

LTE-Advanced

A Practical Systems Approach to Understanding 3GPP LTE Releases 10 and 11 Radio Access Technologies

Sassan Ahmadi



Copyright © 2013, Elsevier Science & Technology. All rights reserved.

LTE-Advanced

This page intentionally left blank

Copyright © 2013. Elsevier Science & Technology. All rights reserved.

LTE-Advanced

A Practical Systems Approach to Understanding the 3GPP LTE Releases 10 and 11 Radio Access Technologies

Sassan Ahmadi



AMSTERDAM • BOSTON • HEIDELBERG • LONDON
NEW YORK • OXFORD • PARIS • SAN DIEGO
SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO
Academic Press is an imprint of Elsevier



Academic Press is an imprint of Elsevier
The Boulevard, Langford Lane, Kidlington, Oxford, OX5 1GB, UK
225 Wyman Street, Waltham, MA 02451, USA
525 B Street, Suite 1800, San Diego, CA 92101-4495, USA

Copyright © 2014 Elsevier Inc. 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 without the prior written permission of the publisher.

Permissions may be sought directly from Elsevier's Science & Technology Rights Department in Oxford, UK: phone (+44) (0) 1865 843830; fax (+44) (0) 1865 853333; email: permissions@elsevier.com. Alternatively, visit the Science and Technology Books website at www.elsevierdirect.com/rights for further information.

Notice

No responsibility is assumed by the publisher for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions or ideas contained in the material herein. Because of rapid advances in the medical sciences, in particular, independent verification of diagnoses and drug dosages should be made.

British Library Cataloguing-in-Publication Data

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

Library of Congress Cataloging-in-Publication Data

A catalog record for this book is available from the Library of Congress

ISBN: 978-0-12-405162-1

For information on all Academic Press publications
visit our website at elsevierdirect.com

Typeset by MPS Limited, Chennai, India
www.adi-mps.com

Printed and bound in United States of America

14 15 16 10 9 8 7 6 5 4 3 2 1

		Working together to grow libraries in developing countries
www.elsevier.com • www.bookaid.org		

Network Architecture

2

CHAPTER CONTENTS

2.1 Overall network architecture	31
2.1.1 General concepts	31
2.1.2 Evolution of the network architecture	38
2.2 EPC architecture	42
2.2.1 Reference model and network interfaces	42
2.2.2 Public land mobile network	43
2.2.3 Packet data network gateway	46
2.2.4 Serving gateway	47
2.2.5 Mobility management entity	47
2.2.6 Policy and charging rules function	48
2.2.7 Home subscriber server	49
2.2.8 Authorization, authentication, and accounting	49
2.2.9 EPC identifiers	50
2.3 Roaming architecture and interworking with other networks	50
2.4 E-UTRAN architecture	57
2.4.1 E-UTRAN functions	59
2.4.2 E-UTRAN interfaces	59
2.4.3 E-UTRAN identifiers	60
2.4.4 Support of HeNB in E-UTRAN	61
2.4.5 Support of RRHs in E-UTRAN	63
2.4.6 Support of relays in E-UTRAN	64
2.5 E-MBMS architecture	67
2.6 Positioning network architecture	71
2.7 Mobility management	74
2.8 Radio resource management	75
2.9 Quality of service/quality of experience	77
2.10 Security in E-UTRAN	78
2.11 Support of voice and messaging services in LTE/LTE-Advanced	80
2.11.1 IMS architecture	80
2.11.2 Single radio voice call continuity	84
2.11.3 Voice over LTE	90
2.11.4 Circuit-switched fallback	94

LTE-Advanced.

© 2014 Elsevier Inc. All rights reserved.

2.12 LIPA, selected IP traffic offload, and IP flow mobility	100
2.12.1 Local IP access	101
2.12.2 Selected IP traffic offload	105
2.12.3 IP flow mobility	107
2.12.4 Wi-Fi offloading	110
2.13 Mobile IP	111
2.13.1 Client-based IP mobility	113
2.13.2 Network-based IP mobility	113
2.14 Network architecture evolution and small cells	114
References	116

3GPP Rel-8 specifications included enhancements to high-speed packet access (HSPA) core and access networks as well as the introduction of the evolved packet system (EPS) which comprises an Internet protocol (IP)-based, all-packet core network with minimal hierarchy in conjunction with a new orthogonal frequency division multiplexing (OFDM)-based radio access network. The EPS is a simplified architecture (i.e., some intermediate network elements were removed or combined together) with an all-IP network support for higher throughput, lower latency radio access and improved mobility between multiple heterogeneous access networks including evolved universal terrestrial radio access (E-UTRA), 3GPP legacy systems (e.g., GERAN or UTRAN), as well as non-3GPP systems, such as Wi-Fi and CDMA2000.

The network architecture of LTE is based on the functional decomposition principle, where underlying features are divided into logical functional entities without specific implementation assumptions about physical network nodes. That is why 3GPP specified an entirely new packet-switched core network architecture to support the evolved universal terrestrial radio access network (E-UTRAN) through reduction in the number of network elements, simpler functionality, and more importantly allowing connections and handover to other access technologies, giving the service providers the ability to deliver a seamless mobility experience.

The introduction of the evolved packet core (EPC) and all-IP network architecture had significant impacts on mobile services, where voice, data, and video communications are built on the IP-based protocols; interworking of the new architectural entities with 2G/3G legacy systems; scalability required by each of the core elements to address the increasing number of direct connections to user terminals, orders of magnitude of bandwidth increase, and terminal mobility; reliability and availability provided by each network entity to ensure service continuity. To address a fundamentally different set of network and service requirements, the EPC design represented a departure from previous mobile networking paradigms.

This chapter describes the design principles of the EPS, network reference model, logical building blocks of the network architecture as well as theoretical

and operational aspects of EPC and LTE/LTE-Advanced radio access network (E-UTRAN) building blocks, functional split, and their interconnections. Under this topic, the network functions, such as mobility management, radio resource management (RRM), quality of service (QoS)/QoE, mobile IP, IP flow mobility, relays and femto-cells, and security, along with practical design/implementation considerations and real-life issues and limitations will be discussed. Furthermore, the operation and management of small cells involving low-power remote radio heads (RRHs) will be covered. The selective IP traffic offload, local IP access (LIPA), and Wi-Fi offloading mechanisms will be discussed, as well. In this chapter, we provide a top-down systems approach to understanding the architecture and operation of the LTE/LTE-Advanced access and core networks. The focus in this chapter is on the network architecture and underlying network functional entities and the network interfaces. The protocols and procedures for network attach/detach, connection establishment/release, mobility management, and other functional details will be discussed in the next chapters.

2.1 Overall network architecture

2.1.1 General concepts

In its most abstract form, the EPS consists of only two nodes in the user-plane, that is, a base station and a core network gateway. The node that performs control-plane functionality is separated from the node that performs user-plane functionality. The basic EPS architecture is illustrated in Figure 2.1. The EPS architecture was designed not only to provide a smooth evolution from the 2G/3G quasi-packet-switched architectures consisting of nodeBs (NBs), radio network controllers (RNCs), SGSNs, and GGSNs, but also to provide support for non-3GPP access (e.g., Wi-Fi), improved policy control and charging, a wider range of QoS capabilities, advanced security/authentication mechanisms, and flexible roaming [70,71].

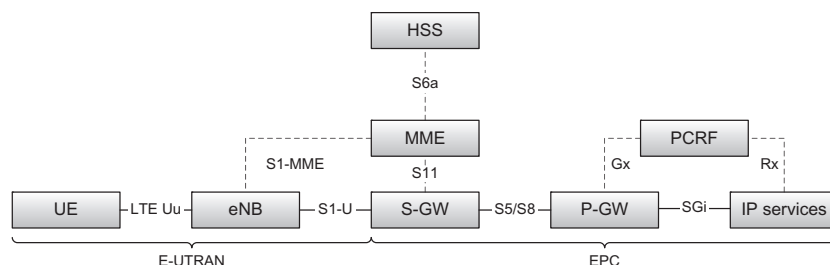


FIGURE 2.1

Generic EPS architecture [70,71].

The introduction of EPC combined with E-UTRAN in many ways represented a radical departure from previous mobile network design philosophies. It heralded the end of the circuit-switched voice-centric design era. 3GPP LTE was meant to use voice over IP (VoIP) for voice services. The EPC treats voice just as other IP-based applications, although an important one that imposes stringent performance and operational requirements [34,37,69].

3GPP LTE was required to match the wired-broadband quality of experience (QoE). This is quite different from providing mainly best-effort low-rate web browsing or short message service (SMS), two data applications for which its predecessor quasi-packet-switched mobile networks were optimized. In LTE, all mobility management functions have been moved to the core network and have become part of mobility management entity (MME). This is a consequence of the split of functions previously performed by the RNC and NB/Base Transceiver Station (BTS). The MME requires a control-plane capacity that is an order of magnitude larger than the SGSN or PDSN, and must ensure interworking with 2G/3G legacy mobile systems (see Figure 2.2).

The long-term evolution of universal mobile telecommunications system (UMTS) systems was expected to provide superior end-to-end QoS management and enforcement in order to deliver multimedia, low-latency, and real-time services, with more granular QoS classes and stricter performance targets. This was going to be achieved while ensuring scalability of users, services, and data sessions. In LTE, service control is provided via the policy and charging rules function (PCRF). This was a major change from previous systems, where service control was realized primarily through user equipment (UE) authentication by the network. The PCRF entity dynamically controls and manages all data sessions and provides appropriate interfaces toward charging and billing systems, as well as enabling new business models. The design of the new air-interface required

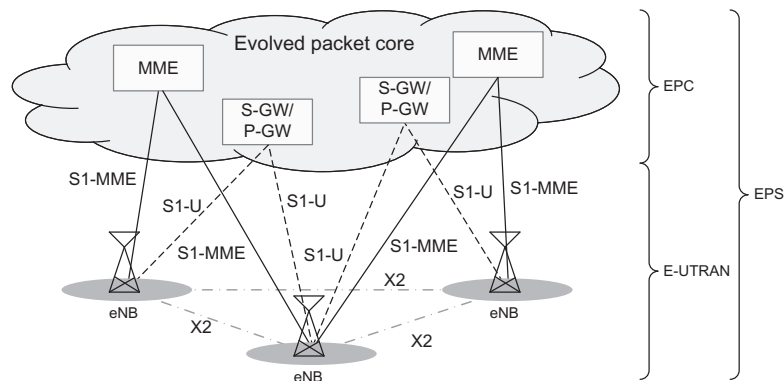


FIGURE 2.2

EPS composition [12].

significantly more capacity in the user-plane and control-plane. The existing 2G/3G mobile core elements cannot satisfactorily address LTE requirements without major upgrades to the platforms. Therefore, the legacy core network implementations must be dramatically upgraded to enable LTE.

While the core network consists of many logical nodes, the access network consists of essentially just one node, the eNB, which provides connections to the UEs. Each of the network elements is interconnected by means of logical interfaces that are standardized in order to allow multi-vendor interoperability. This gives network operators the possibility of sourcing 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 2.3. The EPC and E-UTRAN network elements are described in the next sections.

The EPC design has radically changed some key networking concepts which were used in the previous generations of cellular core networks and successfully addressed a number of technological challenges which had arisen from significant technological improvements in the radio access side. The LTE radio access system provides more efficient use of the spectrum with wider spectral bands reserved for the broadband systems, resulting in greater system capacity and performance.

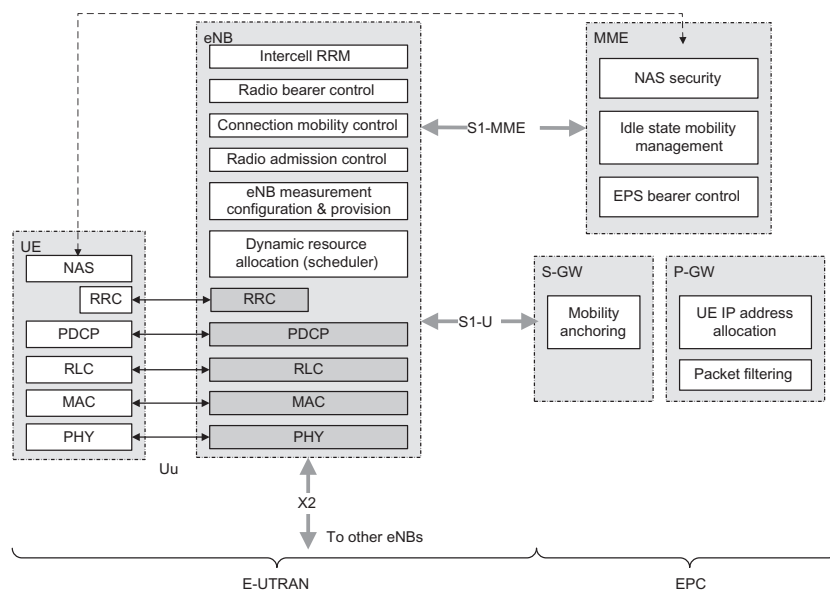


FIGURE 2.3

Functional split between E-UTRAN and EPC entities [12].

At the same time, the core network needs to change in order to provide higher throughput and lower latency; both should come as a result of simplified and improved IP-based network architecture. More specifically, the EPC addressed a number of fundamental IP-related aspects of LTE deployment including the following:

- IP routing and network addressing and real-time management
- Centralized versus distributed network architecture (for MME, serving gateway (S-GW), and packet data network gateway (P-GW) deployment)
- IPv6 (strategy, introduction, and coordination with IPv4)
- End-to-end QoS deployment and coordination with underlying transport
- Layer-2 versus layer-3 connectivity at the transport layer (eNB, S-GW, P-GW, and MME)
- End-to-end security for user-plane and control-plane
- Interconnectivity to external networks and virtual private networks
- Enabling lawful interception of data packets

There are a stringent set of requirements for reliable, scalable, high-performance network elements. Due to the dynamic nature of the user mobility in LTE, coupled with the short duration of multiple data sessions per UE and large-scale deployment targets, the EPC must further address the challenging requirements of dynamic management of mobility, policies, and signaling/data bearers, while ensuring interworking and interoperability with legacy 2G/3G systems.

The EPC is realized through the following four new network elements: (1) S-GW, (2) P-GW, (3) MME, and (4) PCRF. While S-GW, P-GW and MME were introduced in 3GPP Rel-8, PCRF had been introduced in 3GPP Rel-7. The architectures using PCRF have not been widely adopted. The PCRF interoperation with the EPC gateways and the MME is normative in 3GPP Rel-8 and essential for the operation of LTE.

The MME is a node within the EPC that performs signaling and control functions to manage the user access to network connections, assignment of network resources, and management of the mobility states to support tracking, paging, roaming, and handovers. The MME controls all control-plane functions related to subscriber and session management. The MME is the key control-node for the LTE access-network. It is responsible for idle-mode UE tracking and paging procedure. It is involved in the bearer activation/deactivation process and is also responsible for choosing the S-GW for a UE at the initial attach and at the time of intra-LTE handover involving core network node relocation. It is further responsible for user authentication. The non-access stratum (NAS) signaling terminates at the MME and it is also responsible for generation and allocation of temporary identities to UEs. It checks the authorization of the UE to camp on the service provider's public land mobile network (PLMN) and enforces UE roaming restrictions. The MME is the termination point in the network for ciphering/integrity protection for NAS signaling and handles the security key management. Lawful interception of signaling is also supported by the MME. The MME also

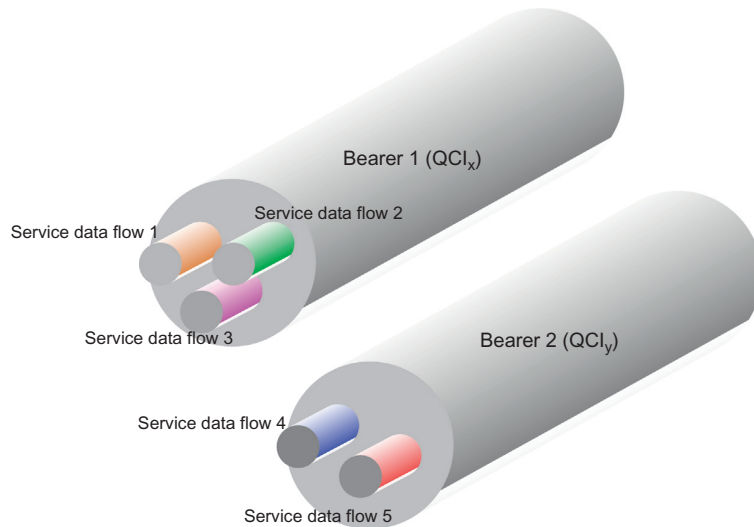
provides the control-plane function for mobility between LTE and 2G/3G access networks with the S3 interface terminating at the MME from the SGSN. The NAS protocols form the control-plane between the UE and the MME and support the UE mobility and the session management procedures to establish and maintain IP connectivity between the UE and a P-GW. They define the rules for a mapping between parameters during intersystem mobility with 3GPP or non-3GPP access networks. They also provide the NAS security by integrity protection and ciphering of NAS signaling messages.

The home subscriber server (HSS) is a central database that contains user- and subscription-related information. The functions of the HSS include mobility management, call and session establishment support, user authentication, and access authorization. The HSS is based on legacy home location register (HLR) and authentication center (AuC) functions in previous 3GPP releases.

The S-GW is a user-plane element whose primary function is to manage user-plane mobility and to act as a demarcation point between the radio access and core networks. The S-GW maintains data paths between eNBs and the P-GW. From a functional perspective, the S-GW is the termination point of the packet data network interface toward the E-UTRAN. When terminals move in the areas served by eNBs in the E-UTRAN, the S-GW serves as a local mobility anchor (LMA). This means that packets are routed through this point for intra E-UTRAN mobility and mobility with other 3GPP technologies, such as GSM and UMTS. The S-GW routes and forwards user data packets, while acting as the mobility anchor for the user-plane during inter-eNB handovers and as the anchor for mobility between LTE and other 3GPP radio access technologies, that is, terminating S4 interface and relaying the traffic between 2G/3G systems and P-GW. For idle mode UEs, the S-GW terminates the downlink data path and triggers paging when downlink data arrives for a UE. It manages and stores UE contexts, for example, parameters of the IP bearer service and network internal routing information. It also performs replication of the user traffic in case of lawful interception.

The P-GW is the termination point of the packet data interface toward the packet data network. As an anchor point for sessions toward the external packet data networks, the P-GW supports policy enforcement features, such as applying operator-defined rules for resource allocation and usage; packet filtering functions, such as packet inspection for application type detection and charging support. The P-GW provides connectivity from the UE to external packet data networks by being the point of exit and entry of traffic for the UE. A UE may have simultaneous connections with more than one P-GW for accessing multiple packet data networks (PDNs). Another key role of the P-GW is to act as the anchor for mobility between 3GPP and non-3GPP technologies, such as CDMA2000.

In LTE, the data-plane traffic is carried via virtual connections known as service data flows (SDFs). The SDFs are carried over bearers that are virtual containers with unique QoS characteristics. [Figure 2.4](#) illustrates an example scenario

**FIGURE 2.4**

Relationship between SDFs and bearers [70,71].

where one or more SDFs are aggregated and assigned to one bearer. One bearer, a data path between a UE and a PDN, has three segments as follows: radio bearer between UE and eNB; data bearer between eNB and S-GW (S1 bearer); and data bearer between S-GW and P-GW (S5 bearer). Figure 2.5 illustrates three segments that constitute an end-to-end bearer. The primary role of a P-GW is QoS enforcement for each of these SDFs, while S-GW dynamically manages the bearers. There are two types of EPS bearers: the *default EPS bearer* and the *dedicated EPS bearer*. The default EPS bearer is established during the attach procedure and allocates an IP address to the UE that does not have a specific QoS (only a nominal QoS). A dedicated EPS bearer is typically established during call setup and after transition from idle mode to the connected mode. It does not allocate any additional IP address to the UE and is linked to a specified default EPS bearer but has a specific QoS class. A QoS class identifier (QCI) is a scalar that is used as a reference to access node-specific parameters that control bearer level packet forwarding (e.g., scheduling weights, admission thresholds, queue management thresholds, and link layer protocol configuration), and are preconfigured by the operator owning the access node (eNB).

As shown in Figure 2.5, the EPS bearer transports packets by an S1 bearer between the S-GW and the eNB, and by a radio bearer between a UE and an eNB. There is a one-to-one correspondence between a radio bearer identifier and an S1 bearer. The 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, and radio link control (RLC) configuration. In order to



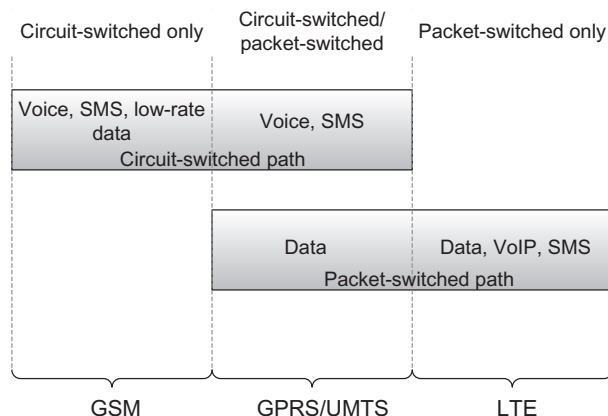
Bearers across different EPS interfaces [70, 71].

provide different bearer-level QoS, a separate EPS bearer should be established. The IP packets must then be filtered and mapped to the appropriate EPS bearers. Packet filtering of different bearers is based on traffic flow templates (TFTs). The TFTs use IP header information, such as source and destination IP addresses and transport control protocol (TCP) port numbers to filter packets, such as VoIP from web-browsing traffic, so that each flow can be mapped to the appropriate EPS bearers with the appropriate QoS settings. 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. When a UE attaches to the network, the UE is assigned an IP address by the P-GW and a default bearer is established, and it remains in effect during the lifetime of the PDN connection in order to provide the UE with always-on 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 policy and charging enforcement function (PCEF) may change these values in interaction with the PCRF or according to local configuration. One or more dedicated bearers can be established at any time during or after completion of the attach procedure. A dedicated bearer can be either a guaranteed bit rate (GBR) or a non-GBR bearer. The default bearer is always a non-GBR bearer. The distinction between default and dedicated bearers should be transparent to the access network. 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. The dedicated bearers may be established by the network based on a trigger from the IP multimedia subsystem (IMS) domain or a request by the UE. The dedicated bearers for a UE may be provided by one or more P-GWs.

In the E-UTRAN side, the evolved nodeB (eNB) provides the E-UTRA user-plane and control-plane protocol termination toward the UE. The eNBs can optionally be interconnected via X2 logical interfaces. The X2 interface may not be necessarily in the form of a point-to-point physical connection between eNBs. The eNBs are connected by means of the S1 interface to the EPC. More specifically, it is connected to the MME by means of the S1-MME reference point, and to the S-GW by means of the S1-U interface. The S1 interface supports a many-to-many relationship between MMEs/S-GWs and eNBs.

2.1.2 Evolution of the network architecture

The global system for mobile communications (GSM) network architecture was based on a circuit-switched design where a direct connection was established between the call-making and call-receiving entities throughout the telecommunication network, that is, radio access and core networks of the mobile operator and the fixed network. The GSM was a voice-centric system providing circuit-switched telephony, SMS, and some very low-rate data services. The general packet radio service (GPRS) enhancement added packet-switched services to the circuit-switched functions, where data was transported in the form of packets without

**FIGURE 2.6**

Evolution of cellular networks and services.

establishing dedicated circuits. This enhancement offered more flexibility and efficiency. In GPRS, voice and SMS are transported via circuit-switched paths; thus, the core network is composed of circuit-switched and packet-switched domains. In UMTS, the dual-domain concept was maintained on the core network side and only a few network elements evolved. In the design of the 3G system long-term evolution, the 3GPP decided to use the IP as the key protocol to transport all services; therefore, the EPC vision did not include a circuit-switched domain and it was an evolution of the packet-switched architecture used in GPRS/UMTS. This paradigm shift in the network design not only had significant impact on the architecture itself but also on the way that the services were provided. Traditional use of circuits to carry voice and short messages needed to be replaced by IP-based solutions in the long term. The evolution of 3GPP network architecture and services are depicted in Figures 2.6–2.8 from different perspectives (services, general architecture, and network elements). As shown in the figures, while the services have migrated toward pure IP-based services, the network architecture has been simplified and redundant entities have been eliminated or integrated into other network elements.

Different options are possible to integrate MME, S-GW, and P-GW logical entities. Each option will result in different handling of control-plane and user-plane packet routing. The choice may depend on operator strategy and/or EPC vendor equipment availability and constraints. Integrating more logical entities together in one physical entity will reduce latency/delay, cost (fewer nodes and interfaces) and signaling load, but will increase the functional complexity per entity/component and will limit the benefit and potential savings achieved by a split architecture, i.e., redundancy, flexibility, and scalability. A possible deployment option may start with some level of integration, to save initial costs, and migrate to more flexible and scalable architecture when the number of users and user traffic grows.

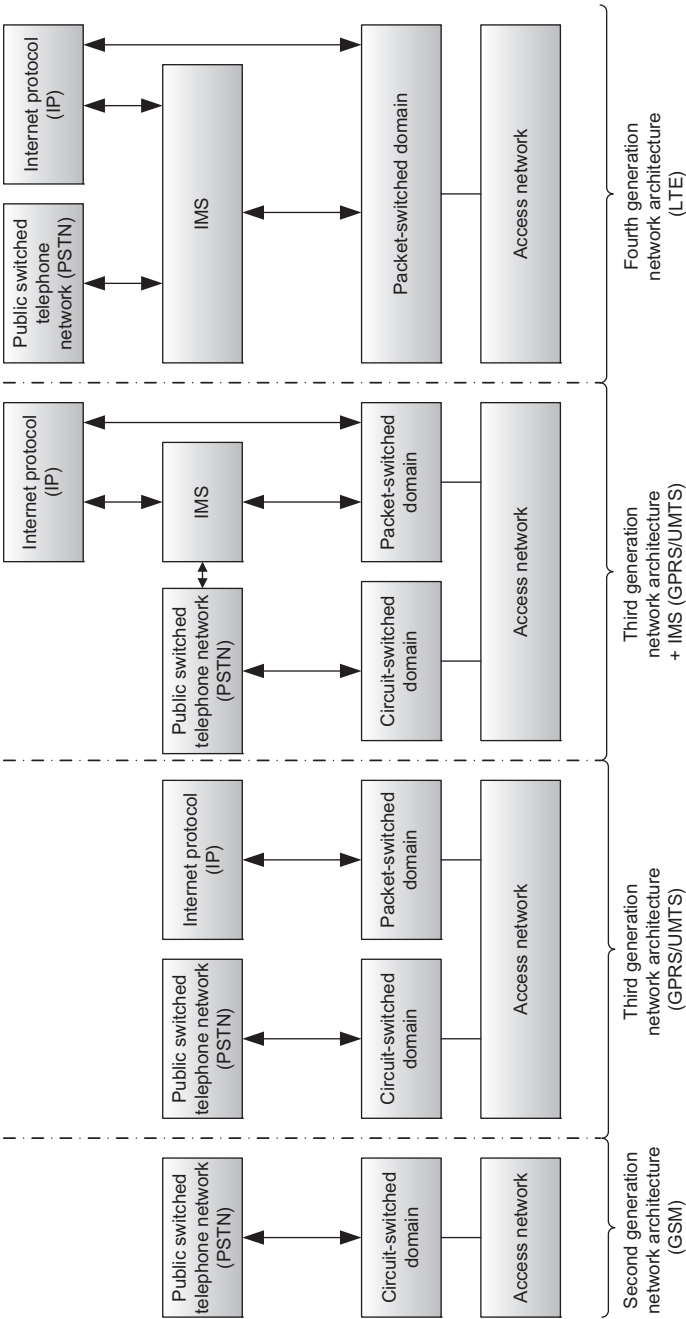


FIGURE 2.7

Evolution of network architecture through cellular generations.

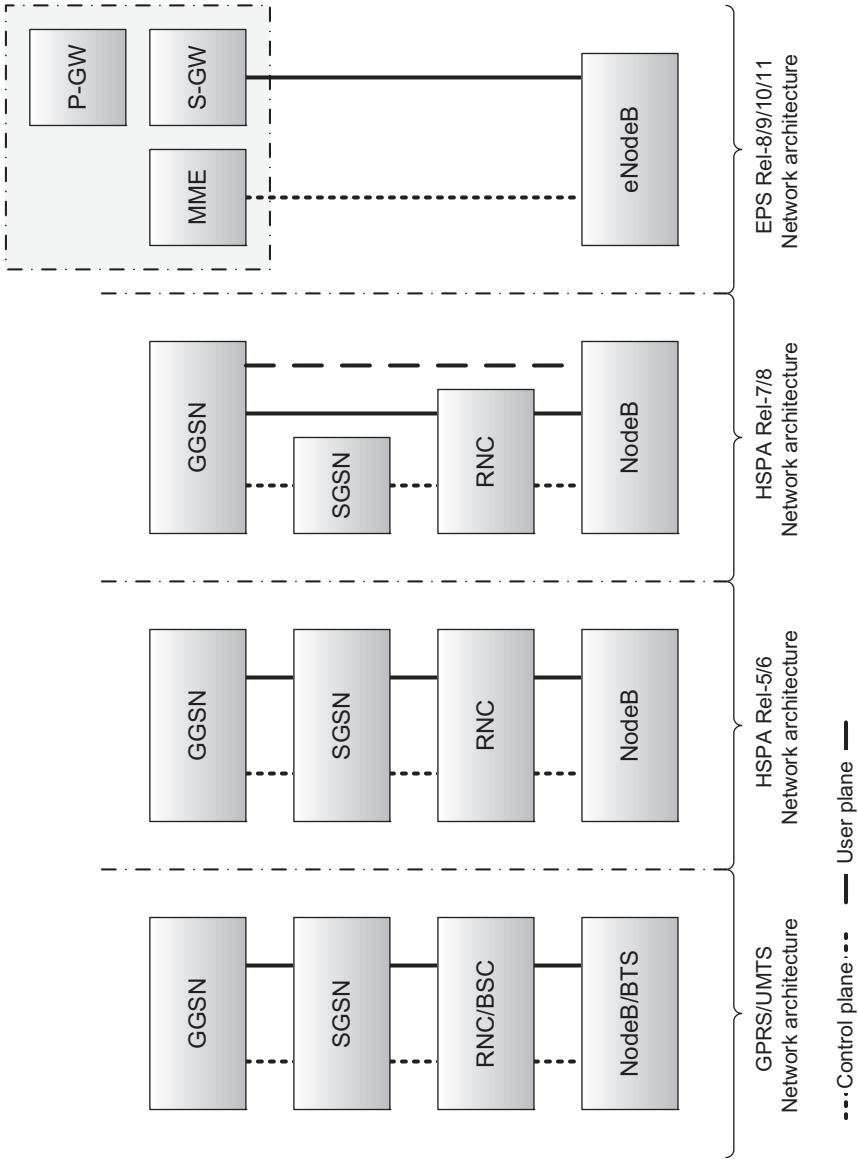


FIGURE 2.8

Evolution of 3GPP network architecture through different releases.

2.2 EPC architecture

2.2.1 Reference model and network interfaces

As mentioned earlier, the EPC is an all-IP mobile core network for the LTE radio access, providing a converged framework for IP-based real-time/non-real-time applications and services. The EPC provides the entire core-network functionality that, in previous mobile network generations, was realized through two separate domains: circuit-switched for voice and packet-switched for data. As shown in Figure 2.7, in LTE, these two distinct mobile core network domains used for separate processing and switching of voice and data are unified in a single IP-based domain. The EPC implementation and deployment is essential for end-to-end IP-based service delivery which complements the LTE radio access network. The EPC promotes the introduction of new services and the enablement of new applications [70,71].

Figure 2.9 shows the EPS reference model, which is a logical representation of the network architecture. The network reference model identifies the functional entities in the architecture and the reference points between the functional elements over which interoperability is achieved. The overall architecture has two distinct components: the access network and the core network. The access network is the E-UTRAN and the core network is known as the EPC. Voice services which were traditionally circuit-switched are handled using IMS infrastructure. Network complexity and latency are reduced as there are fewer nodes in both signaling and data paths. The EPC is further designed to support non-3GPP access for mobile IP. To improve system robustness, security, integrity protection, and ciphering have been added and represented by NAS, which is an additional layer of abstraction to protect important information, such as key and security interworking between 3GPP and non-3GPP networks. Apart from the network entities handling user-plane traffic, the EPC further includes network control entities for storing user subscription information represented by HSS, determining the identity and privileges of a user and tracking activities via authorization, authentication and accounting (AAA) server, and enforcing charging and QoS policies through PCRF. Note that E-UTRAN and EPC together constitute the EPS. The radio access and core networks perform several functions including network access control functions; packet

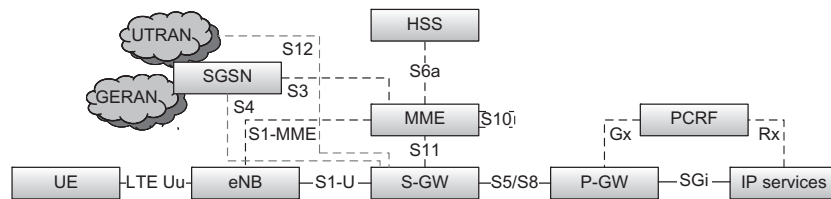


FIGURE 2.9

EPS reference model [6].

routing and transfer functions; mobility management functions; security functions; RRM functions; and network management functions.

Some general principles that were taken into consideration in the design of E-UTRAN architecture as well as the E-UTRAN interfaces are summarized as follows:

- Signaling and data transport networks are logically separated.
- E-UTRAN and EPC functions are fully separated from transport functions. The addressing schemes used in E-UTRAN and EPC are not associated with the addressing schemes of transport functions. The fact that some E-UTRAN or EPC functions may reside in the same physical equipment does not make the transport functions part of the E-UTRAN or the EPC.
- The mobility for RRC connection is fully controlled by the E-UTRAN.
- The interfaces are based on the logical model of the network entities.
- One physical network element can implement multiple logical nodes.

3GPP network architecture defines a reference point as a logical link that connects two groups of functions that reside in different functional entities of the E-UTRAN or EPC. Figure 2.9 illustrates the EPS network reference model and the most commonly used reference points specified by 3GPP as of Rel-11. These reference points are defined in Table 2.1.

2.2.2 Public land mobile network

A PLMN provides land mobile telecommunication services to the public which is established and operated by an administration or private operating agency. A PLMN may be referred to as an extension of networks, e.g., integrated services data digital network (ISDN), corporate, and public PDNs. The PLMN is a collection of mobile switching centers (MSCs) in circuit-switched domain and SGSNs in GPRS core network or SGSNs/MMEs in EPC, in packet-switched domain under a common numbering and routing plan. The MSCs are the functional interfaces between the fixed networks and a PLMN for call setup in circuit-switched domain. The GGSN and the SGSN are the functional interfaces between the fixed telephone networks and a PLMN for packet transmission in a GPRS packet-switched domain. In the case of an EPC packet-switched domain, the P-GW, S-GW, SGSN, and the MME are the functional interfaces between the fixed networks and a PLMN for packet transmission. From a functional point of view, the PLMNs may be considered as independent telecommunication entities even though different PLMNs may be interconnected through an ISDN or public switched telephone network (PSTN) and PDNs for the purpose of forwarding calls or network information. A similar type of interconnection may exist for interaction between the MSCs, SGSNs, or MMEs of a PLMN [1].

The PLMN infrastructure typically consists of a core network and an access network. The core network is logically divided into a circuit-switched domain, a packet-switched domain, and an IMS domain. The access network may include a

Table 2.1 EPS Reference Points [6,33]

Reference Point	Termination Points	Description
S1-MME	eNB and MME	It provides control-plane path between eNB and MME.
S1-U	eNB and S-GW	It enables per-bearer user-plane tunneling and inter-eNB path switching during handover. The S1-U is based on GTP-U protocol. ^a
S3	MME and SGSN	It enables user and bearer information exchange for inter-3GPP access network mobility in idle and/or active mode. This reference point can be used intra-PLMN or inter-PLMN in the case of inter-PLMN handover. The S3 is based on GTP protocol.
S4	S-GW and SGSN	It provides control and mobility support between GPRS core and the 3GPP anchor function of S-GW and is based on GTP protocol. If direct tunneling is not established, it provides the user-plane tunneling capability.
S5	S-GW and P-GW	It provides user-plane tunneling and tunnel management between S-GW and P-GW. It is used for S-GW relocation due to UE mobility and if the S-GW needs to connect to a non-collocated P-GW for PDN connectivity and is based on GTP protocol.
S6a	MME and HSS	It enables transfer of subscription and authentication data for authenticating/authorizing user access to the network services.
Gx	PCRF and P-GW	It provides transfer of QoS policy and charging rules from PCRF to PCEF in the P-GW.
S8	S-GW and P-GW	Inter-PLMN reference point providing user-plane and control-plane between the S-GW in the visited PLMN and the P-GW in the home PLMN. S8 is the inter-PLMN variant of S5 and is based on GTP protocol.
S9	Visited PCRF and home PCRF	It provides transfer of QoS PCC information between the home PCRF and the visited PCRF in order to support local breakout function.
S10	MME and MME	It is used for MME relocation and MME to MME information transfer. This reference point can be used in the form of intra-PLMN or inter-PLMN in the case of inter-PLMN handover.
S11	MME and S-GW	It provides the reference point between MME and S-GW.
S12	UTRAN and S-GW	It provides user-plane tunneling when direct tunnel is established.
S13	MME and EIR	It enables UE identity check procedure between MME and equipment identity register (EIR). ^b

(Continued)

Table 2.1 (Continued)

Reference Point	Termination Points	Description
SGi	P-GW and PDN	It provides the interface between the P-GW and the packet data network. Packet data network may be an operator external public or private packet data network or an intra-operator packet data network as a provision for IMS services.
Rx	PCRF and PDN	This reference point resides between the application function (AF) and the PCRF.

^aGPRS tunneling protocol (GTP) is a group of IP-based communications protocols used to carry GPRS within GSM, UMTS, and LTE networks [35]. In 3GPP architecture, GTP and PMIPv6-based interfaces are specified on various reference points. GTP can be decomposed into separate protocols, GTP-C, GTP-U, and GTP'. GTP-C is used within the GPRS core network for signaling between gateway GPRS support node (GGSN) and serving GPRS support node (SGSN). This allows the SGSN to activate a session on a user's behalf (PDP context activation), to deactivate the same session, to adjust QoS parameters, or to update a session for a subscriber joined from another SGSN. GTP-U is used for carrying user data within the GPRS core network and between the radio access network and the core network. The user data transported can be packets in any of IPv4, IPv6, or PPP formats. The primary bearer plane in LTE across the access and the core networks is the evolved GPRS tunneling protocol (eGTP). The user-plane variant of this protocol eGTP-U provides tunneling support for user data between the eNB and the EPC elements. The eGTP-C (GTPv2-C) protocol is responsible for creating, maintaining, and deleting tunnels on multiple Sx interfaces. It is used for control-plane path management, tunnel management, and mobility management. It also controls forwarding relocation messages, serving radio network subsystem (SRNS) context and creating forward tunnels during inter-LTE handovers [9].

^bThe EIR is often integrated to the HLR. The EIR maintains a list of mobile phones (identified by their IMEI) which are banned from the network or need to be monitored. This is designed to allow tracking of stolen mobile phones. The EIR is a database that contains information about the identity of the mobile equipment that prevents calls from stolen, unauthorized or defective mobile stations [51].

single or multiple radio access networks, such as GERAN, UTRAN, or E-UTRAN. The core network, on the other hand, consists of the circuit-switched and packet-switched domains, where the latter includes GPRS core and EPC. These two domains differ in the way that they support user traffic and may include a number of common entities. A PLMN can implement only one domain or both domains. The circuit-switched domain refers to the set of core network entities providing circuit-switched type of connection for user applications and the related signaling. A circuit-switched connection is a connection for which dedicated network resources are allocated at the time of connection establishment and are released at the end of connection.

The packet-switched domain refers to the set of core network entities that provide a packet-switched type of connection for user applications and the related signaling. A packet-switched connection transports user information in the form of bit packets, where each packet can be routed independently from the others.

The network entities specific to the packet-switched domain are the GPRS support entities, i.e., SGSN and GGSN, and EPS-specific entities comprising P-GW, S-GW, and MME. The IMS domain encompasses those core network elements which enable provisioning and delivery of IP-based multimedia services, such as voice, audio, video, text, or chat over the packet-switched domain.

2.2.3 Packet data network gateway

The P-GW is the entity which terminates the SGi interface toward the packet data network. If a UE has multiple PDN connections, then the UE may be assigned to more than one P-GW. The P-GW provides PDN connectivity to GERAN/UTRAN capable UEs as well as E-UTRAN capable UEs using E-UTRAN, GERAN, or UTRAN access networks. The P-GW provides PDN connectivity to E-UTRAN capable UEs using E-UTRAN only over the S5/S8 interface. It may also provide PDN connectivity to UEs using non-3GPP access networks. The P-GW functions include per-user packet filtering; lawful interception; IP address allocation; transport-level packet marking in the uplink and downlink by setting the differentiated services code point based on the QCI of the associated EPS bearer; uplink/downlink service level charging, gating control, rate enforcement; uplink/downlink rate enforcement based on per access point name-aggregate maximum bit rate (APN-AMBR)¹ parameter; downlink rate enforcement based on the accumulated maximum bit rates of a set of SDFs with the same GBR QCI through rate policing/shaping; and dynamic host configuration protocol (DHCP)v4 (server and client) and DHCPv6 (client and server) functions.² In addition, the P-GW includes uplink/downlink bearer binding (bearer binding is the generic procedure for associating a bearer in the access network to an SDF) and uplink bearer binding verification functions for the GTP-based network interfaces. The P-GW functions also include user-plane anchor for mobility between 3GPP access and non-3GPP access.

¹The APN-AMBR is a subscription parameter stored per APN in the HSS. It limits the aggregate bit rate that can be provided across all non-GBR bearers and across all PDN connections of the same APN (e.g., excess traffic may be discarded by a rate shaping function). Each of those non-GBR bearers could potentially utilize the entire APN-AMBR, for example, when the other non-GBR bearers do not carry any traffic. The GBR bearers are outside the scope of APN-AMBR. The P-GW enforces the APN-AMBR in downlink. Enforcement of APN-AMBR in uplink is done in the UE and additionally in the P-GW [1].

²DHCP is a network protocol that is used to configure devices that are connected to a network so they can communicate on that network using the Internet protocol. It involves clients and a server operating in a client-server model. The clients request configuration settings using the DHCP protocol, such as an IP address, a default route, and one or more domain name system (DNS) server addresses. Once the client implements these settings, the host is able to communicate on the Internet. The DHCP server maintains a database of available IP addresses and configuration information. When the server receives a request from a client, the DHCP server determines the network to which the DHCP client is connected, and then allocates an IP address or prefix that is appropriate for the client, and sends configuration information appropriate for that client. DHCP servers typically grant IP addresses to clients only for a limited interval. DHCP is used for IPv4 and IPv6 [27,74].

2.2.4 Serving gateway

The S-GW is the entity which terminates the interface toward E-UTRAN access network. For each UE associated with the EPS, at any given point of time, there is a single S-GW assigned. The functions of the S-GW include a local mobility anchor point for inter-eNB handover; mobility anchor point for inter-3GPP mobility; EPS connection management (ECM)-IDLE mode downlink packet buffering and initiation of network-triggered service request procedure; lawful interception; packet routing and forwarding; transport-level packet marking in the uplink and downlink; and accounting for interoperator charging; as well as local non-3GPP anchor for the case of roaming when the non-3GPP IP access network is connected to the visited PLMN; event reporting (change of RAT) to the PCRF; uplink and downlink bearer binding toward 3GPP access; uplink bearer binding verification with packet dropping of misbehaving uplink traffic; mobile access gateway (MAG) functions, if proxy mobile IP (PMIP)³ based S5 or S8 is used; and support of necessary functions for enabling GTP/PMIP chaining functions [6,7].

2.2.5 Mobility management entity

The MME is the main control node that processes the NAS signaling between the UE and the core network which is responsible for idle-mode UE tracking and paging procedure as well as bearer establishment and release [10]. The main functions supported by the MME can be classified into two major categories: (1) functions related to bearer management including the establishment, maintenance, and release of the bearers which are handled by the session management function in the NAS protocol; and (2) functions related to connection management including the establishment of the connection and security association (SA) between the network and UE which is handled by the connection or mobility management function in the NAS protocol [11]. The MME is a control-plane entity within EPS which performs mobility management functions including NAS signaling and security; inter-node signaling for mobility between 3GPP access networks; tracking area list management; P-GW and S-GW selection; SGSN selection for handovers to 2G/3G access networks; roaming; and authentication. The MME's bearer management functions include dedicated bearer establishment. In order to enable 3GPP2 access (in a multi-RAT network), the MME should support high rate packet data (HRPD) access node selection and maintenance for handovers to HRPD and transparent transfer of HRPD signaling messages and transfer of status information between E-UTRAN and HRPD access networks [1].

³Network-based mobility management enables IP mobility for a host without requiring its participation in any mobility-related signaling. The network is responsible for managing IP mobility on behalf of the host. The mobility entities in the network are responsible for tracking the movements of the host and initiating the required mobility signaling on its behalf. IETF RFC 5213 specifies a network-based mobility management protocol that is referred to as proxy mobile IPv6 [26].

The MME creates a UE context when a UE is powered on and attaches to the network. It assigns a unique short temporary identity to the UE that 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 relaxes the need to frequently exchange messages with the HSS. In addition, the UE context also holds dynamic information, such as the list of bearers that are established and the terminal capabilities [70,71].

To reduce the overhead in the E-UTRAN and processing complexity in the UE, all UE-related information in the access network including the radio bearers can be released during long periods of user inactivity, which is known as ECM-IDLE state. The MME retains the UE context and the information about the established bearers during these idle periods. To allow the network to contact an idle-mode UE, the UE provides the network with location update whenever it moves out of its current tracking area using a tracking area update procedure. The MME is responsible for tracking user location while the UE is in the ECM-IDLE state. Upon availability of downlink data for an ECM-IDLE UE, the MME sends a paging message to all eNBs in the tracking area where the UE was recently located and the eNBs page the UE over the radio interface. Upon receipt of the paging message, the UE performs a service request procedure, which results in transitioning to the ECM-CONNECTED state. The UE-related information is thereby created in the E-UTRAN and the radio bearers are reestablished. The MME is responsible for the reestablishment of the radio bearers and updating the UE context in the eNB (see Chapters 4 and 5 for more details).

Security functions are the responsibility of the MME for both signaling and user data. When a UE attaches to the network, a mutual authentication of the UE and the network is performed between the UE and the MME/HSS. This authentication procedure establishes the security keys that are used for encryption of the bearers.

2.2.6 Policy and charging rules function

The PCRF acts as a policy decision point for policy and charging control (PCC) of SDFs/applications and IP bearer resources. It was introduced as an improvement in 3GPP Rel-7 network architecture in terms of policy and charging to allow optimization of interactions between the policy and rules functions. This improvement involved a new network node which integrates policy decision function (PDF) and charging rules function (CRF) in the previous releases. 3GPP Rel-8 further enhanced PCRF functionality by expanding the scope of PCC framework to facilitate non-3GPP access to the network (e.g., Wi-Fi or fixed broadband access). In the 3GPP generic PCC model, policy and charging enforcement function (PCEF), which resides in the P-GW, is the functional entity that supports SDF detection, policy enforcement, and flow-based charging. The PCRF provides the QoS authorization (QCI and bit rates) that decides how a certain data flow

will be treated in PCEF and ensures consistency with the user's subscription profile. The application function (AF) represents the network element that supports applications that require dynamic policy and/or charging control.

2.2.7 Home subscriber server

The home subscriber server is the subscription data repository for all permanent user data. The HSS stores the master copy of the subscriber profile, which contains information about the services that are applicable to the user. It also records the location of the user at the level of visited network control node, such as MME. It is the entity containing the subscription-related information to support the network entities that handle calls/sessions. A home network may contain one or several HSSs depending on the number of mobile subscribers, the capacity of the equipment, and the organization of the network. As an example, the HSS provides support to the call control servers in order to complete the routing/roaming procedures by solving authentication, authorization, naming/addressing resolution, location dependencies, etc. The HSS is responsible for holding information related to user identification, numbering and addressing information; user security information; network access control information for authentication and authorization; and user location information at the inter-system level. The HSS supports the user registration and stores inter-system location information as well as the user profile information. The HSS also generates user security information for mutual authentication, communication integrity check, and ciphering. Based on this information, the HSS is also responsible for supporting the call control and session management entities of the different domains and subsystems [1].

2.2.8 Authorization, authentication, and accounting

Trusted and untrusted non-3GPP access networks are IP-based access networks which may use access technologies whose specifications are out of the scope of 3GPP. Whether a non-3GPP access network is trusted or untrusted is not a characteristic of the access network. In non-roaming scenario, it is the home PLMN's operator decision if a non-3GPP IP-based access network is treated as trusted or untrusted. In a roaming scenario, the HSS/3GPP AAA server in home PLMN determines whether a non-3GPP IP access network is viewed as trusted or untrusted. The HSS/3GPP AAA server may consider and utilize the visited PLMN's policy and capability returned from the 3GPP AAA proxy or roaming agreement.

The 3GPP AAA server is located in the home PLMN and provides user support for non-3GPP access with authentication, authorization, and location management in order to allow access to the EPS. It also contains necessary user-related information in order to grant access to a non-3GPP network, and further coordinates the information needed to support mobility between 3GPP and non-3GPP access networks, such as coordination of P-GW information. It interacts with HSS

to maintain consistent information for users supporting mobility and service continuity between 3GPP and non-3GPP networks. The 3GPP AAA proxy provides support for roaming non-3GPP access users in the visited PLMN with the necessary authentication, authorization, and location management services in order to get access to the EPS. It may also provide roaming-related information for support of chaining scenarios. If an S-GW is needed for non-3GPP access in the visited network, the 3GPP AAA proxy selects an S-GW for the UE during initial attach or handover process [2].

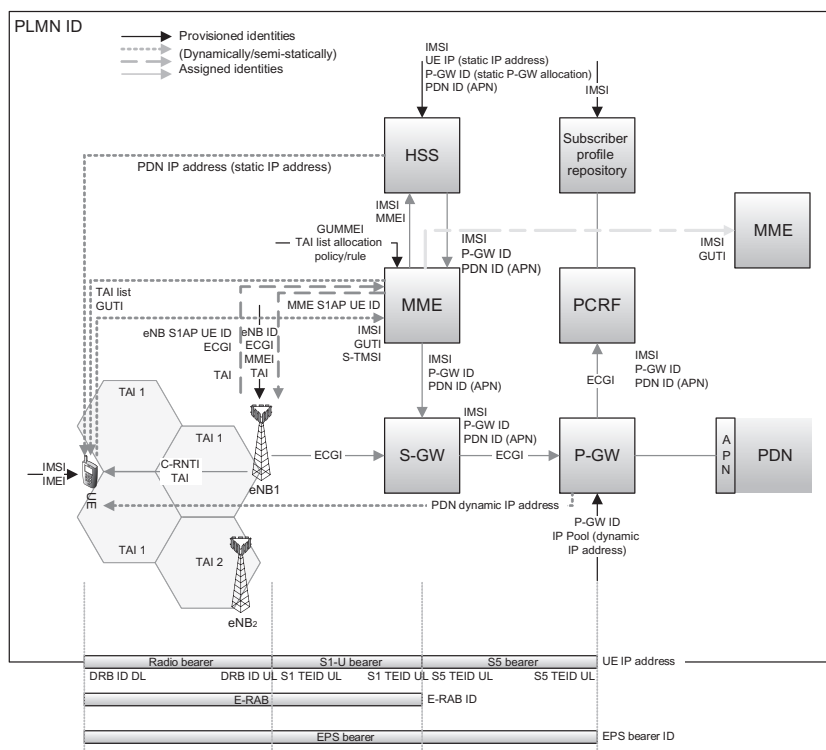
2.2.9 EPC identifiers

Each entity and bearer in E-UTRAN and EPC is identified with a unique identifier. The identifiers are either permanently provisioned, such as the IMSI and IMEI identities that are assigned by an operator to the UE or they are assigned during the lifetime of UE operation in an operator's network. Figure 2.10 illustrates the identities that are either provisioned or dynamically/semi-statically assigned to the bearers and the entities in EPS. These identities are described in Table 2.2 and the next chapters of this book and further specified by 3GPP in 3GPP TS 23.003 specification [2]. The direction of the arrows in the figure indicates the entity which assigns the identifier and the entity to which the identifier is assigned. The bearers and various bearer identifiers depending on the network interface and the flow direction are also shown in the figure.

Note that the protocols over LTE-Uu and S1 interfaces are divided into two categories: user-plane protocols, which are the protocols implementing the actual E-RAB service carrying user data through the access stratum (AS); and control-plane protocols that are the protocols for controlling the E-RABs and the connection between the UE and the network from different aspects including requesting the service, controlling different transmission resources, handover, etc. Furthermore, a mechanism for transparent transfer of NAS messages is included.

2.3 Roaming architecture and interworking with other networks

In cellular communications, roaming is a general term referring to the extension of service continuity and availability in a location that is different from the home location where the service and the subscriber were registered. Roaming functionality ensures that the wireless device is connected to the network without connection interruption while crossing the PLMN boundaries. The roaming is defined as the ability for a cellular user to automatically make or receive voice calls; send or receive data; or access other services including home data services, when moving outside the geographical coverage area of the home network using a visited network. The roaming feature is technically supported by mobility management,

**FIGURE 2.10**

EPS entities (bearers) and corresponding provisioned/assigned identities [40].

authentication, authorization, and billing functions in the network. While a roaming user is connected to the E-UTRAN, MME, and S-GW of the visited LTE network, the EPS allows the P-GW of either the visited or the home network to be used, as shown in Figure 2.11. Using the home network's P-GW allows the user to access the home operator's services while in a visited network. A P-GW in the visited network allows a local breakout to the Internet in the visited network [70,71].

The EPS also supports interworking with and handover to/from other networks using 3GPP/non-3GPP radio access technologies, such as GSM, UMTS, and CDMA2000. The architecture for interworking with 2G/3G GPRS/UMTS networks is shown in Figure 2.12. The S-GW acts as the mobility anchor for interworking 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 mobile WiMAX. The P-GW may also support a PMIP-based interface.

Table 2.2 Description of EPC and E-UTRAN Identities [40,41]

Abbreviation	Identifier	Description	Value/Structure	Permanent/ Temporary
IMSI	International mobile subscriber identity	Unique identification of mobile subscriber in the network. The IMSI is composed of three parts: (1) mobile country code (MCC) consisting of three digits. The MCC uniquely identifies the country of residence of the mobile subscriber; (2) mobile network code (MNC) consisting of two or three digits for GSM/UMTS applications. The MNC identifies the home PLMN of the mobile subscriber. The length of the MNC (two or three digits) depends on the value of the MCC. A mixture of two and three digit MNC codes within a single MCC area is not recommended and is outside the scope of this specification; and (3) mobile subscriber identification number (MSIN) identifying the mobile subscriber within a PLMN.	IMSI (not more than 15 digits) = PLMN ID + MSIN = MCC + MNC + MSIN	Permanent
PLMN ID	Public land mobile network identifier	Unique identification of PLMN	PLMN ID (not more than 6 digits) = MCC + MNC	Permanent
MCC	Mobile country code	Assigned by ITU	3 digits	Permanent
MNC	Mobile network code	Assigned by national authority	2 or 3 digits	Permanent
MSIN	Mobile subscriber identification number	Assigned by operator	9 or 10 digits	Permanent

GUTI	Globally unique temporary UE identity	Identifies a UE (between the UE and the MME) on behalf of IMSI for security reasons. The purpose of the GUTI is to provide an unambiguous identification of the UE that does not reveal the UE or the user's permanent identity in the EPS. It also allows the identification of the MME and network. It can be used by the network and the UE to establish the UE's identity during signaling in the EPS.	GUTI (not more than 80 bits) = GUMMEI + M-TMSI	Temporary
TIN	Temporary identity used in next update	GUTI is stored in TIN parameter of UE's context. TIN indicates which temporary ID will be used in the next update	TIN = GUTI	Temporary
S-TMSI	SAE temporary mobile subscriber identity	To locally identify a UE in short-form within an MME group (unique within an MME pool). The S-TMSI is the shortened form of the GUTI to enable more efficient radio signaling procedures (e.g., paging and service request). For paging purposes, the mobile is paged with the S-TMSI.	S-TMSI (40 bits) = MMEC + M-TMSI	Temporary
M-TMSI	MME mobile subscriber identity	Unique within an MME Since the TMSI has only local significance (i.e., within a VLR and the area controlled by a VLR, or within an SGSN and the area controlled by an SGSN, or within an MME and the area controlled by an MME), the structure and coding of it can be chosen by agreement between operator and manufacturer in order to meet local needs.	32 bits	Temporary
GUMMEI	Globally unique MME identity	To identify an MME uniquely in a global sense.	GUMMEI (not more than 48 bits) = PLMN ID + MMEI	Permanent
MMEI	MME identifier	To identify an MME uniquely within a PLMN	MMEI (24 bits) = MMEGI + MMEC	Permanent

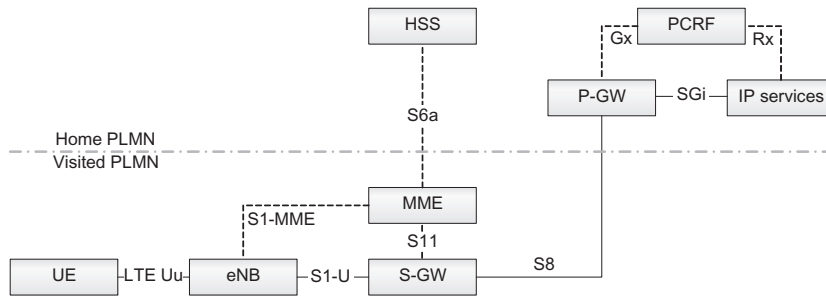
(Continued)

Table 2.2 (Continued)				
Abbreviation	Identifier	Description	Value/Structure	Permanent/ Temporary
MMEGI	MME group identifier	Unique identification of an MME group within a PLMN	16 bits	Permanent
MMEC	MME code	To identify an MME uniquely within an MME group	8 bits	Permanent
IMEI	International mobile equipment identity	To uniquely identify a mobile equipment	IMEI (15 digits) = type approval code (6 digits) + serial number (6 digits) + final assembly code (2 digits) + check digit (1 digit)	Permanent
IMEI/SV	IMEI/software version	Identifies the mobile equipment serial number, hardware model number, and software version number	IMEI/SV (16 digits) = type approval code (6 digits) + serial number (6 digits) + software version (2 digits) + final assembly code (2 digits)	Permanent
ECGI	E-UTRAN cell global identifier	To identify a cell in general (globally unique) EPC can know UE location based on ECGI	ECGI (not more than 52 bits) = PLMN ID + ECI	Permanent
ECI	E-UTRAN cell identifier	To identify a cell within a PLMN	ECI (28 bits) = eNB ID + cell ID	Permanent
Global eNB ID	Global eNB identifier	To globally identify an eNB	PLMN ID + eNB ID	Permanent
eNB ID	eNB identifier	To identify an eNB within a PLMN	20 bits	Permanent
P-GW ID	PDN GW identity	To identify a specific P-GW HSS assigns P-GW for PDN (IP network) connection of each UE	IP address (4 bytes) or fully qualified domain name (variable length)	Permanent
TAI	Tracking area identity	This is the identity used to uniquely identify the tracking areas. The tracking area identity is constructed from the MCC (mobile country code), MNC (mobile network code) and TAC (tracking area code).	TAI (not more than 32 bits) = PLMN ID + TAC P-GW	Permanent

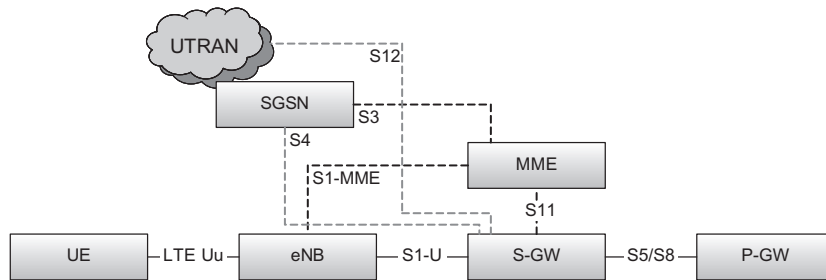
TAC	Tracking area code	To indicate to which tracking area the eNB belongs (per cell) which is unique within a PLMN	16 bits	Permanent
TAI list	Tracking area identity list	UE can move into the cells included in TAI list without location update (TA update) which is globally unique	Variable length	Permanent
PDN ID	Packet data network identity	To identify a PDN (IP network) with which mobile data user communicates PDN identity (APN) is used to determine the P-GW and point of interconnection with a PDN With APN as query parameter to the DNS procedures, the MME will receive a list of candidate P-GWs, and then a P-GW is selected by MME	PDN identity = APN = APN-NI + APN-OI (variable length) APN-NI: APN network identifier APN-OI: APN operator identifier	Permanent
CSG ID	Closed subscriber group ID	A closed subscriber group (CSG) consists of a single cell or a collection of cells within the E-UTRAN and UTRAN that are open to only a certain group of subscribers. A CSG ID is a unique identifier within the scope of PLMN which identifies a CSG in the PLMN associated with a CSG cell or group of CSG cells.	27 bits	Permanent
EPS bearer ID	Evolved packet system bearer identifier	To identify an EPS bearer (default or dedicated) per UE. An EPS bearer identity uniquely identifies an EPS bearer for one UE accessing via E-UTRAN. The EPS bearer identity is allocated by the MME. There is a one-to-one mapping between EPS radio bearer and EPS bearer, and the mapping between EPS radio bearer identity and EPS bearer identity is made by E-UTRAN. The E-RAB ID value, used over S1 and X2 interfaces to identify an E-RAB, is the same as the EPS bearer ID value used to identify the associated EPS bearer.	4 bits	Temporary

(Continued)

Table 2.2 (Continued)				
Abbreviation	Identifier	Description	Value/Structure	Permanent/ Temporary
E-RAB ID	E-UTRAN radio access bearer identifier	To identify an E-RAB per UE	4 bits	Temporary
DRB ID	Data radio bearer identifier	To identify a DRB per UE	4 bits	Temporary
LBI	Linked EPS bearer ID	To identify the default bearer associated with a dedicated EPS bearer	4 bits	Temporary
TEID	Tunnel endpoint identifier	To identify the endpoint of a GTP tunnel when the tunnel is established	32 bits	Temporary
eNB S1AP UE ID	eNB S1AP UE identity	This is a temporary identity used to identify a UE over the S1-MME reference point within the eNB and is unique within the eNB.	32 bits	Temporary
MME S1AP UE ID	MME S1AP UE identity	This is a temporary identity used to identify a UE on the S1-MME reference point within the MME and is unique within the MME.	32 bits	Temporary
C-RNTI	Cell-radio network temporary identifier	To identify a UE uniquely in a cell	0 × 0001 ~ 0 × FFF3 (16 bits)	Temporary

**FIGURE 2.11**

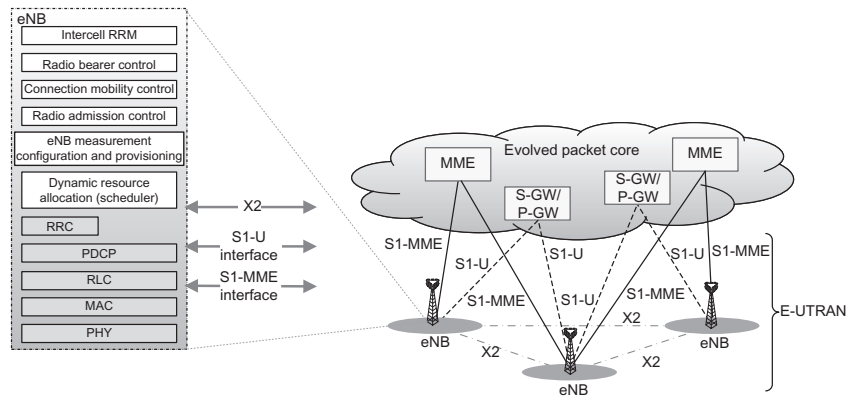
Example roaming architecture [1,70,71].

**FIGURE 2.12**

Interworking of LTE with UMTS network architecture [70,71].

2.4 E-UTRAN architecture

The E-UTRAN access network consists of a network of eNBs, as illustrated in Figure 2.13. For unicast user traffic, there is no centralized controller in the E-UTRAN; thus, the E-UTRAN architecture is considered as a distributed architecture. The eNBs are typically interconnected with each other by means of a logical interface known as X2 and to the EPC by means of the S1 interface. More specifically, the eNB is connected to the MME through S1-MME interface and to the S-GW via S1-U interface. The protocols between the eNBs and the UE are referred to as the AS protocols. The protocols between the EPC and the UE are known as NAS protocols. The E-UTRAN is responsible for all radio-related functions that include RRM, encompassing all functions related to the radio bearers, such as radio bearer control (RBC), radio admission control (RAC), radio mobility control, scheduling and dynamic allocation of radio resources to UEs in both uplink and downlink; header compression, ensuring efficient use of the radio interface by compressing the IP packet headers that could otherwise impose a

**FIGURE 2.13**

E-UTRAN architecture and eNB functional decomposition [12].

significant overhead especially for small-packet applications, such as VoIP; security, encrypting all data and control messages sent over the radio interface; and connectivity to the EPC, consisting of the signaling toward MME and the bearer path toward the S-GW [70,71].

As shown in Figure 2.13, on the network side, all of these functions reside in the eNBs, each of which can be responsible for managing multiple cells. Unlike previous 2G/3G technologies, LTE has integrated the radio controller function into the eNB. This allows close interaction between different protocol layers of the radio access network, thus reducing latency and improving efficiency. Such distributed control structure eliminates the need for a processing-intensive central controller, which in turn has the potential to reduce costs, overhead, latency, and complexity. Given that LTE does not support soft handover, there is no need for a centralized data-combining function in the network. One consequence of the absence of a centralized controller node is that, as the UE moves, the network must transfer all information related to a UE (the UE context) together with any buffered data from one eNB to another. Therefore, there are mechanisms in EPS that are provisioned to avoid data loss during handover (e.g., data forwarding).

An important feature of the S1 interface linking the access network to the core network is known as S1-flex [6]. This is a concept whereby multiple core network nodes (MMEs/S-GWs) can serve a common geographical area, being connected by a mesh network to the set of eNBs in that area. An eNB may be served by multiple MMEs/S-GWs. The set of MME/S-GW nodes that serve a common area is called an MME/S-GW pool, and the area covered by such a pool of MMEs/S-GWs is called a pool area. This concept allows the UEs in the cell or cells controlled by one eNB to be shared among multiple core network nodes, thereby providing a possibility for load balancing (LB) and also eliminating a single point

of failure for the core network nodes. The UE context typically remains in the same MME as long as the UE is located within the same pool area.

2.4.1 E-UTRAN functions

The functions performed and the services provided by an eNB include RRM (RBC, RAC, connection mobility control (CMC), dynamic allocation and scheduling of resources for the UEs in the uplink and downlink); IP header compression and encryption of user data stream; selection of an MME upon UE attachment to the network when no routing to an MME can be determined from the information provided by the UE; routing of user-plane data toward S-GW; scheduling and transmission of paging messages originated from the MME; scheduling and transmission of broadcast information originated from the MME or network operation and management; measurement and measurement reporting configuration for mobility and scheduling; scheduling and transmission of emergency warning messages originated from the MME; closed-subscriber group (femto-cell) handling; and transport-level packet marking in the uplink. A relay-enabled eNB further performs S1/X2 proxy functionality for supporting the relay nodes (RNs) and S11 termination and S-GW/P-GW functionality for supporting the relay nodes [12]. The functional decomposition of an eNB is shown in Figure 2.13.

2.4.2 E-UTRAN interfaces

This section provides general information about the E-UTRAN interfaces (S1 and X2). The E-UTRAN network interface model consists of two main parts: the transport network layer, corresponding to the transport of the access-network data, and the radio network layer, which encompasses the top-level protocols of the radio interface. In addition, each interface has a user-plane and a control-plane component. The user-plane carries all user data as well as application layer signaling, such as session initiation protocol (SIP) or real-time transport protocol (RTP) control protocol (RTCP) packets [22]. The control-plane handles all messages and procedures related to the radio interface supported features. An example of these messages includes control messages for handover management or bearer establishment. The physical layer is common to both user-plane and control-plane. The user-plane and control-plane use specific protocols which allow definition of a different and independent transport stack and bearers for each of the planes.

The control-plane information is more susceptible to security, reliability, and information loss, whereas the user-plane information can rely on simpler and less secured routing protocols. The E-UTRAN network interfaces are open standards defined by 3GPP. This conformity to open standards allows different vendors to manufacture eNBs, deploy them in a single network, and interconnect them over the X2 interface or with MME or S-GW nodes over the S1 interface.

The S1 interface connects the eNB to the EPC. It can be split into control-plane (S1-MME) and user-plane (S1-U). The S1-MME is a signaling interface which

supports a set of functions and procedures between the eNB and the MME. All S1-MME signaling procedures belong to four main groups: (1) bearer-level procedures, corresponding to all procedures related to bearer setup, modification and release, which are typically used during the establishment or the release of a session; (2) handover procedures, which encompass all S1 functions related to user mobility between eNB or with 2G/3G technologies; (3) NAS signaling transport, corresponding to the transport of UE-MME NAS signaling over the S1 interface; and (4) paging procedure, which is used in case of a mobile-terminated call/session. Through the paging procedure, the MME requests the eNB to page a UE in a given set of cells known as a tracking area. The S1-MME interface provides a high level of reliability in order to avoid message retransmission and unnecessary delay in control-plane procedures using stream control transmission protocol (SCTP) [20].

The role of S1-U is to transport user data packets between the eNB and the S-GW. This interface makes use of a very simple GTP over user datagram protocol (UDP)/IP transport protocol which only provides user data encapsulation. There is no flow control or error control, or any mechanism to guarantee data delivery over the S1-U interface. The GTP was inherited from GPRS and UMTS networks. In GPRS networks, GTP is used between the SGSN and the GGSN. In UMTS networks, GTP is also used over the Iu-ps interface between the RNC and the SGSN. S1 application part (S1AP) is the control-plane signaling protocol between the eNB and the MME [13]. It supports the S1-MME interface and utilizes SCTP in the transport layer. The SCTP protocol is for the control-plane, which guarantees delivery of control/signaling messages between the MME and the eNB.

The X2 interface connects multiple eNBs. LTE uses the same protocol structure over both S1 and X2 interfaces, which simplifies the data forwarding operation. The X2 interface has a control-plane (X2-C) and a user-plane (X2-U) component. The X2-U is responsible for the transport of user data packets between eNBs. This interface is only used for a limited period of time, when the terminal moves from the coverage area of one eNB to another and provides buffered packet data forwarding. The X2-U interface utilizes GTP tunneling protocol that is used over the S1-U interface. The X2-C is a signaling interface which supports a set of functions and procedures between eNBs. The X2-C procedures are very limited in number and are all related to user mobility between eNBs, exchange of UE context between nodes including allocated bearers, security information, etc. The need for reliable transport of signaling between nodes results in the use of SCTP protocol for X2-C. X2AP is the control-plane protocol between eNBs over the X2 interface and supports load management and handover coordination between eNBs.

2.4.3 E-UTRAN identifiers

The E-UTRAN (dynamically) assigns unique identities to the UEs at the cell level. These identities temporarily identify the UE resource allocations and

assignments as well as UE-specific control signaling during the active state of the UE in the cell. These identities include C-RNTI, which is a unique identification used for identifying RRC connection and scheduling; semi-persistent scheduling C-RNTI that is a unique identification used for semi-persistent scheduling; temporary C-RNTI, TPC-PUSCH-RNTI that is an identification used for the power control of physical uplink shared channel; TPC-PUCCH-RNTI, which is an identification used for the power control of physical uplink control channel; and random access RNTI for random access procedure and contention resolution. During some transient intervals, the UE is temporarily identified with a random value used for contention resolution purposes.

As shown in [Figure 2.11](#), the E-UTRAN further utilizes unique identities for identifying a specific network entity which include GUMMEI that is used to globally identify the MME. The GUMMEI is constructed from the PLMN identity to which the MME belongs, the group identity of the MME group where the MME is part of, and the MMEC of the MME within the MME group. Note that the GUMMEI or S-TMSI containing the MMEC is provided by the UE to the eNB. The E-UTRAN cell global identifier (ECGI) is used to globally identify the cells. The ECGI is constructed from the PLMN identity where the cell belongs to and the cell identity (CI) of the cell. The PLMN information is periodically broadcasted as part of system information. The eNB identifier (eNB ID) is used to identify eNBs within a PLMN and is contained within the CI of its cells. The global eNB ID is used to globally identify the eNBs, and is constructed from the PLMN identity to which the eNB belongs and the eNB ID. The MCC and MNC are the same as that included in the ECGI (see [Table 2.2](#)).

In addition to the above identities, there are other identifiers related to the support of home eNBs (HeNBs) or femto-cells and relays. The global eNB ID of a relay node is the same as its serving donor eNB (DeNB), which is a relay-enabled eNB. The tracking area identity is used to identify tracking areas and is constructed from the PLMN identity where the tracking area belongs to, and the TAC of the tracking area. The CSG identity (CSG ID) is an identifier of a CSG femto-cell within a PLMN.

As shown in [Figure 2.11](#), there are unique identifiers that are used to identify the data and signaling radio bearers (SRBs) as well as S1 and S5/S8 bearers on the access network that are assigned to UE services. The value of the E-RAB ID used on S1 and X2 interfaces to identify an E-RAB allocated to the UE is the same as the EPS bearer ID value used at the LTE-Uu interface to identify the associated EPS bearer and the NAS layer. The CI, TAC, CSG ID (if any), and one or more PLMN IDs are broadcast periodically in every E-UTRAN cell as part of system information.

2.4.4 Support of HeNB in E-UTRAN

Small cells are low-power wireless access points that operate in licensed spectrum and are managed by an operator. These cells provide improved cellular coverage

and capacity and have applications for homes and enterprises as well as metropolitan public areas. There are various types of small cells including femto-cells, pico-cells, metro-cells and micro-cells, broadly increasing in size from femto-cells (the smallest) to micro-cells (the largest). Small cells also facilitate a new class of mobile services that exploit the technology's ability to detect the presence of, and connect and interact with existing networks [42,56,60].

As shown in Figure 2.14, the E-UTRAN architecture supports deployment of a home eNB gateway (HeNB-GW) to allow an S1 interface between the HeNB and the EPC in order to support a large number of HeNBs in a scalable manner. The HeNB-GW serves as a concentrator for the control-plane, specifically the S1-MME interface. The S1-U interface from the HeNB may be terminated at the HeNB-GW or a direct logical user-plane connection between HeNB and S-GW may be used. Depending on the deployment scenario, the S1 interface in this case is defined as the interface between the HeNB-GW and the core network; between the HeNB and the HeNB-GW; between the HeNB and the core network; or between the eNB and the core network. The HeNB-GW appears to the MME as an eNB, and to the HeNB as an MME. The S1 interface between the HeNB and the EPC is the same, regardless of the HeNB connection to the EPC via a HeNB-GW [12].

The HeNB-GW connects to the EPC such that the inbound and outbound mobility to cells served by the HeNB-GW does not necessarily require inter-MME handovers. One HeNB serves only one cell. In general, the functions supported by the HeNB are the same as those supported by an eNB and the procedures activated between a HeNB and the EPC are similar to those between an eNB and the EPC. Therefore, it can be concluded that HeNB is a (reduced-function and reduced-cost)

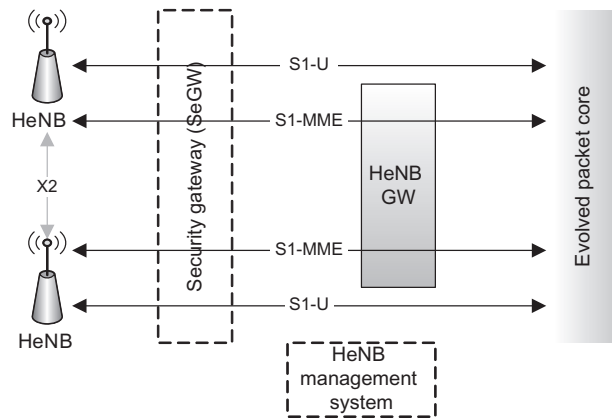


FIGURE 2.14

HeNB logical network architecture [12].

eNB that is installed by the user and not the operator, but is connected to operator's network in the same way that a regular eNB is connected.

3GPP Rel-11 specifications support direct X2-connectivity between HeNBs, independent of whether any of the involved HeNBs are connected to a HeNB-GW [12].

2.4.5 Support of RRHs in E-UTRAN

One recent deployment trend that is foreseen to be widely applied to LTE networks consists of splitting the base station functionalities into a baseband unit, which performs, in particular, the scheduling and the baseband processing functions, and a number of RRHs responsible for all RF transmission/reception aspects. The baseband processing unit is typically located at the center of the cell and is connected via optical fiber to the RRHs. In addition to suppressing the feeder loss (since the power amplifier is immediately close to the antenna), this approach allows a baseband processing unit to manage different radio sites in a central manner. Furthermore, having geographically separated RRHs controlled from the same location enables either centralized baseband processing units jointly managing the operation of several radio sites, or the exchange of very low-latency coordination messages among individual baseband processing units.

Distributed base stations with RRHs can significantly assist mobile operators to overcome cost, performance, and efficiency issues when addressing increasing demand for higher capacities in cellular networks. Multimode radios operating according to GSM, HSPA, LTE, and Wi-Fi standards and advanced software configurability are the key features in the deployment of more flexible and energy-efficient radio networks. Wireless and mobile network operators face the continuing challenge of upgrading their networks to effectively manage high user traffic rates. Mobility and an increased level of multimedia content for end users require end-to-end network adaptations that support both new services and the increased demand for broadband and flat-rate Internet access. In addition, network operators must consider the most cost-effective evolution of the networks toward new generations. Wireless and mobile technology standards are evolving toward higher bandwidth requirements for both peak data rates and cell throughputs. The latest standards HSPA+, mobile WiMAX, and LTE/LTE-Advanced can barely satisfy today's demand for high-performance mobile applications and fast and reliable access to the Internet anywhere and anytime. The network upgrades required to deploy networks based on these standards must balance the limited availability of new spectrum, leverage existing spectrum, and ensure proper operation of the desired wireless standards.

Distributed open base station architecture concepts have evolved in parallel with the evolution of the standards to provide a flexible, low-cost, and scalable modular environment for managing the radio access. For example, the open base

station architecture initiative (OBSAI)⁴ and the common public radio interface (CPRI)⁵ standards introduced standardized interfaces separating the base station server and the RRH part of a base station by an optical fiber.

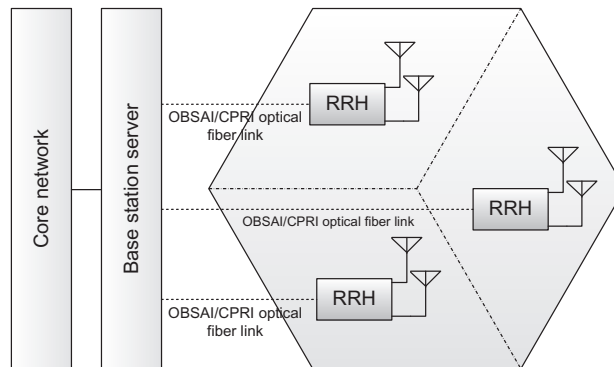
The RRH concept constitutes a fundamental part of a state-of-the-art base station architecture. The RRH-based system implementation is driven by the need to reduce both capital expenditure (CAPEX) and operational expenditure (OPEX) consistently, which allows an optimized and energy-efficient deployment. Figure 2.15 illustrates an architecture where 2G/3G/4G base stations are connected to RRHs over optical fibers. Either CPRI or OBSAI may be used to carry RF signals to the RRH to cover a three-sector cell. The RRH incorporates a large number of digital interfacing and processing functions. It also includes high-performance, efficient, and frequency-agile analog functions, all packaged into a light device with a small mechanical footprint.

2.4.6 Support of relays in E-UTRAN

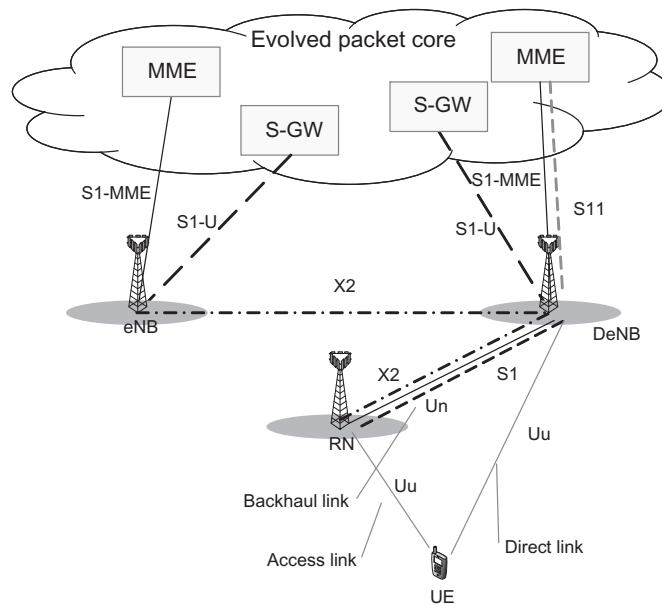
In general, RNs can be used to improve urban or indoor user throughput as well as to extend coverage in rural areas. The concept of relaying is not new but the level of sophistication continues to be enhanced. Figure 2.16 shows a typical scenario where an RN is connected wirelessly to the radio access network via a donor cell. The RN can connect to the Donor cell's eNB (DeNB) in one of two ways: in-band (in-channel), in which case the DeNB-to-RN link shares the same carrier frequency with RN-to-UE links, and out-band, in which case the DeNB-to-RN link does not operate on the same carrier frequency as RN-to-UE links. The most basic relay method is the use of a radio repeater, which receives, amplifies, and then retransmits the downlink and uplink signals to overcome areas of poor coverage. The repeater

⁴The OBSAI family of specifications provides the architecture, functional description, and minimum requirements for integration of a set of common modules into a base transceiver station. It defines an open, standardized internal modular structure of wireless base stations; a set of standard BTS modules with specified form, fit, and function such that BTS vendors can acquire and integrate modules from multiple vendors in an OEM fashion; defines open, standards-based internal digital interfaces between BTS modules to ensure interoperability and compatibility; and supports different access technologies, such as GSM/EDGE, CDMA2000, WCDMA, LTE/LTE-Advanced, or mobile WIMAX [75].

⁵The CPRI is an initiative to define a publicly available specification that standardizes the protocol interface between the radio equipment control (REC) and the radio equipment (RE) in wireless base stations. This allows interoperability of equipments from different vendors and preserves the software investment made by wireless service providers. Conventional base stations are located adjacent to the antenna in a small compartment at the base of the antenna tower. Finding suitable sites can be a challenge because of the footprint required for the compartment, the need for structural reinforcement of rooftops, and the availability of both primary and backup power sources. The CPRI allows the use of a distributed architecture where base stations, containing the REC, are connected to remote radio heads via low-loss fiber links that carry the CPRI data. This architecture reduces costs for service providers because only the remote radio heads containing the RE need to be situated in environmentally challenging locations. The base stations can be centrally located in less problematic locations where footprint, climate, and availability of power are easily managed [38,44].

**FIGURE 2.15**

Example three-sector cell using RRH [38].

**FIGURE 2.16**

Support of relay nodes in E-UTRAN architecture [12].

can be located at the cell-edge or in some other areas of poor coverage. Radio repeaters are relatively simple devices operating purely at the RF level. Typically they receive and retransmit an entire frequency band, so they must be sited carefully. In general, repeaters can improve coverage but do not substantially increase capacity. In an in-band relaying scenario and due to self-interference, the RNs cannot simultaneously transmit on the access link and receive on the backhaul link as well as they cannot concurrently receive on the access and transmit on the backhaul link.

E-UTRAN supports relaying functionality by connecting an RN to a serving eNB, known as the donor eNB, via a modified version of the E-UTRA radio interface, where the modified version is denoted as the Un interface. The RN supports most of the eNB functionalities, implying that it terminates the radio protocols of the E-UTRA radio interface on the access link as well as the S1 and X2 interfaces on the backhaul link. In addition to the eNB functionality, the RN also supports a subset of the UE functionality, e.g., physical layer, layer-2, RRC, and NAS functionality, in order to wirelessly connect to the DeNB. An RN may not use another RN as its DeNB and intercell handover of the RN is not supported.

More advanced relays at layer 2 can decode transmissions before retransmitting them. Traffic can then be forwarded selectively to the UE by the RN, thus minimizing the interference created by legacy relays that forwarded all traffic. Depending on the level at which the protocol stack is terminated in the RN, such types of relay may require the development of relay-specific standards. This can be avoided by extending the protocol stack of the RN up to layer 3 to create a wireless router that operates in the same way that a normal eNB operates, using standard air-interface protocols and performing its own resource allocation and scheduling.

The modified E-UTRAN architecture to support RNs is shown in Figure 2.16. The RN terminates the S1, X2, and Un interfaces. The DeNB provides S1 and X2 proxy functionality between the RN and other eNBs, MMEs, and S-GWs. The S1 and X2 proxy functionality includes forwarding UE-specific S1 and X2 signaling messages as well as GTP data packets between the S1 and X2 interfaces associated with the RN and the S1 and X2 interfaces associated with other network nodes. Due to the proxy functionality, the DeNB appears as an MME (for S1-MME), an eNB (for X2) and an S-GW (for S1-U) to the RN. In further development of relay specifications in 3GPP, the DeNB also implements and provides the S-GW/P-GW functions to the associated RNs. This includes creating a session for the RN and managing EPS bearers for the RN, as well as terminating the S11 interface toward the MME serving the RN. The RN and DeNB also map signaling and data packets to the EPS bearers that are set up for the RN. The mapping is based on the existing QoS mechanisms already defined for the UE and the P-GW. The P-GW functions in the DeNB allocate an IP address for the RN for the operation and management functions which may be different than the S1 IP address of the DeNB [12].

The S1 and X2 user-plane packets are mapped to radio bearers over the Un interface. The mapping can be based on the QCI associated with the UE's EPS bearers. The UE's EPS bearers with similar QoS can be mapped to the same Un radio bearer. There is a unique S1 interface relationship between each RN and its DeNB, and

between the DeNB and each MME in the MME pool. The DeNB processes and forwards all S1 messages between the RN and the MMEs for all UE-dedicated procedures. The processing of S1AP messages includes modifying S1AP UE IDs, transport layer address and GTP TEIDs, leaving other parts of the message unchanged [12].

All non-UE-dedicated S1AP procedures are terminated at the DeNB, and are handled locally between the RN and the DeNB, and between the DeNB and the MME(s). Upon reception of an S1 non-UE-dedicated message from an MME, the DeNB may trigger corresponding S1 non-UE-dedicated procedure(s) to the RN(s). If more than one RN is involved (in a multi-hop scenario), the DeNB may wait and aggregate the response messages from all involved RNs before responding to the MME. Upon reception of an S1 non-UE-dedicated message from an RN, the DeNB may trigger associated S1 non-UE-dedicated procedure(s) to the MME(s). There is a single X2 interface relation between each RN and its DeNB. In addition, the DeNB may have X2 interface connection to neighboring eNBs. The DeNB processes and forwards all X2 messages between the RN and other eNBs for all UE-dedicated procedures. The processing of X2AP messages includes modifying S1/X2AP UE IDs, transport layer address, and GTP TEIDs, leaving other parts of the message intact [12].

All non-UE-dedicated X2AP procedures are terminated at the DeNB, and are handled locally between the RN and the DeNB, and between the DeNB and other eNBs. Upon reception of an X2 non-cell-related non-UE-associated message from RN or neighbor eNB, the DeNB may trigger associated non-UE-dedicated X2AP procedure(s) to the neighbor eNB or RN(s). Upon reception of an X2 cell-related non-UE-dedicated message from RN or neighbor eNB, the DeNB may pass associated information to the neighbor eNB or RN(s) based on the included cell information. If one or more RN(s) are involved, the DeNB may wait and aggregate the response messages from all participating nodes to respond to the originating node. The S1 and X2 interface signaling packets are mapped to radio bearers over the Un interface. The RN connects to the DeNB via the Un interface using the same radio protocols and procedures as a UE connecting to an eNB [12].

2.5 E-MBMS architecture

Enhanced multimedia broadcast multicast service (E-MBMS) is a point-to-multi-point service in which data is transmitted from a single source to multiple destinations over the radio network. Transmitting the same data to multiple recipients allows network resources to be shared. E-MBMS is realized by addition of new functional entities to 3GPP network architecture. In practice, E-MBMS provides two different services: broadcast and multicast.

The broadcast service can be received by any subscriber located in the area in which the service is offered and multicast services can only be received by users subscribed to the service and joined to the multicast group associated with the service. Both services are unidirectional point-to-multi-point transmissions of

multimedia data, text, audio, video, news, etc., from a broadcast multicast service center (BM-SC) to any user located in the service area. For such a service, only the broadcast service providers can be charged, possibly based on the amount of broadcast data size of service area, or broadcast service duration. Multicast is subject to service subscription and requires the end user to explicitly join the group in order to receive the service. Because it is subject to subscription, the multicast service allows the operator to set specific user charging rules for this service.

The E-MBMS content is typically broadcast over a large number of cells that transmit the same data and are required to be synchronized. In that case, the resulting signal will appear to a UE as one transmission over a time-disperse radio channel. This concept is known as multimedia broadcast single frequency network (MBSFN). A group of cells that transmit the same content to multiple users form an MBSFN area. Multiple cells can belong to a single MBSFN area and each cell can be part of up to eight MBSFN areas. There can be up to 256 different MBSFN areas defined in a network, each one with a unique identity. Once defined, MBSFN areas will not change dynamically. It is not required that a terminal simultaneously receive content from multiple MBSFN areas. In other words, an MBSFN area is a geographical area comprising one or several cells that transmit the same content. For example, in Figure 2.17, which illustrates a hypothetical MBSFN network, cells 2 and 3 belong to both MBSFN areas 1 and 2. The MBSFN areas are static and do not vary over time. The usage of MBSFN transmission requires not only time synchronization among the cells participating in an MBSFN area, but also usage of the same set of radio resources in each of the cells for a particular service. This coordination is the responsibility of the multi-cell/multicast coordination entity (MCE), which is a logical node in the radio access network that controls resource allocation and transmission parameters across the cells in the MBSFN area [12,30–33].

Figure 2.18 illustrates the E-MBMS network architecture. As shown in the figure, the MCE can control multiple eNBs, each controlling one or more cells. The BM-SC, located in the core network, is responsible for authorization and authentication of content providers, charging, and overall configuration of the data flow through the core network. The MBMS gateway (MBMS-GW) is a logical node handling multicast of IP packets from BM-SC to all eNBs involved in transmission in an MBSFN area. It also manages session control signaling via the MME. From the BM-SC, the MBMS data is forwarded using IP-based multicast, a method of sending IP packets to multiple receiving network nodes in a single transmission, via the MBMS-GW to the cells. The MBMS is not only efficient from a radio-interface point of view, but also it saves resources in the transport network by avoiding unicast transmission of the same packet to multiple nodes unless necessary. This can lead to significant savings in the transport network.

More specifically, an MBMS-enabled network consists of the following entities and interfaces [12,32]:

- **MCE:** The MCE is a logical entity (this does not preclude the possibility that it may be part of another network element), whose functions are admission

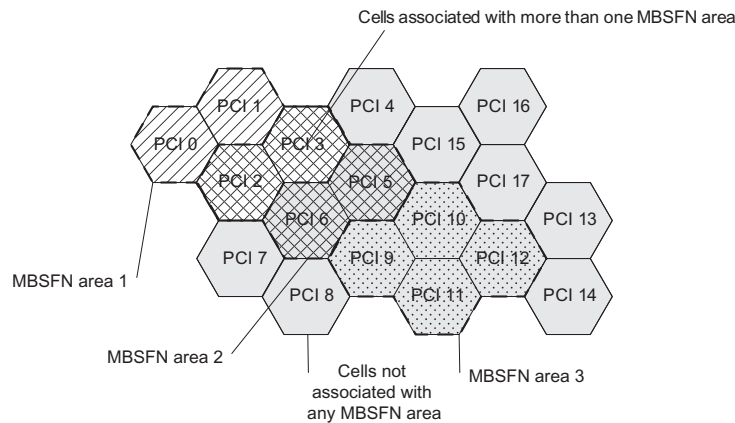


FIGURE 2.17
Illustration of MBSFN areas.

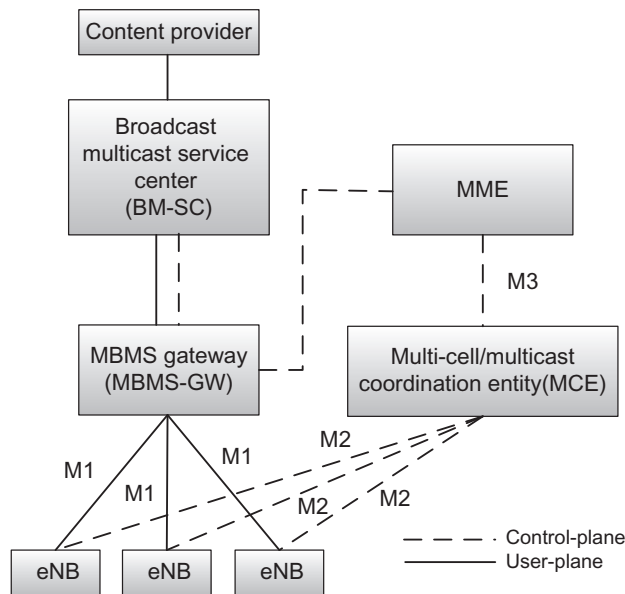


FIGURE 2.18
E-MBMS network architecture [12,32].

control, and allocation of the radio resources used by all eNBs in the MBSFN area for multi-cell MBMS transmissions using MBSFN operation. The MCE decides whether to establish radio bearer(s) for the new MBMS service(s), if the radio resources are not sufficient for the corresponding MBMS service(s), or may allocate radio resources from other radio bearer(s) of ongoing MBMS service(s) according to allocation and retention priority (ARP). In addition to allocation of the time/frequency radio resources, the MCE controls the radio configuration, for example, modulation and coding scheme; counting and acquisition of counting results for MBMS service(s); resumption of MBMS session(s) within MBSFN area(s) based on the ARP and/or the counting results for the corresponding MBMS service(s); suspension of MBMS session(s) within MBSFN area(s) based on the ARP and/or on the counting results for the corresponding MBMS service(s). In the case of distributed MCE architecture, the MCE manages the above functions for a single eNB of an MBSFN.

- **E-MBMS Gateway (MBMS-GW):** The MBMS-GW is a logical entity (this does not preclude the possibility that it may be part of another network element) that is located between the BM-SC and eNBs whose functions include sending/broadcasting of MBMS packets to each eNB participating in the service. The MBMS-GW uses IP multicast as the method of forwarding MBMS user data to the eNBs. The MBMS-GW performs MBMS session control signaling (i.e., session start/update/stop) toward the E-UTRAN via MME.
- **MCE–MME Interface (M3):** The application part of this control-plane interface allows MBMS session control signaling on E-RAB level (i.e., does not convey radio configuration data). The procedures include MBMS session start and stop. The SCTP protocol⁶ is used as signaling transport method.
- **MCE–eNB Interface (M2):** The application part of this control-plane interface conveys radio configuration data for the multi-cell transmission mode eNBs and session control signaling. The SCTP protocol is used as signaling transport.
- **MBMS–GW–eNB Interface (M1):** This user-plane interface carries IP packets between the two network entities and uses IP multicast for point-to-multipoint delivery of user packets.

⁶SCTP is a transport layer protocol, performing a similar role as the commonly used protocols TCP and UDP. It provides some of the service features of both protocols, that is, it is message-oriented like UDP and ensures reliable and in-sequence transport of messages with congestion control like TCP. The protocol is defined by IETF RFC 4960 [20]. The SCTP protocol may be characterized as message-oriented, that is, it transports a sequence of messages comprising a group of bytes, rather than transporting an unbroken stream of bytes as does TCP. As in UDP, in SCTP a sender sends a message in one operation, and that exact message is passed to the receiving application process in one operation. In contrast, TCP is a stream-oriented protocol, transporting streams of bytes reliably and in order.

There are two MBMS architectures defined in 3GPP. In the “distributed MCE architecture,” the MCE is part of the eNB and the M2 interface is maintained between the MCE and the corresponding eNB, whereas in the “centralized MCE architecture” (shown in Figure 2.18), the MCE is a logical entity and can be deployed as a stand-alone physical entity or collocated in another physical entity, such as eNB. In both cases, the M2 interface is maintained between the MCE and all eNB(s) belonging to the corresponding MBSFN area [12].

2.6 Positioning network architecture

3GPP legacy technologies (GSM and UMTS) have supported location services network architecture for the purpose of positioning of mobile devices since 3GPP Rel-4. With the introduction of EPS in 3GPP Rel-8, control-plane location-service architecture for the EPS was introduced in 3GPP Rel-9. Figure 2.19 illustrates the control-plane location service architecture for the EPS, which introduced new interfaces SLg and SLs; and allowed the EPS control-plane element MME to interconnect with the location service core network elements, enabling location services using the positioning functionality provided by the E-UTRAN.

The location service (LCS) architecture follows a client/server model with the gateway mobile location center (GMLC) acting as the location server providing location information to LCS clients. The GMLC sends location requests to the evolved serving mobile location center (E-SMLC) through the MME to retrieve this location information. The E-SMLC is responsible for interaction with the UE through E-UTRAN to obtain the UE position estimate or to obtain position measurements that help the E-SMLC to estimate the UE position. Note that the GMLC interaction over the interfaces connected to it other than SLg was already available before 3GPP Rel-9 for GSM and UMTS access.

The LCS clients can be part of the core network or external to the core network and can also reside in the UE or attached to the UE. Depending on the location of the LCS client, the location request initiated by the LCS client may be a mobile-originated, mobile-terminated, or network-induced location request. Emergency location service is possible even if the UE does not have a valid service subscription due to regulatory requirements. Support of location service-related functionality in the E-UTRAN, MME, and UE are optional. The LCS is applicable to any target UE no matter if the UE supports LCS. The following is a description of various LCS architectural elements and interfaces shown in Figure 2.19 [30]:

- **GMLC:** This entity receives and processes location requests from LCS clients; obtains routing information from HSS; performs registration authorization; communicates information needed for authorization, location service requests, and location information with other GMLC; checks the target UE privacy profile settings; and depending on roaming support, it may assume the role of requesting GMLC, visited GMLC, or home GMLC.

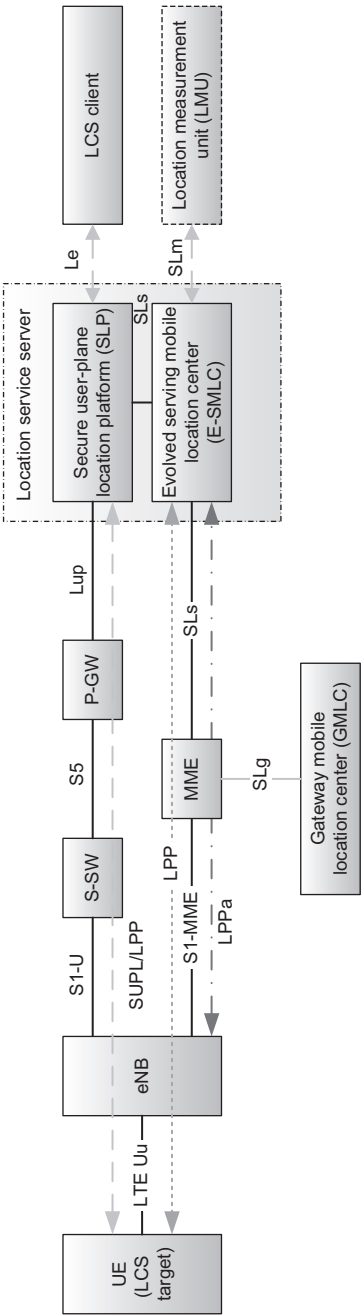


FIGURE 2.19

Positioning network architecture [30].

- **Evolved Serving Mobile Location Center (E-SMLC):** This entity manages the overall coordination and scheduling of resources required for the location determination of a UE that is attached to E-UTRAN. It calculates the final location and estimates velocity as well as the location accuracy. It interacts with the UE to exchange location information applicable to the UE-assisted and the UE-based positioning methods. It interacts with the E-UTRAN to exchange location information applicable to the network-assisted and the network-based positioning methods.
- **Location Measurement Unit (LMU):** The LMU conducts measurements and communicates the measurements to an E-SMLC. All positioning measurements obtained by an LMU are sent to the E-SMLC that requested the measurement report. A UE positioning request may involve measurements by multiple LMUs. The SLM interface between the E-SMLC and the LMU is used for uplink positioning. It is used to transport SLM-AP protocol messages over the E-SMLC-LMU interface. Both stand-alone LMU and LMU integrated into an eNB are supported in LTE Rel-11.

The MME is responsible for UE subscription authorization. It coordinates location services and positioning requests. It handles charging and billing and performs E-SMLC selection (e.g., based on network topology to balance the load on E-SMLC; LCS client type; and requested QoS). It is further responsible for authorization and operation of the LCS services. The HSS is used for storage of location services subscription data and routing information. Two protocols are used for the overall information exchange concerning location services: (1) LTE positioning protocol (LPP); and (2) LTE positioning protocol annex (LPPa). The latter is used for communication between the location server and the eNB. The base station (when using observed time difference of arrival method) is in charge of properly configuring the radio signals that are used by the terminal for positioning measurements, that is, positioning reference signals. It further provides information to the E-SMLC; it enables the device to conduct inter-frequency measurements, if required; and based on the E-SMLC request, it conducts measurements and sends the results to the server. The LPP can be used in user-plane and control-plane methods. The LPP is a point-to-point protocol that allows multiple connections to different devices. The exchanged LPP messages and information can be divided into four categories: (1) UE positioning capability information transfer to the E-SMLC; (2) positioning assisting data delivery from the E-SMLC to the device; (3) location information transfer; and (4) session management.

The SLM-AP protocol, terminated between the E-SMLC and the LMU, is used to support the following functions: (1) delivery of target UE configuration data from the E-SMLC to the LMU; and (2) request positioning measurements from the LMU and delivery of positioning measurements to the E-SMLC. The SLM-AP protocol is directly used between the E-SMLC and the LMU. The MME receives a request for location service associated with a particular target UE from another entity (e.g., GMLC, eNB, or UE), or the MME by itself decides to initiate some

location service on behalf of a particular target UE (e.g., for an IMS emergency call from the UE). The MME then sends a location service request to the E-SMLC. The E-SMLC processes the location service request which may include transferring assisting data to the target UE to help UE-based and/or UE-assisted positioning. The E-SMLC then returns the result of the location service to the MME, for example, a position estimate for the UE and/or an indication of any assisting data transferred to the UE. In the case of a location service requested by an entity other than the MME (e.g., UE, eNB, or E-SMLC), the MME returns the location service result to this entity.

2.7 Mobility management

The NAS is a set of protocols in the EPS. The NAS is used to convey non-radio-access-related signaling between the UE and the MME for an E-UTRAN access. The NAS procedures are grouped in two categories: EPS mobility management (EMM) and ECM. The EMM protocol refers to procedures related to mobility over an E-UTRAN access, authentication, and security. The EMM-specific procedures are UE-initiated. These procedures define attach/detach (to/from the EPC) mechanisms. They also introduce the tracking area update mechanism, which updates the location of the UE within the network. In EPS, a UE initiates a tracking area update when it detects that it enters into a new tracking area. The EMM-specific procedures also define periodic tracking area update. The EMM connection management procedures provide several functions to support the connection of the UE to the network.

The states of a UE with respect to mobility and connection establishment are described by the following NAS states: the EMM state (EMM-DEREGISTERED or EMM-REGISTERED) reflects whether the UE is registered in the MME and the ECM state (ECM-IDLE or ECM-CONNECTED) reflects the connectivity of the UE with the EPC. The transition from ECM-IDLE to ECM-CONNECTED not only involves establishment of an RRC connection, but also includes establishment of an S1 connection. The RRC connection establishment is initiated by the NAS and is completed prior to S1 connection establishment, which means that connectivity in RRC_CONNECTED is initially limited to the exchange of control information between the UE and eNB. The UEs are typically moved to ECM-CONNECTED when becoming active. It should be noted that in LTE, the transition from ECM-IDLE to ECM-CONNECTED is performed within 100 ms and the UEs engaged in intermittent data transfer do not need to stay in ECM-CONNECTED, if the ongoing services can tolerate such transfer delays. The RRC connection release is initiated by the eNB following release of the S1 connection between the eNB and the core network [12,31].

The EMM states describe the mobility management states that result from the mobility management procedures, for example, attach and tracking area update procedures. The EMM has two states: EMM-REGISTERED and

EMM-DEREGISTERED. In EMM-DEREGISTERED state, the EMM context does not have any valid location or routing information for the UE. Since the location of the UE is not known, the UE is not reachable by the MME. In order to avoid performing an authentication and key agreement (AKA)⁷ security procedure in every network attach attempt by the UE, the MME, and the UE may store some UE context in this state. The MME may perform an implicit detach any time after the implicit detach timer expires. The state is changed to EMM-DEREGISTERED in the MME after performing the implicit detach. If all the bearers belonging to a UE are released (e.g., after a handover from E-UTRAN to non-3GPP access), the MME changes the mobility management state of the UE to EMM-DEREGISTERED. This ensures that the UE performs a tracking area update when it reselects E-UTRAN [12,31].

The ECM states describe the signaling connectivity between the UE and the EPC. In EMM-REGISTERED state, there are two EPS connection states: ECM-IDLE and ECM-CONNECTED. In ECM-IDLE, a UE performs cell selection or reselection. There is no NAS signaling connection between the UE and the network. The UE does not have a context in the E-UTRAN and has no S1-MME, and S1-U connection with the network. In EMM-REGISTERED and ECM-IDLE state, the UE performs a tracking area update, if the current TA is not in the list of TAs that the UE has received from the network in order to maintain the registration and enable the MME to page the UE. The UE performs a periodic tracking area update procedure to notify the EPC that it is available [12,31].

2.8 Radio resource management

The purpose of RRM function is to ensure efficient use of the available network resources. In particular, RRM in E-UTRAN provides a means to manage (e.g., assign, reassign, and release) radio resources in single and multi-cell environments. RRM may be treated as a central application at the eNB responsible for interworking between different protocols (RRC, S1AP, and X2AP) so that messages can be properly transferred to different nodes across Uu, S1, and X2 interfaces. RRM may interface with operation and management functions in order to control, monitor, audit, or reset the status due to errors at a protocol stack. The RRM consists of the following main functions [33,66]:

- Radio admission control: The RAC functional module accepts or rejects requests for the establishment of new radio bearers. Admission control is performed according to the required QoS, current system load, and the required service. The RAC function ensures efficient radio resource utilization

⁷AKA is a mechanism which performs authentication and session key distribution in UMTS networks. AKA is a challenge-response-based mechanism that uses symmetric cryptography. AKA is typically run on a UMTS IP multimedia services identity module (ISIM), which resides on a smart card device that also provides tamper resistant storage of shared secrets. AKA is defined in IETF RFC 3310 [21].

by accepting radio bearer requests as long as radio resources are available, and proper QoS enforcement for current sessions by rejecting radio bearer requests, if they cannot be accommodated. It interacts with the RBC module to perform its functions.

- **Radio bearer control:** The RBC functional module manages the establishment, maintenance, and release of radio bearers. It configures the radio resources depending on the current resource availability and usage, as well as QoS requirements of new and ongoing sessions due to mobility or other reasons. The RBC function releases radio resources associated with radio bearers at session termination, handover, or other similar cases. The establishment, maintenance, or release of radio bearers is tied to the configuration of radio resources associated with them.
- **Connection mobility control:** The CMC functional module manages radio resources in the idle and connected modes. In the idle mode, this module defines criteria and algorithms for cell selection, reselection, and location registration that assist the UE in selecting or camping on the best cell. In addition, the eNB broadcasts parameters that configure the UE measurement and reporting procedures. In the connected mode, this module manages the mobility of radio connections without disruption of services. Handover decisions may be based on measurement results reported by the UE, by the eNB, and other parameters configured for each cell. Handover decisions may consider additional information, such as neighbor cell load condition, traffic distribution, transport and hardware resources, and operator-defined policies. Inter-RAT RRM can be one of the subsets of this function responsible for managing the resources in inter-RAT mobility and handovers. In the idle mode, the cell reselection algorithms are controlled by setting of parameters (e.g., thresholds and hysteresis values) that define the best cell and/or determine when the UE should select a new cell. The E-UTRAN broadcasts parameters that configure the UE measurement and reporting procedures.
- **Dynamic resource allocation (DRA) or packet scheduling (PS):** The task of DRA or PS is to allocate and de-allocate resources (including physical resource blocks) to user- and control-plane packets. The scheduling function typically considers the QoS requirements associated with the radio bearers, the channel quality feedback from the UEs, buffer status, inter-cell/intra-cell interference condition, etc. The DRA function may take into account the restrictions or preferences on some of the available resource blocks or resource-block sets due to inter-cell interference coordination (ICIC) considerations.
- **Inter-cell interference coordination:** ICIC function manages radio resources such that intercell interference effects are minimized. The ICIC is essentially a multi-cell RRM function that considers resource usage status and the traffic load conditions across multiple cells.
- **Load balancing:** LB function has the task of handling non-uniform distribution of the traffic load across multiple cells. The purpose of LB function is to control the load distribution in such a manner that radio resources are

efficiently utilized; to maintain the QoS of the current sessions as much as possible; and to minimize the probability of call drops. The LB algorithms may result in handover or cell reselection decisions with the purpose of redistributing traffic from overloaded cells to under-utilized cells.

- **Inter-RAT radio resource management:** Inter-RAT RRM is primarily concerned with the management of radio resources in connection with inter-RAT mobility, notably inter-RAT handover. During inter-RAT handover, the handover decision may take into account the involved RATs resource availability as well as the UE capabilities and the operator policies. Inter-RAT RRM may also include additional functions for inter-RAT load balancing for the idle mode and connected mode UEs.
- **Subscriber profile ID (SPID) for RAT/frequency priority:** The RRM function maps SPID parameters received via the S1 interface to a locally defined configuration in order to apply specific RRM strategies (e.g., to define RRC_IDLE mode priorities and control inter-RAT/inter-frequency handover in RRC_CONNECTED mode). The SPID is an index referring to user information, such as mobility profile and service usage profile. The SPID information is UE-specific and applies to all of its radio bearers. The RRM strategy in E-UTRAN may be based on user-specific information.

2.9 Quality of service/quality of experience

The end-to-end quality of service experienced by an end user results from a combination of elements throughout the protocol stack and E-UTRAN/EPC components. Thus, the performance evaluation of the service requires a detailed performance analysis of the entire network, that is, from the UE up to the application server (AS) or remote UE. The QoE is a subjective measurement of the quality experienced by a user when he/she uses a communication service. In other words, the QoE is defined as the overall acceptability of an application or a service as perceived subjectively by the end user. When evaluating the quality of a service, the objective is to optimize the operation of the network from a perspective which is purely based on objective parameters, or to determine the quality that the user is actually experiencing and its satisfaction level. The QoE takes into account the satisfaction of a user in terms of both content and application performance. In this sense, the introduction of smartphones has been a quantitative leap in user QoE expectations [43,67,68].

The QoE has been traditionally evaluated through subjective tests carried out with the users in order to assess their satisfaction with a mean opinion score value. This approach can be quite expensive, time-consuming, and cannot be used for real-time decision-making to improve the QoE. New methods have been emerged in recent years to estimate the QoE based on certain performance indicators associated with services. A possible solution to instantaneously evaluate the

QoE is to integrate the QoE analysis tools in the mobile terminal itself. If mobile terminals are able to report the measurements to a central server, the QoE assessment process is significantly simplified. Other solutions are focused on including new network elements (e.g., network analyzers and deep packet inspectors) that are responsible for capturing the traffic from a certain service and analyzing its performance. In some cases, the evaluation of video-streaming quality in the mobile terminals is addressed by monitoring objective parameters, such as packet loss rate or jitter.

As mentioned earlier, the E-UTRAN consists of a network of eNBs which are typically interconnected by means of an X2 interface and to the EPC via an S1 interface. The eNB entity plays a critical role in maintaining the end-to-end QoS. The eNB performs QoS-related functions which include admission control and resource reservation; scheduling and distribution of radio resources among the established bearers; and L1/L2 protocol configuration in accordance with the QoS attributes associated with the bearer. The LTE-enabled UEs may support multiple applications at the same time, each having different QoS requirements. This is achieved by establishing different EPS bearers for each QoS flow. The EPS bearers can be classified into two categories based on the characteristics of the QoS they provide: GBR bearers in which resources are permanently allocated and non-GBR bearers which do not guarantee any particular bit rate. In the access network, it is the eNB's responsibility to ensure that the QoS requirements for a particular bearer over the radio interface are met. Each bearer has an associated QCI, characterized by priority, packet delay budget, and admissible packet loss rate; and an ARP used for call admission control. The IP packets mapped to the same EPS bearer receive the same bearer level packet forwarding treatment (e.g., scheduling policy, queue management policy, and rate shaping policy). Thus, the UE is not only responsible for requesting the establishment of EPS bearers for each QoS flow, but also for performing packet filtering in the uplink into different bearers based on TFTs, as P-GW does for the downlink (see [Figures 2.4 and 2.5](#)).

It was mentioned earlier that one EPS bearer/E-RAB is established when the UE connects to a PDN, and that remains established throughout the lifetime of the PDN connection to provide the UE with always-on IP connectivity to that PDN. That bearer is referred to as the default bearer. Any additional EPS bearer/E-RAB that is established to the same PDN is referred to as a dedicated bearer. The default bearer QoS parameters are assigned by the network based upon subscription data. The decision to establish or modify a dedicated bearer can only be taken by the EPC and the bearer level QoS parameter values are always assigned by the EPC.

2.10 Security in E-UTRAN

Security over the E-UTRAN air-interface is provided through cryptographic techniques. The backhaul link from the eNB to the core network uses Internet key

exchange (IKE)⁸ and the IP security protocol (IPsec)⁹ when cryptographic protection is needed. These cryptographic techniques provide end-to-end protection for signaling between the core network and UE. Therefore, the main location where user traffic is exposed to security threats is in the eNB. Moreover, to minimize susceptibility to attacks, the eNB must provide a secure environment that supports the execution of sensitive operations, such as the encryption or decryption of user data and the storage of sensitive data, such as keys for securing UE communication, long-term cryptographic secrets, and vital configuration data. Similarly, the use of sensitive data must be confined to this secure environment. Even with the above security measures in place, one must consider attacks on an eNB, because, if successful, they could give attackers full control of the eNB and its signaling to UEs and other nodes. To limit the effect of a successful attack on one eNB, attackers must not be able to intercept or manipulate user- and control-plane traffic that traverses another eNB, for example, following a handover.

The E-UTRAN and EPS security model is shown in Figure 2.20. The subscriber authentication function in EPS is based on the UMTS AKA protocol. It provides mutual authentication between the UE and core network, ensuring robust charging and guaranteeing that no unauthorized and non-authenticated entities can pose as a valid network node. The EPS AKA provides a root key from which a key hierarchy is derived. The keys in this hierarchy are used to protect signaling and user-plane traffic between the UE and the network. The key hierarchy is derived using cryptographic functions. Furthermore, keys are bound to where, how, and for which purpose they are used. This ensures that keys used for one access network cannot be used in another access network, and that the same key is not used for multiple purposes or with different algorithms. The key hierarchy and bindings also make it possible to routinely and efficiently change the keys used between a UE and eNBs (e.g., during handover) without changing the root key or the keys used to protect signaling between the UE and the core network. For radio-specific signaling, LTE provides integrity, replay protection, and encryption between the UE and the eNB.

The IKE/IPsec protocols can protect the backhaul signaling between the eNB and MME. In addition, LTE-specific protocols provide end-to-end protection of

⁸IKE or IKEv2 is the protocol used to set up a security association in the IPsec protocol suite. The IKE protocol is based on a key-agreement protocol and the Internet security association and key management protocol which is a protocol defined by IETF RFC 2408 for establishing security associations and cryptographic keys in an Internet environment. The IKE uses X.509 certificates for authentication which are either pre-shared or distributed to set up a shared session secret from which cryptographic keys are derived. In addition, a security policy for every peer which will connect must be manually maintained [76].

⁹IPsec is a set of protocols for securing IP-based communications by authenticating and encrypting each IP packet of a data stream. IPsec also includes protocols for establishing mutual authentication between agents at the beginning of the session and negotiation of cryptographic keys to be used during the session. IPsec can be used to protect data flows between a pair of hosts, between a pair of security gateways, or between a security gateway and a host [24,25,52].

signaling between the MME and UE. For user-plane traffic, IKE/IPsec can similarly protect the backhaul from the eNB to the S-GW. Support for integrity, replay protection, and encryption is mandatory in the eNB. The user-plane traffic between the UE and eNB is only protected by encryption, as integrity protection would result in excessive overhead. Nevertheless, it is not possible to intelligently inject traffic on behalf of another user: attackers are essentially blind in the sense that any traffic they try to inject would almost certainly decrypt to invalid information.

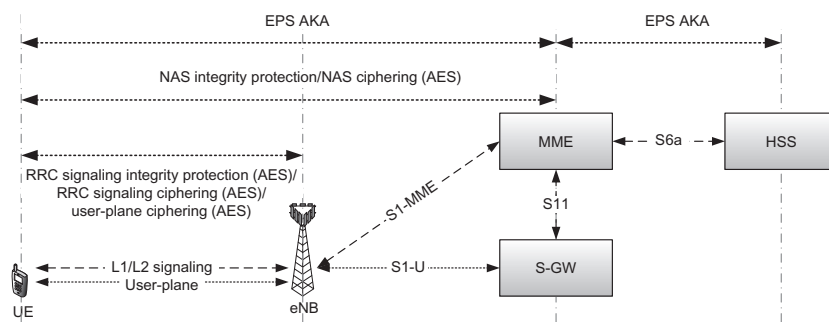
2.11 Support of voice and messaging services in LTE/LTE-Advanced

2.11.1 IMS architecture

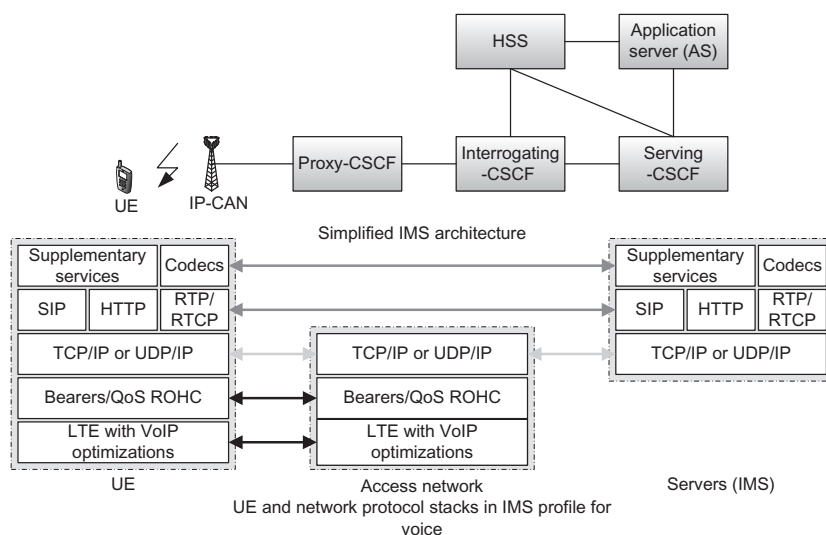
Voice services in LTE are enabled using the IMS which provides a framework for supporting IP-based services and requires new IMS-specific network elements as part of the dedicated core network architecture. The first version of IMS was introduced in 3GPP Rel-5 specifications; however, many enhancements have been made to the baseline in subsequent releases. During the development of the LTE standards, it was assumed that IMS would be commercially available when the LTE networks were deployed, since voice support in LTE relied on the IMS. However, rollout of IMS infrastructure was slower than expected, making voice support in LTE a real challenge for many network operators. A closely related feature to the support of voice is the support of SMS as another key circuit-switched service which was not supported in LTE. In general, the support of voice and messaging services in LTE using IMS remains one of the major targets for the industry. An industry initiative called voice over LTE (VoLTE) was formed in 2010 by the Global System for Mobile Communications Association (GSMA) which has developed a framework for support of voice and SMS over IMS in LTE, including roaming and interconnection issues [53].

The IMS is an access-independent IP connectivity and service control architecture. It provides the framework for IP-based multimedia services in a mobile network and is the most efficient choice to provide VoIP services in LTE. In this section, we focus on voice and messaging services over IMS. A simplified IMS reference architecture is shown in Figure 2.21, which includes interfaces toward legacy and IP-based networks. This architecture provides entities for session management and routing, service support, databases, and interworking.

The IP connectivity access network (IP-CAN) in the case of LTE would consist of the EPS and the E-UTRAN. The call session control function (CSCF)

**FIGURE 2.20**

E-UTRAN and EPS security model [70,71].

**FIGURE 2.21**

Simplified IMS architecture and protocol stack [50,65].

entities are the core components of the IMS. There are three CSCF entities defined in IMS architecture:

- Proxy-CSCF (P-CSCF): The P-CSCF is the first point of contact for a user. The P-CSCF behaves as a proxy, that is, it accepts and forwards the requests.
- Interrogating-CSCF (I-CSCF): The I-CSCF is the contact point within an operator's network for all connections destined to a subscriber.

- **Serving-CSCF (S-CSCF):** The S-CSCF is responsible for handling the registration process, making routing decisions, maintaining sessions, and downloading user information and service profiles from the HSS.

The HSS is the master database for a user. It is comparable to the HLR in a legacy mobile radio network. The HSS contains the subscription-related information required for other network entities that handle calls or sessions. For example, the HSS provides support to the call control servers to complete the routing/roaming procedures by resolving authentication, authorization, naming/addressing resolution, and location dependencies. The application server provides specific IP applications, such as messaging. The purpose of the IMS architecture and the different CSCF entities becomes clearer in the case of roaming. In general, the network operators' internal network structures and user databases are proprietary. The local P-CSCF in the network, which is the point of access for the UE, does not have direct connection to the HSS, rather, it is connected to the HSS via the I-CSCF entity.

The IMS uses a set of IP-based protocols including SIP as a text-based protocol for registration, subscription, notification and initiation of sessions; session description protocol (SDP) as a text-based protocol for negotiating session parameters, such as media type, codec type, bandwidth, IP address and ports, and for media stream setup; RTP and RTCP for transport of real-time applications, such as audio; extensible markup language (XML) configuration access protocol (XCAP), which allows a client to read, write, and modify application configuration data, stored in XML format on a server and maps XML document sub-trees and element attributes to HTTP uniform resource identifiers so that these components can be directly accessed by HTTP; and DHCP for configuring IP addresses for IPv4 and IPv6 [19]. The UDP for messages smaller than 1300 bytes or transport control protocol (TCP) can be used as transport protocols.

As shown in Figure 2.21, the supplementary services requested by the VoLTE profile have been specified in GSMA IR.92 [50,65]. For example, the originating identification presentation (OIP) service provides the terminating user a way to receive identity information to identify the originating user. Another example would be the supplementary service communication hold (HOLD) that enables a user to suspend the reception of media stream(s) from an established IP multimedia session, and resume them at a later time. For supplementary service configuration, the UE and the network must support the XCAP, which is an application layer protocol that allows a client to read, write, and modify application configuration data stored in XML format on a server.

The IMS architecture has been designed to enable operators to provide a wide range of real-time, packet-based services and to track their use in a way that allows both traditional time-based charging as well as packet and service-based charging. It has become increasingly popular both with wire-line and wireless service providers as it is designed to increase carrier revenues, deliver integrated multimedia services, and create an open, standards-based network. The IMS provides a wide range of session border control, including call access control, reachability and security. It also

provides a framework for the deployment of both basic calling services and enhanced services, including multimedia messaging; web integration; presence-based services; and push-to-talk. At the same time, it draws on the traditional telecommunications experience of guaranteed QoS; flexible charging mechanisms (time-based, call-collect, and premium rates); and lawful intercept compliance [50,65].

The IMS decomposes the network infrastructure into separate functions with standardized interfaces between them. Each interface is specified as a “reference point,” which defines both the protocol over the interface and the functions between which it operates. The standards do not mandate which functions should be collocated, as this depends on the scale of the application, and a single device may contain several functions. The IMS architecture specified by 3GPP can be divided into three main layers, each of which is described by a number of equivalent names: service/application layer; control layer; and transport layer [8].

Application Layer: An application layer consists of application servers that host the IMS services and an HSS. The application layer provides a framework for provisioning and management of services and defines standard interfaces to common functionalities including configuration storage; identity management; user status, such as presence and location, which is held by the HSS billing services and provided by a charging gateway function (CGF) control of voice and video calls and messaging, provided by the control layer.

Control Layer: The control layer is located between the application and transport layers. It routes the call signaling, informs the transport layer of permissible traffic, and generates billing information for the use of the network. At the core of this layer is the CSCF, which comprises the following functions. The P-CSCF is the first point of contact for users with the IMS. The P-CSCF is responsible for security of the messages between the network and the user and allocating resources for the media flows. The I-CSCF is the first point of contact from peered networks. The I-CSCF is responsible for querying the HSS to determine the S-CSCF for a user and may also hide the operator’s inner network from peer networks. The S-CSCF is the central entity which is responsible for processing registrations to record the location of each user, user authentication, and call processing including routing of calls to applications. The operation of the S-CSCF is controlled by policy stored in the HSS. This distributed architecture provides an extremely flexible and scalable solution. For example, any of the CSCF functions can generate billing information for each operation. The control layer also controls transport-layer traffic through the resource and admission control subsystem (RACS). This consists of the PDF, which implements local policy on resource usage, for example, to prevent overload of particular access links, and access-RAC function (A-RACF), which controls QoS within the access network.

Transport Layer: The transport layer provides a core QoS-enabled IPv6 network with access from the UE over Wi-Fi and other mobile broadband networks. This infrastructure is designed to provide a wide range of IP multimedia server-based and P2P services. Access into the core network is through border gateways (GGSN/Packet Data Gateway/Broadband Access Server). These entities enforce

policy provided by the IMS core network, controlling traffic flows between the access and core networks. Within the transport layer, the interconnect border control function (I-BCF) controls transport level security and informs the RACS of the resources required for a call the interconnect border gateway function (I-BGF), access network inter-connect border gateway function (A-BGF) border gateway functions provide media relay for hiding endpoint addresses.

2.11.2 Single radio voice call continuity

Single radio voice call continuity (SRVCC) is an LTE feature that allows a VoIP/IMS call in the LTE packet-switched domain to be transferred to a legacy circuit-switched domain (GSM/UMTS or CDMA2000). Let us assume that a new LTE network operator is going to migrate its voice services to VoIP over IMS in conjunction with deployment of an LTE radio access network. Without SRVCC features, this operator would need to provide a ubiquitous LTE coverage in order to have a viable VoIP service. However, an SRVCC-enabled LTE network may not require comprehensive LTE coverage in order to provide voice services across the entire network. The SRVCC feature provides the ability to transition a voice call from the VoIP/IMS packet-switched domain to the legacy circuit-switched domain. Variations of SRVCC have been standardized to support both GSM/UMTS and CDMA2000 circuit-switched domains. For an operator with a legacy cellular network, deployment of IMS-based VoIP services in conjunction with the rollout of an LTE network is possible through use of SRVCC, which offers VoIP subscribers coverage over a much larger area than would typically be available during the rollout of a new network [3,48,54,72,76].

The SRVCC feature functions as follows. When an SRVCC-capable UE engaged in a voice call determines that it is moving away from an LTE coverage area (based on neighboring cell measurements), it notifies the serving eNB. The LTE network determines that the voice call has to be moved to the legacy circuit-switched domain, thus, it notifies the MSC server of the need to switch the voice call from the packet-switched to the circuit-switched domain and initiates handover of the LTE voice bearer to the circuit-switched network. The MSC server establishes a bearer path for the mobile in the legacy network and notifies the IMS core network that the UE's call is moving from the packet-switched to the circuit-switched domain. The IMS core network then enacts the necessary interworking functions to facilitate the call transfer from one domain to another. When the mobile arrives in the legacy network, it switches its internal voice processing from VoIP to legacy circuit-switched voice processing and the call is resumed. If the legacy circuit-switched network also has an associated packet-switched capability and is capable of supporting concurrent circuit-switched/packet-switched operations, the subscriber's data sessions can be handed over to the legacy network in conjunction with switching the voice call from the packet-switched to the circuit-switched domain. In this case, when the voice call is terminated and the mobile reenters the LTE coverage, these packet sessions can be handed over to the LTE network again.

The SRVCC feature may not be suitable for the operators that do not plan to migrate their voice service to VoIP. If an operator does plan to migrate to VoIP and also plans to roll out ubiquitous LTE coverage, then the question of whether or not to adopt SRVCC is more complicated. While SRVCC does not require modifications to the existing radio access network, it does require a significant modification of the operator's legacy core network and further requires full deployment of IMS circuit-switched/packet-switched continuity services. Given the cost of these changes, deployment of SRVCC as an interim measure to allow early rollout of VoIP-based services may not make financial sense.

The SRVCC provides voice call continuity between LTE radio access and circuit-switched mode of other radio access technologies (e.g., CDMA2000, GSM, or UMTS) using IMS. Calls remain anchored in IMS, which means that the circuit-switched connections can be treated as standard IMS sessions and the circuit-switched bearer is used as a media for IMS sessions. Using SRVCC, the UE only has to be capable of transmitting/receiving on one access network at a given time. The SRVCC feature provides a solution for LTE networks which do not support circuit-switched voice and messaging services. Circuit-switched legacy technologies complement the LTE coverage with voice using IMS support [72].

The SRVCC functionality is built on the IMS centralized services framework for delivering voice and messaging services to the users regardless of the type of network to which they are attached, and for maintaining service continuity for moving terminals. The call anchoring is done by means of a voice call continuity application server in the home IMS network. During EPS attach, routing information to the voice call continuity application server is provided from the HSS to the MME. During the EPS attach procedure, the network and the UE indicate their support for SRVCC. Some modifications in the core network elements are required to support SRVCC, for example, the CDMA2000 circuit-switched SRVCC interworking function is required [18].

To support GSM and UMTS, some modifications in the MSC server are required. When the E-UTRAN selects a target cell for SRVCC handover, it needs to indicate to the MME that this handover procedure requires SRVCC. Upon receiving the handover request, the MME triggers the SRVCC procedure with the MSC server. The MSC then initiates the session transfer procedure to IMS and coordinates it with the circuit-switched handover procedure to the target cell. Handling of any non-voice packet-switched bearer is by the packet-switched bearer splitting function in the MME. The handover of non-voice packet-switched bearers, if performed, is according to a regular inter-RAT packet-switched handover procedure. The handover procedure from E-UTRAN (as shown in Figure 2.22) is used to initiate the SRVCC procedure. When SRVCC is enacted, the downlink flow of voice packets is switched toward the target circuit-switched network. The call is moved from the packet-switched to the circuit-switched domain, and the UE switches from VoIP to circuit-switched voice. Figure 2.23 shows the simplified procedure for SRVCC for GERAN or UTRAN.

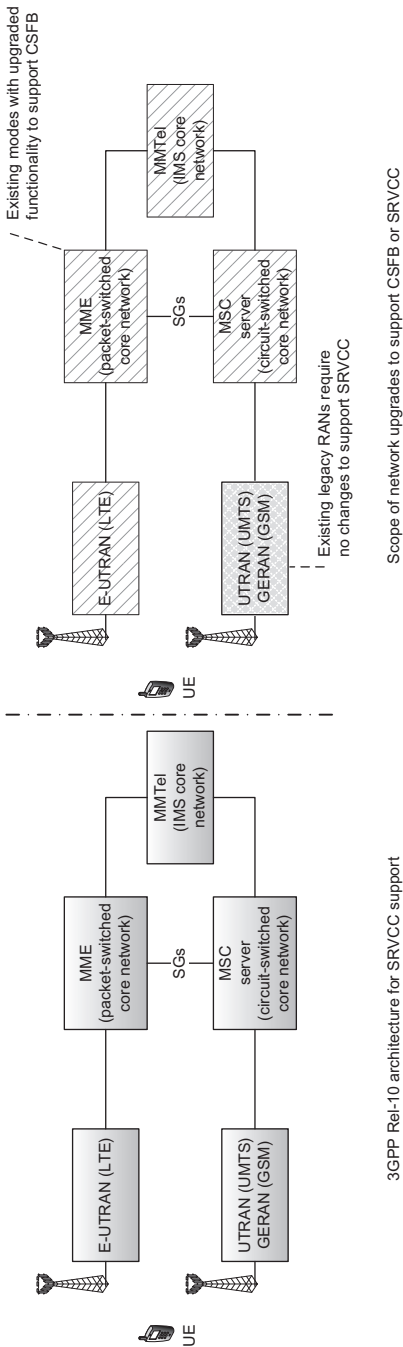


FIGURE 2.22

Network architecture with SRVCC support [72].

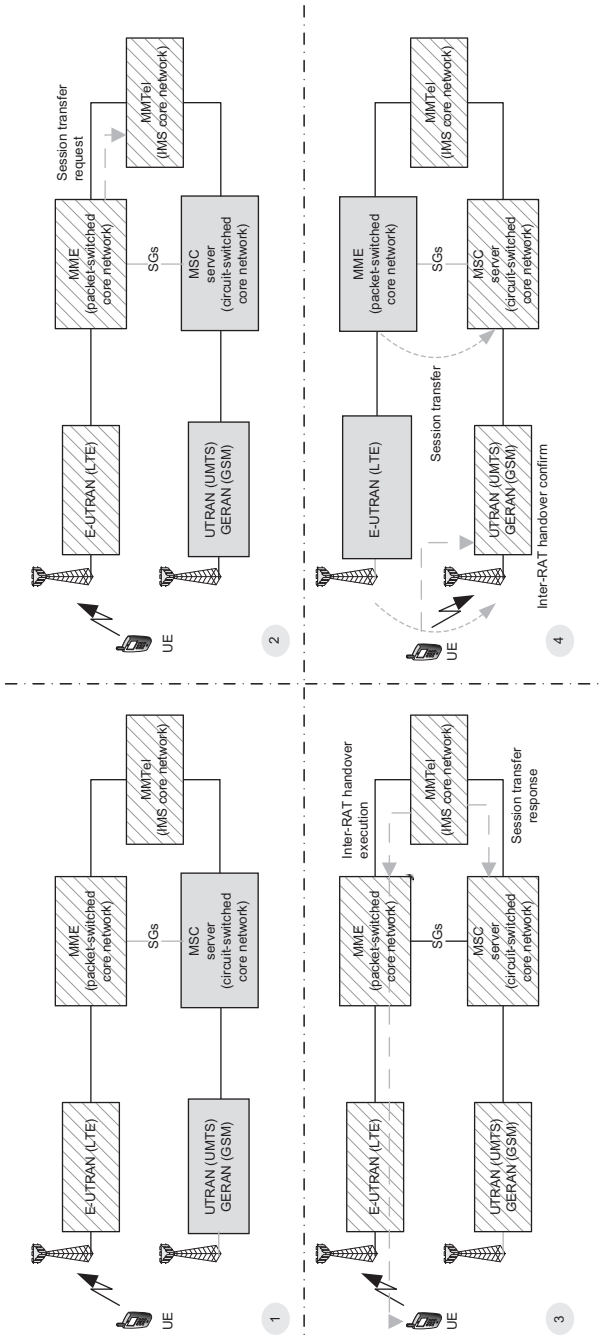


FIGURE 2.23

Call transfer procedure using SRVCC [72].

3GPP Rel-10 architecture has been recommended by GSMA for SRVCC because it reduces both voice interruption time during handover and the dropped call rate compared to earlier configurations. The network controls and moves the UE from E-UTRAN to UTRAN/GERAN as the user moves out of the LTE network coverage area. The SRVCC handover mechanism is entirely network-controlled and calls remain under the control of the IMS core network, which maintains access to subscribed services implemented in the IMS service engine throughout the handover process. 3GPP Rel-10 configuration includes all components needed to manage the time-critical signaling between the user's device and the network, and between network elements within the serving network, including visited networks during roaming. As a result, signaling follows the shortest possible path and is as robust as possible, minimizing voice interruption time caused by switching from the packet-switched core network to the circuit-switched core network, whether the UE is in its home network or roaming. With the industry aligned around the 3GPP standard and GSMA recommendations, SRVCC-enabled user devices and networks will be interoperable, ensuring that solutions work in many scenarios of interest.

As shown in Figure 2.22, the SRVCC functionality can be added to the network by upgrading the mobile soft-switch software as well as the IMS and the LTE/EPC subsystems. The legacy UTRAN/GERAN access networks are not required to be upgraded. When a user with an active voice call using VoLTE moves to an area without LTE coverage, the handover of the call to the network is initiated with inter-RAT handover and is completed with a session transfer. Inter-RAT handover is the conventional method of transferring the UE from E-UTRAN to UTRAN/GERAN. Session transfer is a new mechanism to move access control and voice media anchoring from the EPC to the legacy circuit-switched core network. During the entire voice handover process, the IMS retains control of the user. The handover process is initiated by a session transfer request to the IMS. The IMS/multimedia telephony (MMTel) responds simultaneously with two commands, one to the LTE network and one to the 2G/3G network. The LTE network in which the user's voice call is in progress receives an inter-RAT handover execution command through the MME and the E-UTRAN to instruct the UE to prepare to move to the circuit-switched network. The circuit-switched network, where the user's voice call is being transferred, receives a session transfer response, preparing it to accept the call in progress. With acknowledgments that the commands have been executed, the UE and the IMS, switch to the circuit-switched network to continue the call.

While the implementation of SRVCC does not directly impact the legacy UTRAN/GERAN nodes, an indirect impact arises from methods to help the UE return to the LTE network as soon as the voice call has ended. To help transfer the device back to the LTE network, the legacy RAN can implement functionality to either broadcast LTE system information to the UE, with which the device can perform a cell reselection to an LTE cell after connection release or

release connection to the UE and simultaneously redirect the UE to the LTE network.

The inter-RAT handover and session transfer procedures, as well as the number of sub-systems requiring upgrades, have raised concerns about whether SRVCC can meet the voice interruption time of less than 300 ms. It must be noted that the inter-RAT handover in the E-UTRAN and the session transfer in the core network contribute to an interruption time as they break and make the connection. To address this issue, the SRVCC in 3GPP Rel-10 network architecture has minimized the interruption time by initiating the two procedures simultaneously so that they can be executed in parallel. The studies have shown that the session transfer process can be performed in the order of 10 ms; therefore, the voice call interruption time is mainly affected by the inter-RAT handover latency. The inter-RAT handover delay is measured from the time that the UE receives a network handover command to the time that the UE is uplink synchronized with the new radio access network.

Along with the introduction of the LTE radio access network, 3GPP also standardized SRVCC in Rel-8 specifications to provide seamless service continuity when a UE performs a handover from the E-UTRAN to UTRAN/GERAN. With SRVCC, calls are anchored in the IMS network while the UE is capable of transmitting/receiving on only one of those access networks at a given time, where a call anchored in the IMS core can continue in UMTS/GSM networks and outside of the LTE coverage area. Since its introduction in Rel-8, the SRVCC has evolved with each new release, a brief summary of SRVCC capability and enhancements are noted below [48,49].

- **3GPP Rel-8:** Introduces SRVCC for voice calls that are anchored in the IMS core network from E-UTRAN to CDMA2000 and from E-UTRAN/UTRAN (HSPA) to UTRAN/GERAN circuit-switched. To support this functionality, 3GPP introduced new protocol interface and procedures between MME and MSC for SRVCC from E-UTRAN to UTRAN/GERAN, between SGSN and MSC for SRVCC from UTRAN (HSPA) to UTRAN/GERAN, and between the MME and a 3GPP2-defined interworking function for SRVCC from E-UTRAN to CDMA 2000.
- **3GPP Rel-9:** Introduces the SRVCC support for emergency calls that are anchored in the IMS core network. IMS emergency calls, placed via LTE access, need to continue when SRVCC handover occurs from the LTE network to GSM/UMTS/CDMA2000 networks. This evolution resolves a key regulatory exception. This enhancement supports IMS emergency call continuity from E-UTRAN to CDMA2000 and from E-UTRAN/UTRAN (HSPA) to UTRAN/GERAN circuit-switched network. Functional and interface evolution of EPS entities were needed to support IMS emergency calls with SRVCC.
- **3GPP Rel-10:** Introduces procedures of enhanced SRVCC including support of mid-call feature during SRVCC handover (eSRVCC); support of SRVCC packet-switched to circuit-switched transfer of a call in alerting phase

(aSRVCC); MSC server-assisted mid-call feature enables packet-switched/circuit-switched access transfer for the UEs not using IMS centralized service capabilities, while preserving the provision of mid-call services (inactive sessions or sessions using the conference service). The SRVCC in alerting phase feature adds the ability to perform access transfer of media of an instant message session in packet-switched to circuit-switched direction in alerting phase for access transfers.

- **3GPP Rel-11:** Introduces two new capabilities: single radio video call continuity for 3G-circuit-switched network (vSRVCC); and SRVCC from UTRAN/GERAN to E-UTRAN/HSPA (rSRVCC). The vSRVCC feature provides support of video call handover from E-UTRAN to UTRAN-circuit-switched network for service continuity when the video call is anchored in IMS and the UE is capable of transmitting/receiving on only one of those access networks at a given time. Service continuity from UTRAN/GERAN circuit-switched access to E-UTRAN/HSPA was not specified in 3GPP Rel-8/9/10. To overcome this drawback, 3GPP Rel-11 provided support of voice call continuity from UTRAN/GERAN to E-UTRAN/HSPA. To enable video call transfer from E-UTRAN to UTRAN-circuit-switched network, IMS/EPC is evolved to pass relevant information to the EPC side and S5/S11/Sv/Gx/Gxx interfaces are enhanced for video bearer-related information transfer. To support SRVCC from GERAN to E-UTRAN/HSPA, GERAN specifications are evolved to enable a mobile station and base station sub-system to support seamless service continuity when a mobile station hands over from GERAN circuit-switched access to E-UTRAN/HSPA for a voice call. To support SRVCC from UTRAN to E-UTRAN/HSPA, UTRAN specifications are evolved to enable the RNC to perform rSRVCC handover and to provide relative UE capability information to the RNC.

2.11.3 Voice over LTE

Some of the earliest LTE deployments addressed the lack of circuit-switched voice support in LTE by utilizing a dual-radio solution in which the mobile device was required to simultaneously connect to the E-UTRAN and a legacy circuit-switched network, the latter being used to provide voice and messaging services. The advantage of a simultaneous voice and LTE (SVLTE) solution was that circuit-switched voice and packet-switched data services were supported at the same time. However, an SVLTE-based solution would add significant hardware complexity to the mobile device, given that two completely independent baseband and RF chains were required with obvious cost implications. Furthermore, an SVLTE-based solution has proven to be very power-consuming in practice. On the other hand, since the LTE and the legacy cells do not necessarily have identical footprints, the user experience during mobility can be less than desired. There are several options when it comes to solving the LTE voice services problem. Those range from multi-radio solutions, to solutions which tunnel legacy signaling through LTE, to SIP-based

VoIP solutions, to not using LTE for voice at all, instead relying on legacy networks to provide voice services. The VoLTE solution has developed a framework for optimal support of voice and SMS over IMS in E-UTRAN including roaming and inter-operator issues. VoLTE is based on the existing IMS MMTel¹⁰ concepts and has been specified in GSMA profile IR.92 [45–47]. The GSMA IR.92 (IMS profile for voice and SMS) is a profile as opposed to a technical specification. Therefore, rather than specifying new technologies, this document defines how the framework and services already specified by 3GPP should be configured and utilized in a manner that facilitates interoperability among mobile devices and the serving networks, as well as amongst the networks [64,72,76].

The IR.92 profile requires the mobile device to indicate its support for MMTel-based supplementary services during SIP registration, and if the device supports SMS-over-IP it is also required to be indicated. The terminals are further required to provide their IMEI during SIP registration, which allows the operators to implement management policies in the P-CSCF which are device or device class specific. IR.92-compliant devices are required to support network-initiated deregistration, so that the network can force the mobile device to detach in special circumstances.

IR.92 leverages the existing mechanisms for IMS authentication, that is, IMS AKA. VoLTE-enabled UEs are expected to install a universal integrated circuit card (UICC) equipped with the ISIM application. The ISIM application contains the credentials critical to authentication, such as the subscriber's IP multimedia private identity (IMPI), the home carrier's domain name, and a shared secret key, which is used for encryption. If the ISIM application is absent from the UICC, IMS AKA procedures can still be performed using the credentials contained on the UE's USIM.

A mobile-originated call is handled by an MMTel server. This is accomplished by setting the IMS communication service identifier (ICSI) in the communication invitation to point to the MMTel service, resulting in calls being looped by S-CSCF through the MMTel server in each subscriber's home network. In SIP protocol, which IMS relies on, call control and bearer (transport) control are essentially independent. In the context of the IR.92 profile, there are two situations in

¹⁰IMS MMTel service is a global standard based on the IMS, offering converged, fixed, and mobile real-time multimedia communication using the media capabilities, such as voice, real-time video, text, file transfer and sharing of pictures, audio and video clips. With MMTel, users can add and drop media during a session. One can start with chat, add voice, add another caller, add video, share media and transfer files, and drop any of these without losing, or having to end, the session. MMTel allows a single SIP session to control virtually all MMTel supplementary services and MMTel media. All available media components can easily be accessed or activated within the session. Employing a single session for all media parts means that no additional sessions need to be set up to activate video, to add new users, or to start transferring a file. Although it is possible to manage single-session user scenarios with several sessions, for example, using a circuit-switched voice service that is complemented with a packet-switched video session, a messaging service, or both, there are some benefits in taking MMTel's single-session approach [36].

which it is desirable to have some communication between the two domains. The first is during call session establishment. It is possible for a call-terminating device to accept a call invitation (and to alert the terminating subscriber). This is not normally a problem in legacy circuit-switched systems because call and bearer control are relatively more coupled. In order to minimize the possibility of this occurrence, VoLTE requires the use of SIP pre-conditions framework. This ensures that the terminating subscriber is not notified of the incoming call unless it is certain that there are sufficient resources to support the call (with the required QoS). The second situation in which call and bearer control require some communication is in the context of call drops and call establishment failures. If, for example, communication between the LTE/System Architecture Evolution (SAE) network and the Internet is lost, or if the radio link fails, IR.92 requires the IMS call session to be reestablished or cleared gracefully on the mobile device side and on the network side. VoLTE emergency calls must also be supported by the UE and the network according to regulatory requirements.

The IMS calls are processed by the subscriber's S-CSCF in the home network. The connection to the S-CSCF is, however, indirect and via P-CSCF. Since the P-CSCF varies with time and the mobile device's serving network, an important step in the enablement of voice calling capabilities is the discovery of the P-CSCF.

In legacy systems, telephony features are managed via dedicated signaling messages. Since call control in VoLTE uses completely different signaling, supplementary services and features in VoLTE are managed differently. IR.92 requires the use of XCAP protocol. Using XCAP, the UE connects to the MMTel server via HTTP and utilizes the standard PUT, GET, and DELETE methods to modify XML documents, which describe the supplementary services and their state.

3GPP adaptive multi-rate (AMR) and AMR-wideband (AMR-WB) voice codecs are specified for VoLTE. This choice simplifies interoperability with legacy 3GPP networks, in particular, in the context of transcoder-free operation. In order to ensure interoperability with all legacy telephony networks, in addition to the necessary voice transcoders, support for transmission of dual-tone multi-frequency (DTMF) signaling, which is used for telecommunication signaling over analog telephone lines, is mandatory.

VoLTE requires the use of a particular set of radio bearers, specifically, two signaling radio bearers (SRBs) and two or three data radio bearers (DRBs). The SRBs (SRB1 and SRB2) in E-UTRAN transport the dedicated control channel, which is used for RRC message exchange between the UE and the eNB, as well as NAS message exchange between the UE and the MME. These radio bearers always operate in RLC acknowledged mode. The two (or three) DRBs carry SIP (call control) and XCAP (service management) signaling as well as voice packets. The first DRB corresponds to the default EPS bearer and is used to transport SIP and XCAP signaling messages. The associated QCI index is 5, which means that the bearer operates in non-GBR mode but with high priority, low latency, and very low packet loss. The second DRB corresponds to a dedicated EPS bearer and is used to transport voice packets. The associated QCI index is 1, which

means that the bearer operates in GBR mode with high priority and low latency, but with relatively higher packet loss. There may also be an additional DRB operating with QCI index 8/9 (low-priority, high-latency, and very low packet loss) used for the transport of XCAP messages.

All networks and mobile devices are required to utilize a common APN for VoLTE, which in this case is IMS. Unlike the legacy networks, LTE networks employ an always-on concept which means that devices have PDN connectivity virtually from the moment they attempt their initial attach to the core network. During the initial attach procedure, some devices choose to identify the access point through which they prefer to connect. However, mobile devices are not permitted to name the VoLTE APN during initial attach, that is, to utilize the IMS as their main PDN, rather to establish a connection with the IMS application profile separately. Thus, VoLTE-enabled devices must support multiple simultaneous default EPS bearers. Note that because the VoLTE APN is universal, mobile devices will always connect through the visited PLMN's IMS P-GW.

We mentioned earlier that VoLTE sessions utilize two or three DRBs. This implies the use of one default EPS bearer plus one or two dedicated EPS bearers. The default EPS bearer is always used for SIP signaling, and exactly one dedicated EPS bearer is used for voice packets regardless of the number of active voice media streams. The XCAP signaling may be transported over its own dedicated EPS bearer or it may be multiplexed with the SIP signaling on the default EPS bearer, in which case only two EPS bearers are utilized.

The VoLTE-compatible mobile devices are required to operate in dual-stack IPv4/IPv6 mode (i.e., IPv4 and IPv6 networks run in parallel) by default [28]. If the IMS application profile assigns an IPv6 address, the device is required to prefer that address and specifically to utilize it during P-CSCF discovery. Since IP header overhead in VoIP calls can be considerable, VoLTE mandates IP header compression and the use of RoHC protocol for voice packets.

In the EPS network, mobility is implicitly handled for VoLTE by the defined mobility management mechanisms. A challenge arises during mobility to/from UMTS/GSM. UMTS/GSM utilizes a different call model than VoLTE. The call processing and service delivery entities are the MSC/Visited Location Register (VLR) and HLR rather than the CSCFs and MMTel servers. Furthermore, the bearer/transport channels used to exchange voice packets are circuit-switched and not packet-switched. Part of the solution to this problem is known as IMS centralized services. With IMS centralized services, voice calls are always anchored in the IMS, regardless of the domain from which they are originated. This solves the problem of transfer of call control as the mobile device moves between different radio access technologies; nevertheless, the packet-switched bearer challenge still remains. This problem inspired the concept of packet-switched over circuit-switched in UMTS and GSM, where a circuit-switched bearer is utilized to transport voice packets for calls anchored in the IMS. Voice call continuity during radio handover between LTE and the UMTS/GSM circuit-switched domain is managed by an SRVCC mechanism.

In conjunction with support for voice services, VoLTE-compatible devices also must support sending and receiving SMS over IP and instant messaging services. Furthermore, VoLTE-compatible networks must be capable of functioning as an IP short message gateway. Note that SMS over IP represents an alternative mechanism for providing short messaging services to users. Since mobile devices might support both of these SMS over LTE solutions, the operators are required to provision the preferred SMS transport via the IMS management object [53,64].

2.11.4 Circuit-switched fallback

Circuit-switched fallback (CSFB) is an alternative mechanism to move a subscriber from LTE to a legacy radio access technology in order to use circuit-switched voice services (see Figure 2.24). This function is only available, if LTE network coverage is overlapped by GSM, UMTS, or CDMA2000 coverage. The CSFB was already specified in 3GPP Rel-8, with further enhancements defined in later releases [5]. A number of different CSFB mechanisms are available and there are also differences depending on the radio access technology to which the subscriber falls back. The CSFB to GSM, UMTS, and CDMA2000 is defined, but no mechanisms are specified to support CSFB to both UMTS/GSM and CDMA2000 in the same PLMN. In this section, we describe the CSFB mechanisms for UMTS and GSM followed by CSFB procedure to CDMA2000. As mentioned earlier, depending on the target circuit-switched radio access technology, several CSFB solutions exist. As an example, there are three different solutions for CSFB to UMTS, two using the RRC connection release with redirection mechanism, and one using the packet-switched handover mechanism. For CSFB to GERAN, in addition to the RRC connection release with redirection and the packet-switched handover mechanisms, cell change order with or without network-assisted cell change can be used.

The CSFB affects the radio and the core networks. Figure 2.25 illustrates the modified EPS architecture with support for CSFB. It includes the interfaces between the different radio access network types and the core network entities. Note that UTRAN is the UMTS access network and GERAN is the GSM/EDGE access network. To support circuit-switched services, a connection to the MSC server must be established. The MME interfaces with the MSC server via the SGs interface. The CSFB mechanism is implemented using the SGs interface.

The legacy 2G/3G and the EPS can coexist in the form of hybrid networks in the transition period of network upgrades, where the MME serves LTE users, the SGSN provides service to UMTS users when utilizing data services, and the MSC server serves the users when utilizing circuit-switched voice services. The MSC server connects to the operator's telephony network. To support CSFB signaling and SMS transfer for LTE devices, the MME connects to the MSC server. The architecture shown in Figure 2.25 is a simplified schematic of the hybrid E-UTRAN and UTRAN/GERAN networks. The SGs interface between the MSC server and the MME enables the UE to register with both circuit-switched and

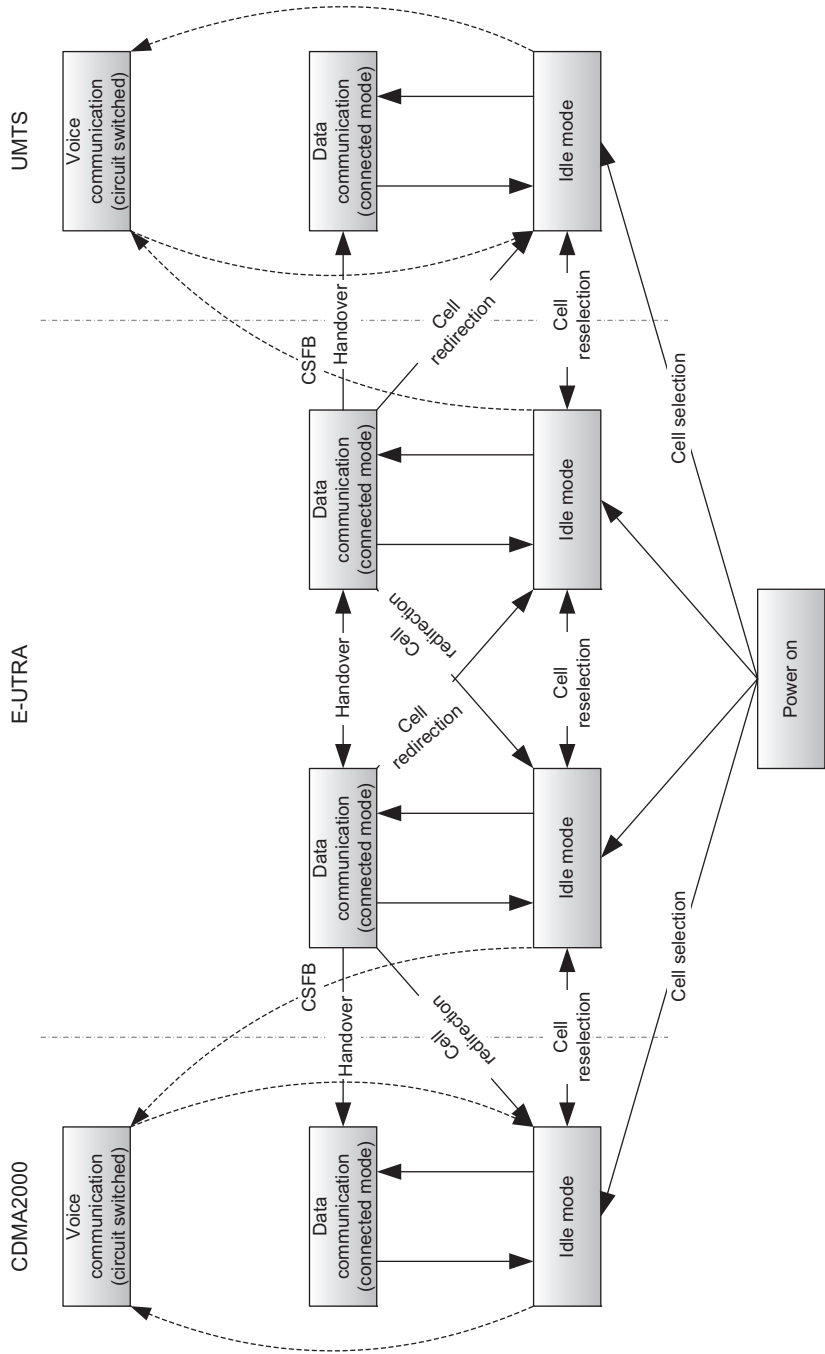


FIGURE 2.24
The concept of CSFB to GERAN or UTRAN [39].

packet-switched networks while on the LTE access network. This interface also enables the delivery of circuit-switched paging messages via the LTE access, as well as SMS, without having the device leave the LTE network. With the default LTE data network connection in operation, a mobile-terminated or incoming circuit-switched voice call triggers a paging message via the LTE network to the UE, as shown in Figure 2.26. The paging message initiates the CSFB operation, as the device sends an extended service request to the network to transition to UMTS/GSM access networks.

Upon transition to circuit-switched mode, the legacy call setup procedures are followed to set up the circuit-switched call. Mobile-originated or outgoing calls follow the same transition from packet-switched LTE to circuit-switched UMTS/GSM, except for the paging step, which is not needed. In UMTS networks, packet-switched data sessions can also move for simultaneous voice and data services. In GSM networks, packet-switched data sessions may be suspended until the voice call ends and the device returns to LTE mode, unless the GSM network supports dual transfer mode, which permits simultaneous voice and data. When the voice call ends, the device returns to LTE mode via idle mode or connected mode mobility procedures, as shown in Figure 2.26.

In EPS network architecture, the CSFB enables the provisioning of voice and traditional circuit-switched domain services. To provide these services, the EPS reuses circuit-switched infrastructure when the UE is served by E-UTRAN. A CSFB-enabled terminal, connected to E-UTRAN, may use GERAN or UTRAN to connect to the circuit-switched domain. This function is only available when the E-UTRAN coverage is overlapped by GERAN or UTRAN coverage. Figure 2.25 illustrates the architecture for CSFB in EPS. The CSFB and IMS-based services can coexist in the same operator's network. Although it is not very straightforward to support CSFB, all participating elements, i.e., UE, MME, MSC, and E-UTRAN must be modified to support additional functionalities.

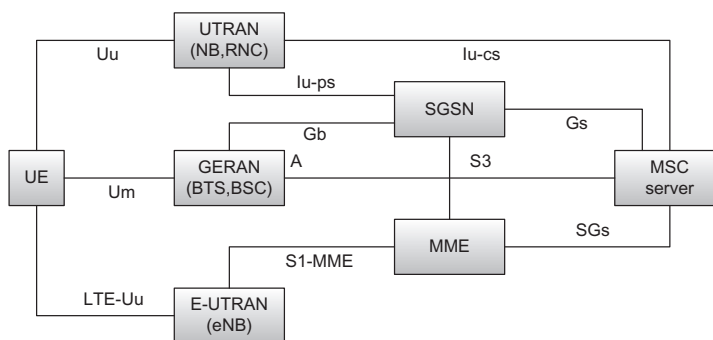


FIGURE 2.25

Modified EPS network architecture to support CSFB [5].

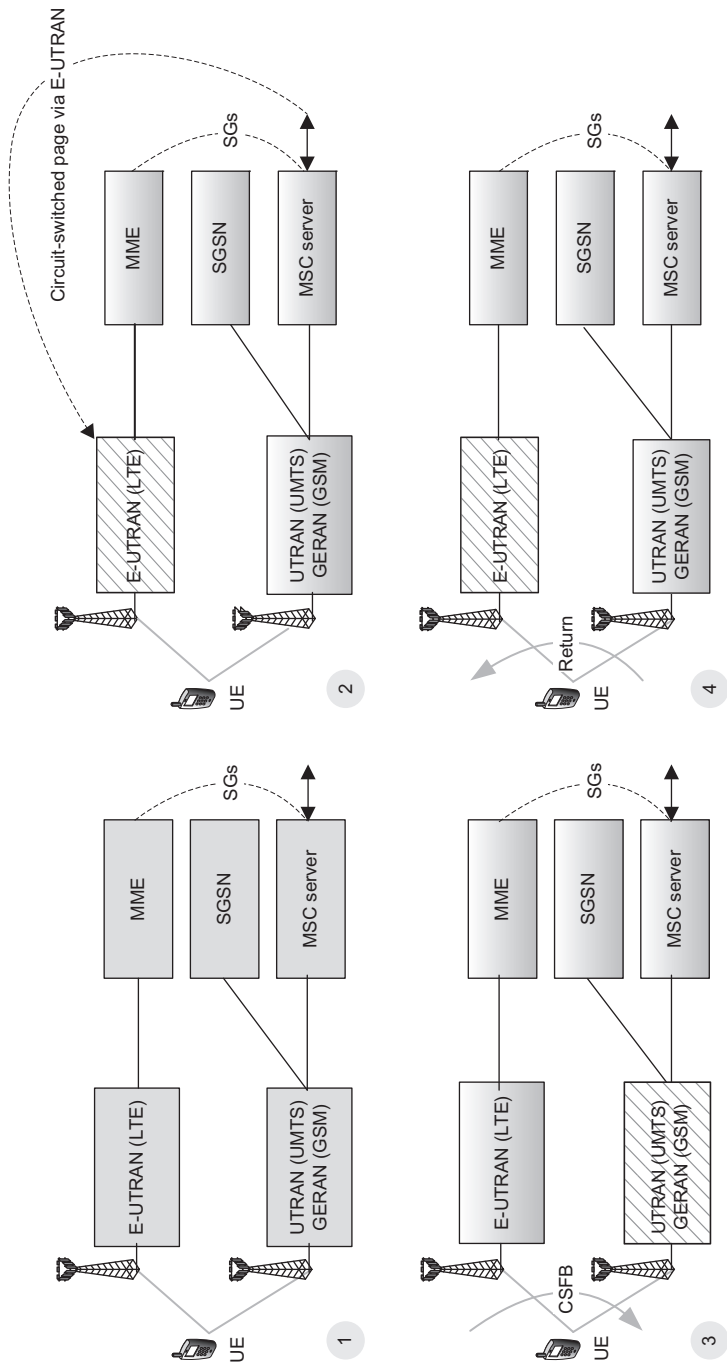


FIGURE 2.26

CSFB call processing in the hybrid network [63].

The CSFB support in EPS includes a new interface SGs, which is the reference point between the MME and MSC server. The SGs interface is used for mobility management and paging procedures between the EPS and the circuit-switched domain. The SGs reference point is also used for delivery of both mobile-originated and mobile-terminated SMS service. The CSFB-enabled network elements are required to support the following additional functions [54]:

- **UE:** Support access to E-UTRAN/EPC as well as access to the circuit-switched domain over GERAN and/or UTRAN; combined procedures for EPS/IMSI attach, update, and detach; and CSFB and SMS procedures for using circuit-switched domain services.
- **MME:** Deriving a VLR number and location area identity (LAI)¹¹ from the GUTI received from the UE or from a default LAI; maintaining of SGs association toward MSC/VLR for an EPS/IMSI attached UE; initiating IMSI detach at EPS detach; initiating paging procedure toward eNB when MSC pages the UE for circuit-switched services; support of SMS procedures; rejecting CSFB call request; and use of the LAI and a hash value from the IMSI to determine the VLR number when multiple MSC/VLRs serve the same LAI.
- **MSC:** Maintaining SGs association toward MME for an EPS/IMSI attached UE and support of SMS procedures specified by 3GPP.
- **E-UTRAN:** Forwarding paging request and SMS to the UE and directing the UE to the target circuit-switched-capable cell.

A CSFB and IMS-capable UE would follow the procedures for domain selection for UE originating session/calls. If a UE is configured to use SMS over IP services and it is registered to IMS, then it would send SMS over IMS, even if it is EPS/IMSI attached. The home operator has the option to activate/deactivate the UE configuration to use SMS over IP by means of device management in order to allow alignment with home PLMN support of SMS over IP. When UE performs CSFB procedure for a mobile-originated call for the purpose of an emergency call, it needs to indicate to the MME that this CSFB request is for an emergency purpose.

When the UE is paged via LTE with an incoming call, or when the user initiates an outgoing call, the device switches from LTE to UMTS/GSM mode. The acquisition of the UMTS/GSM network and setup of the call can be done in one of following two ways: handover or redirection. In the handover procedure, the target cell is prepared in advance and the device can enter that cell directly in the connected mode. Inter-RAT measurements of signal strength may be conducted by the UE while on the LTE system prior to executing the handover. In the redirection procedure, only the target frequency is indicated to the device.

¹¹Each location area of a public land mobile network has its own unique identifier which is known as its LAI. This internationally unique identifier is used for location update of mobile subscribers. It is composed of a three decimal digit mobile country code, a two to three digit mobile network code that identifies the PLMN in that country, and a location area code which is a 16 bit number, thereby allowing 65,536 location areas within one PLMN.

The device is then allowed to select any cell on the indicated frequency, or may even try other frequencies/RATs, if no cell can be found on the target frequency. Once a cell is found, the device initiates normal call setup procedures. Inter-RAT measurements of signal strength are not needed prior to redirection. Consequently, CSFB with redirection may require less time to identify the best cell compared to the handover procedure [49,63,76].

Redirection-based CSFB has variations with differing call setup latencies. In 3GPP Rel-8 basic release and redirection approach, the UE follows LTE Rel-8 procedures and reads the system information which is broadcast through system information block (SIB) messages prior to accessing the target cell. The device can alternatively follow LTE Rel-8 procedures but may read only the mandatory SIBs 1, 3, 5, and 7, and skip other SIBs prior to access to the target cell. In the latter case, the neighbor cell information in SIB11 is delivered to the UE via measurement control messaging once the UE is in the connected mode on the target cell. This approach can be either implicitly supported by the UE and the network, or explicitly through deferred measurement report control signaling.

In 3GPP Rel-9 enhanced connection release with redirection and system information tunneling approach, the device follows LTE Rel-9 procedures, where SIB information can be tunneled from the target access network via the core network to the source access network, and can be included in the redirection message sent to the device. This can avoid reading any SIBs on the target cell. The solutions deployed today are predominantly based on 3GPP Rel-8 connection release with redirection with SIB skipping, in order to achieve acceptable call setup times, reliability, and to simplify deployments. While it is anticipated that 3GPP Rel-9 enhanced connection release with redirection will be deployed soon, a handover-based CSFB approach has not found any traction, since both release with redirection with SIB skipping and enhanced connection release with redirection with system information tunneling can address any call setup latency issues that may arise as a result of the basic release with redirection solution.

Another key issue for the voice call user experience is call setup reliability, which is the ability to successfully establish an incoming or outgoing call on the first attempt, or within a time frame that does not indicate call setup failure. The target is to at least match legacy performance, which is in the 98% range. With CSFB switching between LTE and UMTS networks, there are two primary challenges to call setup success: changes in inter-RAT conditions between measurement and acquisition for handover-based CSFB, and mismatches between LTE and 3G geographic signal coverage areas that require server-side updates. With handover-based CSFB, inter-RAT measurements can change between the time the measurement is conducted using LTE and the time UMTS/GSM/CDMA2000 voice network acquisition is attempted. During this time, the cell identified and prepared for handover may become unavailable, resulting in connection failure [63].

In handover-based CSFB, the measurement is conducted before the inter-RAT transition. If the inter-RAT conditions adversely change, there is a higher probability of handover failure, especially in high mobility scenarios. In practice, it is

shown the failure rate of inter-frequency handover or inter-RAT handover are relatively higher, suggesting that delays between inter-RAT measurement and network acquisition can increase setup failures. In contrast, redirection-based CSFB can result in higher call setup success rates relative to handover-based CSFB, because redirection-based CSFB solutions conduct inter-RAT measurements immediately before attempting access to the identified cell. Since the 3GPP Rel-9 system information tunneling and 3GPP Rel-8 redirection-based CSFB methods have shown satisfactory call setup times and higher reliability, the industry has taken the redirection-based CSFB approach [63].

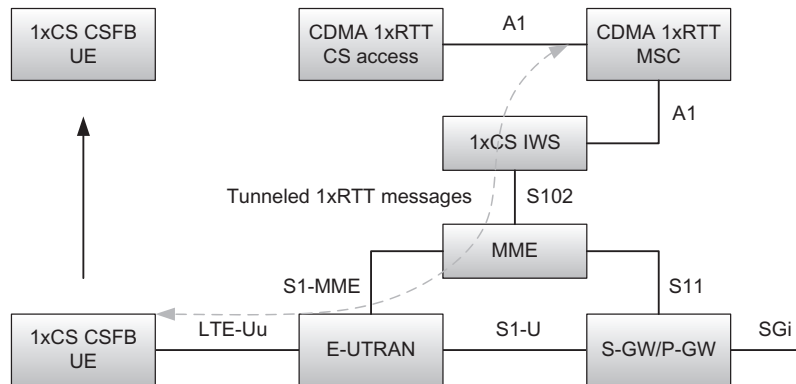
CSFB to CDMA 1xRTT is supported for UEs with both single and dual receiver chain configurations. UEs with single receiver configuration are not able to camp in CDMA 1xRTT when they are active in E-UTRAN. The EPC provides mechanisms for the UE to perform registration procedure with CDMA 1xRTT, receive paging message, and send/receive SMS, while the UE is camped in E-UTRAN. UEs with dual receiver configuration can camp in CDMA 1xRTT while they are active in E-UTRAN, they may not be able to stay in E-UTRAN when they handle a circuit-switched call and/or perform registration signaling, and/or are sending or receiving SMS in CDMA 1xRTT.

The CSFB for CDMA 1xRTT in EPS enables the delivery of circuit-switched domain services (e.g., circuit-switched voice) by reusing CDMA 1xRTT circuit-switched infrastructure when the UE is served by the E-UTRAN. A CSFB-enabled terminal, while connected to E-UTRAN, may register in the CDMA 1xRTT circuit-switched domain in order to be able to use CDMA 1xRTT access to establish one or more circuit-switched services in the circuit-switched domain. The CSFB function is only available where E-UTRAN coverage overlaps with CDMA 1xRTT coverage. 3GPP specifications also specify the architecture for SMS over S102 interface in EPS. The mobile-originated or terminated SMS is tunneled in EPS over S102 and does not cause any CSFB to CDMA 1xRTT and consequently does not require any overlapped CDMA 1xRTT coverage. The CSFB to CDMA 1xRTT and IMS-based services can coexist in the same operator's network [5].

The CSFB to CDMA 1xRTT and SMS over S102 in EPS function are realized by reusing the S102 reference point between the MME and the 1xCS interworking solution (IWS) [16–18]. The reference architecture described in Figure 2.27 is similar to the SRVCC architecture for E-UTRAN to 3GPP2 1xCS, with the additional feature that the S102 session is extended similar to preregistration for S101. The S102 is the reference point between the MME and the 1xCS IWS which provides a tunnel between MME and 3GPP2 1xCS IWS to relay 3GPP2 1xCS signaling messages [14–18].

2.12 LIPA, selected IP traffic offload, and IP flow mobility

Femto-cells have been recently introduced to provide enhanced indoor coverage and to offload macro-cells. 3GPP Rel-10 network architecture introduced three

**FIGURE 2.27**

Reference architecture for CSFB to CDMA 1xRTT [5].

important concepts, namely, LIPA, selected IP traffic offload (SIPTO), and IP flow mobility [57,59].

2.12.1 Local IP access

LIPA is the ability for an IP-enabled UE to directly access a consumer's home local area network as well as the Internet (through the operator's IP core network) using the radio-link of a femto-cell or a home eNB (HeNB). The use of LIPA enables improved performance, more innovative services that interconnect mobile stations and home local area networks, and offloading of traffic from the operator's IP core network which is ultimately through the Internet. Enabling LIPA benefits mobile operators, subscribers, and Internet service providers. For mobile operators, the ability to offer LIPA via an HeNB would mean higher revenues by way of value-added services without having to invest significantly on upgrading network infrastructure to cope with the higher throughput and speed that such applications would demand. For subscribers, LIPA would mean faster and more secure data transfer, since traffic within the home network would not travel outside the subnet. Users could add customized high-speed applications such as file transfer, video streaming, and device sharing without involving the operator's core network, thus avoiding associated bottlenecks and possibly yielding lower data costs.

The requirements for LIPA can be generally classified into two main categories: LIPA to the home network and IP access to the Internet. For LIPA to the home network, the UE should be able to exchange data via the operator's core network and LIPA to the home-based network simultaneously, depending on the destination. The HeNB should decide, based on certain criteria, whether to route the data via the core network or to use LIPA. Furthermore, the UE should

be able to access other devices in the home network via the HeNB using the HeNB only as an access medium. Access to local IP through the HeNB E-UTRAN interface should only be granted to UEs with a valid subscription. This means that no specific modifications need to be done to the UE to support LIPA. In addition, it should be possible for a device in the home-based network to contact a UE via LIPA. Therefore, there is not any impact to the UE in accessing packet data services through the cellular core network, resulting in minimal changes to the protocol stacks in the core network.

IP access to the Internet should be possible without going through the operator's network. Simultaneous access from a UE to both the operator's core network and LIPA to the Internet are also possible. The operator or the HeNB owner, within the limits set by the operator, should be able to enable/disable LIPA to the Internet per HeNB. Furthermore, LIPA to the Internet should support a scenario where the HeNB and backhaul are provided by the same operator, as well as a scenario where the HeNB and backhaul are provided by different operators.

Basic functionality for LIPA was specified in 3GPP Rel-10. LIPA signifies the capability of a UE to obtain access to a local residential/enterprise IP network, subsequently called a local network, that is connected to one or more HeNBs. The LIPA functionality is being extended in 3GPP Rel-12 to allow access to the local network when a UE is in the coverage area of a macro-eNB and to provide mobility support (see Figure 2.28). LIPA allows a UE to work with devices in the local network, e.g., printers, video cameras, or a local web-server. LIPA allows the UE to discover supporting devices and to be discovered, and does not require the local network to be connected to the Internet but achieves IP connectivity with the UE through one or more HeNBs of the mobile operator. 3GPP Rel-10 only specified the support of LIPA when the UE accesses the local network via HeNB.

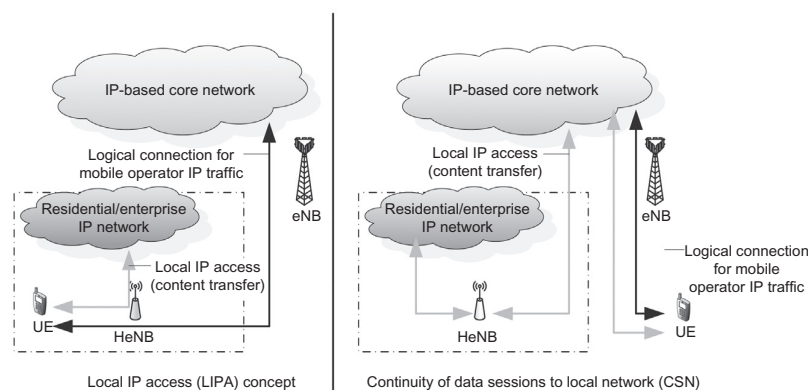


FIGURE 2.28

LIPA and continuity of data sessions concepts [73].



Network architecture to support LIPA or SIPTO over local network with L-GW collocated with the HeNB [73].

An operator may want to further provide access to the local network to a UE that is in the coverage area of a macro network. In 3GPP Rel-10, the UE was required to be able to maintain IP connectivity to the local network when moving between HeNBs within the same local network. However, access to the local network may be lost as a UE moves out of HeNB coverage into the macro network, even if other services (e.g., telephony, data services, and SIPTO) may continue after a handover to the macro network, resulting in an unsatisfactory user experience. A 3GPP Rel-12 study item will allow continuation of data sessions to the local network, when the UE moves between HeNB and the macro network (see [Figure 2.28](#)).

From an architectural perspective, a HeNB subsystem comprises a HeNB entity and it may optionally include a HeNB gateway (HeNB-GW) and a local gateway (L-GW). The LIPA and SIPTO functionalities over the local network are achieved using an L-GW that may be collocated with the HeNB. Figure 2.29 illustrates the network architecture to support LIPA and/or SIPTO over the local network with L-GW function collocated with the HeNB. It is shown that the HeNB subsystem is connected by means of an S1 interface to the EPC, and more specifically to the MME by means of an S1-MME interface and to the S-GW by means of an S1-U interface. When LIPA or SIPTO functions over the local network are activated (when the L-GW is collocated with the HeNB), the L-GW has an S5 interface with the S-GW. The local GW is the gateway toward the IP networks (e.g., residential/enterprise networks and Internet) associated with the HeNB. The L-GW performs some P-GW type functions including UE IP address allocation; DHCPv4 (server and client), and DHCPv6 (client and server).

functions; packet screening, as well as ECM-IDLE mode downlink packet buffering; and ECM-CONNECTED mode direct tunneling toward the HeNB [6].

The architecture for SIPTO over the local network with L-GW function collocated with a HeNB is depicted in Figure 2.29. For LIPA traffic, the addition of an L-GW function in EPS that can be either collocated with the HeNB or as a standalone node can provide partial P-GW and S-GW downlink data buffering functions. The S-GW in the core network serves the core network traffic. The L-S11 interface between the L-GW and the MME is used to manage the session for LIPA traffic. For the LIPA PDN connection, the L-GW needs to be located close to the HeNB, and to establish the connection through an L-S11 interface. When the UE is in the idle mode, the LIPA downlink packets are buffered in the L-GW. Upon UE transition to the connected mode, the packets buffered in L-GW are forwarded over the path between the L-GW and the HeNB. It is assumed that the user-plane does not go through the HeNB-GW for the direct tunnel [6].

The LIPA is realized using the L-GW entity that may be collocated with the HeNB. LIPA is established when the UE requests a new PDN connection to an APN for which LIPA functionality is permitted. In this case, the network selects the L-GW associated with the HeNB and enables a direct user-plane path between the L-GW and the HeNB. A HeNB supporting the LIPA function includes the L-GW address in the UE-related control messaging with the MME. The protocol option (i.e., GTP or PMIP) supported on the S5 interface between L-GW and S-GW is configured by the MME. As of 3GPP Rel-11, no interface between the L-GW and the PCRF is specified, and there is no support for dedicated bearers on the PDN connection used for LIPA. The L-GW rejects any UE requests for bearer resource modification.

The direct user-plane path between the HeNB and the collocated L-GW is enabled with a correlation ID parameter that is associated with the default EPS bearer on a PDN connection used for LIPA. Upon establishment of the default EPS bearer, the MME sets a correlation ID equal to the P-GW TEID (GTP-based S5) or the P-GW generic routing encapsulation (GRE) key (PMIP-based S5). The correlation ID is then signaled by the MME to the HeNB as part of E-RAB establishment and is stored in the E-RAB context in the HeNB. The correlation ID is used in the HeNB for matching the radio bearers with the direct user-plane path connections from the collocated L-GW. If the UE is roaming and if the HSS allows LIPA roaming for this UE in the visited PLMN, then the MME may provide LIPA for this UE.

LIPA is supported for APNs that are valid only when the UE is connected to a specific CSG. The MME releases a LIPA PDN connection to an APN, if it detects that the UE's LIPA CSG authorization data for this APN has changed and the LIPA PDN connection is no longer allowed in the current cell. As mobility of the LIPA PDN connection is not supported in 3GPP Rel-11, the LIPA PDN connection is released when the UE moves away from the HeNB coverage area. Prior to handover procedure, the HeNB requests through an intra-node signaling with the collocated L-GW to release the LIPA PDN connection. The HeNB determines

that the UE has a LIPA PDN connection from the presence of the correlation ID in the UE E-RAB context. The L-GW then initiates and completes the release of the LIPA PDN connection using the P-GW-initiated bearer deactivation procedure. During the handover, the source MME checks whether the LIPA PDN connection has been released. If it has not been released and the handover is an S1-based or an inter-RAT handover, the source MME rejects the handover. The direct signaling from the HeNB to the L-GW is only possible since mobility of the LIPA PDN connection is not supported. During idle state mobility events, the MME/SGSN deactivates the LIPA PDN connection, when it detects that the UE has moved away from the HeNB coverage area [6].

2.12.2 Selected IP traffic offload

The SIPTO refers to the ability to selectively forward different types of traffic via alternative routes to/from the terminal. Home network flavor of SIPTO depends on a HeNB. In this mode, specific traffic, determined by the operator's policy and/or subscription, is transferred to/from eNB directly to Internet/intranet, bypassing the mobile operator access/core network. Macro network flavor of SIPTO covers the ability to offload traffic next to an RNC or S-GW as opposed to traversing the operator's core network. SIPTO extends what LIPA aims to achieve by providing an alternate path for selected traffic. The primary objective is to add capacity to the primary RAN-core network path and reduce the unit cost of transport.

SIPTO for HeNB is similar to LIPA in terms of necessary changes to network capabilities and additional functionality in the form of an L-GW collocated with HeNB (Figure 2.30). Considering substantial complexity and limited benefit of SIPTO for HeNB as opposed to LIPA, 3GPP decided to mark that as a future study item beyond Rel-10. The SIPTO function enables an operator to offload certain types of traffic at a network node close to the UE's point of attachment to the access network. SIPTO beyond E-UTRAN can be achieved by selecting a set of network gateways (S-GW and P-GW) that is geographically/topologically close to a UE's point of attachment. SIPTO beyond E-UTRAN corresponds to a traffic offload through a P-GW located in the mobile operator's core network. SIPTO applies to the non-roaming, provided that appropriate roaming agreements are in place between the operators, and the roaming cases. Offload of traffic for a UE is available for UTRAN and E-UTRAN accesses only. When the UE enters to UTRAN/E-UTRAN from another type of access network (e.g., from GERAN), it is the responsibility of the new SGSN/MME to decide whether to perform deactivation with reactivation request for a given PDN connection, depending on whether SIPTO is permitted for that APN. Realization of SIPTO beyond E-UTRAN relies on the same architecture models and principles that govern the local breakout.

In order to select a set of appropriate network gateways (S-GW and P-GW) based on geographical/topological proximity to the UE, the gateway selection function uses the UE's current location information. In order to allow or prohibit SIPTO on per user and per APN basis by an operator, subscription data in the

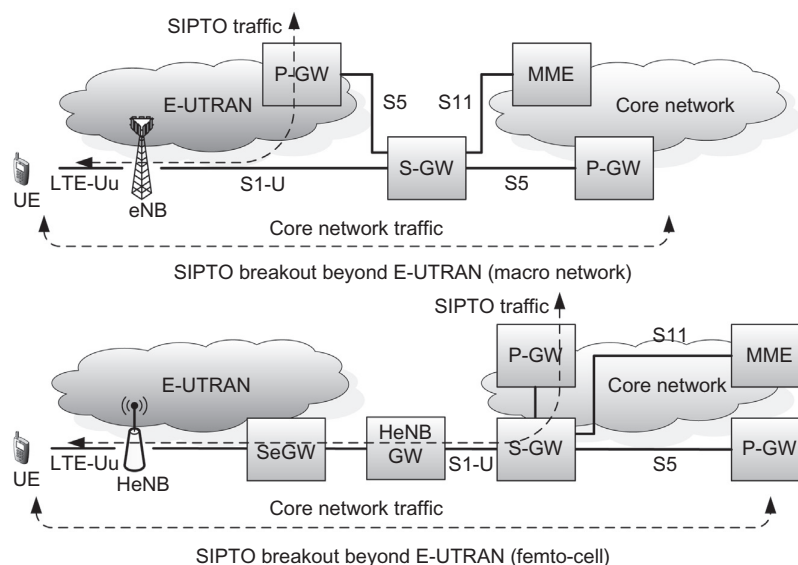
**FIGURE 2.30**

Illustration of SIPTO variants [62].

HSS is configured to indicate to the MME if offload is allowed or prohibited. If the SIPTO permission information from the HSS conflicts with MME's configuration for that UE, then SIPTO is not used. If HSS indicates the visited PLMN address is not allowed, then the MME does not provide SIPTO. The MME may be configured on an APN basis in order to use SIPTO (e.g., to handle the case where the HSS is not configured with SIPTO information for the UE). For SIPTO beyond E-UTRAN, as a result of UE mobility (e.g., detected by the MME at tracking area update procedure), the target MME may redirect a PDN connection toward a different gateway that is more appropriate for the UE's current location.

The SIPTO at the local network function enables an IP-capable UE connected via a HeNB to access a specific IP network or the Internet without the user-plane going through the mobile operator's network. The subscription data in the HSS is configured per user and per APN to indicate to the MME if offload at the local network is allowed. SIPTO at the local network can be achieved by selecting an L-GW function collocated with the HeNB or selecting stand-alone network gateways (with S-GW and L-GW collocated) residing in the local network. In both cases, the selected IP traffic is offloaded via the local network. Specific to the HeNB sub-system, the applicability of SIPTO at the local network does not depend on CSG membership and the feature can be applied to any UE, as long as the HeNB is configured for open or hybrid access mode. There is no interface between the L-GW and the PCRF, and there is no support for dedicated bearers

on the PDN connection used for SIPTO at the local network. The L-GW rejects any UE request for bearer resource modification. In 3GPP Rel-11, SIPTO at the local network is intended for offloading Internet traffic only, thus the L-GW does not provide APN-specific connectivity. Therefore, if the subscription data in the HSS indicates that offload at the local network is allowed, this implies that the related APN is typically used for providing Internet connectivity.

2.12.3 IP flow mobility

An IP flow consists of a set of packets that are exchanged between two nodes and that match a given flow description. In 3GPP Rel-8 mobility architecture, the granularity of access system connectivity was per PDN basis, that is, the UE could not access the same PDN via multiple interfaces, which meant that the UE could not treat individual IP flows within a PDN connection separately; therefore, while performing handover from one access network to another, it had to move all the IP flows within a PDN connection together [57]. In other words, in Rel-8 EPS, the UE could not communicate using multiple radio access networks simultaneously. While the UE could establish a single PDN connection or multiple simultaneous PDN connections, all traffic exchanged by the UE, regardless of the PDN connection it belonged to, was routed through the same radio access network.

IP flow mobility introduced the concept of treating IP flows individually within a PDN connection. It specified that the mobility of a PDN connection is handled per IP flow, which meant that within a PDN connection, the UE could establish IP flows over multiple access networks; selectively remove IP flows from an access system; selectively transfer IP flows between access systems; and transfer all IP flows from a certain access system. The following assumptions are valid while defining the functionality of multi-access PDN connectivity and IP flow mobility: (1) the number of non-3GPP interfaces simultaneously active with a 3GPP interface is limited to one; and (2) all IP flows are toward one, and only one PDN. However, in this framework a single IP flow cannot be split and routed via different access networks, and IP flows cannot be established toward multiple PDNs. The following use cases have been identified for IP flow mobility in 3GPP:

- Adding IP flow to an existing one via a different access network: The UE is reachable through both 3GPP and non-3GPP access networks. It has an active IP flow through the 3GPP (or non-3GPP) access networks. At some point in time, a new IP flow is initiated through non-3GPP (or 3GPP) access toward the same PDN. The non-3GPP (or 3GPP) access of the UE may or may not have any ongoing session through it, as depicted in Figure 2.31.
- Removing IP flow established via a different access network: The UE is reachable through both 3GPP and non-3GPP access networks and has simultaneously active sessions through both RANs. Each of the RANs can have one or more ongoing sessions. At some point in time, one IP flow via non-3GPP (or 3GPP) access is removed.

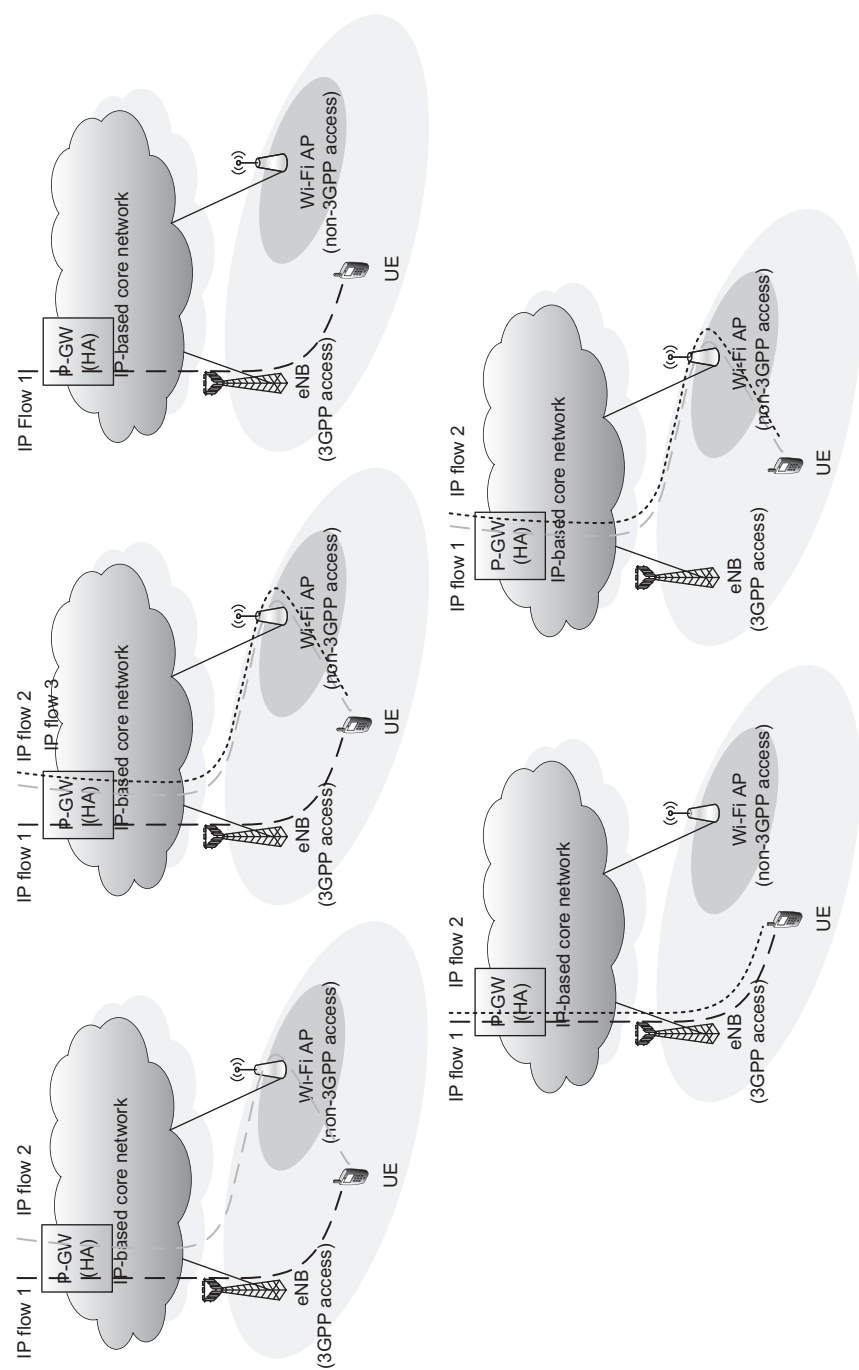


FIGURE 2.31

Adding, removing, and transferring of IP flows in various IP flow mobility scenarios using 3GPP and non-3GPP access networks [57].

- IP flow mobility between access networks: this use case involves selective transfer of IP flows between RANs. The state of the UE's interfaces can be one of the followings: (1) both radio interfaces are active simultaneously; or (2) only one radio interface is active. In the case of both radio interfaces being active simultaneously, the UE is reachable through both 3GPP and non-3GPP RANs and has simultaneously active sessions through both RANs. It has two IP flows through the 3GPP (or non-3GPP) access and one IP flow through the non-3GPP (or 3GPP) RAN. At some point in time, one IP flow is moved from the 3GPP (or non-3GPP) RAN to the non-3GPP (or 3GPP) RAN. The use case is depicted in Figure 2.31. In the case of only one active radio interface, the UE is camping on a 3GPP (or non-3GPP) RAN with at least two active IP flows. The UE is moving toward the coverage of a non-3GPP RAN. At some point in time, the UE detects a non-3GPP (or 3GPP) RAN and moves a subset of active IP flows (i.e., one IP flow) to the newly detected RAN.
- Moving all IP flows from one access network to another: The UE is reachable through both 3GPP and non-3GPP RANs. It has one or more active IP flows through the non-3GPP access. The 3GPP access of the UE may or may not have any ongoing session. The UE is moving away from the coverage of the non-3GPP RAN. At some point in time, due to loss of coverage of the non-3GPP RAN, the UE moves all IP flows going through the non-3GPP access to the 3GPP access.

As mentioned earlier, the granularity of access network connectivity and inter-system mobility is per PDN connection basis. This implies that when a handover occurs, all IP flows belonging to the same PDN connection are moved from the source access system to the target access system. With IP flow mobility, it is possible to have a finer granularity in access network connectivity and inter-system mobility. The handover procedures can be applied to a single or multiple IP flows belonging to the same PDN connection. This implies that some IP flows of one PDN connection can be routed via one access network, while simultaneously some IP flows of the same PDN connection can be routed via another access system. To achieve IP flow mobility, the inter-system mobility signaling is enhanced in order to carry routing filters. The extensions to DSMIPv6 mobility signaling are needed to carry routing filters when the UE is connected to multiple access networks simultaneously [4].

There are two possible approaches to IP flow mobility, client-based and network-based. The former relies on an IP host-centric solution, introducing a mobility client in the host and a mobility agent in the core network. The latter relocates the IP mobility client functionality from the host to the network, thus making the mobile device agnostic to any IP mobility signaling. Client-based IP mobility solutions require the user terminal to be involved in the management of the mobility, running a specialized stack that is able to detect, signal, and react to changes of point of attachment. Dual stack for mobile IPv6 is standardized by the IETF to provide client-based IP mobility support [29]. Network-based IP mobility solutions locate the mobility management control of the terminal in the network.

In this way, the terminal is not required to perform any kind of signaling (e.g., binding updates) to react to changes of its point of attachment to the network, as these changes are transparent to the mobile terminal IP stack. Proxy MIPv6 (PMIPv6) is the protocol standardized by the IETF to provide network-based IP mobility support [59].

2.12.4 Wi-Fi offloading

Wi-Fi offload offers a new way to extend the capacity and coverage of an LTE network. Wi-Fi networks are integrated to complement LTE networks and allow the use of each technology's advantages according to the actual demand. For example, the Wi-Fi is optimized for indoor environments, whereas the LTE network is designed for complete coverage in all areas. Depending on the trust relationship, the Wi-Fi may be integrated as a trusted or a non-trusted access technology. In the newest releases of 3GPP network architecture, a paradigm change has occurred. Instead of a handover for the complete connection, single traffic flows are routed over one access technology while the remaining ones are routed over the other. Therefore, the elaborated QoS feature in LTE may be used for delay or jitter sensitive IP flows, such as voice or video conference, whereas less time critical services may be routed over the cost-effective Wi-Fi when available. Corresponding handover procedures for the partial offload of certain flows are controlled by the network operator to ensure best user experience.

The user authentication is done using USIM credentials and according to LTE authentication procedures. Therefore, in the ideal case, the user does not perceive the offload and only recognizes an enhanced data rate using the mobile services. The Wi-Fi offloading refers to a scenario where part of the IP traffic to or from a device is routed via Wi-Fi access and another part via 3GPP access. Wi-Fi offloading covers both scenarios where the IP traffic via Wi-Fi radio is anchored in the EPC (i.e., seamless offloading) and the scenario where it is not anchored (i.e., non-seamless offloading). There is an access network discovery and selection function (ANDSF) entity to provide the UE with the access network discovery information and the policy on how to use the available access networks, available access networks, and preferred routing of the IP traffic per APN and per IP flow [58,61].

3GPP Rel-8 introduced a client-server-based protocol, dual stack mobile IP (DSMIP), to enable seamless handover between 3G and Wi-Fi. DSMIP is a mobility protocol specified in IETF that provides IP address preservation for IPv4 and IPv6 sessions, allowing the user to roam independently in IPv4 and IPv6 accesses. The solution does not require any support from Wi-Fi access network. If the cellular radio access network supports a home agent (HA), the only new requirement is that the UE (i.e., client) and the HA (i.e., server) are dual-stack capable. Applications and accesses can be either IPv4 or IPv6 compliant depending on the deployment models of the operators. Therefore, the changes required are only made to the UE and GGSN/P-GW. The HA can be implemented as a functionality in the GGSN/P-GW or a stand-alone entity. The HA is the anchor

that binds the permanent identifier of the node (i.e., home address or HoA) with the local address based on the node's location (i.e., care-of-address or CoA). The exposed IP address accessible by the application remains the same.

The DSMIP can be used over trusted and/or untrusted non-3GPP access (e.g., Wi-Fi) and/or 3GPP access (e.g., HSPA and LTE). Wi-Fi mobility in 3GPP Rel-8 can be implemented either using DSMIP over an H1 or S2c interface. The H1 interface is compatible with pre-Rel-8 GGSN and PDG, and the S2c interface is part of the EPS/LTE architecture. The DSMIP provides a better user experience compared to application-based switching, since it preserves the IP address whenever the network is changed. Furthermore, the DSMIP is easily portable to multiple devices since it does not depend on the application.

3G/Wi-Fi seamless offload is a further enhancement to the solution based on DSMIP that enables seamless handover between 3G and Wi-Fi. It also provides the possibility to move selected IP traffic (i.e., HTTP, video, and VoIP) while supporting simultaneous 3G and Wi-Fi access networks. The main components of the solution are DSMIP with extensions for seamless IP flow mobility, ANDSF extensions for providing operator control, and PCC enhancements for QoS [61]. The 3G/Wi-Fi seamless offload solution does not require any PCC support. When IP flow mobility is used and PCC function is deployed, the PCC architecture is enhanced to handle multiple simultaneous connections for a single IP session.

In dual radio scenarios, such as Wi-Fi and cellular, the two access networks are not aware of each other and one access network cannot control the device protocol state in the other access network. Also, Wi-Fi is not a controlled access network and works opportunistically in unlicensed spectrum; therefore, the only element in the network that is known is when a Wi-Fi is available, the radio quality and the performance that can be achieved through that Wi-Fi is available only to the device. As a result, the device is best positioned to decide the connectivity options. In this context, network discovery and selection procedures can be performed by a connection manager in the device. This kind of connectivity management is out of the scope of 3GPP and it is up to the vendors to provide a solution. More intelligent connection managers can search and prioritize the best available links based on predefined requirements, such as QoS, bandwidth, latency, jitter, power consumption, and operator policies. [61].

2.13 Mobile IP

Mobility support within EPS is based on a mobile IP framework. Mobile IP is an IETF protocol that allows mobile users to move from one network to another while maintaining their IP addresses [55]. The two versions of mobile IPs, i.e., Mobile IPv4 and Mobile IPv6, are described in IETF RFC 3344 [23] and IETF RFC 3775 [29], respectively. The mobile IP allows transparent routing of IP datagrams over the Internet. Each mobile node is identified by its home address

(HoA), irrespective of its current location in the Internet. When away from home, a mobile node is associated with a CoA, which provides information about its current location. The mobile IP specifies how a mobile node registers with its HA and how the HA routes datagrams to the mobile node through a tunnel. Using mobile IP, nodes may change their point-of-attachment to the Internet without changing their IP address, allowing the application and transport-layer protocols to seamlessly maintain connection while moving. The general characteristics of mobile IP can be summarized as follows:

- Transparency of user mobility to the transport and application layer protocols.
- Interoperability with stationary hosts running conventional IPs.
- Scalability across the Internet.
- Security by preventing an attacker from impersonating a mobile host.
- Macro mobility by ensuring long-term connection while away from HA.

Node mobility is realized without propagating host-specific routes throughout the Internet. Using mobile IP, a mobile device will have two addresses, that is, a primary or permanent HoA and a secondary or temporary CoA, which is associated with the network that the mobile node is visiting. There are two types of entities in mobile IP:

- An HA that stores information about mobile nodes whose permanent HoA is in the HA's network.
- A foreign agent (FA) that stores information about mobile nodes visiting its network. The FAs also advertise CoAs.

A node that wishes to communicate with the mobile node uses the permanent HoA of the mobile node as the destination address for outbound packets. Since the HoA logically corresponds to the network associated with the HA, conventional IP routing mechanisms forward these packets to the HA. Instead of forwarding these packets to a destination that is physically in the same network as the HA, the HA redirects these packets toward the FA. The HA looks for the CoA in the mobility binding table [55] and then tunnels the packets to the mobile node's CoA by appending a new IP header to the original IP packet, which preserves the original IP header. The packets are detected at the end of the tunnel by removing the IP header added by the HA and are delivered to the mobile node.

The mobile node directly sends packets to the other communicating node through the FA without involvement of the HA using its permanent HoA as the source address for the IP packets, that is, triangular routing. The FA can utilize reverse tunneling by sending the mobile node's packets to the HA, which forwards them to the communicating node. This mechanism is needed in networks whose gateway routers have ingress filtering enabled, and hence the source IP address of the mobile host needs to belong to the subnet of the foreign network; otherwise, the packets would be discarded by the router.

The mobile IP defines an authenticated registration procedure through which a mobile node informs its HA of its CoA, router discovery (which allows mobile

nodes to discover prospective HA and FAs), and the rules for routing packets to and from mobile nodes, including the specification of one mandatory and several optional tunneling mechanisms.

IP mobility enables routing data packets to the destination when moving to a foreign network. A central point is to maintain the IP address of a UE fixed so that there is no need for upper layers (above the IP layer) to adapt to the network change. Consequently, upper layer connections can continue without notification about the receiver's mobility. In general, IP mobility support can be realized with two different approaches: (1) client-based IP mobility; and (2) network-based IP mobility. This distinction does not depend on the underlying radio access technology; rather it is completely handled in the IP stack.

2.13.1 Client-based IP mobility

In the client-based mobility of IPv6, the UE carries the mobility extensions in its own IP stack. Central to this approach is to split the IP address into the HoA, which is the permanent IP address obtained from the home network, and the CoA, which corresponds to a temporary IP address and is obtained from the visited network. The administration of these addresses is done in an HA. As long as the UE is in the home network, the HA routes the data packets directly to the UE using the HoA. When the UE changes the network, it informs the HA with a proxy binding update (PBU) message about its new IP address, which is stored by the HA in the binding cache. This is essentially a lookup table which is searched for each incoming packet. If there is a CoA entry for this UE in the binding cache, the packet is instead forwarded to its CoA. As an optimization, it is also possible to route IP traffic between a correspondent node (CN) and the UE directly; in this case, a binding cache has to be additionally established in the CN itself.

In an extension, mixed networks with both IPv4 and IPv6 are supported with the dual stack mobile IPv6 protocol (DSMIPv6). This is essential to integrate most existing networks built on IPv4 into the new networks, which will be more and more equipped with IPv6 protocols. For a client-based mobility EPC access with Wi-Fi, DSMIPv6 is mandatory. With this approach, it is possible to offload data traffic from an LTE network to a Wi-Fi connection. However, this works only for a complete offload, that is, it is either possible to communicate over the LTE connection or over the Wi-Fi connection, but not over both. The reason is that in this architecture the Wi-Fi network is considered as a foreign network, to which all data packets are forwarded when there is a corresponding entry in the binding cache of the HA.

2.13.2 Network-based IP mobility

A completely different approach to support user mobility is the network-based IP mobility. Contrary to the client-based IP mobility, the network takes all necessary steps to route the data packets to the destination; thus, there is no need for the

client to do any signaling upon network change. There are two approaches for network-based IP mobility: PMIP version 6 (PMIPv6); and the GTP.

The PMIP (PMIPv6) is specified in IETF RFC 5213 [26]. Routing is based on two additional entities, the LMA, which works in a similar way as the HA in the client-based mobility approach, and the MAG, which implements the necessary mobility functions in the visited network. When the UE changes the network, the MAG is contacted by the new base station and informs the LMA about the change of location. Similar to the client-based mobility, this architecture has to be extended in order to implement the IP flow mobility capabilities, that is, moving selected IP flows from one access technology to another, and consequently, installing the required filters for flow routing. In order to send and receive data packets from and to any of its interfaces, the IETF has decided to adapt the logical interface (LIF). This is a software entity which hides the physical interface to the IP layer. This means that the mobile IP stack binds its sessions to the LIF, thus for the UE there is only one single interface to the IP network. The LIF controls the flow mobility in the UE. It is part of the connection manager in the operating system and has no impact on the IP stack. It represents a virtual interface which hides all flow mobility movements to the higher layers. A second aspect to introduce flow mobility to PMIPv6 is to provide signaling extensions to the MAG. This is necessary because the MAG will only forward traffic from and to a UE if the prefix has been delegated to the UE by this MAG. However, in IP flow mobility, this delegation might have been done by a different MAG before the flow handover. Signaling between the LMA and the target MAG resolves this issue.

3GPP has developed and specified the network-based mobility protocol called GTP. The PMIPv6 and client-based mobility were introduced as alternatives in LTE. The HA or LMA in the 3GPP EPC are located in the P-GW for both trusted and untrusted access. This is in contrast to the location of the MAG, where in the trusted access it is located in the Wi-Fi network and in the untrusted access is in the evolved packet data gateway (ePDG) network entity. Consequently, the connection between the LMA and the MAG can always be regarded as trusted, because the ePDG is under control of the EPC network operator. Up to 3GPP Rel-9 network architecture, the Wi-Fi offload was the only way to offload data from the EPC to Wi-Fi. The IP flow mobility was introduced with 3GPP Rel-10 network architecture for client-based mobility. Network-based mobility is not yet supported as of 3GPP Rel-11 network architecture and is being studied in 3GPP Rel-12.

2.14 Network architecture evolution and small cells

While macro-cellular networks will continue to provide essential wide-area coverage and support for high-mobility users, operators have started to look at other solutions to increase capacity in high-traffic areas. Wi-Fi offload is the most widely and successfully adopted solution in areas where subscriber density and

usage is high, such as urban areas and locations such as airports or stadiums. Operators are also exploring compatible solutions such as femto-cells to improve cellular coverage and capacity in residential areas, as well as small cell underlays to address the increasing demand for high capacity per area, and complement and strengthen their Wi-Fi and macro-cellular deployments. Mobile operators can further improve network utilization by actively managing the traffic beyond the radio access network within the core, using content caching, tiered pricing, and policy enforcement. These solutions do not increase the capacity per se, but make data transmission more efficient, allowing operators to process more content within the same network infrastructure. The integration of Wi-Fi and LTE small cells within the cellular core helps operators optimize network utilization across different radio access technologies (RATs), thus providing further improvement in performance, and creating a seamless multi-RAT experience for their subscribers [56,58,60,61]. While no single product or technology by itself can accommodate the current and future increase in data traffic, many viable solutions, working in concert, can be considered. The challenge for mobile operators is not the decision of which solution to select, rather it is how to best integrate multiple technologies within their networks, how to find the right balance to maximize their cumulative benefits, and how to leverage existing assets to facilitate the evolution of their networks.

Macro-cellular networks today typically provide coverage throughout a mobile operator's coverage area. Their capacity is adequate where the number of subscribers per cell site is relatively small. While high density areas account only for a small part of operators' footprint, they are of crucial importance. This is where subscribers are densely concentrated and where they use data services more extensively. The main challenge for mobile operators is to identify and correct capacity shortfalls in these high-traffic areas.

In order to increase network capacity and to meet the increasing demand for mobile wireless applications, a new approach is taken which relies on lower-power, shorter-range equipment, such as Wi-Fi access points or LTE/LTE-Advanced small cells installed closer to subscribers, in dense deployment areas. By limiting the range, the impact of interference is reduced and the capacity per area is increased. A single sector in a macro cell may have a comparable capacity to a Wi-Fi access point or a small cell, but it spreads it over a larger area, leading inevitably to lower capacity per area. Small cell deployments are one of the most viable strategies of mobile operators worldwide as they provide the capacity increase in the near term, along with the flexibility and compact form factor needed for highly localized deployments in high-traffic environments. Yet, there are several challenges that need to be addressed when planning for a small cell underlay network to enable the operator to maximize the benefits of the new network and to limit the costs. These include securing mounting assets where a high number of nodes are required and typically small cells are not mounted on telecommunication assets like cell towers that operators are accustomed to managing or leasing; backhaul for the small cells network for dense urban environments; as

well as integration with the cellular core network. The small cell underlay network must be fully integrated with the EPC and support self-organizing networks to give the operator the ability to manage user data traffic, support mobility, mitigate interference, and implement policy consistently across different radio access networks.

Most mobile operators no longer see Wi-Fi offload and small cells as competing solutions. They realize both are needed, with each playing an important role within a multi-RAT, multi-layer cellular network, spanning from macro base stations at one end, to residential femto cells at the other. Enhanced small cells can be deployed with macro coverage and stand-alone, both indoor and outdoor, and support both ideal and non-ideal backhauls. Enhanced small cells can also be deployed sparsely or densely. An enhanced small cell may benefit from the presence of overlaid macro cells, and it should also work without macro coverage, for example, in indoor scenarios. Cooperative mechanisms, between macro cell and small cell, as well as among small cells, may be beneficial for the above-mentioned use cases.

References

3GPP Specifications¹²

- [1] 3GPP TS 23.002, Technical Specification Group Services and System Aspects; Network architecture (Release 11), December 2012.
- [2] 3GPP TS 23.003, Numbering, Addressing and Identification (Release 11), December 2012.
- [3] 3GPP TS 23.216, Single Radio Voice Call Continuity (SRVCC); Stage 2 (Release 11), December 2012.
- [4] 3GPP TS 23.261, IP Flow Mobility and Seamless Wireless Local Area Network (WLAN) Offload; Stage 2 (Release 11), September 2012.
- [5] 3GPP TS 23.272, Circuit-switched Fallback in Evolved Packet System; Stage 2 (Release 11), December 2012.
- [6] 3GPP TS 23.401, General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access (Release 11), December 2012.
- [7] 3GPP TS 23.402, Architecture Enhancements for non-3GPP Accesses (Release 11), December 2012.
- [8] 3GPP TR 23.882, 3GPP System Architecture Evolution: Report on Technical Options and Conclusions (Release 8), September 2008.
- [9] 3GPP TS 29.274, Evolved General Packet Radio Service (GPRS) Tunneling Control Protocol for Control Plane (GTPv2-C) Stage 3 (Release 11), December 2012.
- [10] 3GPP TS 24.301, Non-Access-Stratum (NAS) protocol for Evolved Packet System (EPS); Stage 3 (Release 11), December 2012.
- [11] 3GPP TS 33.401, 3GPP System Architecture Evolution: Security Architecture (Release 11), December 2012.

¹²3GPP specifications can be found at the following URL: <http://www.3gpp.org/ftp/Specs/archive/>

- [12] 3GPP TS 36.300, Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2 (Release 11), December 2012.
- [13] 3GPP TS 36.413, Evolved Universal Terrestrial Radio Access Network (E-UTRAN); S1 Application Protocol (S1AP), (Release 11), December 2012.

3GPP2 Specifications¹³

- [14] 3GPP2 C.S0024-A v3.0, cdma2000 High Rate Packet Data Air Interface Specification, September 2006.
- [15] 3GPP2 C.S0005-A v6.0, Upper Layer (Layer 3) Signaling Standard for CDMA2000 Spread Spectrum Systems, February 2002.
- [16] 3GPP2 A.S0009-C v4.0, Interoperability Specification (IOS) for High Rate Packet Data (HRPD) Radio Access Network Interfaces with Session Control in the Packet Control Function, April 2011.
- [17] 3GPP2 A.S0013-C v4.0, Interoperability Specification (IOS) for CDMA2000 Access Network Interfaces—part 3 Features, August 2012.
- [18] 3GPP2 X.S0042-A v1.0, Voice Call Continuity between IMS and Circuit Switched System, August 2008.

IETF Specifications¹⁴

- [19] IETF RFC 2460, Internet Protocol, Version 6 (IPv6) Specification, S. Deering, December 1998.
- [20] IETF RFC 4960, Stream Control Transmission Protocol, R. Stewart, September 2007.
- [21] IETF RFC 3310, Hypertext Transfer Protocol (HTTP) Digest Authentication Using Authentication and Key Agreement (AKA), A. Niemi, September 2002.
- [22] IETF RFC 3261, SIP: Session Initiation Protocol, J. Rosenberg et al., June 2002.
- [23] IETF RFC 3344, IP Mobility Support for IPv4, C. Perkins, August 2002.
- [24] IETF RFC 2138, Remote Authentication Dial in User Service (RADIUS), C. Rigney et al., April 1997.
- [25] IETF RFC 3588, Diameter Base Protocol, P. Calhoun et al., September 2003.
- [26] IETF RFC 5213, Proxy Mobile IPv6, S. Gundavelli et al., August 2008.
- [27] IETF RFC 2131, Dynamic Host Configuration Protocol, R. Droms, March 1997.
- [28] IETF RFC 5555, Mobile IPv6 Support for Dual Stack Hosts and Routers, H. Soliman, June 2009.
- [29] IETF RFC 3775, Mobility Support in IPv6, D. Johnson et al., June 2004.

Books

- [30] S. Sesia, I. Toufik, M. Baker, *LTE—The UMTS Long Term Evolution: From Theory to Practice*, second ed., John Wiley & Sons, 2011.
- [31] C. Johnson, *Long Term Evolution in Bullets*, second ed., Create Space Independent Publishing Platform, 2012.
- [32] E. Dahlman, S. Parkvall, J. Skold, *4G, LTE/LTE-Advanced for Mobile Broadband*, Academic Press, 2011.

¹³3GPP2 specifications can be found at the following URL: http://www.3gpp2.org/Public_html/specs/index.cfm

¹⁴IETF specifications can be found at the following URL: <http://www.ietf.org/rfc.html>

- [33] T. Ali-Yahiya, *Understanding LTE and Its Performance*, Springer, 2011.

Articles

- [34] LTE and Beyond: LTE, 4G, MME, P-GW, S-GW, Interfaces and beyond. <<http://www.lteandbeyond.com/>>.
- [35] GPRS Tunneling Protocol, <http://en.wikipedia.org/wiki/GPRS_Tunnelling_Protocol>.
- [36] (IMS) Multimedia Telephony (MMTel), from Wikipedia, <http://en.wikipedia.org/wiki/Multimedia_telephony>.
- [37] K. Suzuki, et al., *Core network (EPC) for LTE*, NTT DoCoMo Tech. J. 13 (3) (2011).
- [38] C.F. Lanzani, et al., Remote Radio Heads and the Evolution Towards 4G Networks, Altera, 2009<<http://www.altera.com>>.
- [39] LTE System Description, <<http://www.sharetechnote.com/>>.
- [40] LTE Identifiers, NMC Consulting Group. <<http://www.nmcgroups.com>>, 2011.
- [41] LTE Identifiers, EventHelix.com Inc. <<http://www.eventhelix.com/>>, 2012.
- [42] Small Cell Forum, <<http://www.smallcellforum.org/resources-technical-papers>>.
- [43] G. Gómez, et al., *Towards aQoE-Driven resource control in LTE and LTE-A networks*, J. Comput. Netw. Commun., (2013).
- [44] Common Public Radio Interface (CPRI), <<http://www.altera.com/technology/high-speed/protocols/cpri/pro-cpri.html>>.
- [45] GSMA, 2010, IR.92 IMS Profile for Voice and SMS V3.0. Available at: <<http://www.gsma.com/newsroom/wp-content/uploads/2012/06/IR9230.pdf>>.
- [46] GSMA, 2011, IR.94 IMS Profile for Conversational Video Service V1.0. Available at: <<http://www.gsma.com/newsroom/wp-content/uploads/2012/03/ir9410.pdf>>.
- [47] GSMA, 2011, IR.64 IMS Service Centralization and Continuity Guidelines V2.0. Available at: <<http://www.gsma.com/newsroom/wp-content/uploads/2012/03/ir6420.pdf>>.
- [48] LteWorld, Evolution of Single Radio Voice Call Continuity. <<http://lteworld.org/blog/evolution-single-radio-voice-call-continuity-srvcc>>, December 2012.
- [49] LteWorld, Understanding CS Fallback in LTE, September 2009.
- [50] IP Multimedia Subsystem, from Wikipedia. <http://en.wikipedia.org/wiki/IP_Multimedia_Subsystem>.
- [51] Network Switching Subsystem, from Wikipedia. <http://en.wikipedia.org/wiki/Network_Switching_Subsystem>.
- [52] Internet Protocol Security (IPsec), <<http://en.wikipedia.org/wiki/Ipsec>>.
- [53] C. Gessner, O. Gerlach, *Voice and SMS in LTE*, Rohde & Schwarz, 2011.
- [54] Motorola White Paper, LTE Inter-technology Mobility: Enabling Mobility between LTE and other Access Technologies, Motorola Inc., November 2008.
- [55] Mobile IP, from Wikipedia, <http://en.wikipedia.org/wiki/Mobile_ip>.
- [56] T. Nakamura, et al., *Trends in small cell enhancements in LTE advanced*, IEEE Commun. Mag. (2013).
- [57] T. Ahmed, et al., Multi access data network connectivity and IP flow mobility in evolved packet system (EPS), Proceedings of WCNC 2010, April 2010.
- [58] A. Schumacher, J. Schlien, *WLAN Traffic Offload in LTE*, Rohde & Schwarz, 2012.
- [59] A. De La Oliva, et al., *IP flow mobility: smart traffic offload for future wireless networks*, IEEE Commun. Mag., 2011.

- [60] Dealing with Density: The Move to Small-Cell Architectures, Ruckus Wireless White Paper, 2011.
- [61] A 3G/LTE Wi-Fi Offload Framework: Connectivity Engine (CnE) to Manage Inter-System Radio Connections and Applications, Qualcomm White Paper, June 2011.
- [62] R. Gupta, N. Rastogi, LTE-Advanced—LIPA and SIPTO, Aricent Group, 2012.
- [63] Circuit-switched fallback, the first phase of voice evolution for mobile LTE devices, Qualcomm White Paper, 2012.
- [64] Anritsu Whitepaper, Voice over LTE (VoLTE), Anritsu Corporation, September 2012.
- [65] G. Bertran, The IP Multimedia Subsystem in Next Generation Networks, May 2007.
- [66] N. Kottapalli, Diameter and LTE Evolved Packet System, Radisys White Paper, September 2011.
- [67] D. Soldani, Bridging QoS and QoS for Mobile Broadband Networks, Huawei Technologies, 2010.
- [68] M. Grea, et al., Quality of Experience Evaluation for Multimedia, Overview of Telecommunications, Telecommunication News, No. 4/2008, April 2008.
- [69] S. Tripathi, et al., LTE E-UTRAN and its Access Side Protocols, Radisys White Paper, September 2011.
- [70] Introduction to Evolved Packet Core, Alcatel-Lucent White Paper, 2009.
- [71] The LTE Network Architecture: A comprehensive tutorial, Alcatel-Lucent White Paper, 2009.
- [72] VoLTE with SRVCC: the second phase of voice evolution for mobile LTE devices, Qualcomm White Paper, October 2012.
- [73] S. Kundalkar, N. Madhur Raj, LIPA: Local IP Access via Home Node B, Radisys White Paper, September 2011.
- [74] Dynamic Host Configuration Protocol, from Wikipedia, <http://en.wikipedia.org/wiki/Dynamic_Host_Configuration_Protocol>.
- [75] Open Base Station Architecture Initiative, from Wikipedia, <<http://en.wikipedia.org/wiki/OBSAI>>.
- [76] Internet Key Exchange, from Wikipedia, <http://en.wikipedia.org/wiki/Internet_Key_Exchange>.

This page intentionally left blank

Copyright © 2013. Elsevier Science & Technology. All rights reserved.