Ronald Kirui

Accessing Cloud Computing Resources over 4G LTE

Helsinki Metropolia University of Applied Sciences

Bachelor of Engineering

Degree Programme in Information Technology

Thesis

22 May 2014

Helsinki
Metropolia
University of Applied Sciences

Author Title Number of Pages Date	Ronald Kirui Accessing Cloud Computing Resources over 4G LTE 66 pages 22 May 2014
Degree	Bachelor of Engineering
Degree Programme	Information Technology
Specialisation option	Communication and Data Networks / Personal Communication
Supervisor	Dr. Tero Nurminen, Principal Lecturer

Over the years, there has been a significant evolution in telecommunication technology, in particular the Third Generation Partnership Project (3GPP) family of telecommunication systems. The demand for broadband wireless Internet has been the main catalyst, hence we now have the fourth generation (4G). Similarly, there has been a paradigm shift in the computing world towards cloud computing, both by individual users as well as private and public institutions. The purpose of this thesis therefore, was to explore the 4G Long Term Evolution (4G LTE) as well as the cloud computing technology and highlight the convergence of these techonologies and their applications in various fields.

Courtesy of the 3GPP, the technical specifications for Long Term Evolution (LTE) and LTE-Advanced (LTE-Advanced) were outlined. In this thesis therefore, these specifications along with other related publications, and in addition, publications about the cloud computing technology were explored for insight. The convergence of 4G and cloud computing, some (4G and cloud computing) solutions for telecommunication operators and a few applications of 4G LTE and cloud computing were brought to the fore.

This thesis as a result, presented the case for the adoption of 4G LTE and cloud computing to complement each other. It has also highlighted available technological solutions for telecommunication operators for cloud computing and also outlined some examples where the technologies can be applied in tandem. In future, experiments can be carried out to test the various applications of these technologies.

Keywords	LTE, LTE-Advanced, EPS, EPC, 4G, Cloud Computing

Helsinki Metropolia University of Applied Sciences

Contents

1 Introduction 1

2 Wireless Communication Evolution 3

2.1 Early mobile communication 3

2.2 Global System for Mobile Communication 3

2.2.1 General Packet Radio Service 5

2.2.2 Enhanced Data Rates for GSM Evolution 6

2.3 Universal Mobile Telephone Systems 7

2.3.1 UMTS Network Architecture 8

2.3.2 UMTS Air Interface and Radio Network 10

2.3.3 High Speed Packet Access 11

2.4 Long Term Evolution, Long Term Evolution-Advanced 12

3 Long Term Evolution 13

3.1 Long Term Evolution Background 13

3.2 LTE Network Architecture 14

3.3 LTE Radio Interface Architecture 15

3.3.1 LTE Transmission Schemes 16

3.3.2 Physical Layer Parameters 18

3.3.3 Multiple Antenna Techniques 19

3.4 Protocol Architecture 21

Figure 12. Control Plane protocol stack [15,24] 22

3.4.1 Radio Resource Control Protocol 22

3.4.2 Packet Data Convergence Protocol 23

3.4.3 Radio Link Control Layer 24

3.4.4 MAC Layer 25

3.4.5 Physical Layer 26

3.5 LTE Channel Structure 27

3.5.1 Logical Channels 28

3.5.2 Transport Channels 29

3.5.3 Physical Channels 30

3.6 LTE-Advanced 32

4 Evolved Packet Core 34

4.1 EPC Functions and Main Elements 34
4.2 EPS Deployment Process 36
4.2.1 Initial Deployment Phase 36
4.2.2 Integration Phase 37
4.2.3 Optimization Phase 38
4.3 LTE User Services 38
4.4 LTE Security 39
4.4.1 Network Access Security 41
4.4.2 Network Domain Security 43
4.4.3 User Domain Security and Other Security Features 44
4.5 Basic Procedures 44

5 Cloud computing 46

5.1 Definition, History and Cloud Computing Features 46
5.2 Cloud Computing Architecture 47
5.2.1 Cloud Service Models 47
5.2.2 Cloud Deployment Models 48
5.2.3 Cloud Computing Implementation Hierarchy 50
5.3 Cloud Components 52
5.4 Benefits and Disadvantages of Cloud Computing 53
5.4.1 Benefits of Cloud Computing 53
5.4.2 Challenges to Cloud Computing Adoption 53

6 Cloud Computing and 4G LTE 55

6.1 Convergence of Cloud Computing and 4G LTE 55
6.2 Cloud Technological Solutions for Telecommunications Operators 56
6.3 Applications of Cloud Computing and 4G 57

7 Conclusion 59

References 60

1 Introduction

Human communication has transformed over the years from basic verbal communication, use of signals such as smoke signals and to the present where it takes many different forms through varied media. With regard to wireless communication, a remarkable milestone was towards the end of the 19th century when Gugliemo Marconi [1, 1] demonstrated the earliest form of wireless communication. His discovery has been exploited in various ways such as the use of radio and television broadcast, radar systems, and mobile communication among others.

Since the inception of the first-generation mobile communication in the 1980s, there has been a constant advancement and evolution to the point where we now are at the fourth-generation (4G). The requirements for 4G as outlined by the International Telecommunication Union (ITU) include for example, high data rates, low latency and user-friendly applications, services and equipment. The Third Generation Partnership Project's (3GPP) Long Term Evolution-Advanced (LTE-Advanced) is one of the technologies which has been approved as 4G technology [2]. At the present, telecommunication companies across the globe are offering and marketing LTE services as 4G which technically are not. Indeed there may be some early adopters of LTE-Advanced but it will take a few more years before it is widely deployed and also for devices which are truly LTE-Advanced capable are available for the masses.

On the other hand, the cloud computing technology, which enables access of IT resources over the Internet, is quickly gaining traction and popularity as an alternative to traditional computing. This is furthered by the fact that mobile devices – which have become part and parcel of our lives – continue to accommodate more capabilities and functionalities. Mobile devices will however face a number of limitations such as storage capacity, processing capability and battery life and this is where the use of the cloud becomes beneficial. By offloading the demanding computational functions to the cloud and utilizing the emerging high speed wireless access technologies to access the cloud, mobile devices can be used as an interface to the cloud, hence mitigating some of its limitations.

The goal of this thesis was to learn and gain a better understanding of how the 4G LTE technology and also the cloud computing technology functions. The various technological aspects of 4G LTE as well as those of cloud computing are discussed, and thereafter, examples of their utilization.

2 Wireless Communication Evolution

2.1 Early mobile communication

Mobile wireless communication came to military use in the early 20th century and later in the mid-1940s it found its use in car-based telephones. It was not until the 1980s when the first-generation cellular networks such as Advanced Mobile Phone System (AMPS), Nordic Mobile Telephony (NMT), and Total Access Communication System (TACS) - which were analog-based systems - were introduced. [3, 8-17] The cellular network systems for these systems were laid out such that a large area was subdivided into small cells, thus allowing for frequency re-use and hence better utilization of allocated frequency bands. An additional benefit for smaller cells was that it meant smaller and cheaper devices could be used to transmit and receive information due to their demand for less power. Although the first-generations cellular system offered handover and roaming, one of their disadvantages was that they were not interoperable between the different systems and therefore across countries.

The second generation, 2G systems were launched in the 1990s and marked a major turning point being a switch from analog systems to digital communication systems. Interim Systems-95 (IS-95) and Interim Systems-136 (IS-136), predominantly in the USA and Global System for Mobile communication (GSM), dominant in Europe are some of the 2G mobile standards. The shift to digital systems – Code Division Multiple Access (CDMA) and Time Division Multiple Access (TDMA) resulted in the possibility of integrated services (fax, data and voice services), better channel utilization due to better compression techniques, better quality by error detection and correction, and secure communication through encryption [3, 8-20]. In addition, higher spectrum efficiency and advanced roaming was possible. 2G networks were primarily designed for voice services and low data rates over a Circuit Switched (CS) network.

2.2 Global System for Mobile Communication

In 1982 The European Conference of Postal and Telecommunications Administrations (CEPT) tasked the Group Special Mobile (GSM) to develop a standard for mobile telephony across Europe in the 900 MHz band. This led to the birth of the Global System

for Mobile Communication (GSM) a few years later, which is so far the most successful mobile communication system. Its success is attributed to many factors such as progressive and backward compatibility evolution, global user roaming, multivendor environment – hence lower costs for users and vendors among other reasons. [4,3]

The initial GSM network consisted of the Radio Access Network (RAN) which comprised of the Base Station Subsystem (BSS) i.e. the Base Transceiver Station (BTS) and Base Station Controller (BSC), and the core network comprised of the Mobile Switching Centre (MSC), Visitor Location Register (VLR), Home Location Register (HLR), Authentication Centre (AuC) and Equipment Identity Register (EIR). In addition, it had Voice Mail Services and Short Message Service Centre (SMSC) as Value Addition Services (VAS).

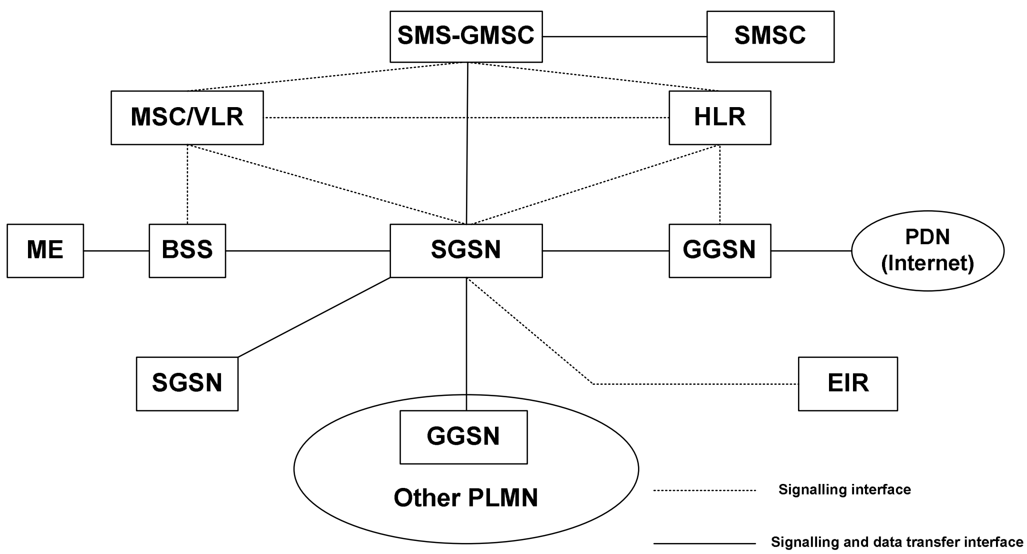

Figure 1. GSM and GPRS architecture. Adapted from Sauter M (2010) [5,20]

In figure 1, the Mobile Equipment (ME) is the device that a user operates to access the GSM network. Its main components are the Subscriber Identification Module (SIM) and the main hardware. The SIM contains the International Mobile Subscriber Identity (IMSI) in addition to other information used by the network to authenticate and authorise the user to the network. The hardware on the other hand is composed of the case, display, battery and the electronics that generates, receives, and processes data received and to be transmitted. It also contains the International Mobile Equipment Identity (IMEI) hardcoded to the device. [4,52]

The section of the network which is responsible for communication with the mobile station (device) is the BSS which consists of the BTS and BSC. The BTS is composed of a radio transceiver and antennas which communicate with the mobile devices. The element charged with radio resource management, channel allocation and controls such as handovers is the BSC. The Network Sub System (NSS), which is also known as the core network, provides the main control and interfacing for the whole mobile system and is comprised of the MSC, HLR, VLR, EIR, AuC, Gateway Mobile Switching Centre (GMSC) and Short Message Service Centre (SMSC). Briefly, the descriptions of the different elements are as follows: [4,11-19]

- MSC: Is the central component of the NSS whose function is to perform the switching function of the network and provide a connection to other networks
- HLR: Stores subscriber information belonging to the MSCs coverage area, current subscriber location and the services they can access
- VLR: Contains information from the subscribers HLR to provide subscribed services to the visiting user that is in addition to all active subscribers in its area.
- EIR: Stores information about the mobile equipment such as the IMEI which is then used to deny access to the network by unauthorised terminals
- AuC: Provides authentication, authorization and encryption parameters, hence ensuring secure communication and proper subscriber identification
- GMSC: Interconnects the cellular network and the Public Switched Telephone Network (PSTN), hence making calls to and from the fixed network possible
- SMSC: Handles SMS to and from the Mobile Station (MS).

2.2.1 General Packet Radio Service

As a result of demand for mobile data and other services, GSM and the other 2G systems evolved to meet these demands; General Packet Radio Service (GPRS) and later Enhanced Data Rates for GSM Evolution (EDGE) for GSM and 1xRTT for IS-95. For GSM in particular, a new Packet Switched (PS) network was developed and overlaid on the initial GSM core network for GPRS. The new network elements added towards the new network are Serving GPRS Support Node (SGSN), Gateway GPRS Support Node and Packet Control Unit as illustrated in figure 1. Short descriptions of these elements are as follows:

- SGSN: Delivers data packets to and from MS. Also includes packet routing and transfer, attach/detach and authentication functions for MS and logical link management
- GGSN: Acts as the interface and router to external packet data networks (PDNs)
- PCU: manages and controls GPRS traffic. [6,235]

The development of GPRS offered benefits both to the user and the provider. To the user, it made it possible to access the Internet with higher speeds (15-45K bps) compared to GSM (up to 9.6 Kbps), to be always connected without having to worry about costs because charging is per bit transferred and the possibility to access new applications. The provider on the other hand benefits from a new avenue for revenue from the higher data rates services and other service differentiation opportunities, and lessons and a path to third-generation (3G). [7,32-35]

2.2.2 Enhanced Data Rates for GSM Evolution

Enhanced Data Rates for GSM Evolution (EDGE) was developed based on the GPRS system to increase throughput speeds. Basically it was a new modulation and coding scheme which used 8-Phase Shift Keying (8PSK). With 8PSK, the EDGE transmission speed can be up to three times faster compared to GSM and GPRS. The Universal Mobile Telephone Systems (UMTS), a successor to GSM provides superior data rates, but EDGE is still operated in parallel to it for extra capacity and speed and for faster transmission speeds in buildings and rural areas where 3G coverage may be limited. The EDGE deployment requires an upgrade to software and limited hardware upgrades to the GSM/GPRS elements. Subscribers also require EDGE capable equipment. [5,70; 6,49]

Table 1. EDGE data rates. Adapted from Andersson C (2001)[8,324]

		Slot Combination data rate Kb/s		
Channel Coding Scheme	**Modulation**	**1 Slot**	**4 Slots**	**8 Slots**
MCS 1	GMSK	8.8	5.2	70.4
MCS 4	GMSK	17.6	70.4	140.4
MCS 5	8PSK	22.4	89.6	179.2
MCS 9	8PSK	59.2	236.8	473.6

Table 1 shows the achievable EDGE data rates for different channel coding schemes i.e. the modulation and slot combinations. [8,324]

2.3 Universal Mobile Telephone Systems

Towards the end of the 1990s, advances in telecommunication and related technologies such as electronics memory and processing capacity led to the design of a new telecommunication systems; the UMTS with capabilities that far exceeds those of GSM. The UMTS was not designed from scratch but rather it inherited some features from the GSM. The UMTS RAN was however a completely new development which was later enhanced to offer broadband Internet with HSPA. Its design combined CS and PS and it offered a multitude of possibilities and services. [5,115]

The Third Generation Partnership Project (3GGP) is the entity responsible for evolving GSM, UMTS and LTE in form of releases. The following is a summary of the 3GPP releases: [5, 115]

Release 99 was the first 3GPP release and contains all the specifications for UMTS that combined GSM and UMTS. UMTS features a redesigned RAN where Wideband CDMA (WCDMA) is introduced in place of CDMA, which means the user is separated by a unique code rather that time slots. Consequently there is increased data bandwidth up to 384 kbps for downlink (DL) and 128 kbps for uplink (UL). There was no major redesign for the CN but mainly software updates in the core elements (i.e. MSC, HLR, AuC etc.). UMTS supports both voice calls and data packet services, but its RAN was mainly designed for high speed packet data service. The combined network elements for GSM and UMTS ensured simplified roaming between the two network systems, provided that the mobile device in use was a dual mode.

Release 4 introduced the Bearer-Independent Core Network concept (BICN) where core network traffic was transported inside IP packets rather than inside circuit switched 64 kbps timeslots.

Release 5 laid the foundation for IP Multimedia Subsystem (IMS) to handle calls and other services via the PS part of the network. In addition, it introduced the High Speed Downlink Packet Access (HSDPA) a new transmission scheme where under ideal conditions, data speeds up to 14.4 Mbps can be attained.

Release 6 Introduced High Speed Uplink Packet Access which enabled a higher uplink speed and also increased the number of maximum simultaneous users.

Release 7 includes specifications for reduced power consumption and faster return to a fully active state based on a feature called Continuous Packet Connectivity (CPC). It also makes further specifications for increased downlink data transfers by introducing Multiple Input Multiple Output (MIMO) techniques and 64 Quadrature Amplitude Modulation (64-QAM) scheme and as result attaining speeds of 28 Mbps and 21 Mbps respectively. In the uplink direction, 16 Quadrature Amplitude Modulation (16QAM) is specified increasing data rates to 11.5 Mbps under ideal conditions.

Release 8 Introduced Long Term Evolution (LTE) – which will be described in more detail in chapters 3 and 4. With regard to UMTS, this release provides specifications for downlink carrier aggregation where adjacent carriers are combined to get a 10 MHz bandwidth. It also gives specifications for simultaneous use of 64 QAM and MIMO for a single carrier operation and consequently, under ideal conditions the possibility of 42 Mbps downlink throughput.

Release 9 Outlines the specifications for aggregation of two adjacent carriers in the uplink direction doubling data rates to up to 20 Mbps. There is also a specification for combination of a dual-carrier operation with the MIMO operation in the downlink direction, hence increased data rates of up to 82 Mbps. Also in the downlink direction carrier aggregation is not limited to only adjacent carriers and therefore carriers in different bands can be combined. Security enhancement measures such as the introduction of the A5/4 security algorithm and the doubling of the length of the Ciphering Key (CK) are also specified in this release.

Release 10 -12 mainly give specifications for LTE and LTE-Advanced, which are addressed in the subsequent chapters.

2.3.1 UMTS Network Architecture

The Universal Mobile Telephone Systems Network Architecture (UMTS) network is made of the radio network, Universal Terrestrial Radio Access Network (UTRAN) and the core network. The UTRAN is composed of NodeB (NB) and Radio Network Controller (RNC). Elements in the core network include the MSC, SGSN, and GGSN among others. Figure 2 illustrates the common GSM/UMTS network. [9,15]

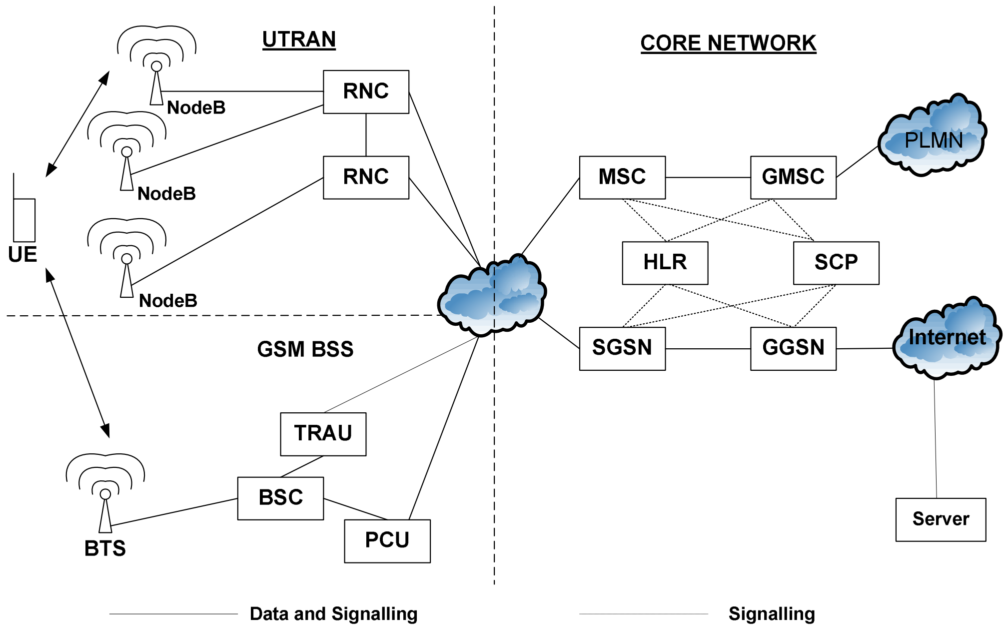

Figure 2.GSM/UMTS Network. Reprinted from Sauter M (2009) [9,15]

The UMTS base station is referred to as NodeB, and it communicates with mobile devices over the air interface. The NB coverage area is divided into sectors – also known as cells – to increase data rates and the number of simultaneous calls. This is achieved by having directional antennas and transceivers for each sector. [9,15] In the UMTS initial years, NodeBs were connected to the RNC by 2 Mbps E-1 links, but due to increase in traffic capacity demands, at present high speed IP-based links (e.g. Digital Subscriber Line (DSL), fibre links or microwave links) are mostly in use [5,150]. The RNC is responsible for establishing radio connections, radio resource management and some mobility management [9,17]. It is also responsible for some security functions such as data encryption and decryption. The MSC is responsible for managing the circuit-switched core network. It handles voice, video and SMS. The GMCS acts as the interface to external networks and hence facilitates calls to other networks.

Some of the functions of the SGSN are subscriber mobility and session management. SGSN keeps track of subscriber location for proper routing of user packets and in addition manages the packet-switched sessions such as access and Quality of Service (QoS). The GGSN on the other hand connects the UMTS to the Internet and is also responsible for assigning Internet Protocol (IP) addresses to the users and for forwarding incoming data to subscribers.

In addition to the above mentioned elements, the UMTS is composed of the following data base elements which are shared with the GSM.

- HLR: Contains a record for each subscriber; the record contains subscription information and the last known subscriber location
- VLR: Holds a copy of the subscriber record in the HLR currently served by the MSC
- AuC: Contains a copy of the secret key contained in the Universal Subscriber Information Module (USIM)
- SMSC: is used to store and forward short messages. [5,14]

2.3.2 UMTS Air Interface and Radio Network

In order to overcome the limitations of a narrow channel bandwidth (200 kHz) which is used in the GSM), the UMTS uses the WCDMA with a 5 MHz bandwidth. This is combined with the use of spreading codes to communicate between the NodeB and the user equipment; it is not only possible to achieve higher transmission speeds but also to make multiple simultaneous transmissions [9,25]. For the purposes of network distinction, neighbouring cell detection, for mobility reasons, power management and network QoS, the radio channel is split up into sub-channels whose access is controlled by the network. These physical channels represented by spreading codes include:

- The Primary Common Control Physical Channel (P-CCPCH)
- The Secondary Common Control Physical Channel (S-CCPCH)
- Physical Random Access Channel (PRACH) and
- Dedicated Physical Data and Control Channels (DPDCH, DPCCH). [9,25-26]

The UE close to the BS requires a small amount of power for data transfer, but the UE in buildings and far from the NodeB require more transmission power. For this reason, the network constantly monitors the air interface connection to ensure efficient power and mobility management by establishing dedicated control channels alongside dedicated traffic channels. With constant monitoring of the air interface, and the mobile device continuously reporting to the network the data reception quality, the network is able to instruct the mobile device on the transmission power it should use. In addition, the network is able to make a decision whether or not a mobile device should be transferred to a neighbouring cell. The transfer is known as a handover and can either be a soft handover or a hard handover.

During a soft handover (i.e. make-before-break), the mobile device has active links to multiple cells at the same time, and hence a transfer from one cell to another is gradual. In a hard handover, on the other hand, once the network detects a more suitable cell, it prepares it prepares the new cell and then instructs the mobile device to change to the new cell. The mobile device thus breaks the old connection and then establishes a new connection based on handover parameters (e.g. frequency, spreading codes etc) sent by the network.

2.3.3 High Speed Packet Access

In Release 5, the 3GPP introduced the specifications for higher bit rates and lower delays in the DL i.e. High Speed Downling Packet Access (HSDPA). This was later followed by specifications for higher uplink speeds and increased maximum simultaneous users i.e. High Speed Uplink Packet Access (HSUPA) in Release 6. These two specifications combined are referred to as High Speed Packet Access (HSPA), and were further refined in Release 7 and also in subsequent releases with specifications for among others:

- faster HSPA and continuous packet connectivity
- reduced power consumption
- short wake time from sleep to active state
- increased data transfer by use of MIMO or higher modulation schemes such as 64 QAM, QPSK, and 16 QAM etc [9,31]

These subsequent releases aimed at improving the air interface and the network architecture are referred to as HSPA+.

In order to achieve the goals set by specifications for HSPA and HSPA+, the following mechanisms were introduced and adopted:

- higher order modulation
- error detection and correction
- MIMO
- continuous packet connectivity and
- radio network enhancements [9,31-44]

The final and most important results of the HSPA adoption is the achieved high data rates. Theoretically, up to 168 Mbps DL and 22 Mbps DL can be achieved by a combi-

nation of higher order modulation, 64 QAM, and 4X4 MIMO. In practice, however, this may not be achievable due to limitations to device capabilities, which determine the MIMO configuration to be used, the modulation technique used and the air link quality.

2.4 Long Term Evolution, Long Term Evolution-Advanced

The continuous increase in demand for higher data rates and better QoS encouraged the 3GPP to develop the LTE. When the specifications for International Mobile Telecommunication Advanced (IMT-Advanced) were set out in March 2008 by International Telecommunication Union Radiocommunication Sector (ITU-R), the 3GPP initiated the LTE-Advanced work item to study and develop a technology solution and components that would meet the ITU-Advanced specifications.

The key recommendations for IMT-Advanced were:

- a high degree of commonality of functionality worldwide while retaining the flexibility to support a wide range of services and applications in a cost efficient manner
- compatibility of services within IMT and with fixed networks
- capability of interworking with other radio access systems
- high-quality mobile services
- user equipment suitable for worldwide use
- user-friendly applications, services and equipment
- worldwide roaming capability
- enhanced peak data rates to support advanced services and applications (100 Mbps for high and 1 Gbps for low mobility were established as targets for research)[10]

LTE-Advanced was submitted to the ITU as a candidate for 4G and was ratified by the ITU-R in the autumn of 2010 as an IMT-Advanced technology [11,11-13].The LTE and LTE-Advanced will be discussed in more detail in chapters 3 and 4.

3 Long Term Evolution

3.1 Long Term Evolution Background

The evolution towards LTE started as early as 2004 when the 3GPP initiated work on the LTE radio interface, and by mid-2005 it released a technical report with the design objectives. Some of the design targets were: high data rates, low user plane latency, requirements for normal capacity and also for peak data rates, flexibility in spectrum usage, and reduced time for state changes. [11,11-12] The motivation for the LTE included: the need to ensure competitiveness of the 3G system for the future, user demand for higher data rates and quality of service, and low system complexity among others.

After a few years of research on technical solutions for the LTE, the 3GPP released the first detailed specification in 2008 in Release 8. Concurrently with the development of the LTE, the 3GPP had another project termed System Architecture Evolution (SAE), tasked with developing the Evolved Packet Core (EPC) whose specifications were released in Release 8 alongside those of LTE. There have been subsequent releases which introduced more functionality and capabilities to LTE and SAE, as briefly described in section 2.3. [11,17]

The 3GPP in order to be compliant with IMT-Advanced requirements made specifications for LTE-Advanced in Release 10. Their main focus among others were:

- increased peak data rates (1 Gbps in the downlink and 500 Mbps in the uplink)
- reduced latency in both the C-Plane and the U-Plane of less than 50 ms
- higher spectral efficiency (30 bps/Hz in the downlink and 15 bps/Hz in the uplink)
- increased performance at cell edges
- increased number of simultaneous active subscribers. [5,272;]

In 2010 LTE-Advanced was released and it introduced enhanced features such as: carrier aggregation to enhance spectrum flexibility, enhanced multi-antenna techniques, support for relay nodes and intercell interference coordination.[11,103] LTE-Advanced

was submitted to ITU-Advanced as a candidate for IMT-Advanced and was officially designated as IMT-Advanced in autumn 2010, hence becoming one of the 4G mobile technologies. [2]

3.2 LTE Network Architecture

The User Equipment (UE), the Evolved UTRAN (E-UTRAN) and the EPC are the three main elements of the LTE network architecture. Figure 3 illustrates the various parts and the interfaces between them.

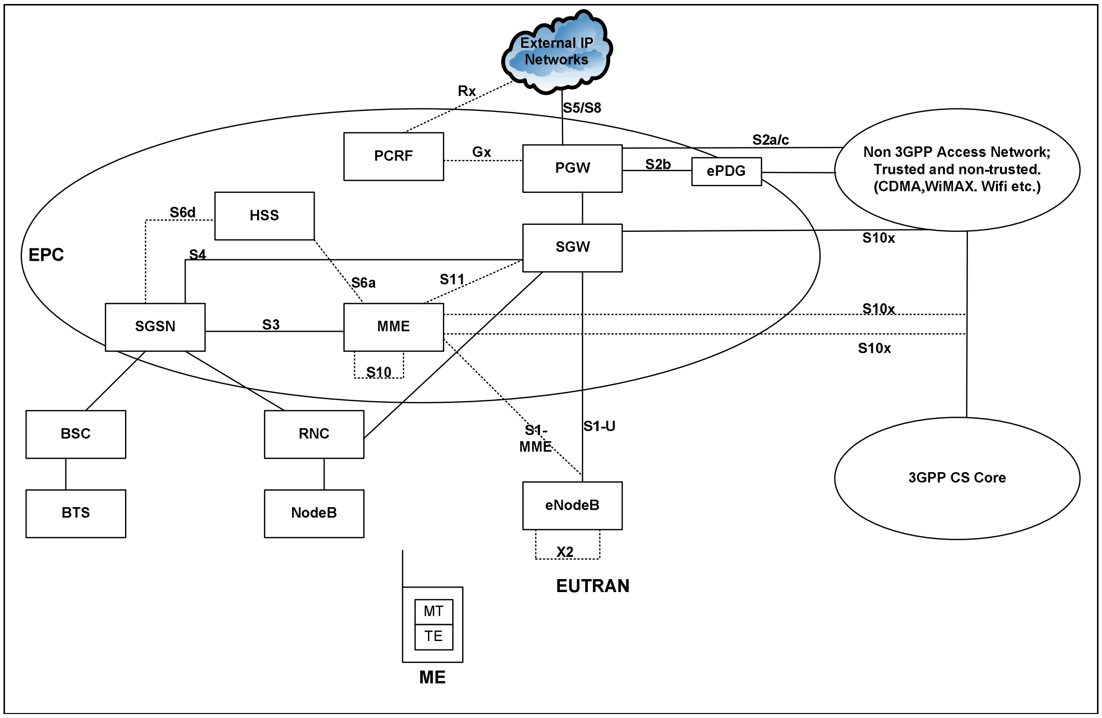

Figure 3. LTE multi-access network architecture

Adapted from Olson M et al (2013) [12,20]

The UE which is also referred to as the mobile device is comprised of: the Mobile Terminal (MT) which handles all communication functions, the Terminal Equipment (TE) which terminates data streams and the Universal Integrated Circuit Card (UICC). The UICC runs the Universal Subscriber Module (USIM) which stores details about the user such as the phone number and security keys. The UE also supports coding and modulation, antenna diversity and MIMO.

In LTE, the functions of UMTS RNC have been moved to the Evolved NodeB (eNodeB) and some other functions to the core network resulting in a much simplified flat architecture. The eNodeB which is a logical node is responsible for functions such as radio resource (air interface) management, mobility functions (e.g. performing handovers and ensuring quality of service). It is connected to the EPC via the S1 interface which is based on IP protocol. The S1 interface is split into two logical parts; the S1 user-plane part (S1-U) responsible for user data and the S1 control-plane (S1-C) responsible for signalling data. S1-U and S1-C connects to Serving Gateway (SGW) and Mobility Management Entity (MME) respectively. [11,111; 9, 46]

The X2 connects the eNodeBs to each other and is responsible for handling handovers for active mobiles, packet forwarding during handovers and may also be used for multi-cell Radio Resource Management (RRM) such as Inter-Cell Interface Co-ordination (ICIC). The EPC which consists of a number of nodes such as SGW, MME, Packet Data Network Gateway (PDN-Gateway or PGW) and Home Subscriber Server (HSS) which together with the eNodeB makes up the SAE are discussed further in section 4.2. [11,111; 9, 46]

3.3 LTE Radio Interface Architecture

One of the requirements for LTE is flexible use of frequency bands. The 3GPP therefore in its technical specifications designed LTE to operate in the frequency band ranging from 700 MHz to 3800 MHz and channel bandwidths from 1.4 MHz to 20 MHz [13,17-18]. This flexibility enables operators in different parts of the world to deploy LTE even with the varying spectrum availability as well as regulation in different jurisdictions.

Furthermore, LTE supports both the Frequency Division Duplex (FDD) and Time Division Duplex (TDD). During the FDD operation, the downlink and uplink transmission is separated into paired frequencies which may be operated either in half-duplex or in full-duplex modes. The Half-duplex FDD, in which transmission and reception are separated in both frequency and time, is particularly ideal at the terminal because of reduced terminal complexity. In the TDD operation, the downlink and uplink uses the same frequency but the transmission takes place in different non-overlapping time slots. [11,101]

3.3.1 LTE Transmission Schemes

Unlike UMTS which used CDMA as the transmission scheme, LTE uses Orthogonal Frequency Division Multiplexing (OFDM). This along with other LTE features such as multi-antenna technology, spectrum flexibility and link adaptation techniques enables LTE to have much improved performance in terms of peak data rates, delay and spectrum efficiency.[14.69] With OFDM, data is transmitted over several closely spaced orthogonal subcarriers. This combined with cyclic prefix (CP) offers advantages such as resistance to effects of fading and multipath and makes it ideal for multi-antenna transmission. The modulation to be used depends on the signal condition and can be QPSK, 16QAM or 64QAM, QPSK being selected under the low Signal to Noise ratio (S/N).

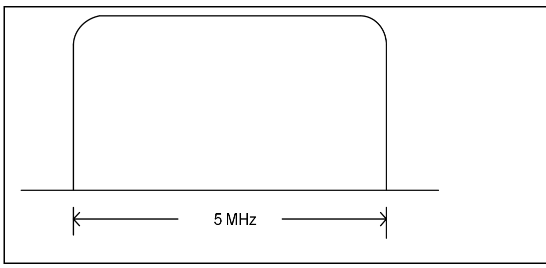

Figure 4.a Single carrier Transmission

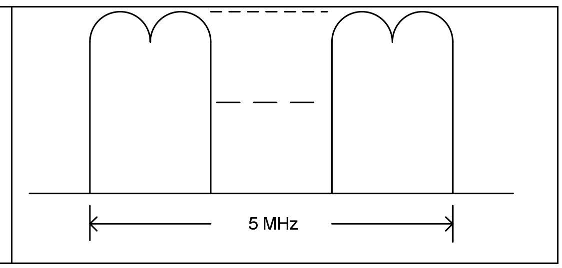

Figure 4.b Several carriers with spacing of X MHz

Figures 4.a and 4.b illustrate the difference between single carrier transmission and multi-carrier transmission using a 5 MHz channel bandwidth.

In the downlink direction, LTE uses Orthogonal Frequency Division Multiple Access (OFDMA) for data transmission. The OFDMA dynamically assigns a set of subcarriers to individual users where each frequency component is modulated with unique information. 3GPP specifies 15 kHz spacing for subcarriers and consequently, the available subcarriers depend on the transmission bandwidth of the system. The OFDMA takes a group of input bits to assemble the subcarriers which are then processed by the Inverse First Furrier Transform (IFFT) to get a time signal as illustrated in figure 5.[9,54;15,42]

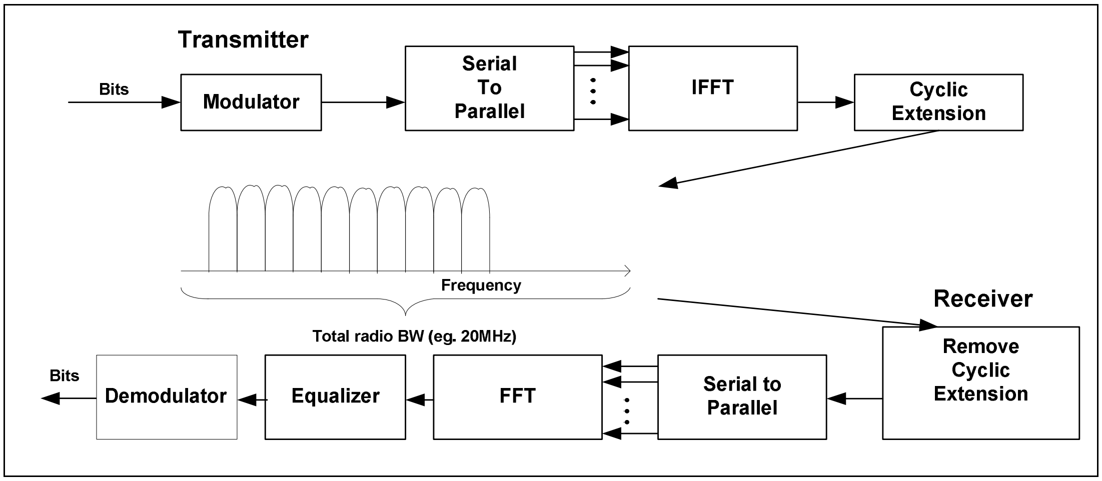

Figure 5. OFDMA transmitter and receiver. Adapted from Holma H (2009) [14,71]

Given the high Peak Average Power Ratio (PAPR) associated with OFDM as well as practical design considerations for UE, the Single-Carrier-FDMA (SC-FDMA) became the preferred option for LTE uplink transmission. SC-FDMA in addition to low PAPR allows for low-complexity equalization and offers flexibility in bandwidth assignments.

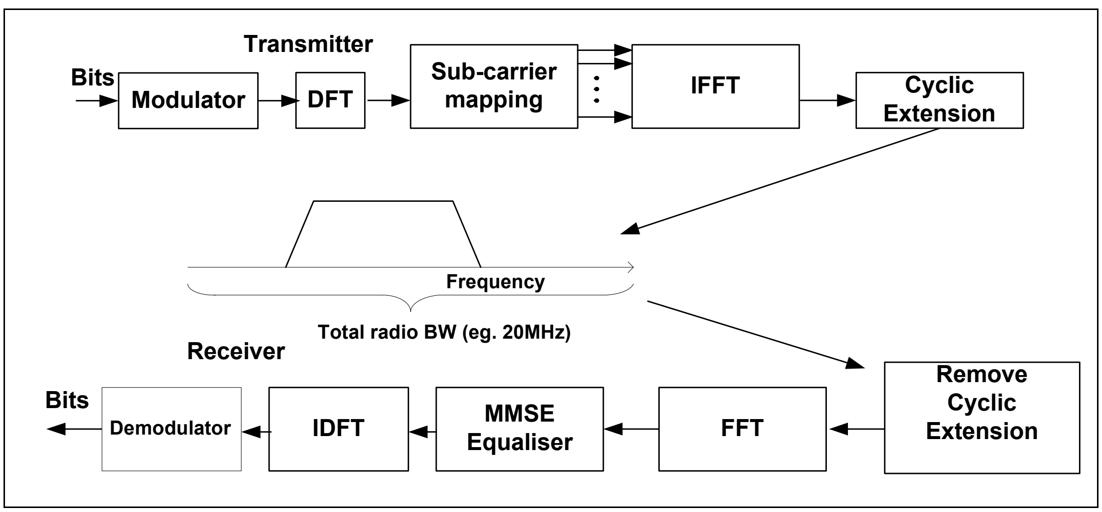

Figure 6. SC-FDMA transmitter and receiver with frequency domain signal generation. Adapted from Holma H, and Toskala A (2009) [14,76] and [15,45;]

In terms of the signal processing, SC-FDMA is similar to OFDMA but for additional steps. In other words, it first runs a Discrete Furrier Transform (DFT) over the group of input bits to spread them over all subcarriers and then uses the result for the IFFT which creates the time signal. [9,52;15,44] This is illustrated in figure 6.

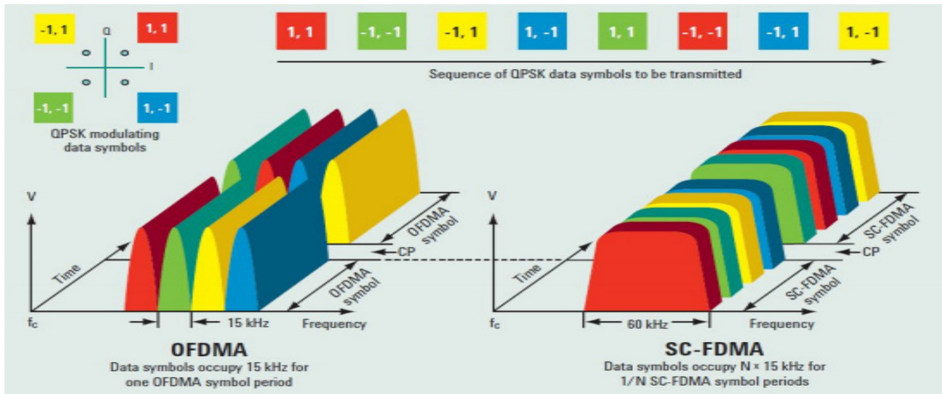

Figure 7. Comparison of OFDMA and SC-FDMA transmission
Reprinted from Rumney M (2009) [16]

From figure 7 we can see that subscriber spacing is 15 kHz. The symbol duration is 66.667us separated by 4.7us cyclic prefix. The CP is transmitted before each OFDM symbol to prevent inter-symbol interference (ISI) due to the different lengths of several transmission paths. 16.67us CP is used for difficult environments.

3.3.2 Physical Layer Parameters

The LTE (downlink and uplink) data transmissions are organized into frames of 10 ms. The frames are divided into 10 sub-frames which are further subdivided into two slots also referred to as Resource Blocks (RB). An RB is the smallest aggregation unit and is composed of 12 subcarriers and 6-7 symbols – the number of symbols depending on the length of the CP as illustrated in figure 8.

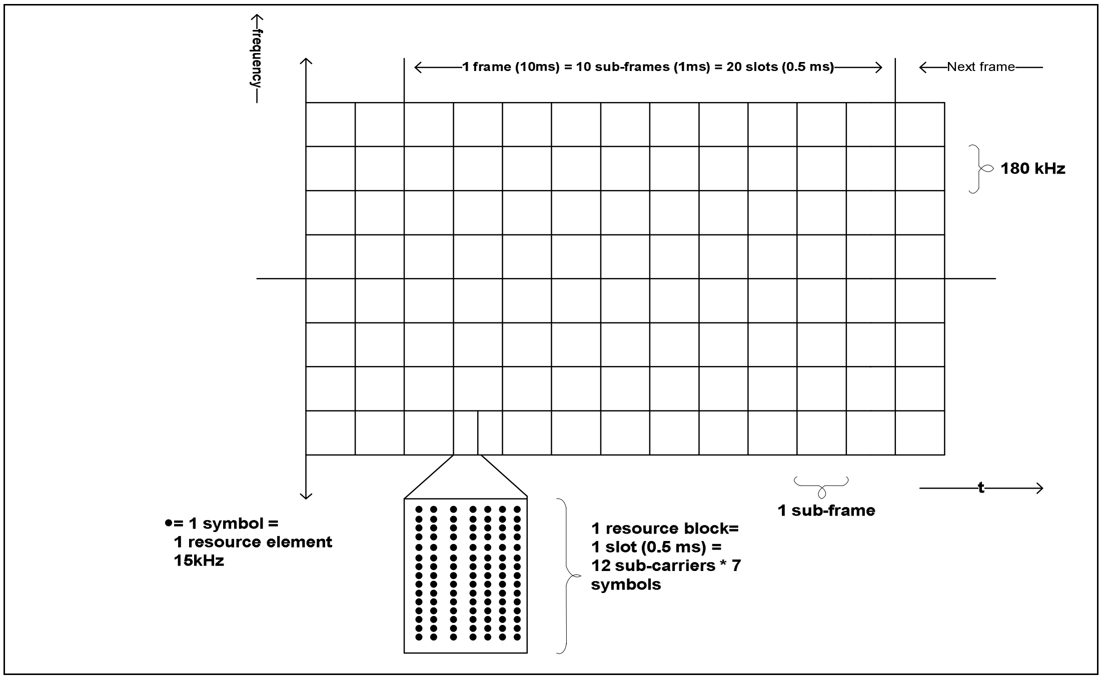

Figure 8. LTE Resource Grid. Adapted from Sauter M (2009) [9,56]

At any given time, a mobile device is allocated at least two resource blocks, and hence the data transmission rate is directly proportional to the number of assigned RBs. It is worth noting that around the centre frequency, some resource elements are used for reference symbols. The pilot symbols (reference symbols) are used by the UE to search for the network during power on and also to find neighbouring cells. In addition they are used for QoS measurements and for error correction.[9,56-57]

3.3.3 Multiple Antenna Techniques

As a means to providing high data rates and efficient spectral use, 3GPP specified multiple antenna transmission techniques provide robustness in radio links and increase data rates under optimal conditions. The eNodeB based on Channel State Information (CSI) is able to select the best multiple antenna technique and also the transmission mode best suitable for the channel condition. A number of multiple transmission techniques are specified for LTE, including spatial diversity, transmit diversity and spatial multiplexing among others. In spatial diversity, multiple antennas are used to improve the quality and reliability of a wireless link. During wireless data transmission, the link may suffer from multipath fading as illustrated in figure 9. Therefore the use of multiple receive antennas alleviates the effect when at least one antennas receives a clear/stronger signal. [4,227-230;11,100]

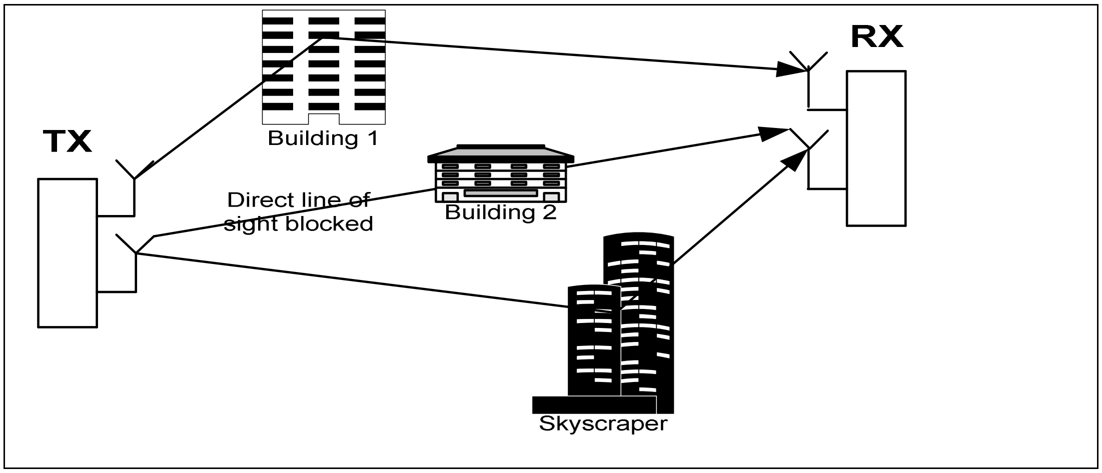

Figure 9. Spartial diversity

Transmit diversity is used to improve the quality and the reliability of the link. It also increases the capacity of the system and the cell range through beam-forming. Spatial multiplexing, sometimes referred to as MIMO, on the other hand has multiple antennas transmitting parallel streams through different antennas. This results in increased bit rate (up to 400 Mbps in DL with 4x4 MIMO using a 20 MHz channel) without the need for extra bandwidth or extra transmission power. [9,60-62;11,100] The 3GPP has specified various MIMO designs: Single Input Multiple Outputs (SIMO), MIMO, Multiple Input Single Output (MISO) as illustrated in figure 10.

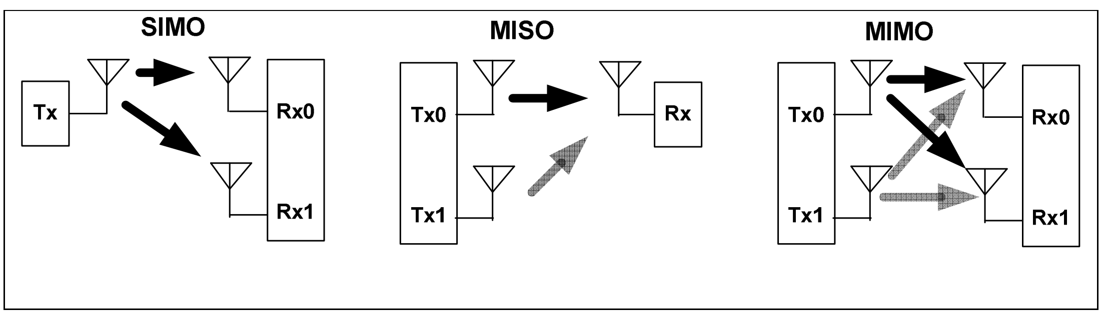

Figure 10. Figure of mode of MIMO. Reprinted from Rumney M (2009) [16,44]

The LTE can support up to 8 transmit antennas and 8 receive antennas in the downlink direction (8x8) and hence the possibility for up to 8 separate transmit streams. In the uplink direction, 4 transmit by 4 receive (4x4) antennas is possible, hence supporting up to 4 multilayer transmission streams. [16,8]

Multi-antenna techniques enable higher data rates and efficient use of radio resources in lightly loaded or small cells. In large and heavily loaded cells, it is best used for single stream beam-forming to enhance the signal quality. [11,100]

3.4 Protocol Architecture

The LTE protocol architecture defines how information flows between the different LTE/SAE elements and is divided into the User-plane (U-plane) and the Control-plane (C-plane). The U-plane is used to deliver and exchange user data while the C-plane is used to exchange signalling messages critical to the UE's connectivity management. [17,21]

In the U-plane, the protocol elements involved are UE, eNodeB, S-GW and P-GW. The U-plane protocol is further stratified into layers composed of the Physical layer (PHY), the Medium Access Control (MAC) layer, the Radio Link Control (RLC) layer and the Packet Data Convergence Protocol (PDCP) layer as seen in figure 11. [17,21]

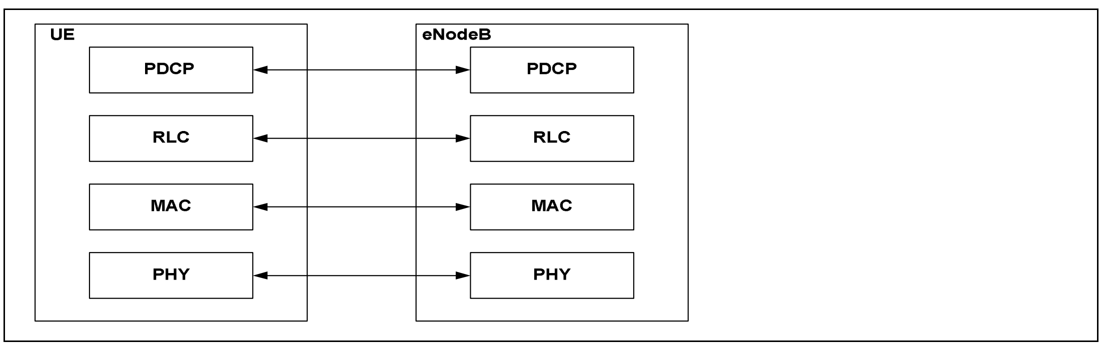

Figure 11. User Plane protocol stack. Reprinted from 3GPP TS 36.300 [15,23]

On the other hand, the network elements involved in the C-plane are the UE, the eNodeB and the MME.

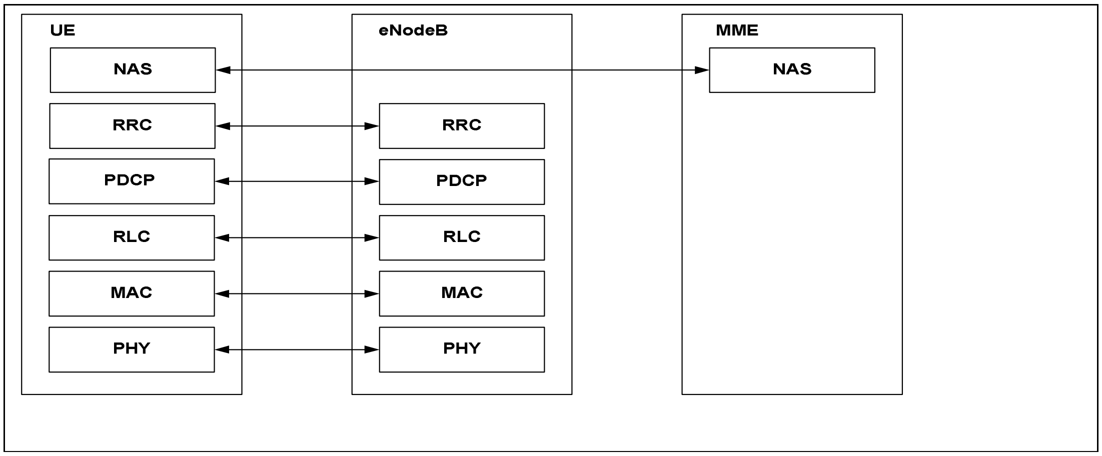

Figure 12. Control Plane protocol stack [15,24]

Similarly its protocol stack is composed of PHY, MAC, RLC, PDCP and in addition, the Radio Resource Control (RRC) layer, as illustrated in figure 12.[17, 21]

3.4.1 Radio Resource Control Protocol

The 3GPP has specified a number of functions for different protocol stack sublayers. The RRC is charged with radio resource control functions which can be categorized into connection management, radio bearer management, mobility management, and signalling connection. Connection management functions include establishing, maintenance and release of the RRC connection between the UE and the EUTRAN. Bearer management is responsible for establishment, configuration, maintenance and release of point-to-point Radio Bearers as well as Radio Bearers for Multimedia Broadcast Multicast Service (MBMS) services. [15,57-58;18,23]

The UE cell selection and reselection and handover procedures are managed by RRC mobility functions. This is achieved by performing measurements and the control of the measurements reporting. Signalling RRC functions include the broadcast of System Information (SI) related to the non-access stratum (NAS) and access stratum (AS), and configuration of signalling radio bearer for RRC, among others. Other RRC functions are QoS management, NAS direct message transfer to/from NAS from/to UE, paging and security functions (e.g. key management).[15,57-58;18,23].

In addition to RRC functions, the RRC protocol states and state transitions are defined for efficient use of the network resources as well as to conserve mobile device battery power. During the RRC_Connected state, data is exchanged between the network and the mobile device. It is however possible for the network to activate the Discontinuous Reception Mode (DRX) in the case of prolonged time of inactivity, whereby the mobile equipment only listens to downlink bandwidth assignments and control commands periodically and is switched off at all other times.[9,63;19,58] The other RRC protocol state, RRC_Idle, occurs in case there is no packet transmission for a significant duration. The radio connection is removed during this state, but the logical and the IP connection is retained. If packets arrive at the MME destined to the UE or the UE needs send data during this state, the MME will send a paging message or the UE requests for connection, respectively, leading to the reestablishment of the radio bearer.[9,63-64;15,58]

3.4.2 Packet Data Convergence Protocol

The PDCP sublayer is located between the RRC and the RLC layers of the protocol stack. Its key functions include header compression and decompression, maintenance of sequence numbering to ensure in-sequence delivery of upper layer Packet Data Units (PDUs), detection of lower layer Service Data Units (SDUs), deciphering and transfer of user and control plane data, and also seamless handovers among others.
In order to protect data (e.g. IP Packet, radio resource control messages and mobility session management messages) from being altered during transmission, the PDCP provides a mechanism whereby the data integrity checksum for each data is calculated before being transmitted.[9,64;19,18-19]

At the receiver, the data is discarded or accepted depending on the integrity of the checksum. Encryption is the other security operation performed at the PDCP layer. An encryption key is calculated by the UE and the eNodeB using the subscriber's secret shared key stored both in the USIM and the HSS. The encryption key is similarly used to cipher IP packets, RRC messages as well as mobility and session management messages which ensure that it is not possible to decode them if they are intercepted during transmission.[9,64;19,18-19]

Due to sensitivity to delay by some data transmissions (e.g. VoIP) and the need to efficiently use the radio interface resources, LTE was designed to support Robust Header Compression (ROHC) which is defined by IEEE in RFC 4995 and RFC 5795. The ROHC framework was the natural choice (for header compression) because of its advantages which include; good compression ratio, built-in feedback mechanisms, which detects compression process corruption, and it has the ability to detect different header types in a packet and apply appropriate compression profile (i.e. compression algorithm).

Some of the profiles specified are the Real Time Protocol (RTP) profile for VoIP, the User Datagram Protocol (UDP) profile for IP and UDP headers, and the Encapsulation Service Payload (ESP) profile for Internet Protocol Security (IPSec) encrypted packet headers.[9,65;19,16-17]

3.4.3 Radio Link Control Layer

The RLC layer lie between the PDCP and the MAC sub layer and it is charged with maintaining Layer 2 data link between the UE and the eNodeB by ensuring not only reliable but also correct delivery of data steams to the receiver. The RLC achieves this by performing functions such as transfer of upper layer PDUs, error correction through Automatic Request (ARQ), duplicate detection, RLC re-establishment, and protocol error detection. For every RLC entity configured in the UE, there is a corresponding peers in the eNodeB and vice versa. It operates in three different modes namely; Transparent Mode (TM), Acknowledged Mode (AM) and Unacknowledged Mode (UM).[17,164;20,8]

While operating in TM, the RLC handles system information messages, paging messages and RRC connection establishment messages on the relevant channels i.e. Broadcast Control Channel (BBCH), Paging Control Channel (PCCH), and Common Control Channel (CCCH). In this mode, it is not necessary to segment and reassemble RLC SDUs because the messages are small enough to fit in a transport block. In the UM RLC SDUs are segmented into RLC PDUs and RLC headers added, a process which is reversed at the receiving peer RLC. Since there are no delivery guarantees in UM, it is suitable for data streams which require timely delivery such as VoIP and video streams. Similarly in AM, RLC SDU segmentation, header addition, and reassembly take

place. In addition however, it has a mechanism to buffer a transmission pending confirmation of receipt by its peer. The receiving RLC employs ARQ functions to detect and reports back to the peer lost RLC data PDUs for retransmission. AM is ideal for data transmission where reliable delivery is of more essence than speed of delivery.[17,164-167;20,8]

3.4.4 MAC Layer

The MAC layer whose main purpose is to control the upper layer's access to the radio resources is located below the RLC layer and above the Physical Layer [18,24].
According to 3GPP's specifications, the MAC layer performs functions such as logical and transport channels mapping, MAC SDUs multiplexing and de-multiplexing, priority handling, error correction through Hybrid ARQ (HARQ), and transport format selection.

Two MAC entities are defined, one in the UE and one in the E-UTRAN. Table 2a lists the transport channels used by MAC and the corresponding direction. [21,11]

Table 2a. Transport channels used by MAC. Reprinted from [21,11]

Transport Channel name	**Acronym**	**DL**	**UL**
Broadcast Channel	BCH	X	
Downlink Shared Channel	DL-SCH	X	
Paging Channel	PCH	X	
Multicast Channel	MCH	X	
Uplink Shared Channel	UL-SCH		X
Random Access Channel	RACH		X

Table 2b. Logical Channels provided by MAC. Reprinted from [21,11]

Logical Channel Name	**Acronym**	**Control channel**	**Traffic Channel**
Broadcast Control Channel	BCCH	X	
Paging Control Channel	PCCH	X	
Common Control Channel	CCCH	X	
Dedicated Control Channel	DCCH	X	
Multicast Control Channel	MCCH	X	
Dedicated Traffic Channel	DTCH		X

Multicast Traffic Channel	MTCH		X

In table 2b, the logical channels provided by MAC are listed. The channels can either be control or traffic channels as indicated. The description of various LTE channels and their mapping is presented in more detail in section 3.5.

3.4.5 Physical Layer

The Physical Layer lies at the bottom of the protocol stack. Its main function is to send user data and control signals between the eNodeB and the UE by employing advanced techniques already introduced in section 3.3 such as OFDMA and SC-FDMA for transmission as well as multiple antenna techniques. The control signals are used for functions such as cell search and synchronization, power control, random access procedures and channel-related procedures and measurements among others.[22,8]

The actual data transmission procedure involves coding, modulation, resource mapping and antenna mapping – a process that is reversed at the receiver. The MAC layer through the various transport channels not only sends data and signals but also controls the physical layer operations. During coding and modulation, the physical layer receives transport blocks, which it in turn adds a Cyclic Redundancy Check (CRC) for error detection purposes. It then performs channel encoding (turbo or convolution) and ensures encoded packet matches physical channel size – a process which is controlled by MAC's HARQ – and may make adjustment to the coding scheme based on the feedback from the receiving peer. After coding, modulation is performed – a process which is controlled by the MAC scheduler.[23,38]

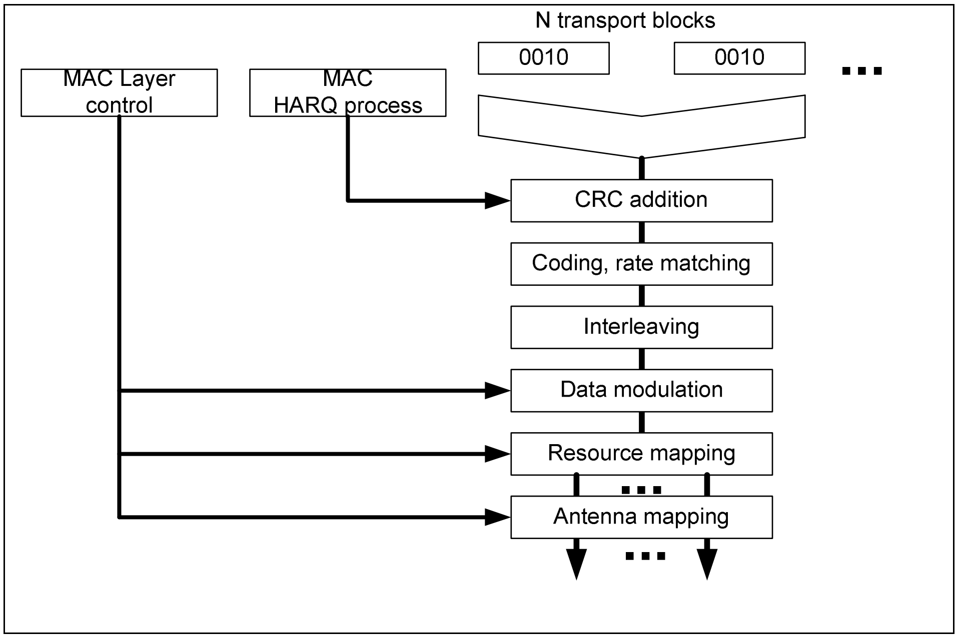

Figure 13. The downlink shared channel PHY model
Reprinted from Shah D S (2010) [23,39]

Data to be transmitted is the segmented and mapped to the resource blocks and then to antenna ports before finally being transmitted as illustrated in figure 13.

3.5 LTE Channel Structure

Channels are interfaces between the LTE layers and are used to segregate data, hence the possibility to support various QoS classes. The LTE having borrowed the channel concept from UMTS preserves the use of the hierarchical channel structure as seen in figure 14.

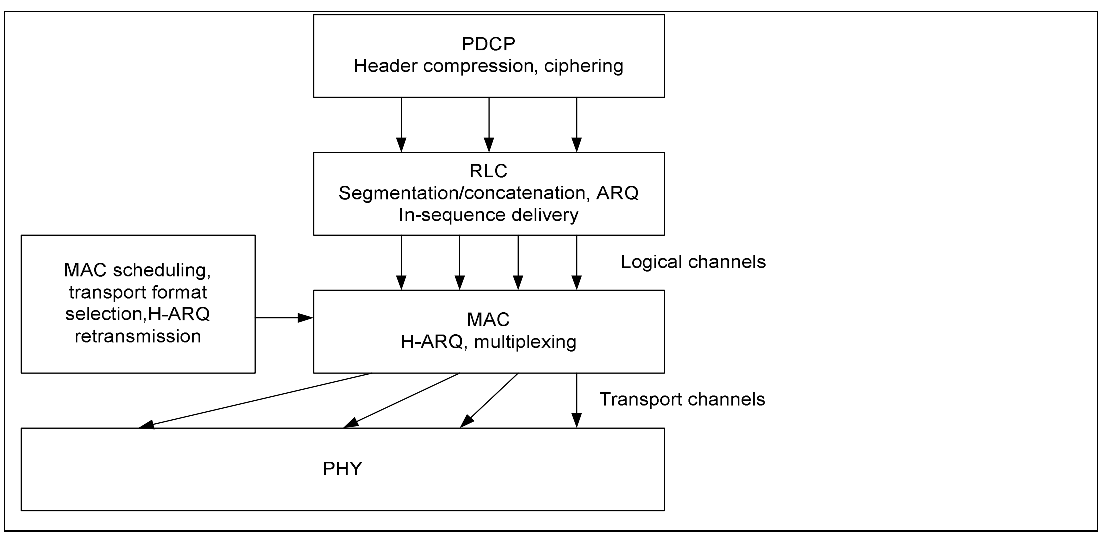

Figure 14. LTE Radio interface protocol architecture. Reprinted from Ghosh A[24]

There are three channel categories each associated with a service access point (SAP), namely the logical channel, the transport channel and the physical channel. In LTE, the transport and logical channels structures are more simplified and fewer than those of UMTS. In addition, the physical layer devices use shared and broadcast channels unlike the use of dedicated channels in UMTS. This model therefore improves radio interface efficiency and can support dynamic resource allocation depending on the QoS requirements and the channel conditions. [24]

3.5.1 Logical Channels

The MAC uses the logical channels to provide services to RLC. Logical channels are categorized into logical control channels and logical traffic channels based on the type of information it carries.The logical control channels are used to transfer control plane information. The Broadcast Control Channel (BCCH) is used to broadcast system control information to the UEs in the cell (i.e. downlink system bandwidth, antenna configuration, reference signal power etc.). It is mapped to Broadcast Channel and Downlink shared channels due to the large amount of info it carries.[11,116]

The Multicast Control Channel (MCCH) is a point-to-point downlink channel for transmitting control information to UE in the cell. It is only used by UEs that receive multicast and broadcast services. Paging information to registered UEs in the cell is transferred by the Paging Control Channel (PCCH). In addition, Common Control Channel (CCCH) is used for transmitting control information between the network and UEs in the absence of an RRC connection - for example during the RRC_Idle state. It is commonly used during random access procedure. On the other hand, the Dedicated Control Channel (DCCH) transmits dedicated control information between the UE and the network when the UE is attached – that is an RRC connection is available. It is a point-to-point and bidirectional channel.[11,116]

Logical traffic channels are used to transfer user plane information includes the Dedicated Traffic Channel (DTCH) and Multicast Traffic Channel (MTCH). The Dedicated Traffic Channel (DTCH) which is a point-to-point bidirectional channel and which can exist in both uplink and downlink direction is used for data transmission between a UE and the network. The Multicast Traffic Channel (MTCH) however is an unidirectional,

point-to-multipoint data channel associated with multicast/broadcast service that transmits data from the network to UEs.[11,116]

3.5.2 Transport Channels

Mainly characterized by how and with what characteristics data is transmitted over the radio interface, transport channels are used by the physical layer (PHY) to offer services to the MAC. Examples of these characteristics are: channel coding scheme, modulation scheme and antenna mapping.

The downlink transport channels are Downlink Shared Channel (DL-SCH), Broadcast Channel (BCH), Broadcast Channel (BCH), Multicast Channel (MCH), and Paging Channel (PCH). The Downlink Shared Channel (DL-SCH) is used for transmitting both control and traffic downlink data, therefore associated with both logical control and logical traffic channels. It supports procedures such as H-ARQ, dynamic link adaptation, dynamic and semi-persistent resource allocation, UE DRX and multicast/broadcast transmission. The downlink channel used to broadcast system information over the entire coverage area of a cell and is associated with BCCH logical channel is the Broadcast Channel (BCH).[11,116;24]

The Multicast Channel (MCH) supports Multicast/Broadcast Single Frequency Network (MBSFN) transmission. MBSFN transmits the same information on the same radio resource from multiple synchronized base stations to multiple UEs, hence used for multicast/broadcast services. It is associated with MCCH and MTCH logical channels. Lastly the Paging Channel (PCH) is mapped to the dynamically allocated physical resources and is required for broadcast over the entire cell coverage area. It is associated with PCCH logical channel, transmitted on the Physical Downlink Shared Channel (PDSCH) and supports UE's DRX.[11,117;24]

In the uplink, the following two transport channels are defined: The Uplink Shared Channel (UL-SCH) which has similar functions as DL-SCH but in the uplink direction is associated with CCH, DCCH, DTCH logical channels. The other one is the Random Access Channel (RACH) is not mapped to any logical channel and is used to transmit data for initial access or during RRC state changes.[11,117;24]

In addition to the transport channels, a number of control information is defined which serves different physical layer procedures. The Downlink Control Information (DCI) which is sent over the Physical Downlink Control Channel transmits information related to downlink/uplink scheduling assignment, modulation and coding scheme, and Transmit Power Control (TPC) commands. The Control Format Indicator (CFI) which is sent over the Physical Control Format Indicator Channel (PCFICH) indicates how many symbols the DCI spans in a given sub-frame.[24]

The H-ARQ carries H-ARQ acknowledgement in response to uplink transmission is sent over the Physical Hybrid ARQ Indicator Channel (PHICH). The Uplink Control Information (UCI) which can be transmitted either over the Physical Uplink Control Channel (PUCCH) or Physical Uplink Shared Channel (PUSCH) is used for measurement indication on the downlink transmission, scheduling request of uplink, and the H-ARQ acknowledgement of downlink transmissions.[11,117-118;24]

3.5.3 Physical Channels

A physical channel corresponds to a set of resource elements in the LTE resource grid. Basic entities that make up a physical channel are resource elements and resource blocks. Physical channels are defined for downlink and uplink.

Downlink physical channels defined include: Physical Downlink Control Channel (PDCCH), Physical Downlink Shared Channel (PDCCH), Physical Broadcast Channel (PBCH), Physical Multicast Channel (PMCH), Physical Hybrid-ARQ Indicator Channel (PHICH), Physical Control Format Indicator Channel (PCFICH). The Physical Downlink Control Channel (PDCCH) which is mapped from the DCI transport channel carries information about transport format and resource allocation related to the DL-SCH and PCH transport channels. It also transports the H-ARQ information related to DL-SCH. In addition, it informs the UE about the transport format, resource allocation and H-ARQ information related to UL-SCH.[11,123-124;24]

The Physical Downlink Shared Channel (PDCCH) is associated with DL-SCH and PCH and carries user data and higher-layer signalling. Additionally the Physical Broadcast Channel (PBCH) carries system information and corresponds to the BSC while the Physical Multicast Channel (PMCH) carries multicast/broadcast information for MBMS

services. The Physical Hybrid-ARQ Indicator Channel (PHICH) is mapped from the HI transport channel and carries H-ARQ ACK/NAKs associated with uplink data transmission whereas the Physical Control Format Indicator Channel (PCFICH) is mapped from the CFI transport channel. It informs the UE about the number of OFDM symbols used for the PDCCH.[11,123-124;24]

Physical Uplink Control Channel (PUCCH), Physical Uplink Shared Channel (PUSCH) and Physical Random Access Channel (PRACH) are the define Uplink Physical Channels. The Physical Uplink Control Channel (PUCCH) carries uplink control information such as CQI, ACK/NAKs for H-ARQ in response to downlink transmission, and uplink scheduling requests. Physical Uplink Shared Channel (PUSCH) corresponds to the UL-SCH transport channel and carries user data and higher-layer signalling.[11,123-124;24]

The random access preamble sent by UEs is carried by the Physical Random Access Channel (PRACH).

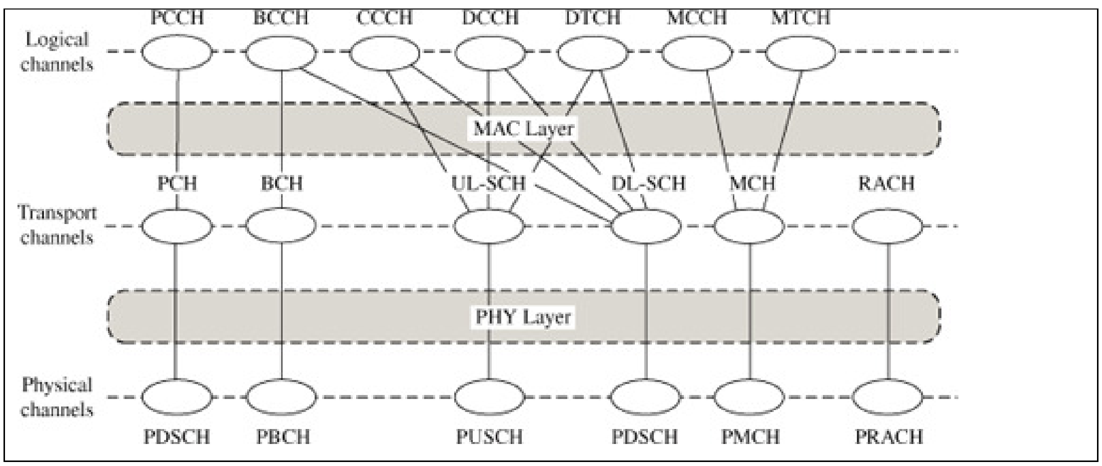

Figure 15. LTE Channel Mapping. Reprinted Ghosh, Arunabha et al (2010)[24]

Figure 15 illustrates the LTE channels and their mappings for both the uplink and the downlink direction.

Furthermore, Physical Signals (i.e. Reference Signal and Synchronization Signal) are defined in the LTE Specification and are only used by the physical layer. The Reference Signal is used for channel estimation and channel quality measurement to allow coherent demodulation and user scheduling respectively. In the downlink, cell-specific refer-

ence signals, MBFSN reference signals and UE-specific reference signals are define. Two types of reference signals, that are demodulation reference signal and sounding reference signal, are defined. On the other hand, synchronization signals (primary synchronization signal and secondary synchronization signal) which are defined only in the downlink and are used during cell search procedures by the UE to complete time and frequency synchronization, and to acquire system parameters.[24]

3.6 LTE-Advanced

As mentioned in section 2.4, the 3GPP in response to the requirements set out by the ITU-R for IMT-Advanced initiated the LTE-Advanced work item. This is culminated in 3GPP LTE Release 10 which proposed technological solutions that complied and even in some aspects surpassed the IMT-Advanced requirements. These enhancements to LTE are briefly discussed in this chapter.

Bandwidth aggregation also referred to as Carrier Aggregation (CA) is meant to address the LTE-Advanced requirement for high peak data rates (see section 3.1). It involves the aggregation of multiple component carriers (upto 5) and jointly used for transmission. Due to the absence of contiguous 100 MHz spectrum required for 1 Gbps peak data rates, component carriers can be non-contiguous making it possible to exploit fragmented spectrum. There are however doubts about its viability because of the cost implications as well as the complexity that it brings to the UE. [11;104, 16;418]

3GPP Release 10 proposes the extention of multi-antenna transmission to support up to eight transmission layers in the downlink and up to four layers in the uplink. In the downlink, the introduction of enhanced reference-signal structure improves various beam-forming solutions. This enables up to 3 Gbps downlink data rates that is with the support for carrier aggregation.[11,104 and 161]

Similarly in the uplink, spartial multiplexing consists of code-book-based scheme hence can be used for transmit-side beam-forming. Data rates of up to 1.5 Gbps can be achieved in the uplink with UL multi-antenna transmission with CA.The higher order MIMO however has its challenges such as increased power consumption, the challenge

of physical space needed for antennas at the UE and the difficulty to achieve the necessary spartial separation of the antennas.[11,104]

Relaying is yet another solution proposed in Release 10 for LTE-Advanced. Relay nodes which are like repeaters are placed at cell edges or areas of poor coverage such as indoors and connects to the donor cell via an in-band LTE-based backhaul. The LTE-based backhauls can however be replaced by optical fibre hence freeing up radio resources to the donor cell.[11,105]

Heterogeneous deployments in the form of femtocells, also known as Home eNodeB (HeNB), is an enhancement to the LTE-Advanced initially proposed and supported in Release 8. Femtocells are deployed over a small area within a larger cell and could operate in the same radio channel (co-channel) as the larger cell or on a dedicated channel. They are mostly deployed indoors but can be deployed outdoors to provide high data rates and capacity in densely populated areas or the rural areas where coverage may be poor or none existant at all.[11,105]

The HeNBs connects to the core network by existing DSL internet connection. Although there are obvious benefits to heterogeneous deployments, there are a number of concerns such as security (i.e to the backhaul, the devices and user authentication), the quality of service (QoS), net neutrality – with regard to backhaul ownership), unnecessary handovers between the macro and femtocell et cetera.[16,424]

4 Evolved Packet Core

4.1 EPC Functions and Main Elements

The EPC was developed to provide a number of functions to LTE which include network access functions such as UE network selection, admission control, authentication and authorization, and charging and policy control. It is also responsible for mobility management, for example UEs idle mode mobility management and user traffic management during roaming. In addition it performs load balancing functions between MMEs to avoid overloading some MMEs. These functions are performed by various network entities which may be implemented as standalone products or different entities combined in a single product.[25,20] These network entities are the MME, SGW, PGW, PCRF and HSS illustrated in figure 16 and whose specific functions are elaborated next.

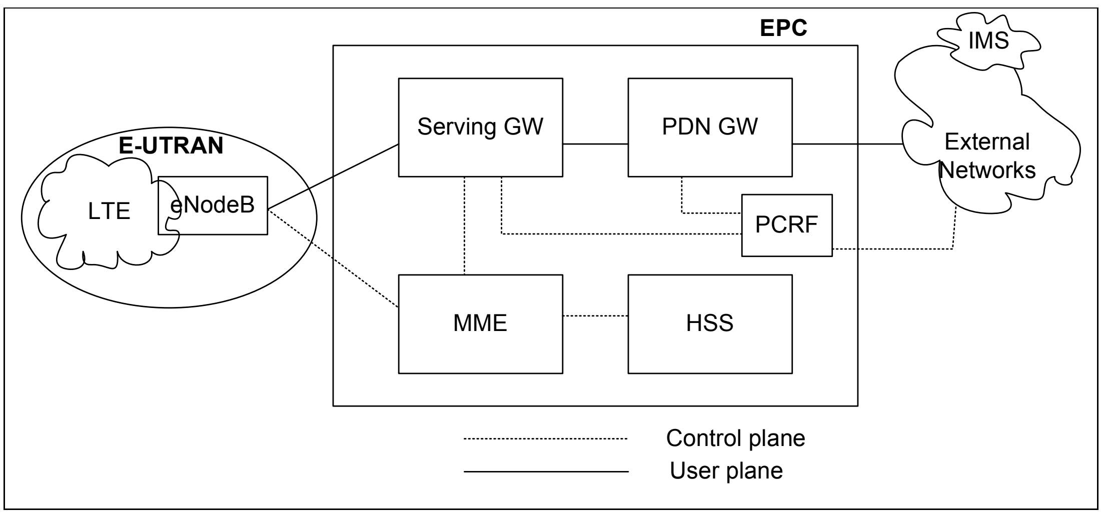

Figure 16.Basic EPC architecture overview. Adapted from Olson et al (2013) [12,26]

The MME is the main node for control of LTE access network. It is tasked with the selection of the SGW for a UE during initial attachment and during handovers. It is also responsible for tracking and paging procdures for UEs in idle mode and the activation and deactivation of bearers on behalf of a UE. Other functions include authentication and authorization in conjunction with HSS, terminating the S6a interface towards the HSS during roaming, roaming restriction enforcements, and providing control-plane functionality for mobility between LTE and 2G/3G access network. Furthermore, it is

responsible for NAS signalling, NAS signalling security, support for relaying function and lawful interception of signalling traffic. [12,368;25,53]

The SGW terminates the interface towards E-UTRAN which means each UE attached to the EPC is associated with a single SGW. It is selected for the UE by MME based on network topology and UE location. It acts as the mobility anchor point for both inter.eNB handover and inter-3GPP mobility, hence tasked with IP packet routing and forwarding. During idle mode, it performs downlink packet buffering and initiates network triggered service request. In order to assist re-ordering functions in eNB, it sends “end markers” packets to the source eNB, source SGSN or source RNC during both inter-NodeB and Inter-RAT handovers. The SGW facilitates access to user traffic during lawful interception.[12,368;25,54]

The PGW contains Access Point Names (APNs) which are logical end points and also mobility anchors of PDN connections and EPS beares which provide connectivity to external PDNs for UEs. There may be multiple APNs each for the PDN that the UE will need to connect to. Since all EPS traffic, inbound and outbound, pass through a PGW, it is from here where packet inspection is performed as well as service level gating control and rate enforcement through rate policing and shaping. The PGW in addition act as the anchor for mobility between 3GPP and non-3GPP technologies such as WiFi and 3GPP2 [12,369;25,54].

The Policy Charging and Rules Function (PCFR) is the policy and charging control element which interfaces to a number of entities such as the Application Function (AF), Subscription Profile Repository (SPR), Charging Systems etc. It takes available information from these entities and in addition configured policies into account and creates service-session-level policy decisions.

These decisions are forwarded to the Policy and Charging Enforcement Function (PCEF) or Bearer Binding and Event Reporting Function (BBERF) for enforcement. It in addition forwards events reports between BBERF, the PCEF and AF. In short, the PCRF ensures user-plane traffic mapping and treatment is in accordance with the subscription profile associated with the end user.[12,369;25,56]

The master database that contains subscription-related information for a given user is the HSS. The information therein supports network entities that handle mobility in th CS domain, PS domain and the IP Multimedia Core Network (IPCN). Furthermore, it generates user security information for mutual authentication, communication integrity check and ciphering. This information is used to determine access authorization and service authorization.[12,272]

4.2 EPS Deployment Process

Given that the EPS was developed to improve several aspects of the existing systems (i.e. GSM, WCDMA), it was designed to allow for internetworking between these systems. This was to ensure that the EPS deployment does not interrupt the existing system and also so that they can be operated in a complementary manner, for example in an area where one system has a superior coverage that the other. Besides Internetworking with 3GPP systems, the 3GPP made specifications for the EPC internetworking with non-3GPP radio access technologies.[12,67]

In figure 3 of section 3.2 the EPC is indicated by an oval which illustrates the possible interconnections with virtually all packet data access networks. From the figure, we can see that there are a number of options for the 3GPP family systems, for CDMA and also both trusted and non-trusted non-3GPP systems. Treating an access network as trusted or non-trusted is the prerogative of the 3GPP network operator. Examples of trusted and non-trusted technologies include fixed Wi-Fi and WiMAX networks. In addition, for the 3GPP technologies (GSM and WCDMA), there are two interface options to choose from, i.e. the s4 interface or the Gn interface.[12,27]

4.2.1 Initial Deployment Phase

A network operator may not necessarily deploy an EPC network with all the nodes and interfaces illustrated in figure 3 of section 3.2. A more likely scenario is where there is a GSM/WCDMA or a CDMA network already in place and the LTE is then rolled out in phases. In the first phase, the operator deploys the various physical entities – a process that involves dimensioning of the new EPC network and individual nodes, planning of the integration of the nodes into IP infrastructure, and the configuration of the IP entities.[12,67]

In addition, the operator utilizes the pooling capabilities of the entities where considerations for balancing the traffic load and network redundancy are made. This is beneficial for the operator because it ensures optimized capacity utilization, high service availability and seamless capacity expansion. [12,68]

The final step in the first phase is to have the network support access by multiple radio access techniques/systems (i.e. GSM/WCDMA or LTE) and such that the UE can utilize the network with a better signal/coverage. This calls for proper planning of the user authentication and authorization mechanism when accessing the network from either RAT. Additionally, the network is set up and configured to ensure traffic from the different RAT is routed to the right network entities. [24,71]

4.2.2 Integration Phase

The second phase in the LTE/EPC deployment involves the integration of the EPC with the existing packet core. This makes inter-system mobility possible as well as roaming by LTE users and as a result improving user service and experience. In order for inter-system mobility to be possible, SGSN serving GSM and WCDMA is interconnected with MME and PGW in the EPC. Importantly, the SGSN should be able to distinguish between devices attached to it that are able to move to LTE and those that cannot so that it can select the correct Gateway (GW) node. [12,71]

To ensure correct GW selection by SGSNs, one option is to upgrade all GGSNs to support P-GW functionalities or to replace the GGSNs with P-GW nodes. Another option is to have SGSNs to distinguish between GGSNs that serve GSM/WCDMA and P-GWs that can serve users over any access. There are two possible ways to achieve this; i.e. to configure Access Point Name (APN) on the UE to point to the preferred external network or based on the SGSN's knowledge of the UE capabilities. The final step to ensure inter-system mobility is to upgrade the GSM/WCDMA network to support fall back by LTE in case of coverage loss. [12,71-74]

To facilitate LTE roaming, the S-GW in the visited network (VPLMN) is connected to the P-GW in the home network (HPLMN) over the S8 interface. In addition, the MME is

connected to HSS in the HPLMN over the S6a interface where Diameter-based signalling is utilized. To provide redundancy, improved scalability, and security for the inter-operator connections, it is recommended that Diameter agents are installed in both VPLMN and HPLMN.[12,76-77]

The final step of the second phase is to make network capacity considerations, which calls for careful monitoring and dimensioning of control signalling load levels. This is necessary because of the divergent user device and applications behaviour in the signalling plane. In addition, over time, the introduction of more LTE users and devices will have an implication on the overall signalling load. [12,75-76]

4.2.3 Optimization Phase

In the third and the final phase of LTE/EPC deployment, the aim is to optimize the various EPS aspects such as network elements, subscriber data management and the overall performance. In terms of network optimization, elements such as the SGSN and MME can be combined because they perform similar roles i.e. signalling to their respective RAT. Additionally, the GGSNs can either be upgraded to SGW/PGWs or phased out all together and replaced with SGW/PGWs. These two steps simplify network operations and also optimize the overall capacity. [12,76]

Subscriber data management can be optimized by having only the HSS handling subscriber data for not only LTE but also GSM/GPRS and WCDMA/HSPA. This is achieved by having the appropriate interconnection between the SGSN and the HSS via the S6d interface. Another aspect to be considered for optimization is the overall network performance during inter-RAN mobility. In order to reduce LTE to WCDMA handover duration, the LTE base station (eNodeB) can be provided with applicable system information of the target WCDMA cell through SGSN-MME signalling which in turn provides the UE. Additional optimization can be achieved by packet handovers whereby a target cell is prepared in advance by for example bearer establishment via inter-RAN signalling. [12,77]

4.3 LTE User Services

The LTE network is designed to only provide IP-based bearer service to a mobile device i.e IP connectivity between a UE and a PDN. This in turn enables a user to access

services such as voice and video telephony, location-based services, sms and mms, email, and roaming services depending on the capabilities of the external PDNs connected to the UE. These services which are grouped into data services and voice services are managed by the IP Multimedia Subsystem (IMS) or may be Over the Top (OTT) services.

Data services include messaging services and machine-to-machine (M2M) communications. The messaging services are implemented either by IP-based solutions such as the IMS or via the CS infrastructure in which case it demands specific mechanisms given that LTE is strictly packet based. M2M communication is exploited for industrial and corporate use or by governmental agencies. M2M can for example be used for corporate fleet management, surveillance systems, remote utility control and meter reading and infrastructure and service management. [12,81]

Voice services on the other hand can either be through Voice over LTE (VoLTE) or CS-Fallback (CSFB). VoLTE utilizes the Multi-media Telephony (MMTel) service of the IMS to offer voice calls. Other than offering enhanced features such as video, text and other multimedia in addition to the voice service, it allows for a handover to legacy systems for example due to lack of LTE coverage or insufficient QoS or bandwidth. Single Radio Voice Call Continuity (SRVCC) is an example of such handover whereby IMS-anchored sessions are transferred to legacy networks (GERAN or UTRAN). [12,89]

CSFB allows an operator to provide data services with EPS while reusing the CS for voice services. The UE during power-up registers in both the EPS and the CS domain but camps in the LTE network. The UE switches from LTE to the CS system whenever it needs to use the CS services i.e make or receive a voice call or message. CSFB provides operators with a migration path from CS systems to the EPS. [12,93]

4.4 LTE Security

Similar to all digital communication systems, the EPS faces a number security threats and vulnerabilities. Some threats specific to the EPS are:

- Threats to user identity and privacy
- Threats to UE tracking and those related to handovers

- Threats to the eNodeB and the backhaul
- Threats to multicast and broadcast signalling.

In addition, threats to abuse of network services, radio protocol, mobility management, and unauthorized EPC network access cannot be ignored. A Denial of Service (DoS) attack can be targeted at either the radio interface or an element in the EPC hence posing a threat to both the EUTRAN and the EPC.[12,157]

With all these threats inherent, it is paramount that the security measures are put in place to ensure the system (EPS) functions properly and is not abused. Given that most of these threats are not new, the 3GPP adopted, enhanced and extended security features from GSM and UMTS systems in line with the new architectural and business requirements of EPS (LTE/SAE). Subscriber authentication, encryption at the radio interface and use of temporary identities are some of the features that have been carried over from GSM and UMTS. The main security measures in EPS as in other information and communication systems in general are control (i.e. authentication and authorization) and integrity and confidentiality measures.

Authentication is the process of validating the identity of an entity whereby it proves that it is what it claims to be. In the EPS, both the user and the network performs mutual authentication by proving access to a shared secret only known to them. Once a user has been authenticated, he/she is granted the rights to access the various services and resources depending on the device's capabilities and service agreement levels. [12,158]

The next step after authentication and the determination of agreement levels is to ensure the secure transmission of both signalling traffic and the user-plane traffic. There are two measures towards this end one being ciphering where data to be transmitted is encrypted to ensure it is only readable by the intended recipient. The other measure is integrity protection which provides a means to detect tampering/modification of data traffic along the transmission path. Additional EPS-related security measures are subscriber information privacy protection and the physical infrastructure security such as that of eNodeBs and HeNBs. [12,158]

The 3GPP in TS 33.401 has grouped the security features of the EPS architecture into security domains. These domains, network access security, network domain security,

user domain security, application domain security, and visibility and configurability of security deal with security threats and offer security solutions specific to each domain. [26,14]

4.4.1 Network Access Security

Security features such as mutual authentication, privacy, and protection of both signalling and user-plane traffic in an access system are what is referred as network access security. In the EUTRAN, the NAS process begins with the communication parties (UE and the network) verifying each other's identities. This is done in such a way that the user identity confidentiality and user location confidentiality is not compromised and also the user cannot be traced by eavesdropping on the radio access link. To achieve this, the user is assigned a temporary identity and also by ciphering signalling or user plane data that may reveal the identity of the user [27,92;28,14].

Once the user has been identified, mutual authentication will be initiated in the HSS/AuC. The mutual authentication known as EPS authentication and Key Agreement (EPS AKA) is based on a shared secret key K, stored in the USIM and the HSS/AuC in the operators' network. The MME then fetches the authentication data (authentication vector – (AV)) from the HSS/AuC and then triggers the AKA protocol with the UE. The MME then authenticates the UE verifying the response (RES) from the UE is equal to the expected response (XRES) from its own AKA procedure. Additional keys are generated now that the UE and the network share a similar Key Access Management Entity (K_{ASME}). These additional keys are used to protect the traffic between the UE and the E-UTRAN such as NAS signalling between UE and MME, RRC signalling between UE and eNodeB, and user-plane traffic between UE and eNodeB.[12,161-166]

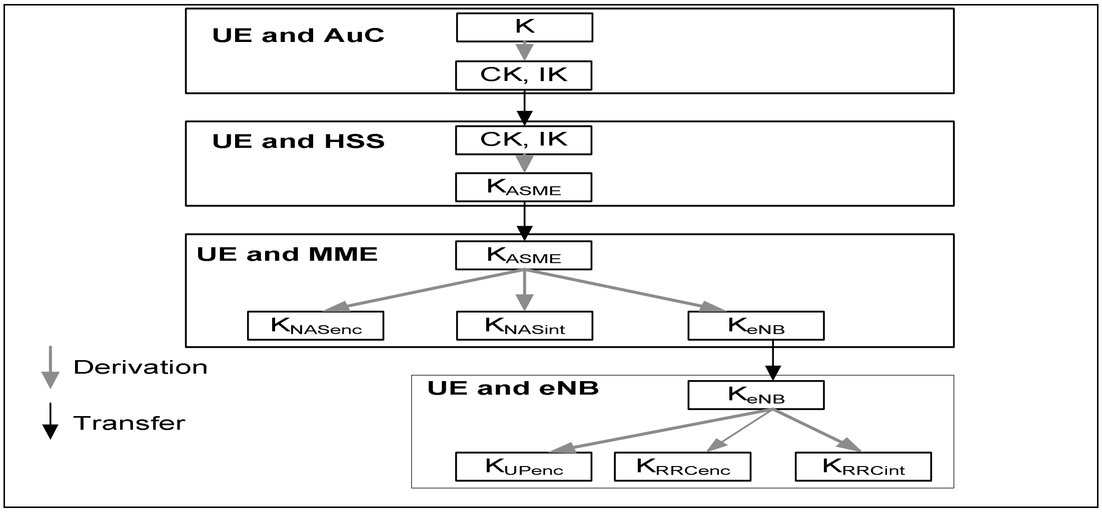

Figure 17. Key Hierarchy for E-UTRAN. Reprinted for Oslon et al (2013) [12,166]

Figure 17 illustrates the key hierarchy of E-UTRAN in EPS. It can be seen that whereas as it is inherited from the UTRAN, it has been enhanced such that there is a strong key separation between networks and key usage. Additional enhancements are larger key sizes and measures for the protection against compromised base stations.

Even with all these security measures in place, the user should be able to make calls during emergencies. The 3GPP, in TS 23.122 and TS 23.167 has specified how the EPS should handle emergency calls in situations where it is not possible to authenticate the UE. Support for inter-RAT IMS calls handover is also specified and also handover from E-UTRAN to HRPD as well as the related security procedures.

As mentioned in section 4.3, legacy 3GPP RAT as well as non-3GPP RAT can be connected to the EPC and consequently, related access security measures have been specified in TS 33.401 and TS 33.42 respectively. With regard to inter-3GPP RAT mobility (i.e. between GERAN/UTRAN and UTRAN), the option would be to perform new authentication and key agreement procedure at every inter-RAT handover. Due to delay implications however, this is not ideal and therefore native or mapped security context is preferred. Utilizing native security context may be possible if the network caches a native security context such that when a UE returns from a different RAT, it is not necessary to rerun a full AKA procedure. On the other hand, in the absence of cached security context, mapping of security context from source RAT to a security context for the target RAT is performed. [12,168]

4.4.2 Network Domain Security

The network domain security (NDS) is subdivided into security domains – networks that are managed by a single administrative entity. The purpose for NDS is to protect traffic between network elements by mutual authentication between the communicating parties, data confidentiality and integrity. Traffic protection is made even more necessary given that it may traverse third-party IP networks or that the network may interface with other operators for roaming purposes.[12,176]

IP-based protocols are secured by Security Gateways (SEGs) which sit on the borders of IP security domains. Before entering or leaving security domains, NDS IP traffic passes through a SEG. Traffic between the SEGs us protected using IPSec ESP in tunnel mode where either Internet Key Exchange version 1 (IKEv1) or IKEv2 is used to set up the IPSec Security Associations (SA). Figure 18 illustrates the NDS architecture. [12,177;29,16]

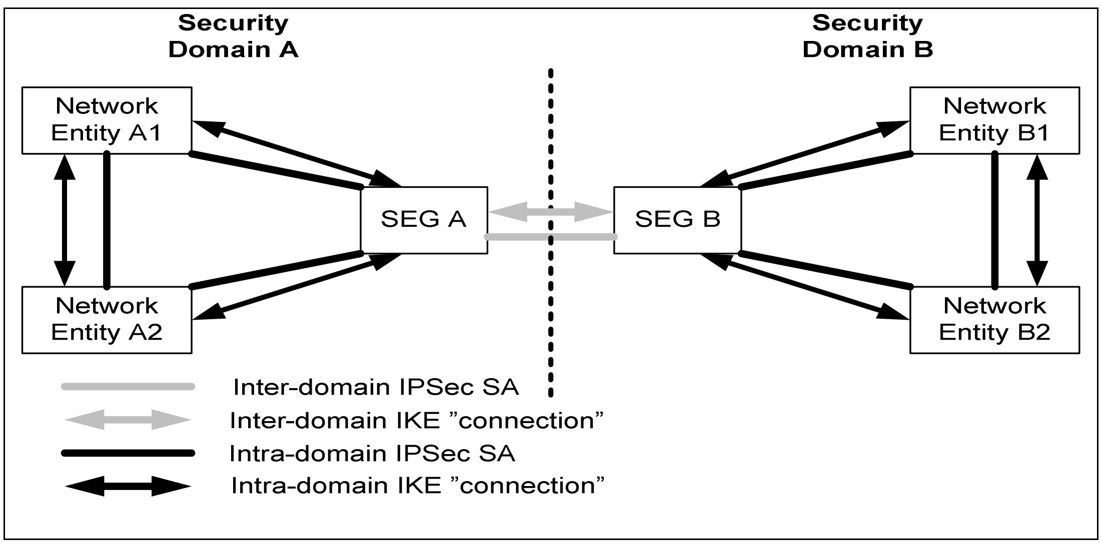

Figure 18. NDS Architecture. Adapted from Olson et al (2013) [12,178] and [29,16]

According to 3GPP specification TS 33.210, in addition to making the use of ESP mandatory, integrity protection, message authentication as well as anti-replay protection is mandated. Other security services provided by NDS/IP that may be used include data integrity, confidentiality, and data origin authentication.[29,10]

4.4.3 User Domain Security and Other Security Features

In the context of user domain security, the security features that secure the physical access to the UE, the USIM is blocked until it has authenticated the user. The USIM stores a shared secret (PIN) which the user enters on the terminal before being granted access.[12,160]

The application domain security enables applications in the user and provider domain to exchange messages securely. It is used by applications such as web browsers and as such they are operated end to end over the user-plane transport of the EPS. This means therefore that the application security is transparent to EPS and that it relies on the trust of the various underlying security domains of the EPS.[12,160]

Visibility and configurability of security features ensures that the user not only is able to learn of security related events but also is able to configure some security related settings. An example of visibility of security is when the user is notified whether or not encryption is enabled on the radio access link. On the other hand, configurability allows the user to enable or disable for example USIM authentication, and accept or reject incoming non-ciphered calls.[12,161]

4.5 Basic Procedures

When a user powers on the UE, it will have to undergo the network attachment procedure. During this procedure, the UE registers with the network (in order to receive services) and the EPS bearer is established and the IP address is assigned.

The initial step is for the UE to scan for synchronization signals in all the frequency bands it supports. It starts the search on the last used channel before switch off and once found and properly decoded, it reads the broadcast information (i.e. cell systems info such as Network ID (PLMN ID)), Tracking Area ID (TA ID), cell ID and radio and core network capabilities. If it is not found, it will select another network based on preferences stored on the SIM card or it will display detected networks for the user to make a choice.[9,66]

After the UE has selected a cell and PLMN, hence synchronized with network in downlink direction, it will perform the Random Access Procedure (RAP) in order to be synchronized in the uplink direction. It does this by sending a short message over the Random Access Channel. RAP can either be contention-based mode in which there is a possibility of network access collision due to the attempt by multiple UE to access the network or non-contention-based which is initiated by the network for example during UE handover from one eNodeB to another. The network responds with a Random Access Response message which contains the assigned Cell-Radio Network Temporary Identifier (C-RNTI) and an initial bandwidth grant (set of resource blocks) of which the UE then uses to send the ATTACH_REQUEST message together with PDN_CONNECTIVITY_REQUEST for PDN (IP) connectivity to the eNodeB which in turn forwards it to the MME.[9,67]

The MME then tries to identify the UE using Temporary Identity (TI) and if there is no corresponding IMSI in the MME, it queries the AuC in the HSS for authentication information. In case the AuC cannot resolve the TI, the MME requests the UE to send its IMSI after which the UE and the network partake in mutual authentication described in section 4.5.1. Upon successful authentication, the eNodeB and the UE activates the air interface while the MME sends and Update Location Request message to the HSS confirming proper authentication of the subscriber. The HSS responds by sending Update Location Acknowledge message which contains subscribers profile data such as authorized connection and QoS, hence terminating the Update Location procedure.[9,67]

In the final step of the attach procedure, the default bearer for data transfer is setup. It is initiated by the MME by sending a create Default Bearer Request message to the PDN-GW via the S-GW. The PDN-GW creates a new entry in its context table and assigns an IP address to the UE which it forwards to the MME again via the S-GW. During this process, a tunnel is created between the PDN-GW and the S-GW for the particular subscriber. The MME then sends the ATTACH_ACCEPT message which contains the UE assigned IP address to the eNodeB which in turn forwards it to the UE via the Radio Bearer Establishment Request Message. On receiving the message, the UE sends an ATTACH_COMPLETE message in the Radio Bearer Establishment Response message to the MME – signalling that the process is complete. At this point the mobile equipment can make data connection to either the Internet or the operators' internal IP network.[9,68]

5 Cloud computing

5.1 Definition, History and Cloud Computing Features

The origin of cloud computing can be traced back to the 1906s during which John McCarthy presented the concept of utility computing and the development ARPANET – the precursor to the Internet – by J.C.R Licklider [30, 1]. In the years that followed, computer information and communication technology has gradually evolved to bring to reality such concepts as grid computing, peer-to-peer computing, client-server computing and service-oriented computing, which are predecessors to cloud computing. In particular, technological advancement in various fields have led to the availability of powerful microprocessors, large and affordable data storage, high capacity data networks, and system virtualization which have had a direct influence on the evolution towards cloud computing.

One can be forgiven for failing to draw a clear distinction between technologies such as grid computing, utility computing and even traditional server-client computing from cloud computing. In essence, these technologies overlap. However, the following features distinguish cloud computing from the rest:

- On-demand self service: This means a user can access the cloud resources or services whenever they need without the need for intervention by the provider. This calls for an effective and user-friendly interface which eliminates interaction between user and provider, hence reduced costs for both parties. Examples of on-demand service providers are Amazon. Google, Salesforce.com among others.
- Broad Network Access: Cloud resources are accessed over the Internet from diverse locations using a variety of client devices such as mobile phones, tablets and laptops. Easy access to reliable high speed Internet, courtesy of fibre optic broadband and wireless broadband such as WiMax and LTE, justifies the use of LTE solutions.
- Resource pooling: The cloud provider should have a large resource base which can be dynamically allocated or reallocated as required. These resources (e.g. storage, memory, virtual machines etc) may be located in different locations but the actual implementation is masked from the user.

- Rapid elasticity: Given that it is not possible to know the user need before hand, the service should therefore be available around the clock and be able to scale up or down depending on user demands.
- Measures service: There should be a mechanism for both the provider and the consumer to monitor, control and report cloud resources usage. This is important especially because the provider has to have a way to bill the user for usage. [31;32]

5.2 Cloud Computing Architecture

There are different classification models for cloud computing which include: Software, Platform and Infrastructure cloud classification, otherwise known as SPI model, Santa Barbara-IBM Cloud Ontology (UCSB-IBM) and Hoff's Cloud Model among others [33]. The SPI model is the widely accepted and supported by most institutions and cloud service providers (CSPs) perhaps given that initially cloud computing services were categorized into service, platform and infrastructure.

5.2.1 Cloud Service Models

As mention in paragraph above, cloud services are generally grouped into three categories: Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS) and Software-as-a-Service (SaaS). Cloud services are therefore made possible by the interoperability of the three service models.

IaaS enables for access to resources such as data storage, networking equipment and computation as per need basis. This means that the customer is able to increase or decrease resource capacity as required and consequently is charged according to the same. Its implementation has been made possible by various advancements in virtualization technologies. Amazon's Elastic Compute Cloud (EC2) is an example of successful implementation of IaaS where customers can rent virtual computers/servers and run their own applications [34,340].

PaaS on the other hand is intended for software developers and provides them with a platform as well as the necessary Application Programming Interfaces (APIs). Through the APIs, web portals or software installed on a computer, a customer (developer) can

build, test, deploy and tune their applications in the cloud over the Internet [33,3]. Google Apps and Force.com are examples of PaaS.

In SaaS, browser-based applications designed to meet specific user needs are built and made accessible over the Internet either free of charge of for a small fee [32,10]. There is a large range of SaaS products and it includes: office productivity products such as Google Docs, Customer Relation Management software, project management software, email software among many others. By subscribing to SaaS services, the customer gets the benefit of not having to worry about software licenses, upgrades or maintenance because it is handled by the provider.

It is worth noting however that different organizations or CSPs may have a different view of how to define the service delivery models perhaps instructed by their products and services differentiation and other reasons. This is as can be seen in figure 19 of section 5.2.3 where in addition to IaaS, PaaS and SaaS, Network as a Service (NaaS) and Communication as a Serive (CaaS) are depicted in the services layer.

5.2.2 Cloud Deployment Models

Cloud service models deal with how cloud services are delivered to the user, whereas cloud deployment models define the physical location of cloud equipment and also who is responsible for its management. Deployment models are functionally not related with service models and can be referred to as external or internal clouds [30,43].
The National Institute of Standards and Technology (NIST) has defined four deployment models:

- Private cloud
- Public cloud
- Hybrid cloud
- Community cloud.

All the four models bear the basic characteristics of cloud computing i.e. access via the Internet, dynamic scaling of virtual resources and their implementations are abstracted to the user [30,44].

In a private cloud, the infrastructure is dedicated to a single organization and located typically on-site or off-site. Its management and operation may be in-house or by a

third party. It offers services similar to the other cloud models, whereas an organization has more control over the data and processes. In addition, due to the more control, security in a private cloud is considered to be higher than in the other models. [30,48] Furthermore, direct resource monitoring, control and maintenance result in improved efficiency and lower operation costs. Its benefits notwithstanding, the initial setup cost implications make it a prohibitive option for start-up or small companies and as such make it a preserve for large wealthy organizations and government institutions.

Companies, institutions and individuals keen on using cloud services can benefit from a public cloud. A public cloud is owned by organizations selling cloud services and is available by subscription to the general public. The Cloud Service Provider (CSP) offers its services such that it can be dynamically provisioned and the subscriber billed based on usage. To the subscriber, this means they have on-demand access to affordable cloud services with minimum investment in IT infrastructure and operational overhead. The CSP on the other hand by offering quality services benefits from a large customer base and in turn is able to invest more in its infrastructure leading to improved quality of services offered.[30,44]

A community cloud is comprised of infrastructure that is shared by a group of organizations with similar interests and requirements (e.g. policy, mission, compliance considerations). Examples of such organizations and institutions are universities and government institutions. It is managed jointly by the organizations or by a third party and can be located in or off site. The shared cost of its establishment and management translates to overall lower cost of owning and utilization of the cloud by the involved organizations. Additional benefits of a community cloud are graceful failure, convenience and control, and environmental sustainability. [30,46]

A cloud infrastructure that combines at least any two of the aforementioned cloud models is referred to hybrid cloud. This kind of arrangement allows organizations the flexibility to utilize public clouds when there are cost-saving benefits and still own a private cloud in which they have control over the security and privacy of their confidential and propriety data. A hybrid cloud operation is made possible by use of APIs shared by the interacting clouds.[30,49]

The different cloud models offer a variety of benefits as well as disadvantages and therefore it is upon the interested individual or organization to carefully consider its options before settling for a particular option.

5.2.3 Cloud Computing Implementation Hierarchy

In order for a cloud system to be considered properly implemented, it should be such that it can be accessed in an easy convenient and secure manner. It should also be possible to be configured and used with ease. Towards this, the cloud implementation architecture is stratified into access layer infrastructure, service layer infrastructure and resource layer infrastructure as illustrated in figure 19.

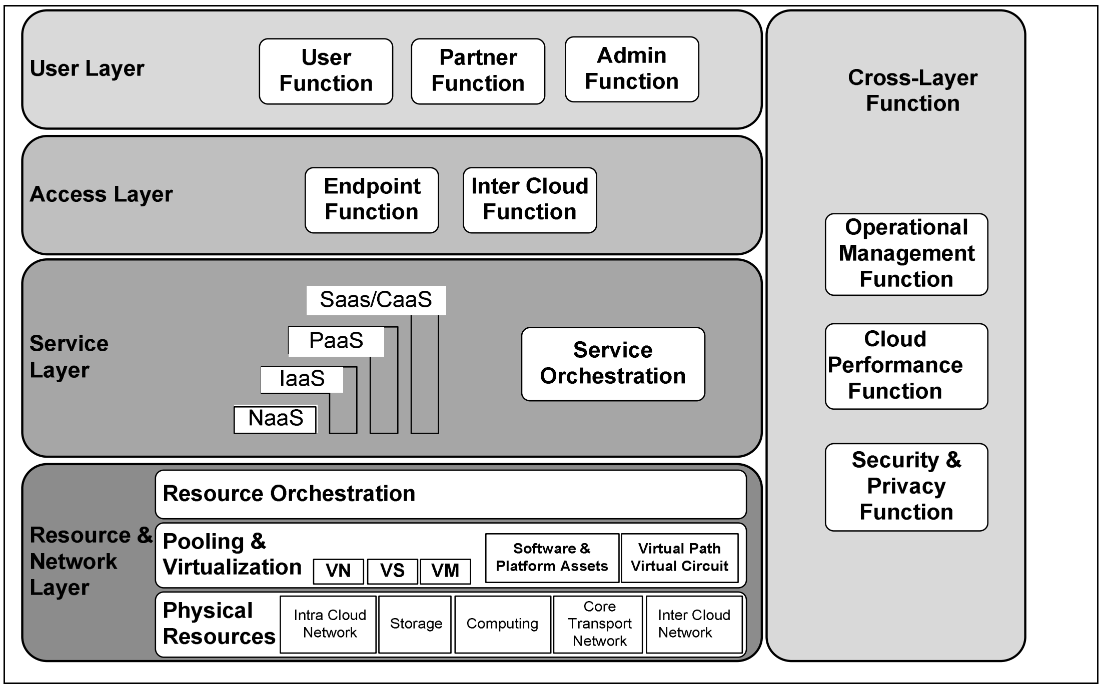

Figure 19. Diagram of cloud functional layers.

Adapted from Focus Group on Cloud Computing [35,8]

The access layer provides the user with the interface to access the various cloud services. It is from this layer that the user is able to request and receive cloud services, utilize the requested service, and administer and monitor cloud resources in a secure manner. In addition, the cloud administrator performs some administrative roles from this layer.[35,5]

The service layer infrastructure organizes/arranges and exposes the various cloud services (i.e. IaaS, PaaS and SaaS) to the user. It is in essence the entity responsible for making a cloud service to be available to the user by managing the infrastructure required for providing the services, running software that implements the services and arranging and delivering the services.[35,6]

The physical layer is where resources such as servers, storage, networking devices (e.g. switches, routers) and the access architecture reside. It provides functions for resource management, monitoring and scheduling hence proving services to the upper layers. It also provides mechanisms for pooling and virtualization of resources to ensure an on-demand and elastic cloud [35,12].

In addition (to the three layers), there are functions that traverse the three layers and a referred to as cross-layer functions. They perform overall system management, monitoring and provides security mechanisms.[35,7] The cross-layer functions include security and privacy, operational management, and cloud performance functions. By monitoring real-time cloud information such as security alert logs, failure alert logs, resource utilization logs and SLA violations, a CSP can take appropriate measures to ensure secure, optimized and efficient service provision.

The cloud hierarchy present is a logical view of the cloud infrastructure. This means that the actual cloud deployment may vary from one vendor to another depending on their own need and their core services.

5.3 Cloud Components

Cloud computing components are grouped into two groups: front-end and back-end components which are connected via the Internet as illustrated in figure 20.

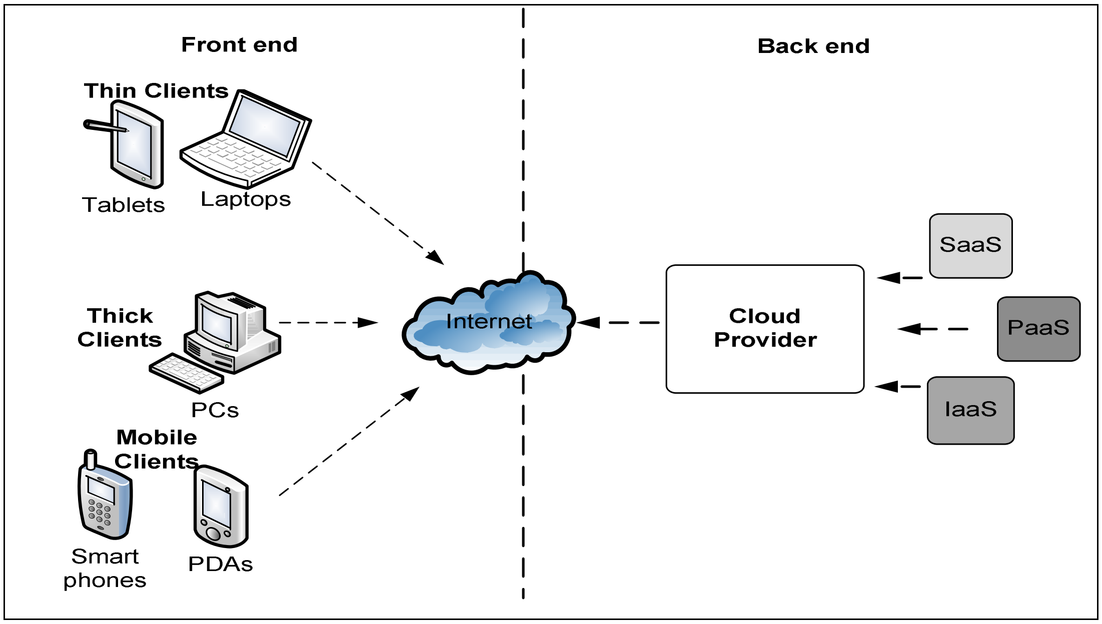

Figure 20. Diagram of cloud infrastructure

The front-end may be comprised of client devices which are categorized into mobile, thick and thin clients or enterprise networks with the relevant applications for accessing the cloud. The role of the front end client devices is to offer a platform to the user to access and manage their information in the cloud. Client devices are installed with interface applications (such as web browsers or proprietary applications) to access the cloud. Examples of client devices are desktop computers, laptops, tablets and smart phones among others. [32, 46] On the other hand, the back-end is basically the components such as applications, data storage systems and servers that together or in groups form a cloud computing system.

5.4 Benefits and Disadvantages of Cloud Computing

5.4.1 Benefits of Cloud Computing

Among the benefits of cloud computing is the ease with which IT infrastructure as well as personell can be managed. By transferring the IT infrastructure into the cloud, the CSP is charged with its management and maintenance, hence freeing up IT personel to dedicate more time to the core enterprise needs.[36,879]

Another advantage of cloud computing is the reduced cost of running the enterprise IT department. For an existing enterprise, the reduced IT workforce as well as the reduced IT operational costs lead to an overall reduction in the operational expenditure. For a small or startup company, it can avoid massive capital expenditure for both infrastructure acquisition and its maintenance as well as the management of an IT department.[37,79]

One of the stated features of cloud computing is its elasticity. The fact that cloud resources can be scaled up or down automatically depending on demand is of a significant benefit to the user. Some enterprises experience periodic peak usage of the IT infrastructure while for others, it is not possible to know in advance what the demand will be like. In both scenarios, the scalability and elasticity of cloud computing means an enterprise can acquire and utilize cloud resources affordably and with relative ease.[38,6]

Related to scalability and elasticity, the possibility of rapid access to cloud resources is a benefit that can not be overlooked. This is as compared to on-site hardware installation and software setup and deployment which typically takes time.[38,6]

5.4.2 Challenges to Cloud Computing Adoption

In as much as the use of cloud computing presents a range of benefits, there is a number of barriers that hinder its adoption by individual users as well as organizations. Based on a number of surveys, security and the related privacy concerns are the

greatest hinderance among IT professionals to utilizating cloud services in their organizations. The security concerns transcend many aspects of security i.e. the network, the host computers, the applications, and data security. In terms of privacy, the uncertainity about who has control as well as access to data in the cloud is of major concern. This is further enhanced by legislations in different regions of the world with regard to the use and export if cryptography and also legislations restricting certain personal information from being stored outside of national bounderies. [38,34]

Cloutage.org which documents known and reported cloud incidents has a record of cloud outages affecting financial institutions, the entertainment industry, social networking services etc. Given the need for around the clock availability of certain enterprise systems and services by organizations, the personel in charge of IT departments are wary not to rush such systems to the cloud due to the uncertainity of its reliability. Since outages may occur due to various reasons such as system failures, accidents, malice or natural disasters, IT professionals demand a degree of guarantee in terms if availability, performance and recovery from severe failures before they can commit to the cloud. [38,36]

Cloud computing being such a relatively new field, the lack of its standardization means cloud applications and data cannot be moved easily across a clolud provider's platform, hence resulting in users getting locked into a single vendor. This is perhaps even in the existence of more affordable alternative CSPs and the cost implications of developing new applications and re-writing data to suit the alternative platform turning out to be an hinderance. To alleviate this, nonetheless, there have been efforts towards standardization with the launch of OpenStack and the release of Open Virtualization Format (OVF). OpenStack provides a ubiquitous open-source cloud computing platform both for public and private clouds [39]. OVF on the other hand is a standard for packaging and distributing software to be run on virtual machines and is portable across different platforms i.e hypervisor-neutral [40,5].

6 Cloud Computing and 4G LTE

6.1 Convergence of Cloud Computing and 4G LTE

The proliferation of mobile and handheld devices with a wireless Internet access coupled with cloud computing, which, as mentioned in section 5.1, may be accessed via high-speed wireless networks (e.g. 4G LTE), is adversely altering the Internet and computing environment. According to Gartner [41], the increase of affordable tablets and their growing capability is accelerating the shift from Personal Computers (PCs) to tablets as consumers' primary computing device and hence they are unlikely to replace their PCs regularly. This implies that users prefer the convenience of Internet services anytime and from anywhere. The same applies to company employees who require remote access to business applications as well as collaborative services.

As noted in the previous paragraph, there is a general trend towards mobile computing, whereas mobile devices continue to face a number of constraints such as storage capacity, processing power and battery life. Battery technologies have not experienced the same evolution as the rest of mobile handsets and consequently the plethora of applications that run on mobile devices quickly drains the battery charge. Besides the resource limitations of mobile devices, unreliable connectivity continues to be a hindrance to the full potential of the utilization of mobile computing.[42]

The combination of cloud computing and 4G LTE alleviates the challenges of mobile cloud computing (MCC). MCC benefits from the overall advantages of cloud computing – described in subsection (5.4) – and worth mentioning are; extending battery life by computation off-loading [42,191], access to more storage capacity, utilization of more powerful processing power and information reliability in case of loss of mobile device. In terms of connectivity, according to Benington S, LTE lends itself to cloud computing for the following reasons: it provides high bandwidth connections, it is IP and Ethernet oriented, it features low latency and ncreases capacity hence putting mobile connectivity at par with fixed broadband. [43]

6.2 Cloud Technological Solutions for Telecommunications Operators

Given the current trend of a shift towards cloud computing, the leading telecommunication equipment manufacturers have come up with innovations to align their products to this reality. This will subsequently ensure that telecommunication operators are well positioned in the cloud hierarchy to not only transmit data through their networks but also be players in the cloud ecosystem. Nokia Solutions and Networks (NSN) and Ericsson are leading the way with Liquid Applications and Ericsson Cloud Systems (ECS) respectively. They have persued different approaches, whereas they both aim to enable operators to participate in cloud service provision and also to bring cloud service closer to the user.

Liquid Applications is a technology in which a base station is used to do more that just send and receive radio signals by enabling it to perform other computational tasks. Liquids Applications which runs on Radio Applications Cloud Server (RACS) incorporated in to the base station aims to spur the creation a range of services that can be delivered from the base station. Such services include: instant video replays, video suveilance and a range of analytic services. It also also aims to accelerate content delivery by extending the capabilities of RACS for acceleration, optimization and storage of third party content. In addition, by enabling the operator to in real-time collect and process network data, the operator can better understand user behaviour and preference therefore tailor the network and services accordingly. [44; 45]

On the other hand, Ericsson's approach is to embrace Software-Defined Networks (SDN) and Network Functions Vertualization (NFV) which is enabled by the Ericsson Cloud System (ECS). This frees up network resources which can then be utilized to offer cloud computing and storage services leading to efficient utilization of network resources and better user experience of cloud based applications. The financial implication of this is htat the provider is able to save operational costs through efficiency and in addition has a new source of revenue through the value-added cloud services. [46]

In summary, Liquid Applications and ECS offers the provider a new avenue for revenue, ease of management of the network entities and cost saving opportunities. To the user, the quality of experience is enhanced through faster access to content due to proximity to content, contextualized services based on real-time data analysis etc.

Based on these two product examples, telecommunication operators have the opportunity to benefit from the cloud computing ecosystem by not only transmitting data but also offering cloud services.

6.3 Applications of Cloud Computing and 4G

Mobile learning, mobile commerce, mobile healthcare and entertainment are some of the areas in which the two technologies combined can be utilized. In the education sector for example, learners can have a fast access to large information from remote locations and for researchers, peer collaboration is enhanced. In the commercial sector, 4G-LTE and cloud computing solutions have been developed to suit both ordinary customers and business customers. An example is TeliaSonera providing Schenker with 4G and cloud services which include Cisco Managed Voice and TeliaSonera Remote User [47].

The generic 4G-LTE and cloud computing implementation is illustrated in figure 21.

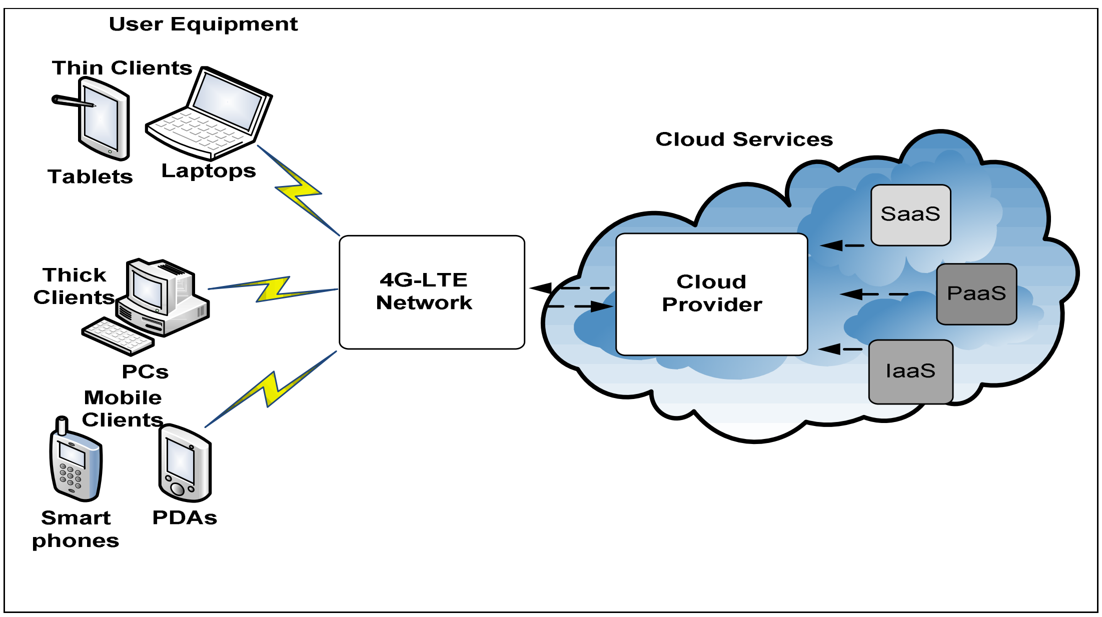

Figure 21. 4G-LTE Implementation

In order to access the cloud services, the user is equipped with an LTE capable device which is installed with appropriate applications and configured accordingly. In addition, the user must have a subscription to an LTE network provider who will in turn provide the Internet access to the cloud services.

As mentioned in section 4.3, M2M can be utilized in a number of ways. In addition to the aforementioned, it can be used to expand healthcare services through connected healthcare devices and appliciations. According to a report by Machina Research, compared to less than 100 million connected devices in 2013, there will be close to 847 million by 2023 in the healthcare sector. The report goes further to state that this will present healthcare organizations with the opportunity to expand the range of services, cut costs and improve the outcome of their services.[48]

Some of the applications of the M2M that are in place in the healthcare sector include Electrocardiography (ECG) units that transmit a patient's vital information to the closest hospital while the patient is in an ambulance still on transit. This aids with the preparation for the patient's arrival by medical personnel in the emergency room ensuring expeditious and correct treatment, hence possibly saving lives. Another example is an established telemedicine network of high-risk patients whose conditions are monitored remotely by health professionals. By use of M2M devices, patients send their medical data to telemedicine centers and therefore appropriate emergency actions can be taken if need arises. The M2M is not restricted to only the transmission of medical data but can be used also for communication in emergency situations. There are wearable devices which can be used to contact emergency centers during a medical emergency. These devices may in addition be equipped with tracking systems which ensures a prompt response to emergencies.[49]

By using the M2M technology, 4G and cloud computing, the benefits to the healthcare provider as well as the receiver include flexibility in the treatment of chronic diseases, healthcare extended to the underserved population, improved patient engagement with healthcare services, reduced risks and improved quality of life.

7 Conclusion

The aim of this thesis was to demonstrate the access of cloud computing resources over the the 4G LTE network. Cloud computing and 4G LTE are fairly new technologies and until recently they have been developing independently. In fact it is just recently that telecommunication equipment manufacturers are developing products which offer solutions for both 4G LTE and cloud computing.

The current consumers demand for information and services on their smart mobile devices, which also a demand for high capacity wireless data networks. It is for this reason that I explored in detail these two technologies and highlighted some applications that illustrate how they are complementary.

Ideally it should have been such that a project had been carried out to demonstrate this but as mentioned, the solutions that couple these technologies together are relatively new and scarce and therefore a challenge and prohibitive to acquire. In the future however, as solutions and the technology become more common, various projects may be carried out to demonstrate various applications.

References

1. Coleman C. An Introduction to Radio Frequency Engineering. Cambridge, UK: Cambridge University Press; 2004.

2. IMT-Advanced standards announced for next-generation mobile technology [Online]. Press Release, 21 October 2010. URL: http://www.itu.int/net/pressoffice/press_releases/2012/02.aspx. Accessed 28 July 2013.

3. Umar A. Mobile Computing and Wireless Communication. NGE solution; 2004.

4. Saily M, Sebire G. GSM/EDGE: Evolution and Performance. West Sussex, UK:Wiley; 2010.

5. Sauter M. From GSM to LTE – An Introduction to Mobile Networks and Mobile Broadband. Haboken, NJ: Wiley; 2010.

6. Halonen T. et al. GSM, GPRS and Edge Performance. West Sussex, UK: Wiley; 2003.

7. Eberspacher J. et al. GSM – Architecture, Protocols and Services. West Sussex, UK: Wiley; 2009.

8. Andersson C. GPRS and 3G Wireless Applications. New York, NY: Wiley; 2001.

9. Sauter M. Beyond 3G – Bringing Networks Terminals and the Web Together. West Sussex, UK: Wiley; 2009.

10. ITU Report. ITU-R M.2134 Requirements Related to Technical Performance for IMT-Advanced Radio Interface(s); 2008.

11. Dahlman E, et al. 4G LTE/LTE-Advanced for Mobile Broadband. Oxford, UK: Academic Press: 2011.

12. Olson M. et al. EPC and 4G Packet Networks. Oxford, UK: Wiley; 2013.

13. 3GPP TS 36.101 Evolved Universal Terestial Radio Access (E-UTRA); User Equipment (UE) radio transmission and reception.

14. Holma H, Toskala A. LTE for UMTS – OFDMA and SC-FDMA Based Radio Access. West Sussex, UK: Wiley; 2009.

15. 3GPP TS 36.300 LTE;Evolved Universal Terrestrial Radio Access (E-UTRA) and evolved Universal Terrestrial Radio Access Network (E-UTRAN);Overall description;Stage 2

16. Rumney M, LTE and the Evolution to 4G wireless. Agilent Technologies; 2009.

17. Yi S. et al. Radio Protocols for LTE and LTE-Advanced. Singapore: Wiley; 2012.

18. Cox, C. Introduction to LTE: LTE, LTE-Advanced, SAE and 4G Mobile Communications – 2nd Edition. Haboken, NJ: Wiley; 2012.

19. 3GPP TS 36.323 LTE;Evolved Universal Terrestrial Radio Access (E-UTRA); Packet Data Convergence Protocol (PDCP) specification.

20. 3GPP TS 36.322 LTE;Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Link Control (RLC) protocol specification.

21. 3GPP TS 36.321 LTE;Evolved Universal Terrestrial Radio Access (E-UTRA); Medium Access Control (MAC) protocol specification.

22. 3GPP TS 36.201 LTE;Evolved Universal Terrestrial Radio Access (E-UTRA); LTE physical layer; General description.

23. Shah D S. A Tutorial on LTE Evolved UTRAN (EUTRAN) and LTE Self Organizing Networks (SON). Arlington, TX: 2010

24. Ghosh A. et al. Fundamentals of LTE. Boston, MA: Prentice Hall; 2010.

25. 3GPP TS 23.401 LTE; General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestial Radio Access Network (E-UTRAN) access.

26. 3GPP TS 33.401 LTE; 3GPP System Architecture Evolution (SAE); Security architecture

27. Forsberg D. et al. LTE Security. West Sussex, UK: Wiley; 2013.

28. 3GPP TS 33.102 3G security; Security architecture.

29. 3GPP TS 33.210 LTE; 3G Security; Network Domain Security (NDS); IP network layer security.

30. Krutz R L, Vines, R D. Cloud Security. Indianapolis, IN: Wiley; 2010.

31. Liu Fang et al. NIST Cloud Computing Reference Architecture – Recommendations of the National Institute of Standards and Technology. Gaithersburg, MD: 2011.

32. Sosinsky B. Cloud Computing Bible. Haboken, NJ: Wiley 2010.

33. Youseff L. et al. Understanding the Cloud Computing Landscape. In: Ahson S.A, Ilyas M, Editors. Cloud Computing and Software Service, Boca Raton, FL: CRC Press; 2011.

34. Jia H. et al. Cloud Types and Services. In: Furht B, Escalnate A, Editors. Handbook of Cloud Computing, New York, NY: Springer; 2010. p.335-354.

35. Focus Group on Cloud Computing Technical Report: Functional Requirements and Reference Architecture; 2012.

36. Jadeja Y, Modi K. Cloud Computing – Concepts, Architecture and Challenges. IC-CEET; 2012.

37. Marks, E A, Lozano R R. Exercutive's Guide to Cloud Computing. Haboken, NJ: John Wiley; 2010.

38. Buyya R et al. Cloud Computing: Principles and Paradigms. Haboken, NJ: John Wiley;2010.

39. OpenStack: The 5-minute Overview. [Online].
URL: http://www.openstack.org. Accessed 29 March 2014

40. Open Virtualization White Paper. [Online] DTMF; 2013.
URL: http://dmtf.org/sites/default/files/standards/documents/DSP2017_2.0.0a.pdf
Accessed: 29 March 2014

41. Gartner Says Worldwide PC, Tablet and Mobile Phone Combined Shipments to Reach 2.4 Billion Units in 2013 [online].
URL: http://www.gartner.com/newsroom/id/2408515. Accessed 15 October 2013.

42. Vallina-Rodriguez N, Crowcroft J. Energy Management Techniques. In: IEEE Communications Survey & Tutorials, Volume 15, Issue: 1; First Quarter 2013.

43. Benington S, Mobile cloud services sustain revenue and growth in New LTE networks. [online]. Tellabs blogs, 28 August 2012.
URL: http://www.tellabs.com/blog/index.cfm/2012/8/28/Mobile-cloud-services-sustain-revenue-and-growth-in-new-LTE-networks. Accessed 20 October 2013.

44. Nokia Liquid Application. [Online]
URL: http://nsn.com/portfolio/liquid-net/intelligent-broadband-management/liquid-applications - liquid apps. Accessed 8 May 2014.

45. Increasing Mobile Operators' Value Proposition With Edge Computing. [Online]. Intel Technical Brief, 2013.

URL: http://www.intel.eu/content/dam/www/public/us/en/documents/technology-briefs/edge-computing-tech-brief.pdf. Accessed 8 May 2014.

46. Ericsson Cloud System. [Online]
URL: http://www.ericsson.com/ourportfolio/telecom-operators/cloud-system. Accessed 8 May 2014

47. TeliaSonera provides Schenker AB with 4G and cloud services [online]. 10 December 2010.
URL:http://www.teliasonera.com/en/newsroom/press-releases/2010/12/teliasonera-provides-schenker-ab-with-4g-and-cloud-services/. Accessed 21 October 2013.

48. Advancing M2M adoption in nthe healthcare industry is transforming medical and personal health devices, business models and patient engagement.[Online] Machina Research news. 07 May 2014.
URL: https://machinaresearch.com/news/press-release-advancing-m2m-adoption-in-the-healthcare-industry-is-transforming-medical-and-personal-health-devices-business-models-and-patient-engagement/ - m2m machine. Accessed 8 May 2014.

49. M2M improves healthcare.[Online] 2014
URL: http://www.telekom.com/media/media-kits/m2-communikation/215740 - m2m. Accessed 8 May 2014