# A TUTORIAL ON LTE EVOLVED UTRAN (EUTRAN) AND LTE SELF ORGANIZING NETWORKS (SON)

by

DHRUV SUNIL SHAH

Presented to the Faculty of the Graduate School of
The University of Texas at Arlington in Partial Fulfillment
of the Requirements
for the Degree of

MASTER OF SCIENCE IN ELECTRICAL ENGINEERING

THE UNIVERSITY OF TEXAS AT ARLINGTON

December 2010

# ACKNOWLEDGEMENTS

I would like to thank my guide, Dr. R. Stephen Gibbs for his continued support and belief in me. This thesis work would not have been possible without his guidance and support.

I am forever indebted to my parents and family for their love, support, and encouragement. Their belief in me during my entire life, especially during my graduate studies made this endeavor simple.

Finally, I would like to thank my friends at UTA, especially my colleagues in Radio Study Group for making this thesis a possibility.

November 19, 2010

# ABSTRACT

# A TUTORIAL ON LTE EVOLVED UTRAN (EUTRAN) AND LTE SELF ORGANIZING NETWORKS (SON)

Dhruv Shah, MS

The University of Texas at Arlington, 2010

Supervising Professor: Stephen Gibbs

The third-generation (3G) cellular communication technology, Universal Mobile Terrestrial System (UMTS), based on Wideband Code-Division Multiple Access (WCDMA), has been widely deployed all over the world providing faster download speeds for data (packet) communications. To further improve the throughput and overall performance of the cellular communication system established by UMTS, the Third Generation Partnership Project (3GPP) in November 2004 launched an ambitious project called the Long Term Evolution (LTE) of UMTS. This would ensure the continued competitiveness of the UMTS in the future. The technical specifications of the LTE project are formally known as the Evolved UMTS Terrestrial Radio Access (E-UTRA) and Evolved UMTS Terrestrial Radio Access Network (EUTRAN).

The main features of LTE are high peak data rate, flexibility of spectrum usage, low latency times, higher capacity per cell, etc. The radio interface of LTE is based on Orthogonal Frequency Division Multiple Access (OFDMA) in the downlink and Single Carrier-Frequency Division Multiple Access (SC-FDMA) in the uplink. LTE undergoes a major design change in its Core Network Architecture. The previously used separate cores for Voice and Data in 3G are

being replaced by a single packet based or an all-IP core in LTE. This evolution of the Core Network is commonly referred to as System Architecture Evolution (SAE).

As we move towards the end of the year 2010, several service providers are set to launch their LTE services in markets across USA. LTE services are already available in some markets in Europe, and the performance so far has been impressive. As a marketing gimmick, LTE is being said and launched as a 4G technology, which in reality is still a 3.9G technology. For LTE to be truly called a 4G technology, it has to undergo some fine improvisations in order to meet the requirements for 4G technology set forth by International Telecommunication Union (ITU). With this goal set in mind, the 3GPP Standards Committee is further developing the LTE standards to meet the requirements set for International Mobile Telecommunications-Advanced (IMT-A) technologies which would be called the Long Term Evolution-Advanced (LTE-A). The 3GPP specification in Release 10 is going to be an LTE-A/IMT-A compatible release. LTE-A will have new requirements and new features for the system, for instance, new Self-Organizing Network (SON) is one of the features.

In this research activity, I have made an attempt to completely explore the Network Architecture in Long Term Evolution, main point of focus being the Evolved UTRAN and the eNodeB (eNB) in LTE. Then we focus on the Self Organizing Networks (SON) concept being implemented in LTE and Handover Optimization Techniques. The LTE SON concept aims at minimizing the human involvement in network maintenance and operation.

# TABLE OF CONTENTS

APPENDIX

## LIST OF ILLUSTRATIONS

Figure Page

# LIST OF TABLES

Table Page

# LIST OF ACRONYMS

3GPP: 3rd Generation Partnership Project.

AES: Advanced Encryption Standard

AN: Access Network

ANR: Automatic Neighbor Relation

AS: Access Stratum

AuC: Authentication Center

BBERF: Bearer Binding and Error Reporting Function

BCCH: Broadcast Control Channel

BCH: Broadcast Channel

BSC: Base Station Center

BTS: Base Transceiver Sub-system

CAPEX: Capital Expenditures

CCCH: Common Control Channel

CDMA: Code Division Multiple Access.

CN: Core Network

CP: Control Plane

CRC: Cyclic Redundancy Check

CS: Circuit Switch

DCCH1: Dedicated Control Channel 1

DCCH2: Dedicated Control Channel 2

DHCP: Dynamic Host Configuration Protocol

DL-SCH: Downlink Shared Channel

DRX: Discontinuous Reception

DTCH1: Dedicated Traffic Channel 1

DTCH2: Dedicated Traffic Channel 2

EDGE: Enhanced Data Rates for GSM Evolution.

eNB: Evolved/Enhanced NodeB

EPC: Evolved Packet Core

E-UTRAN: Enhanced Universal Terrestrial Radio Access Network.

EVDO: Evolution Data Only

GBR: Guaranteed Bit Rate

GGSN: Gateway GPRS Serving Node

GPRS: General Packet Radio Service.

GSM: Global System for Mobile communications.

GTP: GPRS Tunneling Protocol

GUTI: Globally Unique Temporary Identifier

HARQ: Hybrid Automatic Repeat Request

HeNB: Home eNodeB

HLR: Home Location Register

HO: Handover

HPI: Handover Performance Indicators

HSDPA: High Speed Downlink Packet Access.

HSPA: High Speed Packet Access.

HSPA+: Evolved High Speed Packet Access.

HSS: Home Subscription Server

HSUPA: High Speed Uplink Packet Access.

IMS: IP Multimedia Sub-system

IMSI: International Mobile Subscriber Identity

LB: Load Balancing

LTE: Long Term Evolution (for UMTS).

MAC: Medium Access Control

MIMO: Multiple Input Multiple Output

MME: Mobility Management Entity

MRO: Mobility Robustness Optimization

MSC: Mobile Switching Centre

NAS: Non-Access Stratum

NGMN: Next Generation Mobile Networks

OAM: Operations, Administration, and Maintenance

OFDMA: Orthogonal Frequency Division Multiple Access

OPEX: Operational Expenditures

PCCH: Physical Control Channel

PCH: Physical Channel

PCRF: Policy and Charging Resource Function

PDA: Personal Digital Assistant

PDCP: Packet Data Convergence Protocol

PDN: Packet Data Network

P-GW: Packet Data Network Gateway

PHY: Physical Layer

PMIP: Proxy Mobile Internet Protocol

PS: Packet Switched

QCI: QoS Class Identifier.

QoS: Quality of Service.

RACH: Random Access Channel

RAT: Radio Access Technology

RNC: Radio Network Controller

RNS: Radio Network Sub-systems

RoHC: Robust Header Compression

RRM: Radio Resource Management

RSRP: Reference Signal Received Power

RTT: Round Trip Time.

SAE: System Architecture Evolution

SC-FDMA: Single Carrier Frequency Division Multiple Access

SCTP: Stream Controlled Transmission Protocol

SGSN: Serving GPRS Serving Node

S-GW: Serving Gateway

SIP: Session Initiation Protocol

SON: Self Organizing Network

TA: Tracking Area

TE: Terminal Equipment

TFT: Traffic Flow Template

UE: User Equipment

UICC: Universal Integrated Circuit Card

UL-SCH: University Shared Channel

UMTS: Universal Mobile Telecommunication System.

UP: User Plane

USIM: Universal Subscriber Identity Module

UTRAN: UMTS Terrestrial Radio Access Network

VLR: Visitor Location Register

VoIP: Voice over IP

WiMAX: Worldwide Interoperability for Microwave Access

# CHAPTER 1

# LONG TERM EVOLUTION – OVERVIEW AND GOALS

## 1.1 Evolution to LTE

The 2nd Generation Mobile Communication System, commonly referred to as the GSM, was mainly about voice services. Circuit Switching (CS) alone was capable to handle the requirements of the core network. The core network looked pretty simple, its main components being Mobile Switching Centre (MSC), Home Location Register (HLR), Visitor Location Register (VLR), Authentication Center (AuC), Base Station Center (BSC) and Base Transceiver Sub-system (BTS).

As we moved towards 2.5G system, usually referred to as the GPRS, packet data services were made available by introduction of a separate Packet Switched (PS) Network on top of the initial GSM architecture. The GPRS architecture consists of an additional Serving GPRS Serving Node (SGSN) and Gateway GPRS Serving Node (GGSN), which handle data services. As the demand for amount of data & data throughput rate increased, improvements were made to the GPRS system by improving on both the downlink and uplink data rates. It was commonly referred as EDGE which was termed as a 2.75G technology.

In 3GPP R99, UMTS was introduced based on the CDMA technology. UMTS is classified as a 3G technology. Though there was a separate Circuit Switched and Packet Switched domains, significant improvements in data rate were achieved by employing different modulation schemes in downlink and uplink. The UMTS Network was quite different than its predecessors. The Radio Access Network, now called as UMTS Terrestrial Radio Access

Network (UTRAN) was divided in Radio Network Sub-systems (RNS) which consisted of a Radio Network Controller (RNC) and several base stations, called as NodeB.

As the data requirements surmounted, an effort was made to improve the downlink and uplink throughput rates by employing higher modulation techniques. In subsequent releases, 3GPP announced HSDPA, HSUPA, HSPA and HSPA+. The Release 8 of 3GPP was a completely new standard, called Long Term Evolution (LTE), which was based on the System Architecture Evolution (SAE) and Evolved Packet Core (EPC) was implemented in LTE which is completely IP based. Since there are provisions in LTE for inter-operation with legacy systems, there are various paths available to evolve to LTE. An operator with a GPRS/EDGE network can directly evolve to a LTE network. An evolution path to LTE from HSPA/HSPA+ is also available. Non-3GPP systems like CDMA/EVDO can also evolve to LTE. Due to this increased flexibility, LTE is the choice of majority of operators worldwide. Figure 1.1 illustrates the evolution paths to LTE.

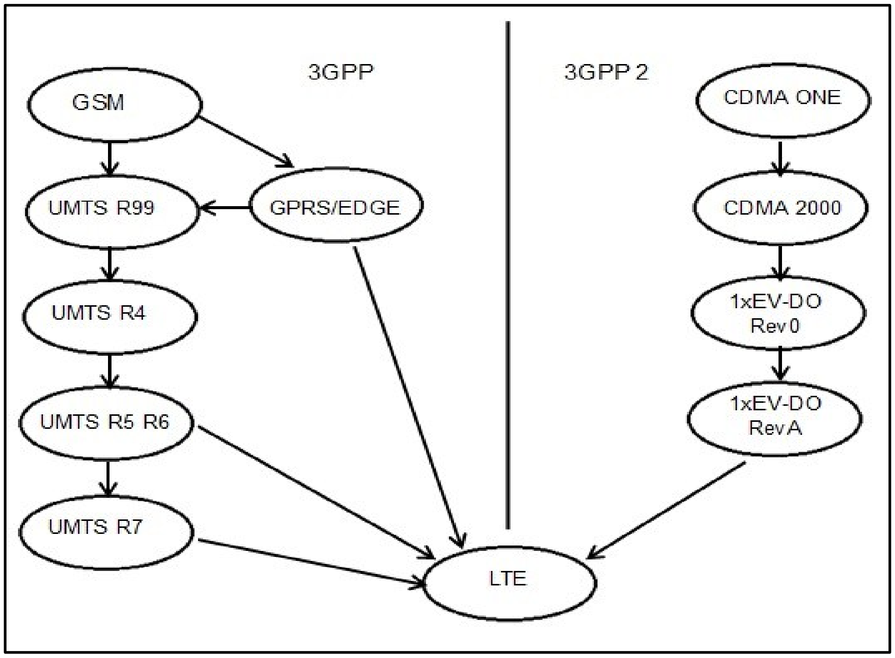


Figure 1.1: Evolution Paths to Long Term Evolution (LTE)

## 1.2 Long Term Evolution (LTE) Goals

As the wireline technologies keep improving over time, so does the wireless technologies. The main driving force and the motivation behind the developing the LTE Standards were need for additional wireless capacity, lower cost wireless data transfer, etc. Following are some of the architectural requirements established by 3GPP Standards Body for LTE:

1. An All-IP based system
2. Flat Architecture for Optimized Payload Path
3. Excellent scalability
4. High level of security in Access Network (AN) as well as Core Network (CN)
5. Simple QoS model
6. Low delay times between nodes
7. Efficient radio usage
8. Flexible spectrum utilization
9. Cost efficient deployment

Many of the above mentioned targets were achieved by implementing a Flat Architecture with less number of nodes. The reduced number of nodes helped in reducing latency times and improved overall performance.

The development towards flat architecture had started in the 3GPP UMTS Release 7 (HSPA+), wherein the Direct Tunnel concept allowed User Plane (UP) to bypass SGSN and directly connect to the GGSN. This feature also helped move RNC functions to NodeB. The figure 1.2 shows the 3GPP architecture evolution towards Flat Architecture.

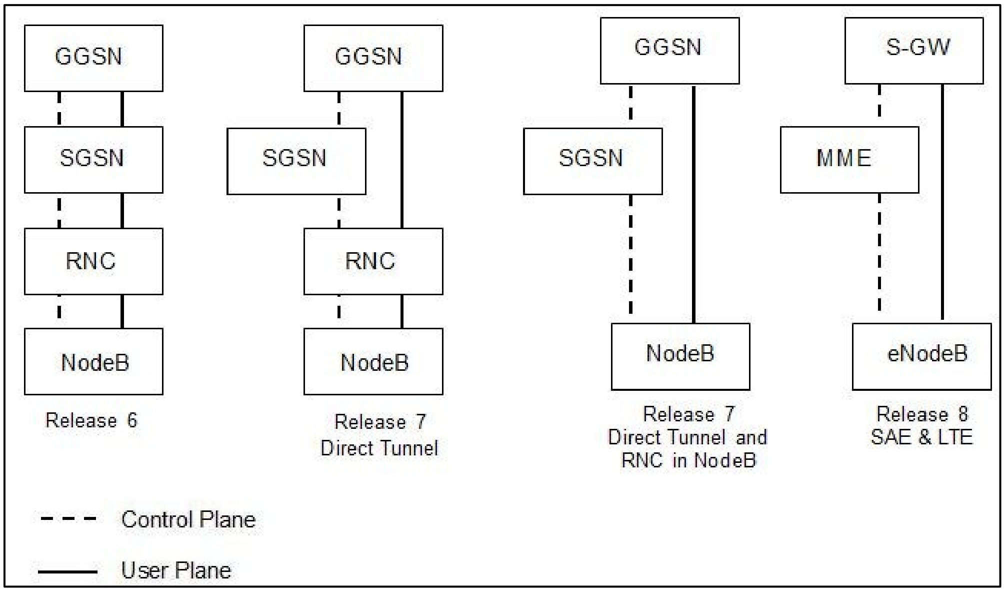


Figure 1.2: 3GPP's Architecture Evolution towards Flat Architecture [12]

LTE was the Release 8 of 3GPP standards, wherein several specifications were laid out regarding the performance and QoS of the new standards. These specifications are listed below:

1. LTE offers a peak downlink data rate of ~100 Mbps ($\eta$ = 5 bits/sec/Hz) and uplink data rate of ~50 Mbps ($\eta$ = 2.5 bits/sec/Hz); for 20 MHz spectrum. These peak data rates can be further increased by employing multiple antenna techniques and techniques like "Beamforming".
2. LTE offers very flexible spectrum usability, from 1.4 MHz to 20 MHz in both Uplink and Downlink.
3. One of the advantages of LTE is its "Flat Architecture". The user plane has just two (2) nodes, which considerably reduces the latency times in User and Control Plane. In LTE User plane, RTT < 5ms; while in the Control Plane, Idle to Active latency < 100ms and Standby to Active latency < 50ms.

4. LTE is 10 times more efficient to transport voice calls than the current GSM deployment. It has the capacity to handle approximately 200 users / cell for 5MHz spectrum.
5. LTE is optimized for high mobility. It provides optimized support for users at 0 – 15kmph, provides high performance for 15 – 120kmph and will support users travelling at speeds above 350kmph and 500kmph in certain frequency bands.
6. LTE provides a high level security in both the Access Network (AN) as well as the Core Network (CN).
7. LTE employs a novel mechanism called *“S1 flex”*. This will provide an efficient utilization of the radio spectrum.
8. LTE seeks to greatly lower the Capital Expenditure (CAPEX) and the Operational Expenditure (OPEX) by using *Self Organizing Networks* (SON). The concepts of LTE SON are explained in detail in Chapter 4.

# CHAPTER 2

# THE LTE NETWORK ARCHITECTURE

## 2.1 Evolved Packet System (EPS) / System Architecture Evolution (SAE)

As with most standardization efforts within 3GPP, the SAE project was developed upon the existing standards; in this case, it was based on the already established 3GPP packet core system standards being used in GSM/GPRS & WCDMA/HSPA. For some operators and vendors, continuity with the existing 3GPP standards was very important as it would benefit them, from the already established packet core networks they had developed, deployed and put in use. For others, they wanted to explore new arenas based on technology developed by other standardization agencies. There were some operators who wished to create an architecture that was not based on the existing 3GPP standards. All these different viewpoints and angles were taken into consideration while working on SAE. As reported in the initial documentation of the 3GPP Technical report 23.882 [23.882], various options were considered and one common goal was established, requirement for a separate control and user plane entities.

EPS (Evolved Packet System) represents the latest stage of evolution of the UMTS standard. Although still a 3G family standard, EPS provides significant improvements, with new radio interface in downlink (OFDMA) & uplink (SC-FDMA) and an evolved architecture for the Core and Access Networks. The two main advantages offered by EPS are:

1. Improved performance in Radio Interfaces characterized by a spectrum efficiency which is twice as large as HSPA Standards.
2. An “all-IP” architecture resulting in a unified and simplified system.

The Evolved Packet System is divided into two main sub-systems:

1. Evolved Packet Core (EPC) – combines the core network entities like MME (Mobility Management Entity), S-GW (Serving Gateway), P-GW (Packet Data Network Gateway), PCRF (Policy and Charging Resource Function) and HSS (Home Subscription Server).
2. Evolved UMTS Terrestrial Radio Access Network (E-UTRAN) – combines multiple eNodeBs.

EPS has provisions for interoperability of LTE with legacy cellular networks (GSM/GPRS/UMTS) and Non-3GPP standards (CDMA/EVDO/WiMAX). This meant that EPC provides a means to bridge different packet core networks. Figure 2.1 illustrates a complete architecture diagram, including inter-operational links between LTE and other systems. While the logical architecture in figure 2.1 may seem quite complex, the EPS consists a few new entities with a large number of new interfaces used to inter-connect the systems together.

Despite the intricate details mentioned in the figure 2.1, the “all-IP” infrastructure of LTE is not clearly visible. This feature is basically contained in the underlying transport network used to run the IP Networks. Most of the interfaces shown in the figure are logical, meaning not all maybe implemented as a separate interface on a physical link. Most of them are software based and may or may not be separate physical interfaces. Also the physical implementation of a particular link may not be directly connected between two nodes. They may pass through different paths or routed through different nodes.

In the next part of this chapter, we focus on the details of EUTRAN and EPC entities.

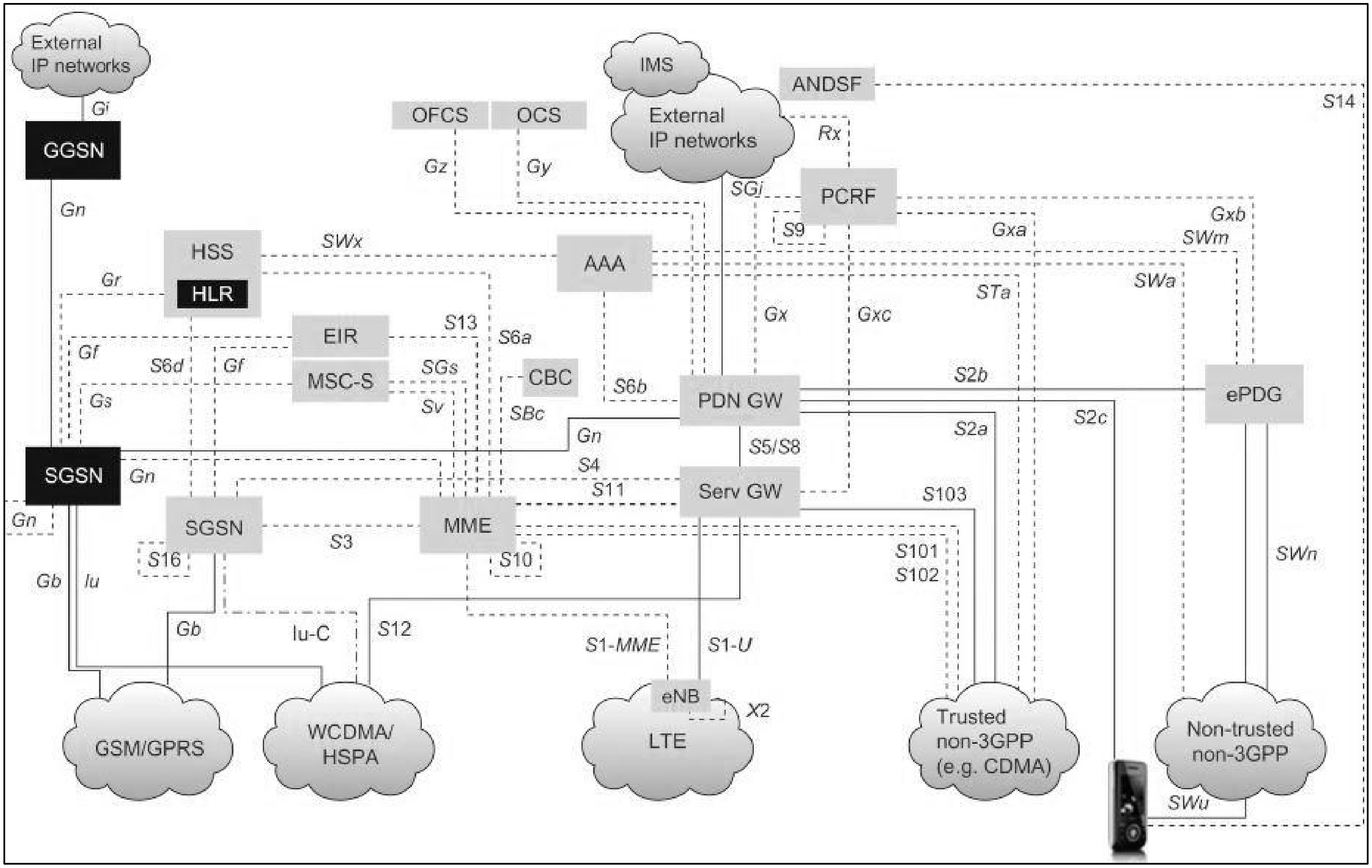


Figure 2.1: Architectural Overview of EPS and other systems [14]

## 2.2 LTE Network Architecture

The figure 2.2 describes the LTE Network Architecture with its basic system configuration and logical nodes. These elements are needed when E-UTRAN is involved in the Access Network.

The architecture of sub-divided in four main sub-systems:

1. Evolved Packet Core (EPC)
2. Evolved Universal Terrestrial Radio Access Network (E-UTRAN)
3. User Equipment (UE)
4. Services Domain

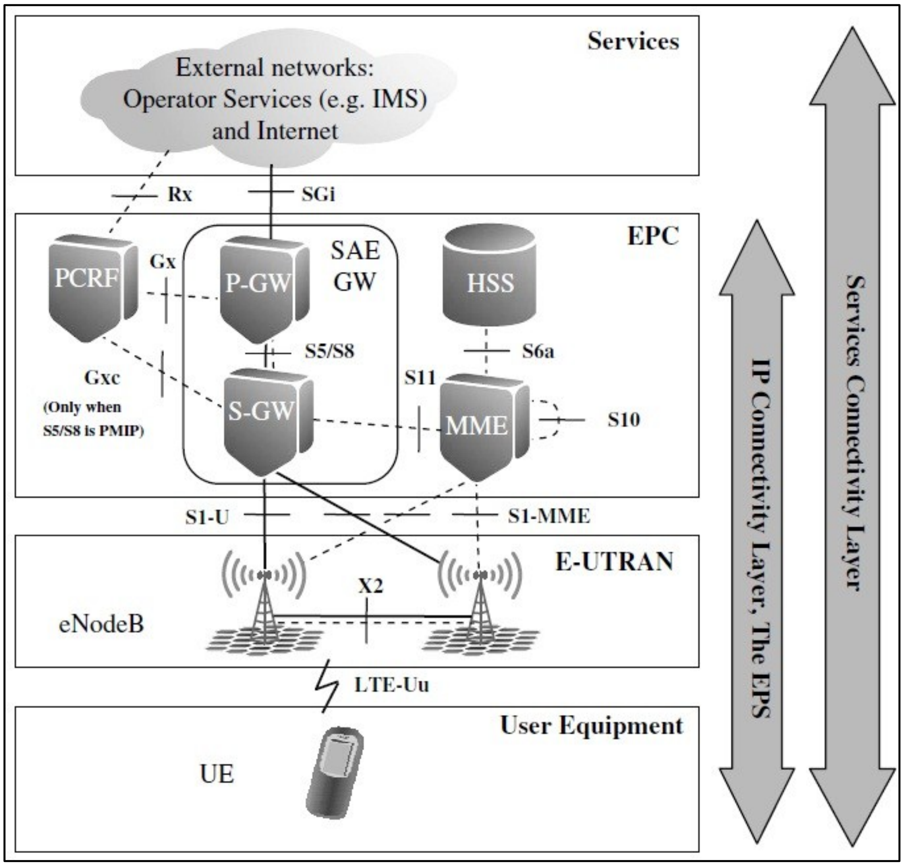


Figure 2.2: LTE Network Architecture [2]

The UE, EUTRAN and EPC combined represent the Evolved Packet System and the IP Connectivity Layer. This layer is optimized for IP connectivity and offers all transport services on top of IP. The IMS (IP Multimedia Subsystem) is an example of the Services offered in LTE.

### *2.2.1 Evolved Packet Core (EPC)*

The Evolved Packet Core in LTE is the equivalent of Core Network in GSM/UMTS System. It contains all the functional core network entities like MME, S-GW, P-GW, PCRF and HSS. It is responsible for voice call management, data call management, billing and other such functions.

#### 2.2.1.1 Mobility Management Entity (MME) [12]

Mobility Management Entity (MME) is the heart of the EPC. MME is in charge of all the Control Plane functions related to subscriber and session management. MME provides a logical direct CP path to the UE which is used as the primary control channel between the network and the UE. Figure 2.3 illustrates the above mentioned.

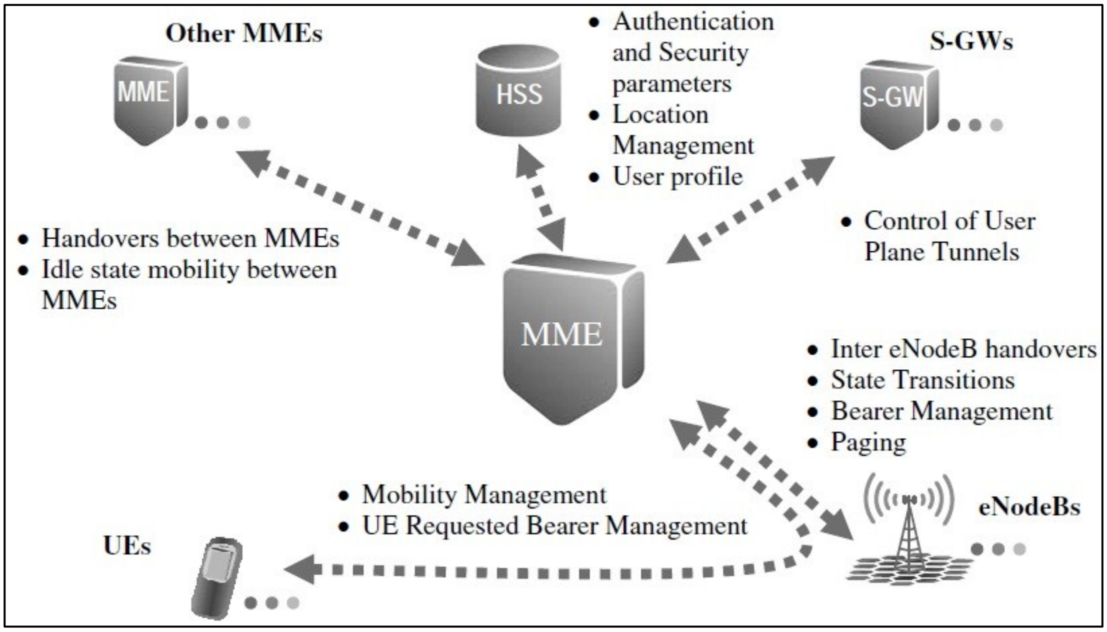


Figure 2.3: MME connections to other logical nodes and functions [12]

MME is the entity in the network responsible for authenticating and allocating resources to the UE when it first connects to the network. To provide additional security to the UE, MME assigns each UE a temporary identity called the "Globally Unique Temporary Identity (GUTI)", which eliminates the need to send IMSI of the UE over radio channels. The GUTI may be periodically refreshed and changed to prevent unauthorized tracking of the UE.

The MME tracks all UEs present in its service area. The MME will keep tracking the UE's location either on an eNB level in case the UE is connected, or at a Tracking Area (TA) level in case the UE is in idle mode.

The MME is also responsible for setting up of resources for the UE. MME does this by retrieving the user profile from HSS and determine what Packet Data Network connections should be allocated to the UE at initial 'attach' point. MME automatically sets up the default bearer, thereby giving UE the basic IP connectivity including CP signaling with the eNB and the S-GW. MME is also involved in setting up the dedicated bearers for the users.

The MME also participates in control signaling for handover of an active mode UE between eNBs, S-GW‟s or MME‟s. MME is involved in every eNB change, since there is no separate RNC to hide most of these events. In principle the MME may be connected to any other MME in the system. Connectivity to a number of HSSs will also need to be supported. The MME may serve a number of UEs at the same time.

#### 2.2.1.2 Serving Gateway (S-GW) [12]

The S-GW is involved mainly in the User Plane (UP) tunnel management, switching and other operations. It is not involved in the Control Plane (CP) operations. S-GW can only handle

its own resources and it allocates them based on requests from MME, P-GW or PCRF. An illustration describing S-GW logical interfaces and primary functions is shown in figure 2.4.
S-GW can use either GTP tunnels or PMIP tunnels for data flow depending on the data bearer setup. S-GW acts as a local mobility anchor during handovers between eNBs. It can monitor data inside the tunnels for Lawful Interception and Charging purposes.

All S-GW connections are “one-to-many”. One S-GW may be serving only a particular geographical area with a limited set of eNBs, and there may be a limited set of MMEs that control that area.

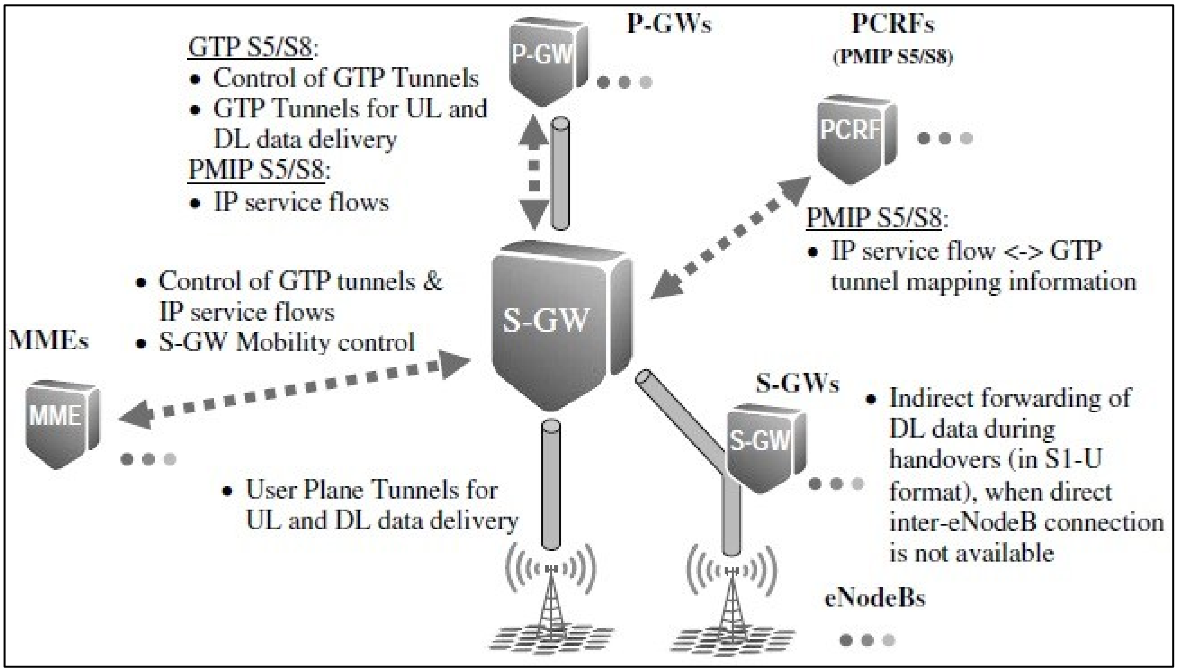


Figure 2.4: S-GW main logical connections and functions [12]

The S-GW should be able to connect to any P-GW in the whole network as the P-GW will not change during mobility, while the S-GW may be relocated. For connections related to one UE, the S-GW will always signal with only one MME and the UP points to one eNB at a time. If one UE is allowed to connect to multiple PDN’s through different P-GW‟s, then the S-GW needs to connect to those separately.

#### 2.2.1.3 Packet Data Network Gateway (P-GW)[12]

PDN-GW (also often abbreviated as P-GW) is the edge router between the EPS and external packet data networks. It acts as the highest level mobility anchor in the EPS and as the IP point of attachment for the UE. It performs traffic gating and filtering functions as required by the service in question. Typically, P-GW assigns an IP address to the UE which it uses for communication with external network. The P-GW performs the required Dynamic Host Configuration Protocol (DHCP) functionality.

P-GW is the highest level mobility anchor in the system. When a UE moves from one S-GW to another, the bearers have to be switched in the P-GW. The P-GW will receive an indication to switch the flows from the new S-GW.

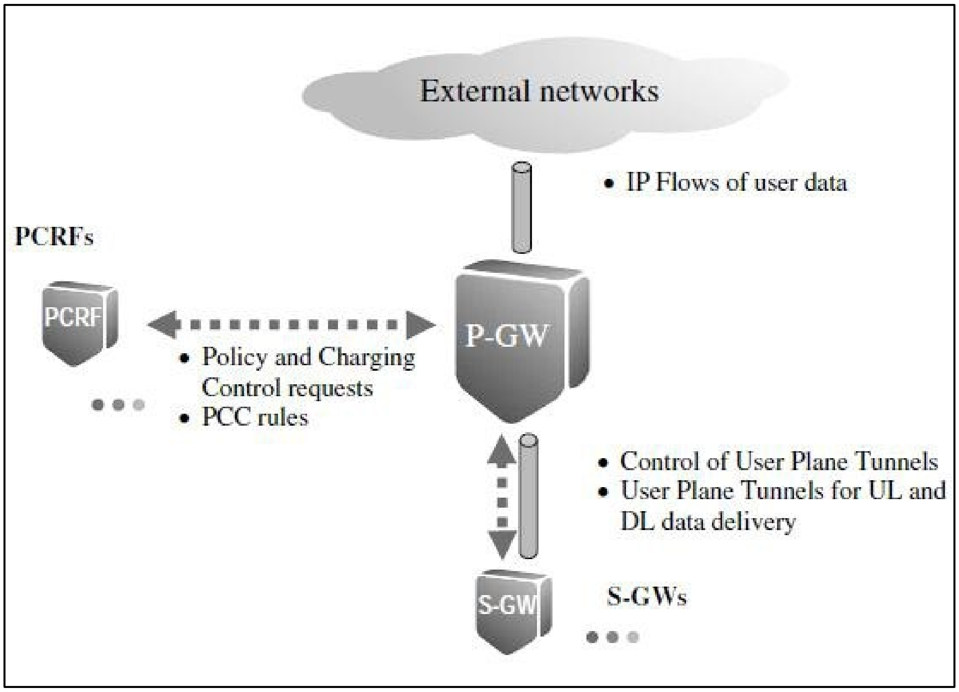


Figure 2.5: Packet Data Network Gateway logical connections and functions [12]

Each P-GW may be connected to one or more PCRF, S-GW and external network. For a given UE that is associated with the P-GW, there is only one S-GW, but connections to many external networks and respectively to many PCRFs may need to be supported, if connectivity to multiple PDNs is supported through one P-GW.

#### 2.2.1.4 Policy and Charging Resource Function (PCRF) [12]

PCRF is the network element that is responsible for Policy and Charging Control (PCC). It makes decisions on how to handle the services in terms of QoS, and provides information to the PCEF located in the P-GW, and if applicable also to the BBERF located in the S-GW, so that appropriate bearers and policing can be set up. The EPC bearers are then set up based on those. The connections between the PCRF and the other nodes are shown in Figure 2.6. Each PCRF may be associated with one or more AF, P-GW and S-GW. There is only one PCRF associated with each PDN connection that a single UE has.

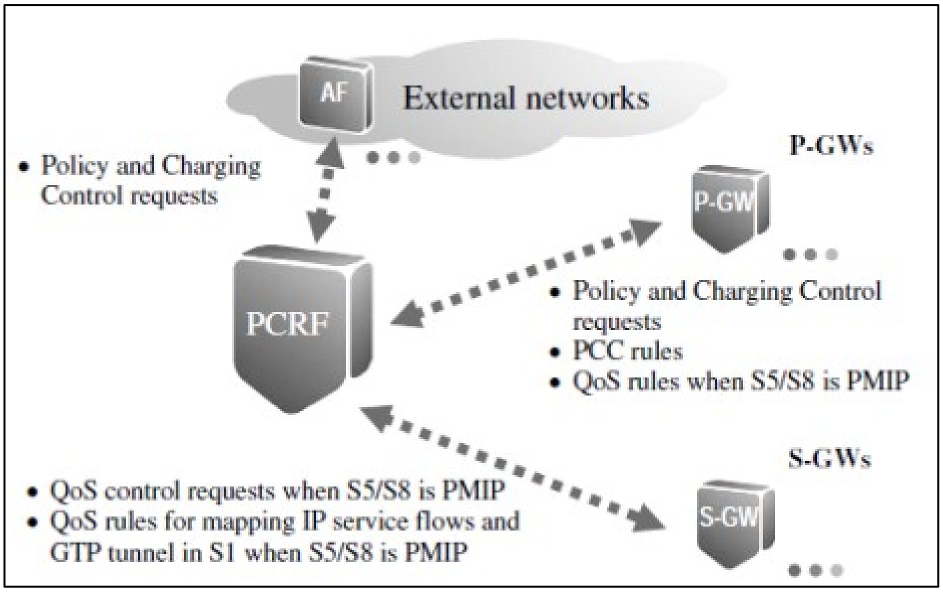


Figure 2.6: PCRF logical connections and main functions [12]

#### 2.2.1.5 Home Subscription Server (HSS) [12]

Home Subscription Server (HSS) is the subscription data repository for all permanent user data. The HSS stores the master copy of the subscriber profile, which contains information about the services that are applicable to the user. It also records the location of the user in the level of visited network control node, such as MME. An overview of the HSS structure is shown in figure 2.7.

For supporting mobility between non-3GPP ANs, the HSS also stores the Identities of those P-GWs that are in use. The permanent key, which is used to calculate the authentication vectors that are sent to a visited network for user authentication and deriving subsequent keys for encryption and integrity protection, is stored in the Authentication Center (AuC), which is typically part of the HSS. In all signaling related to these functions, the HSS interacts with the MME and the HSS will need to be able to connect with every MME in the whole network.

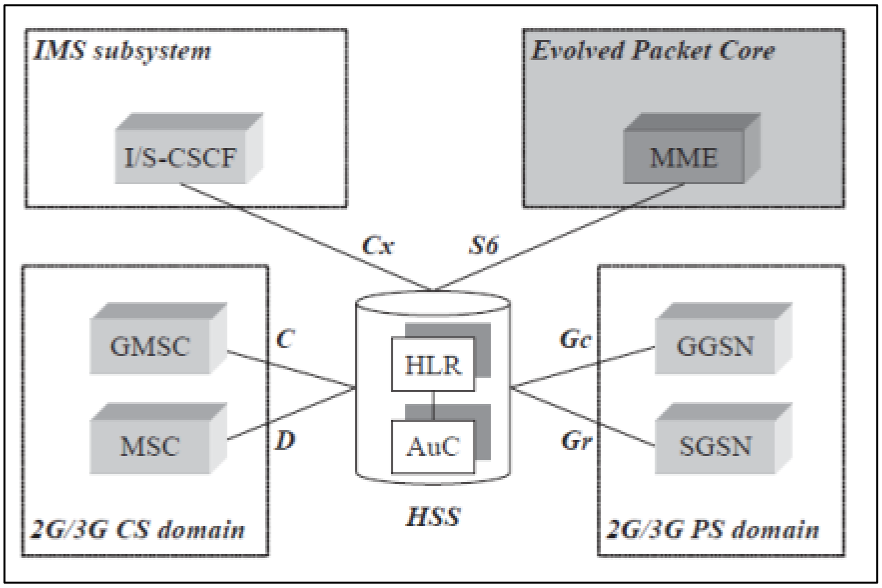


Figure 2.7: HSS structure and external interfaces [14]

### *2.2.2 Evolved UTRAN (E-UTRAN)*

Evolved Node B (eNodeB) is the only physical node present in EUTRAN. The eNodeB in LTE SAE is the equivalent of a radio base station in GSM/GPRS, but with added functionalities. It is in control of all radio related functions in the fixed part of the system. The eNodeBs are typically distributed throughout the coverage region.

Functionally, eNodeB provides bridging between UE and the EPC. It acts as a layer 2 'bridge' in the EPC system. The eNodeB is also a termination point of all the radio protocols towards the UE and acts as data relay between the radio connection and the corresponding IP based connectivity towards the EPC. The eNodeB also performs the functions like ciphering/deciphering of the User Plane data, IP header compression/decompression. Figure 2.8 shows the logical connections of an eNodeB and its major functions.

The eNodeB is responsible for many CP functions. The main features an eNodeB supports are the following:

- Radio Bearer management – this includes Radio Bearer setup & release procedures and also involves RRM functionalities for initial admission control and bearer allocation. This set of functions is controlled by the MME through the S1 interface during session setup, release and modification phases.
- Radio interface transmission and reception – this includes radio channel modulation/demodulation as well as channel coding/decoding.
- Uplink and Downlink Dynamic RRM and data packet scheduling – this is probably the most critical function which requires the eNodeB to cope with many different constraints (like radio-link quality, user priority and requested Quality of Service) so as to be able to multiplex different data flows over the radio interface and make use of available resources in the most efficient way.

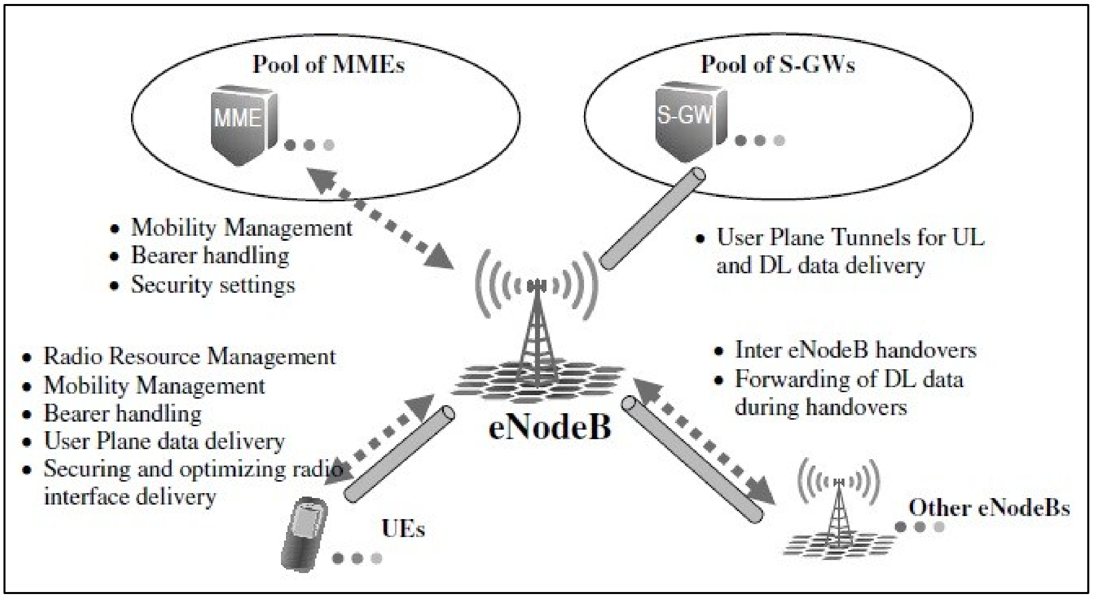


Figure 2.8: eNodeB connections to other logical nodes and main functions [12]

- Mobility management – this function relates to terminal mobility handling while the terminal is in an active state. This function implies radio measurement configuration and processing as well as the handover algorithms for mobility decision and target cell determination. Radio Mobility has to be distinguished from Mobility Management in Idle, which is a feature handled by the Packet Core.
- User data IP header compression and encryption – this item is the key to radio interface data transmission. It answers to the requirements to maintain privacy over the radio interface and transmit IP packets in the most efficient way.
- Network signaling security – because of the sensitivity of signaling messages exchanged between the eNodeB itself and the terminal, or between the MME and the terminal, all this set of information is protected against eavesdropping and alteration.

The eNodeB may be serving multiple UEs at its coverage area, but each UE is connected to only one eNodeB at a time. The eNodeB will need to be connected to those of its neighboring eNodeBs with which a handover may need to be made.

Both MMEs and S-GWs can be pooled together and therefore a set of these nodes is assigned to serve a particular set of eNodeBs. From the perspective of a single eNodeB, this means that the eNodeB may have the need to connect to multiple MMEs and S-GWs. However, each UE will be served by only one MME and S-GW at a time, and the eNodeB has to keep track of this association. This association will never change from a single eNodeB point of view, because MME or S-GW can only change in association with inter-eNodeB handover.

### *2.2.3 User Equipment (UE)*

UE is generally the handheld device that the end users use for their communication needs. Types of devices are explained in in detail below. The main point to be noted is that all the UEs contain the Universal Subscriber Identity Module (USIM) which is a separate module from the rest of the UE, which is often called the Terminal Equipment (TE). USIM is basically an application programmed onto a removable smart card called the Universal Integrated Circuit Card (UICC). USIM is used to identify and authenticate the user and to derive security keys for protecting the radio interface transmission. Functionally the UE is a platform for communication applications, which signal with the network for setting up, maintaining and removing the communication links the end user needs. This includes mobility management functions such as handovers and reporting the terminals location, and in these the UE performs as instructed by the network. Maybe most importantly, the UE provides the user interface to the end user so that applications such as a VoIP client can be used to set up a voice call.

We will now look into various forms of UE's

1) Smartphones: Smartphones encompass the functionality of a regular phone along with features of PDA's. They are rich in multimedia capabilities and provide for various connectivity options like WiFi, Bluetooth, etc.

2) PDA: Personal Digital Assistant, commonly known as PDA's traditionally didn't have any operating system or network connectivity. But over the years, the competitive market gave rise to much more sophisticated devices and PDA's evolved with provisions for WiFi connectivity, smartcard support for data transfers, etc.

3) Internet Tablets / Netbooks / Ultra mobile PCs: The latest craze in the portable devices market is the introduction of Netbooks or UMPCs. They are optimized for internet connectivity. They have provisions for network connectivity via 3G and WiFi/WiMAX chips.

In the 3GPP TS 36.306 specifications, the E-UTRA UE categories have been defined which is shown in the table 3.1

Table 3.1: LTE UE Categories

| **UE Classes** | **Peak Data Rate (Mbps)** | | **Soft Buffer Size (Gbits)** | **Number of MIMO streams** | **Max. DL modulation** | **Max. UL modulation** |
|---|---|---|---|---|---|---|
| | **DL** | **UL** | | | | |
| 1 | 10 | 5 | 0.25 | 1 | 64 QAM | 16 QAM |
| 2 | 50 | 25 | 1.24 | 2 | | |
| 3 | 100 | 50 | 1.24 | 2 | | |
| 4 | 150 | 50 | 1.83 | 2 | | |
| 5 | 300 | 75 | 3.67 | 4 | | 64 QAM |

There are five different UE categories, often referred as UE Classes, which are defined for LTE. But the important point to be noted is that LTE uses far less UE Classes than HSPA. As can be seen in the table above, the UE Class 1 device does not support MIMO functionality but the UE's from Class 2 - 4 will support 2x2 MIMO, whereas UE Class 5 can support 4x4 MIMO. It is important to note that regardless of whatever category a UE belongs to, it has to be capable of receiving transmissions from up to four antenna ports, as the base stations (eNBs) in LTE will have smart antennas (MIMO capabilities). Also, regardless of the UE class, all UE's should have a frontend of 20 MHz in order to receive their allocation anywhere the eNB would like to schedule it.

### *2.2.4 Services Domain*

The Services domain is not a fixed entity in the EPC like the other entities. It may include various sub-systems, which in turn may contain several logical nodes. The following is a possibility of various services provided to LTE EPC

- IMS based operator services: The IP Multimedia Sub-system (IMS) is service machinery that the operator may use to provide services using the Session Initiation Protocol (SIP). IMS has 3GPP defined architecture of its own. It is illustrated in figure 2.9.
- Non-IMS based operator services: The architecture for non-IMS based operator services is not defined in the standards. The operator may simply place a server into their network, and the UEs connect to that via some agreed protocol that is supported by an application in the UE. A video streaming service provided from a streaming server is one such example.
- Other services not provided by the mobile network operator, e.g. services provided through the internet: This architecture is not addressed by the 3GPP standards, and the architecture depends on the service in question. The typical configuration would be that the UE connects

to a server in the internet, e.g. to a web-server for web browsing services, or to a SIP server for internet telephony service (i.e. VoIP).

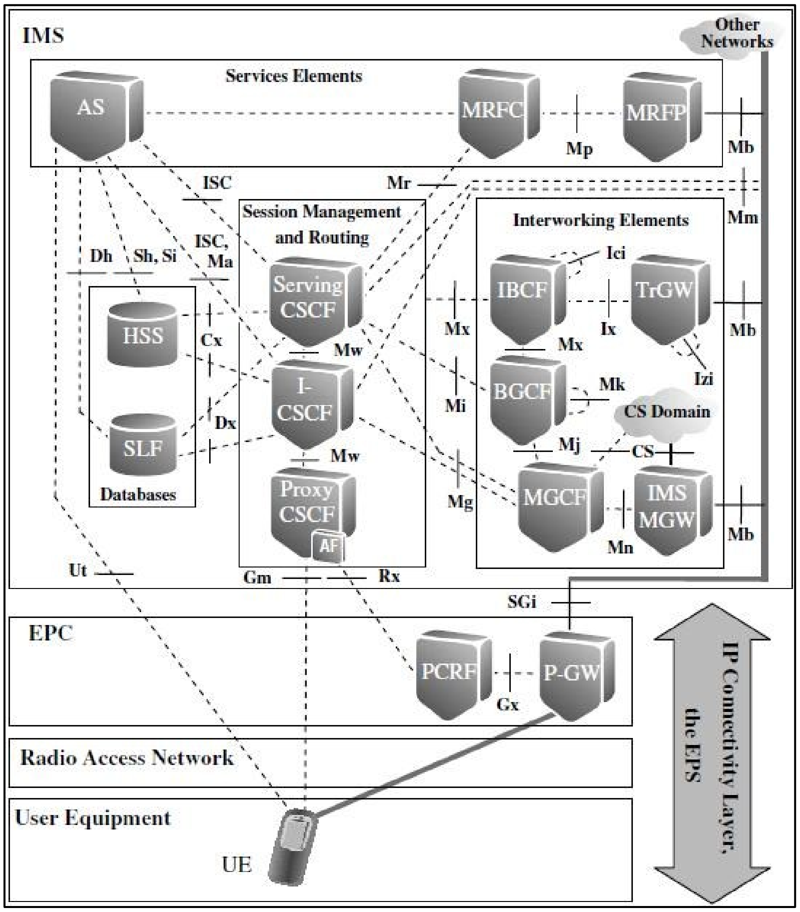


Figure 2.9: Services Domain [12]

## CHAPTER 3

## E-UTRAN & E-UTRAN NETWORK INTERFACES

### 3.1 E-UTRAN (Evolved UMTS Terrestrial Radio Access Network)

The E-UTRAN consists of eNBs, providing the E-UTRA user plane (PDCP/RLC/MAC/PHY) and control plane (RRC) protocol terminations towards the UE. The eNBs are interconnected with each other by means of the X2 interface. The eNBs are also connected by means of the S1 interface to the EPC (Evolved Packet Core), more specifically to the MME (Mobility Management Entity) by means of the S1-MME and to the Serving Gateway (S-GW) by means of the S1-U. The S1 interface supports a many-to-many relation between MMEs / Serving Gateways and eNBs. The E-UTRAN architecture is illustrated in Figure 3.1 below.

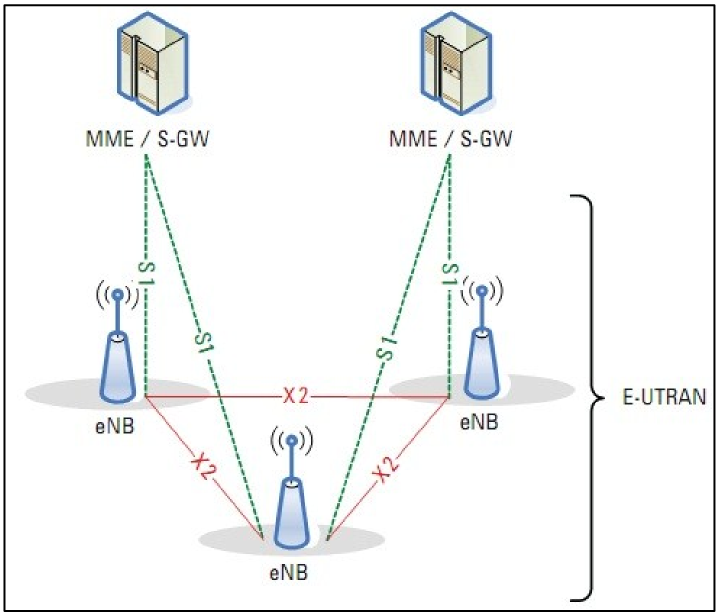


Figure 3.1: EUTRAN Overall Architecture [22]

## 3.2 The eNB

eNodeB plays a very significant role in the EUTRAN. It combines the functions of the UMTS NodeB as well as part of functions of RNC. Hence the term evolved NodeB (eNodeB). We will now see the main functions of an eNB as illustrated in figure 3.2

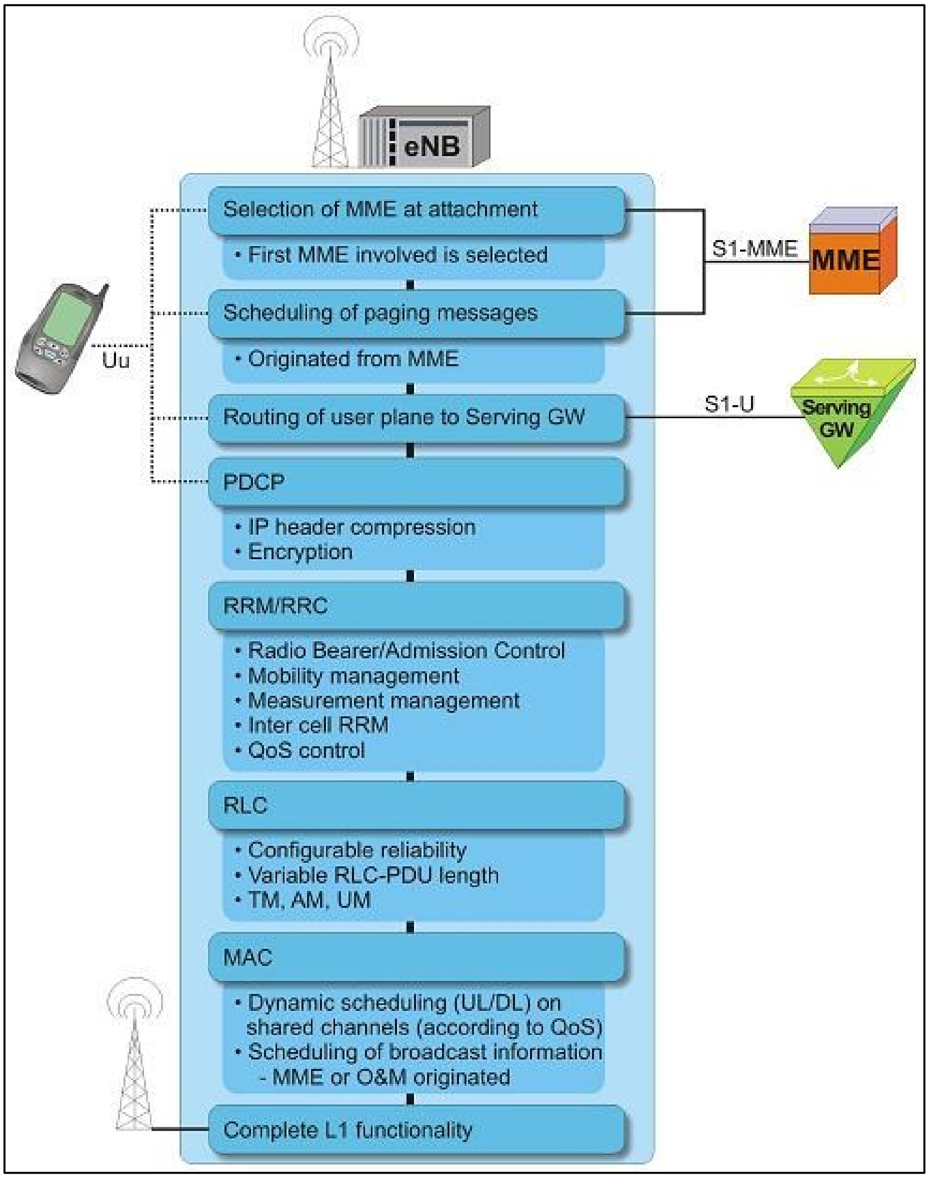


Figure 3.2: key functions of an eNB [23]

The key point to be noted from the above figure is that most of the functions of a RNC in UTRAN now reside in the eNB in an E-UTRAN network; some of which are as follows:

- *Selection of MME at attachment:* S1-flex mechanism allows for a single eNB to be connected to more than one MME. The eNB will select a particular MME for a UE based on the load situation in the network and the operator the UE belongs to.
- *Scheduling of paging messages:* Once the UE is in the idle state, it cannot be tracked on a level of either location area (LA) or routing area (RA) but has to be tracked on a level of tracking area (TA). Then the UE needs to be paged on TA level.
- *Routing of user plane data to Serving GW:* eNB is connected to the S-GW via the S1-U interface. As we will see in the following sections, S1-U uses GPRS Tunneling Protocol (GTP-U) protocol over UDP/IP Stack to transfer user data to and from eNB and S-GW. One of the advantages of using GTP-U is its inherent facility to identify tunnels and also to facilitate intra-3GPP mobility.

## 3.3 LTE Protocol Architecture

The overall radio interface protocol architecture for LTE is described in Figure 3.3, covering only the protocol part of the radio access in LTE. Additionally there are protocols in the core network that are between the UE and the core network but these are transparent to the radio layers and are generally referred to as Non-Access Stratum (NAS) signaling.

The physical layer carries the transport channels provided by the MAC layer. The transport channels, namely PCH, BCH, RACH, DL-SCH and UL-SCH describe how and with what characteristics data are carried on the radio interface on the physical channels. The MAC layer supports the logical channels which are PCCH, BCCH, CCCH, DCCH1, DCCH2, DTCH1 and DTCH2 to the RLC layers which characterizes the type of data to be transmitted. Above the RLC layer there is the PDCP layer, now used for both the control and the user plane, in contrast to WCDMA where it was only used for user plane data. Layer 2 provides upwards radio bearers. RRC signaling messages are carried by Signaling Radio Bearers (SRBs).

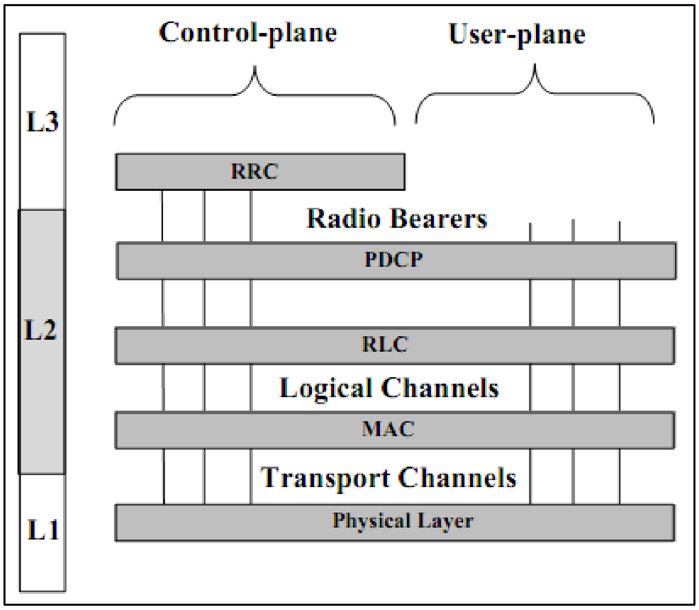


Figure 3.3: Overall LTE Radio Interface Protocol Architecture [14]

User Plane Radio Bearers (RBs) carry the corresponding user data. As described in Figure 3.3 on radio protocol architecture for the control plane, MAC, RLC, PDCP and RRC are all located in the eNodeB.

## 3.4 E-UTRAN Control Plane Architecture

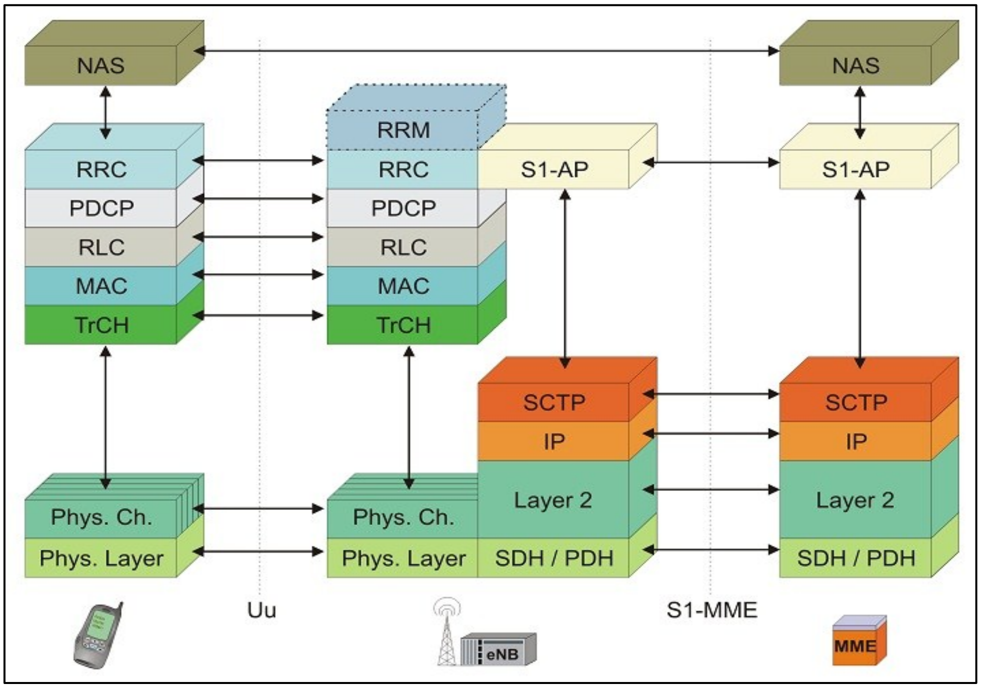


Figure 3.4: E-UTRAN Control Plane Protocol Stack [23]

Key point of this section is that the control plane protocol stacks structure of LTE combines the All IP Network (AIPN) protocol stack and the air-interface protocol stack of UMTS.

### *3.4.1 Radio Resource Control (RRC)*

The RRC (Radio Resource Control) layer is a key signaling protocol which supports many functions between the terminal and the eNodeB. The log list of procedures proposed by the RRC layer can be classified into the following:

#### 3.4.1.1 The key features of RRC

- RRC connection management – which includes the establishment and the release of the RRC connection between the terminal and the eNodeB.

- Establishment and release of radio resources – which relates to the allocation of resources for the transport of signaling messages or user data between the terminal and eNodeB.

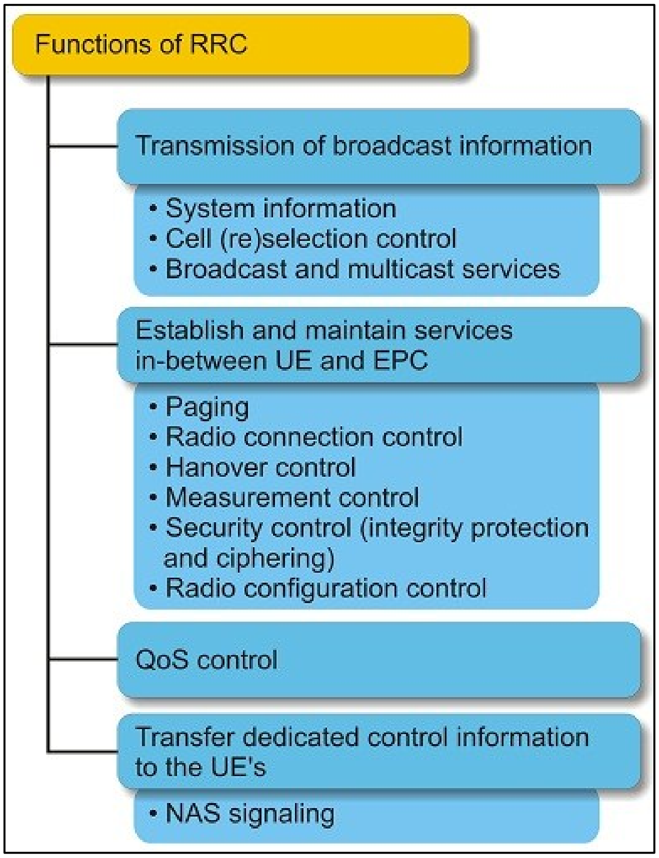


Figure 3.5: Key Features of RRC Layer [23]

- Broadcast of system information – this is performed through the BCCH logical control channel. The information broadcast from the RRC layer is either related to the Access Network (such as radio-related parameters) or to the Packet Core (for information such as the cell corresponding geographical area or network identity)
- Paging – this is performed through the PCCH logical control channel.
- Transmission of signaling messages to and from the EPC – these messages (known as NAS for Non Access Stratum) are transferred to and from the terminal via the RRC; they are, however, treated by RRC as transparent messages.

The RRC also supports a set of functions related to end-user mobility for terminals in RRC Connected state. This includes:

- Measurement control – which refers to the configuration of measurements to be performed by the terminal as well as the method to report them to the eNodeB.
- Support of inter-cell mobility procedures – which are also known as handover
- User context transfer between eNodeB at handover.

### 3.4.1.2 The RRC States

The main function of the RRC protocol is to manage the connection between the terminal and the EUTRAN access network. To achieve this, RRC protocol states have been defined in figure 3.6. Each of them actually corresponds to the states of the connection, and describes how the network and the terminal shall handle special functions like terminal mobility, paging message processing and network system information broadcasting.

In E-UTRAN, the RRC state machine is very simple and limited to two states only: RRC_IDLE, and RRC_CONNECTED.

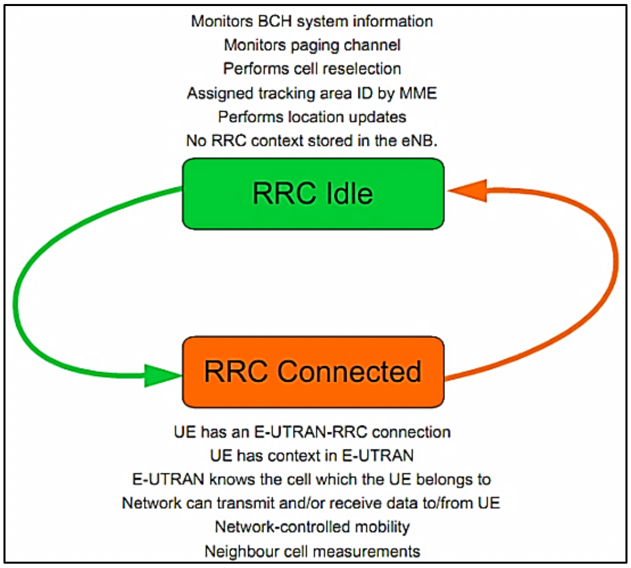


Figure 3.6: The RRC States [24]

In the RRC_IDLE state, there is no connection between the terminal and the eNodeB, meaning that the terminal is actually not known by the E-UTRAN Access Network. The terminal user is inactive from an application-level perspective, which does not mean at all that nothing happens at the radio interface level. Nevertheless, the terminal behavior is specified in order to save as much battery power as possible and is actually limited to three main items:

- Periodic decoding of System Information Broadcast by E-UTRAN – this process is required in case the information is dynamically updated by the network.
- Decoding of paging messages – so that the terminal can further connect to the network in case of an incoming session.
- Cell reselection – the terminal periodically evaluates the best cell it should camp on through its own radio measurements and based on network System Information parameters. When the condition is reached, the terminal autonomously performs a selection of a new serving cell.

Following cell reselection, it may happen that the terminal changes geographical area (or Tracking Area). If this occurs, the terminal is required to update the network in order to be still able to receive paging messages. For that purpose, the terminal has to leave temporarily the RRC_IDLE state so as to be able to exchange the necessary signaling information with the network. When the update procedure is over and if no service has been activated or answered by the user in the meantime, the terminal returns to the RRC_IDLE state.

In the RRC_CONNECTED state, there is an active connection between the terminal and the eNodeB, which implies a communication context being stored within the eNodeB for this terminal. Both sides can exchange user data and or signaling messages over logical channels. Unlike the RRC_IDLE state, the terminal location is known at the cell level. Terminal mobility is under the control of the network using the handover procedure, which decision is based on many possible criteria including measurement reported by the terminal of by the physical layer of the eNodeB itself.

## 3.5 E-UTRAN User Plane Architecture

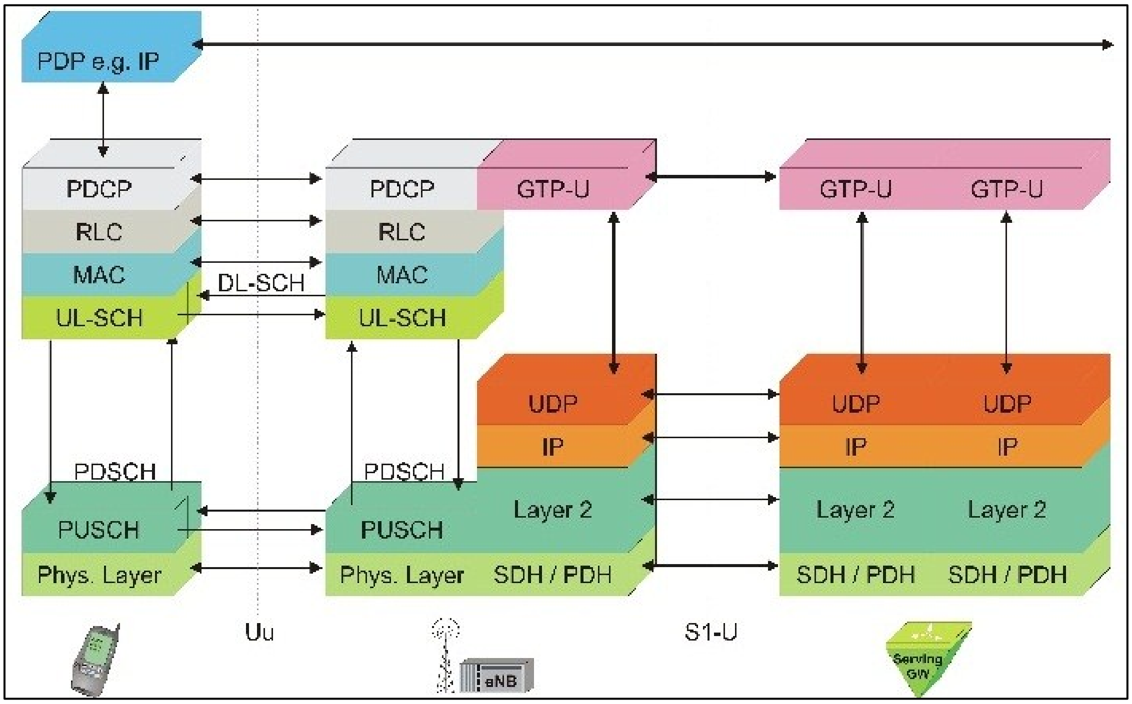


Figure 3.7: E-UTRAN User Plane Architecture [23]

Key point of to be noted from the figure above is that the user plane protocol stack structure of LTE is combining the All IP Network (AIPN) protocol stack and the air-interface protocol stack of UMTS. The main elements of the user plane in E-UTRAN protocol architecture are PDCP, RLC, MAC, PHY layers. Each of them is explained in detail in the following sections.

### *3.5.1 Packet Data Convergence Protocol (PDCP) Layer*

This layer processes Radio Resource Control (RRC) messages in the control plane and Internet Protocol (IP) packets in the user plane. Depending on the radio bearer, the main functions of the PDCP layer are header compression, security (integrity protection and ciphering), and support for reordering and retransmission during handover. There is one PDCP entity per radio bearer.

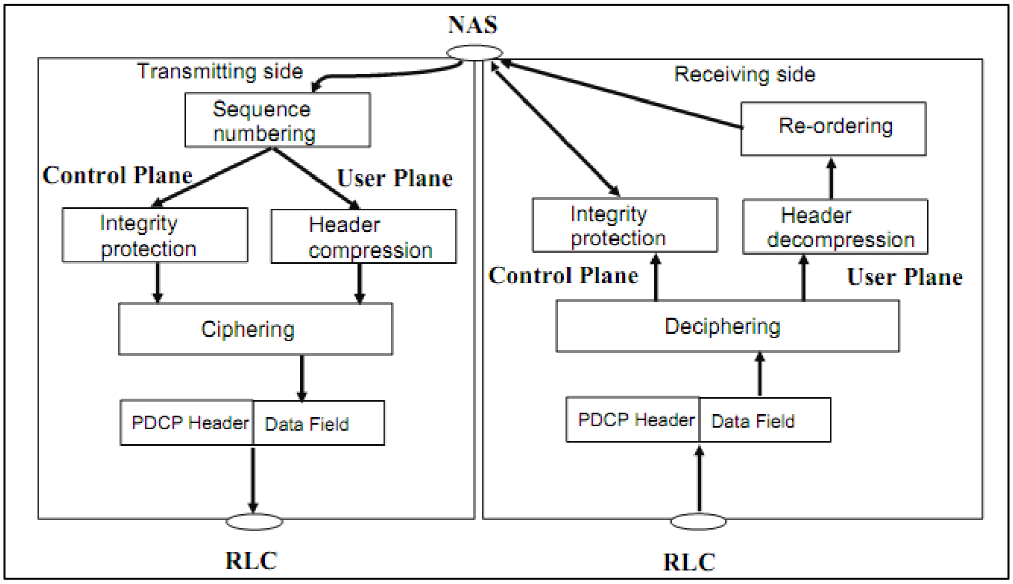


Figure 3.8: PDCP layer operation for the packets which are associated to a PDCP SDU [24]

The PDCP layer manages data streams for the user plane, as well as for the control plane. The architecture of the PDCP layer differs for user plane data and control plane data. The PDCP layer receives PDCP SDUs from the NAS and RRC and after ciphering and other actions, as shown in Figure 3.8 for operation of packets that are associated to a PDCP SDU; the data are forwarded to the RLC layer. Correspondingly, in the receiving side the data are received from the RLC layer. Besides the functionalities listed above, the PDCP layer has specific functions in connection with the handover events (intra-LTE). The PDCP does the in-order delivery function in the downlink direction and detects duplicates. In the uplink direction, PDCP retransmits all the packets which have not been indicated by lower layers to be completed, as the lower layers will flush all the HARQ buffers with handover. In the downlink direction, the PDCP layer will forward the non-delivered packets to the new eNodeB. This is to ensure that no data are lost in connection with a handover event between LTE eNodeBs.

### 3.5.1.1 Robust Header Compression (RoHC)

Since VoIP is a critical application for LTE, provision for Header Compression has been made in the Rel. 8 of 3GPP Standards. Because there is no more circuit switching in LTE, all voice signals must be carried over IP and there is a need for efficiency. Various standards are being specified for use in profiles for robust header compression (ROHC), which provides a tremendous savings in the amount of header that would otherwise have to go over the air. These protocols are designed to work with the packet loss that is typical in wireless networks with higher error rates and longer round trip time. ROHC is defined in IETF RFC 3095, RFC4815 and RFC 3843.

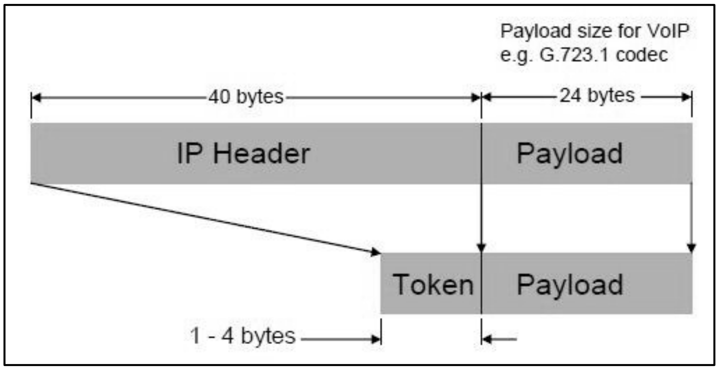
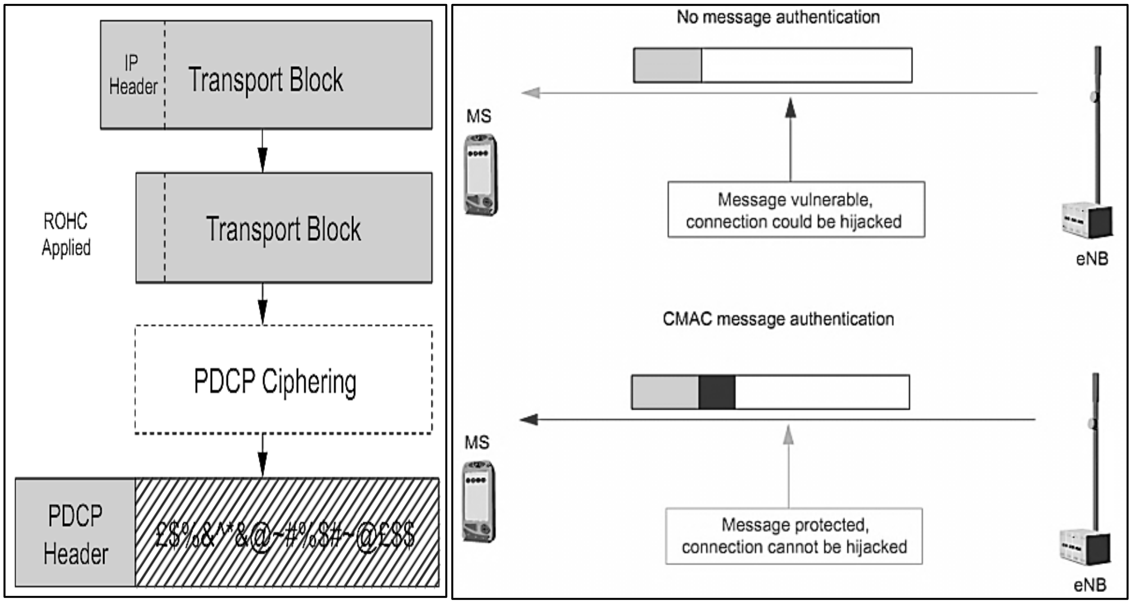


Figure 3.9: Robust Header Compression (RoHC) in PDCP [18]

Typically, for the transport of a VoIP packet which contains a payload of 24 - 32 bytes, the header added will be 40 bytes for the case of IPv4 and 60 bytes for the case of IPv6 - i.e. an overhead of 125% and 188% respectively. By means of RoHC, after the initialization of the header compression entities this overhead can be compressed to four to six bytes, and thus to a relative overhead of 12.5% to 18.8%. This calculation is valid during the active periods, but during silence periods the payload size is smaller so the relative overhead is higher.

### 3.5.1.2 Security (Encryption and Integrity Protection)

Ciphering, both encryption and decryption, also occurs in the PDCP. Security has to occur below the ROHC because the ROHC can only operate on unencrypted packets. It cannot understand an encrypted header. Ciphering protects user plane data, radio resource control (RRC) data and non-access stratum (NAS) data. Processing order in the PDCP is as follows: For the downlink, decryption occurs first, then ROHC decompression. For the uplink, ROHC compression occurs first, then encryption. Details of LTE security architecture are still being defined. The 3GPP System Architecture Working Group 3 (SA3) is responsible for security and has decided to use either Advanced Encryption Standard (AES) or SNOW 3G algorithms. Specific modes for AES are still being determined; AES is a block cipher and has to use specific operational modes to operate in a streaming mode.



(a) (b)

Figure 3.10: (a) PDCP Ciphering function. (b) Integrity Protection [18]

### *3.5.2 Radio Link Control (RLC) Layer*

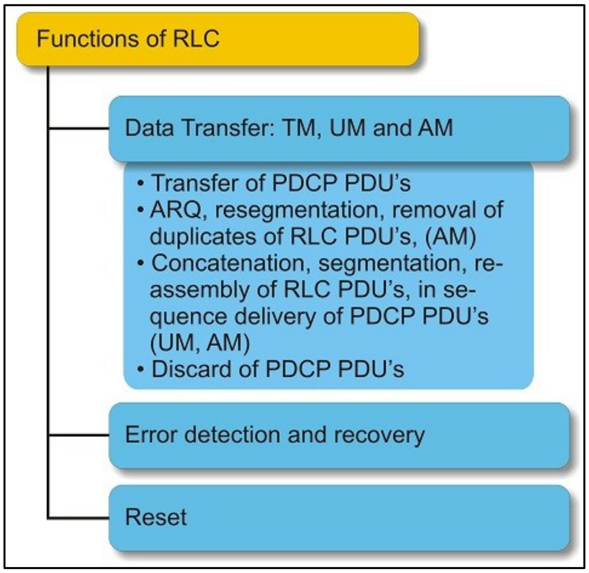


Figure 3.11: Key Features of RLC [23]

The main purpose of the Radio Link Control (RLC) E-UTRAN protocol layer is to receive/deliver a data packet from/to its peer RLC entity. For that purpose, the RLC proposes three modes of transmission TM (Transparent Mode), UM (Unacknowledged Mode) and AM (Acknowledged Mode).

The TM mode is the simplest one, as it does not change or alter the upper layer data. This mode is typically used for BCCH or PCCH logical channel transmissions which require no specific treatment from the RLC layer. The RLC Transparent Mode Entity receives data from the upper layers and simply passes it to the underlying MAC layer. There is no RLC header addition, data segmentation or concatenation. The added value of the UM mode is to allow the detection of packet loss (the receiving entity can detect that a RLC packet has not been received correctly) and provides packet re-ordering and re-assembly. These operations can be performed thanks to the presence of a Sequence Number in the RLC packet header. The UM mode can apply to any Dedicated or Multicast logical channel, depending on the types of application and expected Quality of Service.

Packet re-ordering refers to the re-sequencing of packets in case they have not been received in order (which may happen in the case of HARQ repetition). Packet re-assembly is performed when an upper-layer packet has been segmented by the sending RLC entity before transmission. Finally, the AM mode is the most complex one. In addition to UM mode-supported features, an AM RLC entity is able to ask its peer for packet retransmission in case a loss is detected. This mechanism, specific to the AM mode, is known as ARQ (Automatic Repeat Request). For that reason, the AM mode only applies to DCCH or DTCH logical channels.

### *3.5.3 Medium Access Control (MAC) Layer*

The Medium Access Control (MAC) radio protocol layer of E-UTRAN's main purpose is to provide an efficient coupling between the RLC Layer 2 services and the physical layer. From that perspective, the MAC supports four main functions:

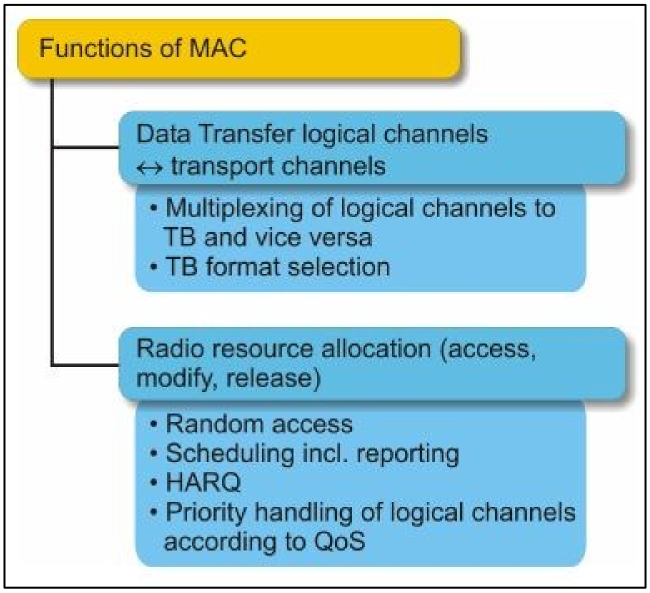


Figure 3.12: Key Features of MAC [23]

- Mapping between logical channels and transport channels – when the standard offers different options for the transport of data for a given logical channel, it is up to the MAC layer to choose the transport channel according to the configuration defined by the operator.

- Transport format selection – this refers to, for example, the choice of Transport Block size and modulation scheme made by the MAC layer and provided as input parameters to the physical layer.
- Priority handling between logical channels of one terminal as well as between terminals.
- Error correction through HARQ (Hybrid ARQ) mechanism.

#### 3.5.3.1 Priority Handling

Priority handling is one of the main functions supported by the MAC layer. Priority handling refers to the process which selects the packets from the different waiting queues to be submitted to the underlying physical layer for transmission on the radio interface. This process is complex, as it takes into account the different flows of information to be transmitted – including pure user data (the DTCH logical channel) as well as signaling initiated by the EUTRAN or the EPC (the DCCH logical channel) – with their relative priority, as well packet repetition in case an already transmitted packet has not been correctly received by the other end. For that reason, the priority handling' part of the MAC layer is tightly coupled with the Hybrid ARQ part. In addition, the MAC layer on the network side is also responsible for uplink priority handling, as it arbitrates between all the uplink scheduling requests from all the terminals which share the same UL-SCH transport channel as shown in figure 3.12. On the terminal side, the MAC layer only mixes flows from the terminal for uplink transmission and has to arbitrate between its own information flows for uplink scheduling requests and transmission. In contrast, for the Downlink Shared Channel, the eNodeB has to consider all the flows (or logical channels) sent to all the users in the cell.

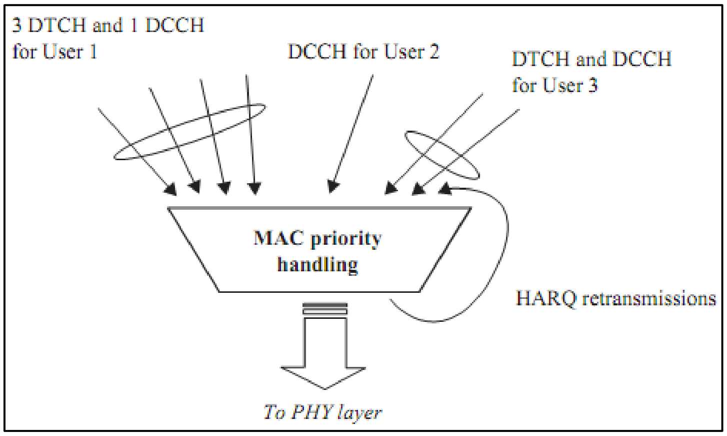


Figure 3.13: Priority handling in the eNodeB MAC layer [18]

### 3.5.3.2 HARQ

The principle of hybrid ARQ is to buffer blocks that were not received correctly and consequently combine the buffered data with retransmissions. The actual method of doing soft combining depends on the HARQ combining scheme selected. In the Chase combining scheme, initial transmission and retransmission are identical. The receiver always combines the full retransmission of the failed block. In the IR (Incremental Redundancy) schemes, new parity bits are transmitted together with the failed block. The receiver receives coded symbols, which introduce new information to the first transmitted block. In E-UTRAN, HARQ is composed of several parallel parts, so that transmission can continue on other processes while one of them is stuck with retransmissions. In the downlink, HARQ is based on asynchronous retransmissions with adaptive transmission parameters. In the uplink, HARQ is based on synchronous retransmissions. In the synchronous scheme, retransmission can only occur at certain sub frame numbers following the first transmission. In the asynchronous scheme, packet retransmissions are not constrained in terms of frame time. Synchronous retransmission is preferred in uplink, because there is less protocol overhead. Synchronous retransmission does not require to explicitly signal the HARQ process number, since it can be deduced from the sub frame number. The HARQ in E-UTRAN is similar to those in 3G HSDPA (for the downlink transmission) and E-DCH/HSUPA (for the uplink transmission).

### *3.5.4 Physical (PHY) layer*

The role of the PHY layer is to provide data transport services on physical channels to the upper RLC and MAC layers. This section describes of the physical layer from a functional perspective and also in terms of interactions with other radio interface layers. Figure 3.14 describes the eNodeB physical layer model in the example of a downlink SCH transport channel as being the most generic scheme. Of course, similar models exist for the uplink (in the terminal), and for all the other transport channels. At each TTI (Transmission Time Interval), the physical layer receives a certain number of Transport Blocks for transmission. To each Transport Block is added a CRC (Cyclic Redundancy Check) or set of bits used by the receiving end to detect transmission errors. The Blocks are then protected by a robust channel-encoding scheme (like convolutional or turbo coding) and size-adapted to make sure the encoded packet matches the physical channel size. This phase is under the control of the MAC HARQ (Hybrid ARQ) process which may adapt the channel coding rate (meaning the robustness to transmission errors) based on the information reported by the receiving entity.

Interleaving is a process to improve robustness to radio transmission errors. When an error occurs on an encoded packet transmitted over the radio interface, it will affect multiple consecutive bits or symbols. On the receiving side, the action of de-interleaving will have the effect of spreading the erroneous bits on the whole transmitted sequence on different Transport Blocks. This will make it easier for the channel decoder to recover the exact bits transmitted initially, as a single block will only be affected by a smaller part.

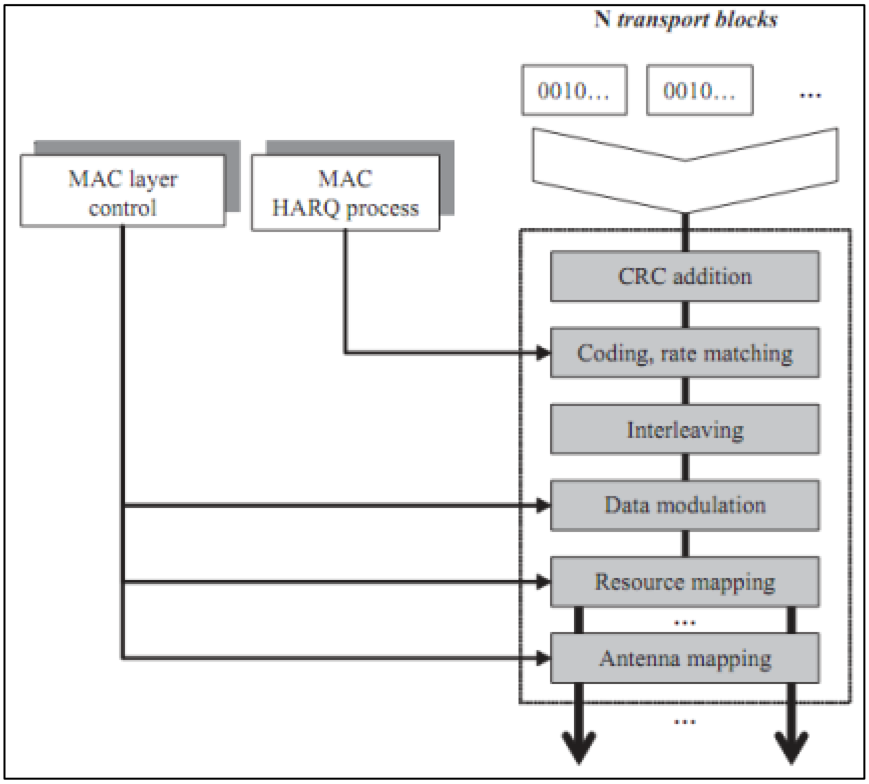


Figure 3.14: The Downlink Shared Channel PHY layer model. [24]

In the data-modulation process, the actual modulation is under the control of the MAC scheduler. Resource mapping relates to the segmentation of transmitted data into resource blocks. Antenna mapping relates to the mapping of resource blocks (as above) on available antenna ports (MIMO).

CRC and interleaving processes are not controlled by higher layers. For those two operations, the PHY layer uses static parameters and algorithms specified by the E-UTRAN standard.

As mentioned above, there exist similar models for other transport channels. However, there are more or fewer subsets of the shared channel model presented above. Transmission over other transport channels such as the PCCH for paging or BCCH for system information broadcasting is not flexible in terms of channel coding or modulation. For this type of transport channel, the E-UTRAN standard does not propose any options or alternatives.

## 3.6 E-UTRAN Network Interfaces

This section provides some general information about the E-UTRAN S1 and X2 network interfaces. Similar to the UTRAN network interface model, the E-UTRAN network interface model is composed of two main parts: the transport network layer – which refers to as the way radio network layer data are transported – and the radio network layer – which encompasses the top-level protocols of the interface In addition to this OSI-like vertical separation, each interface is split between a User Plane and a Control Plane. This can be clearly seen in figure 3.15.

The User plane transports all user data (such as voice or video packets) as well as application level signaling (such as SIP or RTCP packets). The Control plane handles all the messages and procedures related to the interface supported features.

A common example of this includes control messages for the handover management or bearer management. The physical layer is common to both UP and CP. Aside from that, UP and CP use specific protocols which allows for defining a different and independent transport stack and bearers for each of the

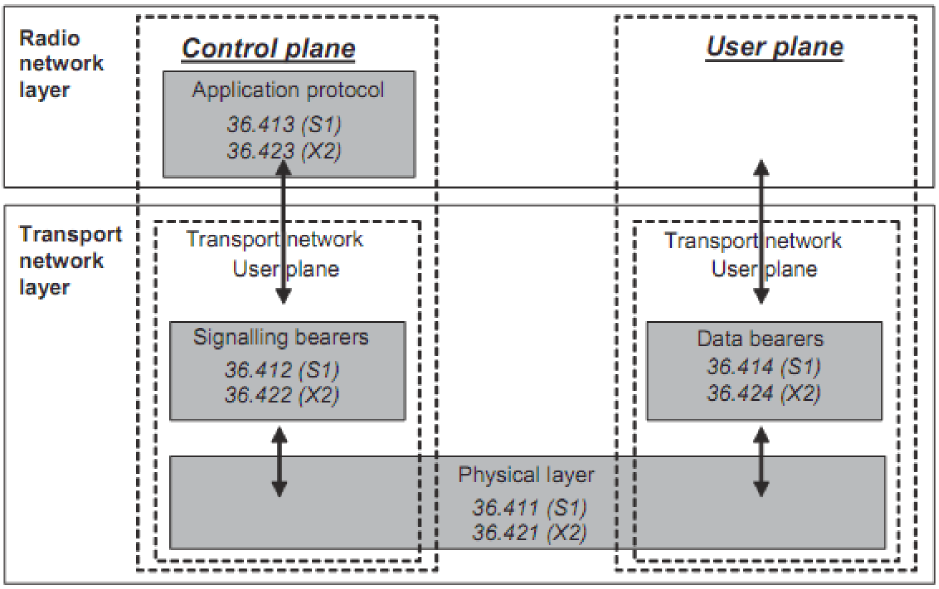


Figure 3.15: The E-UTRAN network interface model. (The numbers 36.4xx indicate the 3GPP Specification for the particular interface) [25]

planes. The CP information is more constrained in terms of security, reliability and data loss, whereas the UP information can rely on simpler and less secured routing protocols. The E-UTRAN network interfaces are open standards defined completely by 3GPP, similar to the UTRAN network interfaces. This conformity to open standards allows different vendors to manufacture eNodeBs, deploy them in a single network and inter-connect them over the X2 interface or with MME or Serving GW nodes over the S1 interface.

### *3.6.1 The S1 Interface [26]*

The S1 interface connects the eNB to the EPC. It is split into Control Plane (S1-MME) and User Plane (S1-U). The details are explained in the following sections.

#### 3.6.1.1 The S1 Control (S1-MME) Plane

The S1-C (or S1-MME) is a signaling interface which supports a set of functions and procedures between the eNodeB and the MME. All the S1-MME signaling procedures belong to four main groups:

- Bearer-level procedures – this set corresponds to all procedures related to bearer setup, modification and release. On the scope of the S1 interface, a bearer corresponds to the S1 segment of a session plus the radio interface path. These procedures are typically used during the establishment or the release of a communication session.
- Handover procedures – which encompasses all the S1 functions related to user mobility between eNodeB or with 2G or 3G 3GPP technologies.
- NAS signaling transport – this corresponds to the transport of terminal–MME signaling over the S1 interface. The terminal–MME signaling is also called NAS (Non Access Stratum signaling), as it is transparent to the eNodeB. Due to the importance of these messages, they are transported over S1-Cusing specific procedures, rather than the nonguaranteed S1-U GTP.
- Paging procedure – which is used in case of user terminated session. Through the paging procedure, the MME request the eNodeB to page to terminal in a given set of cells.

The S1-MME interface shall provide a high level of reliability in order to avoid message retransmission and unnecessary delay in control plane procedure execution.

Depending on transport network deployment, there may be some cases in which the UDP/IP transport is not reliable enough. Besides, in case the transport network is not owned by the mobile radio network operator, it may happen that the transport network QoS cannot be guaranteed all the time. This is the reason why the S1-MME interface makes use of a reliable transport network Layer, which is set up end-to-end (between the eNB and the MME nodes). In the EPS architecture, this service is ensured by SCTP. More details on how SCTP works has been mentioned in Appendix A.

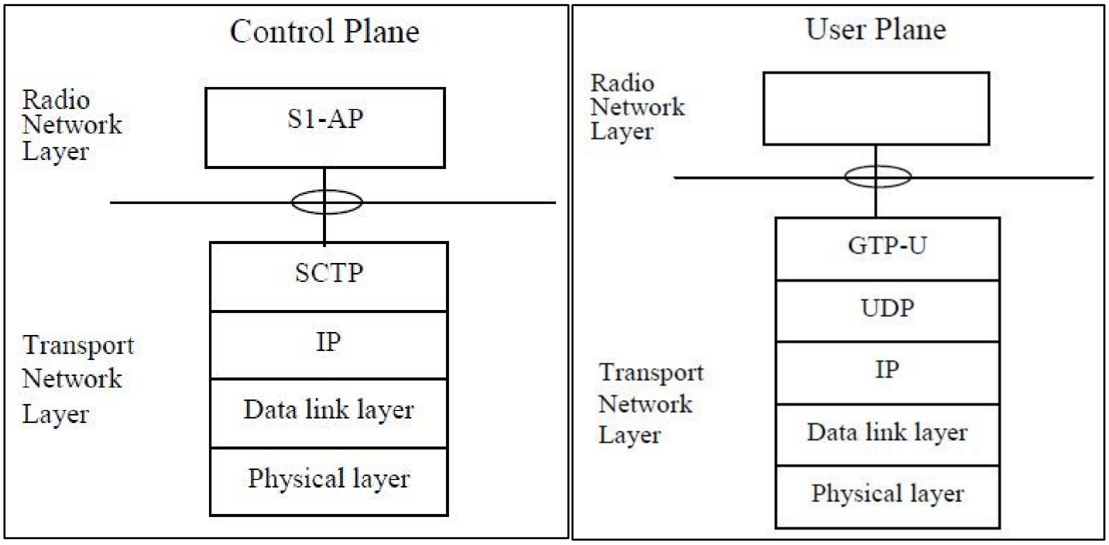


Figure 3.16: S1 – MME and S1 – U interface [26]

#### 3.6.1.2 The S1 User (S1-U) Plane

The S1-U (or S1 User plane interface) role is to transport user data packet between the eNodeB and the Serving GW. This interface makes use of a very simple GTP over UDP/IP transport protocol stack which only provides user data encapsulation. There is no flow control or error control, or any mechanism to guarantee data delivery over the S1-U interface. The GTP (GPRS Tunneling Protocol) is actually inherited from 2G/GPRS and 3G/UMTS networks. In 2G/GPRS networks, GTP is used between GPRS nodes (the SGSN and the GGSN). In 3G/UMTS networks, GTP is also used over the Iu-PS interface (between the RNC and the SGSN).

#### 3.6.1.3 S1-Application Protocol (S1-AP) [28]

S1 Application Part (S1AP) is the control plane signaling protocol between the eNodeB and the Mobility Management Entity (MME). It supports the S1-MME interface and utilizes

SCTP in the transport layer. SCTP is for the control plane, which guarantees delivery of signaling messages between the MME and eNodeB.

The functionality of S1AP is as follows:

- Manage the SAE Bearer
- Transfer initial context
- Provide mobility functions for UE
- Provide paging
- Provide reset
- Transport NAS signaling
- Report errors
- Provide UE context release
- Transfer status

*Features and Benefits of S1-AP*

- Sets up, modifies, and releases SAE bearer planes triggered by the MME or eNodeB
- Establishes an S1UE context in the eNodeB
- Sets up the default IP connectivity
- Sets up one or more SAE bearer(s) if requested by the MME
- Transfers NAS signaling-related information to the eNodeB if required
- Provides UE Capability Information when it is received from the UE to the MME
- Responsible for Mobility Functions for UEs in LTE_ACTIVE in order to enable:
    - Change of eNodeBs within SAE/LTE (Inter MME/Serving SAE-GW Handovers) via the S1 interface (with EPC involvement)
    - Change of RAN nodes between different RATs (Inter-3GPP-RAT Handovers) via the S1 interface (with EPC involvement)

- Provides the EPC the capability to page the UE
- Resets functionality to ensure a well-defined initialization on the S1 interface
- Features Error Indication functionality on the S1 interface to allow proper error reporting / handling in cases where no failure messages are defined
- Features overload functionality to indicate the load situation in the control plane of the S1 interface
- NAS Signaling transport function between the UE and the MME is used to:
    - Transfer NAS signaling related information and to establish the S1 UE context in the eNodeB
    - Transfer NAS signaling related information when the S1 UE context in the eNodeB is already established
- Manages the release of a UE specific context in the eNodeB and the MME
- Transfers PDCP SN Status information from source eNodeB to target eNodeB in support of in-sequence delivery and duplication avoidance for intra LTE handover
- Features a reference application for the S1-MME interface supporting:
    - Bearer setup and release
    - Initial UE context setup
    - NAS Node selection
    - Intra-LTE handover

### *3.6.2 The X2 Interface* [27]

The X2 interface connects multiple eNBs. LTE uses the same protocol structure over both S1 and X2 interfaces, which simplifies data forwarding operation. It is split into Control Plane (X2-C) and User Plane (X2-U). The details are explained in the following sections and are illustrated in figure 3.17.

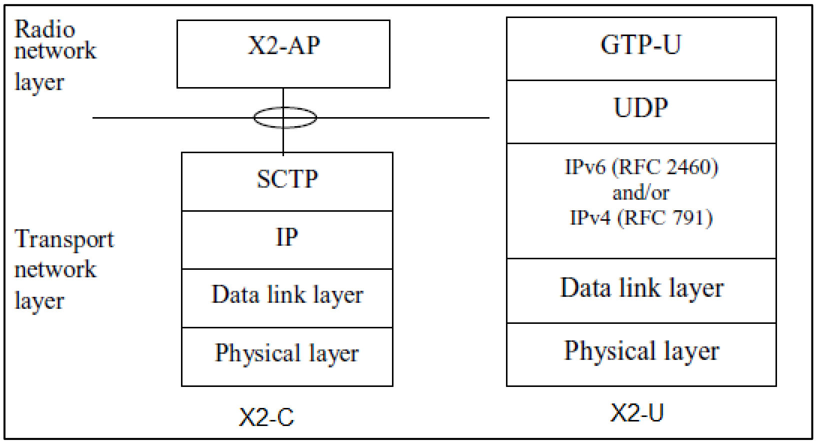


Figure 3.17: X2 Control Plane and X2 User Plane [27]

3.6.2.1 X2 User Plane Interface

The X2-U (or X2 User plane interface) role is to transport user data packets between eNodeBs. This interface is only used for limited periods of time, when the terminal moves from one eNodeB to another, and provides buffered packet data forwarding. X2-U makes use of the same GTP tunneling protocol already used over the S1-U interface.

3.6.2.2 X2 Control Plane Interface

The X2-C (or X2 Control plane interface) is a signaling interface which supports a set of functions and procedures between eNodeBs. The X2-C procedures are very limited in number and are all related to user mobility between eNodeBs, so as to exchange user context information between nodes (including allocated bearers, security material, etc.).

In addition, the X2-C interface proposes the Load Indicator procedure whose purpose is to allow an eNodeB to signal its load condition to neighboring eNodeBs. The detailed use of this function is not further detailed by the standard (as it relates to algorithms under the control of the equipment manufacturer). The aim of this procedure is to help the support of load-balancing management, or to optimize handover thresholds and handover decisions.

The need for a reliable transport of signaling between nodes is the same as over the S1-MME interface. This is the reason why X2-C also uses an SCTP over IP transport layer.

### 3.6.2.3 X2 Application Protocol (X2-AP) [29]

LTE X2AP is the control plane protocol between eNodeBs on the X2 interface; it supports load management and handover coordination between eNodeBs. X2-AP enables Network Equipment Providers of eNodeBs to:

- Accelerate time to market
- Reduce development costs
- Reduce project risk of internally-developed LTE X2AP applications

*Features and Benefits of X2-AP*

The features and benefits of LTE X2-AP are as follows:

- eNodeB functionality
- Mobility Management; allows the eNB to move the responsibility of a certain UE to another eNB. Forwarding of user plane data, Status Transfer, and UE Context Release function are parts of the mobility management.
- Load Management used by eNBs to indicate resource status, overload and traffic load to each other.
- Reporting of General Error Situations allows reporting of general error situations for which function specific error messages have not been defined.
- Resetting the X2 interface
- Setting up the X2 used to exchange necessary data for the eNB for setup the X2 interface and implicitly perform an X2 Reset.

- eNB Configuration Update allows updating of application-level data needed for two eNBs to interoperate correctly over the X2 interface

# CHAPTER 4

# LTE SELF ORGANIZING NETWORKS (SON) [1] [30]

## 4.1 Introduction

Reduction of OPEX, CAPEX and complexity are the key drivers for the development of LTE Self Organizing Networks. The main feature of a Self Organizing Networks is Self (Configuration / Optimization / Healing). Also one of the important derivations of this development is that the operations of a multiple-vendor based system are simplified. It is important that measurements and performance data of different vendors share the same “language”. Such alignment provides ease of network performance analysis and troubleshooting, and reduces efforts in maintaining the network at a properly working state.

A self-optimizing function shall improve the network performance and type of response to dynamic processes in the network. The need for self-optimization arises in the initial stages, especially during the early deployment phase, when the efforts required to set up and optimize a network are significantly large and lead to lengthy periods of network optimization. It is therefore beneficial to have self-configuration and self-optimization algorithms up and running when initial deployment starts.

The key factors for automation of processes in a cellular network operation are:

- The complexity arising in network due to requirements for inter-working of 2G, 3G & 4G and the large number of parameters required by the three systems for optimal inter-operational performance require a lot of effort.
- The Femtocells, widely known as Home eNBs (HeNB) in LTE should have a simple deployment process, something comparable to plug and play.

Figure 4.1 illustrates the process of Network Operation in a LTE Network, first without and then with Self Organizing features. In the previous generation systems (networks without SON features), a network operator had to continuously monitor the network performance and alarms together with network settings. Network configuration then had to be re-planned by human operator based on collected and analysis of the gathered data. Since there are quite many complex operations related to network management in OAM, human-errors are a common occurrence.

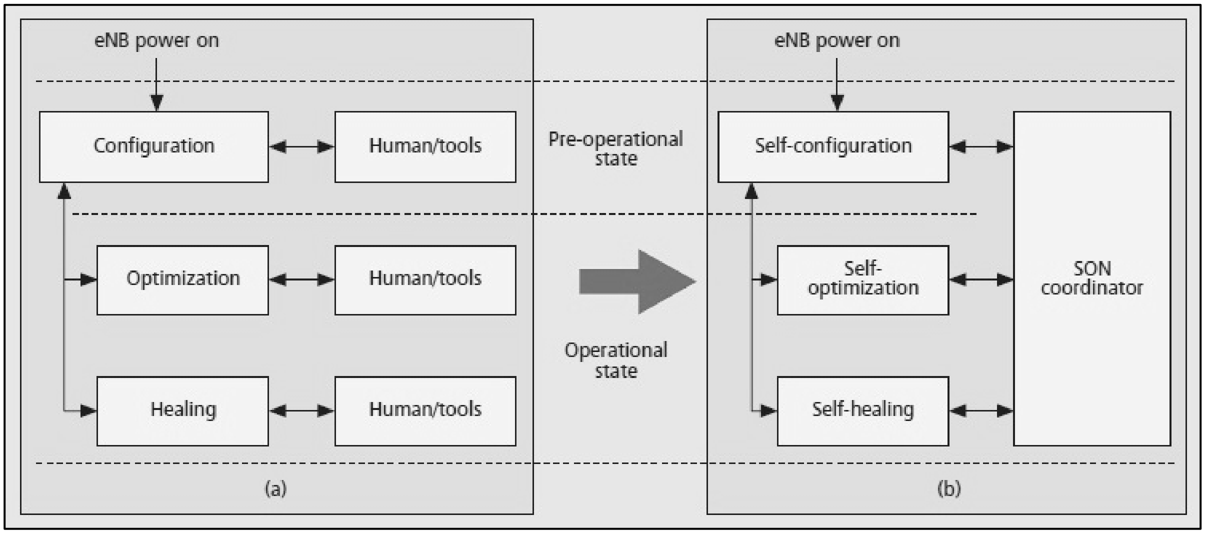


Figure 4.1: LTE Network Operation (a) without SON; (b) with SON feature [2]

SON can be used to reduce OPEX by reducing the requirement for manual workload. The current vision for SON is to have a very minimal human interaction in operations and automate in the network elements, domain manager, and/or management system. SON helps shift the human effort from the lower levels to the higher management level, as seen from Figure 4.1 (b), so that the human operators' main role is to design policies for SON, supervise SON processes and intervene, if needed. OPEX savings are derived from the automated and closed loop functionality of SON. SON configurations can be manually overridden and sufficient provisions should be made for the operator to reconfigure and revert to the original configuration.

## 4.2 SON Development in NGMN and 3GPP

NGMN (Next Generation Mobile Networks) laid the foundations for the initial requirements on Self-Organizing Networks in 2006, and since then several use cases have been defined to cover multiple aspects of the network operations, including planning, deployment, optimization and maintenance. The generation of SON-specific requirements by the NGMN contributed to the adoption of the SON concept by the 3GPP in TR 36.902 and TR 32.500.

### *4.2.1 Self-Configuration*

Self-configuration mechanism is desirable during the pre-operational phases of network elements such as network planning and deployment, which will help reduce the CAPEX. Some Self-Configuration use cases are defined in the following table.

Table 4.1: Self-Configuration Use Cases [30]

| Automatic Neighbor Relation (ANR) function |
|---|
| Automated configuration of Physical Cell ID (PCI) |
| Self-establishment of a new eNB in the network |
| Self-configuration and self-healing of eNBs |

Following illustration from TS32.501 shows a high level logical architecture for Self-Configuration Mechanism

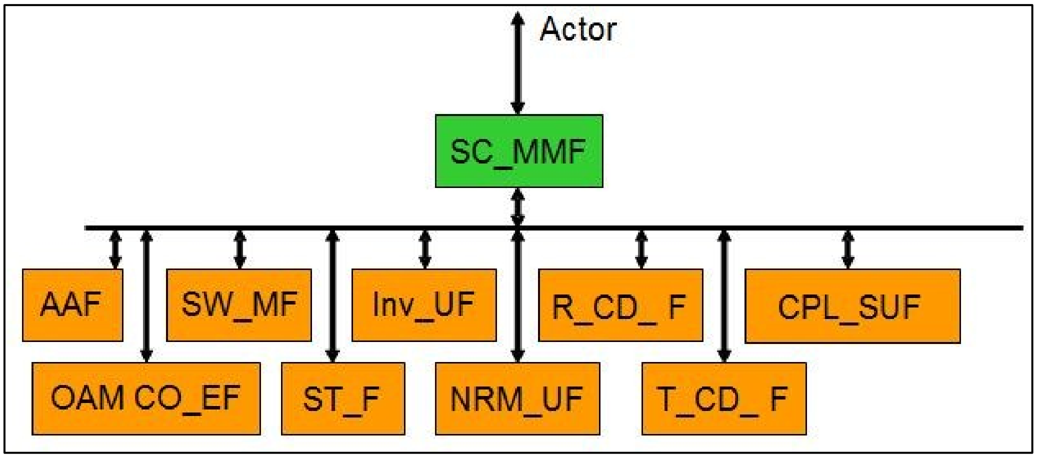


Figure 4.2: Self-Configuration Logical Architecture [3]

SC_MMF: Self-Configuration Monitoring and Management Function

AAF: Address Allocation Function

OAM CO_EF: OAM Connectivity Establishment Function

SW_MF: Software Management Function

Inv_UF: Inventory Update Function

R_CD_ F: Radio Network Configuration Data Function

CPL_SUF: Call Processing Link (CPL) Set Up Function

NRM_UF: NRM IRP Update Function

T_CD_ F: Transport Network Configuration Data Function

### *4.2.2 Self Optimization* [1] [30]

Self-optimization mechanism is desirable during the operational stage so that network operators get benefits of the dynamic optimization, e.g., mobility load balancing to make network more robust against environmental changes as well as the minimization of manual optimization steps to reduce operational costs. The main use cases in self-optimization are:

Table 4.2: Self-Optimization Use Cases [30]

| |
|---|
| Coverage and capacity optimization |
| Energy savings |
| Interference reduction |
| Mobility robustness optimization |
| Mobility load balancing optimization |
| RACH optimization |
| Inter-cell interference coordination |
| Self-establishment of a new eNB in the network |
| Self-configuration and self-healing of eNBs |
| Optimization of parameters due to troubleshooting |
| Continuous optimization due to dynamic changes in the network |
| Optimization of Quality of Service (QoS) related radio parameters |

### *4.2.3 Self-Healing* [1] [30]

The purpose of the Self-healing functionality of SON is to solve or mitigate the faults which could be solved automatically by triggering appropriate recovery actions.

For the fault management functionality, appropriate alarms shall be generated by the faulty network entity for each of the detected faults, regardless of whether it is an automatically detected/automatically cleared or an automatically detected/manually cleared fault.

As described above, alarms can be used to trigger Self-healing mechanisms. The Self-healing function continuously monitors these alarms, and when it is able to resolve which alarm/s could be solved automatically, it gathers necessary information, makes a deep analysis of the issue and then according to the derived results, the mechanism will trigger appropriate recovery actions to solve the fault automatically, if necessary.

For some Self-healing functions which are located in NEs and require more rapid response, the trigger of Self-healing can be the detection of a fault. Hence, when a fault is detected, an appropriate Self-healing Process will be triggered to try to heal the fault automatically.

The Self-healing functionality also monitors the execution of the recovery action/s and decides the next step accordingly. After a Self-healing procedure has ended, the Self-healing functionality shall generate and forward appropriate notifications to inform the IRPManager about the Self-healing result and all the information of the performed recovery actions may be logged.

A much more detailed description with the self-healing procedures is defined in the TS 32.541

## 4.3 SON Architectures [4] [6]

The self-organization functionality can be located as whole functionality block or even split in sub-functionality located in different nodes. Self-Optimization algorithms can be located in OAM or eNB or both of them. According to the location of optimization algorithms, SON can be divided into the three main architecture versions: Centralized SON, Distributed SON and Hybrid SON.

In the following three versions of SON, the self-optimization functionality (SOF) with respect to data acquisition functionality, data processing functionality and configuration management are presented. The data processing can be seen as process with input parameters, appropriate algorithm to process these input resulting in optimized parameters.

### *4.3.1 Centralized SON*

In Centralized SON, optimization algorithms are stored and executed from the OAM System. In such solutions SON functionality resides in a small number of locations, at a high level in the architecture. Figure 4.3 shows an example of Centralized SON.

In Centralized SON, all SON functions are located in OAM systems, so it is easy to deploy them. But since different vendors have their own OAM systems, there is low support for optimization cases among different vendors. And it also does not support those simple and quick optimization cases. To implement Centralized SON, existing Itf-N interface needs to be extended.

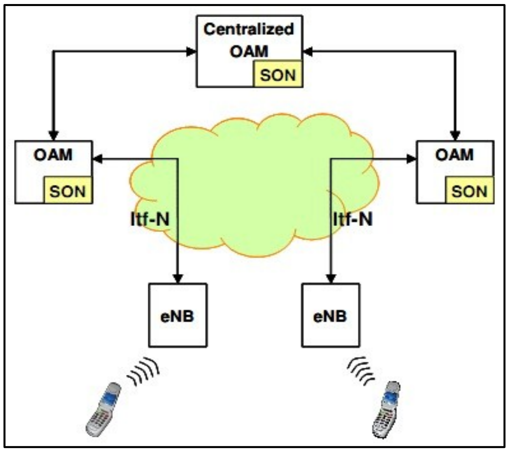


Figure 4.3: Centralized SON Architecture [4]

### *4.3.2 Distributed SON*

In Distributed SON, optimization algorithms are executed in eNB. In such solutions SON functionality resides in many locations at a relatively low level in the architecture. Figure 4.4 shows an example of Distributed SON. In Distributed SON, all SON functions are located in eNB, so it causes a lot of deployment work. And it is also difficult to support complex optimization schemes, which require the coordination of lots of eNBs. But in Distributed SON it is easy to support those cases, which only concern one or two eNBs and require quick optimization responses. For Distributed SON, X2 interface needs to be extended.

### *4.3.3 Hybrid SON*

In Hybrid SON, part of the optimization algorithms are executed in the OAM system, while others are executed in eNB. Figure 4.4 shows an example of Hybrid SON. In Hybrid SON, simple and quick optimization schemes are implemented in eNB and complex optimization schemes are implemented in OAM. So it is very flexible to support different kinds of optimization cases. And it also supports the optimization between different vendors through X2 interface. But on the other hand, it costs lots of deployment effort and interface extension work.

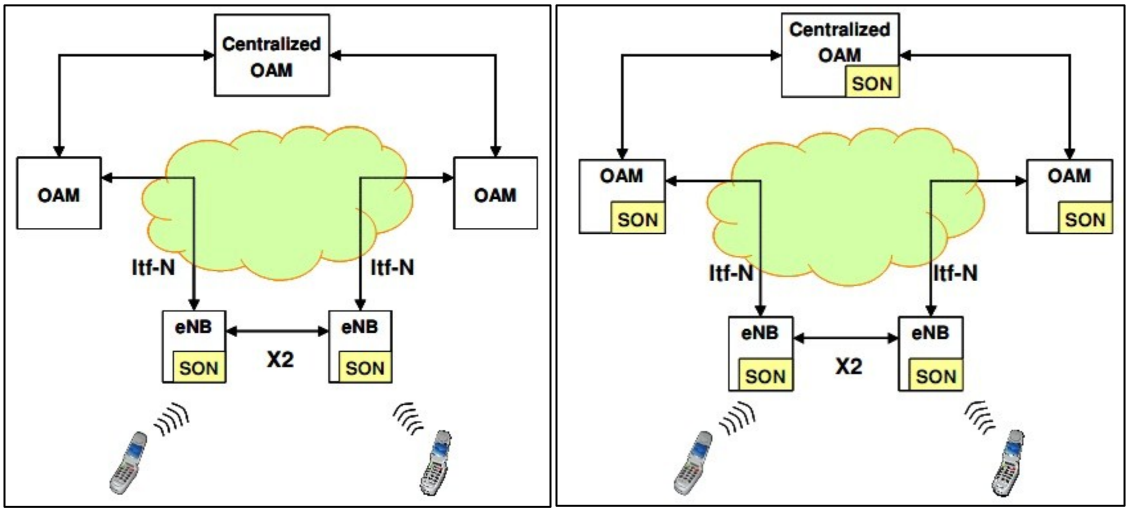


Figure 4.4: SON Architectures – Distributed and Hybrid [4]

## 4.4 SON Use Cases [5] [30]

Up to June 15, 2010, there are already nine use cases approved on 3GPP meetings. Most of them are included in 3GPP TR36.902. The use cases are defined and some of the solutions are still in discussion. Following are some of the use cases.

### *4.4.1 Coverage and Capacity Optimization [4] [30]*

A typical task in day to day network optimization is to optimize its coverage and capacity. The traditional and the most common method are to find the problems by conducting drive tests and use planning tools to define the possible solutions. The purpose of this use case is to find out the coverage and capacity problems automatically through the eNB and UE measurements. It minimizes the human involvement and reduces the feedback delay. The objective here is to optimize the network coverage and maximize the system capacity.

The expected results are presented as follows:

- Continuous coverage
- Increased system capacity
- Reduction in Interference
- Improved cell edge performance
- Minimization of drive tests
- Minimal human involvement in network management and optimization
- Self-healing in case of equipment (e.g. eNB) failure by automatic reconfiguration of surrounding eNBs.

Following are the possible solutions for Coverage and Capacity Optimization. The input variables of Coverage and Capacity Optimization function can be some or all of the below:

- Signal strengths of current and neighboring cells measured in UE.
- UE signaling/reporting
- Timing Advance (TA) parameter.

- Radio Link Failure (RLF) counters
- Coverage triggered mobility counters
- Traffic load distribution measurements

The output consists of optimized radio parameters, which may include downlink transmit power, downlink Reference Signal Power Offset and Antenna tilt.

Following is the procedure:

1. eNB and UE measurement reports are collected.
2. Coverage and capacity problems are detected.
3. Problems detected in the above step are defined, described and passed to a planning tool. The planning tool adjusts the radio related parameters to solve the problems and optimize the coverage and capacity of the system.
4. Adjusted parameters are given to Coverage and Capacity Optimization function.
5. The Optimization function updates the parameters, which are used to deploy and operate the system.

### *4.4.2 Energy Savings* [4] [30]

A critical cost factor for the network operator is the expenses incurred on energy. Cost reduction on energy expenses could be obtained if the capacity offered by the network and the traffic requirement matches at a given instant, almost always.

The main objective of this use case is to achieve energy savings on the basis of cell on/off states. The expected outcome can be defined as cost reductions on OPEX through energy savings.

### *4.4.3 Interference Reduction* [4] [30]

One way to achieve improvement in the network capacity is by interference reduction. This can be done by switching off those cells which are not in use for some time, particularly the home eNBs or the Femtocells. They can be switched off when no activity is detected on the cell for some time.

The objective here is to reduce the interference by switching the cell on/off. The expected outcome is increase in cell capacity and quality through interference reduction.

### *4.4.4 Mobility Robust Optimization* [4] [30]

The process of manually setting up the HO parameters in current 2G & 3G systems is a very time consuming task. Also, manual process proves too costly to update the mobility parameters after the initial deployment.

In some cases, it has been observed that RRM in one particular eNB was able to detect problems and correct the mobility parameters, but there have also been examples where RRM in another eNB cannot resolve problems. Following are the troubleshooting steps:

- *Identification of non-suitable neighbors and its avoidance:* There may be times when source eNB may not be able to detect when a handover was performed to a non-suitable cell. One example of this is RLFs that occur shortly after UE has connected to the target cell.
- *Identify problematic settings of cell selection/reselection parameters:*
- *Minimize handovers immediately after initial RRC connection establishment*: In cases when the idle and active mode mobility parameters are not properly aligned / matched, it may result in large quantity of unwanted handovers shortly after the UE has transitioned from idle to active mode. For such scenarios and cases where the number of handover exceeds an acceptable level, it would be advantageous to control this situation.

The main objective here is to automatically adjust the mobility parameters for cases where the parameters cannot be adjusted automatically by RRM.

Following are the expected outcomes:

- Non-suitable neighbors are identified and avoided.
- Problematic settings of cell are corrected.
- Immediate handovers after initial RRC connection establishment are minimized.

The possible solutions of the Mobility Robust Optimization are first to detect the problems and then to adjust appropriate parameters.

- Problem detection

One solution to detect the mobility problems is to set two counters for each pair of cells to evaluate their pair relationship. One is called too early, which means handover occurs too early between this pair. The other is called too late, which means the handover should be performed earlier between this pair. The eNB collects the counters of each pair of cells and sends them to a centralized entity in OAM. OAM is responsible for adjusting the parameters based on the counter information. Here some scenarios are given to show how to use the two counters.

1. Rapid handover between three cells

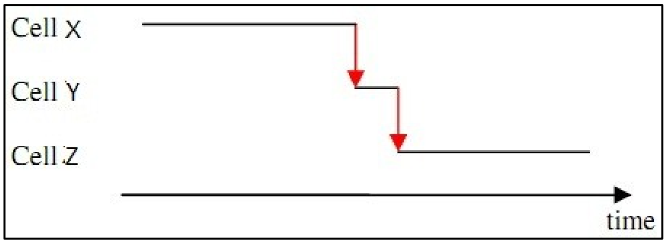


Figure 4.5: Rapid handover between 3 cells [4]

UE performs handover to cell Z shortly after it performs handover from cell X to cell Y. It means that it would be better if UE performs handover directly from cell X to cell Z. So it indicates that the handover is:

- Too early between cell X and cell Y
- Too late between cell X and cell Z

2. Radio Link Failure shortly followed by a radio reestablishment

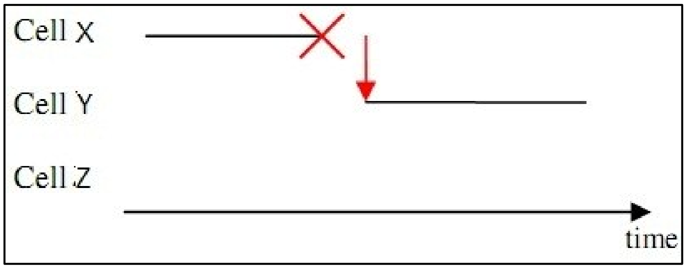


Figure 4.6: RLF followed by link re-establishment [4]

Shortly after the radio link failure in cell X, UE reselects cell Y and reestablishes a radio link. It means that the handover from cell X to cell Y should be performed earlier, before radio link failure. So it indicates that the handover is:

- Too late between cell X and cell Y

3. Radio Link Failure shortly after a handover

Shortly after UE performs the handover from cell X to cell Y, the radio link fails. Then UE reselects cell Z and establishes a new radio link. It means that cell Z is the perfect neighbor for cell X.

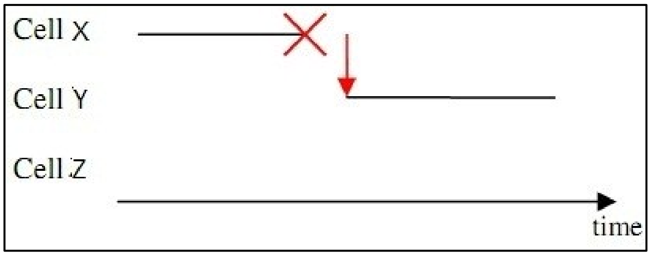


Figure 4.7: RLF shortly after Handover [4]

UE should perform the handover from cell X to cell Z. So it indicates that handover is:

- Too early between cell X and cell Y
- Too late between cell X and cell Z

4. Access failure

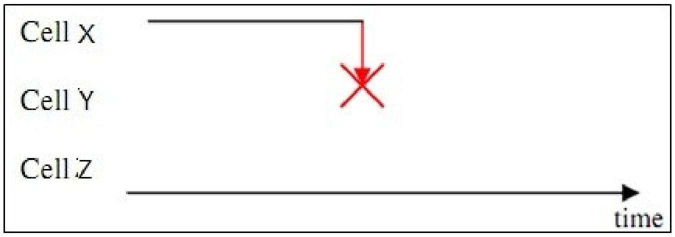


Figure 4.8: Access failure [4]

When UE performs handover from cell X to cell Y, it fails to access to cell Y. It means that the handover has been performed too early. So it indicates that handover is too early between cell X and cell Y.

### *4.4.5 Mobility Load Balancing Optimization* [4] [30]

Self-optimization of the intra-LTE and inter-RAT mobility parameters to the current load in the cell and in the adjacent cells can improve the system capacity compared to static/non-optimized cell reselection/handover parameters and can minimize human intervention in the network management and optimization tasks.

The load balancing shall not affect the user QoS negatively in addition to what a user would experience sat normal mobility without load-balancing. Service capabilities of RATs must be taken into account, and solutions should take into account network deployments with overlay of high-capacity and low-capacity layers where high-capacity layer can have spotty coverage. Objective: Optimization of cell reselection/handover parameters to cope with the unequal traffic load and minimize the number of handovers and redirections needed to achieve the load balancing.

Expected outcome:

- According to the cell reselection and handover mechanisms, part of the UEs at the cell border reselect or hand over to the less congested cell;
- In the new situation the cell load is balanced.
- Increased capacity of the system.
- Minimized human intervention in network management and optimization tasks.

Possible Solutions: To implement Mobility Load Balancing Optimization, the following actions need to be executed:

- Load is measured for each cell in its monitoring eNB. Load information is exchanged between eNBs over X2 interface.
- An algorithm is applied to identify the need to distribute the load between two adjacent cells.
- Handover and/or cell reselection parameters are adjusted in both cells to enable the load balancing and at the same time avoid Ping-Pong effects.

- *Load definition*

The definition of load here has not been decided yet. It can be radio load, transport network load or even the processing load. Even for radio load, it can be split between uplink load and downlink load or split among different QCIs. The definition of load influences the algorithm to distribute the load.

- *Algorithm for load balancing*

An algorithm needs to be defined as to when to balance the load. Due to the different possibilities of load definition, the algorithm can be based on radio load or transport network load or both of them. If the radio load is concerned as the most important factor, it should also be decided whether to differentiate among QCIs. For example, if GBR services have been overloaded but non-GBR services have not, will the load be balanced? If transport network load is also considered with the radio load, which one will have higher priority?

So the problem is how to define the overload situation.

- *Parameters to be adjusted*

According to the cell reselection criterion, UE calculates the R values for serving cell and neighbor cells, ranks the R values and selects the cell with the highest R value. The calculation is defined in 3GPP TS36.304 as follows:

$$\boxed{\begin{aligned} R_s &= Q_{meas,s} + Q_{hysts} \\ R_n &= Q_{meas,n} - Q_{offsets,n} \end{aligned}}$$

'*Qmeas,s*' and '*Qmeas,n*' are the RSRP (Reference Signal Received Power) measured by UE for serving cell and neighbor cells respectively. '*Qhysts'* specifies the hysteresis value for ranking criteria. '*Qoffsets,n'* specifies the offset between serving cell and neighbor cell.

If '*Qhysts*' changes, it will affect the selection relation between serving cell and all the neighbor cells. So if only one pair of cell relation needs to adjusted, it is better to tune '*Qoffsets,n*'.

To avoid mobility problems, the '*Qoffsets,n*' parameter between two cells would be tuned within a proper range. So one of the outputs of Mobility Robust Optimization may be the optimized range of '*Qoffsets,n*' values.

### *4.4.6 RACH (Random Access Channel) Optimization* [4] [30]

In LTE, RACH (Random Access Channel) is an uplink unsynchronized channel, used for initial access or uplink synchronization. The triggers for Random Access procedure include:

- Connection setup
- Radio Link Failure
- Downlink data transmission in uplink unsynchronized state
- Uplink data transmission in uplink unsynchronized state
- Handover

So the Random Access procedure performance influences the call setup delay, handover delay, data resuming delay, call setup success rate and handover success rate. Besides, physical resources for RACH are reserved for its special use. So the configuration for RACH influences the capacity of the whole network.

Necessity for RACH optimization: The performance of Random Access performance is evaluated by its delay and success rate. The performance depends on following factors:

- Population under the cell coverage;
- Call arrival rate;
- Incoming handover rate;
- Whether the cell is at the edge of a tracking area;
- Traffic pattern, as it affects the DRX (Discontinuous Reception) and uplink synchronization states, and hence the need to use RACH.

These factors are affected by network configurations, such as antenna tilt, transmission power and handover threshold, and also by the load of network. If network configurations or load is changed, the performance of Random Access procedure may change greatly, which influences the performance of other procedures, such as call setup, data resuming and handover. Therefore the automatic optimization of RACH would be beneficial.

Possible RACH optimization algorithm: The configurations of RACH include:

- RACH physical resources
- RACH preamble allocation for different sets (dedicated, random-low and random-high)
- RACH persistence level and back-off control
- RACH transmission power control

Measurements are done in eNB, recording random access delay, random access success rate and random access load. The random access load can be indicated by the number of received preambles in a cell in a time interval. It is measured per preamble range (dedicated, random-low and random-high), and averaged over the PRACHs configured in a cell.

Thresholds are set separately for random access delay and success rate. If either of the thresholds is reached, RACH optimization is triggered. First, Random access load is analyzed to check if the random access is overload in any of the three preamble ranges. If one of them is overload, RACH preambles are reallocated among these three preamble ranges. If all of them are overload, more physical resources need to be reserved for RACH. If none of them is overload, other parameters need to be adjusted, such as increasing the transmission power step and distributing the back-off time in a wider range.

### *4.4.7 Automatic Neighbor Relation (ANR)* [4] [30]

In the context of LTE, it is necessary to set up the neighbor relation automatically as much as possible. Because the next generation mobile network is growing more and more complex, it will cause a lot of efforts to configure the neighbor relation relying on traditional configuration methods. ANR function aims at automatic setting of neighbor relation. ANR function relies on UE to report the cells that it has detected but not in the neighbor list. According to the standards, the UE measures and reports the following types of cells:

- The serving cell.
- Listed cells, i.e. cells that are indicated by the E-UTRAN as part of the list of neighboring cells (i.e. as measurement object).
- Detected cells, i.e. cells that are not indicated by the E-UTRAN but detected by the UE. However, E-UTRAN does indicate the carrier frequency.

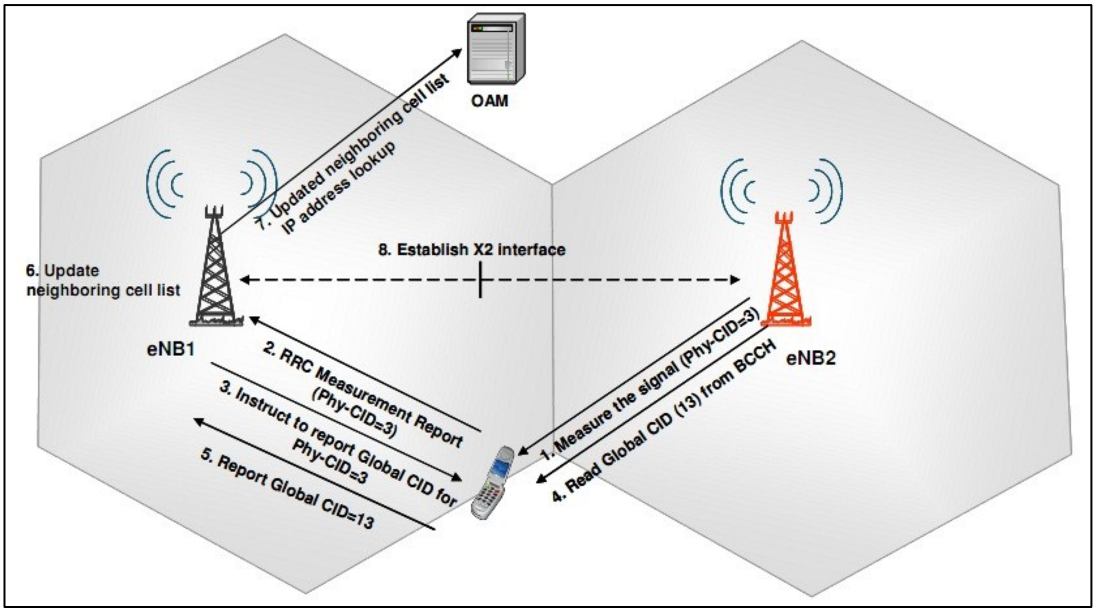


Figure 4.9: ANR Procedure [4]

So the detected cell can be a LTE cell within the same frequency or a LTE cell with a different frequency or even a cell belonging to another RAT. To detect inter-frequency cells or inter-RAT cells, eNB needs to instruct UE to do the measurement on that frequency.

- <u>ANR Procedure</u>

Figure 4.9 gives an example of intra-RAT ANR procedure.

1. UE does the measurement according to the measurement configuration set by E-UTRAN. In this example, UE detects an E-UTRAN cell with Physical ID 3.
2. UE sends the measurement report to the serving cell, using Physical ID to identify different E-UTRAN cells. Here, UE includes the measurements of the cell with Physical ID 3.
3. eNB receives the report and instructs the UE to report Global Cell ID for the cell with Physical ID 3.
4. UE gets the Global Cell ID by reading the BCCH (Broadcast Control Channel) of the detected cell.

5. UE reports the Global Cell ID to the serving cell.
6. The serving eNB updates the neighbor cell list.
7. The serving eNB sends the updated neighbor list to OAM and gets the IP address of the new detected cell from OAM.
8. If required, the serving eNB will establish a new X2 interface with the target eNB.

### *4.4.8 Mobility robustness optimization* [8]

Mobility Robustness Optimization (MRO) encompasses the automated optimization of parameters affecting active mode and idle mode handovers to ensure good end-user quality and performance, while considering possible competing interactions with other SON features such as Automatic Neighbor Relation (ANR) and Load Balancing (LB). While the goal of MRO is the same regardless of radio technology (optimized end-user performance and system capacity), the specific algorithms and parameters vary with technology. The description below is for LTE Release 8 and 9, with its single-link (no macro diversity) approach and X2 interface between eNodeBs.

Whether a distributed or centralized MRO function is implemented it is assumed there will be centralized OA&M control for an operator to enable/disable and configure relevant algorithm settings.

Use Case description:

Manual setting of HO parameters in current 2G/3G systems is a time consuming task and is considered too costly to update the mobility parameters after the initial deployment. For some cases, RRM in one eNB can detect problems and adjust the mobility parameters, but there are also examples where RRM in one eNB cannot resolve problems:

Incorrect HO parameters can negatively affect the user experience and waste network resources by causing Ping-Pong HO, HO failures and Radio Link Failures (RLF). While HO failures that do not lead to RLFs are often recoverable and invisible to the user, RLFs caused by incorrect HO parameter settings have a combined impact on user experience and network resources. Therefore, the main objective of mobility robustness optimization should be reducing the number of HO-related radio link failures. Furthermore, non-optimal configuration of handover parameters, even if it does not result in RLFs, may lead to serious degradation of the service performance. Example of such a situation is incorrect setting of the HO hysteresis, which may be the reason for either Ping-Pong effect or prolonged connection to non-optimal cell. Thus the secondary objective will be reduction of the inefficient use of network resources due to unnecessary or missed handovers.

HO-related failures can be categorized as follows:

- Failures due to too late HO triggering
- Failures due to too early HO triggering
- Failures due to HO to a wrong cell

Additionally cell-reselection parameters not aligned with HO parameters may result in unwanted handovers subsequent to connection setup, which should be avoided by parameter adjustments done by MRO function.

## 4.5 Study of Handover Parameter Optimization [8]

In the currently deployed networks, only manual tuning of handover parameters is available and possible. The main aim of the handover parameter optimization use case in LTE SON is to automatically tune the handover parameters with minimal human involvement. This process is often related to optimizing the neighbor list and tuning the parameters related to handovers.

This handover optimization procedure impacts the following factors:

- Network capacity

- Network Performance due to user mobility.

The main targets of this process are reducing the number of handovers that are initiated but not completed (HO failures), repeated back and forth handovers between two base stations (Ping-Pong HOs) and calls drops (Radio Link Failures). Therefore handover parameter optimization should work towards improving on these objectives.

As stated above, the basic requirements of any Handover parameter optimization algorithm should be:

- To minimize the handover failures – The SON algorithm should be capable of detecting the HO failures, identify the reason, the impact on the overall network performance and take remedial action.
- To reduce the Ping-Pong handovers – The SON feature needs to understand the situation when the back and forth handovers occur, identify the involved parameters and correct the values of these optimal values.
- To increase the load balancing capability of the network – SON feature should regularly monitor the congestion situation in all the cells and dynamically change the HO parameters of the particular cell as well as neighbor cells to ease the congestion without compromising other cells' performance.

As defined in 3GPP R3-071598[7], HO history information in HO request messages can be basis for optimization. Due to the distributed nature of RRM in LTE, Handover stability and avoidance of Ping-Pong Handover becomes a challenge. Especially in a multi-vendor environment, where the Handover algorithms are going to differ, it could prove cumbersome for operators to tune their networks to achieve a reasonable quality with regards to HO stability and drop rate.

As suggested in [7], Introduction of a hysteresis value allows limiting the number of Ping-Pong Handover by simply adjusting the magnitude of the hysteresis. But hysteresis values have to be carefully adjusted. In case of selecting hysteresis value to low, the number of Ping Pong Hand over will be increased. In case of selecting hysteresis value to high, the number of call drops could be increased.

The reason leading to Ping Pong Handover besides distance is field strength which is highly varying, depending on time and position. This is illustrated in figure 4.10

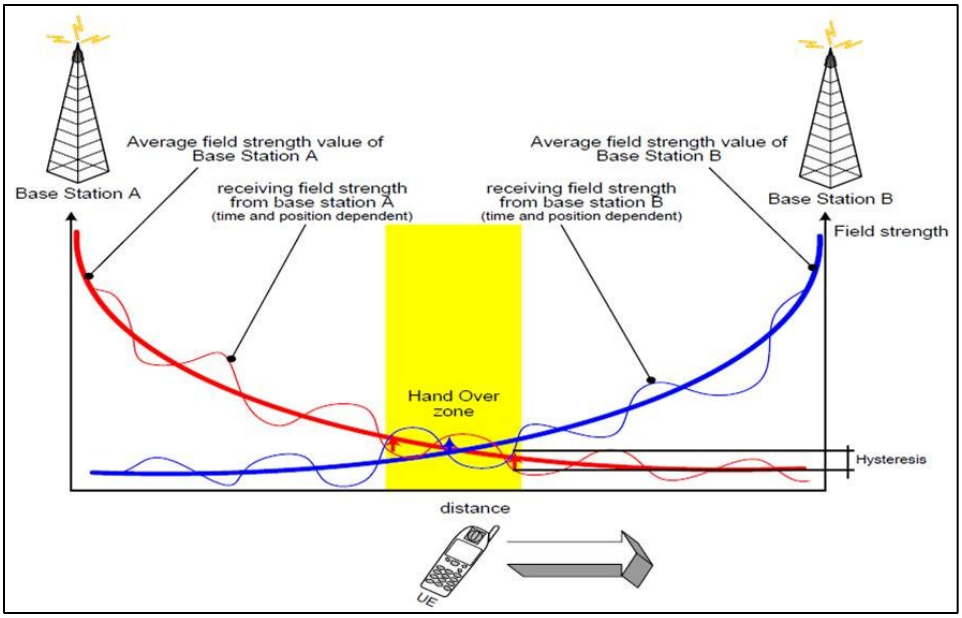


Figure 4.10: Configured hysteresis to low, Ping-Pong Hand Over behavior in Hand Over zone [7]

Regardless of being it a field strength, quality or Power budget Hand over, the highly time and position varying nature of the receiving field strength, requires to include a hysteresis values into the Hand Over decisions.

In the handover preparation phase, measurement reports are sent from the UE to eNB is the prime deciding factor for the eNodeB to make a decision. The report consists of 3 input and 3 output values which is sent to the serving eNodeB.

Also, the threshold value of SNR in source eNodeB should be setup for triggering handover procedure. The formula of the handover triggering is as follows,

Inequality X1-1 (Entering condition)

$$\boxed{Ms - Hyst > Thr}$$

Inequality X1-2 (Leaving condition)

$$\boxed{Ms - Hyst < Thr}$$

Where the variables in the formula are defined as follows:

*Ms*: measurement report from the serving cell, independent of the cell individual offset. Ms is expresses in units of dBm for RSRP and in dB for RSRQ.

*Hyst*: is the hysteresis parameter for this event. Hyst is expressed in dB.

*Thr*: is the threshold parameter for this event. *Thr* is expressed in accordance with units of *Ms*.

As proposed in [8], the Handover Optimization Algorithm will tune the handover control parameters individually for all cells.

Following are the metrics based on which the Handover Optimization Algorithm has been designed in [8]

- Handover Failure Ratio $$\boxed{HP_{HOF} = \frac{N_{ho\_fail}}{N_{ho\_fail} + N_{ho\_succ}}}$$
- Ping Pong Handover Ratio $$\boxed{HP_{pp} = \frac{N_{ho\_pp}}{N_{ho\_pp} + N_{ho\_npp} + N_{ho\_fail}}}$$

- Call Dropping Ratio

$$\boxed{HP_{DC} = \frac{N_{ho\_dropped}}{N_{ho\_accepted}}}$$

- RSRP

$$\boxed{RSRP_{c,ue} = P_c - L_{ue} + L_{fad}}$$

- SINR

$$\boxed{SINR_{c,ue} = RSRP_{conn} - RSRP_{int}}$$

Now we will have a look at the algorithm presented in [8]

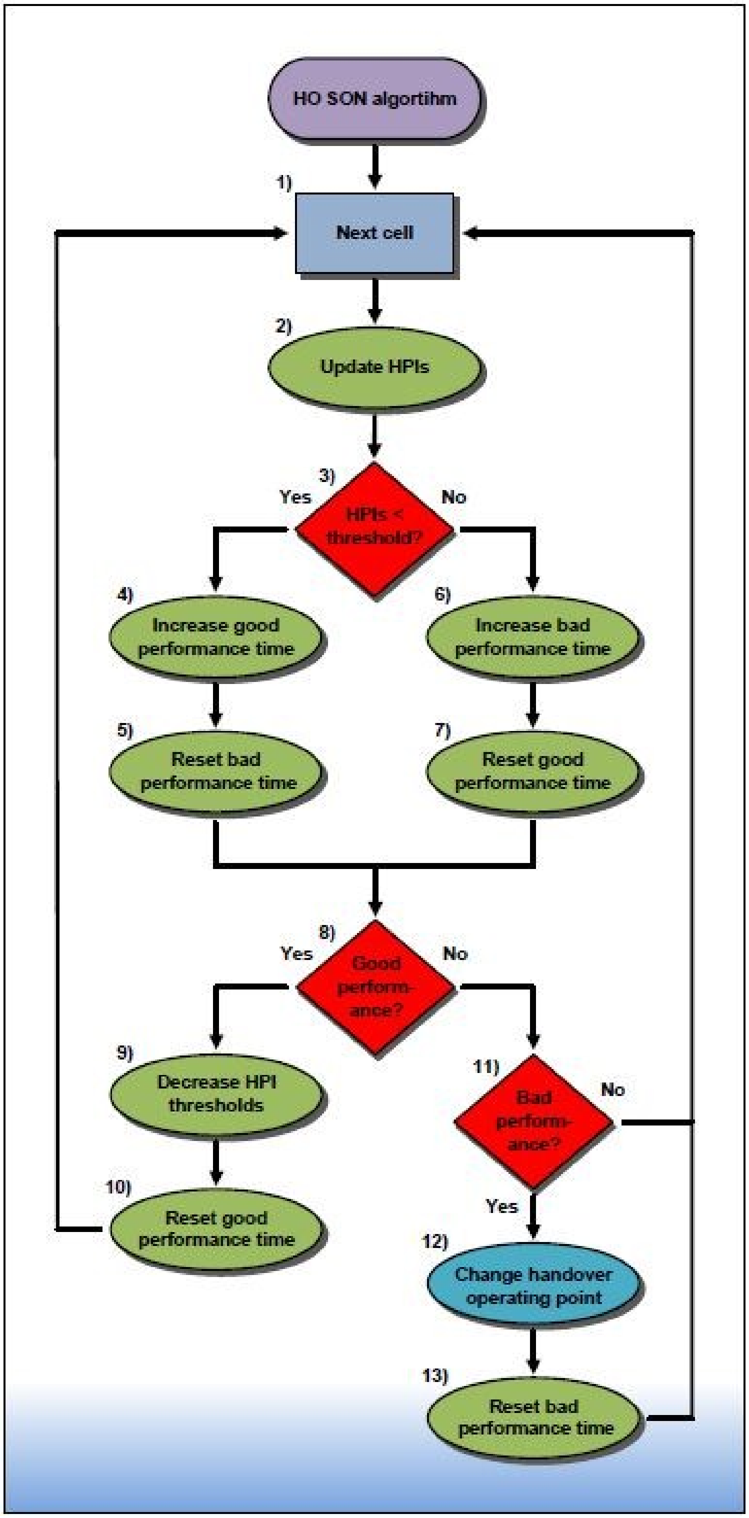


Figure 4.11: HO SON Algorithm. [8]

They define the combination of hysteresis and time-to-trigger as a handover operating point.

According to the handover algorithm shown in figure 4.11, the handover operating points are dependent on the current handover performance and are changed accordingly. In order to adapt to the optimal possible handover performance, the handover performance target thresholds for all handover performance indicators (HPIs), i.e. the handover failure ratio, the Ping-Pong handover ratio and the call dropping ratio. The target thresholds are decreased by 33% if the HPIs stay below the target thresholds for a certain amount of time, called the good performance time. The initial handover performance target thresholds are influenced by an operator policy that defines the importance of the different HPIs. If the performance of more than one HPI is above the HPI target threshold, the handover performance target thresholds are increased by 33% again. If the simulation parameter bad performance time exceeds a given threshold, i.e. the performance of one HPI overshoots the handover performance target for a certain amount of time, the handover operating point of the cell is changed according to the criteria given in Table 4.3. The optimization criteria are derived from system simulations for all valid operating points. The optimization direction, i.e. change of hysteresis and time-to-trigger, is based on the system performance of neighboring handover operating points. The system performance is calculated for every HPI individually. Hence the optimization direction can be leading off from these system simulations. The hysteresis value as well as the time-to-trigger values is changed by one step per handover parameter optimization only. The simulation results on the handover performance are given in the next section.

### *Simulation Results* [8]

The system simulations are based on the simulation parameter settings shown can be looked up from [8]. The operating point with a hysteresis of 6 dB and a time-to-trigger of 0.32 s is the starting operating point for all cells in the network. Later on, a comparison of the handover performance of this fixed operating point to the handover performance of the optimized network.

The operator policy specified in the parameter settings gives the importance of the HPIs. In this case the operator is most interested in avoiding call dropping, i.e. the operator policy value for the HPI is set to 2.

Table 4.3: Optimization Criteria for handover performance indicators [8]

| **Handover Performance Indicator** | **Hysteresis** | **Time to Trigger** | **Optimization** |
|---|---|---|---|
| ***Handover Failure Ratio*** | < 5 db | | ↑ TTT |
| | 5 db - 7 db | | ↑ TTT & ↑ HYS |
| | > 7 db | | ↑ HYS |
| ***Ping Pong Handover ratio*** | < 2.5 db | | ↑ TTT |
| | 2.5 db - 5.5 db | | ↑ TTT & ↑ HYS |
| | > 5.5 db | | ↑ HYS |
| ***Call Dropping Ratio*** | > 6 db | > 0.6 s | ↓ TTT & ↓ HYS |
| | <= 6 db | > 0.6 s | ↓ TTT |
| | > 7.5 db | <= 0.6 s | ↓ TTT & ↓ HYS |
| | 3.5 db - 6.5 db | <= 0.6 s | ↑ HYS |
| | <3.5 db | <= 0.6 s | ↑ TTT & ↑ HYS |

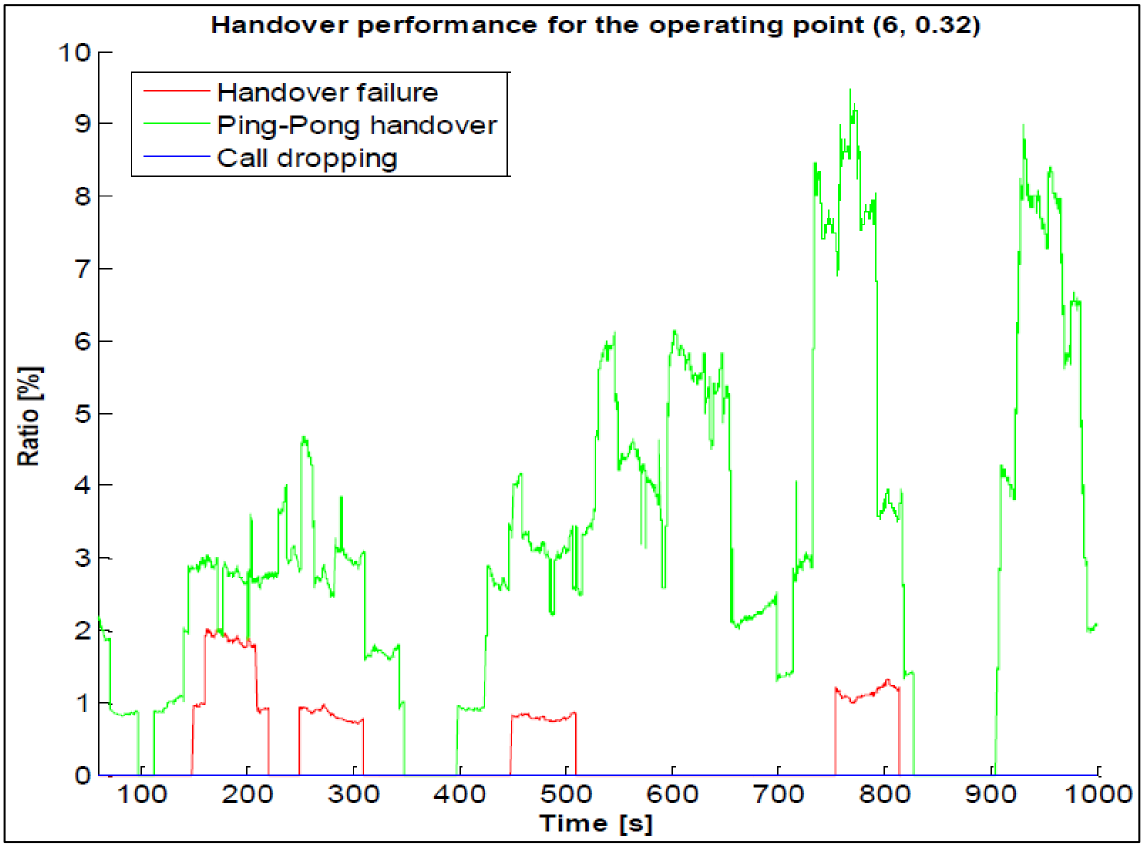


Figure 4.12: Handover performance without SON Algorithm being implemented. [8]

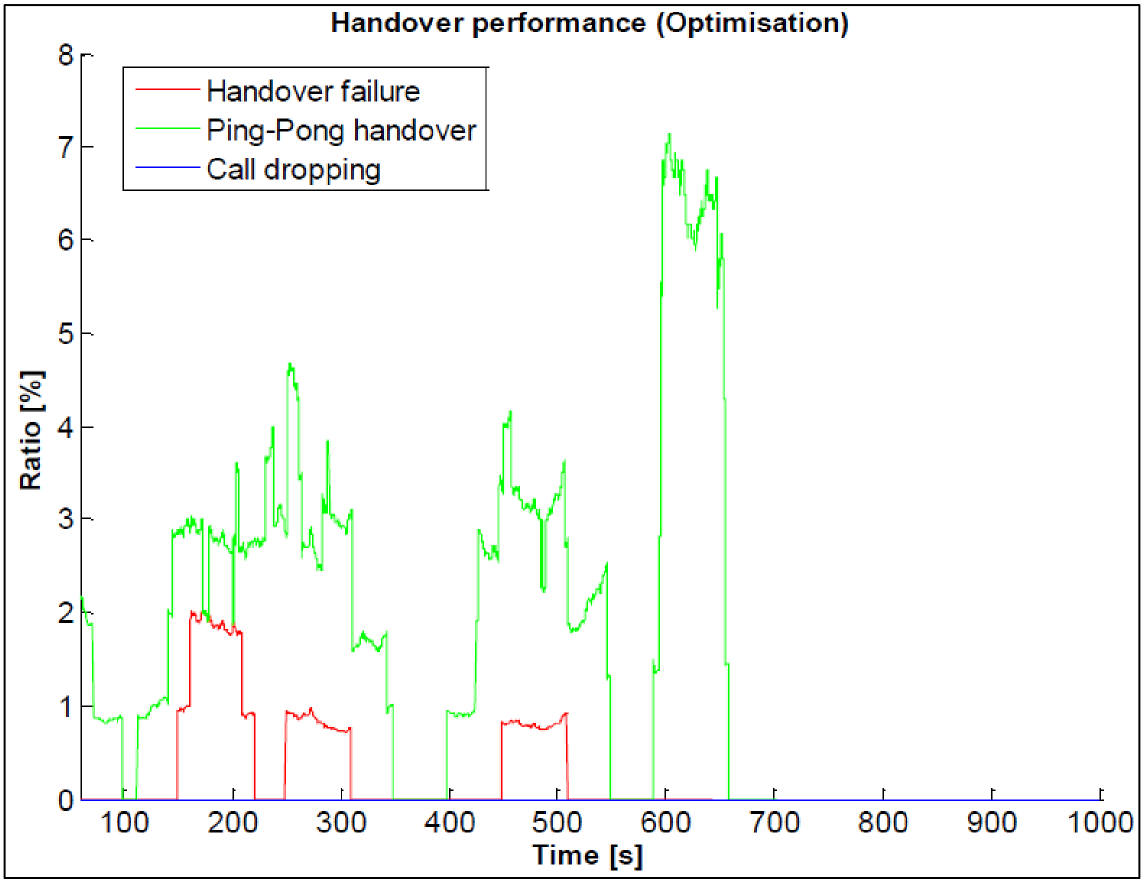


Figure 4.13: Handover Performance after applying SON algorithm. [8]

*Conclusion & Outlook* [8]

The proposed handover optimization algorithm changes the values of the hysteresis and time-to-trigger parameters in an automated manner in response to changes in the network performance. This algorithm takes into account the weighting factor given by the operator policy to different performance metrics (handover failure ratio, call dropping ratio and Ping-Pong handover ratio).

This novel feature makes the SON algorithm flexible and very appealing to operators. The simulation results show that the optimization algorithm increases the system performance significantly. However the current results are limited to the used, realistic, simulation scenario. It has to be proven that the optimization algorithm works for other simulation scenarios as well.

Relevant data for optimization, for both the basic and enhanced solution, are:

- UE local distribution; measured localization data
- Bit error rates
- Radio power receiving levels (UL)
- Radio power receiving levels (DL)
- Traffic load of cells
- Time duration in cells with histogram analyzing
- Radio drops in relation to location
- Radio bad quality in relation to location

Enhanced solutions on HO parameter optimization:

1. Predict the possible handover failures in advance and take action pro-actively to avoid any failures – SON features need to have the capability to gather the user distribution and mobility information. Based on this data, it should predict the possible failures and take evasive actions without causing service deterioration
2. Predict the traffic requirements of cells and adjust the parameters according to the requirements – SON features need to have the capability to predict the near future capacity requirement of cells and determine the possible actions.

The HO Optimization feature has to provide functionalities to improve the HO parameter settings regarding following detailed requirements:

- Perform best radio quality for up and downlink for real-time and non-real-time traffic in the overlapping area of cells to the UE
- Perform the best effort of UE power consumption
- Avoid any bearer drop in the overlapping area of intra-system cells

- At Radio systems edge, perform a HO to other inter radio access technologies, with an optimum time setting before the radio quality or radio power link degradation, leading to interference or perceived customer quality.
- Minimize signaling load with Optimizing/minimizing ping pong HO
- HO decision should be dependent on the QOS profile of users
- Consider the load situation
- Consider service class and parameterization

# CHAPTER 5

# CONCLUSION AND FUTURE WORK

The ever changing and evolving cellular communication market demands for high speed broadband connectivity for mobiles and this future of mobile broadband market depends largely on the success of LTE technology. So far, from the few deployed LTE Networks in the Nordic market, the performance complies with the specifications. We can expect an increase in the data transfer speeds, both uplink and downlink, as and when all the new techniques mentioned in the 3GPP suite of specifications are developed and implemented.

The LTE specifications mention of an evolved NodeB, the NodeB being the UTRAN base- station. I have made an attempt to completely explore the EUTRAN part of the LTE which mainly focuses on the eNodeB and all the incorporated protocols. The EUTRAN Control Plane and EUTRAN User Plane have been explored in this tutorial along with the main protocols like the Radio Resource Control (RRC), Radio Link Control (RLC), Packet Data Convergence Protocol (PDCP), Medium Access Control (MAC) and the Physical Layer (PHY). Due to complete change in the radio part of the LTE, different physical layer specifications are introduced as compared to 3G UMTS.

Then we saw the LTE Self Organizing Network Concept, which is a promising solution as it allows the mobile operators to implement objectives such as robustness, better performance, energy efficiency, continuous coverage, increased system capacity, interference reduction, minimal human involvement in network management and optimization, etc. All this features will prove beneficial to the operators as it provides a reduction in CAPEX as well as OPEX.

A special attention has been given to the mobility robustness use case and a handover parameter optimization algorithm has been discussed. The newly developed handover optimization techniques is smart enough to adjust and tune the parameters which define the process of handover and makes an effort to avoid handover failures as far as possible.

My efforts in this tutorial based on LTE was to try and explain the System Architecture Evolution (SAE), which is the basis on which the project of LTE was started in 2004 and developed during the following years. There’s still a lot to be accomplished in LTE. As a future work, it would be a good learning experience to try and simulate some of the LTE SON use cases in system modelers like OPNET, MATLAB or other capable software.

All eyes are now set on the North American market, where the largest operational LTE network is being deployed and provide mobile broadband experience to the users by the end of year 2010. As per the claims of the operators, it will be important that the live networks perform as per the expectations and lab simulations.

# APPENDIX A

# STREAM CONTROL TRANSMISSION PROTOCOL (SCTP) [13]

We will now look at SCTP or Stream Control Transmission Protocol as given in [13]. SCTP (Stream Control Transmission Protocol) is a reliable and connection-oriented transport layer protocol. It is very similar to the popular and widely used TCP. Similar to TCP, SCTP also employs congestion and flow control mechanisms, detects of data loss, corruption and/or duplication of data and also supports selective retransmission mechanism.

Similar to TCP, SCTP operates in a connected mode, so that an ‘association’ (the term for connection in SCTP) needs to be set up between peers before data transmission can occur. In SCTP, an association is defined by a (Source IP, Source Port, Destination IP, Destination Port) group.

When comparing TCP and SCTP from a functional perspective, SCTP provides two key features which TCP does not support:

- Multi-streaming.
- Multi-homing.

In the SCTP domain, a stream is a unidirectional sequence of user messages to be delivered to upper layers. As a consequence, bi-directional communication between two entities involves at least a pair of streams, one for each direction. The multi-streaming is the feature from which the STCP name is actually derived. It allows setting up several independent streams between two peers. In such a case, when a transmission error occurs on one of the stream, it does not affect data transmission on the other streams.

In contrast, TCP only provides one stream for a given connection between IP peers, which may cause additional data transmission delay when a packet or group of packets is lost. When a transmission loss occurs on a TCP connection, packet delivery is suspended until the missing parts are restored, as in-sequence data delivery (or data sequence preservation) is a key TCP feature.

This important characteristic of TCP is not necessary in all cases. For example, in the case of a multimedia document such as a Web page, multiple parallel streams may be opened in parallel to retrieve the whole page content. Content delivery is more critical than content order for such a kind of application. The same applies to independent signaling flows which are transferred between two network nodes, such as the MME and the eNodeB. The delivery order of each signaling flow (e.g. corresponding to one mobile-network connection) needs to be preserved; however, all the flows can be delivered independently.

The other core added value of SCTP is multi-homing. This allows a SCTP endpoint to be reached through multiple network addresses. The interest of multi-homing is about redundancy, as it improves the resilience when network failures occur. In case of transmission errors, retransmitted packets may be sent to alternate addresses in order to increase the probability of successful transmission.

Of course, there are other differences between TCP and SCTP. The two which are worth mentioning here are:

- SCTP framing: SCTP works at the message level whereas TCP is an octet stream protocol. This was one of the main reasons for 3GPP adoption in E-UTRAN signaling transport.
- SCTP built-in cookie-based protection against denial of service attacks, which is described hereafter.

### *Why SCTP?* [13]

There have been some debates in 3GPP working groups on which transport protocol would be the most suitable for the E-UTRAN Control plane, to support signaling message exchange between network nodes. Among the three most obvious candidates, UDP was quickly ruled out as not being reliable enough. From a high-level perspective, SCTP and TCP are quite close to each other, as they both support reliable and ordered data delivery, as well as congestion control to regulate network data flow. What really made the difference in favor of STCP was the following:

- The multi-streaming feature.
- The fact that SCTP is message-oriented and supports framing of individual messages as opposed to TCP, which is octet stream-oriented and does not preserve transmitted data structure. In SCTP, messages are transmitted as a whole set of bytes (provided the maximum length is not reached) which helps to improve transmission efficiency.

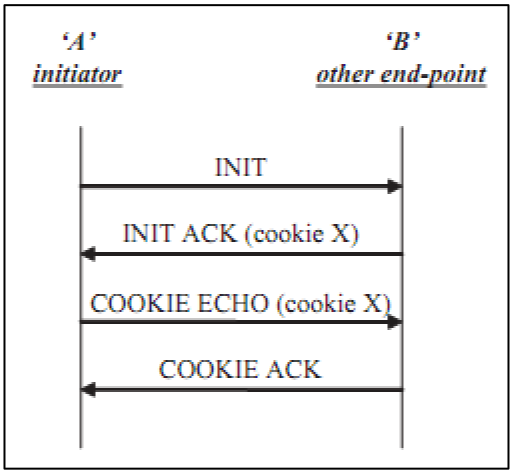


Figure A.1: The steps of SCTP association setup. [13]

- The resilience of SCTP against some types of denial of service attack TCP is vulnerable to, like the 'SYN flood.

The 'SYN flood' attack is special kind of attack, causing a TCP endpoint to receive a connection request to reserve a resource context and memory for some incoming connections that will never be fully set up by the initiator. At some point, this process can exhaust all memory or processing resources in the receiver. To counter this, SCTP makes use of a cookie mechanism. The actual seizure of association resource is only performed once the initiator successfully answers with the correct cookie.

For illustration, Figure describes the four steps of a SCTP association establishment. On reception of the INIT message, the receiver builds a cookie and sends it to the initiator using the INITACK message. To enable the association, the initiator must answer a COOKIE ECHO containing the same cookie as received in the INITACK. Resource reservation related to the association is only performed by the 'B' side on reception of a COOKIE ECHO. At the end, the COOKIE ACK is sent back to the initiator to acknowledge the association setup. Resource attack is prevented by building the cookie in a special way. In principle, the receiver of the INIT message is using a secret key and a hash mechanism to create it, so that on reception of the COOKIE ECHO, it can then validate that the cookie was actually previously generated by the receiver.

This protection is based on the fact that the receiving entity (the B part in the diagram) does not reserve resources or keep context pending during the INIT phase. Resource activation is only performed when a valid COOKIE ECHO message is received. Of course, this assumes the rogue initiator does not process the answers, which is generally the case for denial of service attacks.The cookie structure is not fully specified by the SCTP recommendation, but it may possibly contain a Timestamp corresponding to its creation time.

### *SCTP in E-UTRAN Transport Network*

In the S1 interface (and the same applies to the X2 interface described below), SCTP is used over the usual IP network layer. There is only one association per instance of S1 interface (or eNodeB to MME' relation). Over this association, one SCTP stream is used for all common procedures – such as the paging procedure – between two pieces of equipment. Regarding all dedicated procedures – which include all procedures which apply to a specific communication context – they all are supported over a limited number of SCTP streams.

A much more detailed explanation along-with some demos can be found at http://www3.rad.com/networks/2003/sctp/index.htm

# APPENDIX B

## S1 FLEX MECHANISM [13]

In traditional 2G and 3G cellular networks, the connectivity between the Core Network and Access Network part was defined as a one-to-multi hierarchical relationship: a Core Network node (either the MSC on the Circuit domain or the SGSN in the Packet domain) serves a set of radio Controllers (the 2G BSC or the 3G RNC), and a given controller is only assigned to one Core Network node within a domain. In other words, each Core Network node is connected to its own set of radio Controllers, having no intersections with other sets.

In Release 5 of the 3G/UMTS standard, a new feature was introduced, allowing more flexibility in the inter-connection between Access and Core nodes, breaking the usual network hierarchy. This feature has been introduced from the beginning in the EPS standard and is known as S1-flex

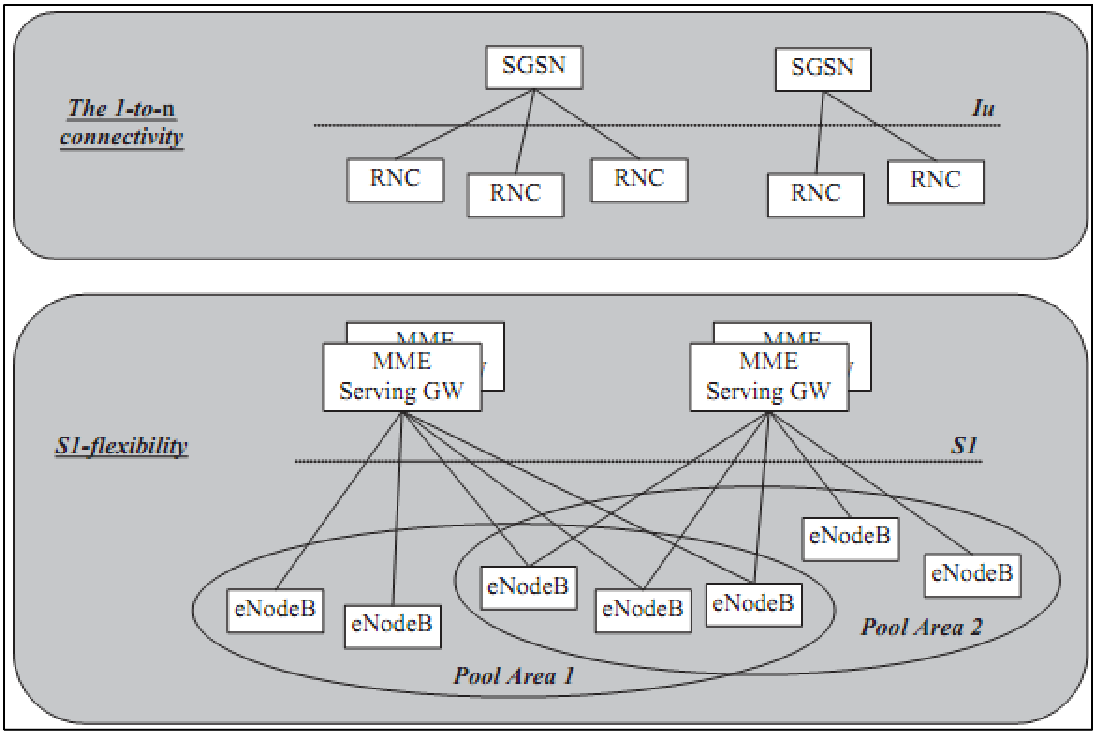


Figure B.1: Traditional Access–Core connectivity and S1-flex. [13]

As represented in Figure B.1, S1-flex allows an eNodeB to be connected to more than one MME or S-GW node. In this picture, MME and S-GW are combined in one node for simplicity, but the S1 flexibility applies to both MME and S-GW, independently. This picture also introduces the notion of pool area. Although an eNodeB can be connected to several MME, a terminal is only associated to one MME at a time, due to the fact that the sessions of a subscriber are always under the control of a single Core Network MME node.

S1 flexibility has many advantages:

- By extending the usual service area seen by a Core Network node, the S1 flexibility allows the reduction of the number of inter-Core Network node handover procedures (in Connected mode) or Tracking Area updates (in Idle mode). This extends the possibility for a MME to maintain the connectivity with a moving terminal, as long as the terminal remains in the same pool area. As a consequence, the S1 flexibility helps to reduce the HSS load generated by the change in MME.
- The S1 flexibility also helps to define network architectures shared by different operators. As an example, part of the E-UTRAN network – represented by a set of eNodeB in a given geographical – may be simultaneously operated by two different business entities. In such a case, when a terminal attempts to register, the eNodeB can forward the initial registration message to the MME, which corresponds to the network operator of the subscriber.
- S1 flexibility allows the network to become more robust to Core Node failure, as the loss of one Core Network node will be compensated for by other nodes associated to the same pool areas. This increased service availability is, however, not dynamic, meaning that in case of failure, on-going communication sessions are not automatically transferred to a new node.

And, at last, S1 flexibility has some advantages as regards to capacity upgrade and network load management. Opening the possibility for an eNodeB to be connected to more than one MME allows balancing and possibly redistributing the load by directing incoming terminal connection requests to less loaded Core Network nodes. The S1 flexibility relies on a new field of information which is actually a sub-part of the temporary subscriber identity (S-TMSI). This new field uniquely identifies a MME in an area served by multiple MMEs and is then used by the eNodeB to direct an initial connection request towards the right MME – or set of MME – in case of network sharing, or to send the terminal initial message to the MME it was registered to.

# APPENDIX C

## CAPEX [1]

## C.1 Introduction [1]

As we can see from [1], CAPEX encompasses the investments needed in order to create future benefits, which includes Radio Access Network (RAN) equipment (eNodeB), core network (MME and S-GW), transmission and transport network (e.g., Ethernet and microwave networks), service layer, equipment roll-out (e.g., integration and testing and in-house solutions), and construction (e.g., site acquisition and civil works). In general there are tradeoffs between QoS and/or GoS, and CAPEX. Typically, QoS and GoS decrease with increasing site-to-site distance. This results in decreased CAPEX since an increasing site-to-site distance implies fewer RAN and transport network equipment.

An approach to estimate the number of network elements (RAN equipment) needed to satisfy requirements on QoS and/or GoS is presented in Section C.2. The introduction of SON may increase the cost of network elements and this will be discussed in Section C.3. A methodology for assessing self-organization algorithms based on their overall CAPEX savings is presented in Section C.4.

## C.2 Estimating number of network elements [1]

Key component of the overall CAPEX is the purchase of network elements, in particular the base stations, needed to provide sufficient capacity. Starting point of the approach, besides assumptions regarding propagation environment, service mix, traffic characteristics, spatial traffic distribution, quality/grade of service requirements, is an assumption regarding the traffic demand per km$^2$

- The approach is most readily formulated and implemented if we assume spatially uniform traffic demand and hence apply hexagonal network layout. The idea is then to start with a widely stretched network with very large cells, and compress this layout until the cells are just small enough to handle the correspondingly captured traffic load with sufficient grade/quality of service. The thus obtained cell area is then easily applied to determine the number of base stations that is needed to cover a given service area, e.g., The Netherlands. Multiplying this number with an assumed typical price of a base station gives us an estimate of the base station-related CAPEX. Given some a priori agreed ratio of supportable base stations per other network element (e.g., MMEs), and the associated purchasing price, we can extend the CAPEX estimate to cover other network elements as well. It is noted that the proposed approach can in principle be applied for any of the above-described capacity definitions, as long as the traffic demand per km$^2$ is expressed in corresponding units. For instance, in case of the 'Maximum supportable traffic load' capacity definition, the capacity is expressed in a supported maximum aggregate call arrival rate per cell, and hence the traffic demand per km$^2$ should then also be expressed in terms of a call arrival rate per km$^2$

- One outcome of self-optimization may be a decrease in CAPEX since less number of sites or cells may be needed to provide the same QoS and/or GoS as a non-optimized network. Self-organization may result in enhancements in QoS and/or GoS and, consequently, an increase in the site-to-site distance and less investment in equipment for capacity extension compared to a non-optimized network. Another aspect is that SON, e.g., by avoiding incorrect operations or by temporarily switching off unused cells or antennas, might positively impact the average lifetime of the equipment. In order to capture changes in lifetime an average lifetime of the equipment should be deduced in both cases if possible and the required CAPEX per year should be compared. Of course, such effects might be hard to quantify in the current research phase, but at least this aspect should be discussed qualitatively.

## C.3 Impact of SON on CAPEX [1]

The introduction of self-organization (configuration, optimization, and healing) may, however, also result in increases in equipment cost. This depends on the self-organization algorithm and a set of factors associated with the algorithm such as computational complexity, network bandwidth requirements to other nodes (over, e.g., X2 and S1), and additional costs related to needed site equipment, e.g., electrical antenna tilt (which is often omitted due to cost savings), and additional circuitry for enabling power savings.

In general, the computational capability (hardware) of a cell or site needs to be dimensioned according to the offered traffic. With the introduction of self-organizing algorithms the computational capabilities of eNodeBs must also encompass self-organization algorithms executing in the eNodeBs. The execution time of a particular algorithm typically depends on the size n of the input data, e.g., number of measurements, and can be asymptotically logarithmic (log n), polynomial ($n^a$), or even exponential ($a^n$). An analysis of asymptotic execution time gives an insight in the processing demand of an algorithm. In addition, self-organization algorithms may require extensive network monitoring in order to ensure that these algorithms are operating well. Reports and counters need to be sent to OSS over the backhaul. A self-organization algorithm may also interact heavily with other nodes in the network (e.g., other eNodeBs) resulting in higher transmission costs. Ideally, these aspects must also be taken into account when evaluating CAPEX savings. The underlying assumption is, however, the savings due to reduced number of sites and RAN equipment is larger than corresponding increases in equipment costs due to additional complexity.

### C.4 Overall analysis of CAPEX [1]

Estimating savings due to enhanced QoS and/or GoS, and increases in equipment cost due to additional software running on the eNodeBs may be difficult to quantify exactly. Instead we have to resort to approximations or qualitative assessments. The following metrics can be used when assessing the CAPEX saving as a result of introducing a particular self-organization algorithm. Here we assume that a network without self-organization is configured with a set of default or standard parameters yielding acceptable performance.

- The number of sites or cells needed to cover a certain area served by a network with and a network without self-organization. A cost associated with a site or cell may be used to estimate CAPEX savings.

- The number of sites or cells needed to provide a certain QoS in a given area served by a network with and a network without self-organization. A cost associated with a site or cell may be used to estimate CAPEX savings.

The following issues should be considered when assessing increases in equipment costs. Since associating a cost with the issues given below is difficult (if feasible at all), a qualitative assessment should be carried out considering:

- The estimated execution time, i.e., to determine the asymptotic execution time of an algorithm.
- The estimated bandwidth, i.e., to determine the bandwidth required over the transport network to other eNodeBs, MME/S-GW, and OSS. This can be estimated in terms of the number of packets sent and the estimated payload of each packet.
- The need for additional equipment, e.g., electrical tilt.

The third issue mentioned above (need for additional equipment), requires an understanding of what RAN functions are typically optional and may be purchased if needed.

The CAPEX savings and expenses listed above should be used as input to form an overall assessment of CAPEX savings associated with a particular self-organization function. For example, if two self-organization algorithms perform equally in terms of reduced number of needed sites, then execution time and bandwidth requirements may serve as indicator for determining which of the two algorithms performs best in terms of CAPEX savings.

# REFERENCES

[1] INFSO-ICT216284 SOCRATES, “Review of Use Cases and Framework”; Deliverable D2.5, v.1.0, March 2009; http://www.fp7-socrates.org/files/Publications/

[2] IEEE Communications Magazine, “Self-Configuration and Self-Optimization for LTE Networks”, February 2010.

[3] TS 32.501, “Telecommunication management; Self-configuration of network elements; Concepts and requirements.”

[4] S, Feng; E Seidel; “Self-Organizing Networks (SON) in 3GPP Long Term Evolution”, Nomor Research GmbH, Munich, Germany, May 2008.

[5] TS 36.902, “E-UTRAN - Self-configuring and self-optimizing network (SON) use cases and solutions.”

[6] R3-061488, “Impact of Self-configuration and self-optimization functionality on architecture & interfaces.”

[7] R3-071598, "Information to be included at Handover request messages, to avoid Ping-Pong Handover”

[8] T Jansen, I Balan, I Moerman, T Kurner; “Handover parameter optimization in LTE self-organizing networks”, SOCRATES, http://www.fp7-socrates.eu.

[9] R3-081165, “Solutions for the Mobility Robustness use case”

[10] 3GPP Technical Specification 36.300, “E-UTRA and E-UTRAN - Overall description (Release 8)”, www.3gpp.org

[11] M. Olsson, et.al, *“SAE and the Evolved Packet Core – Driving the Mobile Broadband Revolution”,* 1st Edition, Elsevier, 2009.

[12] H. Holma, A. Toskala, "*LTE for UMTS-OFDMA and SC-FDMA based radio access",* 1st edition, John Wiley & Sons Ltd., 2009

[13] P. Lescuyer & T. Lucidarme, *"Evolved Packet System (EPS) - The LTE and SAE Evolution of 3G UMTS",* John Wiley & Sons Ltd., 2008

[14] S.Sesia, I.Toufik, M.Baker, *"LTE – The UMTS Long Term Evolution: From Theory to Practice"*, 1st edition, John Wiley & Sons Ltd., 2009.

[15] http://wired-n-wireless.blogspot.com

[16] IEEE Communications Magazine, February 2009.

[17] Rysavy Research, "HSPA to LTE - Advanced: 3GPP Broadband evolution to IMT - Advanced (4G)', 3G Americas, Sept. 2009.

[18] "*Long Term Evolution Protocol Overview*", White Paper, Freescale Semiconductor, 2008

[19] 3GPP Technical Specification 23.978, "All IP Network Specification (Release 7)", www.3gpp.org

[20] 3GPP Technical Specification 33.401, 'System Architecture Evolution (SAE): Security Architecture (Release 8)', www.3gpp.org.

[21] http://www.eventhelix.com/lte/lte-tutorials.htm

[22] 3GPP Technical Specification 36.300, 'E-UTRA and E-UTRAN - Overall description (Release 8)', www.3gpp.org.

[23] "*LTE from A-Z- Technology and Concepts of the 4G 3GPP Standard"*, INACON GmbH, 2009

[24] *"LTE Evolved UTRAN Engineering Overview",* Wray Castle Ltd., 2008

[25] 3GPP Technical Specification 36.410, "E-UTRAN - Architecture description (Release 9)", www.3gpp.org

[26] 3GPP Technical Specification 36.401, "E-UTRAN - S1 layer 1 general aspects and principles (Release 9)", www.3gpp.org

[27] 3GPP Technical Specification 36.420, “E-UTRAN - X2 general aspects and principles (Release 9)”, www.3gpp.org

[28] 3GPP Technical Specification 36.413, “E-UTRAN - S1 Application Protocol (S1AP)”, www.3gpp.org

[29] 3GPP Technical Specification 36.423, “E-UTRAN - X2 Application Protocol (X2AP)”, www.3gpp.org

[30] 3GPP Technical Specification 36.902, “E-UTRAN - Self-configuring and self-optimizing network (SON) use cases and solutions”, www.3gpp.org

# BIOGRAPHICAL INFORMATION

Dhruv Shah is an Electrical Engineering Masters candidate at The University of Texas at Arlington. He holds a Bachelors of Engineering degree in Electronics and Telecommunications from The Mumbai University, India. He presently works at Ericsson, Inc. as an Radio Access Network Engineer. His prime interests are in the broadband wireless technology arena and computer networks. Apart from academia, his interests lie in music, technology and art.