TUTORIAL ON NETWORK QUALITY AND

IMS SERVICES USED

IN LTE

by

AYUSH MAHESHWARI

Presented to the Faculty of 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

ACKNOWLEDGMENTS

It is an honor for me to thank the people who made it possible for me to complete this thesis. I would like to thank my parents and my friends who kept me encouraged throughout this endeavor.

I owe my deepest gratitude to my advisor, Dr. S Gibbs, without whose guidance, patience, enthusiasm and support this thesis would have been incomplete. His experience and knowledge were invaluable in helping me achieve my goals.

I am indebted to my colleagues who shared their knowledge of this technology with me and helped me obtain a sophisticated knowledge of this topic.

I am also grateful to my roommates who had the patience to tolerate my late night working.

November 4, 2010

ABSTRACT

TUTORIAL ON NETWORK QUALITY AND

IMS SERVICES USED

IN LTE

Ayush Maheshwari, MS

University of Texas at Arlington, 2010

Supervising Professor: Ralph Stephen Gibbs

3rd Generation or 3G technologies like WCDMA, HSPA and GPRS are very advanced technologies in themselves. They allow data downlink speeds of up to 7Mbps and will soon be offering up to 14Mbps. The number of people using these networks has crossed 70 million users worldwide and with ever increasing users we need to provide a solution to handle the new and improved services which are being designed daily to work on mobile devices.

Long Term Evolution (LTE), will be one of the two technologies which will meet the ITU's IMT-Advanced requirements for 4G technologies. 3G Partnership Project (3GPP) is the organization behind the design and development of LTE. The design parameters for LTE included a system with higher data rates, low latency and high spectral efficiency. The design that was developed is said to have four times better performance figures than the previous technologies. To achieve these high numbers, LTE is based on Multiple Input Multiple Output (MIMO) antenna technique, inter-cell interference mitigation, low latency structure, Single-

Carrier Frequency Division Multiple Access (SC-FDMA) and Single-Frequency Network (SFN) Broadcast.

In this research documents, we have paid close attention on how services like voice and MBMS have been brought to the users in LTE. Since LTE will be replacing the current telecommunication network, it is very important to see that Voice over LTE (VoLTE) gives better results when comparing with the current telecommunication networks. Also, the high data rates allow many MBMS services to be streamed to the user. LTE also has inherent support for relay systems. In previous technologies, even with high data speeds and good coverage, the users at the cell edge suffered from poor service overall. We have analyzed the conditions that a UE will face when dealing with the new relay architectures.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

CHAPTER 1

LTE NETWORK ARCHITECTURE

1.1 Introduction

LTE was designed keeping in mind the need to change the way the current systems handle voice calls. Unlike its predecessors this technology is completely based on packet switching rather than circuit switching. To accommodate this change, the network architecture had to be redesigned to enable it to perform voice calls as well as data calls with seamless mobility. Packet switching also means taking special care of Quality of Service (QoS), minimal jitter and latency. A packet switched approach for both voice and data calls allow seamless connectivity between the user and the base station (BS). Some of the reasons for SAE are:

- Enabling voice communication to take place with a packet switched network. The new architecture should be able to balance the resources between real-time communication and other services without switching over to circuit switched architecture.
- Support for high data capacity for users demanding complete access to internet using their mobile devices including streaming media, video conference and video downloads.
- Reduction in time required for bearer setup.
- Reduction in delay of packet delivery in a packet switched network.
- Interconnectivity of the current and past wireless technologies.

The evolution of the radio access for LTE came through with Evolved-UMTS Terrestrial Radio Access Network or E-UTRAN. The non radio part of this technology evolved with the System Architecture Evolution (SAE) and this also includes the Evolved Packet Core (EPC). Evolved Packet System (EPS) is the combination of both, LTE and SAE, technologies [1].

EPS is the backbone of the LTE network. It provides the IP connection between the user and PDN for data connection. It is also responsible for voice calls made over IP in LTE. Since multiple subscribers connect for various services to the server at the same time, a QoS has to be maintained which will decide the priority of the calls. For example, voice calls will get higher priority as compared to a file download.

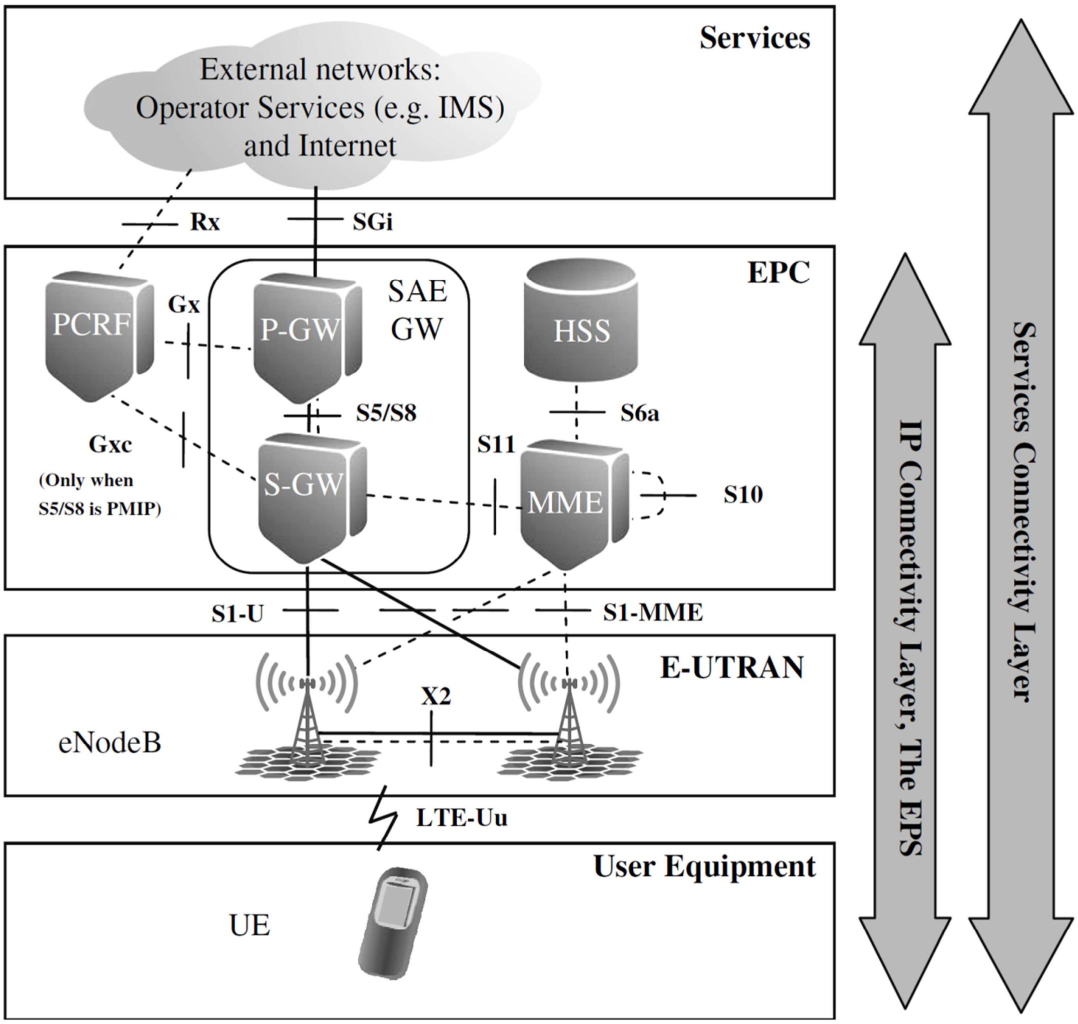


Figure 1.1 Block diagram of EPC with its main entities [2].

1.2 Core Network

The core network comprises of 5 main entities and is responsible for controlling the User Interface (UE) and the establishment of the bearer. This core network is called the EPC and its 5 main entities are:

- Mobility Management Entity (MME)
- PDN Gateway (P-GW)
- Serving Gateway (S-GW)
- Home Subscriber Server (HSS)
- Policy and Charging Resource Function (PCRF)

1.2.1 Mobility Management Entity (MME)

EPC's main control element is MME. It acts as a Control Plane (CP) connection with the UE. This determines the connection of the UE with the rest of the network. The MME is so important that its server would be located in a very secure location at operators premise. MME's functions are:

1.2.1.1 Security and Authentication: The UE's first contact with the server is through the MME. On first contact, the MME starts the authentication process by obtaining the permanent identification from the UE or the previous visited network. The received permanent identification is checked with the Home Subscriber Server (HSS). HSS replies to the MME with the authentication challenge. The MME sends this authentication challenge to the UE which responds to the challenge. The response from the UE is compared with the response from the HSS and if they match, the UE is granted access. This ensures that UEs, with forged identity, do not get access to the network and its resources. This authentication process can be repeated multiple times throughout the lifetime of the connection. MME may also assign the UE a Globally Unique Temporary Identity (GUTI). This minimizes the need to send the permanent UE identity, also called International Mobile Subscriber Identity (IMSI), at regular intervals. To prevent eavesdropping of the communication between the server and the UE, MME also calculates the integrity and ciphering protection keys. These keys are calculated from the master key that MME gets from HSS during initial authentication check of the UE.

1.2.1.2 Mobility Management: When the first contact is made by the UE, MME adds it into a list. MME also signals the HSS in the UE's home network and sets up resources with eNodeB and S-GW. The UE can be either in idle mode or active mode. In the case of the UE being in the active mode, MME can keep track of it by checking the eNodeB at the eNodeB level. If the UE slips into sleep mode then the MME has to keep track of the UE using a group of eNodeBs which is the level of Tracking Area (TA). Once in the sleep mode the UE has to report its location either periodically or when it moves to a new TA. MME controls the handover of the UE when it is in active mode. It also controls the resources for the UE depending on its activity.

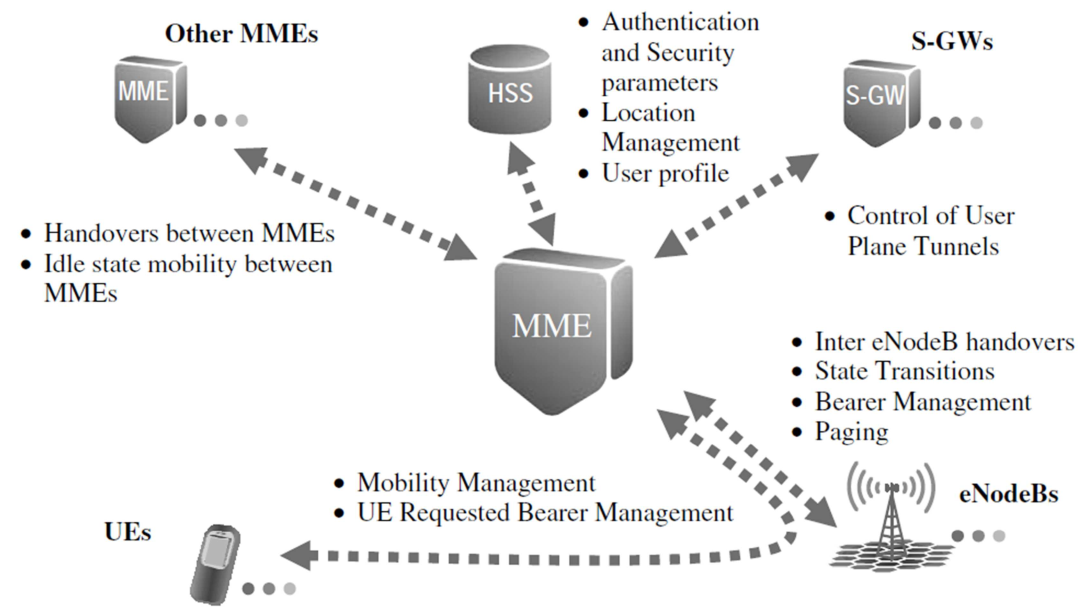


Figure 1.2 MME connections with other logical nodes and functions [2].

1.2.1.3 Connectivity and Subscription: MME is also responsible keeping a record of the UE for the time it is serving that UE. MME may get this information from the HSS when the UE makes the first contact with the MME. The information kept with MME will decide the Packet Data Network (PDN) connection for the UE [2]. After the first contact the MME gives the UE basic connectivity which allows CP signaling with eNodeB and S-GW. MME can change the bearer at a later stage on request from S-GW or from the UE.

1.2.2 Serving Gateway (S-GW)

S-GW is responsible for forwarding all the IP packet traffic generated by the users. This forms the User Plane (UP). When the user hops though various eNodeBs, an IP tunnel is needed to change the route of the IP packets. GPRS Tunneling Protocol (GTP) is used for this purpose. This technology was in place during the time of UMTS as well and is now being reused for LTE. S-GW will handle the GTP for LTE.

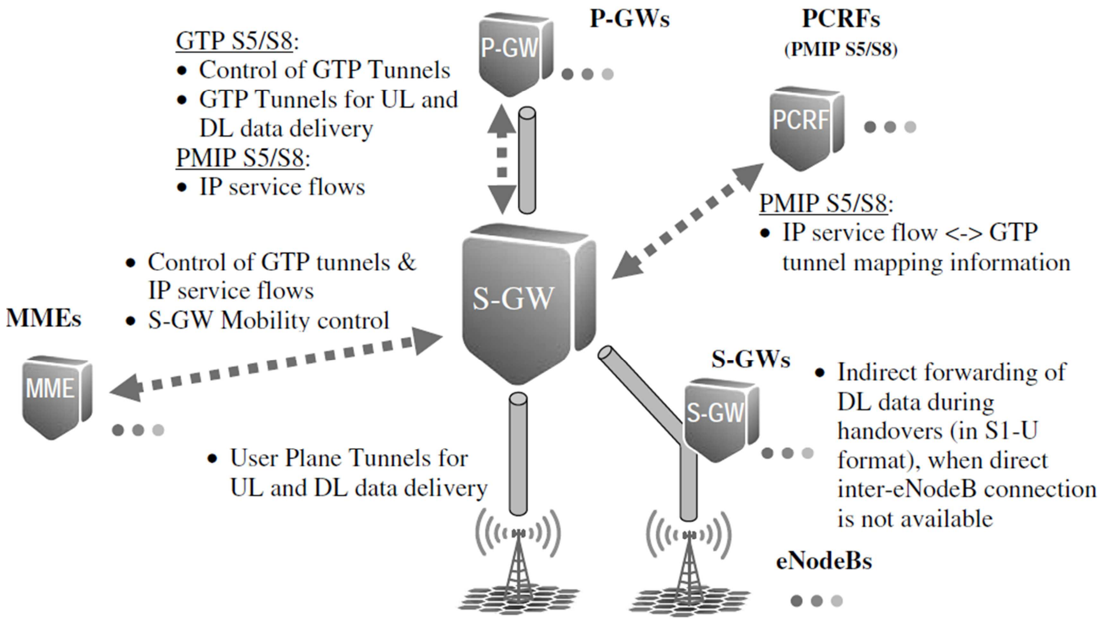


Figure 1.3 Connections of other logical nodes and main functions with S-GW [2].

S-GW does not play a very important role when it comes to controlling functions. It handles most of its own resources. This resource management is also based on requests put in by the MME, PCRF and P-GW. The main controlling agent is MME and as a result when a command for resource modification is received either from P-GW or PCRF, S-GW forwards a copy of this request to the MME.

In the connected mode, the data flow associated with UE is relayed by S-GW through eNodeB and P-GW. This changes dramatically when the UE is in the idle mode. During the idle mode, data path with S-GW is terminated and eNodeB releases all the resources. Hence, when the UE is in idle mode S-GW starts buffering the data. It also notifies MME, which begins the paging process. UE connects back to the network once the paging messages are received and after the connection has been established S-GW sends the buffered messages.

S-GW should have the capability to talk to any P-GW throughout the network. This is required when UE moves to a different location because during mobility the P-GW does not change. For a particular UE, S-GW will relate it with only one MME and one eNodeB except for when the data transfer is through indirect forwarding method. Figure 1.3 also shows how the data transfer occurs during indirect forwarding method. Here the data is forwarded through another S-GW.

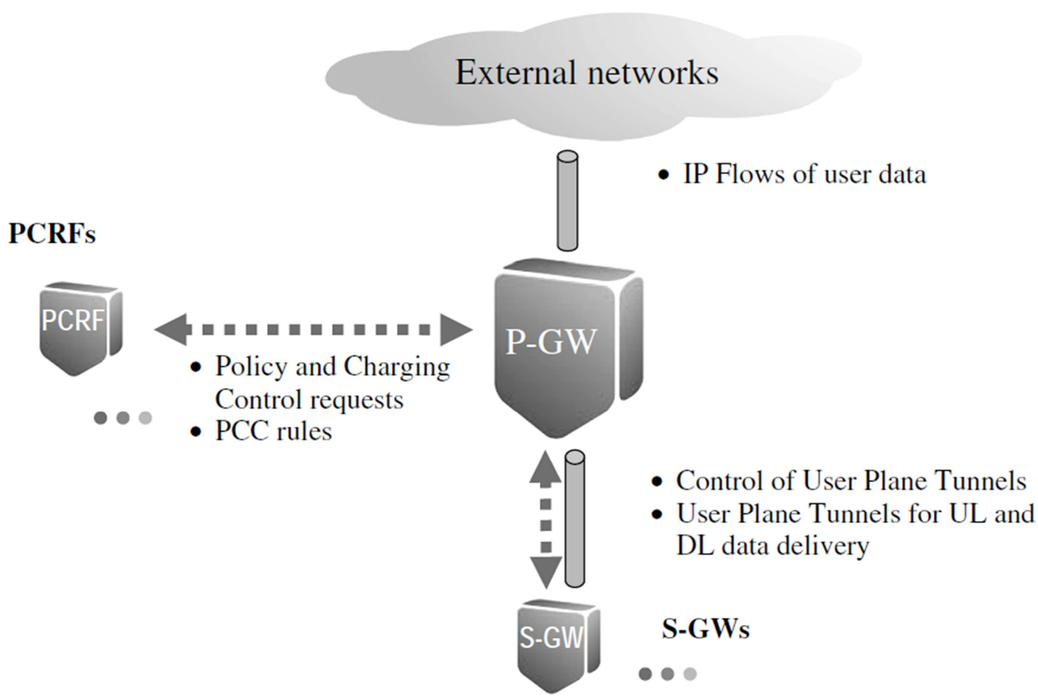


Figure 1.4 Connections to logical nodes and other main functions for a P-GW [2].

1.2.3 Packet Data Network Gateway (P-GW)

P-GW or PDN-GW is the edge router between the EPC and the rest of the world. It is also the last and the highest level of anchor for any UE. For a UE, a P-GW is the only link to the outside world when the UE is on the move. When the UE changes an S-GW, P-GW receives a request from S-GW for a change in bearer. P-GW is also responsible for collection of all the reports and data that will be used for charging purposes.

A P-GW hides the IP address of the UE from the rest of the world. It is also responsible for providing an IP address to the UE from the IP address pool that it administers. If the UE is already connected to an external PDN and is carrying an IP address provided by that PDN then the P-GW has to tunnel all the messages to that network. A P-GW may or may not have Dynamic Host Configuration Protocol (DHCP) functionality. In any case it has to either perform DHCP functionality or use an external source for DHCP before assigning an IP address to the UE.

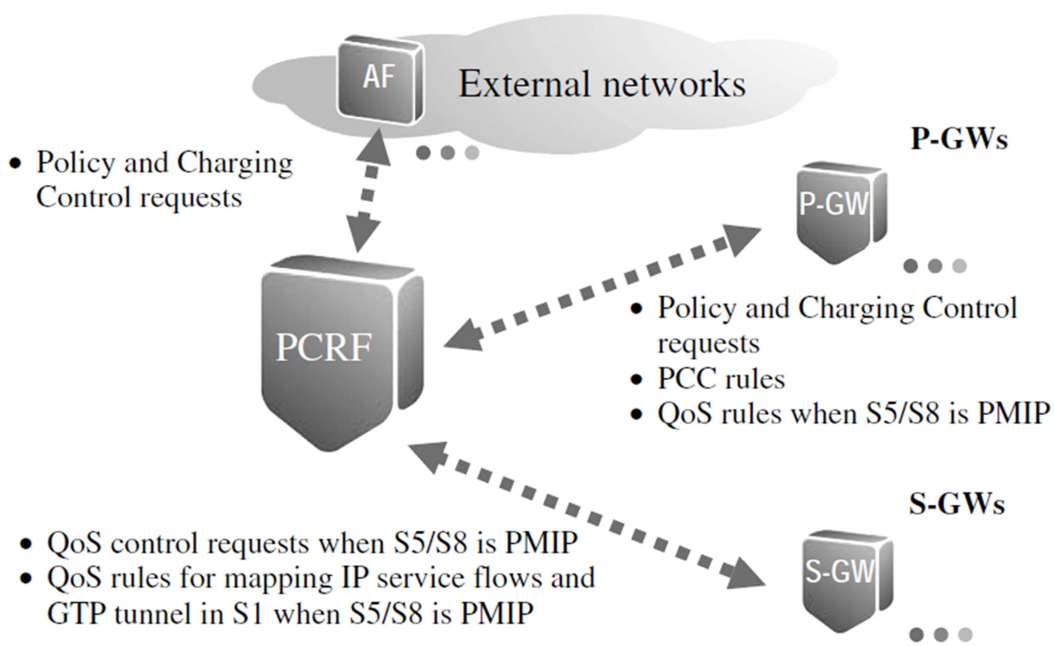


Figure 1.5 PCRF connections with other logical nodes and main functions [2].

1.2.4 Policy and Charging Resource Function (PCRF)

PCRF takes decisions on handling services like QoS. It sends the information collected by the UE to Policy and Charging Enforcement Function which is located in the P-GW. PCRF is responsible for Policy and Charging Control (PCC). PCRF forms a part of the framework defined for PCC [2]. The information that PCRF collects to send to PCC is called as PCC rules. These rules are sent to PCC when the UE requires a new bearer to be setup. A default bearer is setup when UE first comes in contact with the network. More bearer or advanced bearers may be setup later on depending on the needs of the UE. Based on the request received either from S-GW or the P-GW, PCRF will send the PCC rules. For example when the UE sends a request for QoS, the Application Function (AF) pushes the QoS information to PCRF. It then pushed the PCC rules to P-GW. Figure 1.5 shows PCRF connection to other logical nodes. It can be seen that each PCRF is connected to more than one P-GW, S-GW and AF.

1.2.5 Home Subscriber Services (HSS)

Home Subscriber Services is the location where the subscription data of all the subscribers are stored. The subscriber data contains the complete profile of the user and includes information like type of services that the user is subscribing to, type of connections it is allowed to have with the P-GW, if roaming is allowed. It also keeps track of the user's location at MME level. That is, it looks at all the MMEs visited by the UE. The HSS may also keep a record of the current P-GW that is connected to the users. Authentication Center (AuC) is also a part of HSS. AuC stores the keys that are used to encrypt and decrypt the information transfer between the user and the network. It also holds the key which is used for user authentication.

HSS is connected to all the MMEs that the UE can connect to. HSS maintains a record of the MME a UE is connected to at a particular time. Once the UE moves from that MME and a new MME accepts the UE, HSS cancels its record from the previous MME and changes it to the new MME.

1.3 E-UTRAN Node B (eNodeB)

eNodeB is the radio base station in a LTE network. It is responsible for all the functions that are related to the radio. It is also the only node in E-UTRAN. Depending on the coverage sought by the operator, eNodeBs will be distributed at a various points within the network.

eNodeB performs the basic encryption and decryption of the user plane data and also performs IP header compression and decompression. It is also connects the UE with EPC for all data transmission to take place and acts as the termination point of all radio protocols for UE. Radio resource management also falls on the shoulders of eNodeB. It is responsible for controlling the resources required by UE when the UE requests it under special conditions. These conditions may include, but are not limited to, VoIP services, high definition video streaming and other similar high data demanding services.

Another important role performed by eNodeB is the mobility management. It records the measurements made and reported by the UE along with its own measurements of the radio signal between neighboring eNBs. On the basis of these records it can decide about the handover of the UE. After the handover, the eNodeB also exchanges handover information with the new eNodeB and the MME. The new eNodeB is also responsible for routing any and all data from the previous MME to the UE. If the link to the previous MME cannot be found, then it will have to set up a new MME for the UE.

eNodeB has one-to-many relationship with the UEs. That is, one eNodeB can be connected to multiple UE that are within its coverage area but a UE can only be connected to one eNodeB at a time. An eNodeB has a many-to-many relationship when it comes to connection with other eNodeBs. An eNodeB has to be connected to all the eNodeB, via X2 interface, to which a handover can be made by the UE. Although each UE can connect to a single MME and S-GW, each eNodeB has access to a pool of MMEs and S-GWs. During the time a UE is connected to an eNodeB, its connection with the MME and S-GW will not change. So the only time an S-GW or a MME will change is when eNodeB changes during a handover.

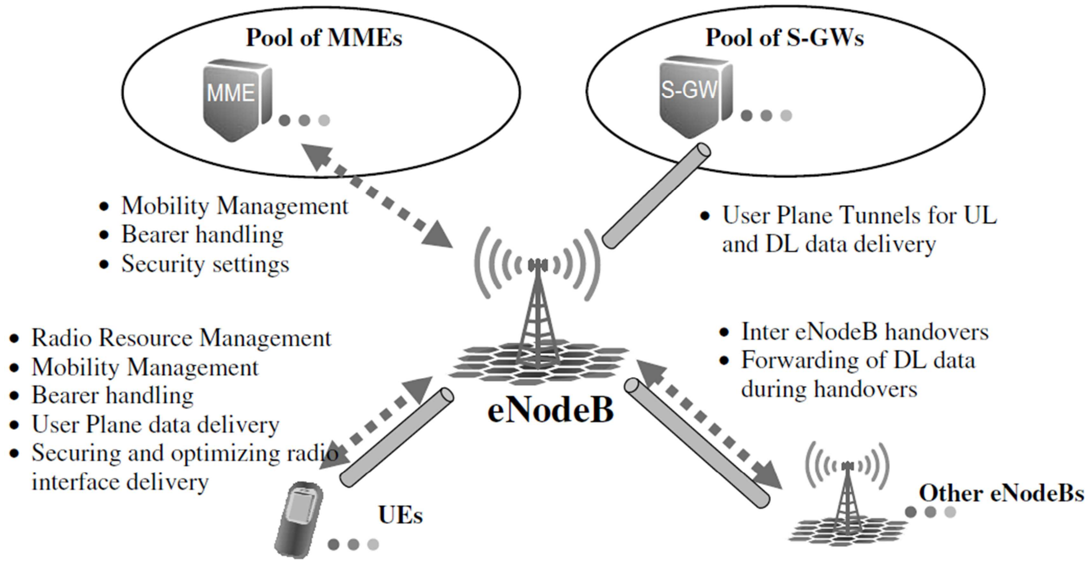


Figure 1.6 Connections of eNodeB with other surrounding entities and logical nodes [2].

1.4 User Equipment (UE)

The mobile devices that are available at the user's end are called UE in LTE. These are the end devices which are used to signal the network that a user wants to connect to the network. After the connection is established it is responsible for maintaining the connection till the user decides to switch off the device or move to a new location. In the case of moving to a new location, the UE will measure the signals from nearby eNodeB and report those back. The report will be used by eNodeB to choose the best eNodeB for the user at its current location. When the user switches the device off, UE will signal the eNodeB for termination of service before switching off completely. eNodeB will act as the BS of current 3G and 2G systems. Instead of Subscriber Identity Module (SIM) cards, used in the current mobile handsets to identify the UE with the user, Universal SIM (USIM) will be used in LTE. Universal Integrated Circuit Card (UICC) will hold the USIM application.

1.5 IP Connectivity Layer and Service Connectivity Layer

IP connectivity layer is the combination of user equipment, E-UTRAN and EPC. This layer is also called the Evolved Packet System (EPS). The role of this layer is to support the IP connectivity and is designed specifically to meet this need. The LTE system is wholly based on packet switched networking while circuit switched networking has been entirely removed. As a result this system plays a vital role in making LTE work the way it was designed.

Since LTE system is based entirely on IP, the service connectivity layer has to go hand in hand with it. It provides support for running services from the IP-Multimedia System (IMS) over IP. An example of this can be providing VoIP services over the IP connectivity layer and interconnecting this service with the traditional PSTN networks, which are circuit switched.

CHAPTER 2

VOICE SERVICES IN LTE NETWORK

2.1 Introduction

Over the years the demand for high speed internet in a mobile network has increased for a number of reasons, even then voice remains the number one priority of all these networks. Until recently, voice was being transmitted in a mobile network on a circuit switched network. With the introduction of LTE, which is an entirely IP based system; the voice traffic will have to be packet switched and over IP. The advantage of using a packet switched network is that it is not application dependant. The mode of transmission is pre-defined and applications have to be designed to run on this method of transmission. This becomes a challenge as it is known that packet switched network is not the best for handling real time data. Special rules have to be applied to packets carrying voice and other real time data for getting the desired results. The rules which allow an IP enabled network to provide such services is called Quality of Service (QoS).

2.2 Voice Requirements For LTE

One of the main problems of real time services is the delay induced in a packet switched network. The delay preferred in a voice conversation is less than 200ms [2]. On the higher side the delay should not exceed 280ms for a satisfactory conversation to take place.

Figure 2.1 shows the relationship between mouth to ear delay and the R-Factor. R factor uses a formula to best judge the call quality from a user’s perspective. It varies from 50 to 100, with 100 allowing the best VoIP conversation and below 50 leaving all the users dissatisfied.

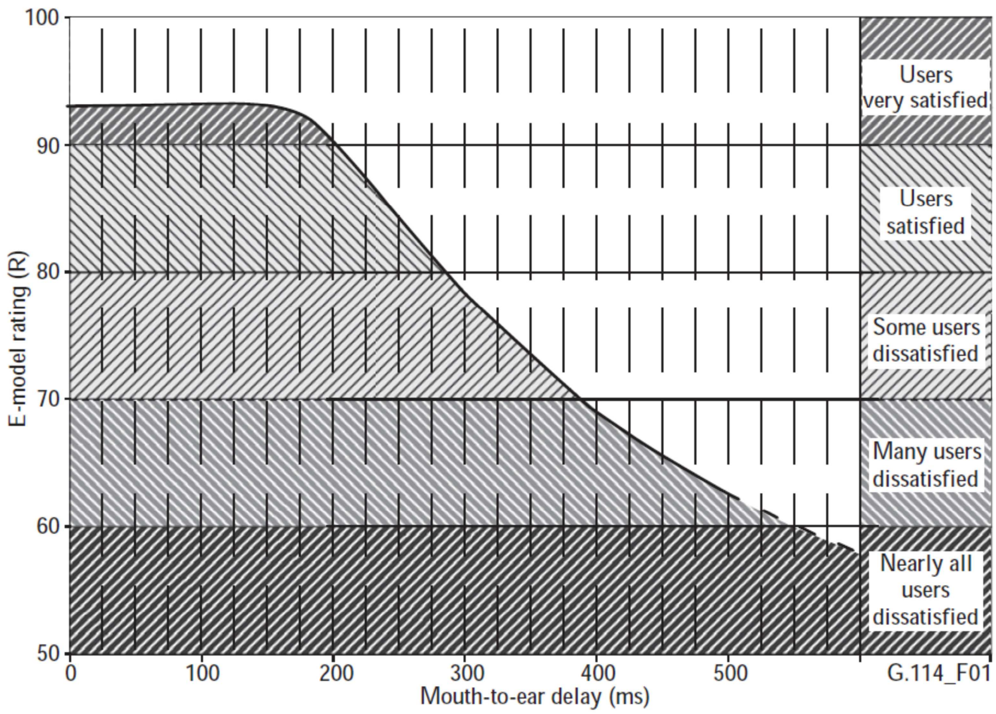


Figure 2.1 Delay and R-Factor ratings [2].

Packet switched networks operate by breaking an information block into smaller sizes called packets. These packets are then numbered and sent over IP. At the destination, a service takes care of out of order packets and arranges them in order once all the packets arrive. However, rearranging the packets at the destination is not the best option for real time services. The IMS will provide information about QoS for the requested service. The radio network must have the resources and the algorithm in place to deliver the required QoS.

The current GSM and WCDMA technologies have evolved to a level where voice call drop rate has been reduced to a mere 0.3% [2]. For voice calls to be successful in LTE, it should provide a comparable voice call drop rate if not less. LTE should also be able to interconnect between earlier networks which are circuit switched networks.

Keeping power consumption in mind, the UE will have to be pinged with ‘keep alive’ messages at regular interval when an active call is not taking place. The frequency of these ‘keep alive’ messages is decided by IMS. If the IMS is within the operator’s networks without firewalls and NAT, then this frequency can be low. On the other hand, if the IMS is behind some firewall or NAT then the ‘keep alive’ message frequency has to be increased. Increasing this will have a negative effect on the battery life of the UE.

2.3 Session Initiation Protocol (SIP) Telephony

SIP is a RFC 3261 defined standard protocol for establishing voice calls over IP. This protocol is used to initiate and terminate a call and also to maintain the session during a voice call between two SIP enabled devices.

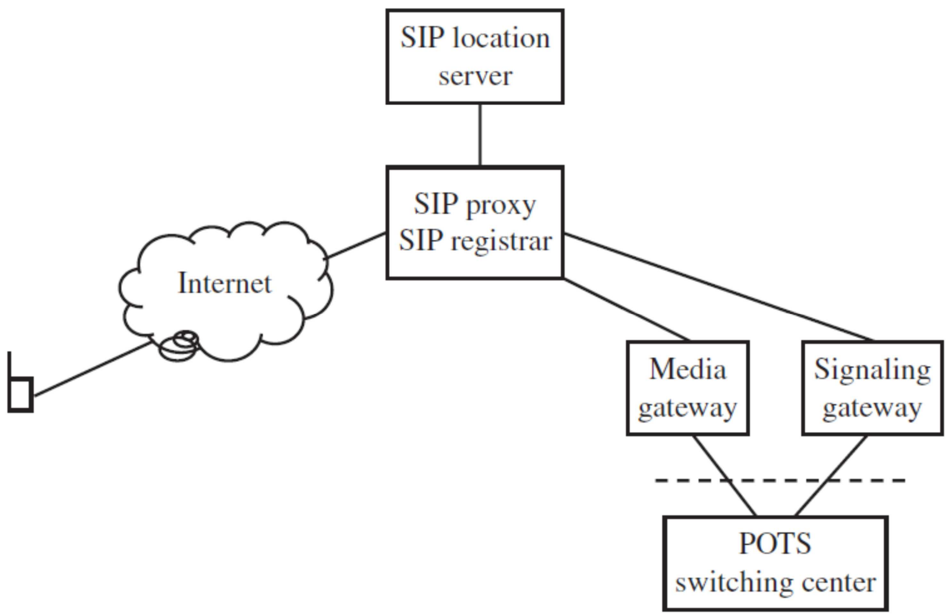


Figure 2.2 SIP network Architecture [3].

With VoIP gaining popularity SIP has gained popularity too. It is evident from the fact that a number of operators are providing SIP enabled modems and phones which can be connected to the internet at one end and phone at the other. Using these calls can be made to almost anywhere. SIP calls can be placed over any network that has support for packet data transmission and since LTE’s architecture is designed exclusively for packet data, SIP can play a major role in realizing LTE’s voice communication goals.

2.3.1 SIP registration

The device which runs the SIP application is called the User Agent (UA). This application can be programmed inside a computer, phone or even a modem. After startup the application will register its user information with the SIP registrar on the network. Registration allows other devices to search the user, make calls and activate a user on the SIP network. To identify a user, a User Resource Identifier (URI) is used. To send a message, the source must know the URI, destination UDP port number and the IP address. When the SIP application first sends its information over to the registrar, it searches the database for the information and subscriptions of the user. The authentication of the user is done via a password provided to the user at the time of registering for the SIP service.

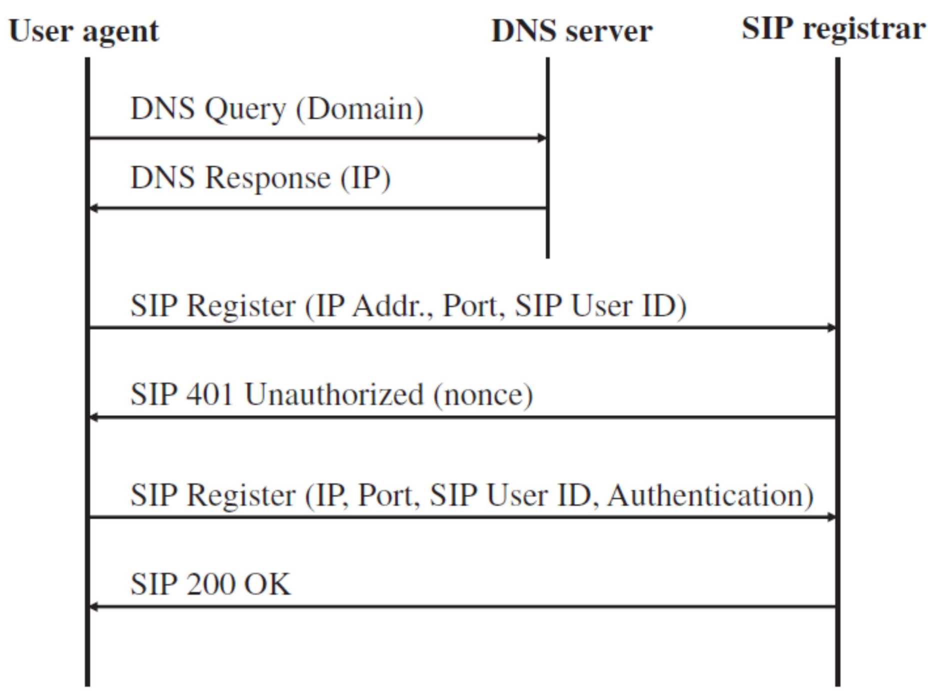


Figure 2.3 SIP registration messages [3].

The first attempt to register is rejected by the registrar and sends a SIP '401 Unauthorized' message back to the UA along with a random value known as 'nonce'. The UA then creates an authentication response based on the nonce value and the initial password. If this matches the calculations of the SIP registrar, access is granted to the UA. The SIP registrar denotes a successful registration by sending a SIP '200 OK' message. It also stores the IP address, user ID, UDP port etc for call establishment at a later stage.

2.3.2 SIP based call establishment

To establish a call, a UA sends SIP ID or a number associated with the user it wants to call, to the SIP proxy. SIP proxy is implemented in SIP registrar but is a separate entity altogether. Its function is to look up all the required information for establishing a call based on the SIP ID or the number sent by the caller.

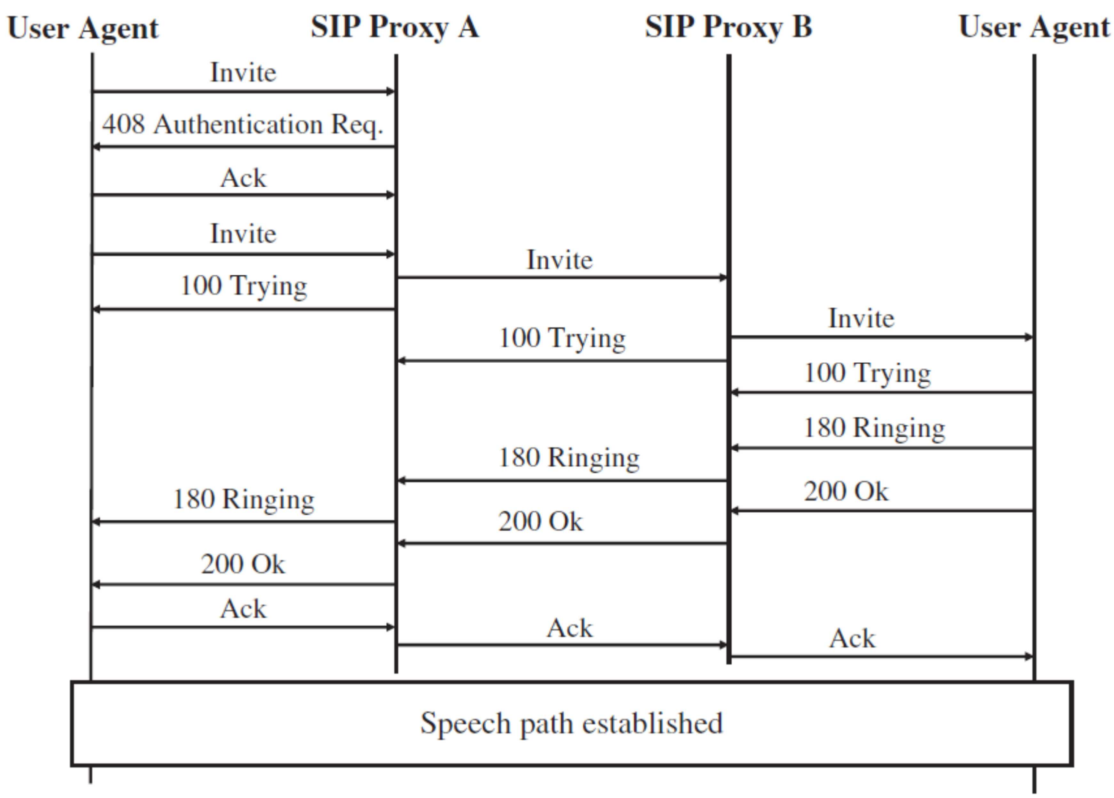


Figure 2.4 Message exchange during SIP call establishment [3].

The first message to the SIP proxy is the invite message which contains the identity of the user the originator wants to call. On receiving the invite message, SIP proxy asks for authentication by the originator UA. The originator drops the invite message by sending an acknowledgement and then resends the invite after calculating the required value using nonce.

After authentication has been verified, SIP proxy starts searching for the destination user. If the destination user is not in the database of a SIP proxy, it will forward the invite to other SIP proxies. Once the user is found in a SIP proxy database, the invite message will be forwarded to it. At the same, time SIP proxy sends a SIP ‘100 Trying’ messages back to the originator. At this point the UA on the call receiving side alerts the user of an incoming call. It also sends a SIP ‘180 Ringing’ message to the SIP proxy which forwards it to the originator. If the user accepts the call, a SIP ‘200 OK’ message is generated and sent to the originator. At this point, the audio data will start flowing between the two users. The audio data will not go through the SIP proxies, which are there only for establishing the call.

After the call establishment Session Description Protocol (SDP) is used for describing the other UA about the type of media that is going to be used in the call. This is a necessary step to prepare the hardware for proper streaming of the audio/video data being used in the call.

2.4 Internet Multimedia Subsystem (IMS) Support For Voice Services

IMS is a subsystem within the network designed for supporting voice communication in GSM and WCDMA networks. IMS is based on concepts around Session Initiation Protocol (SIP). It is a signaling protocol designed to establish and maintain connections for multimedia services like voice calls over IP networks.

IMS has multiple entities within itself which are connected to each other via various interfaces. The core of IMS is called Call Session Control Function (CSCF). All the SIP signals are handled by CSCF which involve controlling the path of the media and starting applications required for a media call. It is split into three entities:

- The Proxy-CSCF (P-CSCF)
- The Serving-CSCF (S-CSCF)
- The Interrogating-SCSF (I-CSCF). [4]

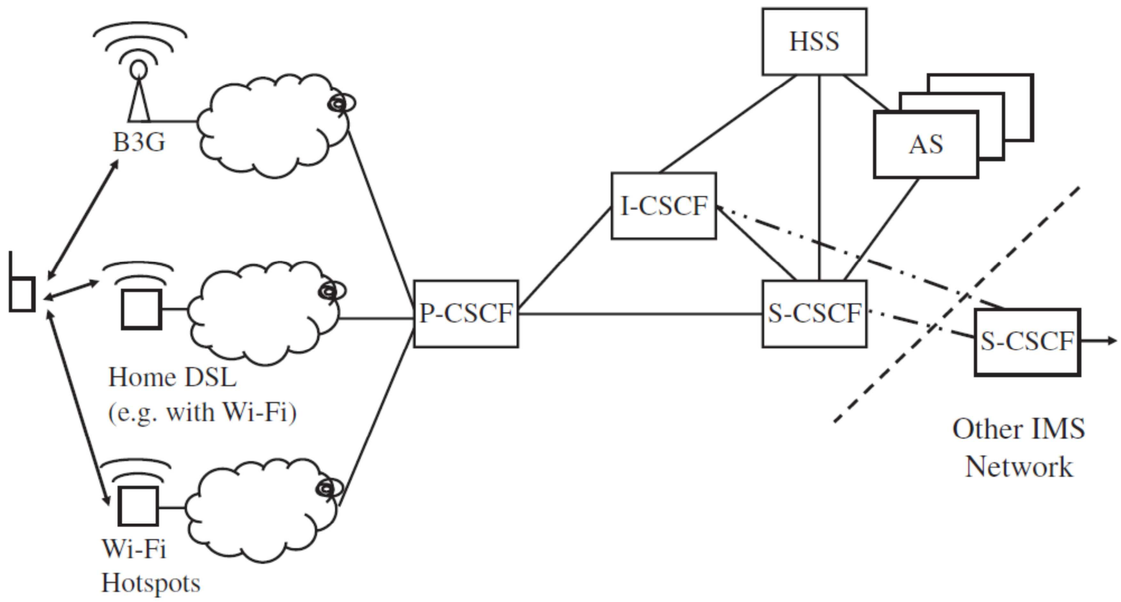


Figure 2.5 Architecture for IMS [3].

2.4.1 The Proxy-CSCF

P-CSCF functions as a SIP proxy and the first contact with the IMS for an IMS terminal. During the registration process with IMS, a P-CSCF is assigned to a user and the assigned P-CSCF remains the same for the user until the next registration. All the signaling messages, to and from the user, pass through P-CSCF. This is necessary as a mobile user can get disconnected while on the move, in such a case P-CSCF ends a call gracefully. In case the user loses connection while on the move, P-CSCF informs the SGSN about the connection loss with the user. SGSN lowers the uplink and downlink bit rate to zero and then stops forwarding packets to the user if the media is still being streamed. This procedure is employed when the user is connected to a GERAN. If the user is connected to E-UTRAN then the eNodeB signals the MME about connection disruption and the MME will forward this message to P-CSCF. P-CSRF then talks to the IMS layer and disconnect gracefully by sending a SIP ‘Bye’ message to the other user.

P-CSCF also has to handle encryption between an IMS terminal and itself. SIP signaling messages are in plain text and can give a lot of information to someone who hacks into the system. IMS requires P-CSCF to establish an IPSec connection between the IMS terminal and itself. Any information between the IMS terminal and the P-CSCF will be encrypted and the operators do not have to worry about encrypting SIP messages.

Since SIP signaling messages are in plain text and they contain a lot of information about functionality, their size is huge. A bigger sized file will take longer to be transmitted over the system and as a result it would induce delay in establishing a call. To minimize this delay, P-CSCF uses Sigcomp to compress the size of SIP signaling messages.

P-CSCF also has an interface with the wireless network it is serving like GGSN for GSM, ASN-GW for WiMAX and MME for LTE. It uses this interface to control and maintain the QoS that its system is providing. It also ensures that high quality link provided between two users is used for the type of media or service agreed upon before setting up the resources. P-CSCF may also generate bills for the subscribers and send it to a billing service for processing.

2.4.2 The Serving-CSCF

S-CSCF is the central node in IMS and essentially a mixture of SIP registrar and SIP proxy. During registration the IMS terminal will use one of the S-CSCF to register to the network and will be connected to the S-CSCF until it disconnects. It will be through S-CSCF that all the SIP requests will be sent.

When the IMS terminal first comes in contact with an IMS network, S-CSCF will obtain the authentication, service and subscription information from a centralized server, known as the HSS. A S-CSCF needs to get this information from the HSS every time an IMS connects because there are many S-CSCFs available in an IMS network. The user is not guaranteed to be connected to the same S-CSCF every time. After the registration, P-CSCF forwards all the messages sent by the IMS terminal to the S-CSCF. It is S-CSCF which decides how the message needs to be dealt with and where it needs to be forwarded.

If the message is a simple voice call establishment request, the S-CSCF will check the SDP messages to confirm that it is a voice call. After checking the database and confirming that the user has voice call privileges, it will analyze the destination address. The destination address can be in a phone number format or it can be a URI. If the destination address is in the form of a tradition phone number, the S-CSCF first checks to see if the destination is over an IMS network or a circuit switched network. If it is a CS network, then the call is forwarded to that network and interworking between IMS voice telephony and CS network needs to be performed to connect the call. On the other hand, destination address is a URI or the phone number was identified as a URI, the S-CSCF will check if the destination user is within the same IMS network or not. If the user is not within the same network, SIP proxy address will be located and the request for call establishment will be sent to that address. If the user is within the same IMS network, S-CSCF will start charging records and the call will be established by the usual SIP procedures.

Even after being the central part of the IMS network, its functions are limited to the following:

- SIP message analysis
- SIP message modification (Used for special services like hidden numbers)
- SIP message routing

All these functions on SIP messages allow S-CSCF to perform the basic functions of establishing a call and providing some special functions like hiding the number from the destination user. If the subscriber wants to use some advanced functions, S-CSCF will have to implement it using an external server called the Application Server (AS).

2.4.3 The Interrogating-CSCF

I-CSCF is an edge SIP function with an IP address advertised on the Domain Name Server (DNS). After the initial contact with the IMS network, the IMS terminal will contact P-CSCF for registering to the IMS network. P-CSCF will need to forward the received registration message to a S-CSCF. Since IMS is a distributed system, the right S-CSCF for the IMS terminal is unknown to the P-CSCF. So P-CSCF forwards this message to the I-CSCF. An I-CSCF for a particular IMS terminal is found using DNS lookup with the information provided by the IMS terminal in the registration request to the IMS network. On receiving the registration request, forwarded by the P-CSCF, I-CSCF will contact HSS and obtain the configuration of the S-CSCF, which will be used for the IMS terminal. The HSS, depending on the user's subscription profile, will send out a response to the I-CSCF with a list of user subscription capable S-CSCF. I-CSCF then selects the best S-CSCF for the user keeping in view the capabilities of the S-CSCF and the subscription and service requested by the IMS terminal.

Since IMS network is a big and distributed system, the S-CSCF of an originator of a call may not be serving the terminating IMS terminal. In this scenario, the SIP 'invite' message is sent to an I-CSCF. The I-CSCF of the call originator looks up the I-CSCF serving the terminating IMS terminal, using the contact information provided by the originator. SIP 'invite' message sent by the originator is forwarded to the terminating I-CSCF, once it is found. The I-CSCF contacts the HSS for subscription information which it obtains from the 'invite' message. HSS responds to the request with the contact details of the S-CSCF assigned to the terminating IMS terminal. If an S-CSCF is not defined for the terminating IMS terminal (this may occur if the device is not registered), a new one will be created. This step is needed to support voicemail function in voice calls on IMS. Once the identity of the S-CSCF is known, I-CSCF moves out of the routing path and does not include its address in the routing path of the SIP messages to the terminating IMS terminal.

2.4.4 Media Servers

Media servers are used to provide functions related to media like media manipulation, audio mixing etc. [5]. These functionality maybe used when features like call conferencing needs to be implemented. Media servers are split into two parts:

- Media Resource Function Processor (MRFP): When the media being sent needs modification like mixing or joining, example during a voice or video conference, then MRFP is called to perform these functions.

- Media Resource Function Controller (MRFC): As the name suggests the MRFC is just a controller for MRFP whenever the services of MRFP are required. It connects with CSCF for controlling MRFP.

2.4.5 Breakout Gateway Control Function

When a S-CSCF decides that DNS that a session cannot be routed using DNS it is routed to a Media Gateway Control Function (MGCF) which takes care of all the circuit switched networks.

2.4.6 Application Server (AS)

Application server communicates with CSCF to deliver services which it hosts and executes. Although most of the services need to be standardized via 3GPP but it can handle other non standard services as well.

For example, if the IMS terminal wants to use the voicemail service while it is unavailable or out of coverage area, the AS will act as a UA for the originator caller. The UA will establish a call as if it was the UA for the destination and take the message. It will then, via S-CSCF, send a message to the destination user informing it of the message stored on the AS server. So application server can act as an active and a passive system. In this case, taking a voice call becomes a passive application while generating a message for the destination user becomes an active application. Another example of the AS as an active application is providing wake up alarm functionality. AS will also be used for instant messaging, push-to-talk and other similar services.

2.5 IMS Registration

IMS registration process is very similar to the registration of a user in to a SIP registrar. It includes some additional protection and authentication steps to prevent any unauthorized access to the network. Before starting a voice call on an IMS network the user agent (UA) needs to register itself and in turn the device with the IMS network.

Once the connection to the P-CSCF is made, which is the first point of contact for a device in an IMS network, the UA sends a SIP registration message. A subscriber needs to have a private user identity, which is often stored on the SIM card on an IP Multimedia Service Identity Module (ISIM). If the user does not have that information on its SIM card, a temporary private user identity will be created.

At the beginning of the process, no S-CSCF is assigned to the user, so the P-CSCF begins by forwarding the registration request to the I-CSCF. I-CSCF contacts the HSS and obtains the subscription and service information and selects a suitable S-CSCF for the UA. S-CSCF analyzes the user's registration message and on finding that the user is unknown, contacts the HSS and obtains the authentication information from it. HSS responds with the following information:

- RAND: a random number
- XRES: this is the value that is generated by the SIM card using RAND, a secret key (pre-stored on the SIM card) and an authentication algorithm. This value is sent by the UA back to the S-CSCF and is compared to the value generated by HSS. The value generated can only be matched if they are created using the same algorithm and the same secret key.
- AUTN: it is possible for attackers to act as an IMS network. To prevent that the HSS generates an authentication token (AUTN) using the RAND and a different algorithm. This token is used by the UA to confirm if the network it is connecting to legitimate.
- IK: the validity of the messages exchanged between the UA and the P-CSCF are checked by a checksum calculated using the Integrity Key (IK) and RAND.
- CK: UA and P-CSCF use this key to encrypt SIP messages sent through the IP tunnel.

The first registration request is cancelled by the S-CSCF and a SIP '401 Unauthorized' message is sent back to the user. This response also contains the IK and CK keys but these keys are removed by the P-CSCF before forwarding the message to the UA. The UA does not need these keys as it can create the keys using the same process the HSS does, security algorithms and RAND.

The response sent by P-CSCF will still contain the AUTN and RAND values. The AUTN will be used by the UA or the SIM card to authenticate the network. If the authentication is valid, the UA will generate the XRES, CK and IK. CK and IK, again, will not be sent to the IMS network and will only be used for validity test and encryption of the messages exchanged between the P-CSCF and the UA. However, the XRES is sent to the P-CSCF along with another request to register to the network.

On receiving the second registration request, the S-CSCF compares the XRES value with the one sent by the HSS. If the value matches, the user is considered as registered. The S-CSCF, then requests the user profile from HSS and sends a SIP '200 OK' message to the UA, informing it about the successfully registration.

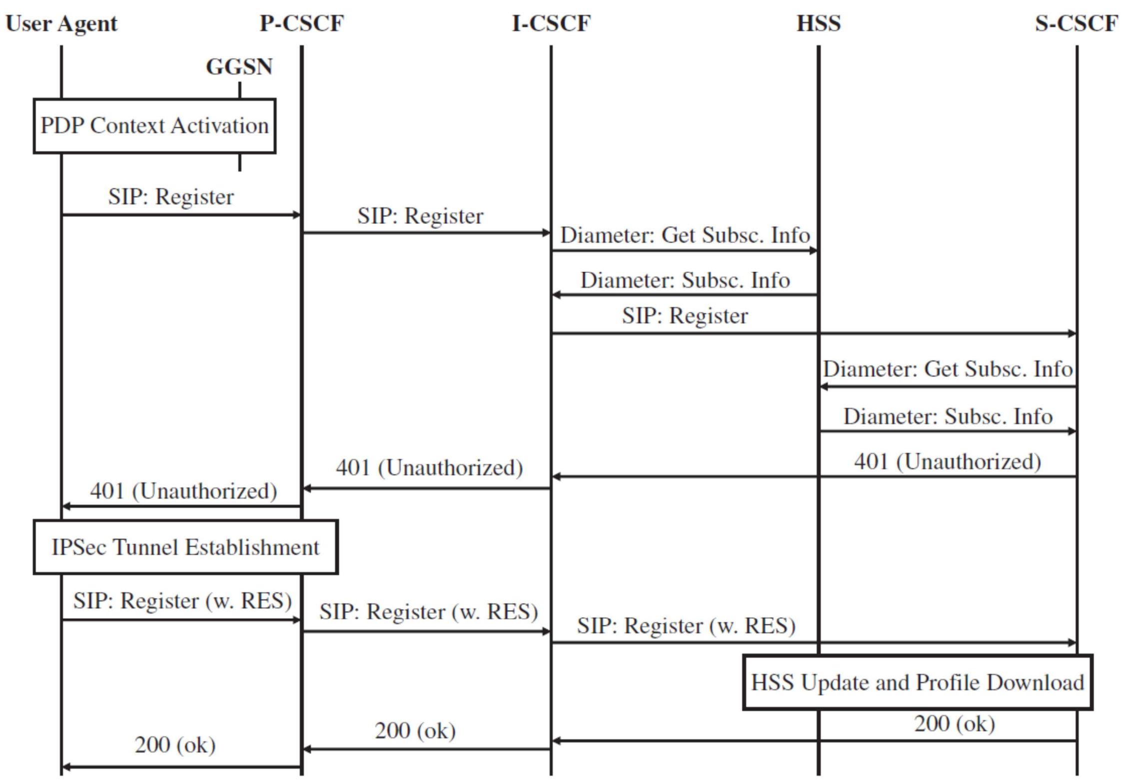


Figure 2.6 Message exchange for IMS registration [3].

A device will stay registered to the IMS network till it deregisters. Deregistration can be done by sending a SIP register message with expiry set to zero [3].

2.6 IMS Session Establishment

An important factor for real time media services to work satisfactorily over wireless networks is that proper resources should be available. The amount of resources that need to be reserved can only be judged by determining the required bandwidth needed for transmission of a particular media. This information can only be obtained after establishment of a session on both the ends. Thus, the process of session and resource establishment is tied together.

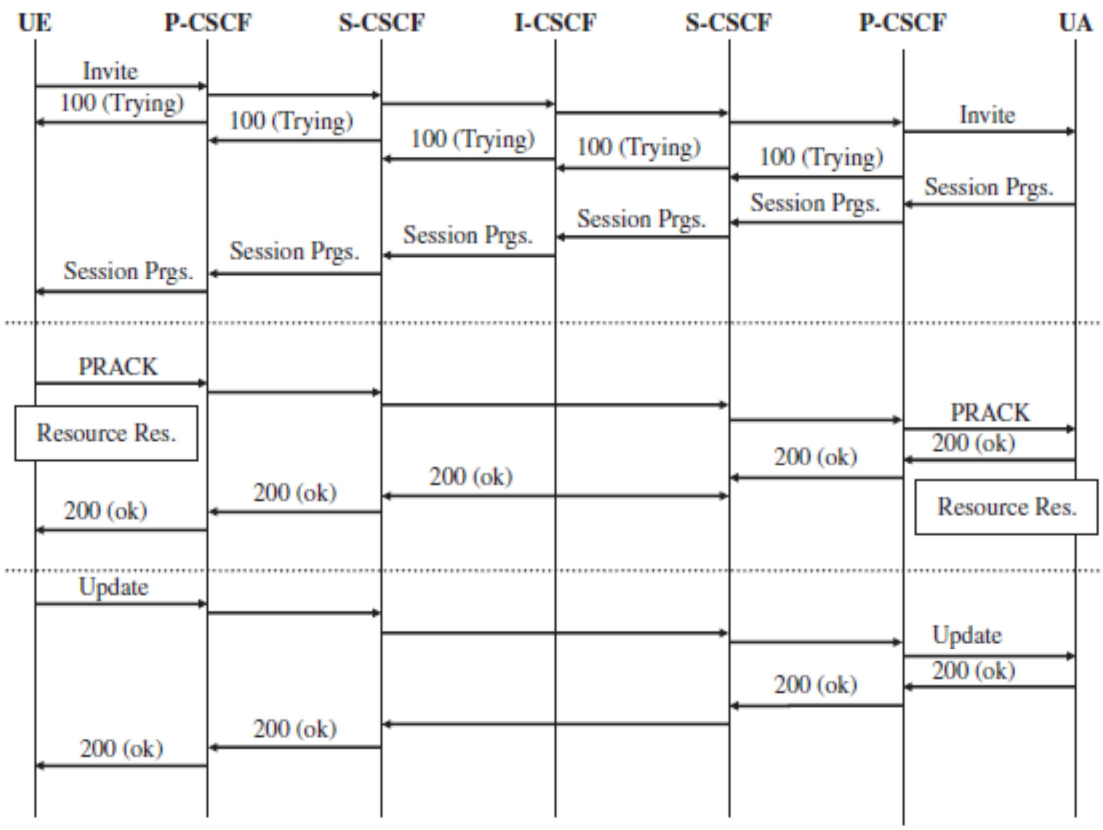


Figure 2.7 Message transactions during an IMS session establishment [3].

Similar to a SIP call establishment, an IMS session establishment begins with the originator sending a SIP 'Invite' to the destination. The invite message also contains the resources that will be required to handle the type of call the caller is planning and also alerts the destination device not to start alerting the user about the call until all the resources have been established. The invitation message also contains the codecs that the caller plans to use during the call. In response the destination sends a SIP ‘183 (Session Progress)’ [3]. The response alerts the call originator that the destination also needs to reserve resources for the call, though it will not be able to reserve the resources until the codecs have been finalized. This response also contains a list of codecs supported by the destination.

The response allows the originator of the call to select a suitable codec and start the resource reservation process at its end. It responds with a provisional acknowledgement (PRACK) message to the destination. This message tells the destination which codec was selected by the originator and it can start reserving resources on its own. In response to PRACK the destination sends a SIP ‘200 OK’ message back to the originator. After receiving the ‘200 OK’ from the destination, the originator will send a SIP ‘Update’ message when its resources have been established. The destination, after the receipt of the SIP ‘Update’ message, will alert the user once its own resource establishment is complete. If for some reason, the resource establishment fails at any end, the session will be terminated and the user will not be notified.

When both ends are ready with the required resources, a SIP ‘180 ringing’ message will be send by the destination to the originator informing the originator that the resources were established properly and the user is being alerted about the call. The originator responds to the ‘180 ringing’ message by making the stream ready to flow once the user accepts it and the destinations sends a ‘200 OK’ in response.

2.7 Understanding Voice Over LTE (VoLTE)

Previous technologies like CDMA, GSM and WCDMA, had support for packet switched services as well as circuit switched services. So delivering voice was to connect to a circuit switched network and use the pre-existing technologies. With LTE, there is no support for circuit switched networks. This makes the task of providing voice calls on LTE a little different from its predecessors. So voice service can be provided in LTE using either the IMS based services or the older architecture of circuit switched networks like in CDMA, GSM or WCDMA.

In LTE, if the IMS based services are used, then a standardized voice call service, VoIP, will be used. It will provide the users with the traditional functionality of a circuit switched voice calling solution along with some additional features. These features may include video calls, chats along with sharing pictures or videos. Before going further, a realization has to be made that since the LTE technology is still new and even though it is growing at a fast pace, it will not be present everywhere where voice calls will be made. The success of VoIP over LTE relies on the fact that initially it will not be replacing the circuit switched network completely. It has to go work with the current technologies till its roots are deep enough to take over the entire voice architecture. New devices need to be designed with LTE architecture in mind. They should also be able to provide support for both voice services till VoIP over LTE becomes the primary choice. This would allow the users to use the desired service depending on the network coverage of LTE. This means that the handover from one technology to another should be extremely seamless.

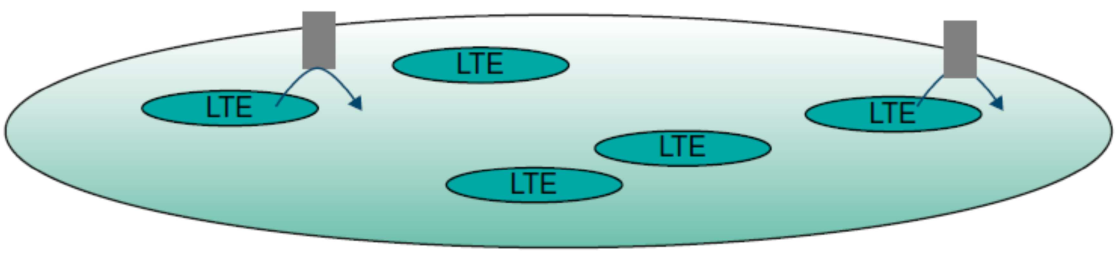


Figure 2.8 Scattered LTE coverage with need of complementing networks [4].

Figure 2.8 shows how new and superior LTE networks will rely on the vast coverage of the existing networks for providing complete voice call coverage. With the above scenario two conditions can arise.

- The call is started in a non LTE network area (light colored area). Since this call was established while in a non LTE network, the call will be on a circuit switched network. Depending on the LTE network and the choice of voice technology present in that network, the call will either be switched over to IMS services or will remain in the traditional circuit switched network.

- The call is established in a LTE covered area and the user moves out to a non LTE covered area. In this case the handover needs to be made. If the non LTE area has the capability to support IMS voice services that there would be a Packet Handover and the voice service will continue to run on IP based network. Otherwise the call will have to be transferred using Single-Radio Voice Call Continuity (SRVCC).

2.8 Single-Radio Voice Call Continuity (SRVCC)

When a call established in a LTE network needs to be handed over to a network which does not support IMS services then the call needs to be handed over using SRVCC. SRVCC allows call handover from a LTE network to a network which handles its voice services using the circuit switched network approach.

The challenge in providing a handover from an IP based voice service to a non IP based voice service is that the handset cannot stay connected to both the types of network while the handover takes place. This means that the handset is connected to only a single radio at a given time and this is the reason it is called as single-radio voice call continuity [4]. Allowing the handset to be connected to both the networks at the same time would require the handset to have a pricey and complex set of filters, antennas and processors. To be able to make such a handover, the handover itself has to be very quick to minimize the interruption in the voice call.

Voice Call Continuity (VCC) was first introduced in Rel-7 for providing handover between WLAN and circuit-switched services. Although in VCC, the handset is connected to both the networks at the same time before handover is carried out. It was possible for the handset to be connected to both the networks at the same time because of the difference in the system properties. WLAN is a low transmitting power system with relatively small coverage while GSM is high power for transmission with a large area of coverage [4].

In 3GPP the following handovers have been specified:

- LTE to GERAN
- LTE to 1xCS

2.8.1 *Voice call from LTE to GSM (E-UTRAN to GERAN)*

A handover made from LTE to GSM impacts the functions in MME, eNodeB and MSC. It also requires support for call continuity in the IMS [4]. A special interface, Sv, has been introduced between MME and MSC for performing handover from LTE to GSM.

For a successful handover, both the networks (GSM and LTE) and UE should support SRVCC. Also, that the session is provided through Service Centralization and Continuity Application Server (SCC AS) [4].

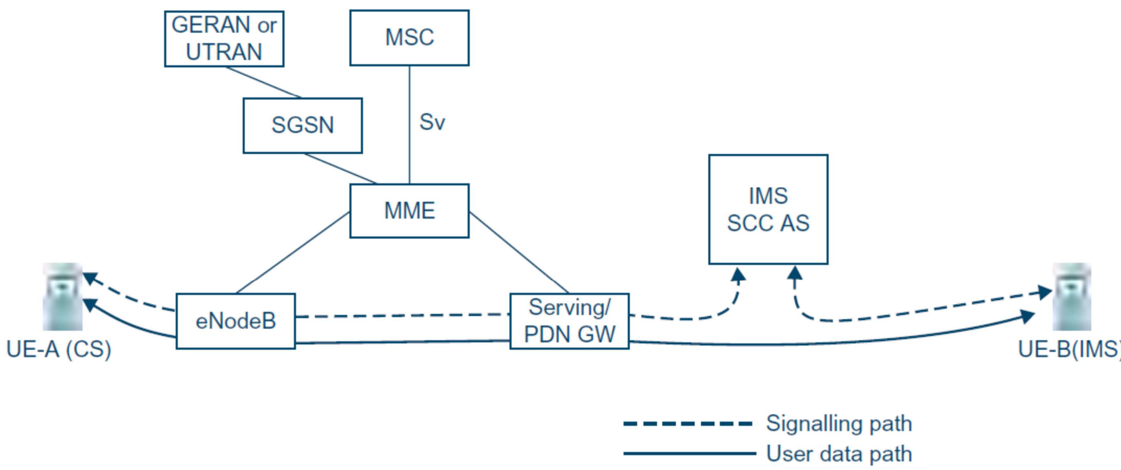


Figure 2.9 UE before SRVCC execution [4].

In figure 2.9, both the UEs are connected to LTE network and are connected through SCC AS. The signaling is provided by IMS and the voice packets are sent from UE-A to UE-B using the IP connectivity through eNodeB and P-GW. When the LTE signal strength starts diminishing in UE-A's area, it signals back to the eNodeB about the change in signal strength. It also sends the signal strength measurements of other potential handover candidates. eNodeB also gathers the following information before initializing the handover procedure:

- The current status of UE with respect to an active voice call.
- SRVCC capability of the UE.
- The support of VoIP on the neighboring non LTE network.

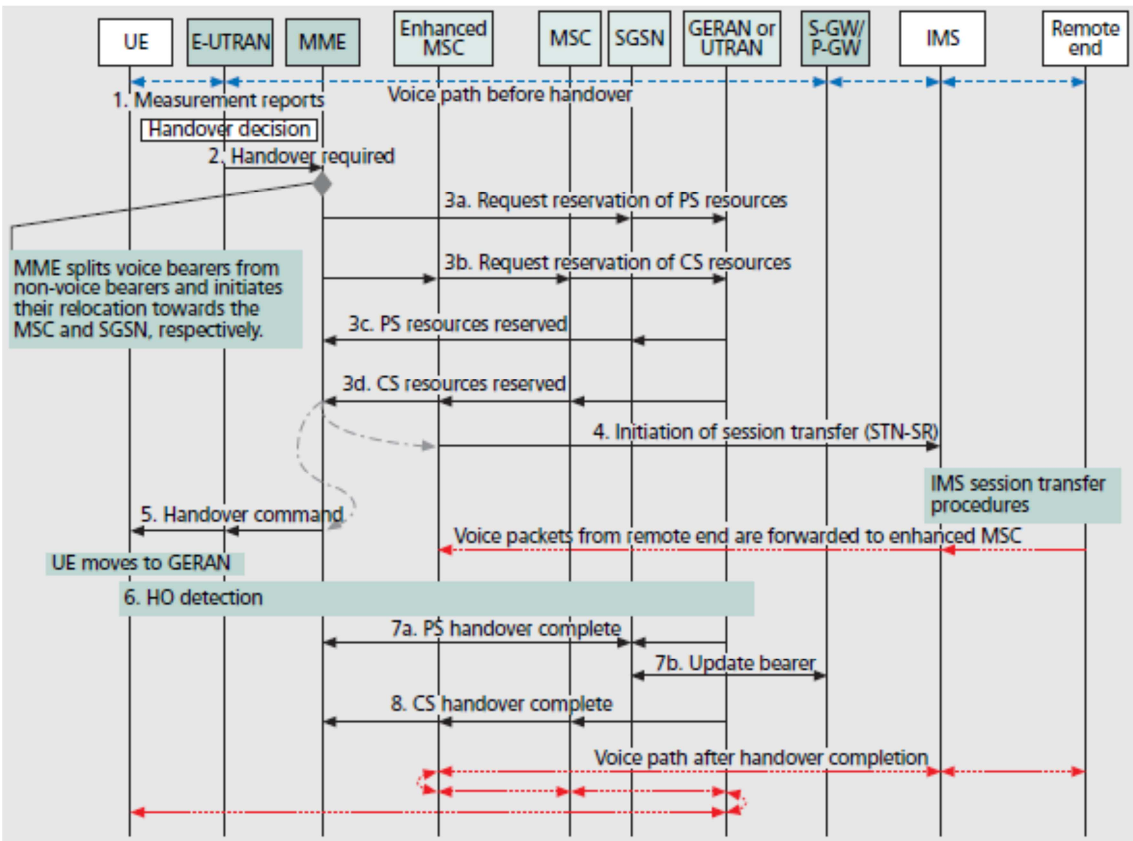


Figure 2.10 Messages transaction during a SRVCC call handover [6].

Once all these information are gathered by the eNodeB, it calculates the best network available to handle the services of the served UE and sends the request when it is time for handover to the MME. In the request it also specifies that the handover needs to be done using SRVCC. These messages are shown in messages 3a – 3d in figure 2.10. The MME needs to know that the handover is going to be SRVCC to enable the right resources for a packet switched to circuit switched handover or vice versa. In this case, the handover is from packet switched to circuit switched network.

Message 4 is a very important message in the SRVCC handover procedure. It is a new voice all request to a special number known as session transfer number for SRVCC (STN-SR). This number is associated with the active voice call between UE-A and UE-B in the IMS server from where the UE is performing a handover. STN-SR request is transmitted to the SCC AS, which matches the STN-SR number with an active voice call. Receiving a STN-SR number indicated SCC AS that the corresponding call needs to be routed through a different network and it starts the redirection process to the new endpoint. STN-SR is a unique number that is generated for each UE and is stored in the HSS. This number is sent to MME by HSS when the UE first connects with the network.

After the resources preparation is complete, MME responds to the handover request by the eNodeB. The eNodeB then transmits this response to the UE in step 5 which contains the required information for handover to be complete. The information includes carrier frequency traffic channel specifications, GERAN cell identity and also the description of the resources allotted to the UE. In the final steps, the UE is detected in the new GSM network and the call is re-established with the UE. The voice gap that will be created during this handover will start from step 4 and will stretch over till step 8.

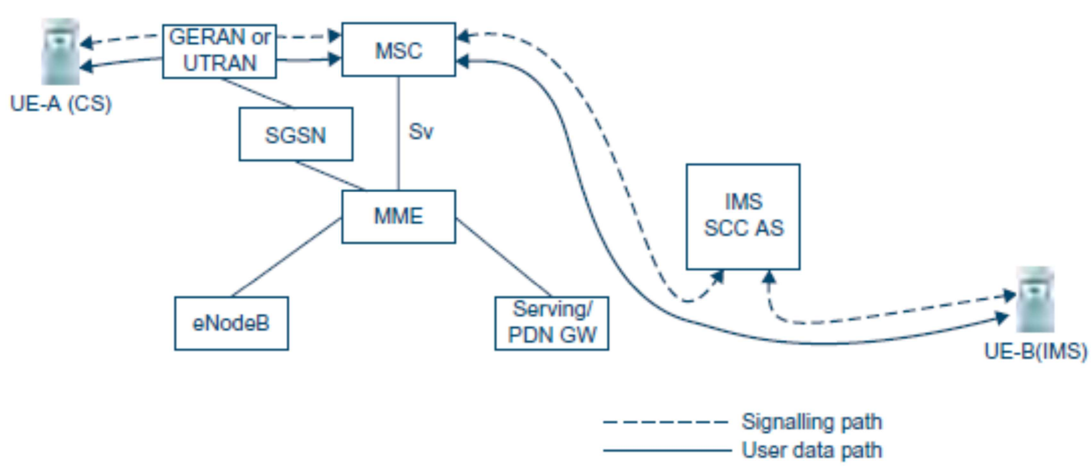


Figure 2.11 UE after the completion of SRVCC handover [4].

Voice packets and non voice packets can be handed over using this method. Though the voice handover is guaranteed, the non voice handover is not because the data speeds in a LTE networks exceeds the speeds of a GSM network.

2.8.2 Voice call from LTE to CDMA2000 (E-UTRAN to 1xCS)

SRVCC between EUTRAN and 1xRTT requires a 1xCS interworking solution (IWS) in the circuit switched domain to interwork with EPS through a new interface called S102 [6]. 1xCS IWS enables a UE with only one radio to communicate, in parallel, with both the source and the destination [7]. In this case it will enable the UE to talk to 1xRTT MSC while it is still connected to the E-UTRAN. Using 1xCS minimizes the voice lag that is created during a handover by establishing a CS access before the actual handover takes place. Special support is needed in MME for S102 interface and the MME merely acts as a relay station for signaling between UE and 1xCS IWS.

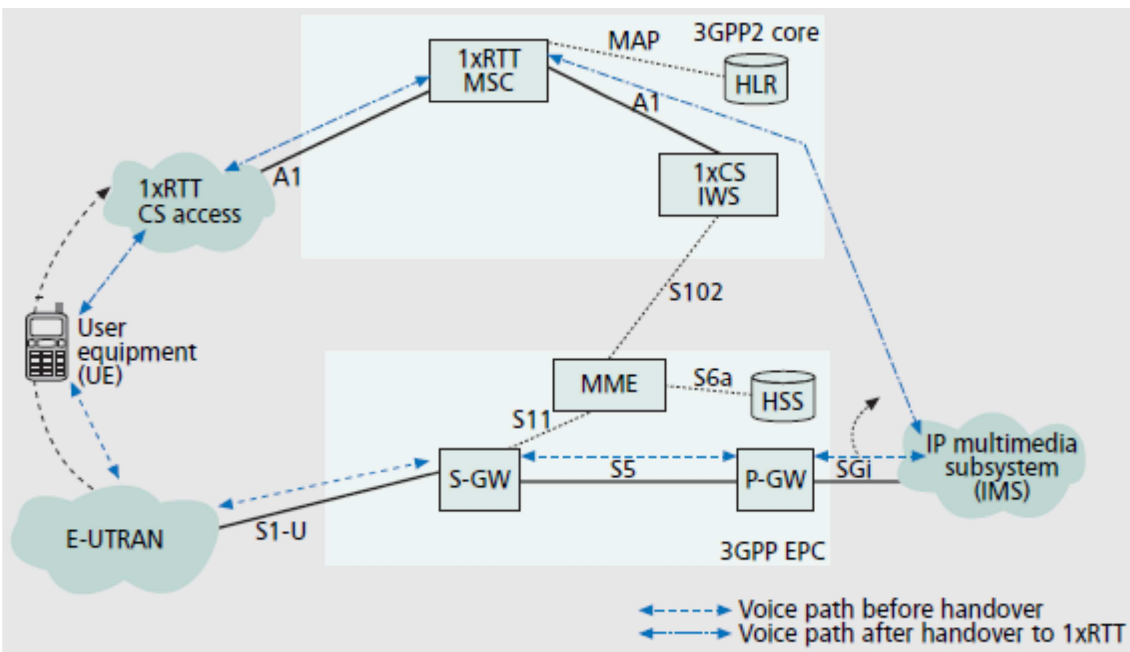


Figure 2.12 Voice call handover from E-UTRAN to 1xRTT architecture [6].

The SRVCC architecture was designed to handover only voice calls, unlike the SRVCC architecture of GSM which can handle voice as well as non voice calls simultaneously. Similar to GERAN SRVCC case, the voice call was established in E-UTRAN using IMS services. Once the request for SRVCC handover is received by the E-UTRAN, it will send the required parameters to the UE. These parameters include the required 1x parameters for a successful handover.

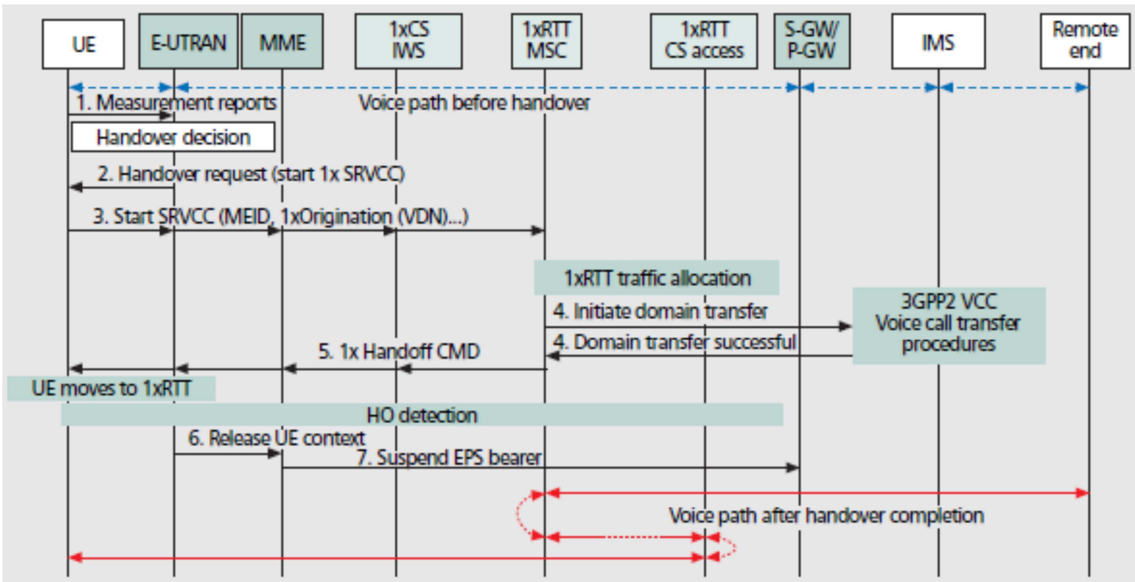


Figure 2.13 Message transactions during SRVCC for 1xRTT [6].

At Step 3 the UE starts the 1xSRVCC process by sending an uplink handover preparation transfer message. Upon receipt of the transfer messages, the MME selects a 1xCS IWS reference cell ID and the traffic channel resources are established. If the resource allocation was successful, 1xCS IWS will send a 1xmessage to the MME signaling the success of resource allocation. MME then sends the 1xmessage and other handover related information to the E-UTRAN which forwards it to the UE. After receiving this message the UE will perform the required change in frequency to connect to 1xRTT and all future voice data will be sent through that system. The EPC will stop carrying any voice related information for that particular UE. E-UTRAN then requests the MME to release UE related information and the S-GW suspends the EPS bearer.

2.9 Circuit Switched Fallback

Initially voice services will not be available on E-UTRAN, so to enable the user of E-UTRAN to be able to use the voice services, the users will be handed over to the traditional circuit switched networks. This handover to a circuit switched network will be a service switched handover. Whenever there is a request for a voice call involving an E-UTRAN user as the source or the destination, the E-UTRAN user will perform a handover. After the handover the call will be established using the same steps as it is done in a circuit switched network. Although this technique can handle voice calls in E-UTRAN, this is only a temporary fix for E-UTRAN and other networks not having compatible voice solutions.

Following characteristics need to be enabled for fallback to circuit switched domain to work:

- Mobility management is tied with the mobility management of EPS.
- UE receives termination command from EPS.
- For paging responses, call handling and originating calls 2G or 3G is used [6].

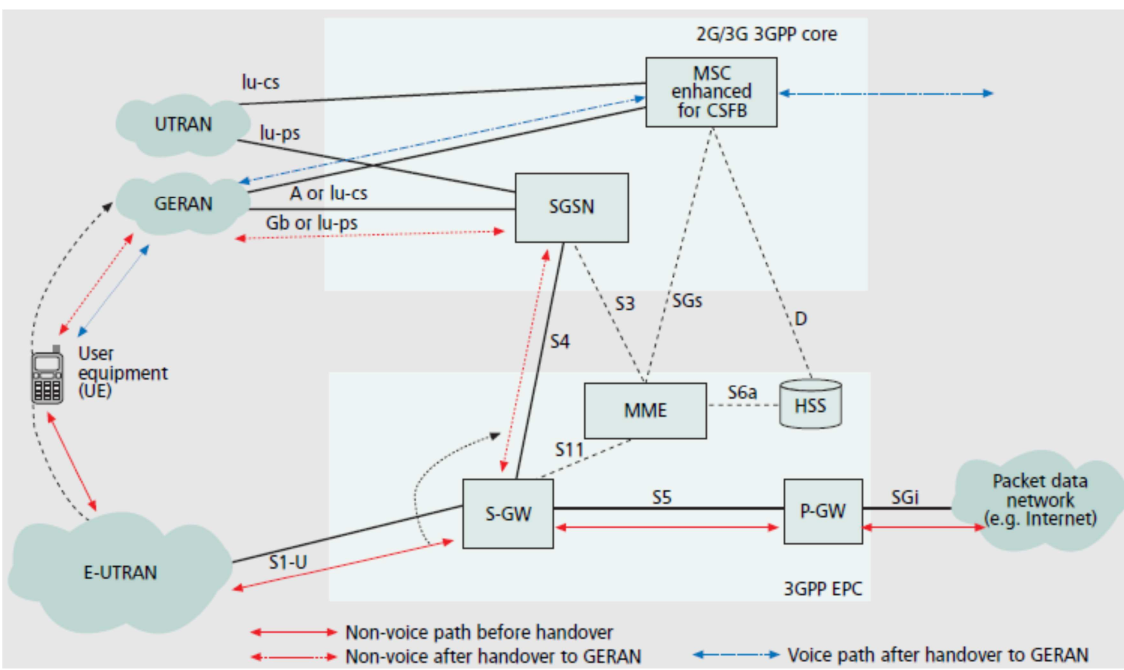


Figure 2.14 Architecture for circuit switched fallback [6].

Similar architecture can be used for CS fallback in 1xRTT. In the figure 2.14 it can be seen that the Mobile Switching Center (MSC) is connected with MME using a new interface called the SGs. This interface was introduced to reuse the concept of mobility management between PS and CS domains. It is important for GERAN to be notified about the location of the UE when it moves to E-UTRAN to deliver any incoming calls to the correct location. The MME performs the tracking of UE when it is in E-UTRAN and MSC uses most of this information to reduce computations.

2.9.1 CS fallback for mobile originating call

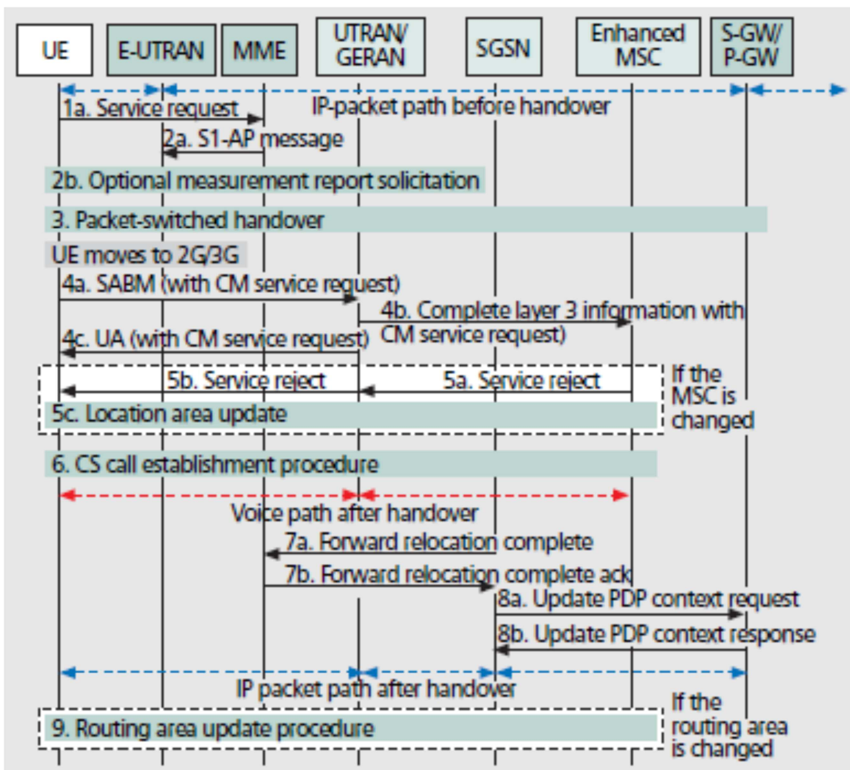


Figure 2.15 Message exchange during CS fallback with UE originating call while in E-UTRAN [6].

Any non voice data transmission taking place at the time the call is placed will either be handed over to the new network or it will be suspended until the UE returns back. If the non voice data transmission is also handed over, then it will operate at the lower speeds of the other network. The new network can also reject handing over of non voice data if it does not possess the capability to allow voice and non voice data transmission simultaneously.

The UE is initially in the E-UTRAN with active IP connectivity. When the UE decides to make a voice call, it sends a service request message to MME. MME checks if the UE is capable of handling a CS fallback and notifies the eNodeB to transfer the UE to a CS enabled network. Before handing over the UE, eNodeB may request RF measurements from UE of the neighboring GERAN or UTRAN networks. eNodeB then decides the best network for UE and performs the handover. This handover is a PS based handover in which the resources are first reserved in UTRAN or GERAN network and then a handover command is sent to the UE to move to the new network. If the location area stored in UE is different from that of the new cell, the UE will perform a location area update (Shown in step 5). After a successful fallback to a CS network, the UE will send a service request to MSC to establish the voice call. These steps are denoted in figure 2.15 from 1 to 4. By steps 7 and 8 the PS bearers are switched to the UTRAN or GERAN network and the UE performs any remaining handover steps.

2.9.2 CS fallback for mobile terminating call

In this scenario, the UE is in E-UTRAN and it receives an incoming call. The call is routed through MSC to which the UE is registered to. In step 1, the MSC sends a paging message to MME via SGs. This message is forwarded to the UE which is still connected to the E-UTRAN. This message will also include the caller’s ID. Including caller’s ID is beneficial because if the user refuses to take the call then the need for CS fallback can be eliminated. If the user accepts the call, it sends a service request to MME. The procedure for CS fallback is similar to the previous scenario in this case. Once the UE has connected to the new cell, it will send a service request to MSC for establishing the call. If the user decides to reject the call, a reject message is returned by the MME to the MSC.

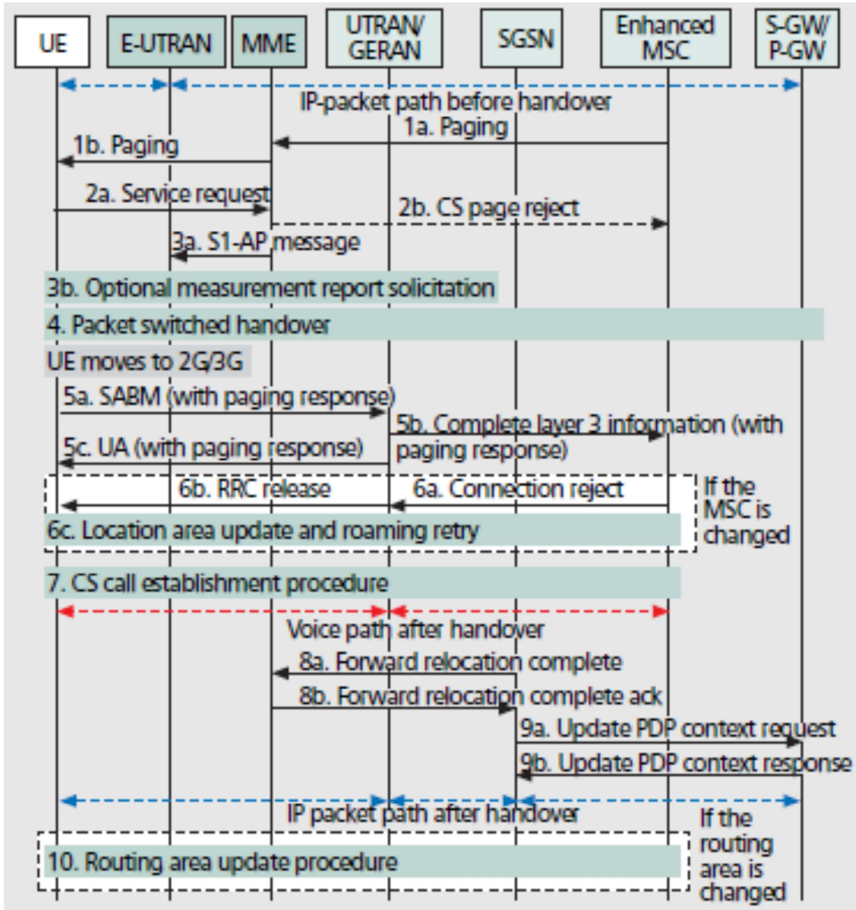


Figure 2.16 Message exchange during CS fallback with UE terminating call while in E-UTRAN [6].

CHAPTER 3

NETWORK QUALITY

3.1 Introduction

The demand for higher data rate and better call quality is omnipresent. These are the two main reasons we look for better technology for wireless and wired communication. This led to the two new technologies which are on the verge of changing the telecommunication industry forever: LTE and WiMAX.

Network coverage and capacity are two things which are of major consideration for any telecommunication technology. No matter how good the technology is if the network supporting that technology is not good enough, the technology is going to fail. Particularly if the technology is new, it will take time to get the same or better coverage as any of the previous technologies. Even with these two new technologies we need to make modifications in the way the networks are laid down.

Along with coverage and capacity, another important factor that needs to be considered is quality. Getting a user to connect to a network and to stream data to the device is very easy. Satisfying the user with the quality of data and service is completely other story. With faster speeds, applications like VoIP, MBMS and other high capacity and real time services will be implemented on today's telecommunication technologies. Since, both WiMAX and LTE are completely IP based networks the mode of data transfer is going to be through packets. These packets can take any possible route to reach their destination which may leave the receiver with a lot of out of order packets. Arranging the packet and providing them to the higher layer will cause a delay and eventually poor service to the user.

These networks need to provide quality of service (QoS) to these packets so that all the real time and time sensitive packets get more importance as compared to other packets on the network. Overall the quality of a network is judged on the basis of good coverage, quality and high capacity.

3.2 EPS bearer for E-UTRAN access

The EPS is responsible for providing the QoS needs for different services supported by LTE. EPS bearer is the central node for providing not just the IP connectivity to users but also providing QoS. EPS bearers or simply called as the bearers, are responsible for providing IP transport channel between the UE and packet data network. A PDN connection has to have at least one EPS bearer. It can also have multiple EPS bearer for different quality of service requirements. Each bearer can be identified by its properties to provide a certain type of QoS. So each bearer has a set if parameters which decide its delay, jitter, bit rates, error rates etc. To have different quality of service in a network, a network needs to have different bearers with different properties. Data sent on a bearer with the same properties will experience the same QoS.

A bearer which is established when a PDN connection first established is called as the default bearer. Usually a default bearer does not have high end QoS specifications and is associated to non real time data needs. A default bearer will remain active throughout the lifetime of the PDN connection. Once a default bearer is deactivated, it will close down the whole PDN connection. After the default bearer has been setup, the service which is being served can request for enhanced bearer setup. These bearers are called as dedicated bearers and may be activated on demand.

3.2.1 User plane

In the user plane, the traffic between P-GW or S-GW and UE is mapped using packet filters. These filters check the IP traffic for specific packets and depending on pre-defined rules, relate the IP traffic with a bearer. Each EPS bearer has its own Traffic Flow Template (TFT) and packet filters are included in the TFT. These packets filters can work on uplink data as well as the downlink data. When a new EPS bearer is established, a new TFT is also created for the bearer. At the time of creation of the EPS bearer and considering the type of services that the bearer is assigned to, filters are added to the TFT.

At the time of start of service, the filters will check the IP packets and if they match, the bearer associated with that filter will carry the user plane for that session. These filters usually define 5 things.

- Source IP address
- Destination IP address
- Source port address
- Destination port address
- Protocol type identifier

Along with the above information a packet filter may be programmed to search for packets based on additional information such as:

- Protocol Number (IPv6/IPv4)
- Subnet mast and remote IP address
- Local port range
- Remote port range
- IPSec parameters
- Type of service

Figure 3.1 shows an example of how TFT is used to filter the packets and provide an appropriate bearer for the service. The procedure starts with UE connecting to the media server in P-GW. The connection is complete when correct QoS parameters are assigned for the service. After this step, packet filters are installed in UE and the P-GW. P-GW, directs all the user data towards the newly established ES bearer. Policy and charging control (PCC) may also be used to ensure that the right QoS and TFT have been established for the service.

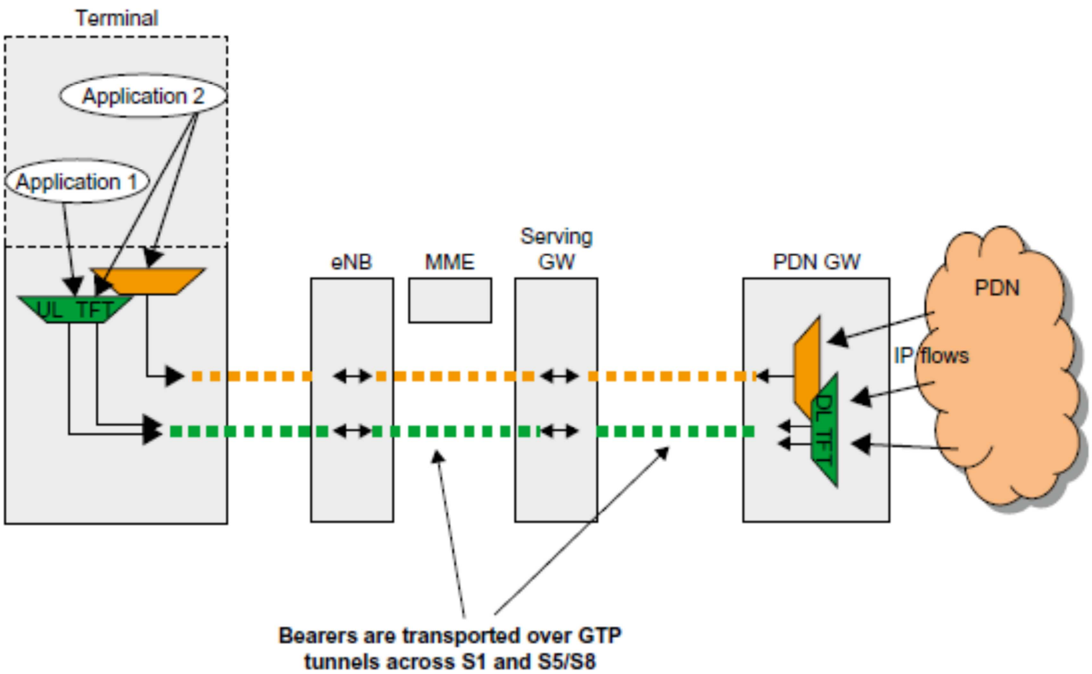


Figure 3.1 Establishment of EPS bearer for a LTE network [4].

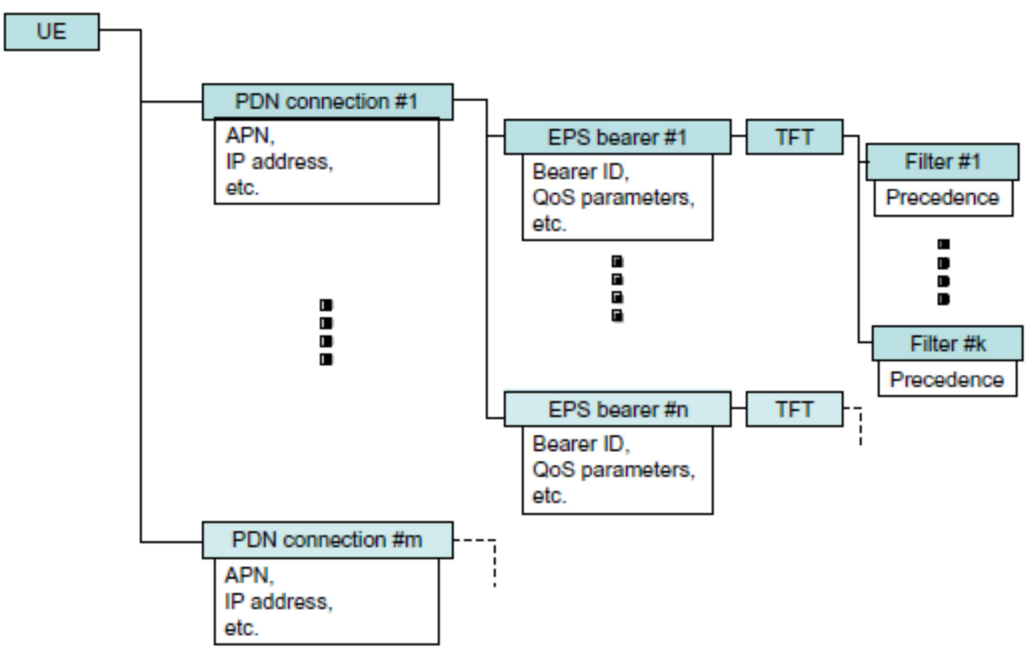


Figure 3.2 Relation between UE, PDN connection, EPS bearer, TFT and packet filters [4].

With the establishment of a new EPS bearer, an EPS bearer context is also created. This context is shared with all the components of the EPS structure which includes S-GW, UE, P-GW, MME and eNB, although the content of the context will be different in each of these nodes.

The user plane traffic belonging to a user will be transported using an encapsulation header called as the GTP-U. Figure 3.3 shows a user plane packet encapsulated using GTP-U.

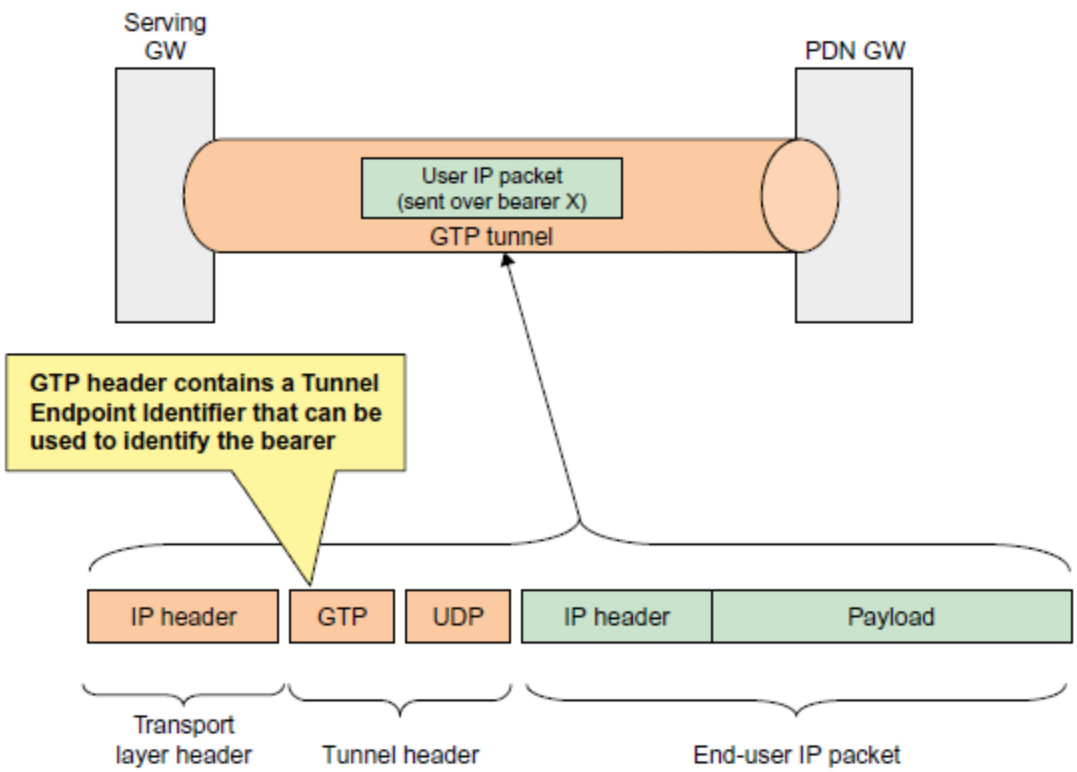


Figure 3.3 User plane packet encapsulated by GTP-U [4].

3.2.2 *Control plane*

The role of the control plane is to activate, modify and deactivate bearers. It also provides QoS parameters and packet filter to bearers. In order to provide better control over the network, the EPS only allows the P-GW to activate or deactivate a certain bearer. It is also responsible for providing QoS and directing user plane data to the appropriate bearer for complete user satisfaction in terms of QoS. In earlier technologies like GSM, it was the UE that was responsible for this role.

3.3 Bearers in Proxy Mobile IP (PMIP)

PMIP was designed by IETF to provide a common access technology which will be independent of the network the user is on. PMIP has no inbuilt support for QoS. So to make it compatible with all the networks which support QoS, certain modifications need to be made to the networks. Figure 3.4 shows how EPS bearers are established for providing QoS when PMIP is used.

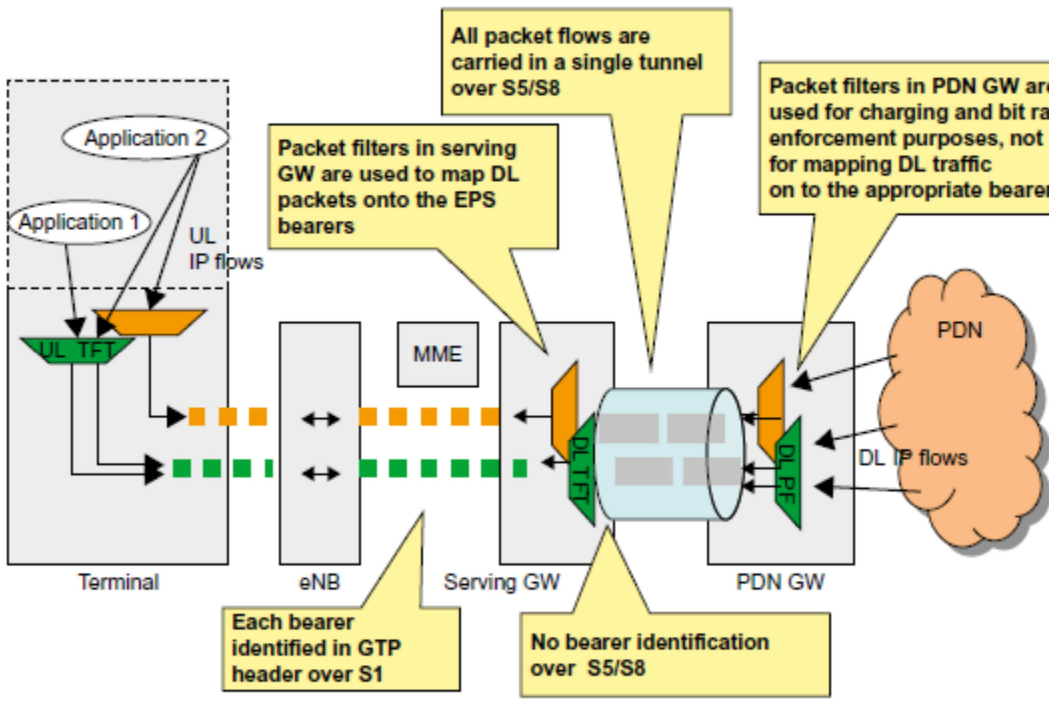


Figure 3.4 Support for QoS in a PMIP bases system [4].

Unlike a GTP based system where the bearers are extended from the P-GW to the UE, PMIP system only has the bearers extended from the S-GW to the UE. As a result the packet filters in P-GW are not used for assigning user data with a bearer. Instead this job is assigned to the S-GW in a PMIP system. An important point to note is the even though P-GW does not perform assignment of bearer for user data, it still needs packets filters. The packet filters allow the P-GW to perform other important functions like bit rate enforcing and charging [4].

3.4 Quality of Service (QoS) in LTE

3.4.1 Differences from GERAN

Packet switched network were not designed with telecommunication and other real time services in mind. They were designed to transport message from the source to destination by the best, shortest and the cheapest route possible. In the last couple of years, the packet network and its advantages have been realized by the telecommunication world and they are trying hard to use the packet network for their services. The concept of QoS has been in the industry for quite some time. It was introduced during the time of GERAN and GPRS.

Back then the UE was in complete control of the resource allotment and QoS request. UE could initiate a new bearer which could provide the required QoS depending on the needs of the UE. Designers learned that this is not the best idea and implemented a more network controlled approach. Now in LTE, the UE can only request for resources from the network. The final decision to grant those resources lies in the hand of the network.

When QoS was first introduced, it had many parameters by which a bearer could be associated and a different QoS class be formed. These parameters involved bit rates, error rates, and transfer delays etc. which lead to the formation of a very complicated algorithm to assign the right QoS to a device or a service. In the latest release by 3GPP, only 3 parameters have been identified. These parameters are:

- QoS class identifier (QCI)
- Allocation and retention priority (ARP)
- Bit rates

3.4.2 QoS parameters for EPS bearers

Most EPS bearers can be associated by 2 QoS parameters, which are QCI and ARP. QCI is used to specify the kind of user plane treatment an IP packet associated with a given bearer should receive while ARP specifies the type of control plane treatment a bearer should receive.

3.4.2.1 QoS Class Identifier (QCI)

QCI is used to filter the IP packets in a network between real packets and non real time packets. Once the packets have been identified as real time and non real time, the network can successfully forward the packets to their respective bearers. QCI is a number which is used to refer to a particular QoS. It does not, in any way, represent the QoS of the network or the service that it is used for. The reference number provided by QCI is used to determine if the packet is real time or not. The QCI parameters will be pre-configured in the nodes by the vendor and are vendor dependent. Some of the QCI values have been standardized and are referenced to specific QoS. These values are shown in table 3.1.

Table 3.1 Standardized QCI values [4].

QCI	Resource Type	Priority	Packet Delay Budget (ms)	Packet Error Loss rate	Services
1	GBR	2	100	10^{-2}	Voice conversation
2	GBR	4	150	10^{-3}	Voice conversation (live streaming)
3	GBR	3	50	10^{-3}	Real time gaming
4	GBR	5	300	10^{-6}	Non conversational video (buffered video)
5	Non-GBR	1	100	10^{-6}	IMS signaling
6	Non-GBR	6	300	10^{-6}	Video (buffered streaming)
7	Non-GBR	7	100	10^{-3}	Interactive gaming
8	Non-GBR	8	300	10^{-6}	Video (buffered streaming)
9	Non-GBR	9	300	10^{-6}	Video (buffered streaming)

3.4.2.2 Allocation and Retention Priority (ARP)

ARP decides the importance of a bearer connection. This helps the network to decide if the requested resources should to be assigned to a particular service or a UE if there is resource crisis. ARP also comes in play when any modification to the service is requested. In

GERAN, only 3 ARP values were supported and were considered sufficient as all the emergency calls were made on the CS network. Since in E-UTRAN all the calls will be made on PS network including emergency calls, it is important to have higher number of priorities. EPS will support 15 ARP values.

A call of higher importance is assigned a lower ARP value while a call of lower importance will be assigned a higher ARP value. As a result, a call with lower ARP value is more likely to get accepted as compared to a call with high ARP value in a time of resource scarcity.

3.4.2.3 Guaranteed Bit Rate (GBR) and non-Guaranteed Bit Rate (non-GBR) bearers

Along with the QoS parameters specified above, the network uses GBR to provide better quality of service when required by certain services. A bearer will either be associated with a GBR or a non-GBR. In the case the bearer is associated with a GBR, it will have a guaranteed bit rate and maximum bit rate (MBR) pre-defined. If the bearer is associated with a non-GBR, it will not have any bit rate pre-defined.

GBR means, that a service running on a bearer with GBR will at least have the specified bandwidth that was pre-defined by the network. This reserved bandwidth is independent of the fact that the service may never use all the bandwidth that it reserves. In other words, a service operating with GBR will not experience any packet loss if the network is under congestion. This is guaranteed because a GBR bearer is subjected to admission control. So if there are not enough resources to support a particular GBR bearer, it will not be setup.

A downside to this reservation is that the resources are reserved in the network even if there is no traffic. Although there is a minimum guaranteed bandwidth reserved for a service operating with GBR, it cannot exceed the maximum specified bit rate. Any traffic which exceeds the MBR will be discarded by the network. The current version of EPC only supports a GBR in which the MBR is equal to itself.

A non-GBR does not guarantee any bit rate and is not associated with any minimum or maximum bandwidth allocation. A non-GBR bearer may experience packet loss if there is any congestion in the network. The resources allotted to a particular non-GBR bearer are dependent on the available resources and the QCI value of the bearer. Although there should be no maximum bandwidth association with a non-GBR bearer, a vendor may still impose a limit on the bit rate calculated by the Aggregate Maximum Bit Rate (AMBR) [4].

3.4.3 APN-AMBR and UE-AMBR

AMBR parameter is defined for the non-GBR bearers. A similar concept was in place for GERAN but the maximum limit was placed on each bearer. This method allows the limit to be placed on the aggregate of non-GBR bearers. The reason for having an aggregate policy rather than individual one is that it is far easier to estimate the bandwidth usage for a bunch of users. APN-AMBR defines the total bit rate which can be used by all of the non-GBR bearers related to a specific APN. This limitation is independent of the number of connections PDN connections.

The UE-AMBR is defined per user and specifies the maximum limit a user can have in terms of total bit rate. UE-AMBR is specified in the subscription profile of the user. If the user is connected to a non-GBR bearer with APN-AMBR as well, the eNB will set the bit rate as the minimum of the two AMBRs. Different AMBR can be assigned for uplink and downlink. So there are total of four different AMBRs that can be defined:

- UL APN-AMBR
- DL APN-AMBR
- UL UE-AMBR
- DL UE-AMBR

3.4.4 User plane handling

With all the parameters defined, the network needs to use these parameters together to get the best possible QoS for its users and services. UE and S-GW or P-GW will take care of filtering the data packets and assigning the right bearer for these packets. They will handle the data packets for uplink and downlink separately.

GW and eNB will enforce admission control and preemptive handling. This allows them to ward off any connection which they find to be too resource consuming and can negatively affect the conditions of the network. In case of congestion, they can reduce some load or limit the entry to the network. They can use ARP to decide which connections to be given more priority and which connections to be neglected at a time when resources are scarce.

eNB and GW also take control of rate monitoring. This function is important to check any misuse of rates that have been assigned to a user or a bearer. They use AMBR and GBR to figure out the maximum and minimum assigned bit rates to users and bearers and accordingly control traffic. Using these value, they can also predict if taking in new connections will lead to congestion or not and can eliminate any new connections based on the results.

3.4.5 Working along with GERAN/UTRAN

As discussed above, the QoS model of a GERAN/UTRAN system is very different from an E-UTRAN system. As a result a bridge is needed between the two technologies to provide QoS on both sides of the networks. Two main alternatives have been defined to handle QoS in GERAN/UTRAN:

- Modify the GERAN/UTRAN networks to use the QoS architecture of the EPS system. This would have to be implemented after doing many changes to the GERAN/UTRAN system. The changed system will also have to be backward compatible for those older devices to be able to use QoS services.
- Leave the QoS architecture of GERAN/UTRAN completely unchanged and specify a new mapping technique to ES based systems.

Since GERAN/UTRAN are widely spread and making changes to them would mean spending a lot of time and money, 3GPP choose the second option of mapping the QoS architecture of GERAN/UTRAN over to EPS based networks.

While mapping rel-99 with rel-9 first thing that should be kept in mind is that ARP for rel-99 has only 3 possible values. In rel-9 there are 15 different values. EPS bearer parameter APR is mapped one to one to GERAN bearer parameter ARP [8]. GBR and MBR values are also mapped one to one for all the GBR bearers. Non-GBR bearer's MBR for a given PDP context is mapped with APN-AMBR. To bring the concept of AMBR in GERAN/UTRAN, the MBR values of all the bearers should not exceed that of EPS bearer parameter AMBR.

In case of a handover from GERAN/UTRAN, the AMBR specified in the EPS QoS profile of the bearer will be given higher preference. So, in this case, any MBR values assigned by the UTRAN/GERAN network will be ignored.

3.5 Relaying in LTE

Relaying has been extended from rel-8 to be used as a tool to improve the network quality by improving the coverage of high data rates, improving capacity at cell edges, allowing temporary deployment of network and to provide coverage in new areas.

A relay node (RN) acts exactly like an eNB from a UE's perspective. A UE does not need to know if it is connected to a RN or directly to eNB. If the UE is connected via a RN to the eNB and the EPC, the connection will look similar to figure 3.5.

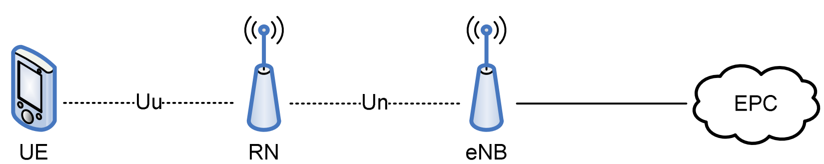


Figure 3.5 UE connected to the LTE network using a RN [9].

As shown in figure, a UE will be connected to the RN. The RN itself is connected to the eNB. This eNB is called as a donor eNB (DeNB).

3.5.1 Relay architecture

3GPP has decided to come out with two variants of the relay architectures. These variants are as follows:

- Architecture A
    - Alt 1: Full L-3 relay, transparent for DeNodeB (Donor eNodeB)
    - Alt 2: Proxy S1/X2

- o Alt 3: RN bearer terminate in DeNodeB
- Architecture B
  - o Alt 4: S1 UP terminated in DeNodeB [10]

3.5.2 Architecture A

Architecture A is based on user plane and control plane of S1 interface (interface between eNB and MME) terminating at the RN. As mentioned before, this architecture has 3 variants.

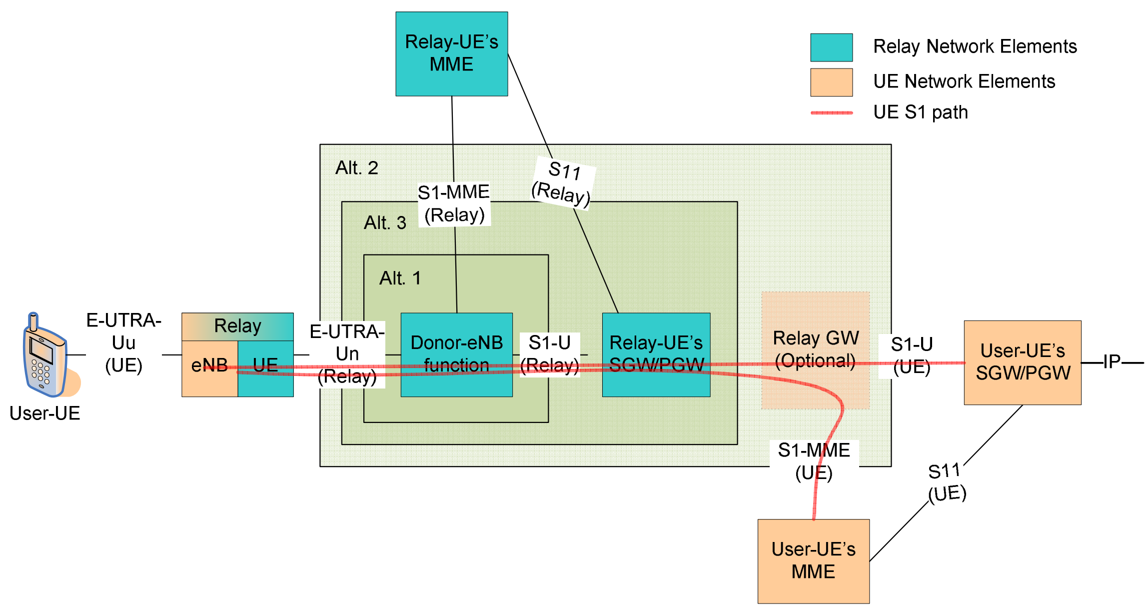


Figure 3.6 Architecture A with all three alternatives [10].

All the three alternatives share the characteristics of Un interface between the RN and the DeNB. For all the three alternatives, the S1-MME interface is unmodified but for alt. 2, it ends in a proxy sense in DeNB while for others it terminates at the RN. X2 interface, which connects DeNB with itself, is also unchanged. The functionality of the DeNB is also the same as any other eNB. All the three approaches for providing architecture for relaying in LTE are transparent to the RN and the UE.

Relay GW shown in the figure 3.6 is an optional and transparent choice of the network designer. It will function as a home eNB GW and is included in alternative 2 but is left out in others.

The Un interface used to connect a RN with a MME needs to be compatible with all the alternatives. This will allow the network designers to make changes to the network depending on the performance with a particular alternative.

3.5.2.1 User Plane Aspects

In architecture A, the UP for S1 interface is terminated at RN. Figure 3.7 shows a UP protocol stack of alternative 1. When a UP packet is sent to the RN to be delivered to a UE served by the RN, it is delivered by the RN GW. In this process the packets are destined to the RN as if it was the UE and they are forwarded to the UE as if they were being sent by the eNB.

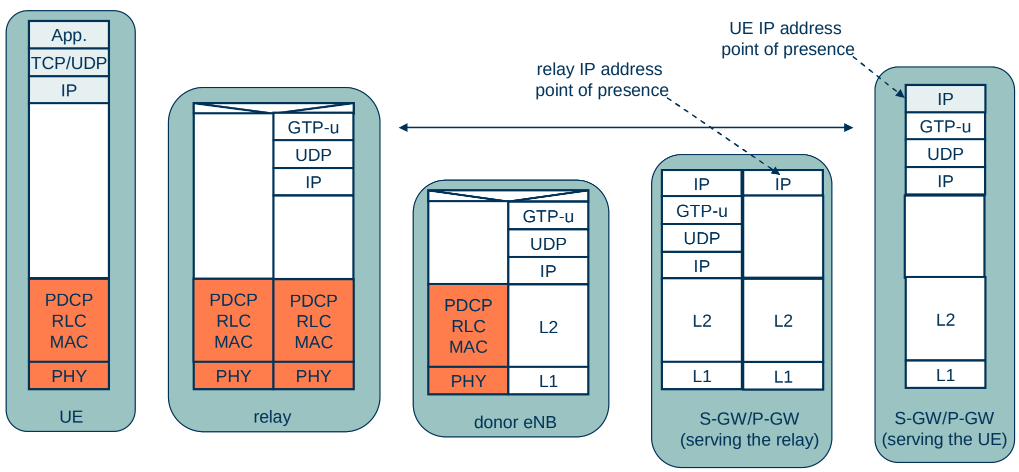


Figure 3.7 User plane protocol stack for alternative 1 [10].

A packet which is destined for a UE will be packed and encapsulated in the appropriate GTP tunnel assigned to the UE by the P-GW serving the UE. The P-GW also classifies an EPS bearer to the UE according to the TFT filtering. The GTP tunnel will be the path between the S-GW to the UE and will also include the RN. The RN which is serving the UE will also have to map the UE bearer with a RN bearer. The P-GW for the RN receives the GTP tunneled packet and using the filtering rules within RN, a new RN bearer will be assigned to the packet. This packet is tunneled again into a second GTP and sent to the DeNB. Since the RN bearers are mapped with EPS bearers, a packet with same QoS will be mapped into the same RN bearer. The DeNB will associate a RN radio bearer wilt the RN tunnel bearer and send the packet to RN. After receiving the packet, RN will assign a UE radio bearer to the packet and forward the packet to UE.

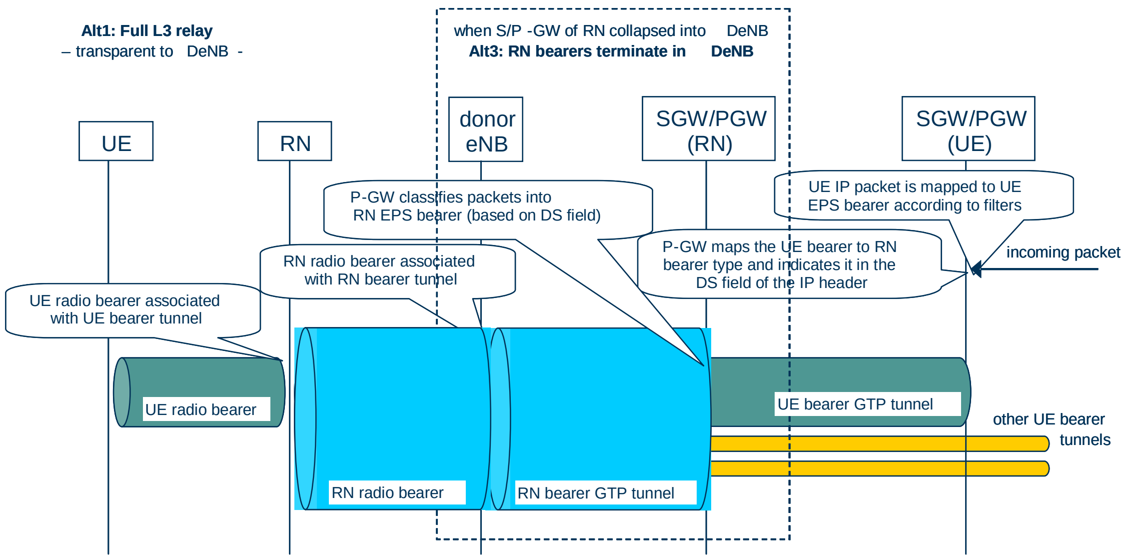


Figure 3.8 Message delivery step for alternative 1 and 3 [10].

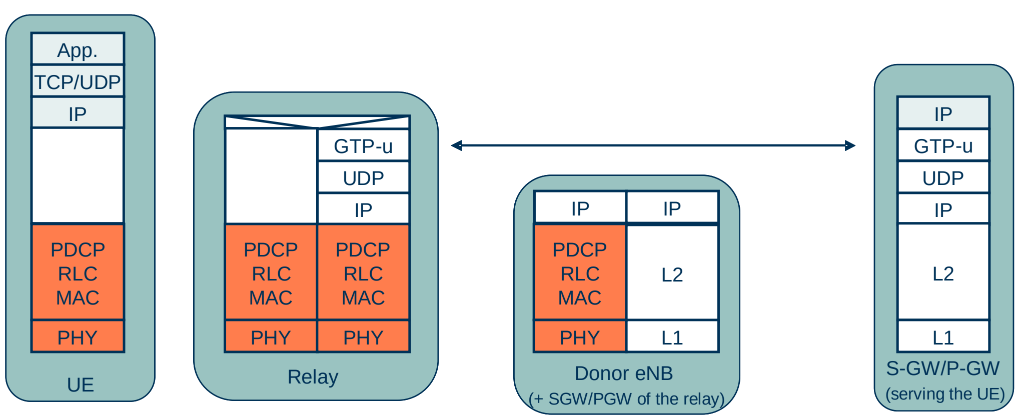


Figure 3.9 User plane stack for alternative 3 [10].

Figure 3.9 shows an optimized user plane protocol stack for alternative 3 in which the P-GW and S-GW are merged for DeNB and RN. This reduces an extra step of going through two GWs. The rest of the functionality for routing and encapsulation remains the same as that in alternative 1.

In alternative 2 of architecture A, a home eNB GW type of functionality has been added to the DeNB. This functionality gives a proxy S1/X2 architecture to alternative 2. In this type of architecture, there is a GTP tunnel per UE, which is going from P-GW/S-GW of the UE to the DeNB. At the DeNB, it switches to another GTP tunnel. This tunnels runs from DeNB to the RN. Figure 3.10 shows the user plane for alternative 2.

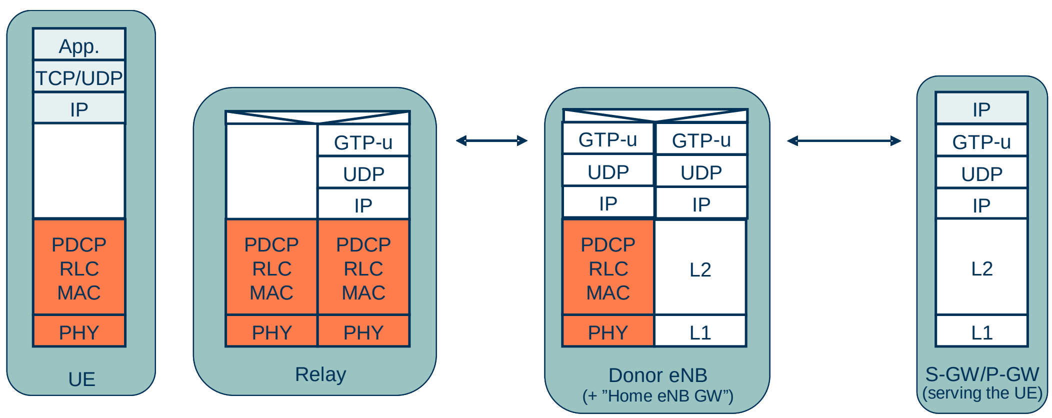


Figure 3.10 User plane protocol stack for alternative 2 [10].

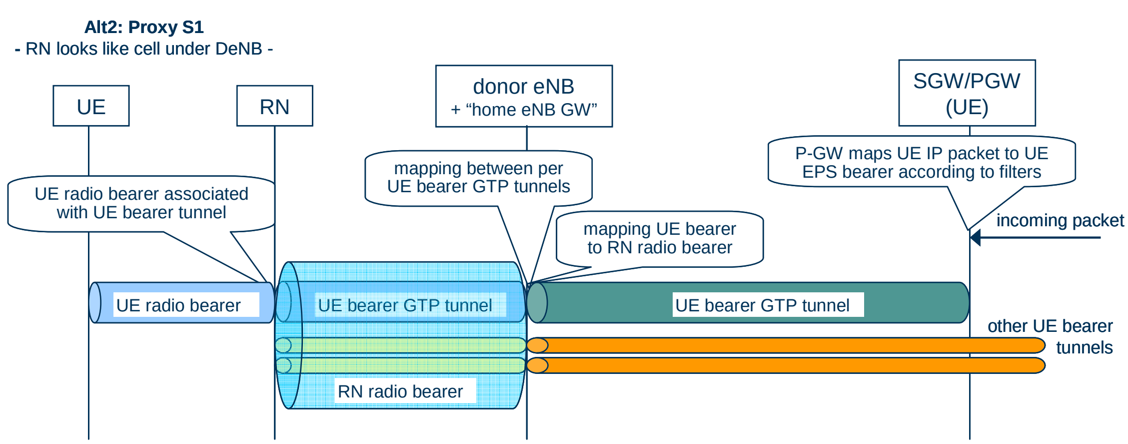


Figure 3.11 Message delivery steps for alternative 2 [10].

The steps for delivering the packets in alternative 2 are shown in figure 3.11. The packet is assigned a UE bearer by the P-GW serving the UE. This packet travels to the DeNB in the corresponding GTP tunnel. Depending on the QCI of the UE bearer, the DeNB assigns a RN radio bearer to the packet. It also changes the UE bearer GTP tunnel, which was assigned to the packet by the P-GW serving the UE, to another GTP tunnel. Even in this case, the UEs that have similar QoS will receive the packets from the RN in the same radio bearer. When the packets reach RN, it assigns a UE radio bearer to the packets and these packets are sent to the UE.

3.5.2.2 Control Plane Aspects

The S1-AP protocol, which is used to send signaling messages between eNB and the MME end at RN. The signaling messages move to MME via the DeNB and the S-GW and P-GW of the relay node. S-GW and P-GW of the relay node act as the user plane nodes for the signaling. The signaling messages sent between the RN and MME will be sent via the user plane EPS bearer of the RN. Figure 3.12 shows how the control stack protocols are implemented for alt. 1 and 3.

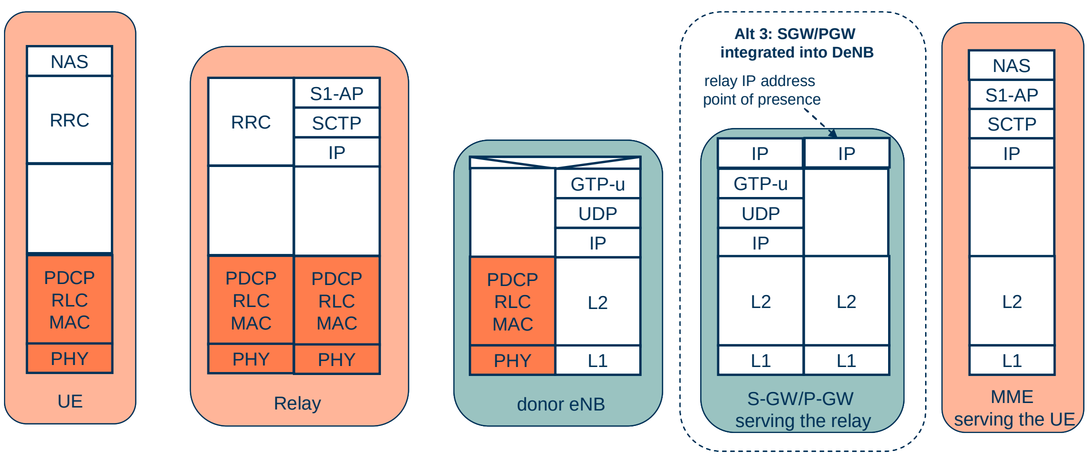


Figure 3.12 Control plane protocol stack for alternative 1 and 3 [10].

There is at least one S1 interface between RN and each MME in the MME pool. In this pool, there is at least one S1 signaling connection for each connected UE on the given S1 interface between the RN and the MME serving the UE. These S1 interface and the signaling connections are transparent to DeNB. The DeNB also has a S1 interface and a S1 signaling connection corresponding to the RN with the MME serving the RN. In this case, the RN is considered as a UE.

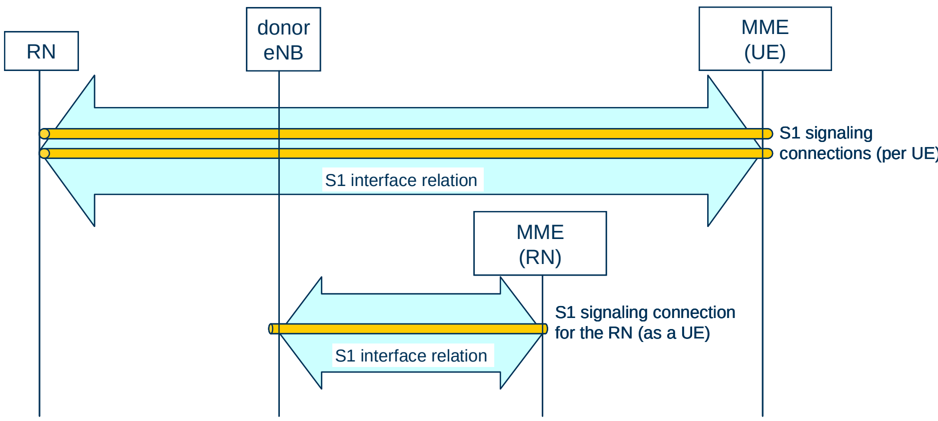


Figure 3.13 S1 interface relation and signaling connection for alternative 1 and 3 [10].

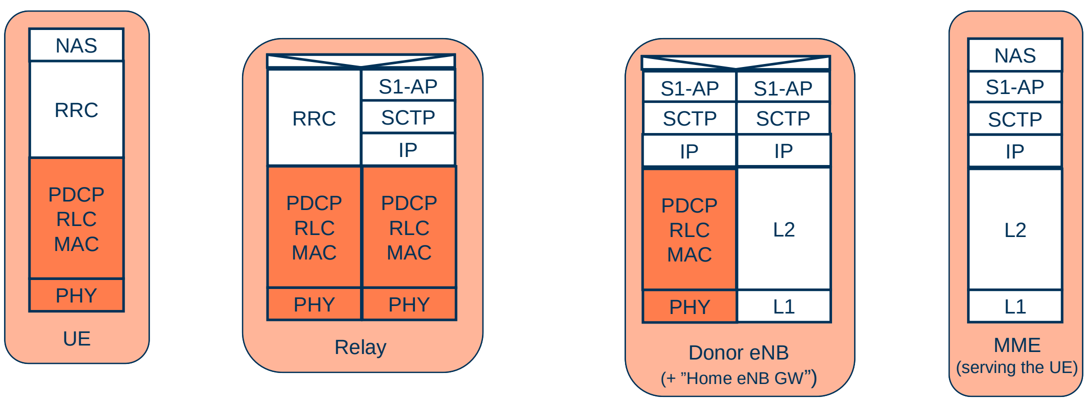


Figure 3.14 Control plane protocol stack for alternative 2 [10].

In case of alternative 2, the S1-AP messages are sent between MME and DeNB and DeNB and RN. When the DeNB receives a packet bound for a certain UE, it will change the UE ID by modifying the S1-AP UE ID of the packet. The rest of the packet is left untouched. The changing of the UE ID corresponds to the S1-AP proxy. From the MME's perspective, the UE is directly connected to the DeNB, while from RN's point of view it is directly connected to the MME.

The S1 interfacing for alternative 2 is shown in figure 3.15, in which there is one S1 interface relation between RN and the DeNB and another between DeNB and MME. All the S1 signaling messages are processed by the DeNB. It can be clearly seen that the DeNB has to maintain two S1 interface connections while the RN only maintains one. The S1 interface relation and signaling connection treat the RN as a UE.

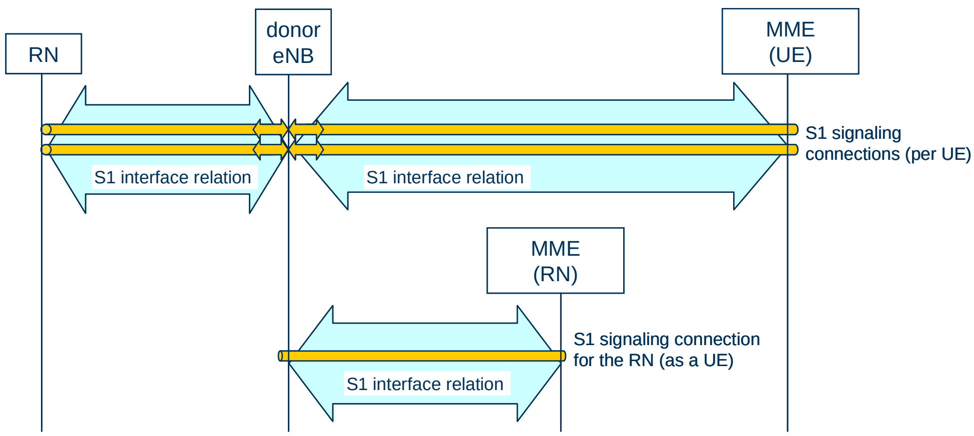


Figure 3.15 S1 interfacing relation and signaling connections for alternative 2 [10].

3.5.3 Signaling for architecture A

3.5.3.1 UE Access Procedure

The attach procedure by a UE to the RN is very similar to the attach procedure followed by a UE if it were to connect to an eNB. The attach procedure for UE in alternative 1 and 3 is shown in the figure below while the attach procedure for alternative 2 is shown in figure 3.17. The procedure is the same for alternative 1 and 3, except that the S-GW and P-GW serving the RN will be combined with the DeNB. The procedure for alternative 2 is also similar to the legacy attach procedure for a UE. The only difference is that the DeNB is also involved in the attach procedure. It is used to send signaling messages between RN and the MME. Since the proxy functionality of DeNB is used in sending the signaling messages, it clearly knows about the UE connecting via the RN.

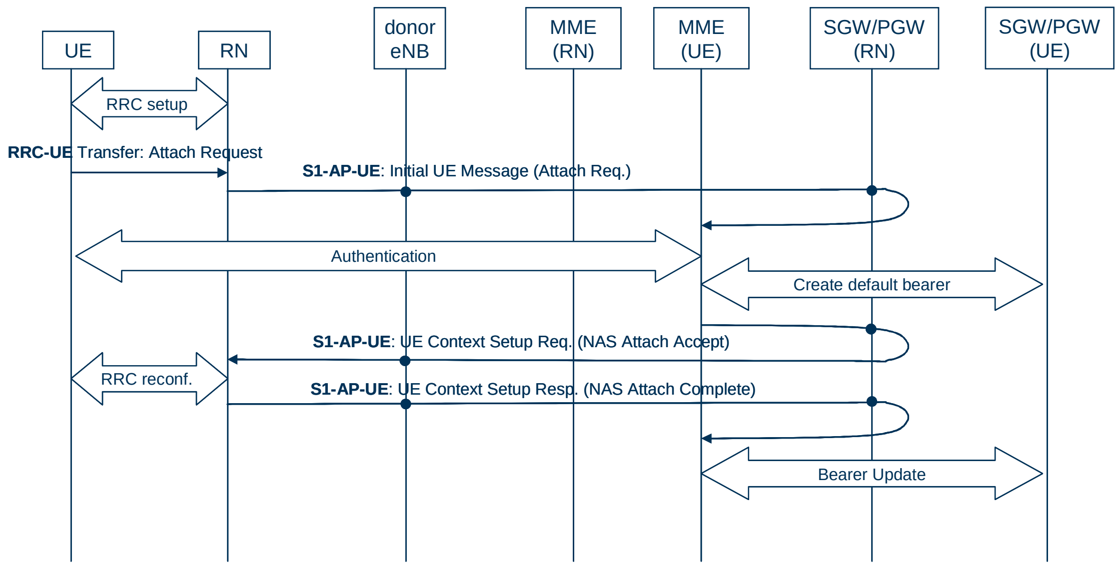


Figure 3.16 Attach procedure of UE with RN for alternative 1 and 3 [10].

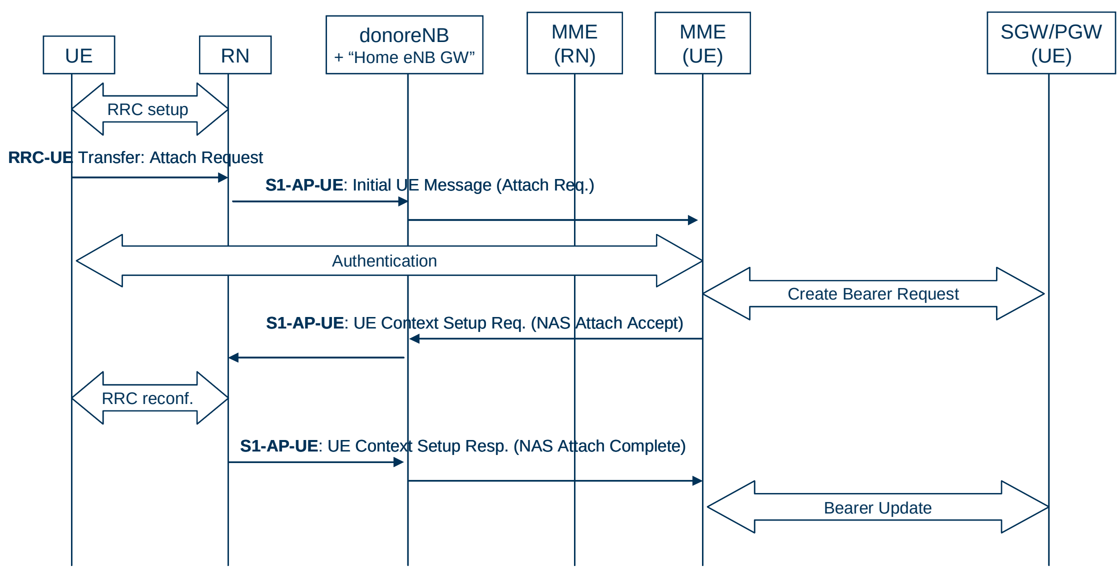


Figure 3.17 Attach procedure of UE with RN for alternative 2 [10].

3.5.3.2 Handover Procedure

The handover performed by the UE under a RN to another eNB or DeNB is shown in the figure 3.18. This procedure is for alternative 1 and 3. The procedure for UE making a handover from one RN to another RN is going to be very similar.

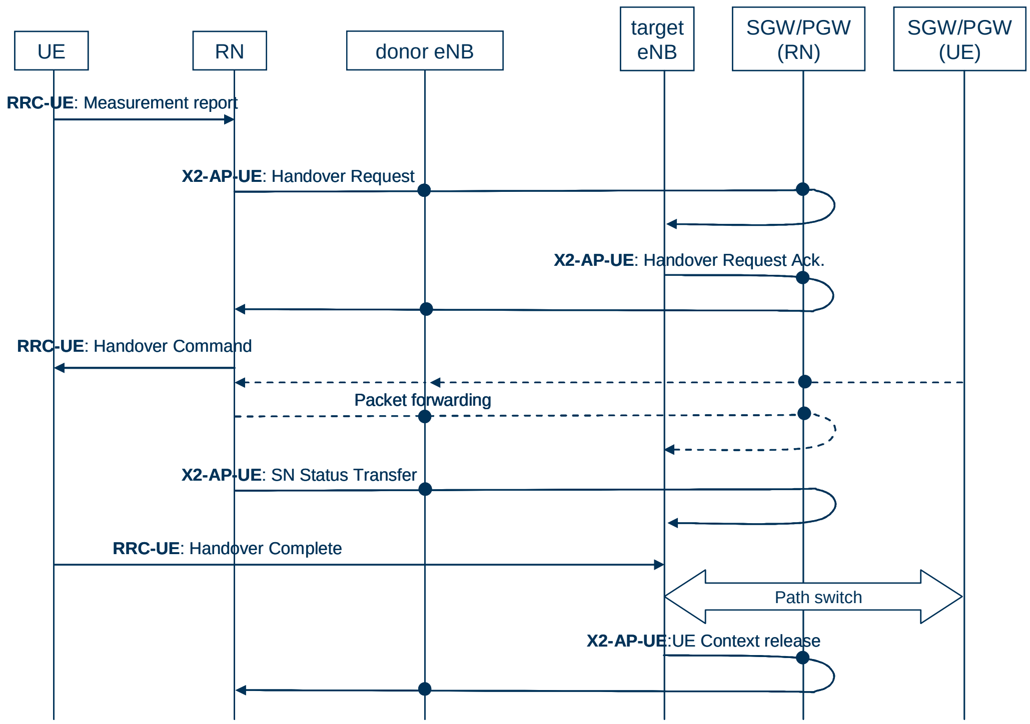


Figure 3.18 X2 handover from RN to target eNB for alternative 1 and 3 [10].

The UE can request for a handover by sending the measurement of the current RN signal and neighboring RN signal strength. Based on the measurements received by the UE, Rn will make the final handover decision. It also selects a target cell for the UE. RN sends the Handover Request message to the target eNB via EPS data bearer. The ES data bearer is provided by the DeNB and the P-GW/S-GW serving the RN. The target eNB receives the request and sends a Handover Request Ack message back to the RN. This message takes the same route backwards. To the target eNB, the handover request was sent by another eNB. The X2 signaling is complete at this point and forwarding tunnels can be established from the RN over the EPS bearer via DeNB and the P-GW/S-GW serving the RN. After the completion of this process, RN will start forwarding the packets to the target eNB.

The start of the handover procedure is same for alternative 2 till the point the UE sends the request for handover along with the network parameters. After that RN sends Handover Request message to the DeNB. The DeNB reads the target cell ID from the message sent by RN. After deciphering the target cell ID, it forwards the request to the target eNB. In this case the RN only needs to maintain one X2 connection with the DeNB. It can send all the handover requests to one DeNB, which will find out the target eNB from the target eNB ID.

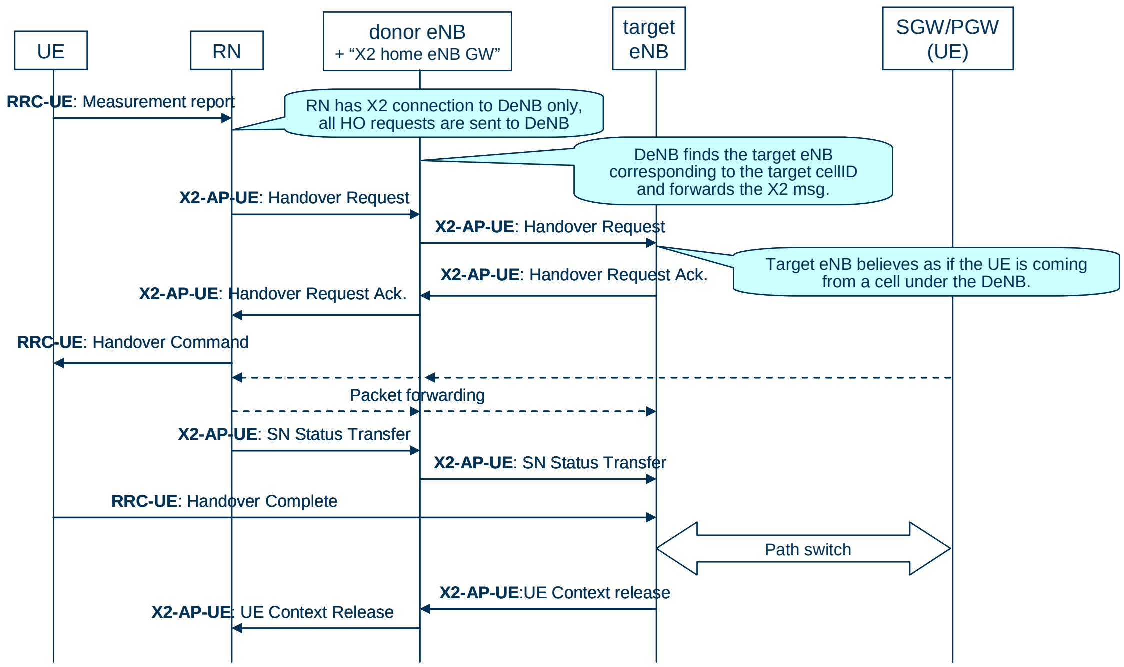


Figure 3.19 X2 handover from RN to target eNB for alternative 2 [10].

The request from DeNB looks like a request from a UE trying to make a handover from a cell within the DeNB to the target eNB. After the X2 signaling is complete, the forwarding tunnels can be established from RN to DeNB to target eNB. DeNB changed the GTP tunnel to maintain the S1-AP proxy.

3.5.3.3 RN Startup Procedure

The RN startup procedure can be divided in two parts for alternative 1 and 3. It is shown in figure 3.20. In the first step the RN connects to the network using the legacy attach procedure. The procedure used by RN is same as that used by a UE to perform the initial attach. The RN also has to go through the authentication procedure before being getting the basic connectivity to the network. When the network gets a request from RN to perform an attach, the network feels it is performing a UE attach procedure. After IP connectivity is established, the O&M authenticated the eNB and loads the configuration data on the RN. From here, RN can establish the required S1 and X2 interfaces and perform the regular functions.

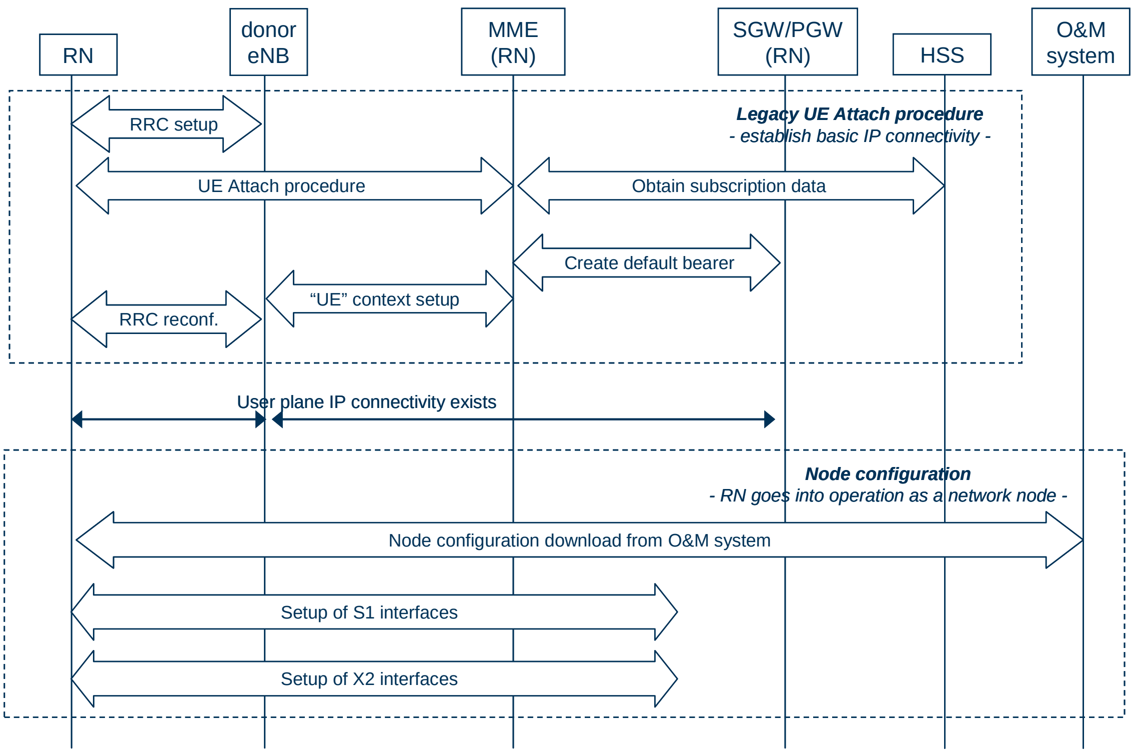


Figure 3.20 RN startup procedures for alternative 1 and 3 [10].

Alternative 2 also follows a similar RN startup procedure. The difference is, since the RN is connected to the MME through DeNB, it only needs to maintain one S1 interface and X2 interface between itself and the DeNB. This number is irrespective of the number of MMEs and eNBs that are present in the network. DeNB will terminate any S1 or X2 signaling initiated by the RN. It will use the existing S1 or X2 connections of the RN to proxy these connections.

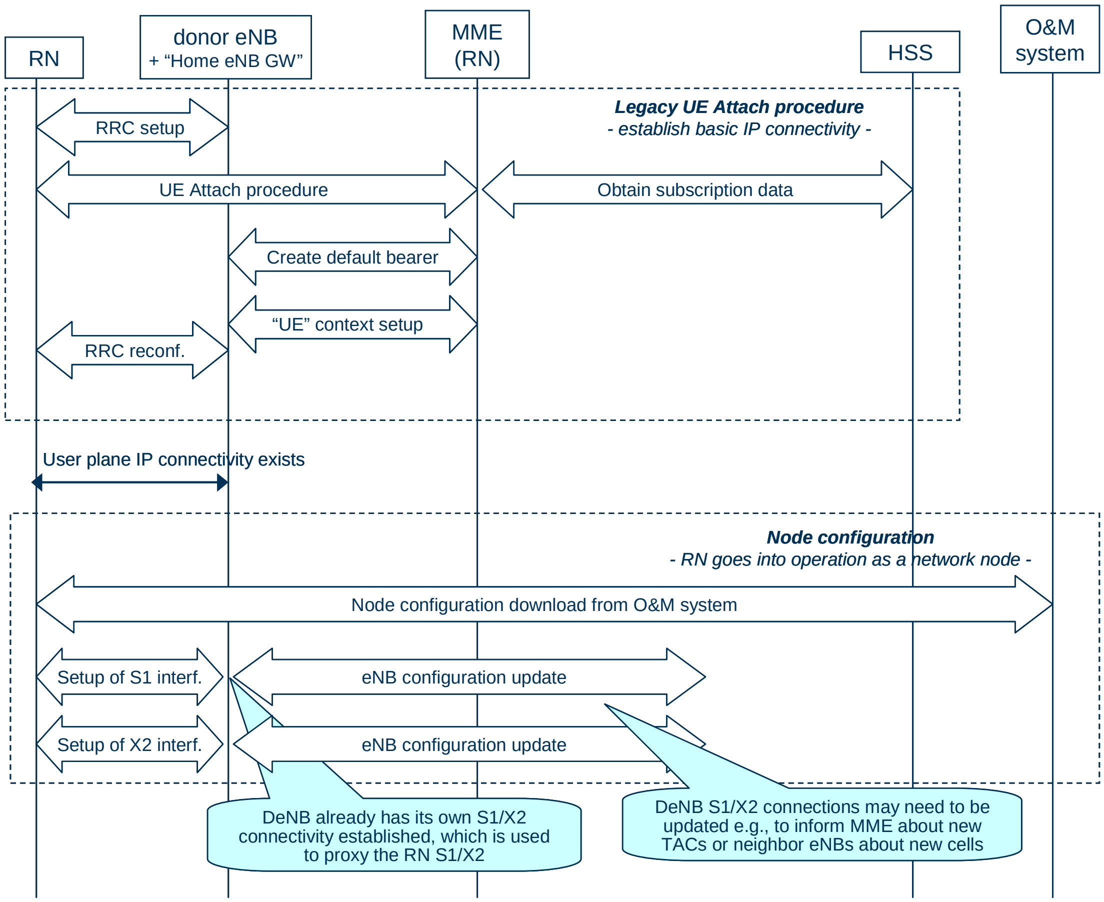


Figure 3.21 RN startup procedure for alternative 2 [10].

3.5.4 Architecture B

In this architecture the S1 connections end at DeNB going towards the EPC. For the neighboring eNB and EPC, the RN is just a cell controlled by the DeNB. DeNB acts as a S1-AP gateway [10].

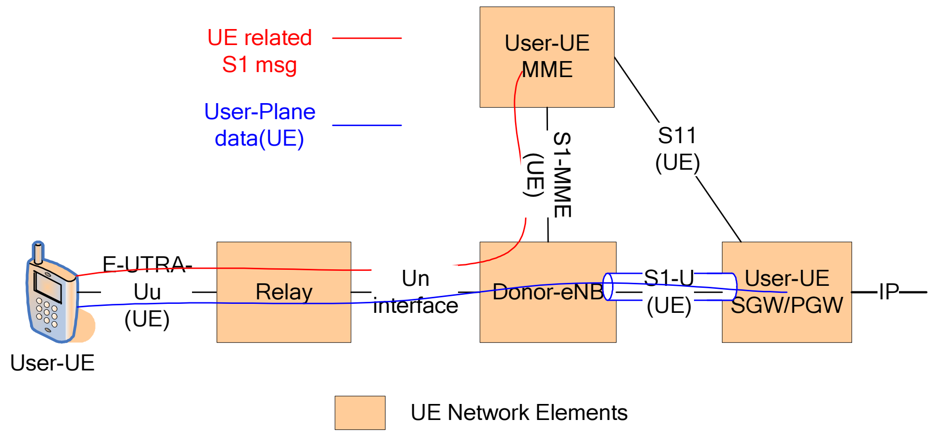


Figure 3.22 Architecture B for RN [10].

3.5.4.1 User Plane Aspects

It can be seen in figure 3.23, that the user plane architecture for alternative 4 is very similar to alternative 2. In this architecture, the DeNB is merged by the S-GW/P-GW of the RN. The S1 interface is terminated at DeNB and the P-GW/S-GW serving the UE, map the incoming packet to the GTP tunnel. The packets are sent to the IP address of the DeNB on EPS bearer of the UE.

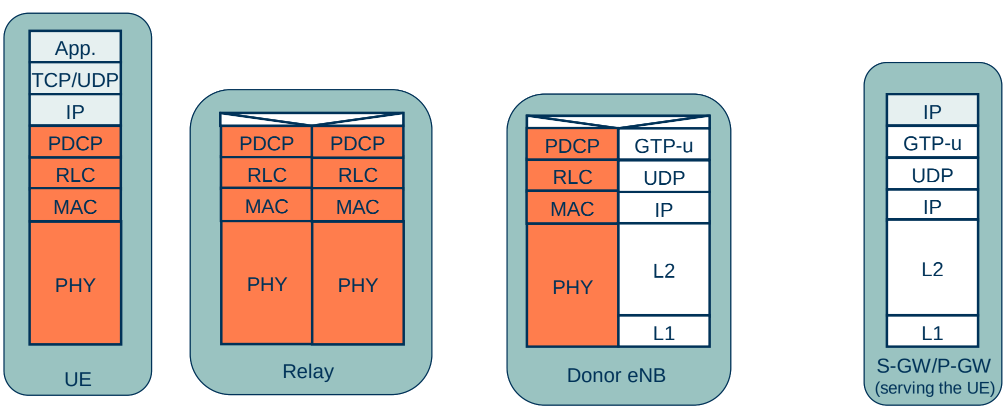


Figure 3.23 User plane stack for alternative 4 [10].

The packets received by the DeNB are stripped of the tunneling and the user IP packets are sent to the RN via the Un radio bearers. These Un bearers are selected based on the EPS bearer. Since there is one to one mapping, each UE bearer needs to be identified by a UE identifier on the Un interface. Additional support needs to be implemented in the current structure to support this. A method that can be adopted is clubbing the similar QoS packets together in one bearer.

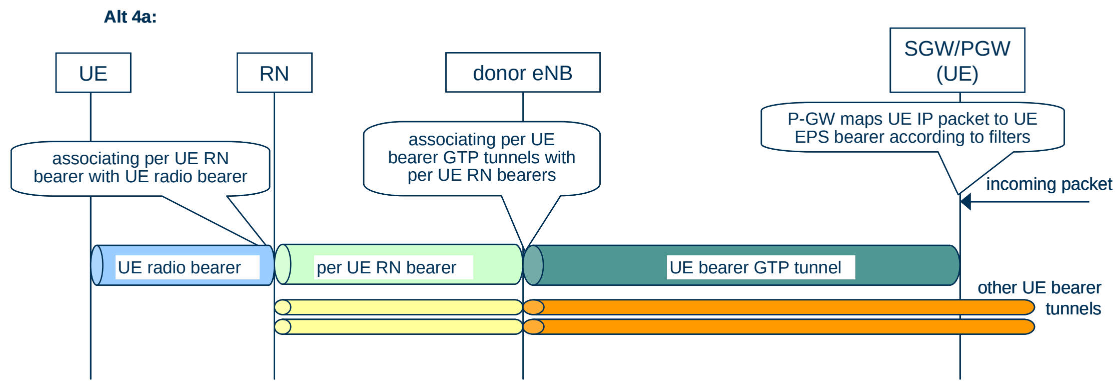


Figure 3.24 Message delivery steps for alternative 4 [10].

3.5.4.2 Control Plane Aspects

MME sends the signaling messages to the DeNB, which opens these messages and modifies the S1-AP UE IDs in the message leaving the rest of the contents untouched. This brings the S1-AP proxy functionality to this architecture. Using the proxy architecture allows the RN to be connected as a UE from the MME's point of view. RN also thinks it is directly connected to the MME. The S1 interface relations and connections are very similar to the alternative 2.

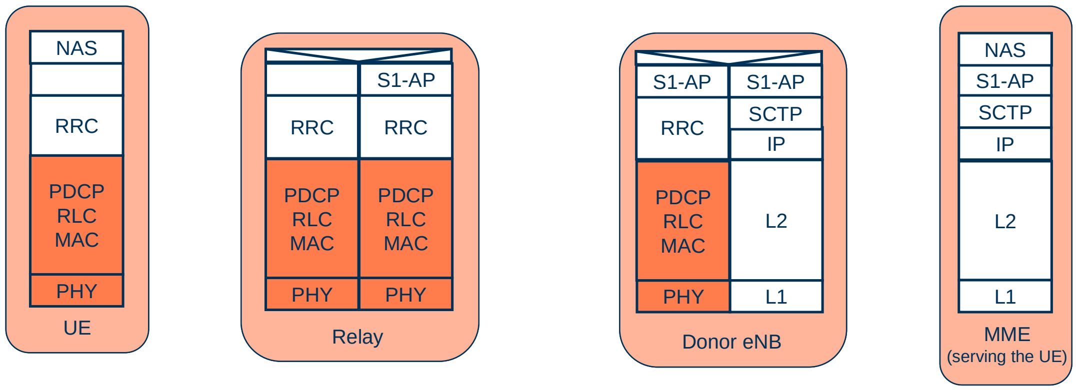


Figure 3.25 Control plane stack for alternative 4 [10].

3.5.5 Signaling for architecture B

3.5.5.1 UE Access Procedure

The UE attach procedure for alternative 4 also follows the legacy attach procedure of a UE connecting to the network. Again the DeNB is involved in the attach procedure and is used for relaying signaling S1 messages between the RN and MME.

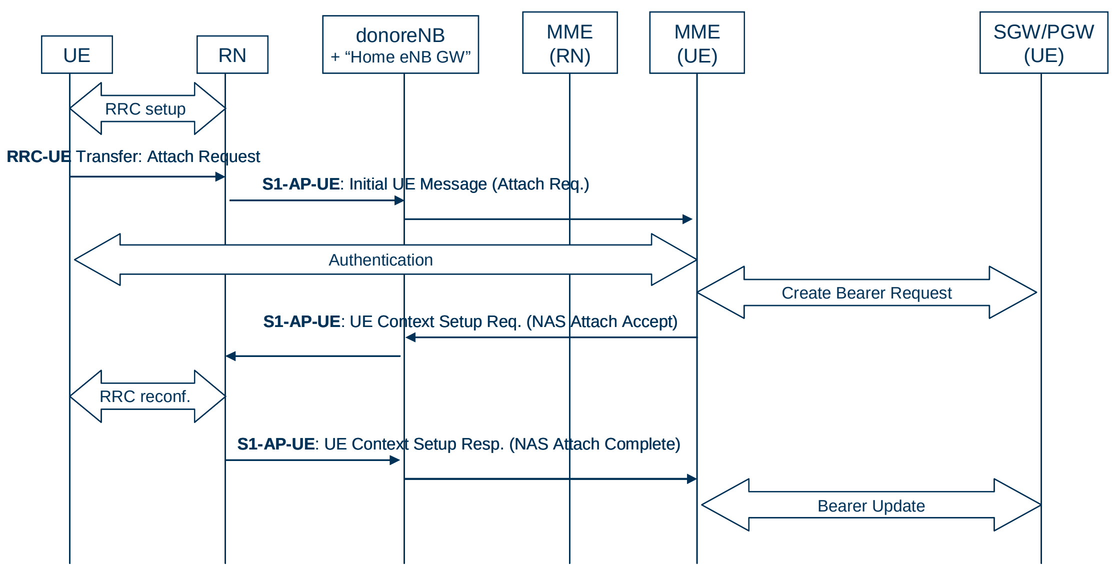


Figure 3.26 UE attach procedure for alternative 4 [10].

3.5.5.2 Handover Procedure

The start of the handover procedure is same for alternative 4 is very similar to that of alternative 2. The procedure starts with UE sending a request for handover along with the radio measurements. On receiving the Handover Request message, RN sends it to the DeNB. The DeNB reads the target cell ID from the message sent by RN. After deciphering the target cell ID, it forwards the request to the target eNB. In this case, the RN only needs to maintain one X2 connection with the DeNB. It can send all the handover requests to one DeNB, which will find out the target eNB from the target eNB ID.

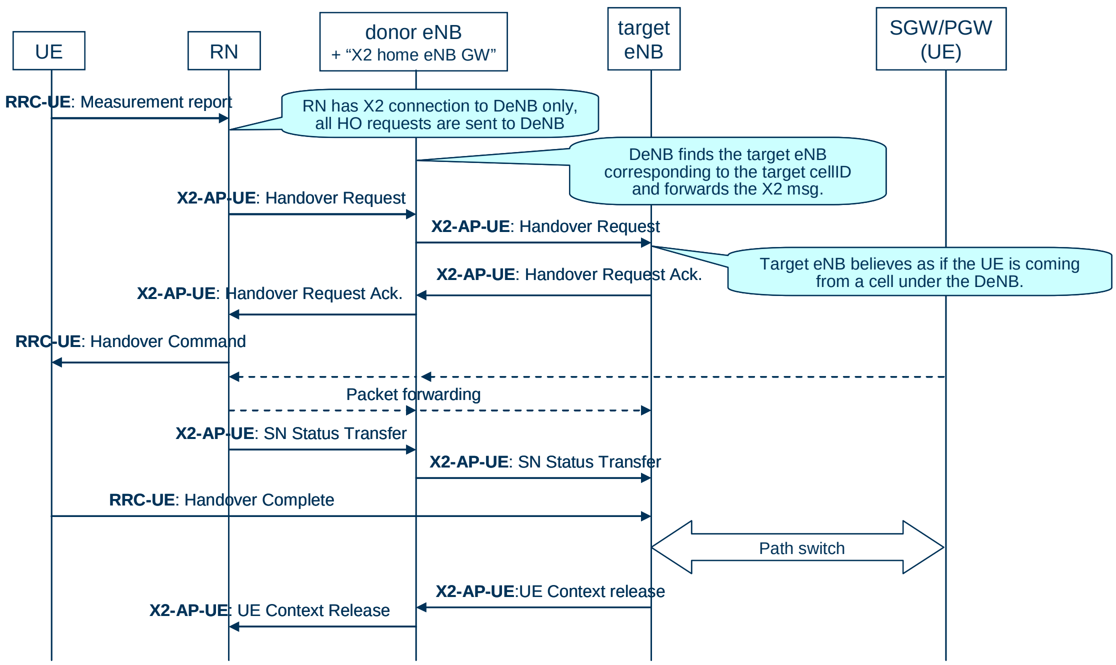


Figure 3.27 Handover procedure of RN to a target eNB for alternative 4 [10]

3.5.5.3 RN Startup Procedure

Alternative 4 follows a RN startup procedure very similar to alternative 2. Even in alternative 4, since the RN is connected to the MME through DeNB, it only needs to maintain one S1 interface and X2 interface between itself and the DeNB. This number is irrespective of the number of MMEs and eNBs that are present in the network. DeNB will terminate any S1 or

X2 signaling initiated by the RN. It will use the existing S1 or X2 connections of the RN to proxy these connections.

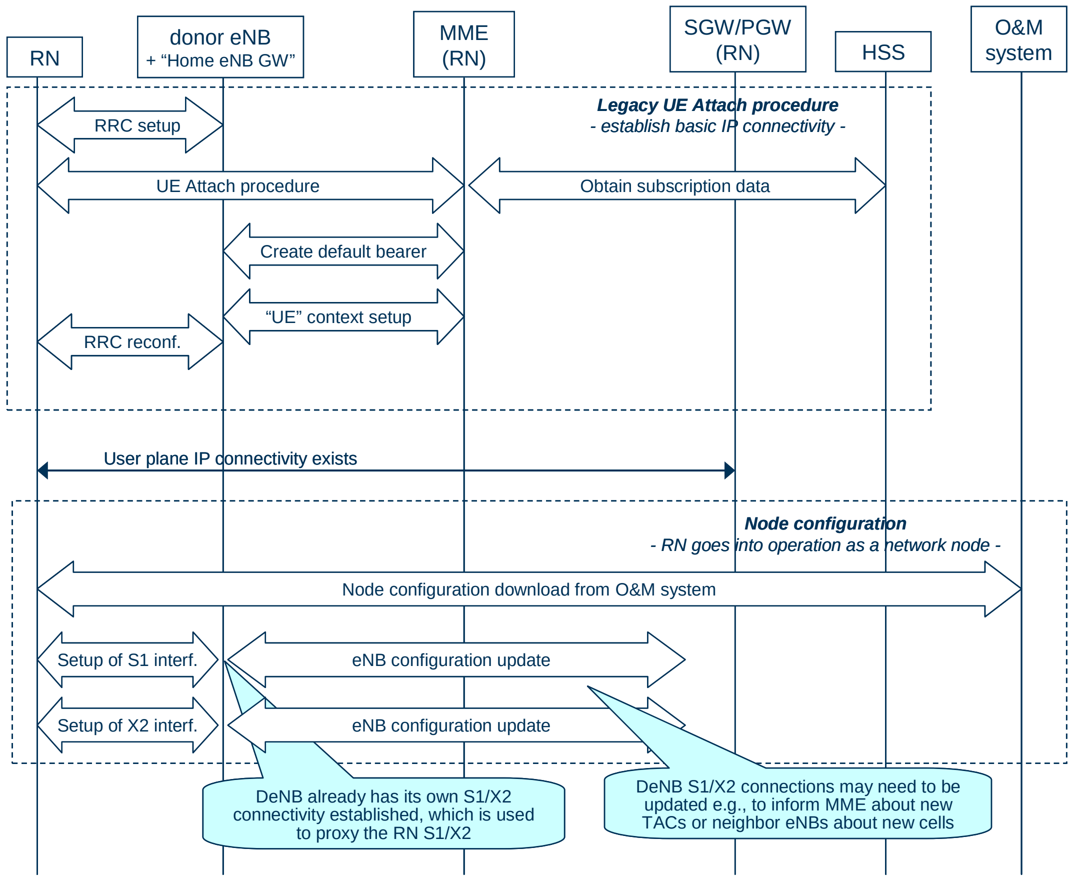


Figure 3.28 RN startup procedure for alternative 4 [10].

3.6 Multi-hop Architecture

The multi-hop relay architecture is based on the principle of using one or more RN in conjunction with the eNB and the UE. A connection is said to be established using a RN when there is one or more RN between a UE and the eNB. A single eNB can serve multiple RN and in turn even larger number of UEs. This connection of a UE via a RN is called as a multi-hop chain. In a multi-hop chain there will be n-1 RN, where n is the number of hops (data from UE to RN and RN to eNB is considered as 2 hops). The number of hops that can be allowed will affect the capacity of the network adversely but help in increasing the coverage. So this number has to be carefully selected so that using relays to improve the coverage does not become too expensive but at the same time, the areas that are covered by the RN do not have degraded network capacity. 3GPP has limited the value of n as 2.

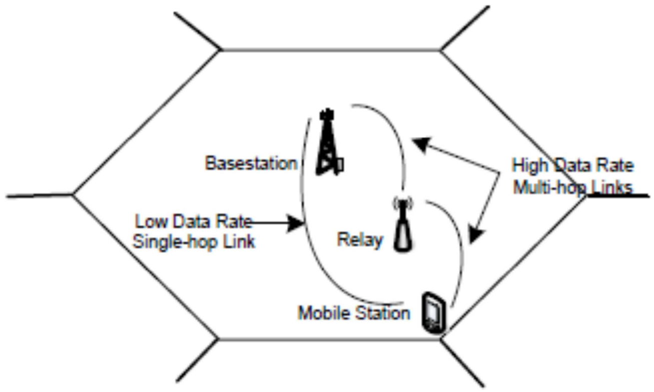


Figure 3.29 Relay node used to improve capacity in a LTE network [11].

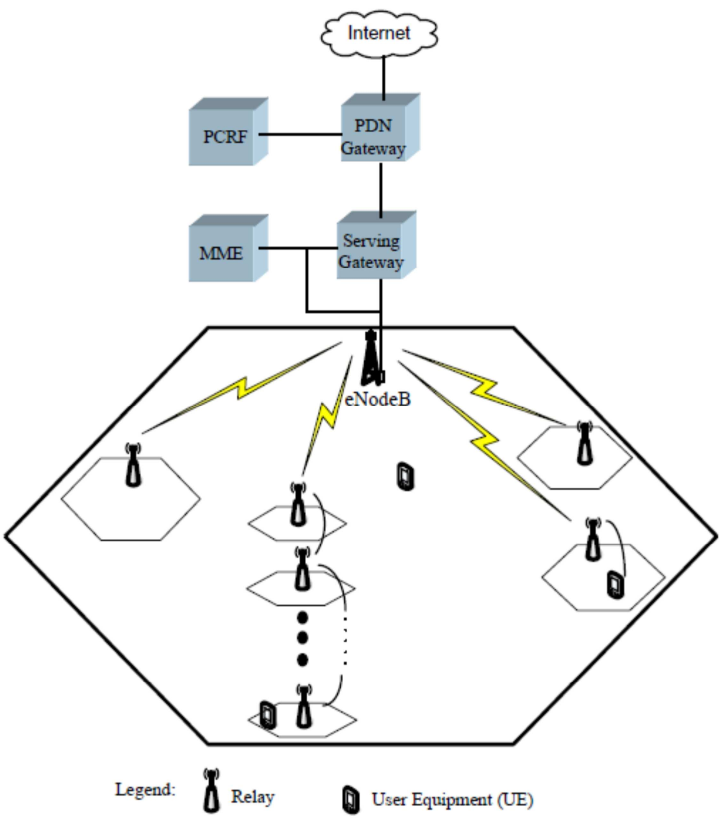


Figure 3.30 Placement of RN in a LTE network [11].

Figure 3.29 shows a UE connected to an eNB via RN. Since the RN is closer to the UE as compared to the eNB, the amount of air loses will be significantly less. This would improve the overall capacity of the network. Other entities of the LTE network architecture have been explained in details but the RN needs further explanation. Figure 3.30 shows where the RN acts in the LTE network.

3.7 Relay Node Theory

RN for LTE and WiMAX are used as repeater/bridges and not just amplifiers. As repeaters/bridges they can change parameters which are best for the environment in real time. This even involves changing modulation for best performance and lower losses.

When looking from a UE point of view, the RN is just like any BS. It has no knowledge about a hop between itself and the real BS because the RN can provide almost all the features a BS can.

Though the resources for connection between RN and UE are provided by the BS, RN is free to choose any channel for linkage. The BS cannot and should not decide for the RN as the environment between BS and RN maybe different from the environment between RN and UE.

3.8 Relay Strategies

3.8.1 In band relay

The link between the eNB and the RN uses the same frequency as used by the RN and UE i.e. the frequency used in the first hop and the second hop is the same. This relay strategy uses time domain to separate the two signals.

3.8.2 Out band relay

The frequencies used in both the hops are different. Both the transmissions can occur at the same time. Hence, an out band relay system requires more frequency resources as compared to the in band relay system.

3.8.3 *Amplify and Forward (AF)*

This is also called as Layer 1 (L1) relay. This is a very simple relay which receives the signal, amplifies it, reshapes it and forwards it to the next hop. The transceivers at both the points can operate on either, in band configuration or out band configuration. If the out band configuration is used, then a frequency reuse factor of 1 will not be possible. Additional resources will have to be assigned by the eNodeB. The resources will be different from each other for eNB to RN and RN to UE. This relay system can be used to increase the coverage of a network. The main disadvantage with L1 relay is that it amplifies signal as well as noise if noise is present in the system. Thus, AF relay provides best performance in environment without noise.

3.8.4 *Decode and Forward (DF)*

DF relay requires the signal to be decoded at RN. The decoded signal is forwarded to the UE after desired parameter changes are implemented making it the best choice to be implemented in noisy situations. The drawback of using DF is that it introduces delay due to decoding and encoding of the data at RN. DF can be further used at Layer 2 (L2) or Layer 3 (L3) with the same combination as L1 of in band or out band.

- Layer 3 (L3) relay configuration: A L3 configuration was proposed by Hoymann in [12] called the Self-backhauling relay. In a self-backhauling relay the relay acts like a base station except for lower transmit power and smaller cell radius. In self-backhauling relay the RN is connected to the BS using a LTE radio interface instead of wires in a traditional base station. This link between the eNB and the RN is called backhaul.

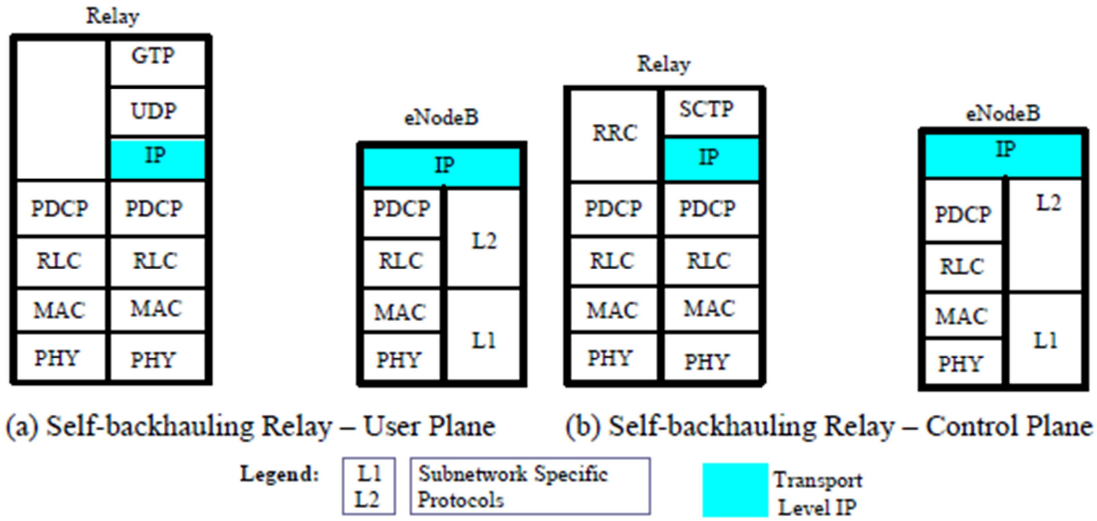


Figure 3.31 Self-backhauling relay user plane and control plane [11].

Different technologies can also be used to make this backhaul connection but it is preferred that the backhaul should be using the same technology. Figure 3.31 shows the user plane and control plane for this configuration.

In the user plane eNB acts as a router between serving gateway and RN. eNB has to differentiate between packets that are going to be sent directly to UE and those going via RN. The packets going through RN will bypass GTP and UDP; these steps will be taken at RN. Thus the traditional method of IP packet classification based on GTP tunnel end-points cannot be implemented and a new QoS mechanism has to be developed.

Although the self-backhauling relay solution looks attractive, it cannot be implemented as LTE radio protocol stack is designed for single hop communication.

- Layer 2 (L2) relay configuration: L2 protocol consists of 3 sub layers, namely Radio Link Control (RLC), Media Access Control (MAC) and Packet Data Convergence Protocol (PDCP). L2 protocols can be used as edge to edge or per hop basis. If we use them in per hop basis, they work similar to self-backhauling configuration. The difference is that L2 relays forward PDCP packets instead of IP packets. This enables IP packet analyzing by GTP at eNB. Due to this we can use GTP tunnel end-points for IP packet classification.

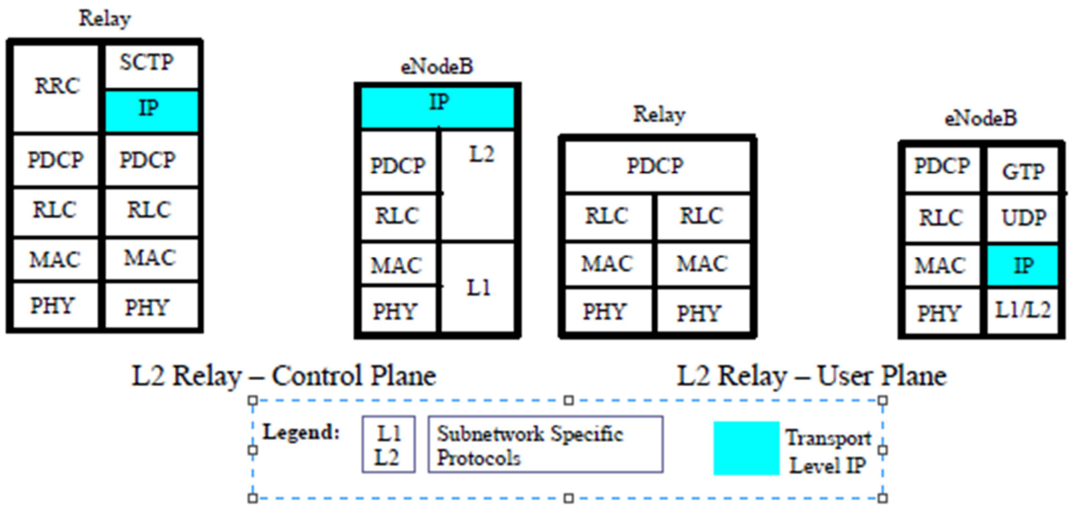


Figure 3.32 L2 user plane and control plane – per hop basis [11].

3.9 Types of Connections Allowed

In current wireless systems, the devices closer to the BS would receiver better speeds compared to the devices at cell edge. This happens due to path loss and power issues. To overcome this problem, we introduce relays. The types of connections possible with relays in a cell are as follows:

- Connection between UE and BS
- Connection between BS and RN
- Connection between UE and RN

Connection between UE and the BS remains similar to the current systems which is a wireless connection using OFDM. Since the connection is wireless it allows the network to be more flexible than the current networks. Implementing such a connection does not require a lot of planning and this network can be changed depending on traffic needs. Even though the final destination is the UE, the BS does not need to know about the final destination. For the BS, the RN is the final destination and it should serve the RN as it would serve a UE.

Connection between UE and RN is similar to the connection between RN and BS. Channels for this connection are provided by the BS but the choice of the channel is of RN. The RN can even modify the modulation technique between itself and the UE. The modulation used here can be different from the modulation used between the BS and the RN.

The UE will be given a choice to connect directly to the BS or via a RN. This idea allows higher speeds to be delivered to the UE that are close to cell edge than the current configuration. Although devices connected directly to the BS will experience slower speeds than the conventional setup, devices which are closer to the cell edge will experience a noticeable increase in the data rate. The decrease in speed will be due to the resource allotment to the RNs while the increase is due to lower pass loss gains due to the RNs.

CHAPTER 4

MULTIMEDIA BROADCAST MULTICAST SERVICE IN LTE NETWORK

4.1 Introduction

Multimedia Broadcast Multicast Service (MBMS) is an old broadcasting service. It is a point to multipoint data transmission technology, which means it was designed to use one source to serve multiple slaves (devices). One of the benefits of using this service over a conventional broadcasting service is the ability to get user interaction. The interaction can be limited of selecting the content to be played in a multicast mode, but can be expanded to get user feedback like rating for a movie. Using MBMS, the user is able to interact with the service.

Some of the major applications of this service are related to broadcasting/multicasting of videos. Live TV on mobile has already been implemented using the current technology on GERAN and UMTS networks. With the improved architecture of the LTE networks and better speed and bandwidth it was decided to bring MBMS to LTE networks as well.

4.2 MBMS Modes

4.2.1 MBMS broadcast mode

Broadcast mode is sending the data from a single source to multiple devices (point-to-multipoint) of multimedia data which includes but is not limited to text, audio, pictures and video. This data is sent from one source to all the devices subscribing to that source within a broadcast area. This technique saves the resources of the network by transmitting the data only once and serving many with that information. Further efficiency can be achieved by modifying the MBMS architecture to adapt its data to different scenarios experienced in a LTE network. MBMS service can adapt to these changing scenarios by modifying/reducing its bit rate, allowing more subscriber to subscribe at the same time.

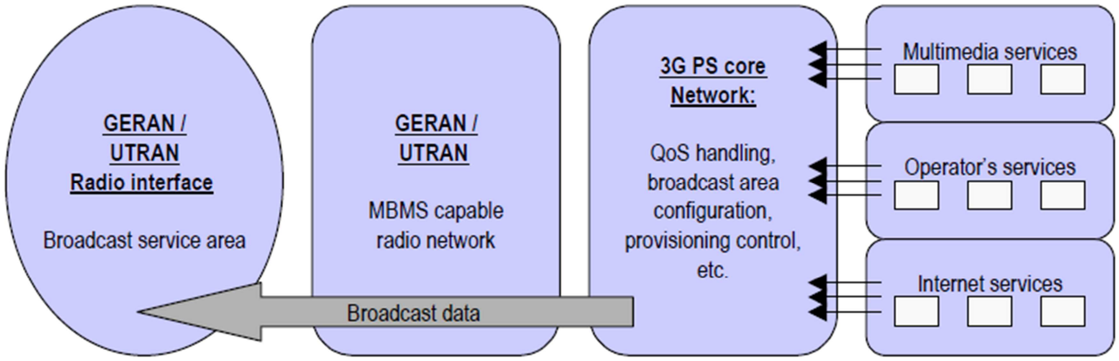


Figure 4.1 A sample broadcasting service network [13].

This broadcast mode should not be confused with the Cell Broadcast Service (CBS) which is used to transmit very low bit rate services like messages. This broadcast service, although can be used to transmit messages, is used to transmit higher bit rate multimedia services like audio, video and pictures. A sample use of this service can be to send advertisements to the user when entering a new service zone or sending a welcome message. Since this is broadcasting service, there is no activation required by the user side. If default settings are left on the device, the user will automatically receive such messages when the case for the messages is triggered. This is one of the differences between broadcast mode and the multicast mode. To consume power and for other reasons, the user should have the capability to turn such a service off.

4.2.2 MBMS multicast mode

This mode is very similar to the broadcast mode described in 4.2.1. This mode also allows a single source to distribute data to all the devices connected to it within a specified multicast group. This mode is also designed to save the network resources by transmitting data on a common channel. However, in this mode the source has a choice of sending data to only a selective part of the network within the multicast service area.

In multicast mode the services are not automatically activated and require users to subscribe to a particular set of services for the network to start sending the data. This is completely different from the broadcasting mode in which the users are sent the data unless they manually turn the service off. Since this is a service that the user has to request, this service maybe subject to charges.

This multicast mode should not be confused with the multicast offered in the IP networks, although there needs to be interoperability between the two to maximize the user experience with both services. This will also allow the best use of IP services platform to make the content and applications available to users and will allow services to be delivered in a more efficient way.

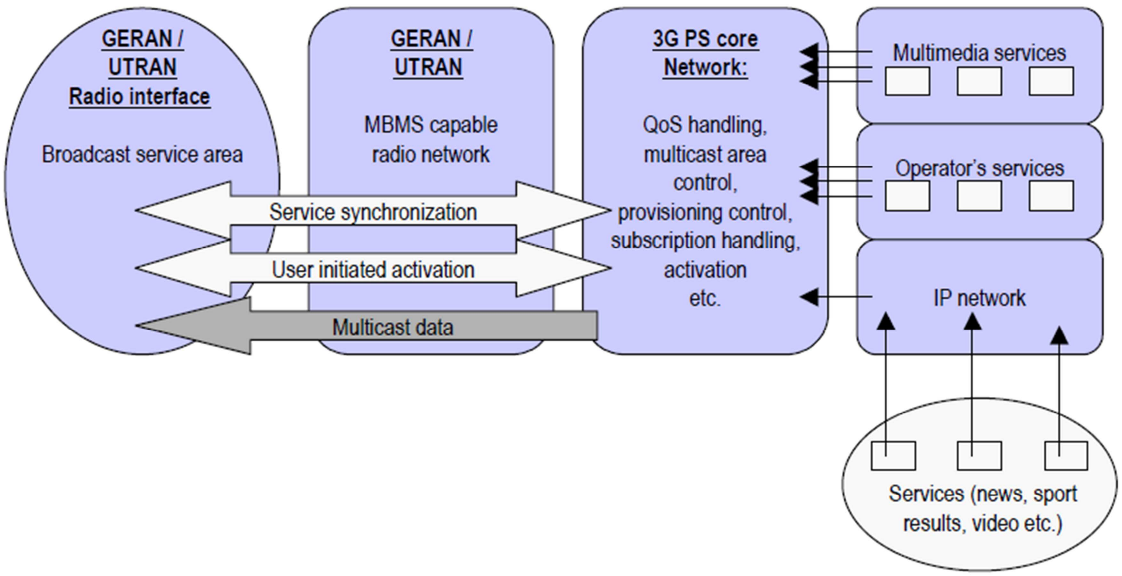


Figure 4.2 Sample multicast mode [13].

4.2.2.1 Subscription And Reception In Multicast Mode

Since a user is not automatically subscribed to a multicast service, the subscription process is explained ahead. The user may be subscribed to a multicast group which will be identified by a unique number. After a successful subscription request, the user will be part of that multicast group. The type of subscription can be of various types and will be defined by the network operator. The user may be subscribed to the service for as long as the user wants, subscription may also be based on time which can vary in hours to days to years. The subscription may also be only a onetime thing based on the media and subscription profile.

All the multicast services that are available in the user's location will have to broadcast their information including content and availability. The user, depending on its interest, can select any multicast group. The multicast group should provide access to the user as soon as possible after the request for service was received by the multicast group. The user can be subscribed automatically depending on the subscription of the user with the Home Environment. Home Environment has the capability to add the user to any multicast group that the user approved to be connected to once it was available in the user's area.

When the transmission is available for telecast, the multicast group should start telecasting immediately after the user has joined its group. If the transmission is supposed to start at a later stage, the transmission will begin at the regular time. The network can also skip the process of making a multicast transmission if the number of users do not justify saving of resources with a multicast transmission. Once the stream is available for the multicast, the user will be able to see it on its device. The user has the option to terminate the stream at any point and also to join in at any time. It can also decide to continue receiving the data from the multicast group or leave the group completely.

If the user decides to unsubscribe to the multicast group completely, the user will not receive any further service notification from the multicast group. The Home Environment should also be able to remove the user from a multicast group depending on the user's subscription profile.

4.3 MBMS Basic Requirements

4.3.1 MBMS broadcasting and multicast requirements

The operators who share a network should be able to provide one or more broadcasting services for their own subscribers and all partner roaming devices. Each broadcast service will have a pre-configured broadcast area. This area will be specific to the broadcast service depending on the content being broadcast. The broadcast area can overlap with other broadcast areas.

The broadcasting services should be able to serve the local subscribers with information like weather, traffic and local news. This is possible if the content sent by the broadcasting service is independent of the other broadcasting services. This allows location specific information to be sent to users in a particular geographic area. The users in one particular geographical area should not be bogged with the information. The users should not have to select their area and also should be sent information of other broadcasting stations. It is also possible that the users from operator A receive different broadcasting information as compared to subscribers of operator B even though both the users and the operators are in the same geographical area.

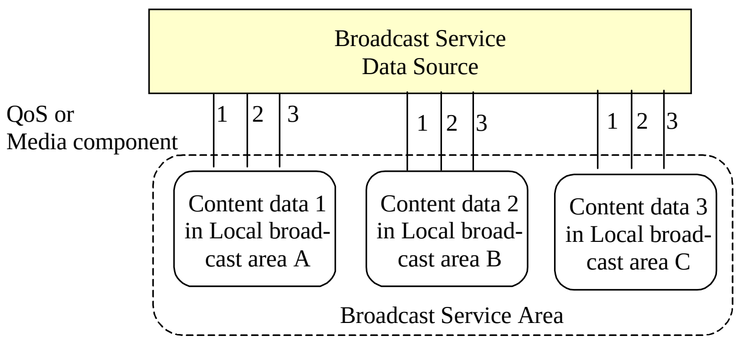


Figure 4.3 Localized and individual content broadcasting.

When the user is roaming on a partner network the user will be able to request for content on the local network. It can subscribe to this local broadcasting/multicasting service as permitted by the subscription profile of the user and its privileges in the roaming network. If the user is subscribed to location specific services like weather and traffic, the same class of service should be provided to the user when it is on the visiting network.

The user, depending on the capabilities of the UE, will be able to save the content which it received via broadcast or multicast. The user will also be able to share this information with other users on different networks via traditional services like MMS, email, voice calls etc.

While streaming television services the user will be able to switch between different contents within the television service. Switching the content from one to another should not take more than a second to switch. This time can be further reduced with the help of better codecs.

4.3.2 User requirements

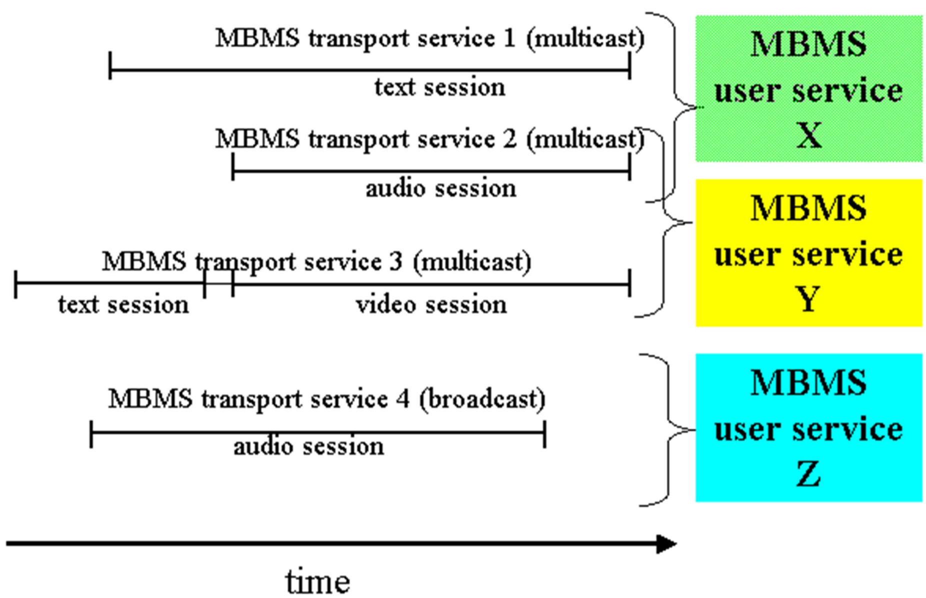


Figure 4.4 One user using multiple MBMS services simultaneously [14].

The user should be able to move freely and transparently while enjoying the MBMS service. At the same time, the user should be able to receive the same broadcasting/multicasting information on the new cell as it was receiving in the previous cell. Thus, the handover during a MBMS service should be seamless and with minimum delay or jitter. At the same time the user should be able to refuse the service in the new cell, if one exists specific to that cell. Depending on the device user by the user, the user will be able to get multiple subscriptions parallel or one after the other.

4.4 MBMS Architecture for EPS

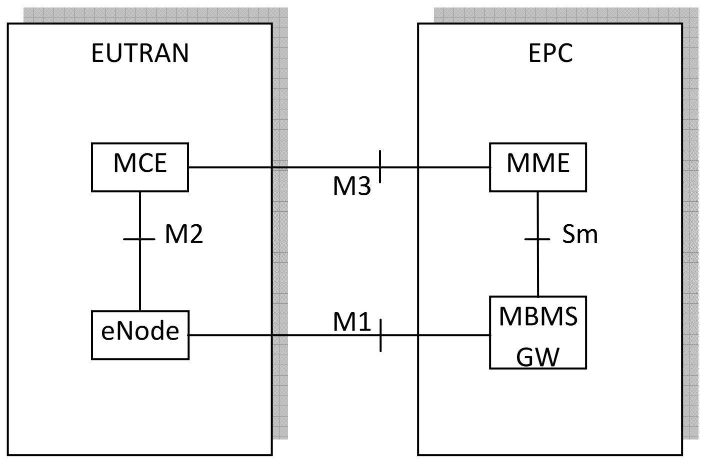


Figure 4.5 The general architecture for MBMS services in E-UTRAN.

4.4.1 MBMS Gateway (MBMS GW)

MBMS GW is the entry point for all the multicast as well as broadcast information. It is present in the EPC and may be co-located with other entities likes Broadcast Multicast Service Center (BM-SC). In a given network one or more of MBMS GW may be used. Each MBMS GW is connected to an eNB via the M1 interface and to MCE via the M3 interface. One MBMS GW can be used to serve multiple eNBs.

A MBMS GW is used in the network to provide control plane and user plane interfaces. A control plane interface may be provided to an entity using MBMS bearer through the SGmb reference point while a user plane interface can be provided using the SGi-mb reference point. It is also responsible for IP multicast distribution of MBMS user plane data to eNB and RNCs through the M1 interface. This is done by providing an IP multicast address to the eNB and the RNCs. These entities can connect to the assigned IP multicast address to join and receive the required data. The IP multicast address is provided for eNB by MME and for RNC by SGSN. MBMS GW also has the capability to communicate with other control plane entities like MME, SGSN and BM-SC.

4.4.2 Broadcast-Multimedia Service Center (BM-SC)

BM-SC performs many user service specific services which include delivery and provisioning. It can also serve as an entry point for MBMS content that is being transmitted by a certain provider. Within a network the role of BM-SC can also be to schedule certain services and deliver the MBMS transmissions. It is also used for authorization and initiation of MBMS bearer services. BM-SC is consists of 5 sub functions:

- Membership function
- Session and transmission function
- Proxy and transport function
- Service announcement function
- Security function

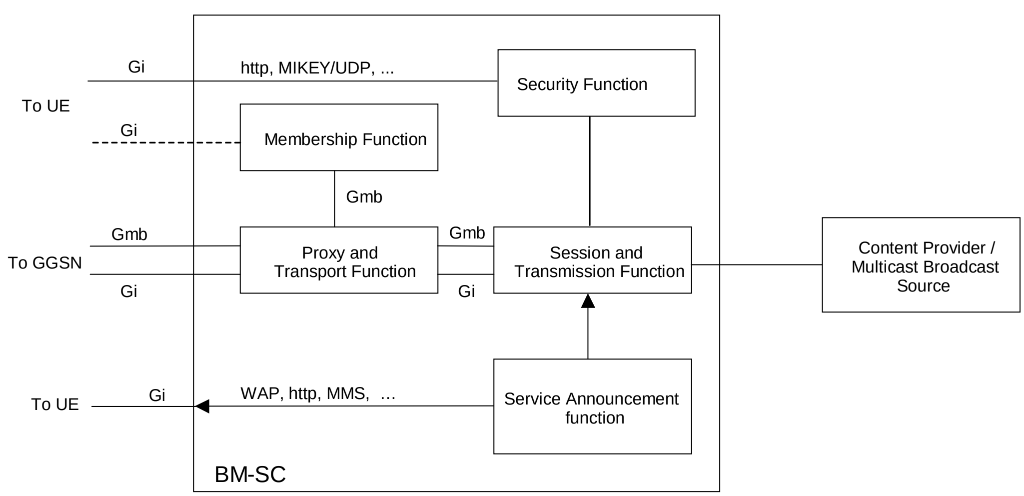


Figure 4.6 Functional structure for BM-SC [15].

4.4.2.1 Membership Function

Membership function authorizes and activates any UE which seeks to access the MBMS services. It has the subscription data of the UE and can specify which services are allowed to be streamed to the UE. Along with keeping a track of all the profiles of the UE, when the service starts rolling, it also keeps a record of the charges incurred in streaming of the service and are later charged to the user.

4.4.2.2 Session and Transmission Function

This function is responsible for establishing MBMS sessions for transmissions. It can also schedule retransmission of the services in case the UE is not reachable or for any other reason the message was not delivered successfully to the UE. To distinguish the retransmitted session from the original transmission, this function labels the retransmitted session. This label is called as the MBMS Session Identifier is 2-3 octets long. This label is passed at the application layer embedded within the content. A shortened version of the label, containing only the MSB of the octet, can also be sent during the session start request message. It also assigns a Temporary Mobile Group Identity (TMGI) to the multicast session.

The BM-SC session and transmission function will provide the GGSN with any transportation parameters like QoS and service area. It should also be able to initiate and terminate the MBMS bearer resource before and after the transmission of the MBMS data.

4.4.2.3 Proxy and Transport Function

Proxy and transportation function is a responsible for signaling over Gmb reference points between the GGSN other BM-SC sub-functions [16]. It should also be able to handle different MBMS service provision by multiple physical network elements. The signaling and the routing of these signals should be transparent to the GGSN. Another role of this function is to generate charging records for the content providers. At the start of the session, the name of the content provider is provided to the proxy and transport function over Gmb.

4.4.2.4 Service Announcement Function

As a service announcement function, its role is to announce all the multicast and broadcasting services within the network to the respective UEs. In these service announcements, it describes the type of media (audio and video codecs) to be used, the time and duration of the media. It will also include multicast service identifications to identify each service from one another.

4.4.2.5 Security Function

The security function is provided for maintaining the integrity and confidentiality of the user. The security is provided by distributing MBMS keys to authorize the UE and the services requested by it.

4.4.3 Multicell/Multicast Coordination Entity (MCE)

MCE performs MBMS content and resource management. Coordination of multiple cells is very difficult with the flat architecture of LTE. MCE plays a vital role in coordination for multicell transmissions. It also reserves time and frequency radio resources which are used by eNBs in the MBSFN area. MCE has a hand in deciding the coding and modulation schemes for the best possible experience by a user. In GERAN and other previous technology, scheduling and other configuration roles were handled by the eNB. These have been moved to a central location which is MCE.

4.4.4 M1 interface (MBMS GW - eNB)

M1 is an interface which is present between MBMS GW and eNB as a user interface. M1 interface is used to deliver all the IP multicast downlink packets. However, it cannot be used to uplink any user packets to the radio network layer. It manages the IP multicast groups. MME is responsible for assigning IP addresses to IP multicast groups. These IP multicast addresses are assigned at the start of the session and are released when the session is terminated. MBMS GW uses M1 interface to advertise these addresses to the eNB. The eNB joins the required session to receive the user plane data at the start of the session and terminates this connection when the session is over.

The messages transmitted over M1 are simultaneously transmitted by the eNB. To synchronize the content with the data transmission of the other eNBs SYNC protocol is used over M1 interface. M1 allows additional data to be transmitted which helps eNBs to identify the timing for radio frame transmission and detect packet loss [17]. All the packets travelling over the M1 interface contain the SYNC protocol information.

4.4.5 M2 interface (MCE - eNB)

This is a purely control plane interface between the MCE and eNB. MCE can connect to multiple eNB and these connections are on M2 interface. This interface is used to convey session control information to the eNB from MCE. The information may also contain radio configuration data for multicell Multimedia Broadcast Single Frequency Network (MBSFN) transmission. By passing this additional information on the M2 interface, the network makes sure that the eNB has the required information for allowing synchronized data to be transmitted to the UEs. This information also allows the RLC/MAC to be configured appropriately. Streaming Control Transmission Protocol (SCTP) is used over M2.

M2 interface is responsible for service context management function which involves establishment of the overall initial service. This will include, MBMS E-RAB context, M2 signaling connection ID etc. It is also responsible for releasing the context of the previous services which may be present in the eNB after the termination of the service. MBMS E-RAB is the service management function which handles the establishment and release of the resources by E-UTRAN for data transport. This is triggered by MCE and requires specific QoS information from eNB.

4.4.6 M3 interface (MME - MCE)

Again, a control plane interface used to provide signaling between the MME and MCE. This plays a vital role in fulfilling the M2 Application Protocol (M2AP) functions. It also sends signals for starting and stopping of service. M3 is part of the service context management function and supports the establishment of necessary E-MBMS service contexts like E-MBMS E-RAB context, M3 Signaling connection ID etc. The service context establishment request is given by MME. Like M2 service context management, M3 service context management also releases any resources in the MCE for previous transmission.

4.5 Channels for MBMS

In Rel 6, three channels were defined for MBMS to allow it to perform the necessary functions.

4.5.1 MBMS traffic channel (MTCH)

This channel was designed to carry application level data for point-to-multipoint transmission [18]. Each MBMS service will be assigned one MTCH.

4.5.2 MBMS control channel (MCCH)

MCCH is a logical channel used to send signaling messages for MTCH reception. One MCCH is used in every MBMS enabled cell. Each MCCH can be used to convey the signaling for multiple MTCH. MCCH information is transmitted using variable number of consecutive transmission time intervals. Figure 4.7 shows MCCH transmission schedule.

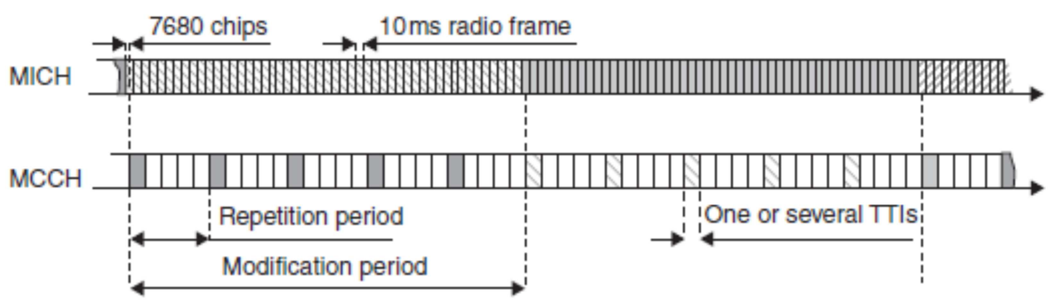


Figure 4.7 MCCH transmission schedule [18].

In the modification period, the critical information does not change and is repeatedly transmitted [18]. This ensures mobility in the cell. If a UE moves to another cell and misses the first transmission, it will not have to wait for the modification to receive the MCCH information.

4.5.3 MBMS scheduling channel (MSCH)

The purpose of MSCH is to enable the UEs to perform discontinuous reception of the MTCH. MSCH contains data which inform the UE in which TTI will a specific service be transmitted [3g evolution]. This allows the UE to conserve battery.

4.6 Single Frequency Network for MBMS in LTE

Since LTE uses OFDM radio interface, it can transmit multicast and broadcast data as a multicell transmission over a synchronized single frequency network known as Multimedia Broadcast Single Frequency Network (MBSFN) [19]. The basic principle behind MBSFN is to transmit MBMS data simultaneously from multiple cells. These cells are tightly time synchronized. From the UE point of view, the data transmitted by multiple cells will behave as multiple versions of the same data with different delay. If the synchronization is kept tight and the information reaches the UE within the cyclic prefix at the start of the message, then there will be no Inter Symbol Interference (ISI). To the UE, this type of transmission will be similar to a transmission from a very large cell having multipath components. The UE does not even need to know about the multiple cells or the number of cells used for transmission of the MBMS data.

4.6.1 MBSFN synchronization area

A geographical area where all eNBs can be used to transmit MBMS data synchronously is called as MBFSN synchronization area. This area is part of the MBMS service area, an area where MBMS services can be provided to the user. Within the MBSFN synchronization area, there is a MBSFN area. This area comprises of a group of cells that are coordinated for a MBMS transmission.

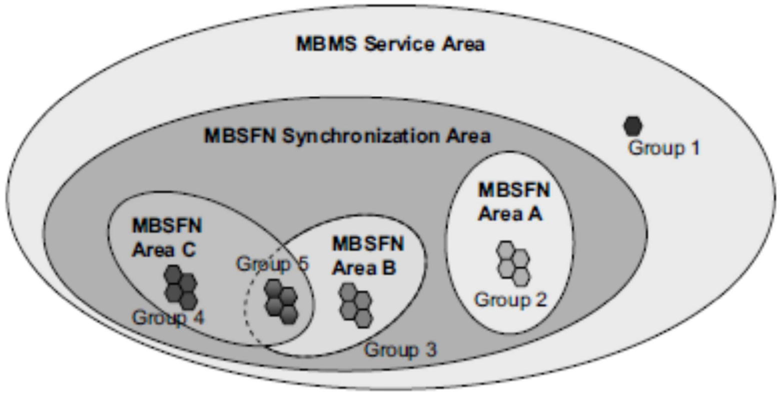


Figure 4.8 Demonstration of a sample MBMS service area scenario [19].

Figure 4.8 shows a sample MBMS service area where MBSFN is a part of the service area. As shown, there can be multiple MBSFN areas within a MBSFN synchronization area. These MBSFN areas can overlap with other areas but still have different transmitted content. In the above figure, group 1 is a single cell transmitting MBMS information to the users. This cell is not a part of the MBSFN synchronization area and cannot perform as a MBSFN node. Although there are many advantages of using MBSFN mode of transmission, there will be provision of single cell transmission in LTE. This is to allow normal transmission of MBMS services when tight synchronization cannot be achieved. The MCE decides at the time of setup of MBMS service whether it is going to be over single cell or MBSFN transmission.

Group 2,3 and 4 belong to the MBSFN synchronization area and are capable of performing MBSFN transmissions. They will all transmit in their respective areas A, B and C. Group 5 is overlapping area A and area B. Till the current release of the LTE specification by 3GPP, it is not known how the resources would be handled in a case of overlapping MBSFN area. MBMS content will be broadcast according to the geographical area it is servicing. The resources of group 5 will not be available for transmission of content specific to area which is not overlapped.

4.6.2 MBSFN user data flow synchronization

To synchronize the user data sent from different eNBs require a SYNC protocol which operates on the M1 interface between BM-SC and eNB. BM-SC includes a time stamp within the SYNC PDU packet as a part of the SYNC protocol procedure. This time stamp is based on a common time reference and start of the first synchronization period available for the BM-SC and the corresponding eNB [20]. An eNB will send the user data packets on the air interface based on this time stamp.

BM-SC sets the time stamp for all the SYNC packets. The time stamp is same for all SYNC packets in one synchronization sequence of a particular MBMS service. This time stamp is dependent on the maximum delay in transmission from BM-SC to the furthest eNB, arrival time of data packet, length of synchronization sequence used and other delays like processing time taken by eNB. BM-SC does not know when the Transmission Time Interval (TTI) will start, but it knows the length of the sequence, a delay point is specified by BM-SC. If the user data starts in between the delay point, it will be given the time stamp value of first data packet.

Based on SYNC parameters, the eNB is able to find the downlink timing and also check if any SYNC packets were lost in transmission from BM-SC to the eNB. eNB can find the size of the SYNC packets lost if the packet loss occurs with consecutive packets or if only one SYNC packet is lost. eNB can then ask for retransmission before sending the packets to Radio Link Control (RLC) for processing.

BM-SC sends a user data frame to the eNB after each synchronization sequence which contains the counter information like ‘total number of packet counter’ and ‘total number of octet’. This information is sent without any MBMS payload and signifies the end of synchronization sequence. These messages can be sent multiple times to increase reliability of the service.

4.7 MBMS Procedures

4.7.1 MBMS session start procedure

MBMS session start procedure comes into play once the content provider is ready to start streaming.

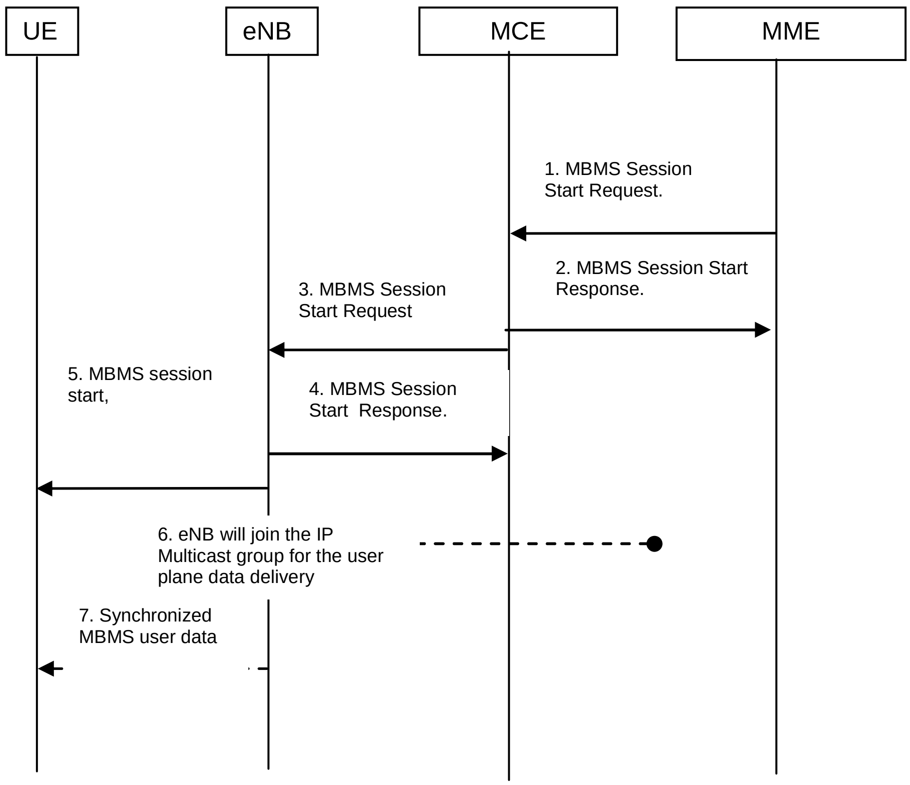
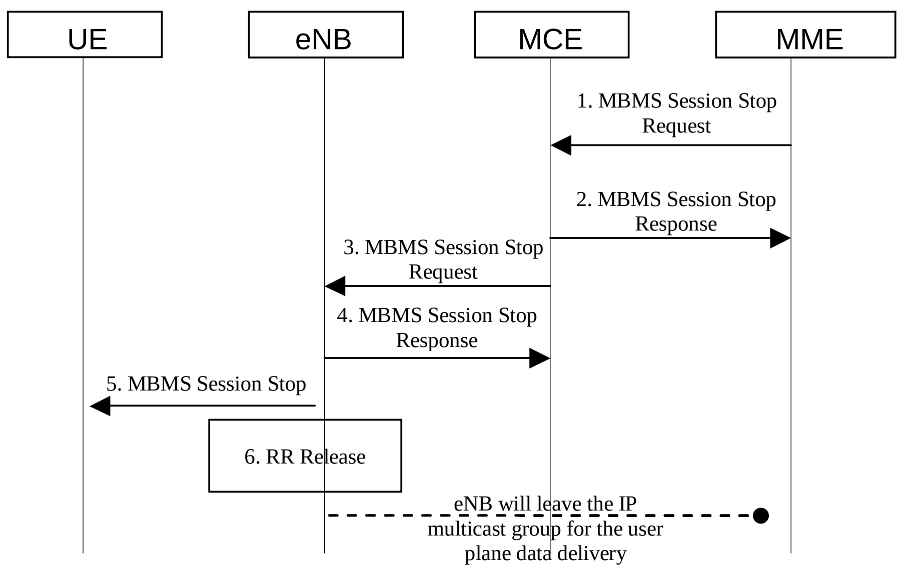


Figure 4.9 Message interactions during Session Start Procedure for E-UTRAN [16].

To start the session, MME sends a Session Start Request message to the MCE controlling the eNB in the area of transmission, to inform about the awaiting transmission. It also provides information about QoS, MBMS service areas, session identifier and other such session attributes. These attributes are saved in the MME. MCE responds with a Session Start Response message after checking the available resources in the area it controls. This message contains the okay for MME to send the data to MCE. In case the MCE finds that the resources are not enough to setup a MBMS service, it will decline the request by MME and will refrain from forwarding the request to eNBs. The saved session attributes are provided to the eNBs with a Session Start Request message. eNB also stores the session parameters and sends a Session Start Request message to the UE. eNB indicates the start of MBMS session by updating the Multicast Control Channel (MCCH) change notification and updates the MCCH information which carries the configuration information for MBMS service. eNB then joins the IP multicast group to start receiving the user data from upstream nodes. It then sends the MBMS data to the radio interface at determined time [20].

4.7.2 MBMS session stop procedure



4.10 Messages exchanged during a MBMS session stop procedure for E-UTRAN [16].

This is a procedure to signal the UE about the end of the MBMS service. This allows the UE to end all E-UTRAN Radio Access Bearer (E-RAB) associated with the MBMS service. To stop the session, MME sends a Session Stop Request to the MCE. MCE responds to this message with a Session Stop Response message. It also forwards the Session Stop Request message to the eNB. eNB responds with a Session Stop Response message and forwards the Session Stop Request message to UE and removes any service configurations associated with the particular MBMS service in the updated MCCH message. All eNB which were serving the stopped MBMS service will release the corresponding E-RAB and leaves the multicast group.

4.7.3 MBMS registration procedure

Registration is the process of downstream MBMS nodes request from the upstream nodes attributes and data related to a particular MBMS session so that it can be distributed further downstream. This procedure results in the development of a MBMS bearer context in all the nodes along a distribution path. However, this does not result in bearer plane establishment, which occurs only with the session start procedure.

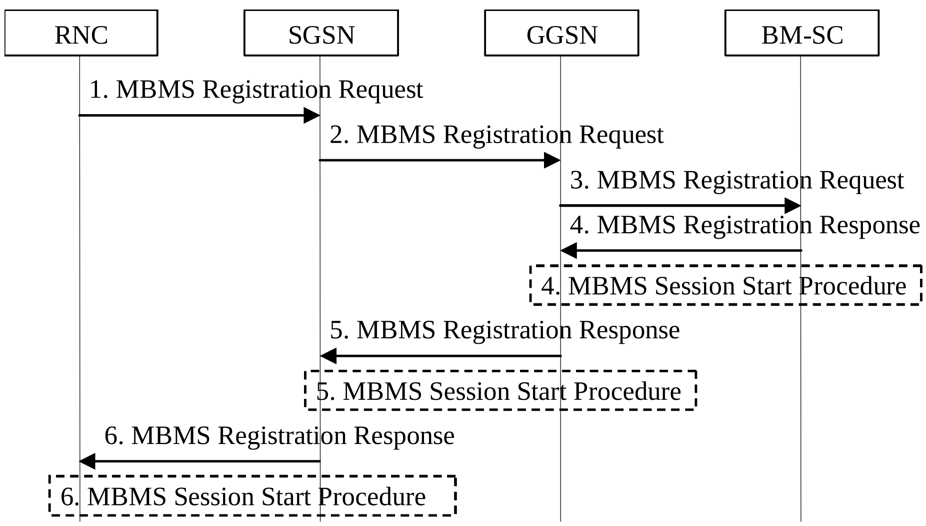


Figure 4.11 Messages exchanged during a MBMS registration procedure [16].

MBMS registration procedure is initiated only when one of the following conditions are satisfied:

- When a MBMS bearer context is not setup but the first MBMS UE context for a particular MBMS bearer service is initiated in SGSN or GGSN.
- When a MBMS bearer context is not setup and a MBMS Registration Request for a particular MBMS bearer service is received.
- When a RNC detects that an UE is interested in a particular MBMS bearer service and is connected to it.

If the last case is true, the RNC to which the UE is connected sends a MBMS Registration Request message to its parent SGSN.

4.7.4 MBMS de-registration procedure

By de-registering, a downstream node lets the upstream node that it is no longer interested in receiving the signaling, session attributes and data for a particular MBMS service.

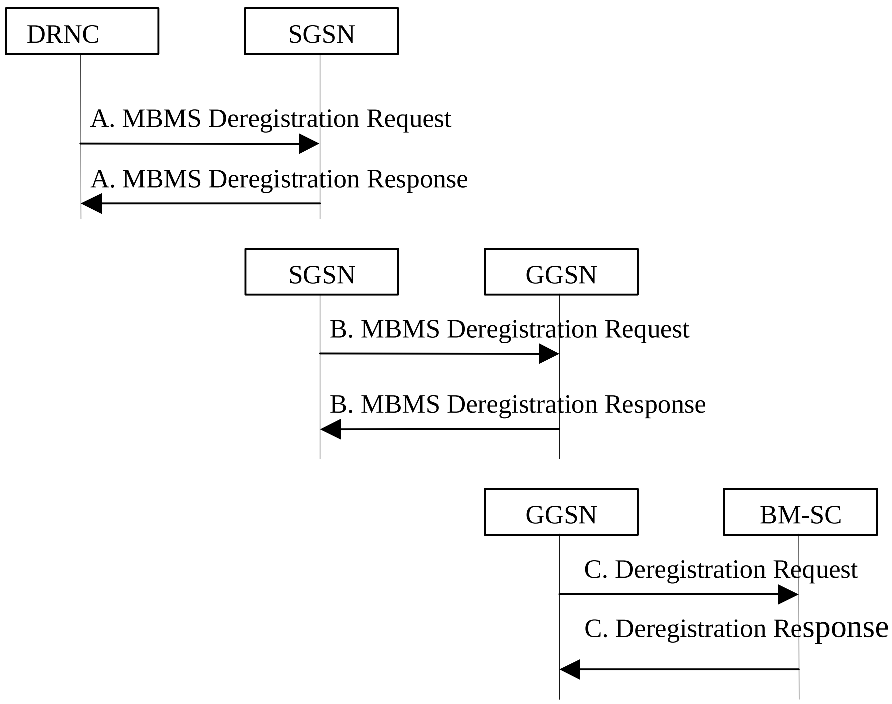


Figure 4.12 Messages exchanged during the de-registration procedure [16].

The de-registration procedure can be started by GSGN, SGSN/MME or RNC depending on the conditions stated below:

- The procedure is started by SGSN or GGSN in GERAN and MME in E-UTRAN when the MBMS UE context is deleted from the node for a particular MBMS bearer service.
- When the last node, registered in the downstream node list, de-registers for a particular bearer service and it has no corresponding MBMS UE context.

- By a RNC, registered at a SGSN, when there is no UE connected to it which is interested in a particular MBMS service.

When one of the above conditions is satisfied, the RNC will send a MBMS Deregistration Request to its parent SGSN. The RNC can also decide to wait for a pre-specified time before sending the de-registration request to the SGSN to save resources and other signaling procedure in case an interested UE decides to join back in. The SGSN removes the RNC from the list of downstream nodes and sends a confirmation in the form of MBMS Deregistration Response. When the list of downstream nodes in a SGSN has no entries left in it, SGSN will send a MBMS Deregistration Request to its upstream GGSN. The GGSN responds to this by sending a MBMS Deregistration Response and also removes the identifier for SGSN from its list of downstream nodes. Any bearer plane established between the GGSN and SGSN for this particular MBMS service will be removed after the above step.

When there are no more nodes in the downstream list of a particular GGSN, it will request to be deregistered from BM-SC by a MBMS Deregistration Request. Again, this message will contain the IP multicast address and other session attributes. The BM-SC removes the GGSN from its own downstream node list and confirms by sending a MBMS Deregistration Request message back to GGSN.

4.7.4.1 BM-SC Initiated De-registration Procedure

In a situation when the service is terminated, the deregistration is initiated by the BM-SC. The procedure results in removing any bearer context associated with a particular MBMS service which was terminated along all the nodes downstream in a distribution tree. The procedure starts with BM-SC sending a MBMS Deregistration Request message to all the GGSN that are present in the downstream nodes list for a particular MBMS service. This indicated to all the GGSN that the session was terminated and any resources allocated for providing the service downstream should be released.

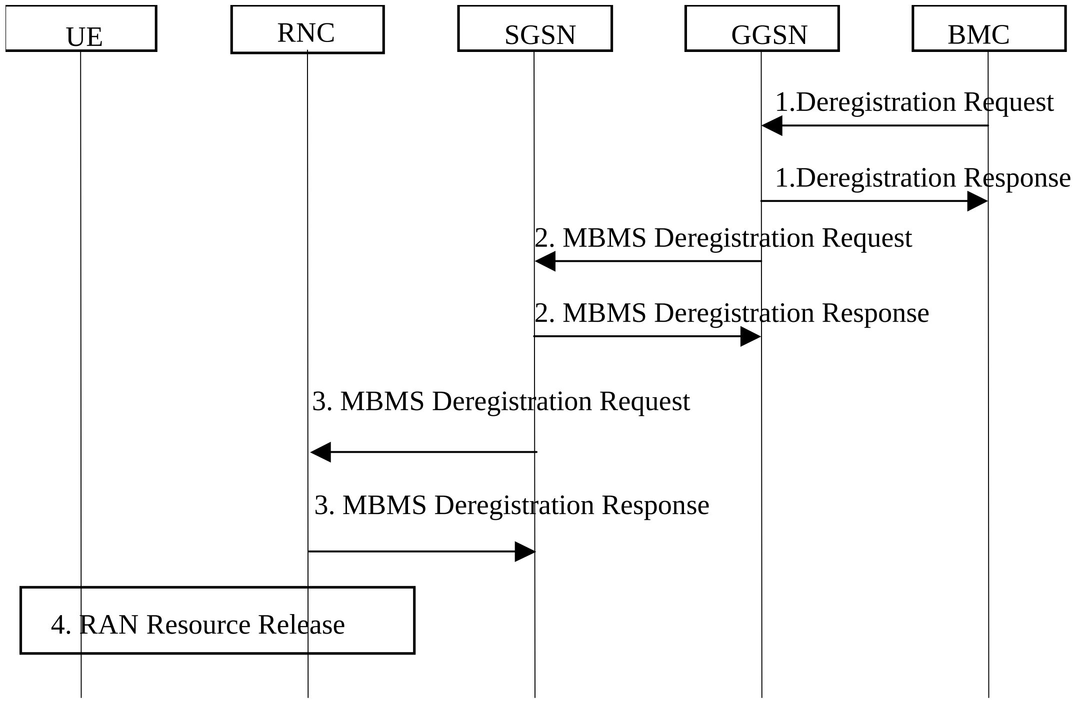


Figure 4.13 Messages exchanged during BM-SC initiated de-registration procedure [16].

All the GGSN respond with a MBMS Deregistration Response message and release any resources associated with the session. After completion of this step, each GGSN sends a MBMS Deregistration Request message to all the SSGN listed in its downstream node list for the terminated MBMS service. After replying to this request with a MBMS Deregistration Response, all the SSGN sends a Deregistration request to all the RNCs on its list of downstream nodes for the particular MBMS bearer context. The RNC responds to this message and releases all the resources it had allocated to the UEs for the purpose of connecting to the particular MBMS service. The radio access network may specify to the UEs that the service has been deactivated allowing the UE to terminate any application running on it related to the MBMS service.

4.8 M2 Application Protocol (M2AP)

The following are the functions specified for the M2AP in 3GPP specifications TS 36.443:

- Handling the MBMS session which includes session start, stop and modification. It also includes the modification and configuration of basic radio transmission parameters.
- Provide scheduling information through MCCH to the eNB.
- Provide general error notification for which function specific error notification messages were not defined.
- Reset and setup M2 interface between eNB.
- Update eNB and MCE configurations needed at the time of M2 setup.

4.8.1 MBMS session start

This procedure is used to notify UE about an upcoming MBMS bearer service and to setup the MBMS E-RAB. This start procedure is triggered by MCE.

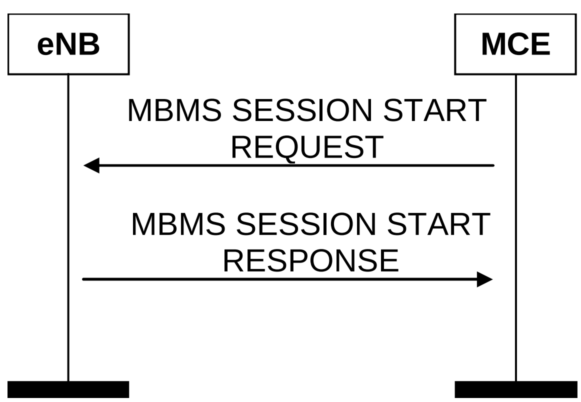


Figure 4.14 Successful MBMS session start procedure [22].

MCE sends a MBMS Session Start request message to the eNB. If the eNB has the required resources it will respond with a MBMS Session Start Rresponse message. After the response message, the eNB will join the IP multicast address.

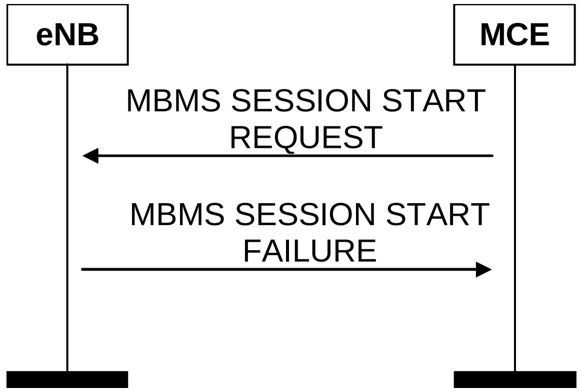


Figure 4.15 Unsuccessful MBMS session start procedure [22].

If for any reason the eNB is not able to process the request correctly, it will respond with a MBMS Session Start Failure message to the MCE.

4.8.2 MBMS session stop

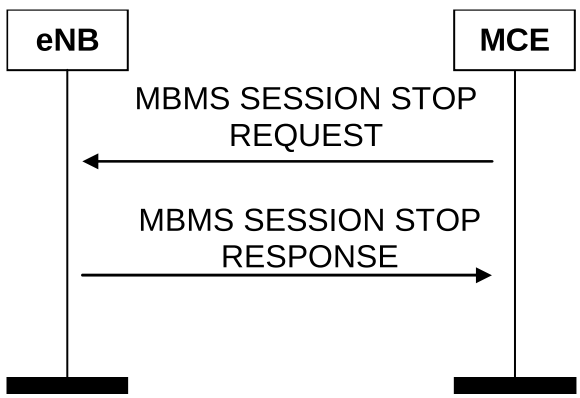


Figure 4.16 Successful MBMS session stop procedure [22].

The purpose of a MBMS session stop procedure is to inform the eNB about the end of a MBMS service so that the eNB can release any E-RAB which was established for the particular MBMS service. Again this procedure is also started by the MCE. MCE sends the MBMS Session Stop Request message to the eNB to which the eNB replies with a MBMS Session Stop Response. The eNB then proceeds to remove any MBMS bearer context information that it may contain and it will remove itself from the IP multicast group.

4.8.3 MBMS scheduling information

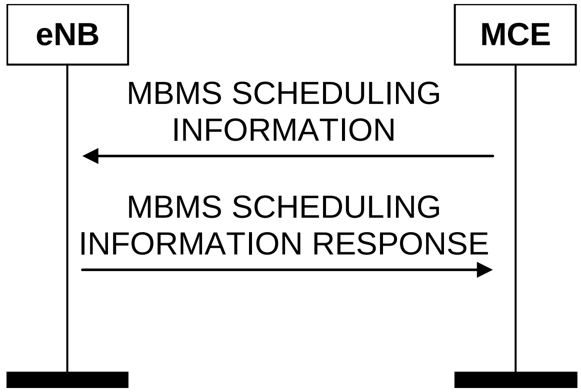


Figure 4.17 Successful MBMS scheduling information procedure [20].

MBMS scheduling procedure is initiated by the MCE to provide MCCH related information to the eNB. The procedure is started with MCE sending a MBMS Scheduling Information message to the eNB. eNB will store the MBSFN area configuration information from the MBMS Scheduling Information message provided by the MCE and apply the modification to the MCCH. The eNB will also schedule the MBMS services to appear in MCCH according to the order defined.

4.8.4 M2 setup procedure

M2 setup procedure is defined to establish the link between an eNB and MCE to exchange information and operate correctly on the M2 interface. It also configures the MCCH related information on BCCH for all the cells that are controlled by the eNB. This procedure erases any pre-existing MBMS related service context and signaling connections.

eNB starts the procedure by sending a M2 Setup Request message to the MCE. This message includes the required data, like information about the cells that are interested in receiving the MBMS service data transmission. The response to this message is a M2 Setup Response message from MCE which includes the configuration parameters for MCCH for all the list of cells provided by the eNB.

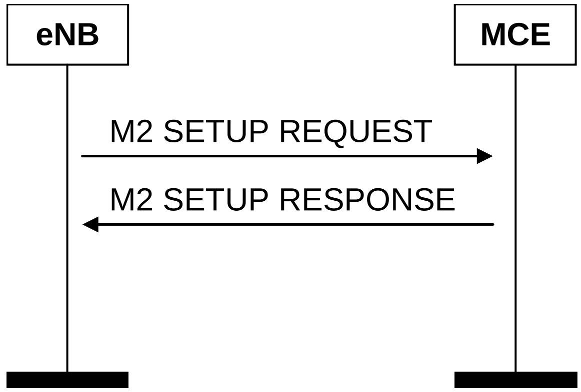


Figure 4.18 Successful M2 setup procedure [20].

eNB starts the procedure by sending a M2 Setup Request message to the MCE. This message includes the required data, like information about the cells that are interested in receiving the MBMS service data transmission. The response to this message is a M2 Setup Response message from MCE which includes the configuration parameters for MCCH for all the list of cells provided by the eNB.

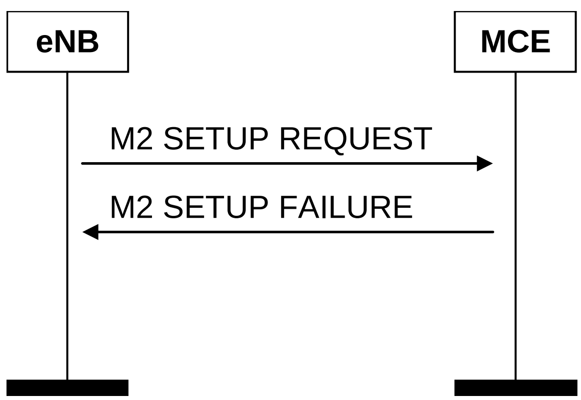


Figure 4.19 Unsuccessful M2 setup procedure [20].

The information provided by the MCE is forwarded to all the interested cells by the eNB. Once this process is complete, all the cells are ready for the particular MBMS transmission.

If for some reason the MCE cannot accept the setup message it will respond with a M2 Setup Failure message as shown in figure 4.16. This message will also contain the time to wait interval. This is the minimum time the eNB will have to wait before retrying to start the setup process.

4.8.5 M2AP reset procedures

In the event of a failure between the MCE and eNB links, it will be required to re-initialize the MBMS service contexts. The reset procedures in such a case are given below.

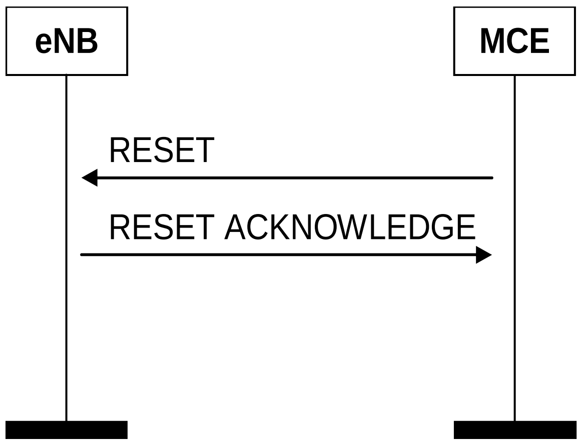


Figure 4.20 Successful reset procedure for M2AP initiated by MCE [20].

Figure 4.20 shows the reset procedure started by a MCE in case of loss of transaction information between MCE and eNB. When the eNB receives a Reset message from the MCE, it releases all resources allocated on M2 and M1 interfaces and also removes any MBMS service contexts. After releasing all the resources and service contexts the eNB will reply with a Reset Acknowledgment message to the MCE. An eNB can send the acknowledgment message before the actual release of all resources take place. If the reset message contains a list of the MBMS services, then the eNB will only reset the resources associated with those MBMS services.

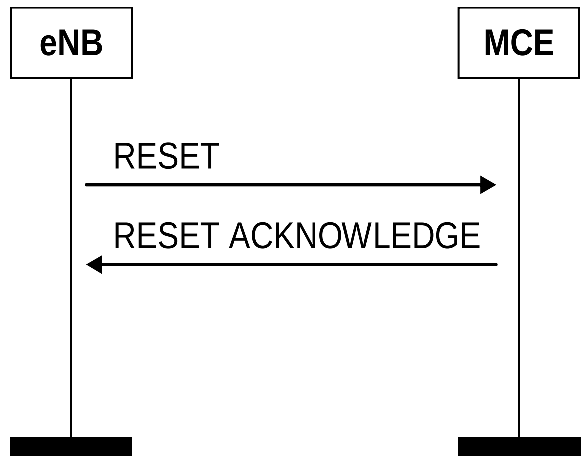


Figure 4.21 Successful reset procedure for M2AP initiated by eNB [10].

If the failure of the transaction information occurs at the eNB end, the reset message will be sent by eNB to the MCE. Figure 4.21 shows the messages exchanged during the reset procedure initiated by the eNB. When the MCE receives a Reset message, it will remove all MBMS service context on M2. Once the resources have been released, it will send an acknowledgement message to the eNB.

In the reset message if the eNB sends the MBMS service ID that need to be reset, then the MCE will reset only those MBMS services.

4.9 M3 Application Protocol (M3AP)

M3AP has been defined with following functions in the 3GPP technical specification number 36.444:

- Session management by starting, stopping and updating MBMS session.
- Reset functionality to make sure that M3 interface between MCE and MME is initialized properly.
- Error indication functionality to make sure that errors are reported and handled properly in case of a failure message.

4.9.1 MBMS session start procedure

The session start procedure allows the MME to get the information about the preparedness of E-UTRAN to handle a new MBMS service and to see if the required E-RAB have been established for the MBMS service.

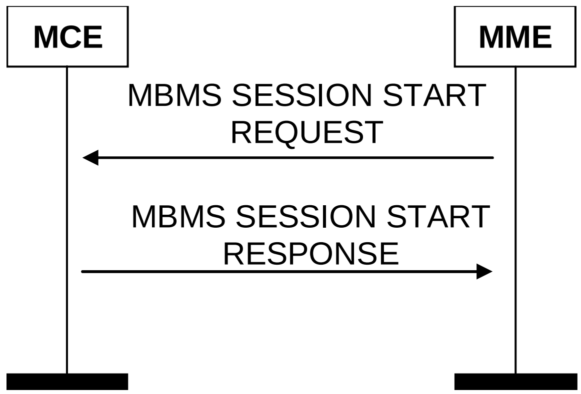


Figure 4.22 Successful MBMS session start procedure [23].

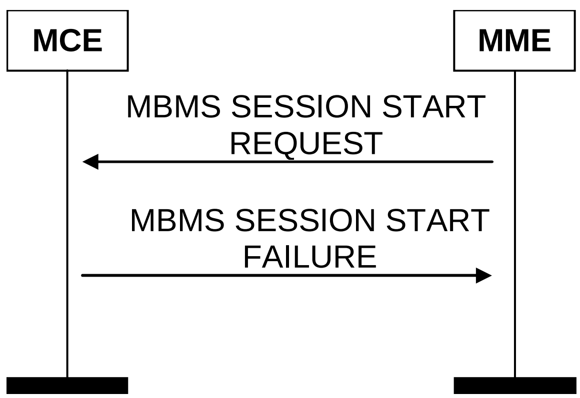


Figure 4.23 Unsuccessful MBMS session start procedure [23].

MBMS session start procedure is initiated by MME by sending a MBMS Session Start Request message to the MCE. MCE will use the MBMS E-RAB QoS parameters to determine whether the E-UTRAN has enough resources to handle the required QoS specifications. It will respond to the MME's request with a MBMS Session Start Response message in which it will provide the result of the request E-RAB analysis.

Figure 4.23 shows the response of the MCE when it discovers that the E-UTRAN will not be able to handle the QoS requested by the MME. It sends a MBMS Session Start Failure message to the MME.

4.9.2 MBMS session stop procedure

The MBMS session stop procedure is in place to let the E-UTRAN know of the end of a particular MBMS service so that the E-UTRAN can release any resources it allocated for that particular MBMS service.

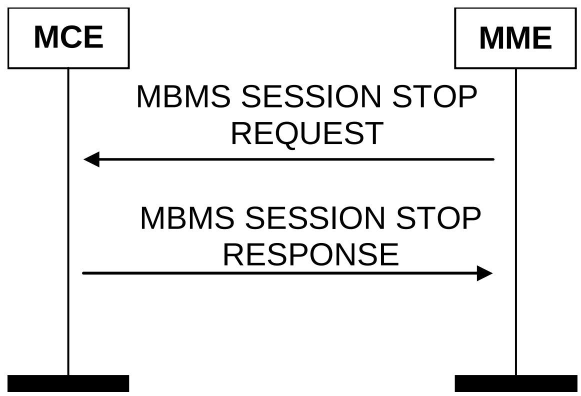


Figure 4.24 Successful MBMS session stop procedure [23].

This process is also initiated by the MME. Figure 4.24 shows that the MME sends a MBMS Session Stop Request message to MCE which replies to it by a MBMS Session Stop Response message. It will also remove any affected resources and MBMS bearer context.

4.9.3 Error indication

Error indication can be initiated by either side of a M3 interface depending on which side detected an error first. Once a node detects an error it can send an Error Indication message to the other side to let it know about the error. Figure 4.25 shows an Error Indication message from the MME side while figure 4.26 shows an Error Indication message from the MCE side.

Figure 4.25 Successful error indication procedure initiated by the MME [23].

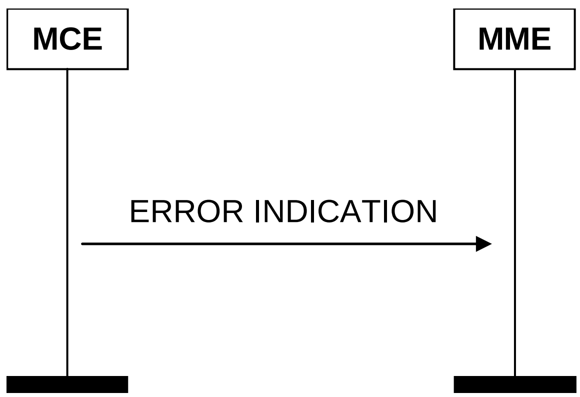


Figure 4.26 Successful error indication procedure initiated by MCE [23].

4.9.4 Reset procedures

Figure 4.27 shows the reset procedure initiated by MME. A reset procedure can be initiated by either side to re-initialize the E-UTRAN or apart of E-UTRAN M3AP MBMS related context.

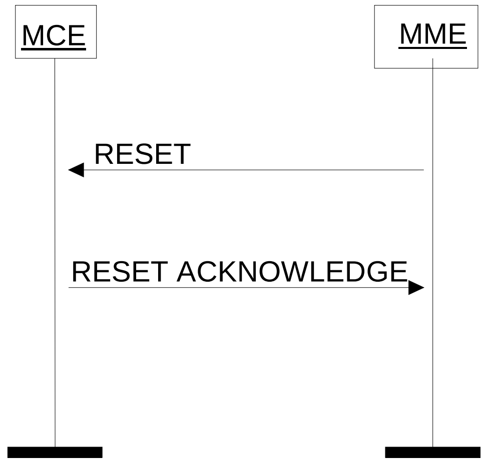


Figure 4.27 Successful MME initiated reset procedure [23].

If there is loss of some or all transaction information which leads to a failure at the MME side, it will let the MCE know of the failure and request for a reset. On receipt of the reset message, the MCE will remove all the allocated resources on the M3 related to the MBMS service indicated in the reset message. The service can be indicated by providing the MBMS M3AP IDs. After all the resources have been removed by MCE, it will respond back to the MME with a Reset Acknowledge message. MCE shall not wait for all the resources to be setup again before sending a reset acknowledge message.

In the event of a failure on the MCE side, it will send a reset message to the MME. Once the MME receives the reset message it will remove all the resources associated with all the indicated MBMS services. After the procedure of removing all the resources is completed, the MME will send a Reset Acknowledgment message. Figure 4.28 shows the procedure for sending a MCE initiated reset message.

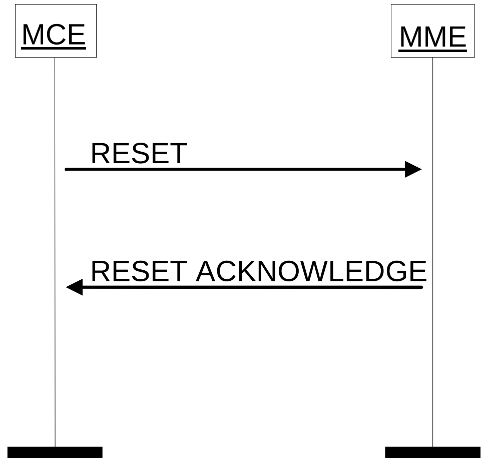


Figure 4.28 Successful reset message initiated by the MCE [23].

CHAPTER 5

CONCLUSION AND FUTURE WORK

5.1 Conclusion

LTE is a very new technology with the first few networks coming into existence for customer usage in late 2010 in the US. Great care and effort went into creating the specification documents which describe the details of LTE.

One of the problems that have been taken care of in LTE is the provision of relay architecture within the network. Even though LTE promises to provide higher data rate and greater coverage with the help of OFDMA, MIMO and advanced error control techniques, we need a technique for providing the same at cell edges. This has been a very big issue with older technologies as they did not have inherent support for relay networks.

Many different relay techniques have been used in the past. Analog repeaters have been used in the past and are one of the simplest techniques to be used for a relay node. In chapter 3, we took a look at the Amplify and Forward technique, which applies linear transformation to the received signal before forwarding it. We also looked at the Decode and Forwarding technique where the signal is decoded and re-coded before transmission. We have also understood the way a UE will act with the new relay architectures and the services and tasks presented to a RN.

During my work, I had the opportunity to read and learn about these techniques. As future work, it would be a very good learning experience if the techniques listed here can be simulated in a real world environment. These can also be simulated on software like MATLAB® or OPNET.

As LTE will be used to support voice networks initially and replace them later on, it is crucial that it live up to the performance of the current voice networks or perform even better.

With the current voice networks the voice call rates have gone down to 0.3%, LTE systems need to perform better than this along with increase in voice quality and jitter. QoS has to be implemented very carefully so that it does not use up all the resources by providing high call quality for voice call leaving the data connections with poor uplink and downlink speeds. It will also have to provide reliable emergency calls without a switch to the CS network.

With the first few LTE systems rolling out in US in late 2010, it will be interesting to see how the networks hold up in the real world scenario. All the research and theoretical claims of the LTE system needs to be verified by testing the deployed networks for promised speeds and services.

5.1 Future Work

5.2.1 *Test scenarios*

During my internship with Motorola, I had the opportunity to work on their WiMAX network to perform a QoS analysis for VoIP calls. Since the tests performed and the results obtained are strictly company confidential, I will not be able to share those results here. Based on my experience and work done at Motorola, I can suggest a few test scenarios that can be performed with LTE networks for VoIP or VoLTE.

Depending on the device that will be used to test the VoIP calls, a tool will be required to remove any LTE frames which will be added to the main transmitted frames. These frames are required for transmission of the packets within the LTE network. A VoIP frame will look similar to figure 5.1 after an LTE frame is attached to it.

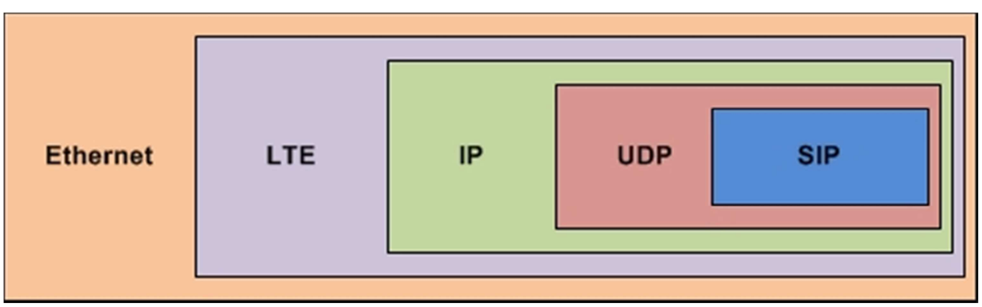


Figure 5.1 VoIP packet encapsulated in LTE frame.

LTE is not as widespread as other protocols that are used in VoIP calls. As a result, software available to read such a frame format is not readily available. For reading these packets and QoS analysis, without LTE frame in it, the following tools can be used:

- OmniPeek
- ManageEngine
- Wireshark
- VQManager

To read a packet with an LTE frame in it, the software would need an add-on. The problem with having a common add-on for all is that every company will have a unique frame for its LTE network and it is not feasible to design a software or add-on specific to companies. The companies themselves can design an add-on, but most of the software are licensed and will not allow companies to make changes in them.

An easier way would be to create a tool which can be run separately on all packets to remove the LTE frame and leave the rest of the packet untouched. This allows the above software to read the frame without any difficulty. The tool can be made using any programming language.

To use the tool, a packet capture (PCAP) file will be required. This file can be obtained using Wireshark. A PCAP file can be opened as a text file and the tool would do the same by opening the file as a text file. It would then identify and remove the LTE frame. It is important to note that when the LTE frame has been removed, the value in Ethernet frame, which is associated with the next frame, will have to be changed to IP frame instead of LTE frame.

The testing for QoS will require at least two UEs connected to two laptops running Wireshark or the UEs running Wireshark on their own (if possible). One UE acts as the originator of call, while the other acts as a terminator. Separate UEs are needed to get the correct packet loss estimate.

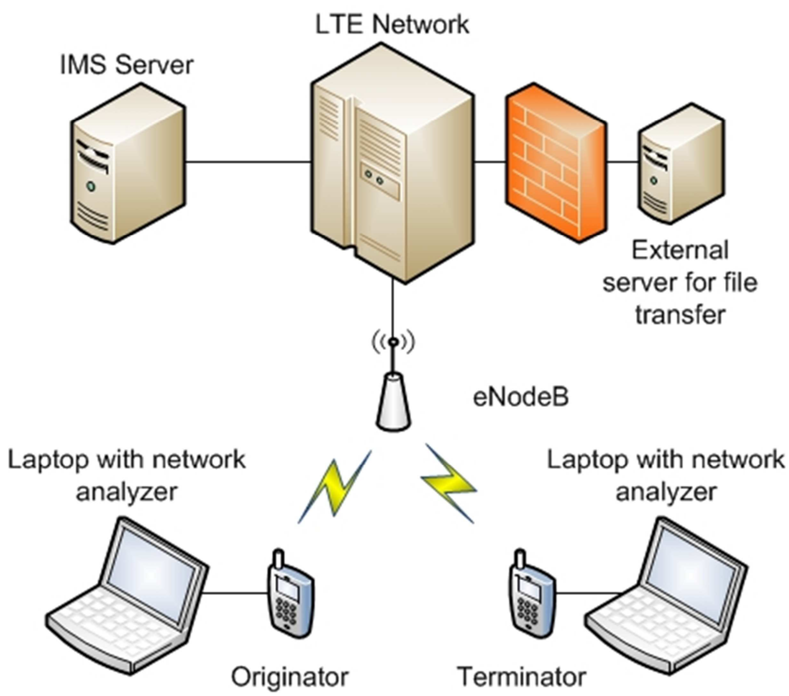


Figure 5.2 QCI testing with file transfer.

The packets captured by Wireshark will be run through the tool to remove LTE frames from it. The UE will also be connected to the LTE network through eNB. To perform a test with load on network, a voice call will be established while downloading a large file from an external network. Figure 5.2 shows QCI testing with file transfer. If the QoS is configured properly, the packets for voice will get higher preference and will have minimum delay. These tests should also be performed with different codecs to see which codec performs the best.

In figure 5.2, the tests are performed with one QCI value which is changed after the test is complete. For a better result, the test should be performed with two or more QCI values using four or more UEs. This will eliminate the environmental conditions affecting the result.

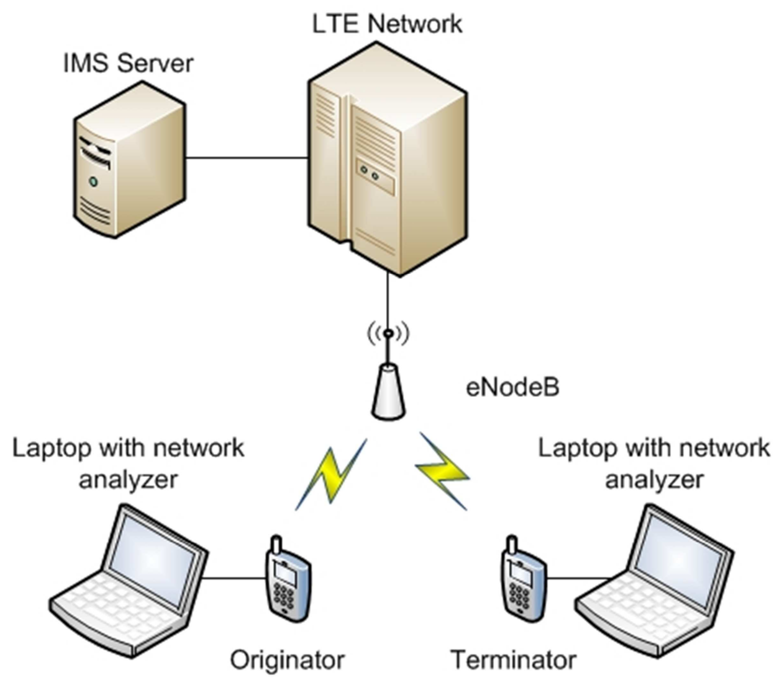


Figure 5.3 QCI testing without load on the network.

The tests should also be performed without a file transfer. This allows the testing of the basic QoS rating. If the basic QoS rating is configured perfectly, the MOS scores for both (basic QoS and high end QoS) QoS ratings should be similar. A sample setup is shown in figure 5.3.

The Mean Opinion Score (MOS) obtained for a QCI value fit for VoIP calls should be compared with MOS scores for best effort transmission.

APPENDIX A

ABBREVIATIONS

AMBR: Average Maximum Bit Rate

APN: Access Point Name

ARP: Allocation and Retention Priority

AuC: Authentication center

AUTN: Authentication Token

BS: Base Station

BM-SC: Broadcast Multicast Service Center

CP: Control Plane

DeNB: Donor eNodeB

DHCP: Dynamic Host

E-UTRAN: Enhanced Universal Terrestrial Radio Access Network

E-RAB: E-UTRAN Radio Access Bearer

EPC: Evolved packet core

EPS: evolved packet system

GERAN: GSM edge radio access network

GGSN: Gateway GPRS Serving Node

GSM: Global System for Mobile communication

GTP: GPRS Tunneling Protocol

GUTI: Globally Unique Temporary Identifier

HSS: Home Subscriber Server

IK: Integrity Key

IMSI: International Mobile Subscriber Identity

MBMS: Multimedia Broadcast Multimedia Service

MBMS GW: MBMS Gateway

MBSFN: Multimedia Broadcast Single Frequency Network

MCE: Multicell/Multicast Coordination Entity

MME: Mobility Management Entity

MOS: Mean Opinion Score

MSC: Mobile Switching Center

MSP: MCH Scheduling Period

O&M: Operation and Maintainence

P-GW: Packet Gateway

PCAP: Packet Capture

PDN: Packet Data Network

PCC: Policy and Charging Control

PCEF: Policy and Charging Enforcement Function

PCRF: Policy and Charging Rules Function

PMIP: Proxy Mobile IP

QCI: QoS Class Identifier

QoS: Quality of Service

RAND: Random number

RLC: Radio Link Control

RN: Relay Node

S-GW: Serving Gateway

SABM: Set Asynchronous Balance Mode

SAE: System Architecture Evolution

SCC AS: Service Centralization and Continuity Application Server

SDP: Session Description Protocol

SIP: Session Initiation Protocol

SGSN: Serving GPRS Support Node

SRVCC: Single Radio Voice call continuity

STCP: Streaming Control Transmission Protocol

STN-SR: Session Transfer Number for SRVCC

TA: Tracking Area

TMGI: Temporary Mobile Group Identity

TNL: Transport Network Layer

TTI: Transmission Time Interval

UE: User Equipment

UMTS: Universal Mobile Telecommunication System

UP: User Plane

URI: User Resource Identifier

VoIP: Voice over IP

VCC: Voice call continuity

WCDMA: Wideband Code Division Multiple Access

REFERENCES

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

[2] H. Holma, A. Toskala, *LTE, for UMTS - OFDMA and SC-FDMA Based Radio Access*, John Wiley and Sons Ltd., 2009.

[3] M. Sauter, *Beyond 3G – Bringing Networks, Terminals and the Web Together*, John Wiley and Sons Ltd. 2009.

[4] M. Olsson, S. Sultana, S. Rommer, L. Frid and C. Mulligan, *SAE and the Evolved Packet Core, Driving the Mobile Revolution,*Academic Press, 2009.

[5] IP Multimedia Subsystems [Available at: http://en.wikipedia.org/wiki/IP_Multimedia_Subsystem]

[6] A. K. Salkintzis, M. Hammer, I. Tanaka, C. Wong, *Voice Call Handover Mechanism in Next-Generation 3GPP Systems,* IEEE Communication Magazine, 2009, pp. 46-56.

[7] 3GPP, Technical Specification Group Services and System Aspects, *Single Radio Voice Call Continuity (SRVCC), Stage 2*, Release 9, *TS* 23.216 v9.4.0, 2010.

[8] 3GPP, Technical Specification Group Services and System Aspects, *General Packet Radio Service (GPRS) Enhancement for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) Access,* Release 10, TS 23.401 v10.1.0, 2010.

[9] 3GPP, Technical Specification Group Radio Access Network, *Feasibility Study for Further Advancement for E-UTRAN (LTE-Advanced),* Release 9, TS 36.912 v9.3.0, 2010.

[10] 3GPP, Technical Specification Group Radio Access Network, *Evolved Universal Terrestrial Radio Access (E-UTRA), Relay Architectures for E-UTRA (LTE-Advanced),* Release 9, TS 36.806 v9.0.0, 2010.

[11] A. Lo, I. Niemegeers, *Multi-hop Relay Architectures for 3GPP LTE-Advance*, IEEE 9th MICC, 2009.

[12] C. Hoymann, A. Racz, N. Johansson, Johan Lundsjo, *A Self backhauling Solution for LTEAdvanced*, WWRF, 2008.

[13] *Multimedia Broadcast/Multicast Service (MBMS),* White paper, TeliaSonera, August 2004. [Available at: www.medialab.sonera.fl]

[14] 3GPP, Technical Specification Group Services and System Aspects, *Multimedia Broadcast/Multicast Service (MBMS) user services, Stage1,* Release 9, TS 22.246 v9.0.0, 2009.

[15] 3GPP, Technical Specification Group Services and System Aspects, *Multimedia Broadcast/Multicast Service (MBMS), Protocols and Codecs,* Release 9, TS 26.346 v9.3.0, 2010.

[16] 3GPP, Technical Specification Group Services and System Aspects, *Multimedia Broadcast/Multicast Service (MBMS), Architecture and Functional Description,* Release 9, TS 23.246 v9.0.0, 2009.

[17] 3GPP, Technical Specification Group Radio Access Network, *Evolved Universal Terrestrial Radio Access Network (E-UTRAN),General Aspects and Principles for Interfacing Supporting Multimedia Broadcast Multicast Services (MBMS) within E-UTRAN,* Release 9, TS 36.440 v9.1.0, 2010.

[18] E. Dahlman, S. Parkvall, J. Sköld and P. Beming, *3G Evolution, HSPA and LTE for Mobile Broadband,* Academic Press, 2007.

[19] S. Sesia, I. Toufik and M. Baker, *LTE – The UMTS Long Term Evolution, From Theory to Practice,* John Wiley and Sons Ltd, 2009.

[20] 3GPP, Technical Specification Group Radio Access Network, *Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN), Overall description, Stage 2,* Release 10, TS 36.300 v10.0.0, 2010.

[21] 3GPP, Technical Specification Group Radio Access Network, *Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN), Overall description, Stage 2,* Release 10, TS 36.300 v10.0.0, 2010.

[22] 3GPP, Technical Specification Group Radio Access Network, *Evolved Universal Terrestrial Radio Access Network (E-UTRAN), M2 Application Protocol (M2AP),* Release 9, TS 36.443 v9.2.0, 2010.

[23] 3GPP, Technical Specification Group Radio Access Network, *Evolved Universal Terrestrial Radio Access Network (E-UTRAN), M3 Application Protocol (M3AP),* Release 9, TS 36.444 v9.2.0, 2010.

[24] 3GPP, Technical Specification Group Services and System Aspects, *Multimedia Broadcast/Multicast Service (MBMS), Architecture and Functional Description,* Release 6, TS 23.846 v6.1.0, 2002.

[25] 3GPP, Technical Specification Group Radio Access Network, *Introduction of the Multimedi Broadcast Multicast Service (MBMS) in the Radio Access Network (RAN), Stage 2,* Release 9, TS 25.346 v9.0.0, 2009.

[26] 3GPP, Technical Specification Group Services and System Aspects, 3G Security, Security of *Multimedia Broadcast/Multicast Service (MBMS),* Release 9, TS 33.246 v9.0.0, 2009.

[27] 3GPP, Technical Specification Group Radio Access Network, *Evolved Universal Terrestrial Radio Access Network (E-UTRAN), Layer 1 for Interfaces Supporting Multimedia Broadcast Multicast Services (MBMS) within E-UTRAN,* Release 9, TS 36.441 v9.0.0, 2009.

[28] 3GPP, Technical Specification Group Radio Access Network, *Evolved Universal Terrestrial Radio Access Network (E-UTRAN), Signaling Transport for Interfaces Supporting Multimedia Broadcast Multicast Services (MBMS) within E-UTRAN,* Release 9, TS 36.442 v9.1.0, 2010.

[29] 3GPP, Technical Specification Group Services and System Aspects, *Multimedia Broadcast/Multicast Service (MBMS), Stage 1,* Release 9, TS 22.146 v9.0.0, 2008.

[30] 3GPP, Technical Specification Group Services and System Aspects, *Circuit Switched (CS) Fallback in Evolved Packet System (EPS), Stage 2,* Release 9, TS 23.272 v9.4.0, 2010.

BIOGRAPHICAL INFORMATION

Ayush Maheshwari was born on May 20th, 1986 in Uttar Pradesh, India. He pursued his Bachelor's degree from Mumbai University in Instrumentation Engineering and graduated in May 2008. His keen interest in telecommunications brought him to US in fall 2008, where he joined the University of Texas at Arlington as a Master's candidate in Electrical Engineering. While completing his Master's degree, he interned at Motorola as a part of the VoIP and Multimedia Team and got his first experience with 4G technology. He completed his Master's degree in December 2010. His interests lie in next generation networks, wireless communication and mobile devices.