed in Section 3.2.

2.4 Quality of Service and EPS Bearers

In a typical case, multiple applications may be running in a UE at any time, each one having different QoS requirements. For example, a UE can be engaged in a VoIP call while at the same time browsing a web page or downloading an FTP file. VoIP has more stringent requirements for QoS in terms of delay and delay jitter than web browsing and FTP, while the latter requires a much lower packet loss rate. In order to support multiple QoS requirements, different bearers are set up within EPS, each being associated with a QoS.

Broadly, bearers can be classified into two categories based on the nature of the QoS they provide:

- **Minimum Guaranteed Bit Rate (GBR) bearers** which can be used for applications such as VoIP. These have an associated GBR value for which dedicated transmission resources are permanently allocated (e.g. by an admission control function in the eNodeB) at bearer establishment/modification. Bit rates higher than the GBR may be allowed for a GBR bearer if resources are available. In such cases, a Maximum Bit Rate (MBR) parameter, which can also be associated with a GBR bearer, sets an upper limit on the bit rate which can be expected from a GBR bearer.

- **Non-GBR bearers** which do not guarantee any particular bit rate. These can be used for applications such as web browsing or FTP transfer. For these bearers, no bandwidth resources are allocated permanently to the bearer.

In the access network, it is the responsibility of the eNodeB to ensure the necessary QoS for a bearer over the radio interface. Each bearer has an associated QoS Class Identifier (QCI), and an Allocation and Retention Priority (ARP).

Each QCI is characterized by priority, packet delay budget and acceptable packet loss rate. The QCI label for a bearer determines how it is handled in the eNodeB. Only a dozen such QCIs have been standardized so that vendors can all have the same understanding of the underlying service characteristics and thus provide the corresponding treatment, including queue management, conditioning and policing strategy. This ensures that an LTE operator can expect uniform traffic handling behaviour throughout the network regardless of the manufacturers of the eNodeB equipment. The set of standardized QCIs and their characteristics (from which the PCRF in an EPS can select) is provided in Table 2.1 (from Section 6.1.7, in [5]). The QCI table specifies values for the priority handling, acceptable delay budget and packet error loss rate for each QCI label.

The priority and packet delay budget (and to some extent the acceptable packet loss rate) from the QCI label determine the RLC mode configuration (see Section 4.3.1), and how the scheduler in the MAC (Section 4.4.2.1) handles packets sent over the bearer (e.g. in terms of

Table 2.1 Standardized QoS Class Identifiers (QCIs) for LTE.

QCI	Resource type	Priority	Packet delay budget (ms)	Packet error loss rate	Example services
1	GBR	2	100	10^{-2}	Conversational voice
2	GBR	4	150	10^{-3}	Conversational video (live streaming)
3	GBR	5	300	10^{-6}	Non-conversational video (buffered streaming)
4	GBR	3	50	10^{-3}	Real time gaming
5	Non-GBR	1	100	10^{-6}	IMS signalling
6	Non-GBR	7	100	10^{-3}	Voice, video (live streaming), interactive gaming
7	Non-GBR	6	300	10^{-6}	Video (buffered streaming)
8	Non-GBR	8	300	10^{-6}	TCP-based (e.g. WWW, e-mail) chat, FTP, p2p file sharing, progressive video, etc.
9	Non-GBR	9	300	10^{-6}	

scheduling policy, queue management policy and rate shaping policy). For example, a packet with a higher priority can be expected to be scheduled before a packet with lower priority. For bearers with a low acceptable loss rate, an Acknowledged Mode (AM) can be used within the RLC protocol layer to ensure that packets are delivered successfully across the radio interface (see Section 4.3.1.3).

The ARP of a bearer is used for call admission control – i.e. to decide whether or not the requested bearer should be established in case of radio congestion. It also governs the prioritization of the bearer for pre-emption with respect to a new bearer establishment request. Once successfully established, a bearer's ARP does not have any impact on the bearer-level packet forwarding treatment (e.g. for scheduling and rate control). Such packet forwarding treatment should be solely determined by the other bearer level QoS parameters such as QCI, GBR and MBR.

An EPS bearer has to cross multiple interfaces as shown in Figure 2.8 – the S5/S8 interface from the P-GW to the S-GW, the S1 interface from the S-GW to the eNodeB, and the radio interface (also known as the LTE-Uu interface) from the eNodeB to the UE. Across each interface, the EPS bearer is mapped onto a lower layer bearer, each with its own bearer identity. Each node must keep track of the binding between the bearer IDs across its different interfaces.

An S5/S8 bearer transports the packets of an EPS bearer between a P-GW and a S-GW. The S-GW stores a one-to-one mapping between an S1 bearer and an S5/S8 bearer. The bearer is identified by the GTP tunnel ID across both interfaces.

An S1 bearer transports the packets of an EPS bearer between a S-GW and an eNodeB. A radio bearer [6] transports the packets of an EPS bearer between a UE and an eNodeB.

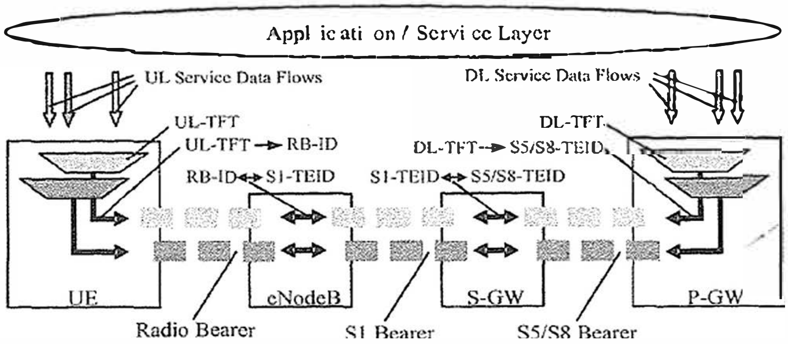


Figure 2.8 LTE/SAE bearers across the different interfaces. Reproduced by permission of © 3GPP.

An eNodeB stores a one-to-one mapping between a radio bearer ID and an S1 bearer to create the mapping between the two.

IP packets mapped to the same EPS bearer receive the same bearer-level packet forwarding treatment (e.g. scheduling policy, queue management policy, rate shaping policy, RLC configuration). Providing different bearer-level QoS thus requires that a separate EPS bearer is established for each QoS flow, and user IP packets must be filtered into the different EPS bearers.

Packet filtering into different bearers is based on Traffic Flow Templates (TFTs). The TFTs use IP header information such as source and destination IP addresses and Transmission Control Protocol (TCP) port numbers to filter packets such as VoIP from web browsing traffic so that each can be sent down the respective bearers with appropriate QoS. An UpLink TFT (UL TFT) associated with each bearer in the UE filters IP packets to EPS bearers in the uplink direction. A DownLink TFT (DL TFT) in the P-GW is a similar set of downlink packet filters.

As part of the procedure by which a UE attaches to the network, the UE is assigned an IP address by the P-GW and at least one bearer is established. This is called the default bearer, and it remains established throughout the lifetime of the PDN connection in order to provide the UE with always-on IP connectivity to that PDN. The initial bearer-level QoS parameter values of the default bearer are assigned by the MME, based on subscription data retrieved from the HSS. The PCEF may change these values in interaction with the PCRF or according to local configuration. Additional bearers called dedicated bearers can also be established at any time during or after completion of the attach procedure. A dedicated bearer can be either a GBR or a non-GBR bearer, (the default bearer always has to be a non-GBR bearer since it is permanently established). The distinction between default and dedicated bearers should be transparent to the access network (e.g. E-UTRAN). Each bearer has an associated QoS, and if more than one bearer is established for a given UE, then each bearer must also be associated with appropriate TFTs. These dedicated bearers could be established by the network, based for example on a trigger from the IMS domain, or they could be requested by the UE. The dedicated bearers for a UE may be provided by one or more P-GWs.

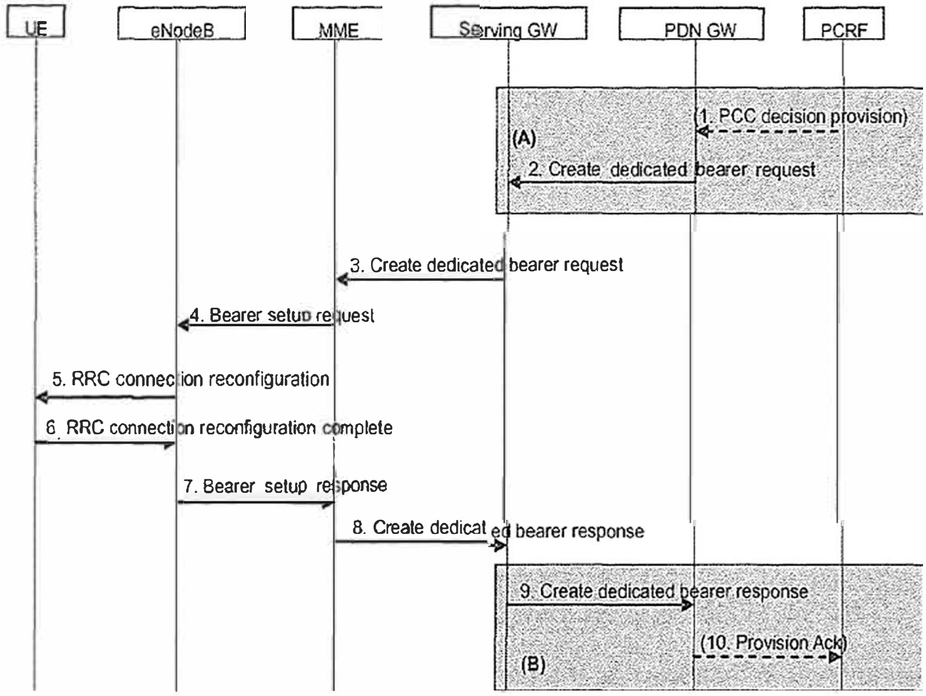


Figure 2.9 An example message flow for a LTE/SAE bearer establishment. Reproduced by permission of © 3GPP.

The bearer-level QoS parameter values for dedicated bearers are received by the P-GW from the PCRF and forwarded to the S-GW. The MME only transparently forwards those values received from the S-GW over the S11 reference point to the E-UTRAN.

2.4.1 Bearer Establishment Procedure

This section describes an example of the end-to-end bearer establishment procedure across the network nodes using the functionality described in the above sections.

A typical bearer establishment flow is shown in Figure 2.9. Each of the messages is described below.

When a bearer is established, the bearers across each of the interfaces discussed above are established.

The PCRF sends a 'PCC[3] Decision Provision' message indicating the required QoS for the bearer to the P-GW. The P-GW uses this QoS policy to assign the bearer-level QoS parameters. The P-GW then sends a 'Create Dedicated Bearer Request' message including the QoS and UL TFT to be used in the UE to the S-GW.

---

[3] Policy Control and Charging.

The S-GW forwards the Create Dedicated Bearer Request message (including bearer QoS, UL TFT and S1-bearer ID) to the MME (message 3 in Figure 2.9).

The MME then builds a set of session management configuration information including the UL TFT and the EPS bearer identity, and includes it in the 'Bearer Setup Request' message which it sends to the eNodeB (message 4 in Figure 2.9). The session management configuration is NAS information and is therefore sent transparently by the eNodeB to the UE.

The Bearer Setup Request also provides the QoS of the bearer to the eNodeB; this information is used by the eNodeB for call admission control and also to ensure the necessary QoS by appropriate scheduling of the user's IP packets. The eNodeB maps the EPS bearer QoS to the radio bearer QoS. It then signals a 'RRC Connection Reconfiguration' message (including the radio bearer QoS, session management configuration and EPS radio bearer identity) to the UE to set up the radio bearer (message 5 in Figure 2.9). The RRC Connection Reconfiguration message contains all the configuration parameters for the radio interface. This is mainly for the configuration of the Layer 2 (the PDCP, RLC and MAC parameters), but also the Layer 1 parameters required for the UE to initialize the protocol stack.

Messages 6 to 10 are the corresponding response messages to confirm that the bearers have been set up correctly.

2.5 The E-UTRAN Network Interfaces: S1 Interface

The provision of Self-Optimizing Networks (SONs) is one of the key objectives of LTE. Indeed, self-optimization of the network is a high priority for network operators, as a tool to derive the best performance from the network in a cost-effective manner, especially in changing radio propagation environments. Therefore SON has been placed as a cornerstone from the beginning around which all X2 and S1 procedures have been designed.

The S1 interface connects the eNodeB to the EPC. It is split into two interfaces, one for the control plane and the other for the user plane. The protocol structure for the S1 and the functionality provided over S1 are discussed in more detail below.

2.5.1 Protocol Structure Over S1

The protocol structure over S1 is based on a full IP transport stack with no dependency on legacy SS7[4] network configuration as used in GSM or UMTS networks. This simplification provides one expected area of savings on operational expenditure when LTE networks are deployed.

2.5.1.1 Control Plane

Figure 2.10 shows the protocol structure of the S1 control plane which is based on the well-known Stream Control Transmission Protocol / IP (SCTP/IP) stack.

---

[4] Signalling System #7 (SS7) is a communications protocol defined by the International Telecommunication Union (ITU) Telecommunication Standardization Sector (ITU-T) with a main purpose of setting up and tearing down telephone calls. Other uses include Short Message Service (SMS), number translation, prepaid billing mechanisms, and many other services.

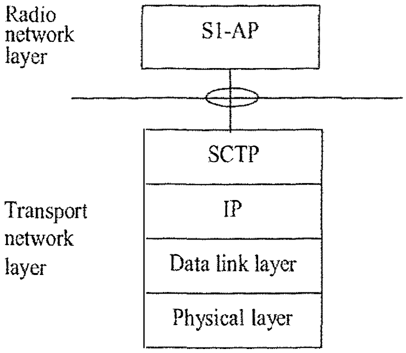


Figure 2.10 S1-MME control plane protocol stack. Reproduced by permission of © 3GPP.

The SCTP protocol is well known for its advanced features inherited from TCP which ensure the required reliable delivery of the signalling messages. In addition it makes it possible to benefit from improved features such as the handling of multistreams to implement transport network redundancy easily and avoid head-of-line blocking or multihoming (see 'IETF RFC4960' [7]).

A further simplification in LTE (compared to the UMTS Iu interface, for example) is the direct mapping of S1-AP (S1 Application Protocol) on top of SCTP. This results in a simplified protocol stack compared to UMTS with no intermediate connection management protocol. The individual connections are directly handled at the application layer. Multiplexing takes place between S1-AP and SCTP whereby each stream of an SCTP association is multiplexed with the signalling traffic of multiple individual connections.

One further area of flexibility brought with LTE lies in the lower layer protocols for which full optionality has been left regarding the choice of the IP version and the choice of the data link layer. For example, this enables the operator to start deployment using IP version 4 with the data link tailored to the network deployment scenario.

2.5.1.2 User Plane

Figure 2.11 gives the protocol structure of the S1 user plane, which is based on the GTP/UDP[5]/IP stack which is already well known from UMTS networks.

One of the advantages of using GTP-User plane (GTP-U) is its inherent facility to identify tunnels and also to facilitate intra-3GPP mobility.

The IP version number and the data link layer have been left fully optional, as for the control plane stack.

A transport bearer is identified by the GTP tunnel endpoints and the IP address (source Tunnelling End ID (TEID), destination TEID, source IP address, destination IP address).

The S-GW sends downlink packets of a given bearer to the eNodeB IP address (received in S1-AP) associated to that particular bearer. Similarly, the eNodeB sends upstream packets of a given bearer to the EPC IP address (received in S1-AP) associated to that particular bearer.

---

[5] User Datagram Protocol.

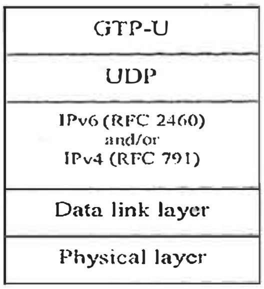


Figure 2.11 S1-U user plane protocol stack. Reproduced by permission of © 3GPP.

Vendor-specific traffic categories (e.g. real-time traffic) can be mapped onto Differentiated Services (Diffserv) code points (e.g. expedited forwarding) by network O&M (Operation and Maintenance) configuration to manage QoS differentiation between the bearers.

2.5.2 Initiation Over S1

The initialization of the S1-MME control plane interface starts with the identification of the MMEs to which the eNodeB must connect, followed by the setting up of the Transport Network Layer (TNL).

With the support of the S1-flex function in LTE, an eNodeB must initiate an S1 interface towards each MME node of the pool area to which it belongs. This list of MME nodes of the pool together with an initial corresponding remote IP address can be directly configured in the eNodeB at deployment (although other means may also be used). The eNodeB then initiates the TNL establishment with that IP address. Only one SCTP association is established between one eNodeB and one MME.

During the establishment of the SCTP association, the two nodes negotiate the maximum number of streams which will be used over that association. However, multiple pairs of streams (note that a stream is unidirectional and therefore pairs must be used) are typically used in order to avoid the head-of-line blocking issue mentioned above. Among these pairs of streams, one particular pair must be reserved by the two nodes for the signalling of the common procedures (i.e. those which are not specific to one UE). The other streams are used for the sole purpose of the dedicated procedures (i.e. those which are specific to one UE).

Once the TNL has been established, some basic application-level configuration data for the system operation is automatically exchanged between the eNodeB and the MME through an 'S1 SETUP' procedure initiated by the eNodeB. This procedure constitutes one example of a network self-configuration process provided in LTE to reduce the configuration effort for network operators compared to the more usual manual configuration procedures of earlier systems.

An example of such basic application data which can be configured automatically via the S1 SETUP procedure is the tracking area identities. These identities are very important for the system operation because the tracking areas correspond to the zones in which UEs are paged, and their mapping to eNodeBs must remain consistent between the E-UTRAN and the EPC. Thus, once all the tracking area identities which are to be broadcast over the radio interface have been configured within each and every eNodeB, they are sent automatically to all the relevant MME nodes of the pool area within the S1 SETUP REQUEST message of this procedure. The same applies for the broadcast list of PLMNs which is used in the case of a network being shared by several operators (each having its own PLMN ID which needs to be broadcast for the UEs to recognize it). This saves a significant amount of configuration effort in the core network, avoids the risk of human error, and ensures that the E-UTRAN and EPC configurations regarding tracking areas and PLMNs are aligned.

Once the S1 SETUP procedure has been completed, the S1 interface is operational.

2.5.3 Context Management Over S1

Within each pool area, a UE is associated to one particular MME for all its communications during its stay in this pool area. This creates a context in this MME for the UE. This particular MME is selected by the NAS Node Selection Function (NNSF) in the first eNodeB from which the UE entered the pool.

Whenever the UE becomes active (i.e. makes a transition from idle to active mode) under the coverage of a particular eNodeB in the pool area, the MME provides the UE context information to this eNodeB using the 'INITIAL CONTEXT SETUP REQUEST' message (see Figure 2.12). This enables the eNodeB in turn to create a context and manage the UE for the duration of its activity in active mode.

Even though the setup of bearers is otherwise relevant to a dedicated 'Bearer Management' procedure described below, the creation of the eNodeB context by the INITIAL CONTEXT SETUP procedure also includes the creation of one or several bearers including the default bearer.

At the next transition back to idle mode following a 'UE CONTEXT RELEASE' message sent from the MME, the eNodeB context is erased and only the MME context remains.

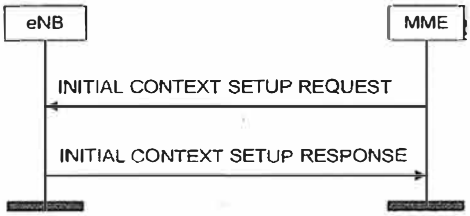


Figure 2.12 Initial context setup procedure. Reproduced by permission of © 3GPP.

2.5.4 Bearer Management Over S1

LTE uses independent dedicated procedures respectively covering the setup, modification and release of bearers. For each bearer requested to be set up, the transport layer address and the tunnel endpoint are provided to the eNodeB in the 'BEARER SETUP REQUEST' message to indicate the termination of the bearer in the S-GW where uplink user plane data must be sent. Conversely, the eNodeB indicates in the 'BEARER SETUP RESPONSE' message the termination of the bearer in the eNodeB where the downlink user plane data must be sent.

For each bearer, the QoS parameters (see Section 2.4 above) requested for the bearer are also indicated. Independently of the standardized QCI values, it is also still possible to use extra proprietary labels for the fast introduction of new services if vendors and operators agree upon them.

2.5.5 Paging Over S1

As mentioned in Section 2.5.3, in order to re-establish a connection towards a UE in idle mode, the MME distributes a paging request to the relevant eNodeBs based on the tracking areas where the UE is expected to be located. When receiving the 'PAGING REQUEST' message, the eNodeB sends a page over the radio interface in the cells which are contained within one of the tracking areas provided in that message.

The UE is normally paged using its SAE-Temporary Mobile Subscriber Identity (S-TMSI). The 'PAGING REQUEST' message also contains a UE identity index value in order for the eNodeB to calculate the paging occasions at which the UE will switch on its receiver to listen for paging messages (see Section 3.4).

2.5.6 Mobility Over S1

LTE/SAE supports mobility within LTE/SAE, and also mobility to other systems using both 3GPP specified and non-3GPP technologies. The mobility procedures over the radio interface are defined in Section 3.2. These mobility procedures also involve the network interfaces. The sections below discuss the procedures over S1 to support mobility. Mobility procedures from the point of view of the UE are outlined in Chapter 13.

2.5.6.1 Intra-LTE Mobility

There are two types of handover procedure in LTE for UEs in active mode: the S1-handover procedure and the X2-handover procedure.

For intra-LTE mobility, the X2-handover procedure is normally used for the inter-eNodeB handover (described in Section 2.6.3). However, when there is no X2 interface between the two eNodeBs, or if the source eNodeB has been configured to initiate handover towards a particular target eNodeB via the S1 interface, then an S1-handover will be triggered.

The S1-handover procedure has been designed in a very similar way to the UMTS Serving Radio Network Subsystem (SRNS) relocation procedure and is shown in Figure 2.13: it consists of a preparation phase involving the core network, where the resources are first prepared at the target side (steps 2 to 8), followed by an execution phase (steps 8 to 12) and a completion phase (after step 13).

Compared to UMTS, the main difference is the introduction of the 'STATUS TRANSFER' message sent by the source eNodeB (steps 10 and 11). This message has been added in order to carry some PDCP status information that is needed at the target eNodeB in cases when PDCP status preservation applies for the S1-handover (see Section 4.2.4); this is in alignment with the information which is sent within the X2 'STATUS TRANSFER' message used for the X2-handover (see below). As a result of this alignment, the handling of the handover by

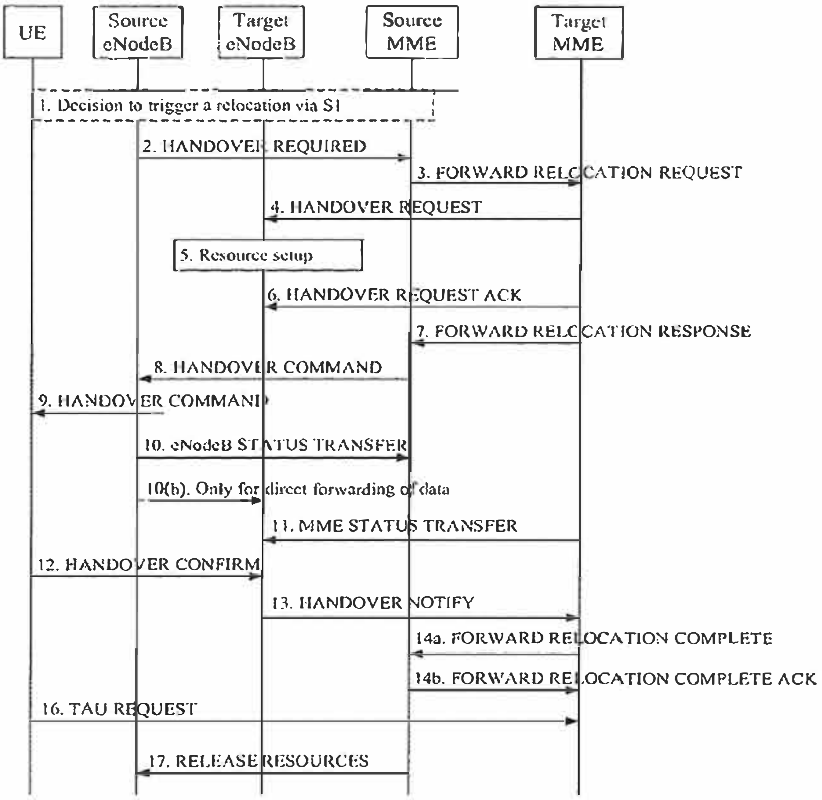


Figure 2.13 S1-based handover procedure. Reproduced by permission of © 3GPP.

the target eNodeB as seen from the UE is exactly the same, regardless of the type of handover (S1 or X2) the network had decided to use.

The Status Transfer procedure is assumed to be triggered in parallel with the start of data forwarding after the source eNodeB has received the 'HANDOVER COMMAND' message from the source MME. This data forwarding can be either direct or indirect, depending on the availability of a direct path for the user plane data between the source eNodeB and the target eNodeB.

The 'HANDOVER NOTIFY' message (step 13), which is sent later by the target eNodeB when the arrival of the UE at the target side is confirmed, is forwarded by the MME to trigger the update of the path switch in the S-GW towards the target eNodeB. In contrast to the X2-handover, the message is not acknowledged and the resources at the source side are released later upon reception of a 'RELEASE RESOURCE' message directly triggered from the source MME (step 17 in Figure 2.13).

2.5.6.2 Inter-Radio Access Technologies (RAT) Mobility

One key element of the design of the first release of LTE is the need to co-exist with other technologies.

For mobility from LTE towards UMTS, the handover process can reuse the S1-handover procedures described above, with the exception of the STATUS TRANSFER message which is not needed at steps 10 and 11 since no PDCP context is continued.

For mobility towards CDMA2000, dedicated uplink and downlink procedures have been introduced in LTE. They essentially aim at tunnelling the CDMA2000 signalling between the UE and the CDMA2000 system over the S1 interface, without being interpreted by the eNodeB on the way. The UPLINK S1 CDMA2000 TUNNELLING message presented in Figure 2.14 also includes the RAT type in order to identify which CDMA2000 RAT the tunnelled CDMA2000 message is associated with in order for the message to be routed to the correct node within the CDMA2000 system.

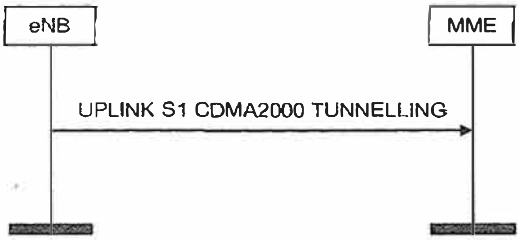


Figure 2.14 Uplink S1 CDMA2000 tunnelling procedure. Reproduced by permission of © 3GPP.

2.5.7 Load Management Over S1

Three types of load management procedures apply over S1: a normal 'load balancing' procedure to distribute the traffic, an 'overload' procedure to overcome a sudden peak in the loading and a 'load rebalancing' procedure to partially/fully offload an MME.

The MME load balancing procedure aims to distribute the traffic to the MMEs in the pool evenly according to their respective capacities. To achieve that goal, the procedure relies on the normal NNSF present in each eNodeB as part of the S1-flex function. Provided that suitable weight factors corresponding to the capacity of each MME node are available in the eNodeBs beforehand, a weighted NNSF done by each and every eNodeB in the network normally achieves a statistically balanced distribution of load among the MME nodes without further action. However, specific actions are still required for some particular scenarios:

- If a new MME node is introduced (or removed), it may be necessary temporarily to increase (or decrease) the weight factor normally corresponding to the capacity of this node in order to make it catch more (or less) traffic at the beginning until it reaches an adequate level of load.

- In case of an unexpected peak in the loading, an OVERLOAD message can be sent over the S1 interface by the overloaded MME. When received by an eNodeB, this message calls for a temporary restriction of a certain type of traffic. An MME can adjust the reduction of traffic it desires by defining the number of eNodeBs to which it sends the OVERLOAD message and by defining the types of traffic subject to restriction.
- Finally, if the MME wants to force rapidly the offload of part or all of its UEs, it will use the rebalancing function. This function forces the UEs to reattach to another MME by using a specific 'cause value' in the UE Release Command S1 message. In a first step it applies to idle mode UEs and in a second step it may also apply to UEs in connected mode (if the full MME offload is desired, e.g. for maintenance reasons).

2.6 The E-UTRAN Network Interfaces: X2 Interface

The X2 interface is used to inter-connect eNodeBs. The protocol structure for the X2 interface and the functionality provided over X2 are discussed below.

2.6.1 Protocol Structure Over X2

The control plane and user plane protocol stacks over the X2 interface are the same as over the S1 interface, as shown in Figures 2.15 and 2.16 respectively (with the exception that in Figure 2.15 the X2-AP is substituted for the S1-AP). This also means again that the choice of the IP version and the data link layer are fully optional. The use of the same protocol structure over both interfaces provides advantages such as simplifying the data forwarding operation.

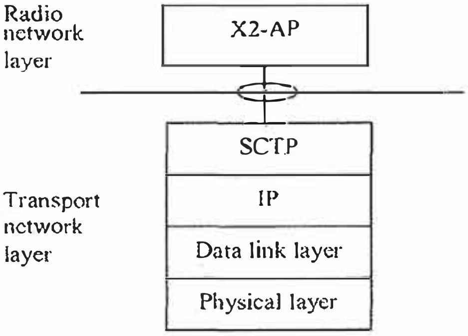


Figure 2.15 X2 signalling bearer protocol stack. Reproduced by permission of © 3GPP.

2.6.2 Initiation Over X2

The X2 interface may be established between one eNodeB and some of its neighbour eNodeBs in order to exchange signalling information when needed. However, a full mesh

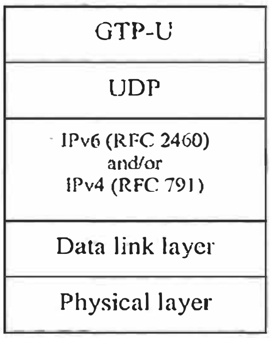


Figure 2.16 Transport network layer for data streams over X2. Reproduced by permission of © 3GPP.

is not mandated in an E-UTRAN network. Two types of information may typically need to be exchanged over X2 to drive the establishment of an X2 interface between two eNodeBs: load or interference related information (see Section 2.6.4) and handover related information (see mobility in Section 2.6.3).

Because these two types of information are fully independent of one another, it is possible that an X2 interface may be present between two eNodeBs for the purpose of exchanging load or interference information, even though the X2-handover procedure is not used to handover UEs between those eNodeBs. (In such a case, the S1-handover procedure is used instead.)

The initialization of the X2 interface starts with the identification of a suitable neighbour followed by the setting up of the TNL.

The identification of a suitable neighbour may be done by configuration, or alternatively a function known as the Automatic Neighbour Relation Function (ANRF) may be used. This function makes use of the UEs to identify the useful neighbour eNodeBs: an eNodeB may ask a UE to read the global cell identity from the broadcast information of another eNodeB for which the UE has identified the physical cell identity during the new cell identification procedure (see Section 7.2).

The ANRF is another example of a SON process introduced successfully in LTE. Through this self-optimizing process, UEs and eNodeB measurements are used to auto-tune the network.

Once a suitable neighbour has been identified, the initiating eNodeB can further set up the TNL using the transport layer address of this neighbour – either as retrieved from the network or locally configured.

Once the TNL has been set up, the initiating eNodeB must trigger the X2 setup procedure. This procedure enables an automatic exchange of application level configuration data relevant to the X2 interface, similar to the S1 setup procedure already described in Section 2.5.2. For example, each eNodeB reports within the X2 SETUP REQUEST message to a neighbour eNodeB information about each cell it manages, such as the cell's physical identity, the frequency band, the tracking area identity and/or the associated PLMNs.

This automatic data exchange in the X2 setup procedure is also the core of another SON feature: the automatic self-configuration of the Physical Cell Identities (PCIs). Under this new SON feature, the O&M system can provide the eNodeBs with either a list of possible

PCI values to use or a specific PCI value. In the first case, in order to avoid collisions, the eNodeB should use a PCI which is not already used in its neighbourhood. Because the PCI information is included in the LTE X2 setup procedure, while

detecting a neighbour cell by the ANR function an eNodeB can also discover all the PCI values used in the neighbourhood of that cell and consequently eliminate those values from the list of suitable PCIs to start with.

Once the X2 setup procedure has been completed, the X2 interface is operational.

2.6.3 Mobility Over X2

Handover via the X2 interface is triggered by default unless there is no X2 interface established or the source eNodeB is configured to use the S1-handover instead.

The X2-handover procedure is illustrated in Figure 2.17. Like the S1-handover, it is also composed of a preparation phase (steps 4 to 6), an execution phase (steps 7 to 9) and a completion phase (after step 9).

The key features of the X2-handover for intra-LTE handover are:

- The handover is directly performed between two eNodeBs. This makes the preparation phase quick.
- Data forwarding may be operated per bearer in order to minimize data loss.
- The MME is only informed at the end of the handover procedure once the handover is successful, in order to trigger the path switch.
- The release of resources at the source side is directly triggered from the target eNodeB.

For those bearers for which in-sequence delivery of packets is required, the STATUS TRANSFER message (step 8) provides the Sequence Number (SN) and the Hyper Frame Number (HFN) which the target eNodeB should assign to the first packet with no sequence number yet assigned that it must deliver. This first packet can either be one received over the target S1 path or one received over X2 if data forwarding over X2 is used (see below). When it sends the STATUS TRANSFER message, the source eNodeB freezes its transmitter/receiver status – i.e. it stops assigning PDCP SNs to downlink packets and stops delivering uplink packets to the EPC.

Mobility over X2 can be categorized according to its resilience to packet loss: the handover can be said 'seamless' if it minimizes the interruption time during the move of the UE, or 'lossless' if it tolerates no loss of packets at all. These two modes use data forwarding of user plane downlink packets. The source eNodeB may decide to operate one of these two modes on a per-EPS-bearer basis, based on the QoS received over S1 for this bearer (see Section 2.5.4) and the service at stake.

2.6.3.1 Seamless Handover

If the source eNodeB selects the seamless mode for one bearer, it proposes to the target eNodeB in the HANDOVER REQUEST message to establish a GTP tunnel to operate the downlink data forwarding. If the target eNodeB accepts, it indicates in the HANDOVER REQUEST ACK message the tunnel endpoint where the forwarded data is expected to

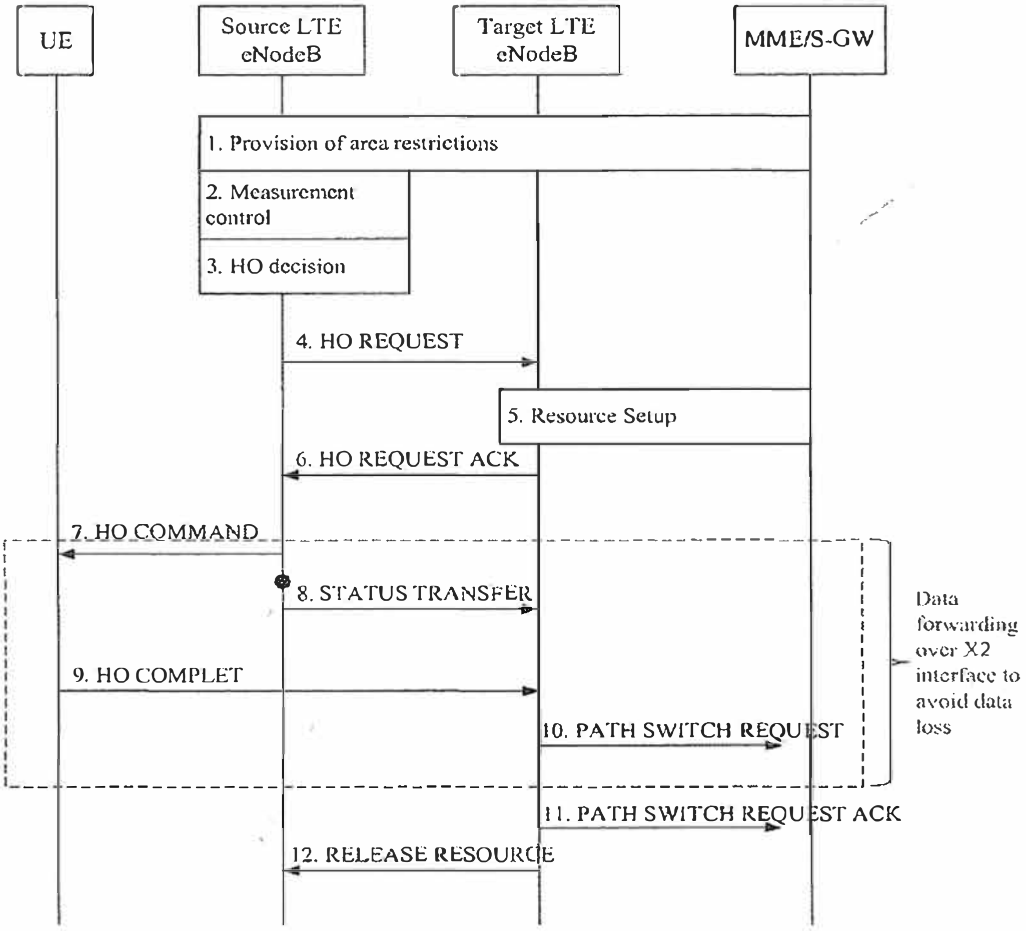


Figure 2.17 X2-based handover procedure.

be received. This tunnel endpoint may be different from the one set up as the termination point of the new bearer established over the target S1.

Upon reception of the HANDOVER REQUEST ACK message, the source eNodeB can start to forward the data freshly arriving over the source S1 path towards the indicated tunnel endpoint in parallel to sending the handover trigger to the UE over the radio interface. This forwarded data is thus available at the target eNodeB to be delivered to the UE as early as possible.

When forwarding is in operation and in-sequence delivery of packets is required, the target eNodeB is assumed to deliver first the packets forwarded over X2 before delivering the first ones received over the target S1 path once the S1 path switch has been done. The end of the forwarding is signalled over X2 to the target eNodeB by the reception of some 'special GTP packets' which the S-GW has inserted over the source S1 path just before switching this S1 path; these are then forwarded by the source eNodeB over X2 like any other regular packets.

2.6.3.2 Lossless Handover

If the source eNodeB selects the lossless mode for one bearer, it will additionally forward over X2 those user plane downlink packets which it has PDCP processed but are still buffered locally because they have not yet been delivered and acknowledged by the UE. These packets are forwarded together with their assigned PDCP SN included in a GTP extension header field. They are sent over X2 prior to the fresh arriving packets from the source S1 path. The same mechanisms described above for the seamless handover are used for the GTP tunnel establishment. The end of forwarding is also handled in the same way, since in-sequence packet delivery applies to lossless handovers. In addition, the target eNodeB must ensure that all the packets – including the ones received with sequence number over X2 – are delivered in-sequence at the target side. Further details of seamless and lossless handover are described in Section 4.2.

Selective retransmission. A new feature in LTE is the optimization of the radio by selective retransmission. When lossless handover is operated, the target eNodeB may, however, not deliver over the radio interface some of the forwarded downlink packets received over X2 if it is informed by the UE that those packets have already been received at the source side (see Section 4.2.6). This is called downlink selective retransmission.

Similarly in the uplink, the target eNodeB may desire that the UE does not retransmit packets already received earlier at the source side by the source eNodeB, for example to avoid wasting radio resources. To operate this uplink selective retransmission scheme for one bearer, it is necessary that the source eNodeB forwards to the target eNodeB, over another new GTP tunnel, those user plane uplink packets which it has received out of sequence. The target eNodeB must first request the source eNodeB to establish this new forwarding tunnel by including in the HANDOVER REQUEST ACK message a GTP tunnel endpoint where it expects the forwarded uplink packets to be received. The source eNodeB must, if possible, then indicate in the STATUS TRANSFER message for this bearer the list of SNs corresponding to the forwarded packets which are to be expected. This list helps the target eNodeB to inform the UE earlier of the packets not to be retransmitted, making the uplink selective retransmission overall scheme faster (see also Section 4.2.6).

2.6.3.3 Multiple Preparation

Another new feature of the LTE handover procedure is 'multiple preparation'. This feature enables the source eNodeB to trigger the handover preparation procedure towards multiple candidate target eNodeBs. Even though only one of the candidates is indicated as target to the UE, this makes recovery faster in case the UE fails on this target and connects to one of the other prepared candidate eNodeBs. The source eNodeB receives only one RELEASE RESOURCE message from the final selected eNodeB.

Regardless of whether multiple or single preparation is used, the handover can be cancelled during or after the preparation phase. If the multiple preparation feature is operated, it is recommended that upon reception of the RELEASE RESOURCE message the source eNodeB triggers a 'cancel' procedure towards each of the non-selected prepared eNodeBs.

2.6.4 Load and Interference Management Over X2

The exchange of load information between eNodeBs is of key importance in the flat architecture used in LTE, as there is no central RRM node as was the case, for example, in UMTS with the RNC.

The exchange of load information falls into two categories depending on the purpose it serves.

- The exchange of load information can serve for the (X2) load balancing process in which case the relevant frequency of exchange is rather low (in the order of seconds);
- The exchange of load information can serve to optimize some RRM processes such as interference coordination (as discussed in Section 12.5), in which case the frequency of exchange is rather high (in the order of tens of milliseconds).

2.6.4.1 Load Balancing

Like the ANRF SON function described in Section 2.6.2, load balancing is another aspect of SON built into the design of LTE. The objective of load balancing is to counteract local traffic load imbalance between neighbouring cells with the aim of improving the overall system capacity. One solution is to optimize the cell reselection/handover parameters (such as thresholds and hysteresis) between neighbouring cells autonomously upon detection of an imbalance (see also Section 13.6).

In order to detect an imbalance, it is necessary to compare the load of the cells and therefore to exchange information about them between neighbouring eNodeBs.

The cell load information exchanged can be of different types: radio measurements corresponding to the usage of physical resource blocks, possibly partitioned into real-time and non-real-time traffic; or generic measurements representing non-radio-related resource usage such as processing or hardware occupancy.

A client-server mechanism is used for the load information exchange: the RESOURCE STATUS RESPONSE/UPDATE message is used to report the load information over the X2 interface between one requesting eNodeB (client) and the eNodeBs which have subscribed to this request (servers). The reporting of the load is periodic and according to the periodicity expressed in the RESOURCE STATUS REQUEST message.

2.6.4.2 Interference Management

A separate Load Indication procedure is used over the X2 interface for the exchange of load information related to interference management as shown in Figure 2.18. As these measurements have direct influence on some RRM real-time processes, the frequency of reporting via this procedure may be high.

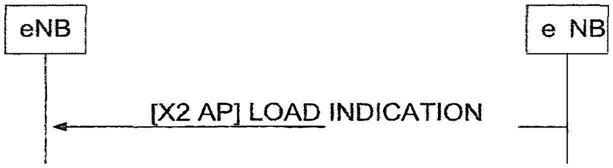


Figure 2.18 The LOAD INDICATION over X2 interface. Reproduced by permission of © 3GPP.

For the uplink interference, two indicators can be provided within the LOAD INDICATION message: a 'High Interference Indicator' and an 'Overload Indicator'. The usage of these indicators is explained in Section 12.5.

2.6.5 UE Historical Information Over X2

Historical UE information constitutes another example of a feature designed to support SON which is embedded in the design of LTE. It is part of the X2-handover procedure.

Historical UE information consists, for example, of the last few cells visited by the UE, together with the time spent in each one. This information is propagated from one eNodeB to another within the HANDOVER REQUEST messages and can be used to determine the occurrence of ping-pong between two or three cells for instance. The length of the history information can be configured for more flexibility.

More generally, the Historical UE information consists of some RRM information which is passed from the source eNodeB to the target eNodeB within the HANDOVER REQUEST message to assist the RRM management of a UE. The information can be partitioned into two types:

- UE RRM-related information, passed over X2 within the RRC transparent container;
- Cell RRM-related information, passed over X2 directly as an information element of the X2 AP HANDOVER REQUEST message itself.

2.7 Summary

The EPS provides UEs with IP connectivity to the packet data network. The EPS supports multiple data flows with different quality of service per UE for applications that need guaranteed delay and bit rate such as VoIP as well as best effort web browsing.

In this chapter we have seen an overview of the EPS network architecture, including the functionalities provided by the E-UTRAN access network and the evolved packet CN.

It can be seen that the concept of EPS bearers, together with their associated quality of service attributes, provide a powerful tool for the provision of a variety of simultaneous services to the end user.

From the perspective of the network operator, the LTE system is also breaking new ground in terms of its degree of support for self-optimization and self-configuration of the network via the X2, S1 and Uu interfaces, to facilitate deployment.

References[6]

[1] 3GPP Technical Specification 24.301, 'Non-Access-Stratum (NAS) protocol for Evolved Packet System (EPS); Stage 3 (Release 8)', www.3gpp.org.

[2] 3GPP Technical Specification 33.401, 'System Architecture Evolution (SAE): Security Architecture (Release 8)', www.3gpp.org.

[3] 3GPP Technical Specification 23.402, 'Architecture enhancements for non-3GPP accesses (Release 8)', www.3gpp.org.

[4] 3GPP Technical Specification 29.060, 'General Packet Radio Service (GPRS); GPRS Tunnelling Protocol (GTP) across the Gn and Gp interface (Release 8)', www.3gpp.org.

[5] 3GPP Technical Specification 23.203, 'Policy and charging control architecture (Release 8)', www.3gpp.org.

[6] 3GPP Technical Specification 36.300, 'Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2 (Release 8)', www.3gpp.org.

[7] Request for Comments 4960 The Internet Engineering Task Force (IETF), Network Working Group, 'Stream Control Transmission Protocol', http://www.ietf.org.

---

[6] All web sites confirmed 18th December 2008.

9

Downlink Physical Data and Control Channels

Matthew Baker and Tim Moulsley

9.1 Introduction

Chapters 7 and 8 have described the signals which enable User Equipment (UEs) to synchronize with the network and estimate the downlink radio channel in order to be able to demodulate data. In this chapter the downlink physical channels which transport the data are reviewed. This is followed by an explanation of the control-signalling channels which support the data channels by indicating the particular time-frequency transmission resources to which the data is mapped and the format in which the data itself is transmitted.

9.2 Downlink Data-Transporting Channels

9.2.1 Physical Broadcast Channel (PBCH)

In typical cellular systems the basic system information which allows the other channels in the cell to be configured and operated is carried by a Broadcast Channel (BCH). Therefore the achievable coverage for reception of the BCH is crucial to the successful operation of such cellular communication systems; LTE is no exception. As already noted in Chapter 3, the broadcast system information is divided into two categories:

LTE – The UMTS Long Term Evolution: from Theory to Practice Stefania Sesia, Issam Toufik and Matthew Baker

- The 'Master Information Block' (MIB), which consists of a limited number of the most frequently transmitted parameters essential for initial access to the cell,[1] is carried on the Physical Broadcast Channel (PBCH).
- The other System Information Blocks (SIBs) which, at the physical layer, are multiplexed together with unicast data transmitted on the Downlink Shared Channel as discussed in Section 9.2.2.2.

This section focuses in particular on the PBCH, which has some unique design requirements:

- Detectable without prior knowledge of system bandwidth;
- Low system overhead;
- Reliable reception right to the edge of the LTE cells;
- Decodable with low latency and low impact on UE battery life.

We review here the ways in which these requirements have influenced the design selected for the PBCH in LTE, the overall structure of which is shown in Figure 9.1.

Detectability without the UE having prior knowledge of the system bandwidth is achieved by mapping the PBCH only to the central 72 subcarriers of the OFDM signal (which corresponds to the minimum possible LTE system bandwidth), regardless of the actual system bandwidth. The UE will have first identified the system centre-frequency from the synchronization signals as described in Section 7.

Low system overhead for the PBCH is achieved by deliberately keeping the amount of information carried on the PBCH to a minimum, since achieving stringent coverage requirements for a large quantity of data would result in a high system overhead. The size of the MIB is therefore just 14 bits, and, since it is repeated every 40 ms, this corresponds to a data rate on the PBCH of just 350 bps.

The main mechanisms employed to facilitate reliable reception of the PBCH in LTE are time diversity, forward error correction coding and antenna diversity.

Time diversity is exploited by spreading out the transmission of each MIB on the PBCH over a period of 40 ms. This significantly reduces the likelihood of a whole MIB being lost in a fade in the radio propagation channel, even when the mobile terminal is moving at pedestrian speeds.

The forward error correction coding for the PBCH uses a convolutional coder, as the number of information bits to be coded is small; the details of the convolutional coder are explained in Section 10.3.3. The basic code rate is 1/3, after which a high degree of repetition of the systematic (i.e. information) bits and parity bits is used, such that each MIB is coded at a very low code-rate (1/48 over a 40 ms period) to give strong error protection.

Antenna diversity may be utilized at both the eNodeB and the UE. The UE performance requirements specified for LTE assume that all UEs can achieve a level of decoding performance commensurate with dual-antenna receive diversity (although it is recognized that in low-frequency deployments, such as below 1 GHz, the advantage obtained from

---

[1] The MIB information consists of the downlink system bandwidth, the PHICH structure (Physical Hybrid ARQ Indicator Channel, see Section 9.3.2.4), and the most-significant eight bits of the System Frame Number – the remaining two bits of the System Frame Number being gleaned from the 40 ms periodicity of the PBCH.

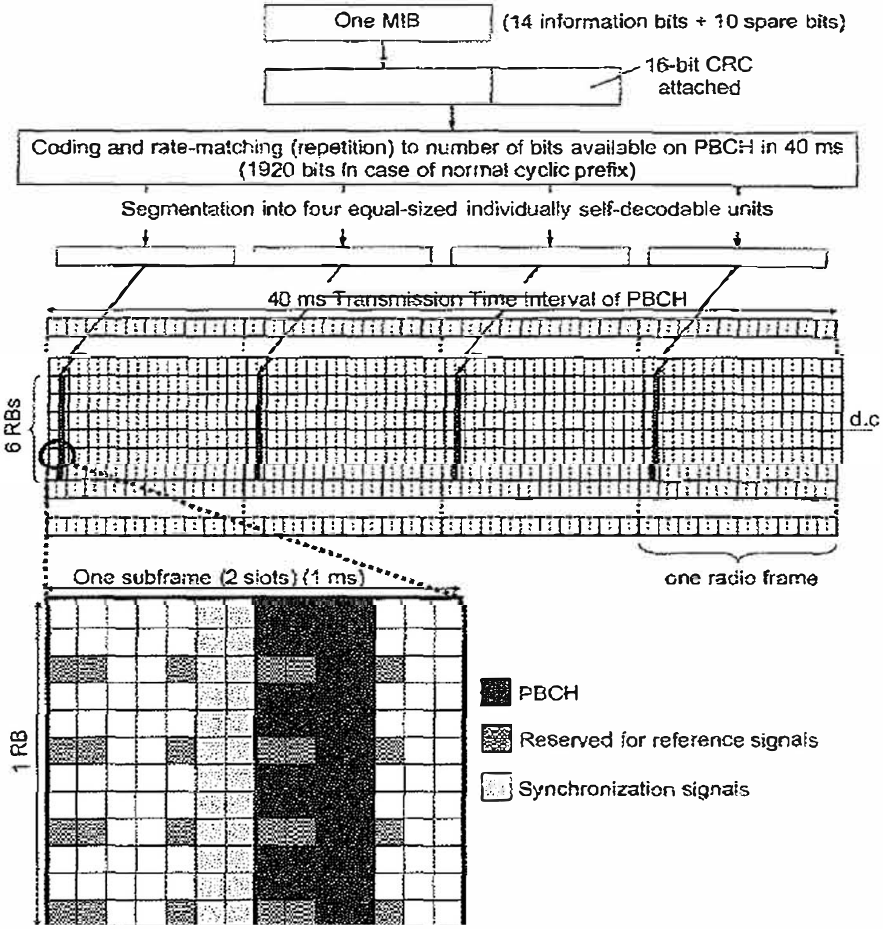


Figure 9.1 PBCH structure.

receive antenna diversity is reduced due to the correspondingly higher correlation between the antennas); this enables LTE system planners to rely on this level of performance being common to all UEs, thereby enabling wider cell coverage to be achieved with fewer cell sites than would otherwise be possible. Transmit antenna diversity may be also employed at the eNodeB to further improve coverage, depending on the capability of the eNodeB; eNodeBs with two or four transmit antenna ports transmit the PBCH using a Space-Frequency Block Code (SFBC), details of which are explained in Section 11.2.2.1.

The exact set of resource elements used by the PBCH is independent of the number of transmit antenna ports used by the eNodeB; any resource elements which may be used for reference signal transmission are avoided by the PBCH, irrespective of the actual number of transmit antenna ports deployed at the eNodeB. The number of transmit antenna ports used by the eNodeB must be determined blindly by the UE, by performing the decoding for each SFBC scheme corresponding to the different possible numbers of transmit antenna ports (namely one, two or four). This discovery of the number of transmit antenna ports is further facilitated by the fact that the Cyclic Redundancy Check (CRC) on each MIB is masked with a codeword representing the number of transmit antenna ports.

Finally, achieving low latency and a low impact on UE battery life is also facilitated by the design of the coding outlined above: the low code rate with repetition enables the full set of coded bits to be divided into four subsets, each of which is self-decodable in its own right. Each of these subsets of the coded bits is then transmitted in a different one of the four radio frames during the 40 ms transmission period, as shown in Figure 9.1. This means that if the Signal to Interference Ratio (SIR) of the radio channel is sufficiently good to allow the UE to decode the MIB correctly from the transmission in less than four radio frames, then the UE does not need to receive the other parts of the PBCH transmission in the remainder of the 40 ms period; on the other hand, if the SIR is low, the UE can receive further parts of the MIB transmission, soft-combining each part with those received already, until successful decoding is achieved.

The timing of the 40 ms transmission interval for each MIB on the PBCH is not indicated explicitly to the UE; this is determined implicitly from the scrambling and bit positions, which are re-initialized every 40 ms. The UE can therefore initially determine the 40 ms timing by performing four separate decodings of the PBCH using each of the four possible phases of the PBCH scrambling code, checking the CRC for each decoding.

When a UE initially attempts to access a cell by reading the PBCH, a variety of approaches may be taken to carry out the necessary blind decodings. A simple approach is always to perform the decoding using a soft combination of the PBCH over four radio frames, advancing a 40 ms sliding window one radio frame at a time until the window aligns with the 40 ms period of the PBCH and the decoding succeeds. However, this would result in a 40–70 ms delay before the PBCH can be decoded. A faster approach would be to attempt to decode the PBCH from the first single radio frame, which should be possible provided the SIR is sufficiently high; if the decoding fails for all four possible scrambling code phases, the PBCH from the first frame could be soft-combined with the PBCH bits received in the next frame – there is a 3-in-4 chance that the two frames contain data from the same transport block. If decoding still fails, a third radio frame could be combined, and failing that a fourth. It is evident that the latter approach may be much faster (potentially taking only 10 ms), but on the other hand requires slightly more complex logic.

9.2.2 Physical Downlink Shared Channel (PDSCH)

The Physical Downlink Shared Channel (PDSCH) is the main data-bearing downlink channel in LTE. It is used for all user data, as well as for broadcast system information which is not carried on the PBCH, and for paging messages – there is no specific physical layer paging channel in the LTE system. In this section, the use of the PDSCH for user data is explained; the use of the PDSCH for system information and paging is covered in the next section.

Data is transmitted on the PDSCH in units known as *transport blocks*, each of which corresponds to a MAC-layer Protocol Data Unit (PDU) as described in Section 4.4. Transport blocks may be passed down from the MAC layer to the physical layer once per Transmission Time Interval (TTI), where a TTI is 1 ms, corresponding to the subframe duration.

9.2.2.1 General Use of the PDSCH

When employed for user data, one or, at most, two transport blocks can be transmitted per UE per subframe, depending on the transmission mode selected for the PDSCH for each UE. The transmission mode configures the multi-antenna technique usually applied:

Transmission Mode 1: Transmission from a single eNodeB antenna port;

Transmission Mode 2: Transmit diversity (see Section 11.2.2.1);

Transmission Mode 3: Open-loop spatial multiplexing (see Section 11.2.2.3);

Transmission Mode 4: Closed-loop spatial multiplexing (see Section 11.2.2.3);

Transmission Mode 5: Multi-user Multiple-Input Multiple-Output (MIMO) (see Section 11.2.3);

Transmission Mode 6: Closed-loop rank-1 precoding (see Section 11.2.2.3);

Transmission Mode 7: Transmission using UE-specific reference signals (see Sections 11.2.2.2 and 8.2).

With the exception of transmission mode 7, the phase reference for demodulating the PDSCH is given by the cell-specific Reference Signals (RSs) described in Section 8.2, and the number of eNodeB antenna ports used for transmission of the PDSCH is the same as the number of antenna ports used in the cell for the PBCH. In transmission mode 7, UE-specific RSs (also covered in Section 8.2) provide the phase reference for the PDSCH. The configured transmission mode also affects the transmission of the associated downlink control signalling, as described in Section 9.3, and the channel quality feedback from the UE (see Section 10.2.1).

After channel coding (see Section 10.3.2) and mapping to spatial layers according to the selected transmission mode, the coded data bits are mapped to modulation symbols depending on the modulation scheme selected for the current radio channel conditions and required data rate. The modulation order may be varied between two bits per symbol (using QPSK (Quadrature Phase Shift Keying)) and six bits per symbol (using 64QAM (Quadrature Amplitude Modulation)). Support for reception of 64QAM modulation is mandatory for all classes of LTE UE and is designed to achieve the high peak downlink data rates that are required for LTE. The available modulation schemes are illustrated in Figure 9.2 by means of their constellation diagrams.

The resource elements used for the PDSCH can be any which are not reserved for other purposes (i.e. reference signals, synchronization signals, PBCH and control signalling). Thus when the control signalling informs a UE that a particular pair of resource blocks[2] in a subframe are allocated to that UE, it is only the *available* resource elements within those resource blocks which actually carry PDSCH data.

Normally the allocation of pairs of resource blocks to PDSCH transmission for a particular UE is signalled to the UE by means of dynamic control signalling transmitted at the start of

---

[2] The term 'pair of resource blocks' here means a pair of resource blocks which occupy the same set of 12 subcarriers and are contiguous in time, thus having a duration of one subframe.

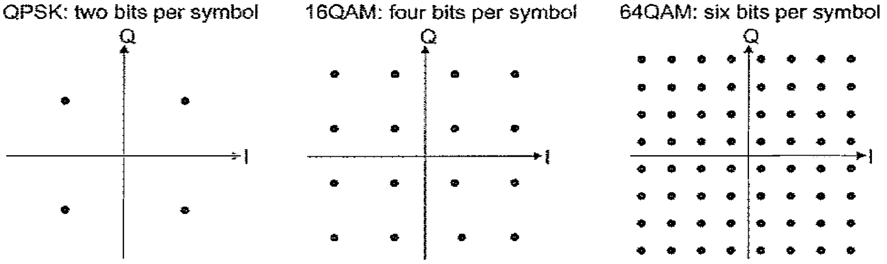


Figure 9.2 Constellations of modulation schemes applicable to PDSCH transmission.

the relevant subframe using the Physical Downlink Control Channel (PDCCH), as described in Section 9.3.

The mapping of data to physical resource blocks can be carried out in one of two ways: *localized mapping* and *distributed mapping*.

Localized resource mapping entails allocating all the available resource elements in a pair of resource blocks to the same UE. This is suitable for most scenarios, including the use of dynamic channel-dependent scheduling of resource blocks according to frequency-specific channel quality information reported by the UE (see Sections 10.2.1 and 12.4).

Distributed resource mapping entails separating in frequency the two physical resource blocks comprising each pair, as shown in Figure 9.3. This is a useful means of obtaining frequency diversity for small amounts of data which would otherwise be constrained to a narrow part of the downlink bandwidth and would therefore be more susceptible to narrow-band fading. An amount of data corresponding to up to two pairs of resource blocks may be transmitted to a UE in this way. An example of a typical use for this transmission mode could be a Voice-over-IP (VoIP) service, where, in order to minimize overhead, certain frequency resources may be 'persistently-scheduled' (see Section 4.4.2.1) – in other words, certain resource blocks in the frequency domain are allocated on a periodic basis to a specific UE by RRC signalling rather than by dynamic PDCCH signalling. This means that the resources allocated are not able to benefit from dynamic channel-dependent scheduling and therefore the frequency diversity which is achieved through distributed mapping is a useful tool to improve performance. Moreover, as the amount of data to be transmitted per UE for a VoIP service is small (typically sufficient to occupy only one or two pairs of resource blocks in a given subframe), the degree of frequency diversity obtainable via localized scheduling is very limited.

When data is mapped using the distributed mode, a frequency-hop occurs at the slot boundary in the middle of the subframe. This results in the block of data for a given UE being transmitted on one set of 12 subcarriers in the first half of the subframe and on a different set of 12 subcarriers in the second half of the subframe. This is illustrated in Figure 9.3.

The potential increase in the number of VoIP users which can be accommodated in a cell as a result of using distributed resource mapping as opposed to localized resource mapping is illustrated by way of example in Figure 9.4. In this example, the main simulation parameters are as given in Table 9.1.

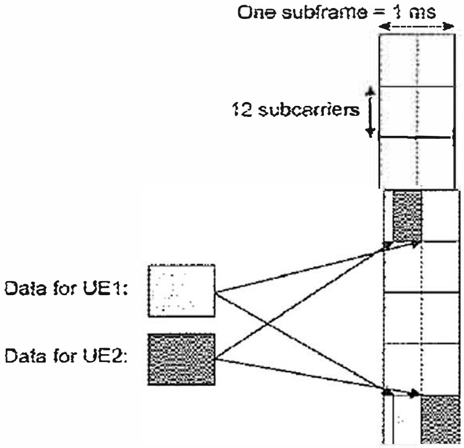


Figure 9.3 Frequency-distributed data mapping in LTE downlink.

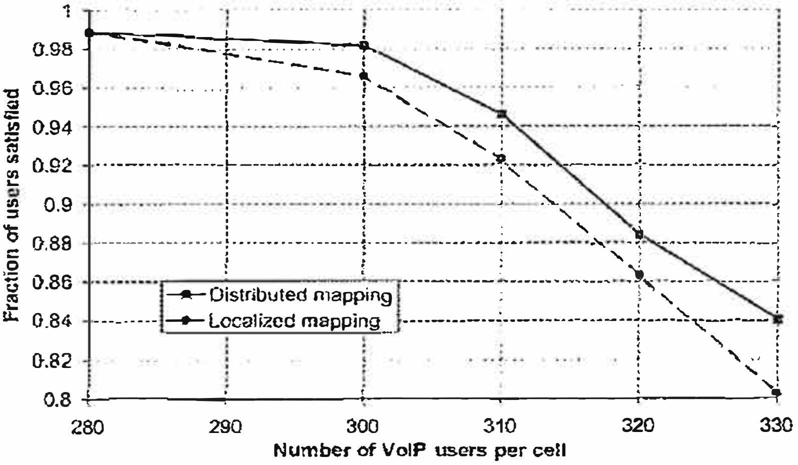


Figure 9.4 Example of increase in VoIP capacity arising from frequency-distributed resource mapping.

9.2.2.2 Special Uses of the PDSCH

As noted above, the PDSCH is used for some special purposes in addition to normal user data transmission.

One such use is sometimes referred to as the 'Dynamic BCH'. This consists of all the broadcast system information (i.e. SIBs) that is not carried on the PBCH. The resource blocks used for broadcast data of this sort are indicated by signalling messages on the

Table 9.1 Key simulation parameters for Figure 9.4.

Parameter	Value
Carrier frequency	2 GHz
Bandwidth	5 MHz
Channel model	Urban micro, 3 km/h
Total eNodeB transmit power	43 dBm
VoIP model	12.2 kbps; voice activity factor 0.43
Modulation and coding scheme	Fixed: QPSK, code rate 2/3
Satisfaction criterion	98% packets within 50 ms

PDCCH in the same way as for other PDSCH data, except that the identity indicated on the PDCCH is not the identity of a specific UE but is, rather, a designated broadcast identity known as the System Information Radio Network Temporary Identifier (SI-RNTI), which is fixed in the specifications (see Section 7.1 of [1]) and therefore known a priori to all UEs. Some constraints exist as to which subframes may be used for particular system information messages on the PDSCH; these are explained in Section 3.2.2.

Another special use of the PDSCH is for paging, as no separate physical channel is provided in LTE for this purpose. In previous systems such as WCDMA, a special 'Paging Indicator Channel' was provided, which was specially designed to enable the UE to wake up its receiver periodically for a very short period of time, in order to minimize the impact on battery life; on detecting a paging indicator (typically for a group of UEs), the UE would then keep its receiver switched on to receive a longer message indicating the exact identity of the UE being paged. By contrast, in LTE the PDCCH signalling is already very short in duration, and therefore the impact on UE battery life of monitoring the PDCCH from time to time is low. Therefore the normal PDCCH signalling can be used to carry the equivalent of a paging indicator, with the detailed paging information being carried on the PDSCH in a resource block indicated by the PDCCH. In a similar way to broadcast data, paging indicators on the PDCCH use a single fixed identifier, in this case the Paging RNTI (P-RNTI). Rather than providing different paging identifiers for different groups of UEs, different UEs monitor different subframes for their paging messages, as described in Section 3.4.

9.2.3 Physical Multicast Channel (PMCH)

Although Multimedia Broadcast and Multicast Services (MBMS) are not included in the first release of the LTE specifications, nonetheless the physical layer structure to support MBMS is defined ready for deployment in a later release. All UEs must be aware of the possible existence of MBMS transmissions at the physical layer, in order to enable such transmissions to be introduced later in a backward-compatible way.

The basic structure of the Physical Multicast Channel (PMCH) is very similar to the PDSCH. However, the PMCH is designed for 'single-frequency network' operation, whereby multiple cells transmit the same modulated symbols with very tight time-synchronization, ideally so that the signals from different cells are received within the duration of the cyclic prefix. This is known as MBSFN (MBMS Single Frequency Network) operation, and is discussed in more detail in Section 14.3. As the channel in MBSFN operation is in effect a composite channel from multiple cells, it is necessary for the UE to perform a separate channel estimate for MBSFN reception from that performed for reception of data from a single cell. Therefore, in order to avoid the need to mix normal reference symbols and reference symbols for MBSFN in the same subframe, frequency-division multiplexing of the PMCH and PDSCH is not permitted within a given subframe; instead, certain subframes may be specifically designated for MBSFN, and it is in these subframes that the PMCH would be transmitted.

The key differences from PDSCH in respect of the PMCH are as follows:

- The dynamic control signalling (PDCCH and PHICH – see Section 9.3) cannot occupy more than two OFDM symbols in an MBSFN subframe. The PDCCH is used only for uplink resource grants and not for the PMCH, as the scheduling of MBSFN data on the PMCH is carried out by higher-layer signalling.

- The pattern of reference symbols embedded in the PMCH is different from that in the PDSCH, as discussed in Chapter 8. (Note, however, that the common reference symbol pattern embedded in the OFDM symbols carrying control signalling at the start of each subframe remains the same as in the non-MBSFN subframes.)

- The extended cyclic prefix is always used. Note, however, that if the non-MBSFN subframes use the normal cyclic prefix, then the normal cyclic prefix is also used in the OFDM symbols used for the control signalling at the start of each MBSFN subframe. This results in there being some spare time samples whose usage is unspecified between the end of the last control signalling symbol and the first PMCH symbol, the PMCH remaining aligned with the end of the subframe; the eNodeB may transmit an undefined signal (e.g. a cyclic extension) during these time samples, or alternatively it may switch off its transmitter – the UE cannot assume anything about the transmitted signal during these samples.

The latter two features are designed so that a UE making measurements on a neighbouring cell does not need to know in advance the allocation of MBSFN and non-MBSFN subframes. The UE can take advantage of the fact that the first two OFDM symbols in all subframes use the same cyclic prefix duration and reference symbol pattern.

The exact pattern of MBSFN subframes in a cell is indicated in the system information carried on the part of the broadcast channel mapped to the PDSCH. The system information also indicates whether the pattern of MBSFN subframes in neighbouring cells is the same as or different from that in the current cell; however, if the pattern in the neighbouring cell is different, the UE can only ascertain the pattern by reading the system information of that cell.

Further details of multicast and broadcast operation in LTE are explained in Chapter 14.

9.3 Downlink Control Channels

9.3.1 Requirements for Control Channel Design

The control channels in LTE are provided in order to support efficient data transmission. In common with other wireless systems, the control channels convey physical layer signals or messages which cannot be carried sufficiently efficiently, quickly or conveniently by higher layers. The design of the control channels transmitted in the LTE downlink aims to balance a number of somewhat conflicting requirements, the most important of which are discussed below.

9.3.1.1 Physical Layer Signalling to Support the MAC Layer

The general requirement to support Medium Access Control (MAC) operation is very similar to that in WCDMA, but there are a number of differences of detail, mainly arising from the frequency domain resource allocation supported in the LTE multiple access schemes.

The use of the uplink transmission resources on the Physical Uplink Shared Channel (PUSCH) is determined dynamically by an uplink scheduling process in the eNodeB, and therefore physical layer signalling must be provided to indicate to UEs which time/frequency resources they have been granted permission to use.

The eNodeB also schedules downlink transmissions on the PDSCH, and therefore similar physical layer messages from the eNodeB are needed to indicate which resources in the frequency domain contain the downlink data transmissions intended for particular UEs, together with parameters such as the modulation and code rate used for the data. Explicit signalling of this kind avoids the considerable additional complexity which would arise if UEs needed to search for their data among all the possible combinations of data packet size, format and resource allocation.

In order to facilitate efficient operation of Hybrid Automatic Repeat reQuest (HARQ) and ensure that uplink transmissions are made at appropriate power levels, further physical layer signals are also needed to convey acknowledgements of uplink data packets received by the eNodeB, and power control commands to adjust the uplink transmission power (as explained in Section 20.3).

9.3.1.2 Flexibility, Overhead and Complexity

The LTE physical layer specification is intended to allow operation in any system bandwidth from six resource blocks (1.08 MHz) to 110 resource blocks (19.8 MHz). It is also designed to support a range of scenarios including, for example, just a few users in a cell each demanding high data rates, or very many users with low data rates. Considering the possibility that both uplink resource grants and downlink resource allocations could be required for every UE in each subframe, the number of control channel messages carrying resource information could be as many as a couple of hundred if every resource allocation were as small as one resource block. Since every additional control channel message implies additional overhead which will consume downlink resources, it is desirable that the control channel is designed to minimize unnecessary overhead for any given signalling load, whatever the system bandwidth.

Similar considerations apply to the signalling of HARQ acknowledgements for each uplink packet transmission.

Furthermore, as in any mobile communication system, the power consumption of the terminals is an important consideration for LTE. Therefore, the control signalling must be designed so that the necessary scaleability and flexibility is achieved without undue decoding complexity.

9.3.1.3 Coverage and Robustness

In order to achieve good coverage it must be possible to configure the system so that the control channels can be received with sufficient reliability over a substantial part of every cell. This can be seen by considering, as an example, messages indicating resource allocation. If any such messages are not received correctly, then the corresponding data transmission will also fail, with a direct and proportionate impact on throughput efficiency. Techniques such as channel coding and frequency diversity can be used to make the control channels more robust. However, in order to make good use of system resources, it is desirable to be able to adapt the transmission parameters of the control signalling for different UEs or groups of UEs, so that lower code rates and higher power levels are only applied for those UEs for which it is necessary (e.g. near the cell border, where signal levels are likely to be low and interference from other cells high).

Also, it is desirable to avoid unintended reception of control channels from other cells, for example, by applying cell-specific randomization, particularly under conditions of high inter-cell interference.

9.3.1.4 System-Related Design Aspects

Since the different parts of LTE are intended to provide a complete system, some aspects of control channel design cannot be considered in isolation.

A basic design decision in LTE is that a control channel is intended to be transmitted in a single cell to a particular UE (or in some cases a group of UEs). In addition, in order to minimize signalling latency when indicating a resource allocation, a control channel transmission should be completed within one subframe. Therefore, in order to reach multiple UEs in a cell within a subframe, it must be possible to transmit multiple control channels within the duration of a single subframe. However, in those cases where the information sent via the control channels is intended for reception by more than one UE (for example, when relating to the transmission of a SIB on the PDSCH), it is more efficient to arrange for all the UEs to receive a single transmission rather than to transmit the same information to each UE individually. This requires that both common and dedicated control channel messages should be supported.

As noted previously, some scenarios may be characterized by low average data rates to a large number of UEs, for example when supporting VoIP traffic. If the data arrives at the eNodeB on a regular basis, as is typical for VoIP, then it is possible to predict in advance when resources will need to be allocated in the downlink or granted in the uplink. In such cases the number of control channel messages which need to be sent can be dramatically reduced by means of 'persistent scheduling' as discussed in Section 4.4.2.1.

9.3.2 Control Channel Structure and Contents

In this section we describe the downlink control channel features selected for inclusion in the LTE specifications and give some additional background on the design decisions. In general, the downlink control channels can be configured to occupy the first 1, 2 or 3 OFDM symbols in a subframe, extending over the entire system bandwidth as shown in Figure 9.5. There are two special cases: in subframes containing MBSFN transmissions there may be 0, 1 or 2 symbols for control signalling, while for narrow system bandwidths (less than 10 resource

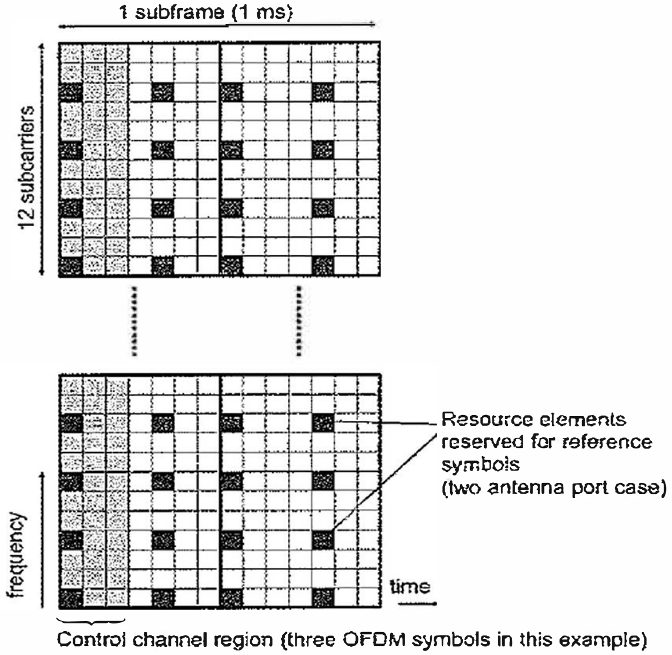


Figure 9.5 Time-frequency region used for downlink control signalling.

blocks) the number of control symbols is increased, and may be 2, 3 or 4 to ensure sufficient coverage at the cell border. This flexibility allows the control channel overhead to be adjusted according to the particular system configuration, traffic scenario and channel conditions.

9.3.2.1 Physical Control Format Indicator Channel (PCFICH)

The PCFICH carries a Control Format Indicator (CFI) which indicates the number of OFDM symbols (i.e. normally 1, 2 or 3) used for transmission of control channel information in each subframe. In principle the UE could deduce the value of the CFI without a channel such as the PCFICH, for example by multiple attempts to decode the control channels assuming each possible number of symbols, but this would result in significant additional processing load.

For carriers dedicated to MBSFN there are no physical control channels, so the PCFICH is not present in these cases.

Three different CFI values are used in the first version of LTE, and a fourth codeword is reserved for future use. In order to make the CFI sufficiently robust each codeword is 32 bits in length, as shown in Table 9.2. These 32 bits are mapped to 16 resource elements using QPSK modulation.

The PCFICH is transmitted on the same set of antenna ports as the PBCH, with transmit diversity being applied if more than one antenna port is used.

In order to achieve frequency diversity, the 16 resource elements carrying the PCFICH are distributed across the frequency domain. This is done according to a predefined pattern in the

Table 9.2 Control Format Indicator codewords.

CFI	CFI codeword $< b_0, b_1, \ldots, b_{31} >$
1	< 0, 1, 1, 0, 1, 1, 0, 1, 1, 0, 1, 1, 0, 1, 1, 0, 1, 1, 0, 1, 1, 0, 1, 1, 0, 1, 1, 0, 1, 1, 0, 1 >
2	< 1, 0, 1, 1, 0, 1, 1, 0, 1, 1, 0, 1, 1, 0, 1, 1, 0, 1, 1, 0, 1, 1, 0, 1, 1, 0, 1, 1, 0, 1, 1, 0 >
3	< 1, 1, 0, 1, 1, 0, 1, 1, 0, 1, 1, 0, 1, 1, 0, 1, 1, 0, 1, 1, 0, 1, 1, 0, 1, 1, 0, 1, 1, 0, 1, 1 >
4 (Reserved)	< 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0 >

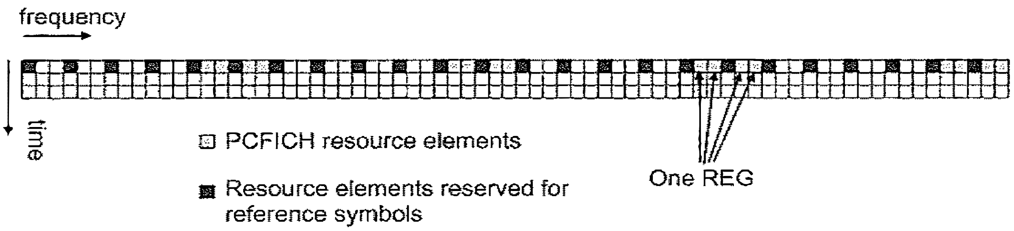


Figure 9.6 PCFICH mapping to Resource Element Groups (REGs).

first OFDM symbol in each downlink subframe (see Figure 9.6), so that the UEs can always locate the PCFICH information, which is a prerequisite to being able to decode the rest of the control signalling.

To minimize the possibility of confusion with PCFICH information from a neighbouring cell, a cell-specific frequency offset is applied to the positions of the PCFICH resource elements; this offset depends on the Physical Cell ID, which is deduced from the primary and secondary synchronization signals as explained in Section 7.2. In addition, a cell-specific scrambling sequence (again a function of the Physical Cell ID) is applied to the CFI codewords, so that the UE can preferentially receive the PCFICH from the desired cell.

9.3.2.2 Physical Downlink Control Channel (PDCCH)

A PDCCH carries a message known as Downlink Control Information (DCI), which includes resource assignments and other control information for a UE or group of UEs. In general, several PDCCHs can be transmitted in a subframe.

Each PDCCH is transmitted using one or more so-called Control Channel Elements (CCEs), where each CCE corresponds to nine sets of four physical resource elements known as Resource Element Groups (REGs). Four QPSK symbols are mapped to each REG. The resource elements occupied by reference symbols are not included within the REGs, which means that the total number of REGs in a given OFDM symbol depends on whether or not cell-specific reference signals are present. The concept of REGs (i.e. mapping in groups of four resource elements) is also used for the other downlink control channels (the PCFICH and PHICH).

Four PDCCH formats are supported, as listed in Table 9.3.

Table 9.3 PDCCH formats.

PDCCH format	Number of CCEs (n)	Number of REGs	Number of PDCCH bits
0	1	9	72
1	2	18	144
2	4	36	288
3	8	72	576

CCEs are numbered and used consecutively, and, to simplify the decoding process, a PDCCH with a format consisting of n CCEs may only start with a CCE with a number equal to a multiple of n.

The number of CCEs used for transmission of a particular PDCCH is determined by the eNodeB according to the channel conditions. For example, if the PDCCH is intended for a UE with a good downlink channel (e.g. close to the eNodeB), then one CCE is likely to be sufficient. However, for a UE with a poor channel (e.g. near the cell border) then eight CCEs may be required in order to achieve sufficient robustness. In addition, the power level of a PDCCH may be adjusted to match the channel conditions.

9.3.2.3 Formats for Downlink Control Information

The control channel messages are required to convey various pieces of information, but the useful content depends on the specific case of system deployment and operation. For example, if the infrastructure does not support MIMO, or if a UE is configured in a transmission mode which does not involve MIMO, there is no need to signal the parameters which are only required for MIMO transmissions. In order to minimize the signalling overhead it is therefore desirable that several different message formats are available, each containing the minimum payload required for a particular scenario. However, to avoid too much complexity in implementation and testing, it is desirable not to specify too many formats. The set of Downlink Control Information (DCI) message formats in Table 9.4 is specified in the first version of LTE. These are designed to cover the most useful cases. Additional formats may be defined in future.

In general the number of bits required for resource assignment depends on the system bandwidth, and therefore the message sizes also vary with the system bandwidth. Table 9.4 gives the number of bits in a PDCCH for uplink and downlink bandwidths of 50 resource blocks, corresponding to a spectrum allocation of about 10 MHz. In order to avoid additional complexity at the UE receiver, Formats 0 and 1A are designed to be always the same size. However, since these messages may have different numbers of bits, for example if the uplink and downlink bandwidths are different, leading to different numbers of bits required for indicating resource assignments, the smaller format size is extended by adding padding bits to be the same size as the larger.

The information content of the different DCI message formats is listed below.

Format 0. DCI Format 0 is used for the transmission of resource grants for the PUSCH. The following information is transmitted:

- Flag to differentiate between Format 0 and Format 1A
- Resource block grant
- Modulation and coding scheme
- HARQ information and redundancy version
- Power control command for scheduled PUSCH
- Request for transmission of an aperiodic CQI report (see Section 10.2.1).

Table 9.4 Supported DCI formats.

DCI format	Purpose	Number of bits including CRC (for a system bandwidth of 50 RBs and four antennas at eNodeB)
0	PUSCH grants	42
1	PDSCH assignments with a single codeword	47
1A	PDSCH assignments using a compact format	42
1B	PDSCH assignments for rank-1 transmission	46
1C	PDSCH assignments using a very compact format	26
1D	PDSCH assignments for multi-user MIMO	46
2	PDSCH assignments for closed-loop MIMO operation	62
2A	PDSCH assignments for open-loop MIMO operation	58
3	Transmit Power Control (TPC) commands for multiple users for PUCCH and PUSCH with 2-bit power adjustments	42
3A	Transmit Power Control (TPC) commands for multiple users for PUCCH and PUSCH with 1-bit power adjustments	42

Format 1. DCI Format 1 is used for the transmission of resource assignments for single codeword PDSCH transmissions (transmission modes 1, 2 and 7 (see Section 9.2.2.1)). The following information is transmitted:

- Resource allocation type (see Section 9.3.3.1)
- Resource block assignment

- Modulation and coding scheme
- HARQ information
- Power control command for Physical Uplink Control Channel (PUCCH).

Format 1A. DCI Format 1A is used for compact signalling of resource assignments for single codeword PDSCH transmissions, and for allocating a dedicated preamble signature to a UE for contention-free random access (see Section 19.3.2). The following information is transmitted:

- Flag to differentiate between Format 0 and Format 1A
- Flag to indicate that the distributed mapping mode (see Section 9.2.2.1) is used for the PDSCH transmission (otherwise the allocation is a contiguous set of physical resource blocks)
- Resource block assignment
- Modulation and coding scheme
- HARQ information
- Power control command for PUCCH.

Format 1B. DCI Format 1B is used for compact signalling of resource assignments for PDSCH transmissions using closed loop precoding with rank-1 transmission (transmission mode 6). The information transmitted is the same as in Format 1A, but with the addition of an indicator of the precoding vector applied for the PDSCH transmission.

Format 1C. DCI Format 1C is used for very compact transmission of PDSCH assignments. When format 1C is used, the PDSCH transmission is constrained to using QPSK modulation. This is used, for example, for signalling paging messages and some broadcast system information messages (see Section 9.2.2.2). The following information is transmitted:

- Resource block assignment
- Modulation and coding scheme
- Redundancy version.

Format 1D. DCI Format 1D is used for compact signalling of resource assignments for PDSCH transmissions using multi-user MIMO (transmission mode 5). The information transmitted is the same as in Format 1B, but instead of one of the bits of the precoding vector indicators, there is a single bit to indicate whether a power offset is applied to the data symbols. This feature is needed to show whether or not the transmission power is shared between two UEs. Future versions of LTE may extend this to the case of power sharing between larger numbers of UEs.

Format 2. DCI Format 2 is used for the transmission of resource assignments for PDSCH for closed-loop MIMO operation (transmission mode 4). The following information is transmitted:

- Resource allocation type (see Section 9.3.3.1)
- Resource block assignment
- Power control command for PUCCH
- HARQ information
- Modulation and coding schemes for each codeword
- Number of spatial layers
- Precoding information.

Format 2A. DCI Format 2A is used for the transmission of resource assignments for PDSCH for open-loop MIMO operation (transmission mode 3). The information transmitted is the same as for Format 2, except that if the eNodeB has two transmit antenna ports, there is no precoding information, and for four antenna ports two bits are used to indicate the transmission rank.

Formats 3 and 3A. DCI Formats 3 and 3A are used for the transmission of power control commands for PUCCH and PUSCH with 2-bit or 1-bit power adjustments respectively. These DCI formats contain individual power control commands for a group of UEs.

CRC attachment. In order that the UE can identify whether it has received a PDCCH transmission correctly, error detection is provided by means of a 16-bit CRC appended to each PDCCH. Furthermore, it is necessary that the UE can identify which PDCCH(s) are intended for it. This could in theory be achieved by adding an identifier to the PDCCH payload; however, it turns out to be more efficient to scramble the CRC with the 'UE identity', which saves the additional payload but at the cost of a small increase in the probability of falsely detecting a PDCCH intended for another UE.

In addition, for UEs which support antenna selection for uplink transmissions (see Section 17.5), the requested antenna may be indicated using Format 0 by applying an antenna-specific mask to the CRC. This has the advantage that the same size of DCI message can be used, irrespective of whether antenna selection is used.

PDCCH construction. In order to provide robustness against transmission errors, the PDCCH information bits are coded as described in Section 10.3.3.The set of coded and rate-matched bits for each PDCCH are then scrambled with a cell-specific scrambling sequence; this reduces the possibility of confusion with PDCCH transmissions from neighbouring cells. The scrambled bits are mapped to blocks of four QPSK symbols (REGs). Interleaving is applied to these symbol blocks, to provide frequency diversity, followed by mapping to the available physical resource elements on the set of OFDM symbols indicated by the PCFICH.

This mapping process excludes the resource elements reserved for reference signals and the other control channels (PCFICH and PHICH).

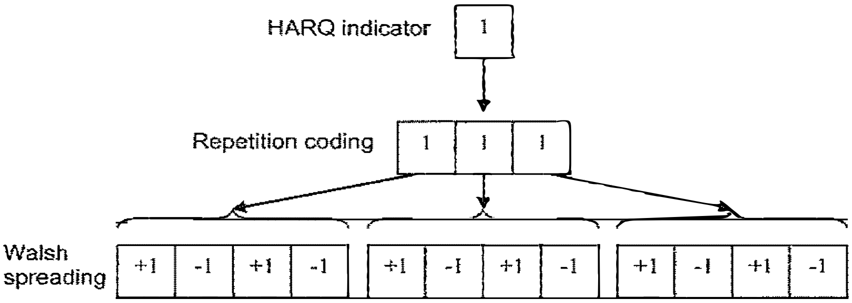
The PDCCHs are transmitted on the same set of antenna ports as the PBCH, and transmit diversity is applied if more than one antenna port is used.

9.3.2.4 Physical Hybrid ARQ Indicator Channel (PHICH)

The PHICH carries the HARQ ACK/NACK, which indicates whether the eNodeB has correctly received a transmission on the PUSCH. The HARQ indicator is set to 0 for a positive ACKnowledgement (ACK) and 1 for a Negative ACKnowledgement (NACK). This information is repeated in each of three BPSK (Binary Phase Shift Keying) symbols.

Multiple PHICHs are mapped to the same set of resource elements. These constitute a PHICH group, where different PHICHs within the same PHICH group are separated through different complex orthogonal Walsh sequences. The sequence length is four for the normal cyclic prefix (or two in the case of the extended cyclic prefix). As the sequences are complex, the number of PHICHs in a group (i.e. the number of UEs receiving their acknowledgements on the same set of downlink resource elements) can be up to twice the sequence length. A cell-specific scrambling sequence is applied.

Factor-3 repetition coding is applied for robustness, resulting in three instances of the orthogonal Walsh code being transmitted for each ACK or NACK. The error rate on the PHICH is intended to be of the order of 10^{-2} for ACKs and as low as 10^{-4} for NACKs. The resulting PHICH construction, including repetition and orthogonal spreading, is shown in Figure 9.7.



Each quadruplet is then mapped to an OFDM symbol

Figure 9.7 PHICH signal construction.

The PHICH duration, in terms of the number of OFDM symbols used in the time domain, is configurable (by an indication transmitted on the PBCH), normally to either one or three OFDM symbols. In some special cases[3] the three-OFDM-symbol duration is reduced to two OFDM symbols. As the PHICH cannot extend into the PDSCH transmission region, the

---

[3] The special cases when the PHICH duration is two OFDM symbols are (i) MBSFN subframes on mixed carriers supporting MBSFN and unicast data, and (ii) the second and seventh subframes in case of frame structure type 2 for Time Division Duplex (TDD) operation.

duration configured for the PHICH puts a lower limit on the size of the control channel region at the start of each subframe (as signalled by the PCFICH).

Finally, each of the three instances of the orthogonal code of a PHICH transmission is mapped to a REG on one of the first three OFDM symbols of each subframe,[4] in such a way that each PHICH is partly transmitted on each of the available OFDM symbols. This mapping is illustrated in Figure 9.8 for each possible PHICH duration.

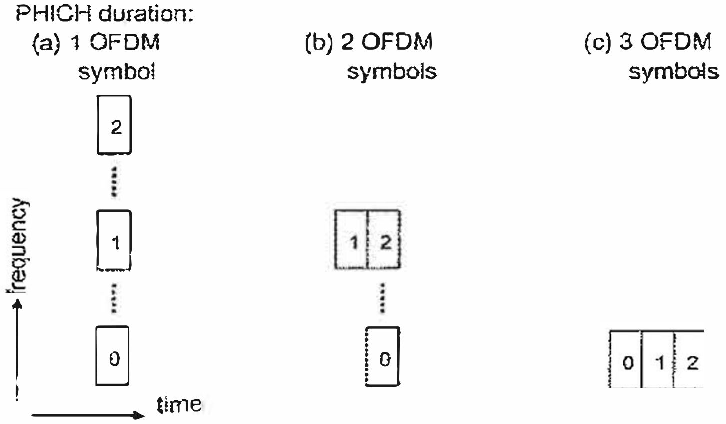


Figure 9.8 Examples of the mapping of the three instances of a PHICH orthogonal code to OFDM symbols, depending on the configured PHICH duration.

The PBCH also signals the number of PHICH groups configured in the cell, which enables the UEs to deduce to which remaining resource elements in the control region the PDCCHs are mapped.

In order to obviate the need for additional signalling to indicate which PHICH carries the ACK/NACK response for each PUSCH transmission, the PHICH index is implicitly associated with the index of the lowest uplink resource block used for the corresponding PUSCH transmission. This relationship is such that adjacent PUSCH resource blocks are associated with PHICHs in different PHICH groups, to enable some degree of load balancing. However, this mechanism alone is not sufficient to enable multiple UEs to be allocated the same resource blocks for a PUSCH transmission, as occurs in the case of uplink multi-user MIMO (see Section 17.5); in this case, different cyclic shifts of the uplink demodulation reference signals are configured for the different UEs which are allocated the same time-frequency PUSCH resources, and the same cyclic shift index is then used to shift the PHICH allocations in the downlink so that each UE will receive its ACK or NACK on a different PHICH. This mapping of the PHICH allocations is illustrated in Figure 9.9.

The PHICHs are transmitted on the same set of antenna ports as the PBCH, and transmit diversity is applied if more than one antenna port is used.

---

[4] The mapping avoids resource elements used for reference symbols or PCFICH.

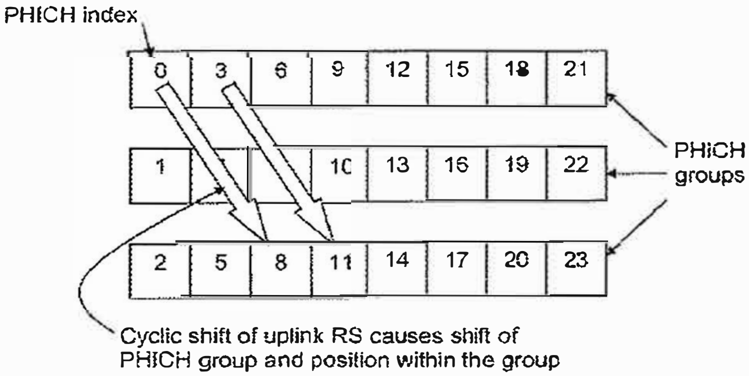


Figure 9.9 Indexing of PHICHs within PHICH groups, and shifting in case of cyclic shifting of the uplink demodulation reference signals.

9.3.3 Control Channel Operation

9.3.3.1 Resource Allocation

Conveying indications of physical layer resource allocation is one of the major functions provided by the PDCCHs. However, the exact use of the PDCCHs depends on the design of the algorithms implemented in the eNodeB. Nevertheless, it is possible to outline some general principles which are likely to be followed in typical systems.

In each subframe, PDCCHs indicate the frequency domain resource allocations. As discussed in Section 9.2.2.1, resource allocations are normally localized, meaning that a Physical Resource Block (PRB) in the first half of a subframe is paired with the PRB at the same frequency in the second half of the subframe. For simplicity, the explanation here is in terms of the first half subframe only.

The main design challenge for the signalling of frequency domain resource allocations (in terms of a set of resource blocks) is to find a good compromise between flexibility and signalling overhead. The most flexible, and arguably the simplest, approach is to send each UE a bitmap in which each bit indicates a particular PRB. This would work well for small system bandwidths, but for large system bandwidths (i.e. up to 110 PRBs) the bitmap would need 110 bits, which would be a prohibitive overhead – particularly for transmission of small packets, where the PDCCH message could be larger than the data packet! One possible solution would be to send a combined resource allocation message to all UEs, but this was rejected on the grounds of the high power needed to reach all UEs reliably, including those at the cell edges. A number of different approaches were considered by 3GPP, and those adopted are listed in Table 9.5.

Further details of the different resource allocation methods are given below.

Resource allocation Type 0. In resource allocations of Type 0, a bitmap indicates the Resource Block Groups (RBGs) which are allocated to the scheduled UE, where a RBG is a set of consecutive PRBs. The RBG size (P) is a function of the system bandwidth as shown

Table 9.5 Methods for indicating resource allocation.

Method	UL/DL	Description	Number of bits required (see main text for definitions)
Direct bitmap	DL	The bitmap comprises 1 bit per RB. This method is the only one applicable when the bandwidth is less than 10 resource blocks.	$N_{\mathrm{RB}}^{\mathrm{DL}}$
Bitmap: 'Type 0'	DL	The bitmap addresses Resource Block Groups (RBGs), where the group size (2, 3 or 4) depends on the system bandwidth.	$\lceil N_{\mathrm{RB}}^{\mathrm{DL}}/P\rceil$
Bitmap: 'Type 1'	DL	The bitmap addresses individual resource blocks in a subset of RBGs. The number of subsets (2, 3, or 4) depends on the system bandwidth. The number of bits is arranged to be the same as for Type 0, so the same DCI format can carry either type of allocation.	$\lceil N_{\mathrm{RB}}^{\mathrm{DL}}/P\rceil$
Contiguous allocations: 'Type 2'	DL or UL	Any possible arrangement of contiguous resource block allocations can be signalled in terms of a starting position and number of resource blocks.	$\lceil\log_2(N_{\mathrm{RB}}^{\mathrm{DL}}(N_{\mathrm{RB}}^{\mathrm{DL}}+1))\rceil$ or $\lceil\log_2(N_{\mathrm{RB}}^{\mathrm{UL}}(N_{\mathrm{RB}}^{\mathrm{UL}}+1))\rceil$
Distributed allocations	DL	In the downlink a limited set of resource allocations can be signalled where the resource blocks are scattered across the frequency domain and shared between two UEs. For convenience the number of bits available for using this method is the same as for contiguous allocations Type 2, and the same DCI format can carry either type of allocation.	$\lceil\log_2(N_{\mathrm{RB}}^{\mathrm{DL}}(N_{\mathrm{RB}}^{\mathrm{DL}}+1))\rceil$

in Table 9.6. The total number of RBGs (N_{RBG}) for a downlink system bandwidth of $N_{\mathrm{RB}}^{\mathrm{DL}}$ PRBs is given by $N_{\mathrm{RBG}} = \lceil N_{\mathrm{RB}}^{\mathrm{DL}}/P\rceil$. An example for the case of $N_{\mathrm{RB}}^{\mathrm{DL}} = 25$, $N_{\mathrm{RBG}} = 13$ and $P = 2$ is shown in Figure 9.10, where each bit in the bitmap indicates a pair of PRBs (i.e. two PRBs which are adjacent in frequency).

Resource allocation Type 1. In resource allocations of Type 1, individual PRBs can be addressed (but only within a subset of the PRBs available). The bitmap used is slightly