IMS in 5G

Analysis of IMS based communication services in the 5G network

Manasa Dattatreya

IMS in 5G

Analysis of IMS based communication services in the 5G network

by

Manasa Dattatreya

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on August 19, 2020 at 10:00 AM.

Student number:	4749634	
Project Duration:	October, 2019 – August, 2020	
Thesis committee:	Dr. Ir. Eric Smeitink,	KPN, TU Delft
	Dr. Ir. Qing Wang,	TU Delft
	Ir. Rogier Noldus,	Ericsson, TU Delft

An electronic version of this thesis is available at `http://repository.tudelft.nl/`.

Acknowledgements

This thesis marks the completion of my Master's Degree in Telecommunication and Sensing Systems at Delft University of Technology. This success has been possible because of the support of a lot of people.

First, I would like to thank Ir. Rogier Noldus for agreeing to mentor me and giving me the wonderful experience of working at Ericsson. I have always received timely guidance, critical feedback and immense support on all fronts for which, I will forever be grateful. I would also like to thank Mr. René van der Mast, my manager at Ericsson for supporting me throughout the time spent at Ericsson.

Next, I would like to thank Dr. Ir. Eric Smeitink and Dr. Ir. Qing Wang for giving their valuable feedback and agreeing to be a part of my thesis committee.

I would also like to thank Dr. Ir. Edgar van Boven for his support and guidance in finalizing the next phase of my life after my masters.

I would like to extend my gratitude to all my friends who have been a part of my journey. The past three years have been enjoyable because of them.

Last but not the least, I would like to extend a big thank you to my family who have seen the highs and lows during my study. They have always stood by my side, forgiven my mistakes and encouraged me to keep moving forward. I dedicate this success to them.

Manasa Dattatreya, August 2020

Abstract

The thesis focuses on investigating the role of IP Multimedia Subsystem (IMS) in 5G networks. IMS already plays a very important role in enabling a wide range of real-time multimedia communication services such as basic phone calls and messaging in the LTE network. Addition of application servers on top of the IMS core can provide enhanced functionalities like presence, advanced messaging and SIP trunking. IMS guarantees quality, security and reliability of multimedia services when serving users without the installation of any application as well as flexibility over access. This sets IMS apart from other third party applications found on the internet. IMS was created with the idea of being adaptable to the evolving technology. With the implementation of the 5G network underway, a study of the impact of this new architecture on the IMS services is imminent. The thesis focuses on the study of different network elements participating in an IMS service as well as investigating IMS voice and video calls over the 5G network.

The thesis consists of two parts: The first part involves exploring the role of IMS in 5G networks. A brief overview of the evolution of mobile networks will help understand the differences between 1G/2G/3G/4G. Following this, IMS and its role in enabling multimedia services in the LTE network is explored. Next, the 5G System Architecture is explained along with components and their functions. Next, a comparison between the elements in 4G and 5G provides a clear understanding of the technological evolution and the procedure involved. The next steps would include understanding the IMS call flow in LTE networks. This would provide a good foundation to understand how the IMS services will be provided over the 5G network.

The second part of the thesis includes testing the voice and video services over Ericsson's 5G network at their Rijen office. As a starting step, the voice and video services are tested using WebRTC. Upon succeeding in the peer-to-peer test, the next step is establishing connectivity to Ericsson's IMS network at Kista, Sweden. This will allow the testing of IMS voice and video services over the 5G network. The results are recorded and analysed. These are in agreement with the theoretical expectations.

Contents

List of Figures

List of Tables

1 Introduction

This chapter introduces the topic and the objectives of the thesis. In Section 1.1, an introduction to 5G is provided. Section 1.2 gives a brief introduction to the IP Multimedia System (IMS). Section 1.3 presents the objectives of this thesis. Lastly, Section 1.4 presents the outline of the thesis report.

1.1. Fifth Generation Mobile Networks (5G)

The number of mobile communication devices has drastically increased over the past few years. The ever-growing demand for faster and better-quality content from the internet has driven the development of 5G. The Fourth generation mobile networks (4G) continues to exist during the development of 5G. The acquisition of the right spectrum to implement the 5G radio access network has been one of the most important tasks. The development of 5G is driven by the different use cases it will support. Enhanced Mobile Broadband (eMBB), Ultra Reliable Low Latency Communication (URLLC) and massive Machine type communications (mMTC) are the three major drivers in the development of 5G. eMBB includes support for broadband access in dense areas providing services like pervasive video and 50+ Mbps speed made available everywhere even at cell edges. URLLC includes use cases like automation in automobiles and factories, and remote control in healthcare applications. mMTC includes use cases like smart sensors and Internet of Things (IoT) which support a large number of devices.

As much as 5G would provide new business opportunities for the telecom operators, it will also greatly impact the vertical industries by providing extremely good connectivity throughout the product cycle. This in turn, necessitates having a unified 5G standard for collaborations between different industries. 5G promises speeds between 50 Mbps and 1 Gbps and an average Round Trip Time (RTT) of 1-10ms [1]. These are the characteristics that will play a crucial role in connecting smart devices and enabling use cases like smart cities, smart agriculture etc. Some of the use cases and applications are presented in Figure 1.1.

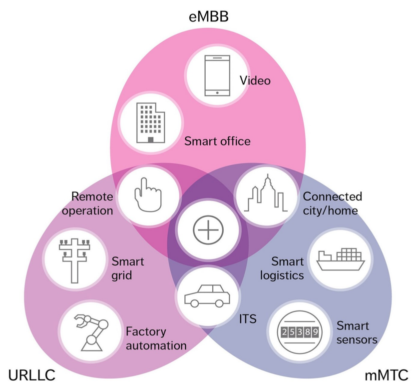

Figure 1.1: Use cases and applications of 5G [2]

1.2. IP Multimedia Subsystem (IMS)

IP Multimedia Subsystem (IMS) is an architectural framework offering different types of IP multimedia services with predictable Quality of Service (QoS). IMS is a way of delivering multimedia (voice, video, data) regardless of the device or the access medium. It enables appropriate charging of the multimedia sessions. It focuses on providing the service without being concerned about how the specific access network is operating. The control plane and the user plane are separated [3]. The IMS architecture proves to be advantageous particularly to the Service Providers (SP). It separates the services from the access network thereby enabling them to develop and deploy applications with reduced cost and complexity. IMS architecture and its components are described in detail in Chapter 4.

1.3. Objectives

The structure of the thesis is divided into two parts:

- Investigating the role of IMS in 5G networks. It includes a detailed study of the following:
  - The relevant technical specifications and standards: 3GPP, GSMA, IETF, ITU-T
  - IMS in the context of 4G
  - 5G architecture and different deployment options
  - Transition from 4G to 5G
  - IMS call flow in 4G and 5G
- Experiments
  - Understanding the setup of the 5G network at Ericsson, Rijen
  - Establishing a peer to peer connection (messaging/audio/video) using WebRTC
  - Extending this setup to connect to the Ericsson IMS network in Kista, Sweden
  - Measuring the performance over the 4G and 5G radio access links with repeated tests. Parameters under consideration are bitrate, jitter, packet loss and latency which are indicative of the Quality of Service
  - Comparing the performance of different IMS services (audio/video/messaging) and analyzing the results

1.4. Thesis Outline

The thesis report constitutes a total of eight chapters. Chapter 2 covers the literature review for this thesis. Chapter 3 explains the evolution of mobile communication technologies from 1G to 4G along with a brief insight on their respective architecture. Chapter 4 explains the need for IMS along with its architecture. It also explains how VoLTE and IMS make calls and other IM services possible. Chapter 5 explains the 5G System, the spectrum options available, the different deployment options, and the transition from 4G to 5G. Following this, the concepts of CUPS, SDN and XHaul are briefly explained. Voice migration to 5G is briefly explained as a conclusion to the chapter. Chapter 6 explains the network setup at Ericsson's 5G garage in Rijen, the role of WebRTC and the Web Communication Gateway in the experiments. Chapter 7 is dedicated to the practical part of this thesis that includes the experiments, the results and the analysis. Chapter 8 presents the conclusions and future work.

2

Literature Overview

2.1. Research Motivation

According to GSMA, as of July 2019, 185 operators have launched Voice services over Long Term Evolution (4G) (VoLTE) services (explained in Chapter 4) across 93 countries [4]. Out of the 7.9 billion subscribers, fewer than 50% are on non-VoLTE LTE and less than 20% are on VoLTE. The remaining subscribers are still served by the 3G/2G or the circuit-Switched networks. With 4G radio, voice will run over the VoLTE network. As the operators migrate to 5G New Radio (NR), voice will continue to be supported by an IMS core. Therefore, IMS will continue to be the core infrastructure for voice across wireline and wireless operators. With the implementation of Voice over New Radio (VoNR), the operators will also be pushed to extend their services beyond voice to provide features like enhanced messaging like Rich Communication Suite (RCS), business services such as unified communication and open Application Programmer's Interfaces (APIs). This also triggers a requirement to switch to virtualized and cloud-native networks to be able to compete in the market with new service creation. IMS beyond voice is utilized to a very small extent in rich messaging and business services. By extending the services beyond rich messaging, mobile operators can bring in more revenue and provide additional services like audio/video conferencing, video sharing, file sharing, presence and group chat [4]. Exploring the impact that the fifth generation of mobile networks will have on these IP Multimedia services is the motivation for this thesis.

2.2. Related Research

Different standardization bodies were formed to develop a baseline for requirements that need to be fulfilled by the service providers to maintain consistency and quality across the technologies developed in the industries. Third Generation Partnership Project (3GPP) is responsible for the development of standards for Global System for Mobile Communication (2G/2.5G), Universal Mobile Telecommunication System (UMTS), Long Term Evolution (LTE), 5G New Radio (NR), 5G Core (5GC) and IMS. The different aspects of each of these systems like nodes, protocols and interfaces are specified in the Technical Specifications (TS). The Internet Engineering Task Force (IETF) is responsible for developing the architecture for the Internet. IMS uses different protocols for establishing a session and media transfer. Session Initiation Protocol (SIP) and Diameter Protocol are used for signaling. Real Time Protocol (RTP) and Real Time Control Protocol (RTCP) are used for voice or video media transmission. These were developed by IETF. GSM Association is another organisation that is an consortium of different mobile operators in the telecom industry. Standards for VoLTE and RCS are defined by GSMA.

Like GSM/2G/3G/LTE networks, 5G will provide a number of value propositions like delivery with consistent experience, access on multiple interfaces/devices, trusted and reliable communications, resilient networks, responsiveness and real-time communications. From the consumer perspective, 5G will provide higher data rates and lower latency. From the enterprise perspective, the consistency of the experience will be the best aspect of 5G. Mobilizing and automating industries, and different processes between them (Machine Type Communication and Internet of Things) will start due to the onset of 5G

[1]. The enhancement on Critical Communication, mMTC, Internet of Things (IoT), Vehicle to anything (V2X) communications, Mission Critical and features related to WLAN and unlicensed spectrum in the initial phase of the 5G System (5GS) is defined in [5]. 5GS refers to the complete network with NR as the RAN and a 5G core. The use cases for 5G are presented in [1] and [5].

The current day networks will undergo changes to incorporate the more recent technology like Software Defined Networking (SDN) and Network Function Virtualization (NFV). These technologies will ensure better utilisation of existing resources while providing new services at the same time. The services provided by the mobile operators will evolve to improve both in quality and capability. Social media and video traffic will dominate the data traffic volume [1].

The architecture for the 5G System (5GS) i.e a combination of 5G NR RAN and 5GC is described in [6]. It describes the service based architecture of 5G as well as the reference point representation between any two network functions. Service-Based representation indicate that the Network Functions (NF) within the control plane allow access to the services offered by other NF upon authorization. The names for the interfaces in a service-based representation are according to the services offered by the NF (e.g. Npcf for Policy and Charging Function, Nudm for Unified Data Management). The Reference-point representation shows how different NFs interact with each other. The interface between two NFs is represented by means of an interface name (e.g. N1 is the interface between the User Equipment (UE) and the Access Mobility Function (AMF), N2 is the interface between Radio Access Network (RAN) and AMF). The deployment of 5G Standalone (SA) network requires the intermediate support of the Non Standalone Architecture (NSA) where the Enhanced Packet Core (EPC) interworks with the 5GC. NSA refers to the 5G radio network in combination with the EPC while SA refers to a full-fledged 5GS with a 5G radio network and a 5G core (5GC). Understanding both architectures is an important aspect of the 4G to 5G migration. The interworking possibilities in different network scenarios (roaming/non-roaming/non-3GPP access) are explained in [6]. The control plane and the user plane are separated. In continuation to this, [6] also describes the IMS voice support over PS session over 3GPP/non-3GPP access and also explains the different ways services are provided when the subscriber is in a roaming network. The aspects of EPS Fallback or RAT fallback when supporting IMS voice is also covered in [6]. 5G system architecture is proving to be a major driver in the process of migrating the IMS based services from 4G to 5G.

The architecture of a mobile network will see a shift due to the inception of 5G. The telecommunications industry is now looking more towards a cloud-based architecture. The main drivers for the technological shift to 5G are Service-based Architecture (SBA), Network Slicing, Statelessness and Common Core. SBA enables the addition of new services without impacting any of the existing ones thereby making scalability simple and seamless. The process and storage pertaining to a particular session are stored in a separate database defined as a Network Function. This eliminates the concept of the UE Context during the session which is called stateful. Network Slicing categorizes the required NFs for a particular service to offer to a dedicated network allowing the fast build up of a network and improving the operational efficiency. With a common core, it will be possible to integrate the different access technologies via a common interface (e.g. N26 interface between EPC and 5GC) [7] [8].

The different 5G system procedures relating to (not limited to) connection, registration, mobility management, session management, handover and policy framework are defined and described in [9]. Flow diagrams for each of these procedures give a clear idea about the activities that participate in the process of serving a subscriber connected to the network. Some examples of the procedures are Protocol Data Unit (PDU) session establishment, Session Managemenet Function (SMF) selection , User Plane Function (UPF) selection, PDU session modification and Session Release. These procedures help develop a clear understanding of how the services are provided to the users over the network.

A large extent of the information available on IMS is in the context of VoLTE. One among the many Technical Specifications (TS) available for IMS is presented in [10]. The document defines the IP Multimedia Core Network Subsystem which includes the necessary elements to support the IMS services. A complete overview of the IMS network and the way the services are operational in a network is presented in [10]. All the procedures and requirements that are to be met are explained in depth. The TS defines the mechanisms to enable support for IMS multimedia applications. In addition to the TS specified in [10], there are other TS's that give a clear picture of IMS and its interaction with the mobile core. The signalling in the IM services takes place using SIP. The end-to-end (E2E) call procedures like UE registration, authentication, call initiation, call session establishment, session modification and call release work together in a flow to connect two users and enable a call between them using IMS. There are different procedures that are taking place at different nodes in the IMS network. Each node has a significant role to play. These procedures need to be known in order to understand how the operators provide the mentioned IMS services to the customer(s) [11].

Spirent technologies has also written a whitepaper on the IMS architecture [12] which highlights how telephony and IP services, access technologies, service types, location information and control functions can be converged by IMS.

The integration of IMS in 5G is backed by a generic understanding of the migration from 4G EPC to 5GS. Potential enhancements to the IMS for integrating with the 5GC functions are identified in [13] . Some of the key issues along with the possible solutions for the integration between 4G and 5G to be successful are presented in [13]. The network slicing feature in 5GC is suggested to be used for different IMS services. Anytime a new technology like 5G is being rolled out, the primary target is to provide as much coverage as possible with predefined QoS for different services. In areas where 5G coverage would be difficult, a seamless integration with the underlying 4G/3G networks is necessary at least during the migration phase. From the architectural point of view, the different ways in which a call could get handed over from 5GS to UTRAN needs to be under consideration. One way involves the tight interworking between 5GC and EPC resulting in a handover from NR to E-UTRAN. One possible solution for such a handover is over the N26 reference point between the MME and AMF. This would ensure appropriate handover procedures between 5GC and EPC. Next, the call can then be handed over from the E-UTRAN to UTRAN over the already existing network infrastructure, ensuring connectivity between different radio access technologies. The reverse handover procedure from the UTRAN to 5GS also needs to happen to ensure seamless connectivity. Different issues relating to Single Radio Voice Call Continuity (SRVCC) are collected and documented along with the possible resolutions in [14]. The issues highlighted in [14] would help the suppliers design/modify their product or the network to target these preliminary issues.

The standardization of a technology usually involves study and preparation of a technical report before the issuing of a technical specification. Technical Reports (TR's) include potential issues and possible solutions to these issues along with the impact the solutions have on the existing elements. The TS provide a lot of information on different aspects of the technology and are constantly updated. There aren't a lot of scientific papers available that address IMS in 5G. These specifications therefore prove to be a great source of information. However, the knowledge is mostly theoretical. The implementation of these specifications are in the hands of the developer. While this allows flexibility to the network operators, it is a tedious task to go through a large number of these TS. It might also make it harder for beginners to understand the technology.

Ericsson, Huawei, Nokia are a few companies that are leading the evolution of 5G technology. Ericsson has issued a series of documents relating to 5G Voice [15] [16] [17] [18] [19]. These documents explain the network evolution aspects for 5G voice services. The series provides a very clear overview as to how 5G Voice can be realised. Voice will pave the path for other IM services to be implemented within the network. Implementation of 5G voice will therefore be an important milestone.

All the stages of the 5G network evolution must be capable of supporting voice for smartphones and other voice centric devices. It explains the different deployment options (explained in Chapter 5). The major challenge is to achieve seamless voice support during the interworking between 5G, 4G and 2G/3G. 5G networks will be able to provide limited coverage initially until they are completely rolled out. 4G VoLTE is already capable of interworking with the underlying 2G/3G network. Enabling the interworking between 5G and 4G will expand the area where voice services are provided to customers. The migration from 4G to 5G starts with a Non-Standalone Architecture (NSA) supporting NR and E-UTRAN support at the RAN and EPC at the core. This is termed as Voice over E-UTRAN NR Dual Connectivity (EN-DC). The operators are able to provide NR as the RAN whilst allowing the support of IMS services over E-UTRAN/4G. The bearer establishment for IMS services is being taken care of by the LTE radio access. This type of deployment in the transitional stage results in cost-effectiveness for the operators by reusing the existing 4G infrastructure until IMS signalling, voice and video over NR is introduced [15].

Voice over NR with Evolved Packet System (EPS) fallback (FB) , also called NSA Option 3 (explained in Chapter 5) provides seamless voice coverage across the whole network. 4G P-Gateway (P-Gw) function is split into Session Management Function (SMF) and User Plane Function (UPF) in the 5GS. It would include tight coupling of the Mobility Management Entity (MME) in EPS with the Access and Mobility Management Function (AMF) in the 5GS over the N26 interface. The Home Subscriber System (HSS)/ Unified Data Management (UDM) are combined. A smart phone's Radio Access Technology (RAT) is set to "NR Preferred" indicating that the UE is on NR by default. When a call is initiated, there is a handover from the 5GS to EPS without exceeding the voice gap KPI ensuring QoS. The reasons for the Fallback (FB) to LTE could be the lack of voice support over NR or temporary lack of radio resources in NR for voice [16]. More details on this is presented in Chapter 5. Voice over NR (VoNR) implies that the NR Standalone (SA) Option 2 (explained in Chapter 5) is considered. NR coverage is controlled by a 5G core (5GC) that supports voice. NR gNodeB which is the base station in 5G, will play a major role in the QoS flow establishment and might also initiate an inter-system handover to EPS later. This type of handover to EPS will also allow to perform 4G Single Radio Voice Call Continuity (SRVCC) if 4G deployment supports Circuit Switched (CS) coverage [17].

An important difference to note here is the difference in the terminology between fallback, handover and SRVCC. A fallback is triggered at the initiation of a call indicating a change in the RAT. A handover is triggered when a UE is already on a call, and moves to another location or moves out of coverage with no change in the anchor Mobile Switching Center (MSC) and the subscription. SRVCC means the fallback of a call to CS network. The SIP connection is moved to the CS network from Media Gateway Control Function (MGCF).

The recommendations for the deployment of 5G voice solutions in areas with different NR coverage situations, different subscriptions and different device generations is presented in [18]. A step-wise approach for the roll out of 5G voice utilising different parts of the spectrum in urban, suburban and rural scenarios is discussed. Mid and high-bands are available initially for deploying NR. Voice in EPS with EN-DC is recommended to tackle spotty coverage issues when using these bands. Coverage can be extended by deploying NR in low and mid-bands.

SMS and Emergency Calling would work over 5G access referred to as SMS over IP (SMSoIP) using IMS and SMS over 5G Non-Access Stratum (5G NAS) [19]. Both require a 5GS UE supporting voice, but the choice of which to be utilized is operator-preference. Emergency call in the case of EN-DC is supported over LTE connected to the EPC without any impact on the IMS. For 5GS, there are four possibilities for an emergency call procedure:

- UE Reselection: The User Equipment (UE) re-selects E-UTRAN or other access in the case of an emergency call attempt if there is no mandatory requirement for NR support.
- EPS fallback: Service Request for Emergency: This setup requires the support of VoNR. The UE uses a service-request for emergency towards the AMF which if supported, works with the gNB to perform the EPS fallback. There is no need for the NR and 5GC to support emergency services themselves but they must support the initiation of the service-request.
- EPS Fallback-QoS triggered for emergency: UE falls back to LTE during emergency when the QoS flow for voice is attempted to be established.
- Emergency Call over NR: This situation involves use of NR for emergency calls in combination with seamless mobility between 5GS and EPS using IRAT handover. A pre-requisite to this is the support of emergency features by the NR and the 5GC .

In conclusion, the first two methods mentioned use 4G for emergency call establishment. The last two solutions differ depending whether a radio bearer used for the voice media is established on NR and any regulatory requirement regarding PDN connection establishment, IMS registration and other procedures has to be fulfilled [19].

Huawei has also issued a technical white paper on Voice over 5G [20]. A detailed overview of the 5G voice evolution aspects are covered in the paper. Codec development, EPS Fall Back, VoNR are a few highlights of the paper. Notable are the sections that describe the possible paths this evolution in VoNR can take and how each of them will add to the technological advancement. The protocol stack involved in different 5G voice solutions that is VoNR, EPS Fallback, VoLTE and RAT Fallback, are also presented. This provides clarity on how the voice service is provided. The difference in terminology between Vo5G and VoNR is important. Deployment of voice services over NR with dependency on EPC is referred to as VoNR (NSA). Deploying the full fledged voice service over NR and involving the 5G core is referred to as Vo5G (SA).

Along the lines of providing coverage and handing over the calls between different radio technologies, [21] proposes efficient handover techniques that guarantee communication continuity and service quality which in turn determines the performance of the network. The paper points out the problems in the mobility management of the 5G networks and proposes Handover Decision and Handover Execution algorithms to effectively complete the handover process with guaranteed Quality of Service (QoS). In situations where the 5G network coverage is not fully available, there is a need to fall back to LTE (or even 3G) creating a need to have efficient handover algorithms that will be able to provide a certain level of quality assigned to the user based on the subscription. The challenge lies in an intensive cell deployment for the 5G networks and the ability to improve and optimize the handover process.

The white papers and various 3GPP specifications help get some clarity on EPS, IMS and 5GS. 5G voice implementation will pave the way for the implementation of other IMS services like (not limited to) instant messaging, presence, video conferencing and file sharing. The evaluation of some of these services (voice, video and instant messaging) in the existing NSA based network at Ericsson is the end goal of this thesis. Indicators like latency, packet loss and jitter are used to observe the performance of the services over the 5G network. The expected values for these are provided in [1]. The performance of the services is measured against these values and analyzed.

3

Evolution and Architecture

This chapter focuses on giving a high level description on the evolution of mobile communication technologies over the last couple of decades (Figure 3.1). Section 3.1-3.3 summarizes the different phases of evolution between 1G/2G/3G. Section 3.4 explains the 4G mobile network, its architecture and functional entities.

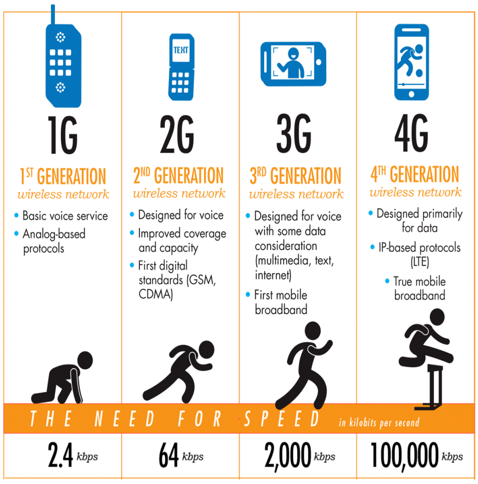

Figure 3.1: Mobile Network Evolution [22]

3.1. First Generation Mobile Networks (1G)

1G or the first generation of mobile networks refers to the network with the capability of making a simple voice call [23]. Figure 3.2 represents the network architecture of 1G. Advanced Mobile Phone System (AMPS) is one of the standards in 1G. It is an analog system and provided a voice only network operating on the 800MHz band. It utilises Frequency Division Multiple Access (FDMA) to provide communication to a very limited number of subscribers over distances almost greater than 40kms. The frequency spectrum is divided and each user is allocated a separate channel for the duration of the call. The Mobile Switching Telephone Office (MTSO) manages the routing of the phone calls. It contains the Mobile Switching Center (MSC) which contains the switching circuitry. The MSC enables connectivity to the Public Switched Telephony Network (PSTN) which is traditionally circuit-switched (CS). This PSTN then redirects the call to the corresponding MSC, to the MTSO and finally, to the mobile station (MS) at the destination [24] [25].

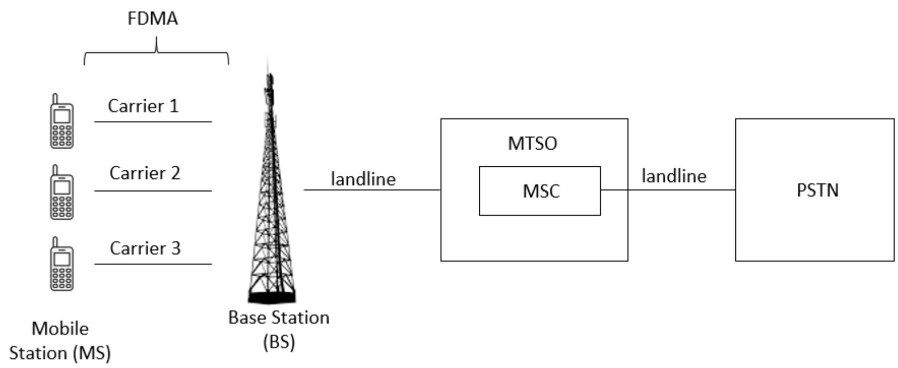

Figure 3.2: High level overview of 1G - AMPS

3.2. Second Generation Mobile Networks (2G)

One of the standards in 2G is Global System for Mobile (GSM) Communications. GSM proved to be a catalyst in the evolution of the mobile networks. Communication occurs between the same or different Public Land Mobile Networks (PLMN). A PLMN is provided by a mobile operator (for e.g. KPN in the Netherlands). It comprises of the Radio Access Network (RAN) and the Core Network (CN). A number of functional entities make communication possible and they are depicted in Figure 3.3. A Mobile Station (MS) is the combination of a Universal Integrated Circuit Card (UICC) with a Subscriber Identity Module (SIM) application and a mobile equipment. The SIM allows the network to identify the subscriber. MS is connected over air to the RAN consisting of the Base Transceiver Station (BTS) and the Base Station Controller (BSC). The CN consists of a Mobile Switching Center (MSC). The Home Location Register (HLR) contains the user profiles and subscriptions and permits or denies the attachment of a MS to the home network. If a subscriber has moved to a PLMN other than the one he initially registered at, it is referred to as a Visiting PLMN (VPLMN). The subscriber will attach to an MSC in a foreign PLMN. The subscriber information is then temporarily stored in a Visitor Location Register (VLR). VLR is integrated with the MSC. The VLR will retrieve the subscriber information from the HLR. An Authentication Center (AuC) provides the authentication of the SIM card of a subscriber connected to the network. Gateway MSC (GMSC) handles the connectivity to/from networks that are either within or external to the home network by routing the calls to the MSC to which the subscriber is attached. 2G

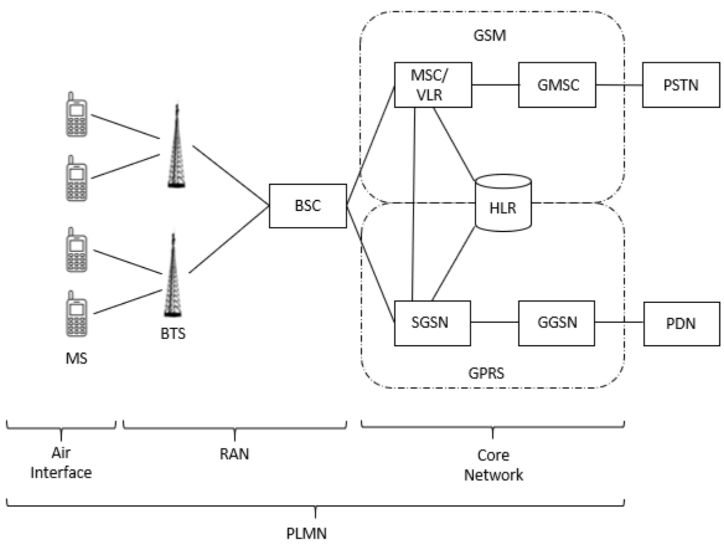

Figure 3.3: High level overview of 2G - GSM/GPRS architecture [26]

uses Gaussian Minimum Shift Keying (GMSK) with Time Division Multiple Access (TDMA) signalling over Frequency Division Duplex (FDD) carriers. While the architecture remains the same, the data rates improved and High Speed Circuit Switched Data (HSCSD) was introduced.

The evolution from 2G to 3G happened in different steps. General Packet Radio Service (GPRS) was introduced which used aggregation of carriers, provided high speed packet-switched applications like always-on internet access. GPRS is also referred to as 2.5G. GPRS marked the point when the technology shifted from Circuit Switched (CS) to Packet Switched (PS). The architecture includes nodes like Serving GPRS Support Node (SGSN) and Gateway GPRS Support Node (GGSN). GGSN provides IP connectivity to the Packet Data Networks (PDN). PDN here refers to a network providing internet services to the UE. SGSN and MSC work together in order to provide CS voice and PS data services to the users. At this point, the MS is attached to both CS and PS services. Taking this architecture one step further, Enhanced Data Rate for Global Evolution (EDGE) was introduced. It used a different modulation technique to provide three times more data rate than GPRS in the same bandwidth [27].

3.3. Third Generation Mobile Networks (3G)

3G or the third generation of mobile networks proved to be an upgrade in terms of network and data transmission. Feature phones were equipped with faster communication in terms of larger capacity for emails, faster web browsing, video streaming and security. The 3G network is called Universal Mobile Telecommunication Systems (UMTS) and was standardised by 3GPP. The radio network is referred to as Universal Terrestrial Radio Access (UTRA). UMTS offered greater spectral efficiency over GSM with two modes: Frequency Division Duplex (FDD) and Time Division Duplex (TDD). Code Division Multiple Access (CDMA) is used in both modes. FDD CDMA is also referred to as Wideband CDMA (W-CDMA). Different requirements were set depending on different use cases like Rural Outdoor, Suburban Outdoor and Indoor/Low Range outdoor. This type of a distinction based on the application is also followed in the 5G networks. UMTS and other technologies that followed were always ensured to be backward compatible. This ensured inter-operability between older and newer technologies[28]. The architecture of a UMTS network is depicted in Figure 3.4. The overall radio access network i.e Universal

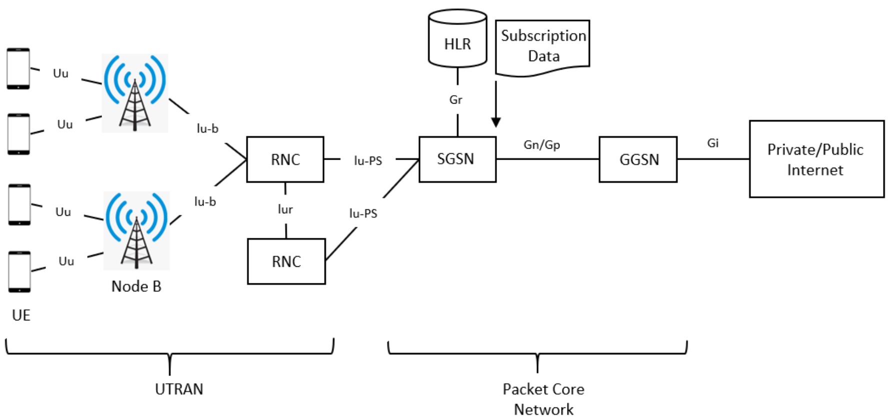

Figure 3.4: 3G - UMTS/UTRAN [29]

Terrestrial Radio Access Network (UTRAN) consists of User Equipment (UE), group of NodeB's and Radio Network Controllers (RNC). UE is a 3G network terminology that refers to either a mobile phone with a UICC and SIM (as mentioned in GSM and referred to as USIM in 3G) or any other terminal device that can connect to the network. NodeB in 3G is the equivalent of BTS in 2G. RNC in 3G is the aggregation point for NodeB's and is the equivalent of BSC in 2G.

The core network elements include SGSN and the GGSN, which are similar to the 2G network. SGSN is responsible for access control, security functions and location tracking of UE. HLR has the same function as in GSM architecture. GGSN also handles billing functions and carries out filtering and firewall protection.

3GPP Release 99 is the first standard for UMTS/W-CDMA. Subsequent releases presented improved data rates eventually leading up to the development of 4G. 3GPP Release 5 (in 2002) introduced High Speed Download Packet Access (HSDPA) and 3GPP Release 6 (in 2004-2005) introduced High Speed Uplink Packet Access (HSUPA) both of which led to the development of High Speed Packet Access (HSPA) and further of HSPA+. New concepts like Multiple Input Multiple Output (MIMO) antennas, multiple carriers and modulation techniques continued to evolve and provide improved performance over their predecessors. Continuous evolution of UMTS resulted in the development of the Fourth generation of mobile networks [30].

3.4. Fourth Generation Mobile Networks (4G)

4G or the fourth-generation mobile networks uses Internet Protocol (IP) completely and there is no CS anymore. This technology focused on solving high demand for data rates by services like gaming, streaming and web browsing. The aim to reduce the delay and latency in network services while increasing the spectral efficiency of the network led to the development of a project called Long Term Evolution (LTE) which involved the study of a new air interface called Evolved-UTRA (E-UTRA). E-UTRA after evolving from UMTS became a new radio access technology. UMTS focussed on CS and PS data while LTE is purely PS system. LTE on one hand, is similar to UMTS in some aspects. On the other it offers significant improvement over UMTS. LTE again was designed to be backward compatible. As a result of which, it is inter-operable with the previous generations of mobile networks. This meant that the users could move out of a LTE coverage area and still continue to receive service. 3GPP continued to standardise the technologies and LTE-Advanced (LTE-A) came into development. Next in the evolution was LTE Advanced Pro. However, the true 4G network was considered as LTE-Advanced, a successor to LTE. The existing LTE-A network is described by the architectural diagram shown below in figure 3.5.

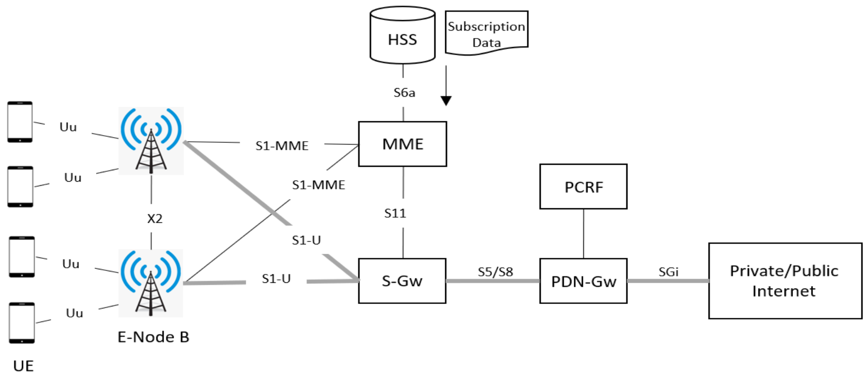

Figure 3.5: EPS Architecture [31]

LTE/E-UTRAN and the Enhanced Packet Core (EPC) are together referred to as Enhanced Packet System (EPS). EPC consists of different elements like MME, S-Gw, PDN-Gw and HSS. Each element of the architecture and its function is listed as follows [32] [33] :

- User Equipment (UE): The device with the user. The Uu interface connecting the UE to the E-NodeB in Figure 3.5 is the LTE version of the interface mentioned previously in 3G.
- Enhanced-NodeB (E-NodeB/eNB) : This is the 4G base station and provides the user with radio interfaces and performs Radio Resource Management, header compression, security modulation, interleaving, handover and re-transmission control. eNB's are connected together through an interface called the X2 interface. There are two S1 reference points: S1-MME and S1-U. S1-MME carries all the control plane signaling information. S1-U carries the payload (data plane) information.
- Mobility Management Entity (MME) : E-NodeB's connect to the MME as shown in Figure 3.5. MME is the mobility server and acts as a registrar for access to the core network. It operates in the control plane and is responsible for sending data to the eNB's. MME also helps with bearer establishment procedures. It also performs user profile authentication in association with the Home Subscriber System (HSS).

- Serving-Gateway (S-Gw) : This acts as the terminating interface towards the E-UTRAN. The control plane signaling traverses through the S1-MME reference point from the eNB to the MME. The user plane payload traverses from the eNB to the S-Gw over the S1-U reference point bypassing the MME. S-Gw serves as a local mobility anchor point of data connections for inter-eNB handover and handles user data tunnels between the eNBs and PDN-Gw which is explained next.
- Packet Data Network-Gateway (PDN-Gw/P-Gw) : A PDN refers to the network that is established and operated to provide data transmission services to the public (like the Internet). P-Gw provides UE with access to a PDN by assigning an IP address from the address space of the PDN. It serves as the mobility anchor point for handover between 3GPP and non-3GPP and also performs policy enforcement, packet filtering and charging based on the rules set by the policy and charging functions. PDN-Gw acts as the gateway router to the internet.
- Home Subscriber System (HSS) : It is the central database which stores user profiles, and provides user authentication information and user profiles to the MME. It is similar to the HLR in the GSM architecture. HSS also keeps a track of the location information of the subscribers allowing them to connect to the nearest eNB.
- Policy Charging Control Function (PCRF) : This node determines the policy rules for the data bearers. It supports service data flow detection, policy enforcement and flow-based charging in a centralized manner.

4
IP Multimedia Subsystems

The IP Multimedia Subsystem (IMS) is an architectural framework that allows access to IP multimedia services from any terminal. It has been standardised by 3GPP. The IP Multimedia Subsystem consists of two parts- IP Connectivity Access Networks (IP-CAN) and IM Core Network (CN) subsystem. IP-CAN consists of mobile core network and the associated radio access networks (for e.g. EPS, GPRS with UTRAN). The IP Multimedia Core Network (CN) subsystem consists of all the CN elements that are necessary to provide these multimedia services to the subscribers. These multimedia services include voice, video, messaging, data and web technologies for both wireline and wireless users. Internet Protocol (IP) is the transport protocol used to access the IMS network [10].

The goal of developing IMS was to standardise the way different multimedia and voice applications are accessed from the wireline or wireless terminals, across the industry. IMS is able to interact with the legacy network, handing over calls in both directions as per requirement. IMS caters to two kinds of requirements in the market: first one, provisioning of standardized services that allowed two users to connect without any restrictions from the operators that both have. Second one, providing innovative and differentiating services to the user. IMS is capable of guaranteeing a certain Quality of Service (QoS) when using its services. Session Initiation Protocol (SIP) is used as the core signaling protocol. SIP is utilised in different network elements to fulfil different roles regarding service triggering, authentication and authorization, media plane resources invocation and other essential network properties [34].

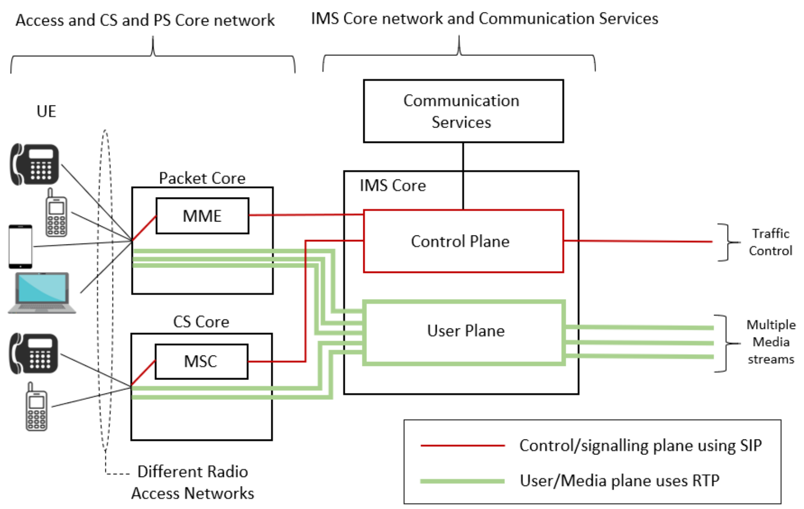

Figure 4.1: IMS Control and User plane separation

IMS has a flat architecture (presented in Figure 4.1) and allows for separate end-to-end control plane and user plane. IMS ensures a strict demarcation between the Access network, Packet core network and the Multimedia Communications Framework. IMS provides great flexibility in terms of delivery of application services. It is device-independent.

As shown in the Figure 4.1, the control plane and the user plane are separated when accessing IMS services. It is also important to note that, SIP messages are transmitted over the control plane (in red) through the MME/MSC. The media stream (like voice or video) uses Real-Time Protocol (RTP) and is transmitted over the user plane (in green). The media payload traverses from the UE through the Public Data Network Gateway (PDN-GW) and directly to the corresponding gateway in the IMS core. The payload does not pass through the Mobility Management Entity (MME) or the Mobile Switching Center (MSC). More details on this is presented in Section 4.2. IMS core (Call Session Control Function (CSCF)) is responsible for handling the connectivity between two users. The range of services offered by IMS includes (not limited to) video conferencing, file sharing, instant messaging, presence and SMS. The IMS CN is on the session/control layer while the communication services offered are on the application layer. Depending on the requests, the corresponding communication service is invoked through the SIP-Application server (SIP-AS) and the user is served. The role of AS is explained in Section 4.1.

The biggest advantage offered by IMS is the QoS guarantees when using any of its services. In case of 4G, these benefits worked in favour of the voice services. As an extension, similar concepts have been applied for the development of services for 5G networks as well which is explained in Chapter 5.

Section 4.1 describes the different elements in IMS CN. Section 4.2 explains how EPC and IMS are connected to serve users and Section 4.4 explains the Voice over LTE (VoLTE) originating and terminating procedure. Section 4.5 describes what is QoS and how it is measured.

4.1. IMS Architecture

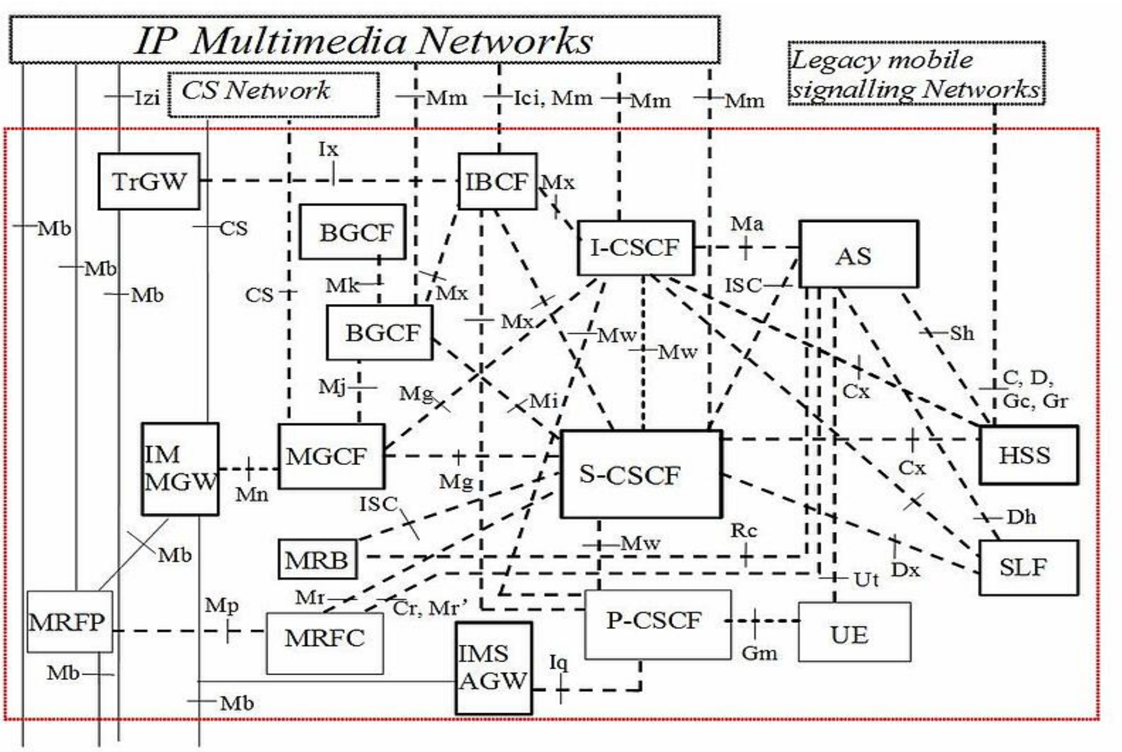

Figure 4.2: IMS Core architecture [10]

IMS CN consists of different functional entities that work together to support IMS services for a given user. However, it is important to note that all these functional entities/nodes are logical in nature. There will be multiple copies of these nodes implmented in a real network. For the sake of simplicity, it is assumed that there is just one copy of these nodes in the IMS core. The different nodes and their functional description is listed as follows [35]:

- Call Session Control Function (CSCF): Depending on the task at hand, CSCF takes on different roles. Logically it is split into the Proxy-CSCF (P-CSCF), Serving-CSCF (S-CSCF) and the Interrogating CSCF (I-CSCF).
  - P-CSCF is the first point of contact for the UE within the IMS. It is the signaling proxy between the UE and the IMS CN.
  - S-CSCF is the registrar. S-CSCF has the IP address of the subscribers upon registration in the IMS. Session origination and termination is handled at the S-CSCF.
  - I-CSCF is mainly the contact point within an operator's network for all IMS connections destined to a subscriber of that network operator or a roaming subscriber currently located within that network operator's service area.

  Apart from these there is also a possibility to provision an Emergency-CSCF (E-CSCF) which handles the emergency sessions and routes the emergency requests to the right emergency center.

- Home Subscriber System (HSS): The Figure 4.3 shows an HSS associated with the EPS (EPS-HSS) and another associated with the IMS (IMS-HSS). The EPS-HSS along with the Authentication Center (AuC) allows for EPS attachment of the UE. In a similar way, the UE is required to be registered for IMS services. The authentication for IMS registration through the IP-CAN and the P/S-CSCF is done by the IMS-HSS. AuC uses challenge and response to exchange the authentication credentials between both HSS. In some deployments, the HSS of both EPS and the IMS can be combined and used for authentication.

- Subscriber Locator Function (SLF): There maybe multiple HSS deployed in an IMS network. When the need arises, it is important to be able to locate the subscriber information in the correct HSS. This process is assisted by the SLF which contains a record of all the IMS subscribers in the network along with a pointer to the HSS where the subscriber information is stored.

- IMS-Media Gateway (MGW): MGW terminates bearer channels from a CS network and media streams from a packet network. It supports media conversion, bearer control and payload processing. It also interacts with the MGCF (explained later) for resource control, owns and handles resources such as echo cancellers and may need to have codecs.

- IMS Access Gateway (IMS AGW): IMS network is accessed through an IP infrastructure. The SIP signaling will traverse through the IMS-Application Level Gateway (IMS-ALG) in the P-CSCF. The media entity will traverse through the IMS-AGW towards the remote party. In short, IMS-AGW the point of entry in the IMS network for the user plane data when the IMS is accessed through an IP access network.

- Breakout Gateway Control Function (BGCF): BGCF determines the next hop for routing the SIP message. It also determines the external network in which PSTN/CS Domain breakout is to occur or selects the corresponding MGCF (explained later) within the network.

- Interconnection Border Control Function (IBCF): An IBCF provides application specific functions at the SIP/SDP protocol to perform interconnection between two operator domains. IPv6 and IPv4 SIP applications, network topology hiding, controlling transport plane functions, screening of SIP signaling information, selecting the appropriate signaling interconnect and generation of charging data records are some of the functions performed by the IBCF.

- Transition Gateway (Tr-GW): This is located within the media path and controlled by an IBCF and provides functions like network address/port translation and Ipv4/IPv6 protocol translation. Tr-GW is also referred to as the Network-SBG (N-SBG).

- Media Gateway Control Function (MGCF): It controls part of the call state that pertains to connection control for media channels in an MGW. MGCF also communicates with CSCF, BGCF and circuit switched network entities. It is also responsible for determining the next hop depending on the routing number for incoming calls from legacy networks. Protocol conversion between PSTN/CS network protocols and the IMS call control protocols is taken care of by the MGCF. It may also forward out of band information received in MGCF to the CSCF/MGW.
- Media Resource Broker (MRB): The MRB supports the sharing of a pool of heterogeneous MRF resources by multiple heterogeneous applications. MRB assigns specific suitable MRF resources to calls as requested by the consuming applications, based on MRF attributes.
- Multimedia Resource Function Controller (MRFC): It controls the media stream resources in the MRFP (explained next), interprets the information coming from an AS, S-CSCF and control MRFP and then it generates Call Detail Records (CDRs).
- Multimedia Resource Function Processor (MRFP): It controls the bearers on the Mb reference point, provides resources to be controlled by the MRFC, mixes incoming streams, sources media streams, processes media streams and performs floor control.
- Application Server (AS): SIP AS, Open Service Access (OSA AS) or Customized Applications for Mobile Networks Enhanced Logic (CAMEL) IM-SSF (IM-Servoce Switching Function) offers value added IM services. It may reside in the user's home network or in a third party (network or stand-alone AS) location. AS may impact SIP session on behalf of the services supported by the operator's network and may also host and execute services.

Apart from the above mentioned entities, there is an Access Session Border Gateway (A-SBG) which is not shown in the Figure 4.2. A-SBG allows IMS network access from wireline terminals. It consists of the Session Border Controller (SBC) on the control plane and a Session Gateway (SG) on the user plane. Every link or connection between the IMS entities are referenced by different reference points. All the reference points presented in Figure 4.2 are explained in detail in [35].

4.2. IMS architecture in Voice over LTE (VoLTE)

IMS provides voice services, referred to as VoLTE, over the IP network. If a video service is offered, it is referred to as Video Over LTE (ViLTE). A VoLTE-enabled UE, EPS and IMS Core are the different parts of the network involved in a VoLTE session. The different elements and their connectivity is presented in the Figure 4.3.

A UE connects to the EPS using a Universal Integrated Circuit Card (UICC) which has a Subscriber Identity Module (SIM) application running on top of it (also explained in Section 3.2). The SIM application contains the International Mobile Subscriber Identity Number (IMSI) and an authentication key. The UICC and SIM are together referred to as Universal Subscriber Identity Module (USIM). Through the USIM, the network is able to identify the registered subscriber on its network. Similarly, UE requires an identity when accessing the IMS network. This is supported by another application running on the UICC called the IMS Subscriber Identity Module (ISIM). This contains the authentication key for authenticating an IMS subscriber. Using USIM and ISIM the UE is able to attach to EPS and further register itself on the IMS network.

The EPS provides the IP connectivity between the UE and IMS as shown in the Figure 4.3. When the UE is switched on, a default bearer (in grey) is set up for internet access with QoS Class Identifier (QCI) value 9 (more on QCI in Section 4.5). When a VoLTE service is requested by the UE, another Default Bearer with QCI value 5 is set up for IMS signaling (in red). The SIP messages are exchanged over this default bearer in the control plane. The EPC simply forwards the SIP messages back and forth between the UE and P-CSCF. PDN-GW is responsible for exchanging data between the UE and the external networks over the SGi interface. SIP messages from the UE follow the same path through the PDN-GW and are forwarded to the P-CSCF. PCRF triggers the establishment of the dedicated bearer upon receiving an indication from the P-CSCF. PDN-GW is then triggered by the PCRF for bearer establishment for voice and video call with QCI values 1 and 2 respectively. Control plane primarily uses SIP while the user plane uses protocols like Real Time Protocol (RTP) or Real Time Streaming Protocol(RTSP).

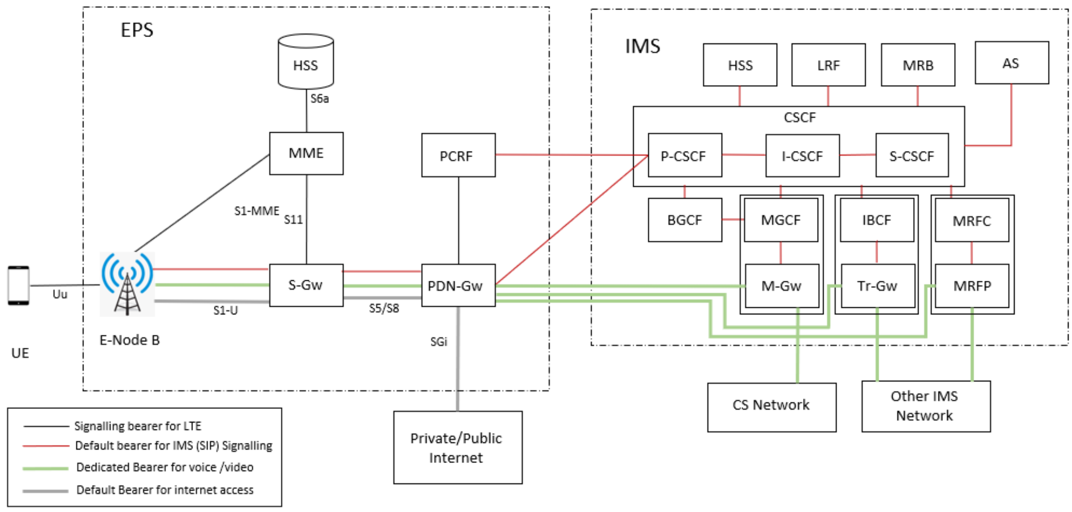

Figure 4.3: IMS in LTE Network [36]

The AS will provide the requested service to the UE. Examples of AS are Multimedia Telephony AS (MMTel-AS) which enables telephony services (voice and video)and the Service Centralisation and continuation AS (SCC-AS) which enables access mobility (LTE, WiFi and CS). Upon receiving a request from the UE, the P-CSCF will forward the data to the AS which contains the information about the services enabled for the subscriber. XCAP is the protocol used for the communication between the AS and the UE.

When the VoLTE call is directed towards another user, the control plane determines the location and the path to reach the terminating user. Once this is discovered, the voice bearer will be established between the originating and the terminating users through the SBG. If the UE requests a video service, there will be two bearers set up. A dedicated bearer each for voice and video. Depending on the service requested, additional bearers will be established when needed.

The detailed end-to-end VoLTE call procedure is explained in detail in Section 4.4. This will give a clear idea on how the VoLTE call is established which will in turn help understand the call procedure in case of 5G systems.

4.3. Roaming

A UE requesting a service is termed as the originating UE and the network associated with the UE is called the Originating Network. The UE which receives the request from the first UE through the CN is referred to as the terminating UE and the network associated with the terminating UE is called as the Terminating Network. After the UE is attached to a UMTS network, a default bearer is setup for the session and an IP address is allocated to the UE. This bearer is referred to as the Packet Data Protocol (PDP) Context or Bearer Context. The bearer setup in EPS is explained briefly in Section 4.2. Considering the VoLTE call scenario, the UE is EPS attached, the default bearer for IMS signaling is established and the dedicated bearer has been setup. There are different ways in which the VoLTE call can be terminated. It is also referred to as the breakout of the call either to the internet/ to CS networks/ to other IMS networks.

There are two kinds of networks depending on the access network a UE is attached to. As explained in Section 3.2, Home Public Land Mobile Networks (HPLMN) is where the subscriber is registered. Visitor Public Land Mobile Networks (VPLMN) refers to the network that the user is attached to when he/she is roaming outside the home network.

There are two ways in which IMS services are provided when the UE is attached to a VPLMN [37]:

- Local Break Out – VPLMN Routing (LBO-VR): The UE is attached to the VPLMN and the corresponding P-CSCF in the VPLMN connects the UE to the IMS in the home network. The VPLMN IMS and the HPLMN IMS are connected through the Border Gateway (B-GW). When the UE is in a VPLMN and makes a request, the PDP/bearer context is established between the UE and the P-GW/GGSN through the S-GW/SGSN. The P-CSCF in the VPLMN will control the session. As a result, the UE is served by the IMS in the VPLMN and not the IMS in HPLMN. The Figure 4.4 presents the pictorial representation of the LBO-VPLMN routing.

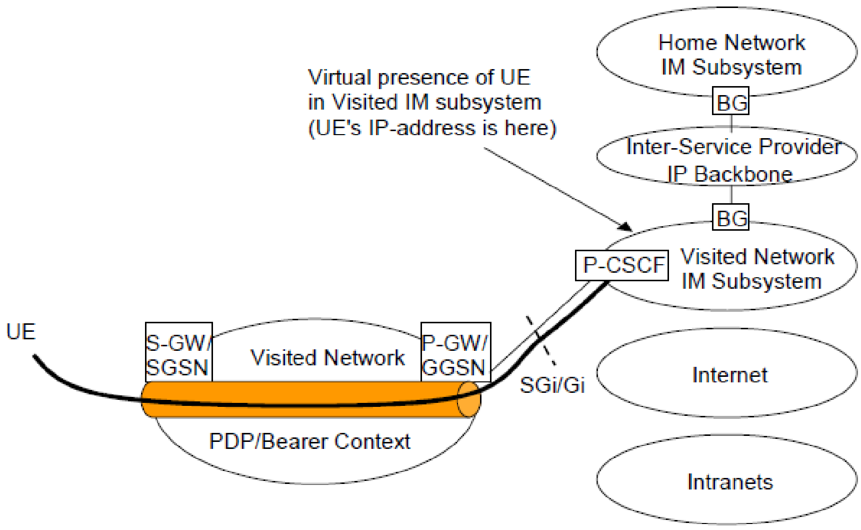

Figure 4.4: Local Break Out [37]

- S8 Home Routing (S8HR): S8 refers to the interface between the S-GW and the P-GW in case of roaming networks. The VPLMN IMS does not serve the UE in VPLMN. The PDP/bearer context is established between the UE in the VPLMN and the VPLMN B-GW passing through the S-GW/SGSN. The B-GW's of the VPLMN and HPLMN are connected over the IP backbone. The HPLMN B-GW then establishes the connection to the IMS CN in the HPLMN through the P-GW/SGSN. In short, S8HR refers to the procedure in which the HPLMN IMS is serving a UE in the VPLMN. The Figure 4.5 is a pictorial representation of S8HR [37].

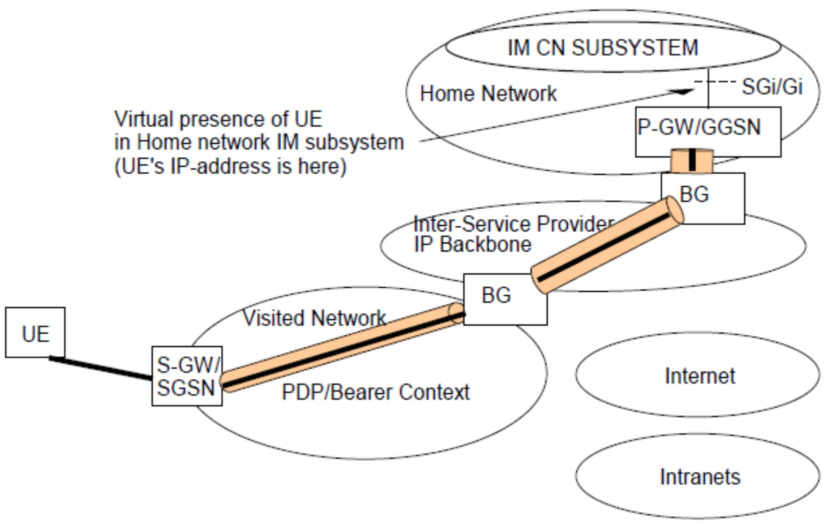

Figure 4.5: S8 Home Routing [37]

4.4. VoLTE Call Procedure

An end-to-end VoLTE session requires different procedures for attachment and registration, detachment and deregistration, call originating procedure and call terminating procedure. These different procedures are briefly explained in the following sections.

4.4.1. VoLTE EPS Attach and Registration

IMS services are made available to the subscriber in two steps. EPS attachment of the UE and then IMS Registration. As explained briefly in Section 4.2, when the UE attaches to the EPS a default bearer is established. When the UE requests access to IMS, it will need to register first. Registration ensures the establishment of a communication session using a Public User Identity (PUI). A SIP registrar creates and maintains a binding which is later used to forward the session to the IP address of the subscriber. The IMS subscriber data which is provisioned in the HSS has two items used for identification [34] :

- IMS Public User Identity (IMPU): IMPU identifies the IMS subscriber in a public network when establishing a communication session with that subscriber. IMPU is also the identity of the session originator. Universal Resource Identifier (URI) is the format used for IMPU. URI constitutes a string of characters that uniquely identify a source. Some examples of IMPU are listed as follows:
  - sip:alice.finch@xyz.se
  - tel:+31123456789
  - sip:+31123456789@xyz.nl
  - sip:121110987654321@ims.xyz001.xyy163.abcnetwork.org

  The format *sip:+31123456789@xyz.nl* is later used for the experiment in Chapter 7 because the IMS network was in a remote location and had to be accessed through the public network.

- IMS Private User Identity (IMPI): A subscriber's service profile and authentication method is determined by IMPI along with IMPU. Some examples for IMPI are listed as follows:
  - alice.finch@xyz.se
  - +31123456789@xyz.nl
  - 121110987654321@ims.xyz001.xyy163.abcnetwork.org

Registration procedure involves authorization and authentication of the subscriber. Authorization is performed by HSS while authentication is carried out by HSS and S-CSCF together. Authorization identifies the users that can use the services they are subscribed to. Detailed information about the registration process is provided in [10]. After the IMS registration procedure, additional bearers will be setup for IMS signaling and data. The detailed procedure for VoLTE attach is defined in [38] and the detailed procedure for IMS registration is provided in [10].A message sequence diagram of the VoLTE attach and IMS registration is provided in Appendix A.

4.4.2. VoLTE UE Initiated Detach and De-Registration

The VoLTE UE will automatically deregister itself from IMS by EPS Detach. Detach means that the UE is no longer attached to neither the network nor to the bearer that was previously allocated for it. Switching the UE off will result in a detachment from the network. De-registration occurs before detachment procedure. IMS de-registration means that no SIP signaling is being supported by the EPS. The binding for the UE is removed by the S-CSCF. HSS then removes the S-CSCF address from the subscriber's record. The P-CSCF will also discard the subscriber data [34]. De-registration procedure is defined in detail in [39] and VoLTE Detach procedure is defined in [10]. A detailed message sequence diagram of IMS de-registration and EPS detach is provided in Appendix A.

4.4.3. End-to-end Call Procedure between two UE's

A VoLTE session between two registered users is supported by SIP. The session establishment between the two UE's is a series of exchanging transactions. Each transaction has a request and a response. Depending on which UE is requesting and which UE is responding, the UE take the role of a client or a server. A SIP message contains a Session Description Protocol (SDP) offer within itself. SDP conveys information about the media capabilities and formats supported between the participants in a multimedia session. Based on the information provided by SDP, the users agree on a set of codecs and capabilities for the session. The session flow procedures vary depending on whether it is a Mobile Originating (MO), a Mobile Terminating session (MT), PSTN-Originating or Terminating (PSTN-O/T), Non-IMS Originating or Terminating (NI-O/T) and AS-Originating or Terminating (AS-O/T). These procedures further differ depending on whether both users are located in the home network or whether one of them is roaming. The different procedures for session flows are listed in [10]. For the sake of this thesis, the focus lies on MO and MT, both in the home network. The procedure for the same is listed in the next section.

Mobile Origination Procedure in the home network

It is assumed that the UE in the originating network (UE-O) is already attached to the EPS and registered in the IMS network. Once these conditions are satisfied, Figure 4.6 presents the message sequence on the originating side of the network. For the sake of simplicity, the EPC nodes are not shown in the figure.

- 1 The SIP Invite contains the SDP offer that is sent from UE-O to the P-CSCF. It contains different parameters for the session like the codecs to be used and the QoS requirements.
- 2 P-CSCF receives the SIP Invite and verifies that the user is registered. It then forwards the SIP Invite to the S-CSCF whose address has been stored during the Registration procedure.
- 3-4 The HSS sends the subscription information to the S-CSCF after registration which contains an Initial Filter Criteria (IFC). The IFC contains a list of one or more Service Point Triggers (SPT). An SPT matches a request and the corresponding AS that will serve the session. Depending on the service request, the S-CSCF will forward the SIP Invite to the corresponding AS. The AS processes the request and returns the Invite back to the S-CSCF.
- 5 The SIP Invite is forwarded to its next hop. This could be either an I-CSCF in the same network or in another network.
- 6-7 SIP 183 session progress message is sent to the S-CSCF and forwarded to the P-CSCF.
- 8-9 P-CSCF requests authorization from the PCRF (not shown in the figure) to reserve radio resources. After receiving a response from the PCRF, UE-O reserves the resources required for the media session.
- 10 SIP 183 Progress message is forwarded from the P-CSCF to UE-O.
- 11-15 The UE-O sends a SIP PRACK message in response to the SIP 183 Progress message.
- 16-18 The originating network receives a SIP 200 OK message in response to the SIP PRACK message.
- 19-22 After the resources are reserved, the UE-O sends a SIP Update message to the terminating UE (UE-T).
- 23-26 A SIP 200 OK message is received in response to the SIP Update message. This is also an indication that the resources are available at UE-T.
- 27-30 The SIP 180 Ringing response is received at UE-O and the user is alerted with a ringtone.
- 31-33 A SIP 200 OK message is received at the S-CSCF and P-CSCF.
- 34 The P-CSCF will now enable the bidirectional media flow between both UE's. A SIP 200 OK message is then sent to UE-O.

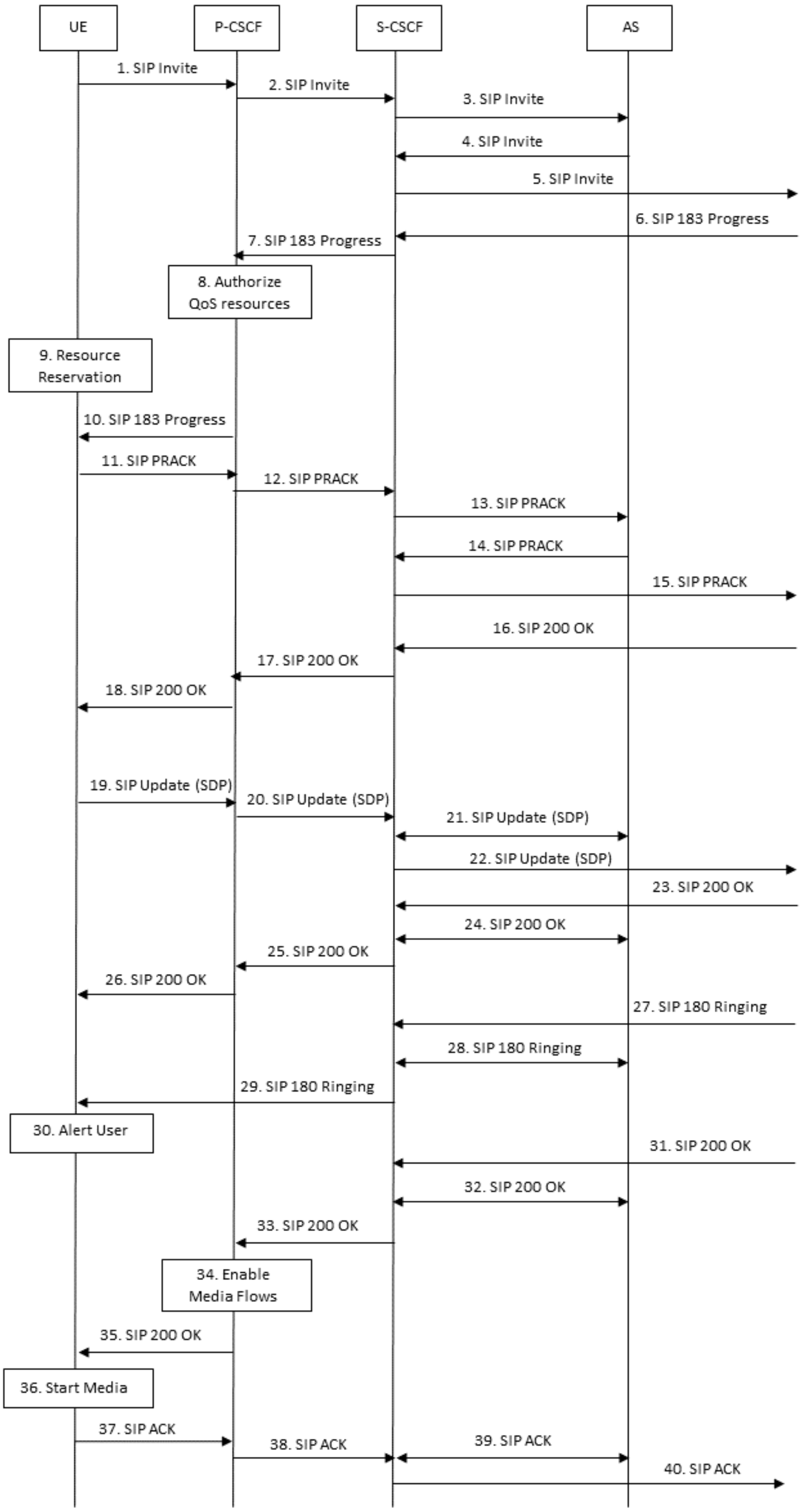

Figure 4.6: Mobile Origination Procedure [10]

- 35 UE-O starts the media transmission over the established bearers, which is voice in this case. The media is transmitted over RTP.
- 37-40 A SIP ACK message is sent by UE-O towards UE-T.

Mobile Terminating in the home network
The UE in the terminating network (UE-T) is also assumed to be EPS attached and IMS registered. Figure 4.7 presents the message sequence on the terminating end. Again, for the sake of simplicity, the EPC components are omitted in the figure. The detailed sequence can be found in [10].

- 1 UE-O sends a SIP Invite containing the SDP offer towards the S-SCSCF of the UE-T.
- 2-5 S-CSCF authorizes the requested SDP depending on the subscription of the user. It further applies IFC and triggers the AS depending on the service requested.
- 6 S-CSCF forwards the SIP Invite to the P-CSCF whose address is stored during the Registration procedure.
- 7 The P-CSCF forwards the SIP Invite to the UE. P-CSCF also forwards the session information to the PCRF (not shown in the figure) to invoke any policies that are applicable to the session.
- 8 The UE-T sends a SIP 183 Progress to the P-CSCF in response to the SIP Invite. The media supported by the UE-T is sent to the UE-O.
- 9 The P-CSCF interacts with the PCRF to get authorization for the resources required for the session.
- 10-12 P-CSCF then forwards the SIP 183 Progress message to the S-CSCF. The S-CSCF sends the message to AS and receives it back. After this, the message is sent towards UE-O.
- 13-16 A SIP PRACK message is received from UE-O in response to the SIP 183 Progress. This SIP PRACK is received first by the S-CSCF which sends it to the AS and receives it back. S-CSCF then forwards it to the P-CSCF which in turn forwards it to the UE.
- 17 UE-T responds to the SIP PRACK with a SIP 200 OK sent to the P-CSCF.
- 18 UE-T further proceeds to reserve radio resources to support the session.
- 19-21 P-CSCF forwards the SIP 200 OK to the S-CSCF and towards the UE-O.
- 22-25 SIP Update message is received from the UE-O at the S-CSCF which is then sent to the AS. After receiving the SIP Update from AS, the message is forwarded to P-CSCF and further to the UE-T.
- 26 UE-T is alerted about the incoming session.
- 27-31 UE-T sends a SIP 200 OK message towards UE-O as a response to the SIP Update message.
- 32-35 SIP 180 Ringing message is sent from UE-T towards UE-O.
- 36 UE-T sends a SIP 200 OK to the P-CSCF after the user has answered.
- 37 Bidirectional media flows are now enabled for the session on the UE-T end as well.
- 38 UE-T starts the media flow for the session.
- 39-41 P-CSCF forwards a SIP 200 OK to the S-CSCF and on to the UE-O confirming the establishment of the media flows.
- 42-45 A SIP ACK is received from UE-O as an acknowledgment to receiving the SIP 200 OK message.

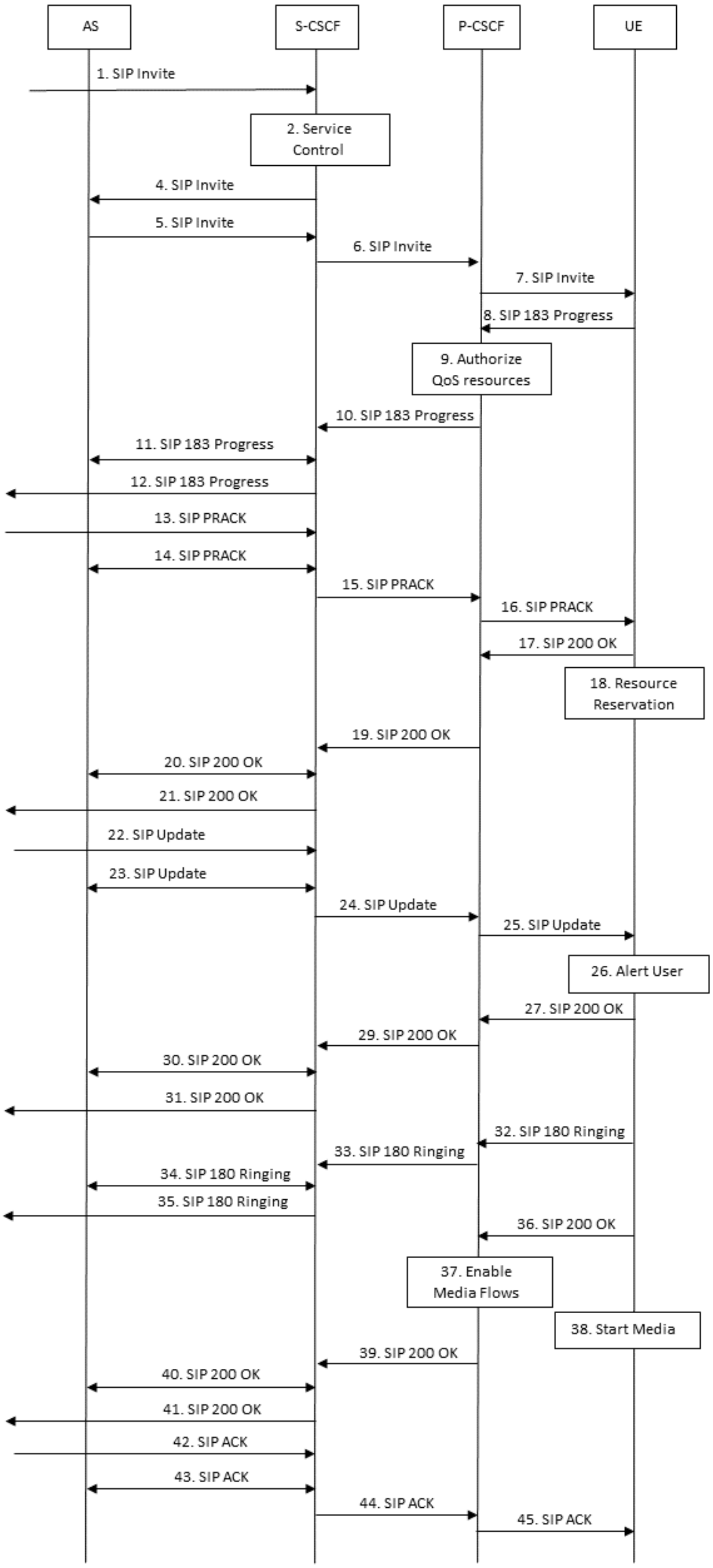

Figure 4.7: Mobile Termination Procedure [10]

4.5. Quality of Service- QoS

IMS was developed to provide a common framework for accessing different communication services. IMS can guarantee certain level of quality when providing these services to the users. This section will explain the different functional entities that take part in mapping the Quality of Service (QoS) parameters. The functions used to map the QoS parameters are present at the AS, PCRF, Policy and Charging Enforcement Function (PCEF) and the UE. The functions convert the QoS parameters from one format to other along the path of the call. The establishment/modification of the IP-Connectivity Access Network (IP-CAN) bearers can be either UE-initiated or network-initiated. Both types of requests are sent to the PCEF for authorization [38].

The connectivity between the UE and an external PLMN network is referred to as the PDN Connectivity service. E-UTRAN is connected to the EPC by an EPS bearer. This bearer provides IP connectivity between the S-GW and the P-GW. The EPS bearer is responsible for the QoS control in the EPC/E-UTRAN. Traffic mapped to the same bearer receive the same bearer level packet forwarding treatment. So, if a bearer needs to handle packets differently, separate bearers need to be provided.

When the UE connects to a PDN, an EPS bearer is established which continues to exist throughout the lifetime of the PDN Connection. This bearer is always on and is referred to as the default bearer. Any additional bearers that are established for the same PDN Connection is referred to as a dedicated bearer. Different bearer-level parameters determine the QoS of the service. QoS Class Identifier (QCI) is one such parameter. QCI is a scalar number that used as a reference to access node-specific parameters that control the bearer level packet forwarding treatment. It is configured by the operator owning the access node (eg: eNodeB).

The following are the different performance characteristics:

- Resource type - Guaranteed Bit Rate (GBR) or Non-GBR
- Priority
- Packet delay Budget
- Packet error loss rate
- Maximum Data Burst Volume (for some GBR QCI's)
- Data Average Window (for some GBR QCI's)

Each Service Data Flow (SDF) is associated with one QCI. Multiple SDFs with the same QCI can be treated as a single traffic aggregate – SDF aggregate. There are a few standardized QCI Characteristics associated with standardized QCI values. This ensures that the applications/services mapped to the QCI receive the same minimum level of QoS in the multivendor networks and case of roaming, independent of the UE's access i.e, 3GPP or non-3GPP [38] [40]. One to one mapping of the QCI values to the standardized characteristics is depicted in Appendix B.

5

Fifth Generation Mobile Networks

The delivery of communication services to subscribers was the goal of the previous generations of mobile networks. The Fifth Generation of Mobile Networks (5G) will be capable of supporting more than just communication services. The development of 5G is influenced by three main drivers as mentioned in Chapter 1 which are Enhanced Mobile Broad Band (EMBB), Massive Machine Type Communications (MMTC) and Ultra Reliable Low Latency Communication (URLLC). These drivers will create possibilities for new use cases like smart cities, smart agriculture, logistics and public safety agencies depending on their data traffic profiles [40] [41].

5G networks differ from the traditional network in different aspects. 5G supports a new radio access technology called New Radio (NR). Therefore, there a need for devices that support NR and its frequencies. 5G supports a Service-Based Architecture (SBA) in comparison with the traditional network.This means that core entity is implemented as a Network Functions (NFs) which offer services to other NFs or to authorized users. Such an implementation makes the network modular and promotes reusability of components. SBA also enables virtual deployment which allows for fully scalable NFs. This results in better utilization of the available hardware resources while servicing a growing network. The focus of the thesis is to explore the role that IP Multimedia Subsystem (IMS) plays in a 5G network. This makes it important to understand the different parts of the 5G network and how different entities interact with each other.

This chapter follows a sequential flow starting with describing 5G radio access network followed by explaining the 5G core network. The Section 5.1 starts with describing the different spectrum options for 5G followed by the 5G System (5GS) Architecture in Section 5.2. Next, Section 5.3 presents the different deployment options for 5G. Section 5.4 tries to compare the elements of Enhanced Packet System (EPS) and 5GS in order to help better understand the evolution process from 4G to 5G. Section 5.5 explains how the transport network evolves in a 5G network. Section 5.6 presents the different options for voice services in 5G.

5.1. Spectrum options for 5G

5GS offers a better balance between throughput, coverage, quality and latency. Different frequency bands are utilised to achieve coverage. The usage of the bands depends on the country where it is being deployed and the available spectrum. The use cases also determine the type of band that will be used depending on the amount and the type of the traffic that needs to be sent. The different frequency bands chosen for 5G radio are grouped into three categories and are listed as follows [42] [43]:

- Low-frequency band: This is also referred to as the sub-1GHz band. These bands are capable of supporting the Internet of Things (IoT) services and provided extended mobile broadband coverage from urban to suburban and rural areas. Lower frequency implies longer wavelength and therefore larger propagation distances. This property makes the frequencies ideal for wide area and outside-in coverage as well as deep indoor coverage.

- Mid-frequency band: This band includes frequencies between 1 GHz to 2.6 GHz and 3.5GHz to 8 GHz. Current mobile broadband spectrum is within this range. This band allows different coverage and capacity benefits. Majority of the commercial 5G networks lie in this band within the 3.3 GHz to 3.8 GHz range. 1800 Mhz, 2.3 GHz and 2.6 GHz are the bands preferred for use by operators. The mid-band provides better wide area and indoor coverage when compared to the High-band Spectrum.

- High-frequency band: This fulfills the expected growth in speed, capacity, lower latency and quality promised by 5G network. Higher frequency means shorter wavelength. Shorter wavelength implies a smaller cell size. Therefore, coverage is reduced when compared to the other two bands. The frequencies preferred in this band are 26 GHz, 28GHz and 40 GHz. It is suitable for both indoor and outdoor deployments .

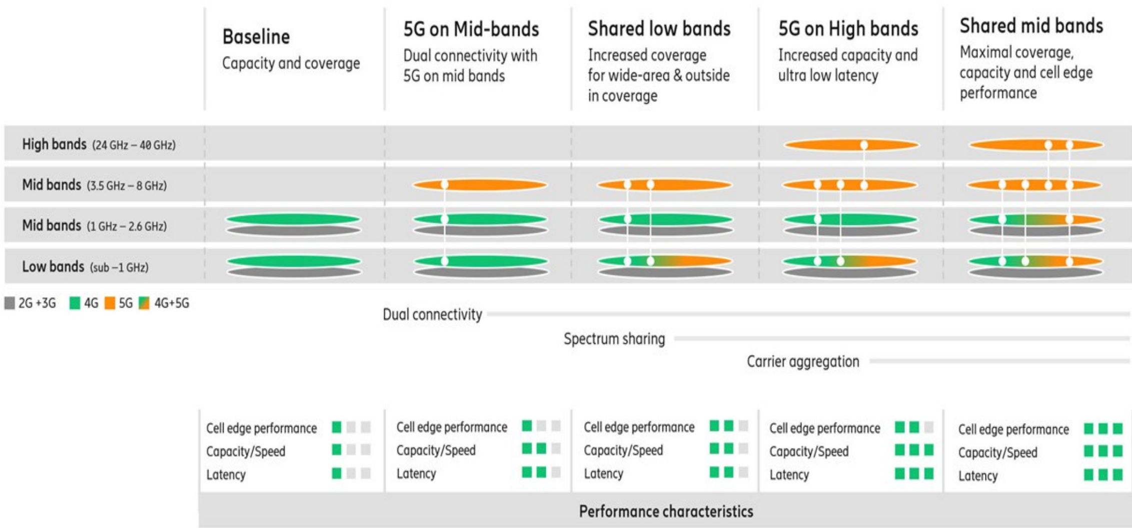

Figure 5.1: Available frequency bands for 5G deployment [18]

Figure 5.1 gives a pictorial representation for the available bands and the coverage build-out for 5G. The end goal would be to get the maximum performance given the existing resources. Spectrum is further classified into Licensed (meant for exclusive use), Shared, and Unlicensed (for shared use). NR-Spectrum Sharing would maximize spectral efficiency and higher traffic loads [18] [6].

5.2. 5G System Architecture

5G core network has a number of different functional elements, most of which are virtualized functions. This means that the elements are not standalone hardware but are software processes. Virtualization makes scalability very easy when compared to traditional entities. 5G core is a series of virtualized elements. A brief explanation about Network Function Virtualisation (NFV) and its architecture is explained in Appendix C.

The 5G system architecture is presented in figure 5.2. The different roles of the Network Functions (NF) (not limited to) are listed as follows [6]:

- Access and Mobility Management Function (AMF): This is the functional entity responsible for the management of registration and authentication of the subscriber. AMF is aware of about the cell/tracking area that the subscriber is present in depending on whether the user is connected to the network or idle. This way, AMF is able to manage the connection, rechability and mobility of the subscriber. AMF is also the point for lawful intercept for control plane signaling.

- Session Management Function (SMF): Session Management Function is responsible for carrying out tasks like establishment, modify or release of the PDU sessions. SMF works with Policy Charging Function (PCF) (explained later in this section) to determine if the data session is allowed or not. The charging data collection and support of charging interfaces is supported by SMF

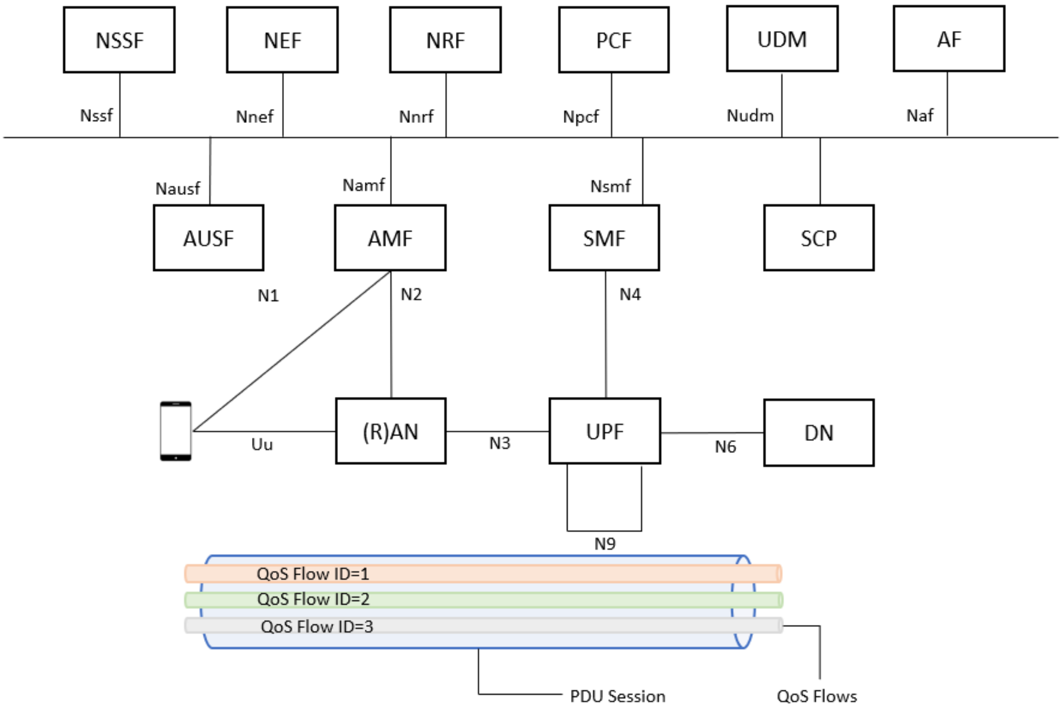

Figure 5.2: 5G network architecture and functional entities

in combination with the PCF. SMF is also involved with establishing the PDU session connectivity between the UE and the external network through the User Plane Function (UPF) (explained later in this section). SMF also chooses the UPF to route the PDU session and also allocates an IPv4 or IPv6 address during an IP session.

- User Plane Function (UPF): When a subscriber is moving within the radio access network, the gNB that the UE is attached to also changes. UPF will always remain the anchor point into the core network. The user plane connectivity will always exist between the gNB and the UPF in the core network. UPF is also responsible for enforcing QoS and policies associated with the session/user.

- Policy Charging Function (PCF): PCF implements policy control on a dynamic basis depending on the conditions existing in the network. SMF checks with PCF to determine any network conditions that might impact the user experience. PCF can also alter mobility and session related service aspects on a dynamic basis. PCF is connected to the external data network as well. It collects session related information and forwards it to the 5G core network. PCF provides policy rules to control plane functions to enforce them. It can also access subscription information relevant for policy decisions in a Unified Data Repository (UDR) (explained later in this section).

- Unified Data Management (UDM): UDM is the central repository of subscriber information. UDM is responsible for access authorization as it holds security keys. It is also involved in registration and mobility management. UDM is aware of which AMF is allocated to the user. UDM contains the subscriber profile, which is indicated to the AMF and the SMF which then determine the capabilities that the subscriber can/cannot make use of.

- Network Exposure Function (NEF): NEF is responsible for the secure exposure of capabilities and events to third parties or Application Functions (AF) (explained later in this section). NEF provides a means for the AF to securely provide service specific information to the 3GPP network. NEF provides translation for the information exchanged with the AF and also handles masking of network and user sensitive information when communicating with external AF's.

- Network Repository Function (NRF): NRF allows the discovery of the services offered by other NF's. In order to do this, it maintains the NF profile of available NF instances and their supported services. The status of NF instances that are newly registered/updated/deregistered is notified to the subscribed NF service consumer.
- Authentication Server Function (AUSF): This supports authentication for 3GPP access and untrusted non-3GPP access. It also stores the authentication keys.
- Application Function (AF): AF supports Application influence on traffic routing, accessing NEF. It interacts with the policy framework for policy control and supports IMS interaction with 5GC
- Unified Data Management (UDM): When a subscriber connects to the network, the UDM generates the authentication credentials to authenticate the user and also handle user identification. UDM handles the subscription management. Depending on the subscription data of the users, the UDM authorizes access. UDM is responsible for storing information about the NF (e.g. serving AMF or serving SMF) that is serving the UE in a PDU session.
- Network Slice Selection Function (NSSF): NSSF supports the selection of Network Slice instances that is serving the UE. NSSF determines the permitted Network Slice Selection Assistance Information (NSSAI) which is used to identify a network slice. NSSF also determines the AMF set to serve the UE.
- Data Network (DN): DN refers to the operator services, Internet access or 3rd party services. When voice services are requested, IMS is the DN.
- Service Communication Proxy (SCP): An NF service is a capability that is exposed by an NF (called the NF Service Producer) to another NF (called the NF Service Consumer) through a service based interface. When the NF's are not communicating with each other directly, the SCP provides an indirect means of communication between these producers and consumers. The NF service consumer may delegate the discovery of the target NF service producer to the SCP for such a method of indirect communication. As a result of being in the middle, the SCP also participates in message forwarding and routing to the destination NF/NF service.

All of the information between the UE and DN is carried by means of a PDU session. A PDU session is unique to each device. Each PDU session has one or more QoS flows. Each flow might have different QoS requirements and an associated QoS ID. The default QoS flow will have a certain level of quality but no packet filtering capabilities. When additional QoS flows with different QoS requirements are setup, the packets will then be filtered and treated differently. In case of IMS voice services, there will be one QoS flow associated with signalling and another one carrying the voice traffic. An example of a PDU session and the different QoS flows associated with it are presented in the Figure 5.2.

5.3. Deployment Options

5G allows integration of different radio access technologies and core networks. NR can co-exist with Evolved Universal Terrestrial Radio Access Network (E-UTRAN) and EPC and also independently with a 5G core. There are different configurations possible which makes the migration process so easy. These different configurations have been classified into two categories: Non-Standalone (NSA) and Standalone (SA). NSA combines different radio access technologies (4G LTE and NR) with Enhanced Packet Core (EPC). SA uses NR as the radio access technology and a 5G core. There are six variations within NSA and SA defined in 3GPP. Within the SA category, there are three variants shown in the Figure 5.3:

- Option 1: This variant uses EPC and Long Term Evolution (LTE) eNodeB (eNB) access
- Option 2: This variant involves a 5GC and NR gNodeB (gNB) access
- Option 5: A 5G core with LTE ng-eNB access. Here, ng-eNB represents the enhanced eNB providing LTE radio access while still connecting to 5G network

Within the NSA there are three more variants too. They are also presented in the Figure 5.3. The three options are:

- Option 3: This variant involves an EPC and both LTE eNB and gNB in the access. The eNB acts as the master node while NR enhanced-gNB (en-gNB) acts as the secondary node. A gNB that has been modified to support E-UTRAN is referred to as a en-gNB.
- Option 4: This variant uses a 5GC and both NR gNB and LTE ng-eNB in the access. The NR gNB is the master node while the ng-eNB acts as the secondary node.
- Option 7: The core is 5G and the access includes both LTE ng-eNB acting as the master node and the NR gNB acting as the secondary node.

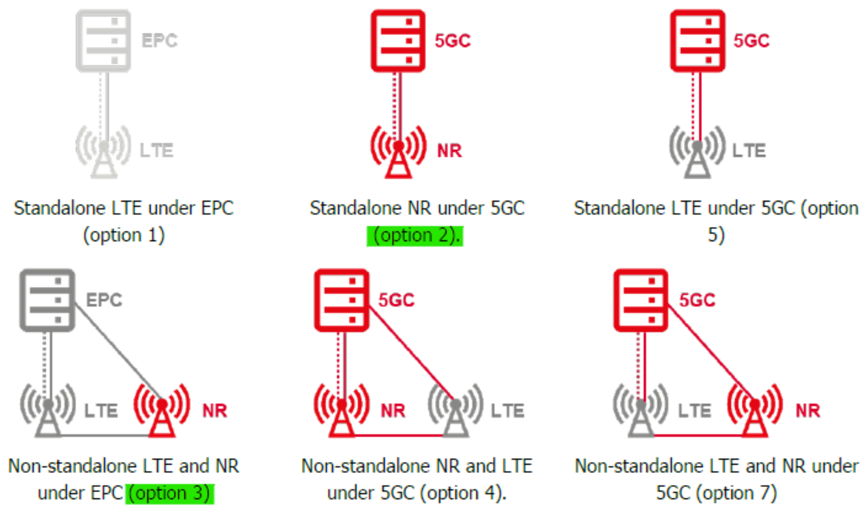

Figure 5.3: Different SA and NSA deployment options [41]

NSA Option 3 or SA Option 2 (highlighted in Figure 5.3) are the early deployment options whose standardisation has already been completed [40].

SA Option 2 with NR and a 5GC offers a great advantage in comparison to LTE by being open, flexible and having a service-based architecture, the features presented in the beginning of this chapter. SA is capable of providing all the benefits of a complete 5GS with a NR gNB based RAN and a 5G core. Although, the deployment of 5GS will require some level of interworking with EPS for service continuity. The existing LTE eNB is required to support Inter-Radio Access Technology (Inter-RAT) mobility that will allow the device to switch between LTE and NR. An AMF-MME interface might be required for smooth transition of services over the two technologies. The interworking between 5GS and 2G/3G CS is not considered during the initial stages of the deployment.

NSA Option 3 combines LTE and NR Radio access enabling compatible devices to have dual connectivity. It is referred to as E-UTRAN New Radio Dual Connectivity (EN-DC). This setup made the development of the specifications a little bit easier as the new addition was that of NR becoming a part of E-UTRAN connected to the EPC. The spectrum above 6GHz (explained in Section 5.1) will provide for large bandwidths for the operators to deliver high throughput in hotspots but coverage area would be reduced due to signal attenuation. The data throughput per 5G subscriber is expected to increase via NR and LTE Dual Connectivity. This kind of deployment does not involve a 5G Core (5GC) [41].

NSA Option 3 has three more variations depending on the control plane and user plane connectivity between the LTE eNB, gNB and the EPC. These variations are referred as Option 3, Option 3a and Option 3x and are presented in Figure 5.4. The explanation for each is listed as follows [44] [45] :

- Option 3: The gNB and the EPC are not connected directly. The eNB will act as the intermediary for all communication. The control plane and user plane connectivity from the UE to the EPC is

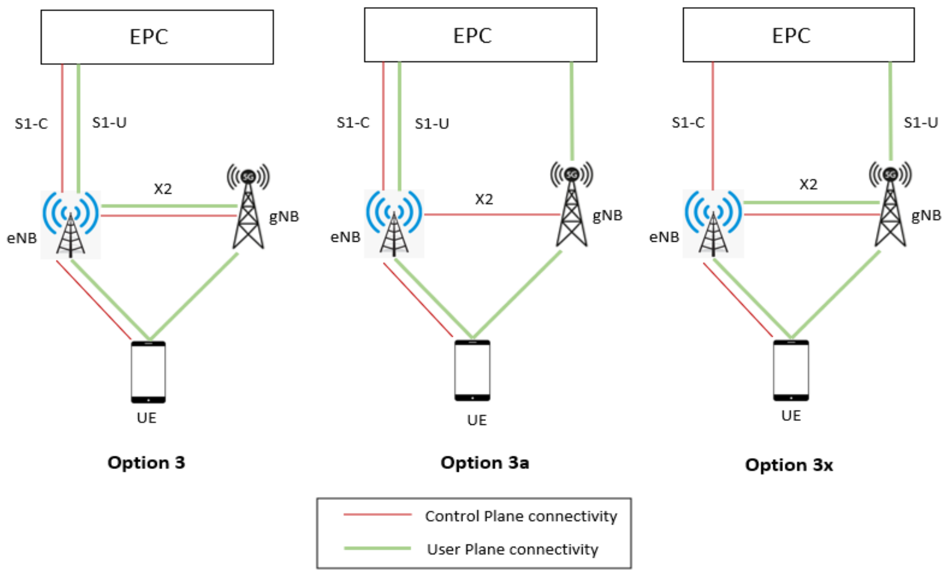

Figure 5.4: Variants in Option 3 Non Standalone Architecture

through the eNB. X2 interface carries the user plane traffic between the eNB and gNB. Traffic is split at the eNB. The control plane signalling is carried by E-UTRAN. User data will utilise both NR and E-UTRAN on the radio access side. The data traffic through gNB will still be routed through the eNB towards the EPC. This setup would result in a lot of traffic on the X2 interface that connects the eNB and the gNB. This NSA model allows users to benefit from NR on the radio access part while still being served by the EPC. However, the S1-U interface would require an increased bandwidth to support the increased traffic load.

- Option 3a: The X2 interface between the eNB and gNB has only the control plane traffic passing through it resulting in a lower traffic load on X2 when compared to Option 3. There is separate user plane connectvity from both eNB and gNB to the EPC. An example to explain this would be during a VoLTE call. The VoLTE traffic will use LTE while the data traffic will use NR. Traffic between the two is split at the EPC (S-GW).

- Option 3x: User data traffic is routed through to gNB towards the EPC while the control plane signalling is routed through the eNB. The X2 interface between both eNB and gNB exists and that allows for the possibility to use that link to send a part of the data. This is typically useful in scenarios where the connection is slower. During the early stages of deployment, this model has the least impact on the existing network and is therefore, capable of providing good service continuity. For the purpose of simplicity, whenever we are referring to NSA in the rest of the thesis, it would imply NSA Option 3x. The network for testing communication services for the purpose of this thesis follows this model. More details on this is presented in Chapter 6.

5.4. Comparison between 4G and 5G

When comparing the different functional entities in the EPC and the 5GC, there is some similarity and also differences that can be noted between the two architectures. Roles of the entities in EPC is either combined into a single NF or split into two or NFs in 5G. This section compares each entity in the EPC and 5GC with the goal of better understanding both the technologies [46].

Figure 5.5 presents the traditional entities of the EPS and their equivalent in the 5G network. The aim of this comparison is to give an idea about both the technologies. NFs in 5G can correspond to one or more elements in 4G. The E-UTRA and eNB constitute the RAN in 4G while NR and gNB constitute the access network in 5G. The roles of each EPC element is explained in detail Chapter 3. UE authentication and mobility management part of the MME in EPC is managed by the AMF in

4G	5G
Radio Access Network	
E-UTRA	NR
eNB	gNB
Core Network Elements	
MME	AMF SMF
SGW-C PGW-C	SMF
SGW-U PGW-U	UPF
HSS AuC	UDM, UDR AUSF
PCRF	PCF

Figure 5.5: One-to-one mapping of 4G and 5G entities

5GC. The session management is handled by the SMF. MME interacts with HSS and AuC while AMF interacts with AUSF to obtain authentication vectors.

In EPC, the S-GW and P-Gw have separate control plane and user plane. In 5G, the control plane functions of the S-GW and P-GW is managed by the SMF and the user plane functions are managed by the UPF.

In 4G EPC, MME, HSS and AUC manage the subscriber data and general authentication vectors. In 5G, subscriber related information, application specific data and policy data are all stored in the Unified Data Repository (UDR). UDR stores structured data that can be later introduced to the NFs. It is different from the Unstructured Data Storage Function (UDSF) which, as the name suggests, is intended for storage and retrieval of unstructured data. UDM acts as the bridge so that the relevant data is accessible to the AMF and the SMF. It is also responsible for the generation of the authentication vectors when requested by the AUSF [6].

PCRF in EPC is equivalent to the PCF in 5G. PCF like the PCRF handles the user plane resources and provide appropriate charging related information during a session. The PCRF relies on AF or on the subscriber data to apply a policy or rule. The PCF takes inputs from a lot of entities in the 5GC like the AMF, SMF, AF, UDR, Charging Function (CHF) and many more. The PCRF is directly connected to the P-GW while the PCF is connected to the SMF which then connects to the UPF. As a result, PCF creates the policies and rules and SMF will apply it on the user plane by communicating with the UPF. The PCRF manages mainly the QoS and charging for each session. In addition to the function carried out by the PCRF, PCF can also verify the right to access of a particular service depending on the geographical area of the subscriber [47].

5.5. CUPS, Xhaul and SDN in 5G networks

Figure 5.6 represents a 5G network. TThe different elements are categorized into:

- Fronthaul: Radio Units (RU) in the figure represent the radio transceiver and the mobile base station associated with it (eNB in case of a 4G network and gNB in a 5G network). A group of

RU's are connected to the Distributed Units (DU's). The radio access network, group of connected eNB's and groups of gNB's are together referred to as the Fronthaul. The UE connects to this radio side of the network.

- Midhaul: The 5G cloud RAN architecture is split into Distributed Units (DU's) and Centralized Units (CU's). In case of a distributed setup, the DU is collocated with the RU. In a centralized RAN setup, the DU is hosted in an Edge cloud data center or a regional office and can be collocated with CU. A number of DU/CU or just DU's are connected using an Ethernet carrier and are referred to as the Midhaul.
- Backhaul: The connectivity to the Packet core and interconnectivity between different CU's is referred to as the backhaul. A group of 4G eNB's over an aggregation network connecting to the packet core is also a part of the backhaul. Ethernet is used to connect to the packet core in the backhaul.

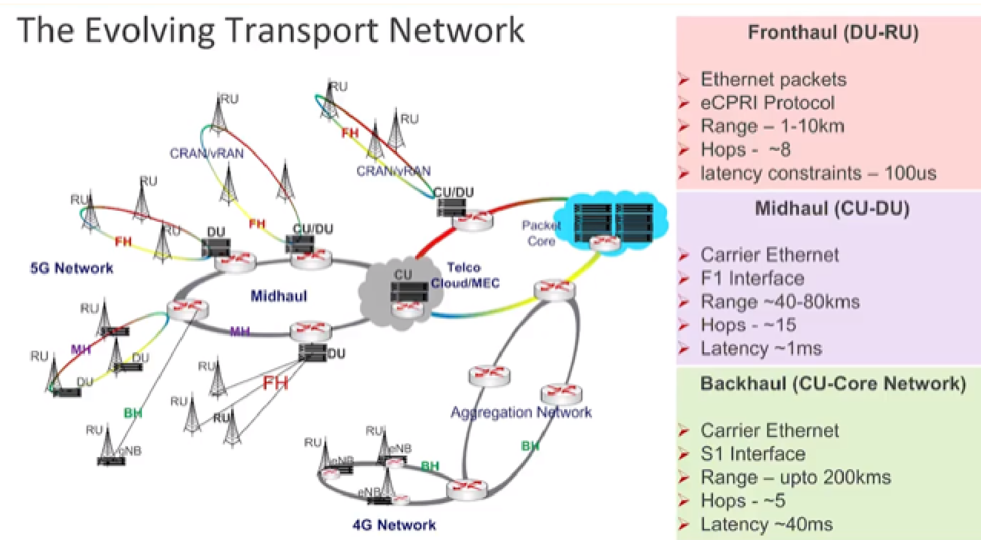

Figure 5.6: Transport network divided into Fronthaul, MidHaul and Backhaul in a 5G network [48]

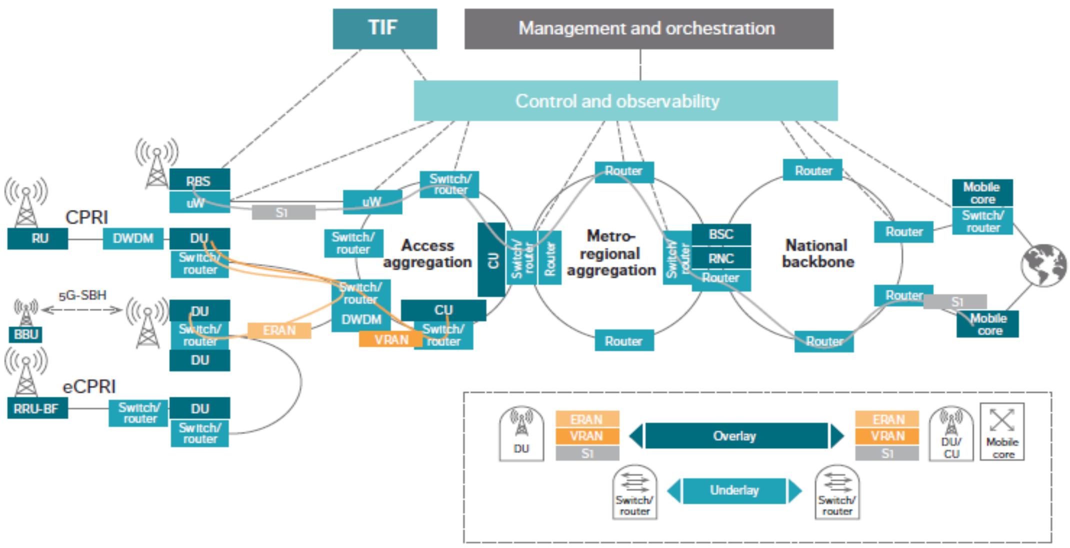

Figure 5.7: Overlay, Underlay and Management and Orchestration layers in the Transport Network [49]

The next step is to understand how the transport network needs to be modified to suit the upcoming 5G needs. Control plane and user plane separation helps achieve this goal. Transport network joins together different pieces of the RAN and the mobile core network. There is therefore a need for the transport network to be flexible, intelligent and automated. The increase in flexibility and dynamic behavior allows the addition or removal of transport connectivity. In order to provide this level of flexibility [49] suggests a solution using Software Defined Networking (SDN) along with a Transport Intelligent Function (TIF). The infrastructure constitutes an SDN-controlled overlay for different RAN and user services. The SDN with the TIF constructs a dynamically controlled and orchestrated transport network requiring minimum manual interference. TIF aids automation in the overlay in establishing dynamic connectivity with its knowledge of the RAN connections and user services. TIF also helps in optimization/reconfiguration of the network. The distributed control plane in the underlay maintains the basic infrastructure and adds redundancy in case of network failures. The four different domains: RAN, Transport, Mobile core, Control and Observability, and Management and Orchestration are represented in the Figure 5.7 [49]. TIF represents a cross-domain function between RAN and Transport domain.

5.6. Voice Migration to 5GS

Voice service in the 5GS will be tightly coupled with an existing VoLTE deployment. NR might be deployed but might not have full coverage. Circuit-Switched Fallback has not yet been standardised for 5G. Therefore, VoLTE will play an important role in service continuity. Three models have been developed to fully realize Voice services in 5G. The transition phase to enabling IMS in the 5G network aims at having minimum impact on the existing network [50] [51].

5.6.1. Voice in EN-DC: Option 3

LTE RAN is upgraded to support E-UTRA-NR Dual Connectivity (EN-DC). The connectivity for EN-DC is presented in Figure 5.8. NR is introduced as the new Radio Access Technology (RAT) being served by the EPC. The EPC and the IMS Core Network (CN) are not aware of the changes in the Radio Access Technology (RAT). SIP signalling and the user plane data principles in VoLTE remain the same. A PDN connection to the IMS APN with a default EPS bearer is used for SIP signalling (QCI=5) and a dedicated bearer for voice media (QCI=1) will be setup for VoLTE [15]. VoLTE is supported by LTE radio and E-UTRAN. NR is used to support other traffic. A detailed explanation about EN-DC connectivity is also explained as NSA Option 3 in Section 5.3.

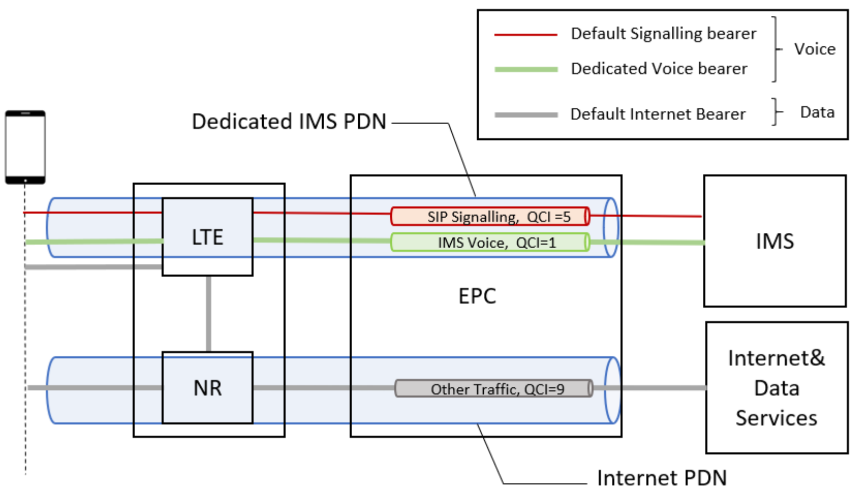

Figure 5.8: Voice in EN-DC [15]

5.6.2. Voice with EPS Fallback

This solution involves incorporating both the 5G core and the EPC. The UE must be attached to the 5GC and registered for IMS over NG-RAN. NR is also the preferred RAN. NR is used for data just like the case of EN-DC. During an LTE voice call, QoS flow establishment is triggered and a fallback to EPS occurs for both voice and data bearers. The UE is anchored to the same UPF and SMF irrespective of the RAN it is connected to. This is done by preserving the IP address when moving between 5G and 4G access. This setup consists of a SMF fulfilling the control plane functions of P-GW and the UPF fulfilling the user plane functions of the P-GW. As a result of this, the IP flows are through the same gateway enabling the UE to maintain the IP Address. The interworking between EPC and 5GC plays a very important role. The N26 interface between MME and AMF will act as the link between the EPC and the 5GC allowing the transfer of the relevant information between them.

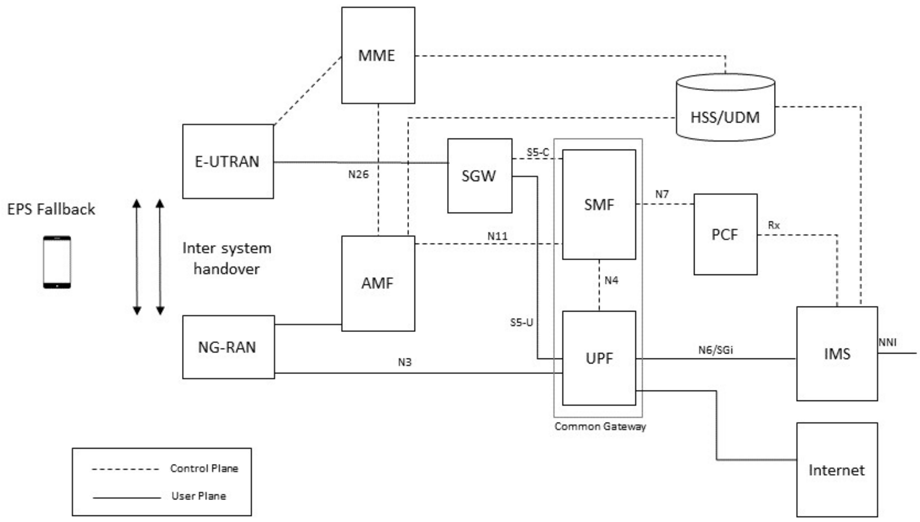

Figure 5.9: Voice with EPS fallback [16]

Typical deployment scenario includes the following aspects:

- NR coverage overlapping with the LTE coverage
- wide-area LTE coverage supporting VoLTE
- VoLTE may interwork with underlying 2G/3G network including Sigle Radio Voice Call Continuity (SRVCC). LTE and/or CS handles emergency services [16].

Some of the notable features presented in Figure 5.9 are:

- Tight interworking between EPC ad 5GC as NR won't be deployed in full network coverage.
- Single registration functionality between the UE and the core.
- SMF and UPF will replace the P-Gw (control and user plane separated) in the new 5GS.
- Interworking between MME and AMF through N26.
- Combined HSS/UDM and PCF with Rx.

5.6.3. Voice over NR: Option 2

The architecture is similar to that of EPS Fallback except for the voice capabilities being supported on NR itself. This is the case of NR SA. The RAN and core, are both 5G. There will be voice support in the NG-RAN and IMS. Coverage area determines the core which will serve the user. Typical deployment scenario includes the following aspects:

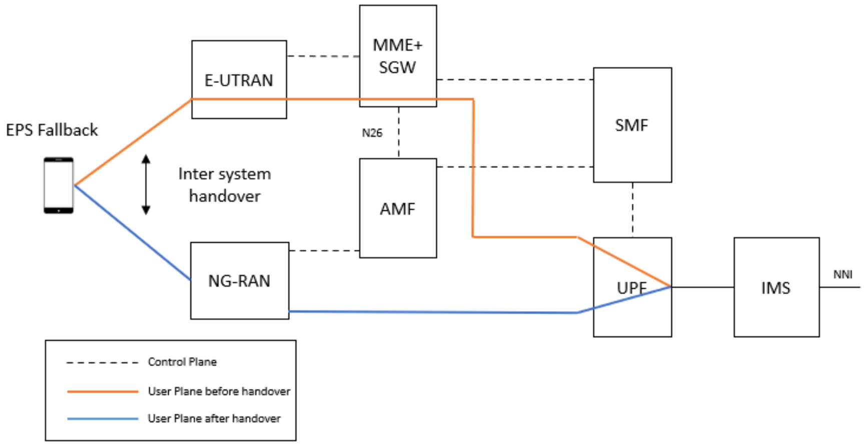

Figure 5.10: Voice over NR [16]

- NR coverage is controlled by a 5GC which supports voice.
- EPC supports LTE coverage and VoLTE.
- Emergency services are all supported on NR (and LTE) if needed.

The handovers in this case can occur both ways, 5GS to EPS or vice versa. The measurement reports generated by the UE decide whether the call needs to be handover to the 5GS or the EPS. This makes voice support on both NR and LTE a must. This setup, like Voice with EPS Fallback requires that the SMF and UPF support S5 and N26 interface resulting in a tight interworking between EPS and 5GS. Figure 5.10 represents the case of a handover from 5GS to EPS.

6

5G Network Setup at Ericsson Rijen

Ericsson's 5G network in Rijen is also called the +31 Network. The setup on site in Rijen is the NSA architecture and is shown in the Figure 6.1.

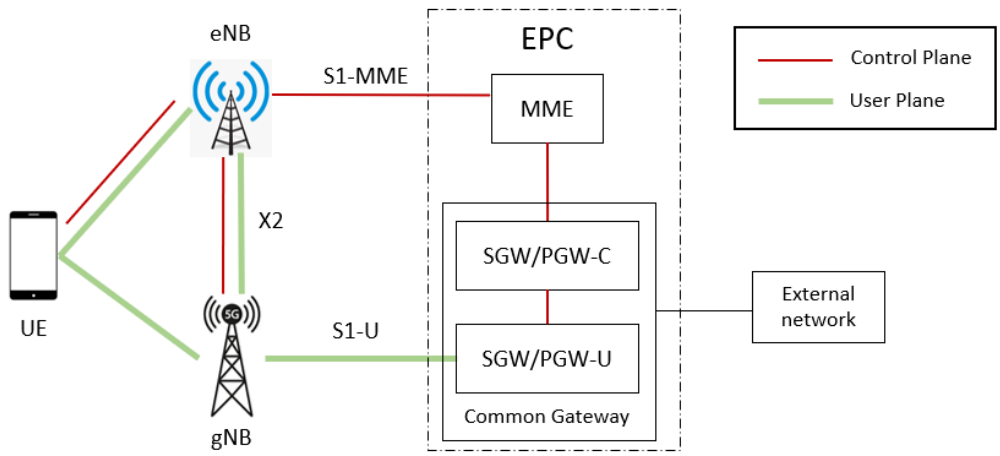

Figure 6.1: NSA architecture at Ericsson, Rijen

The User Equipment (UE) is capable of connecting to both gNodeB (gNB) and eNodeB (eNB). A dual connecting UE is a prerequisite when connecting to an NSA network. When UE connects to the gNB, it uses New Radio (NR) as access. If the UE connects to the eNB, it uses Evolved Universal Terrestrial Radio Access (E-UTRA) as access. The architecture employs Control and User Plane Separation (CUPS). Between the UE and both nodeB's, the control plane handles the signalling. The signalling data to the EPC core is sent via the eNB. The user plane data travels from the UE to the EPC through gNB. eNB is the master node resulting in the UE attachment to the LTE network. Once this is successful, the Core Network will check whether the UE is authorized to access both 4G and 5G. After authorization, the eNB and the gNB communicate via the X2 interface and proceed with bearer activation through the gNB.

The PDU session establishment will be from the UE to the S-GW through both the nodeB's. In the setup at Rijen, the S-GW and P-GW are combined into one Common Gateway. The control plane and user plane functions are still separated in the Common Gateway. Other 4G EPC elements will require some upgrades in order to support 5G NR. The upgrades are seen at the MME, S-GW/P-GW, HSS and PCRF. EPC when mentioned in the context of 5G implies the upgraded EPC which supports NR. The modifications to accommodate NR doesn't have a major impact on the core network's hardware.

The +31 network consists of three areas. For the purpose of this thesis, it is important to understand how the +31 network is able to provide connectivity for the devices. This is explained in Sections 6.1-6.4. This is followed by understanding how the remote IMS network located in Sweden can be accessed using the +31 Network in Section 6.6 and Section 6.7. Figure 6.2 presents a high level understanding of the connectivity between different components. Each component is explained in detail in the following sections.

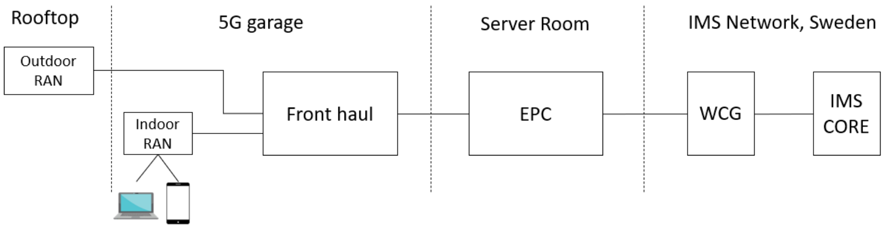

Figure 6.2: Overview of the network for the experiment

6.1. Rooftop (Outdoor)

The outdoor RAN includes the Base Band Unit (BBU), Remote Radio Unit (RRU) and Antenna. An antenna integrated Radio unit (AIR 6488) placed on the rooftop provides the 5G radio access. Both the LTE and the NR are at 3.5GHz. It is a 64TR Time Division Duplex (TDD) Advanced Antenna System (AAS) for NR. It utilizes the beam forming and Multi-User-Multiple Input Multiple Output (MU-MIMO) technology. It provides enhanced coverage, capacity and network performance. The RRUs placed outside provide coverage there. A GPS antenna is also positioned. The different components of the RAN (outdoor/indoor) installed at Ericsson's office is depicted in the Figure 6.3.

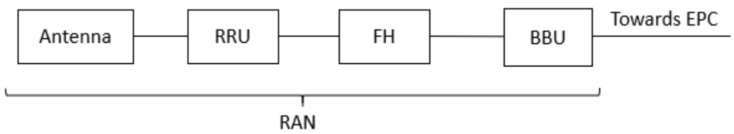

Figure 6.3: Different units connected in the RAN

6.2. 5G Garage

This is the indoor area which is provisioned to allow users to connect to the 5G network and carry out experiments. Front Haul (FH) connects the RRU and the BBU's. One BBU operates at 3.5GHz and the other at 5GHz. An optical fiber connects the Front Haul to the Back Haul (BH) and further to the EPC. An intermediate router for monitoring and analysis is also present. The laptops used in the experiment connect to the RAN here.

6.3. Server Room

6.4. Servers and Applications

6.5. WebRTC

Real-time communication involves the real-time transfer of generic data, voice and video between peers. WebRTC enables this kind of communication between peers. Such peer-to-peer (P2P) connections are enabled using browsers. This eliminates the need to use a third-party or native application or any other plugin on their devices (phone, laptop or tablet). Anybody with access to a web browser supporting WebRTC will be capable of sending or receiving real-time data like voice or video data. Mozilla and Google have developed several Application Program Interface (API's) that can be used on their browsers to serve this purpose. The API's can be generalized into two categories. One includes the API's that allow the applications to access the device's microphone/camera. The other one is the set of API's that enable P2P connectivity [52] [53] [54].

In this thesis, WebRTC is used to enable peer-to-peer communication through the IMS network. In a typical LTE network, when a UE attempts to access the IMS network, a default signalling bearer and one or more dedicated bearers are established for real-time services. In a similar way, when WebRTC Client tries to access the IMS network through a link or a URL, a WebRTC server provides the interface (usually a javascript (JS) code) that will help establish the P2P connectivity. The signalling plane between the client and the IMS is managed by this WebRTC server. Once the signalling is complete, the media plane is established directly between the client and the IMS GW. More explanation on this is provided in Section 6.7.

6.6. Ericsson Web Communication Gateway (WCG)

Ericsson Web Communication Gateway provides Web Access (WA) and acts as a Non-SIM Device Server (NSDS). Figure 6.4 gives a detailed description of the components of the WCG and how the clients connect to the IMS core using the WCG.

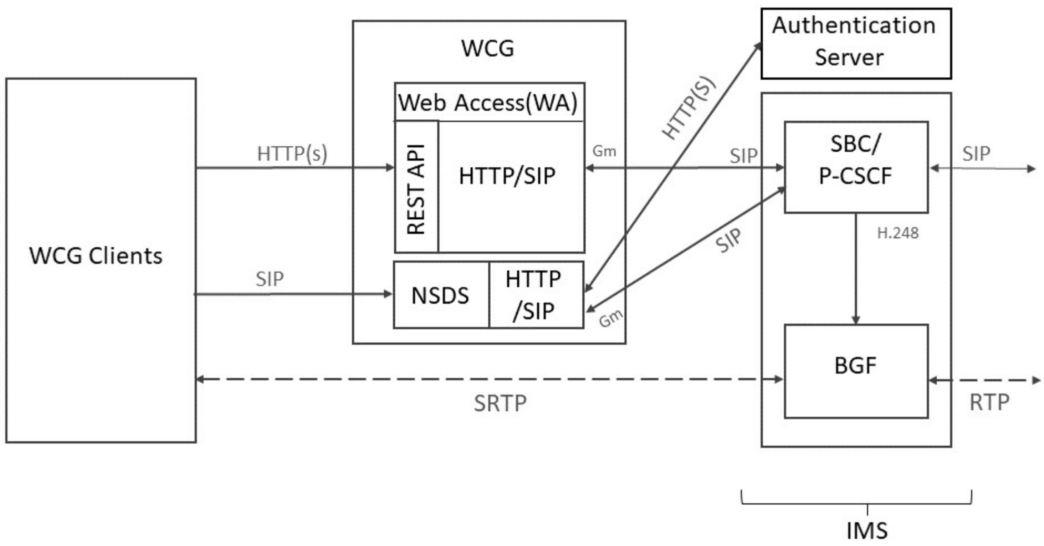

Figure 6.4: Ericsson WCG Overview

WA is responsible for the connectivity between the Web and IMS. HTTP-based clients can connect to an IMS network through the WA. The NSDS provides a Back-to-back user Agent (B2BUA) functionality and allows a SIP client to access the IMS core network after using the credentials provided by an external Authentication server. This Authentication Server verifies the credentials provided by the client and retrieves the credentials of the IMS core network to enable access.

The WCG offers RESTful HTTP interface to web clients. The client-side requests arriving at the WCG are converted into SIP, MSRP, XCAP or DNS to perform operations on the network side. With respect to this thesis project, HTTP/REST protocols used for signaling between the web client and the WCG, and the SIP protocol which deals with the UDP and TCP traffic are of relevance. WA converts the HTTP requests coming in from the client into SIP before sending it to the P-CSCF/SBC over the Gm reference point as shown in Figure 6.4.

WebSocket API enables a duplex connection between client (browser). This API eliminates polling the server for a response. Messages are sent to the server and event-driven responses are received. In the initial stages of writing a chat application for P2P connectivity, WebSockets (Socket.io) were used to establish connectivity between a client and a Node.js server. This setup was purely to test the webRTC functionality on two browsers connected over the internet. The idea was to test out P2P connectivity between two browsers over the internet connection and later incorporate the IMS network connectivity. The WCG also supports WebSockets, but it is not represented in Figure 6.4.

6.7. WebRTC and IMS

A WebRTC client can access the IMS network and is referred to as the WebRTC IMS Client (WIC). The media plane feature negotiation utilizes Session Description Protocol (SDP) offer/answer. The WebRTC IMS architecture is as shown in the Figure 6.5. More information on the reference points can be found in [10].

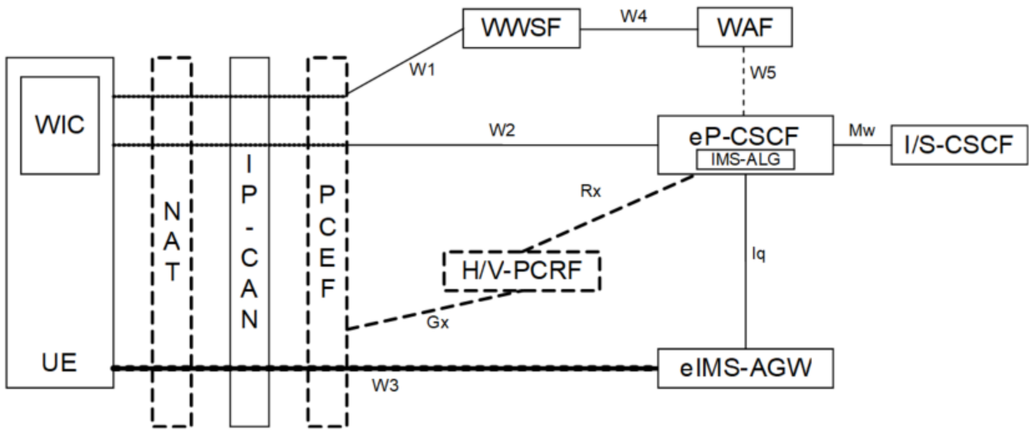

Figure 6.5: WebRTC IMS architecture and reference model [10]

WebRTC Web Server Function (WWSF) is the web server that a user agent (located in Rijen) contacts and it is located within the operator network or within a third-party network. It provides the web page presenting the user interface for IMS access. A JavaScript (JS) WIC application is downloaded on to the browser on the UE. The WebRTC Authorisation Function (WAF) issues an authorization token to the WWSF. It may use this token itself or trust the user identity provided by the WWSF when a WIC is requesting access to the IMS network. The P-CSCF is situated in the operator network and enhanced to support WebRTC and is referred to as eP-CSCF. This the signaling end point from the client side [10].

In case of a non-cellular access, a Network Address Translation (NAT) might be required to establish and maintain connections to the internet. Private networks employ this technique to hide the private IP address and access the internet using another public IP address. When WIC's try to connect to the IMS network, there might be a need to do a NAT traversal (NAT-T). NAT-T enables P2P communication. In the experiment setup used for this thesis, the IMS core network is situated in Ericsson, Sweden. It is accessed using a public IP address and there is a need for NAT-T. Enhanced IMS Access Gateway (eIMS-AGW) is the standard IMS-AGW enhanced to support WebRTC which provides NAT-T support including Interactive Connectivity Establishment (ICE). eIMS-AGW also performs other functions like supporting the media plane interworking and optimization as needed by the WIC's. It supports de-multiplexing the RTP and RTCP streams towards the core network and may also be used to transcode audio and video codecs depending on what is supported by the browser. PCEF and PCRF shown in Figure 6.5 are present only for EPC access with QoS depending on the configuration. W1, W2, W3, W4, W5, Rx, Gx, Mw and Iq are the different reference points between the corresponding elements shown in Figure 6.5.

6.8. Final Test Setup

Figure 6.6 presents the detailed network setup which is used for the experiments in this thesis. The results are presented in Chapter 7.

The experimental setup consists of the following two devices connecting to the IMS core through the

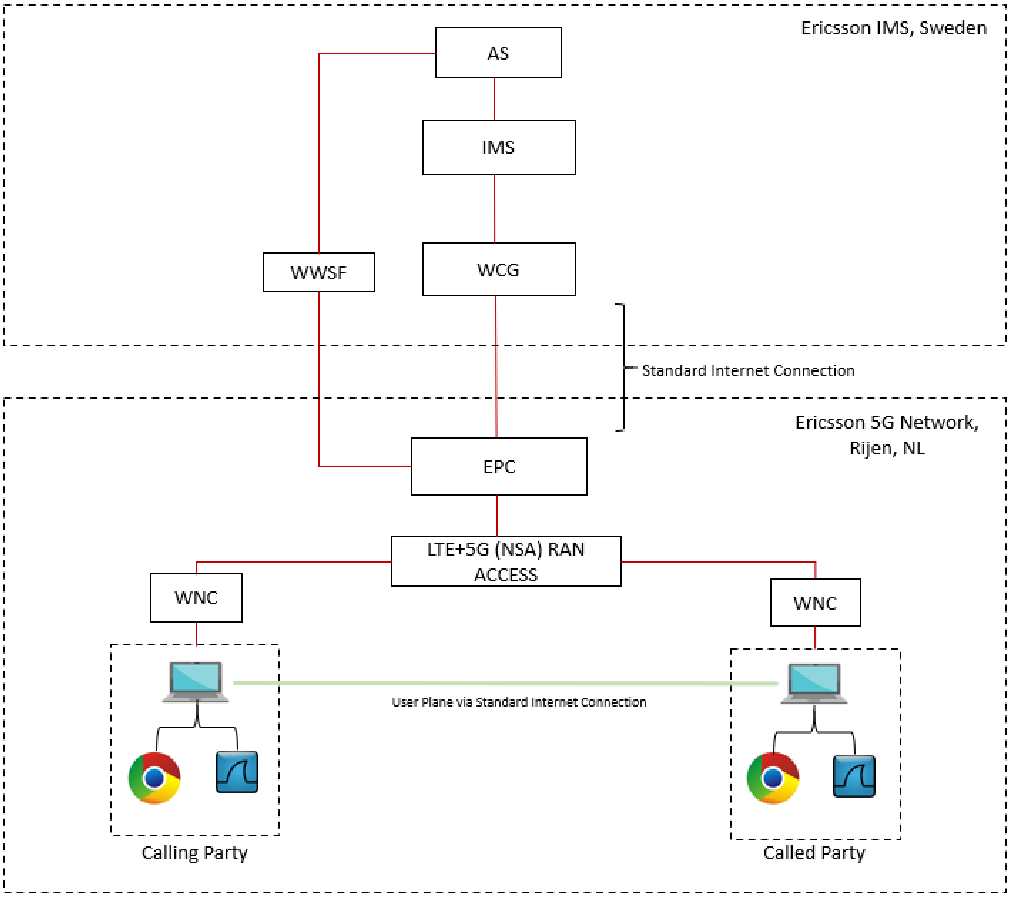

Figure 6.6: Test Setup for the experiments

wifi Access point:

- Laptop 1 - Windows 10, Google Chrome
- Laptop 2 - Windows 10, Google Chrome

Both laptops have Google Chrome installed on them to provide the WebRTC access to IMS. For the measurements, Wireshark has been installed on both the systems to sniff the incoming and outgoing packets. WebRTC supports a tool called "webrtc-internals" which provides information on the ongoing WebRTC sessions on Chrome. A dump file is generated which is later analyzed and presented in the results in Chapter 7.

Each laptop is connected to a Wistron NeWeb Corporation (WNC) device that acts as a hotspot and allows connection to the 5G network on the access side. The WNC is attached to the network through the 5G antenna located in the 5G garage. The connectivity between the 5G garage and the 5G EPC located in the Server Room has been explained earlier in this chapter.

With the laptop connected to the 5G network, the next step is to the IMS network located in Sweden. In order to do this, we first need to connect to the Ericsson WCG. The access to the WCG is provided through a URL. Access to the WCG requires a username and password. One is used for the calling party and another for the called party. A Session Traversal Utilities NAT (STUN) server is configured to obtain an external IP address and a Traversal Using Relays around NAT (TURN) server is configured to direct traffic if P2P connection fails.

After a successful login, the interface for using the supported IMS services appears. There are three IMS services that are tested in this thesis:

- Voice

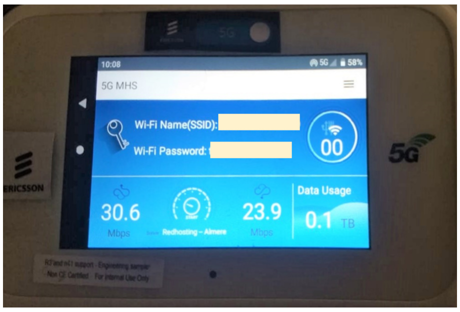

Figure 6.7: WNC device providing 5G access to laptop

- Video (Voice and Video)

- Instant Messaging

When trying to test a service, for e.g. voice, the destination address/number needs to be entered. Section 4.4 explains the formats supported. With the IMS network located remotely, an IMS Public User Identity (IMPU) is used in the format similar to "*sip:+31123456789@xyz.nl*". After this, the call can be placed to the second user. A snapshot taken during a voice and another during a video call is shown in Figure 6.8.

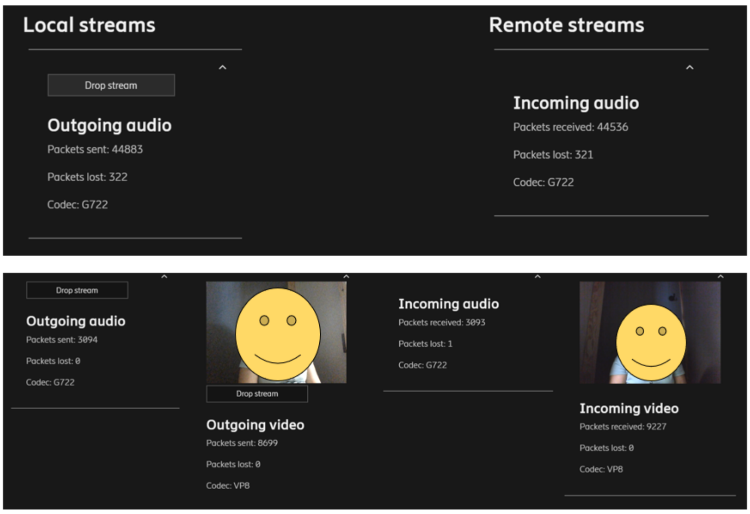

Figure 6.8: Voice call snapshot (top), Video call snap (bottom)

Wireshark and webrtc-internals help record the data of each session. The data analysis is done and graphs are generated for different QoS parameters.

7

Results

The test for each of the IMS services is carried out multiple times. Wireshark is used for packet sniffing. A WebRTC tool called ”testRTC” provides additional information for the analysis of the dump file that was saved during the running of each of the tests. The performance of one test (referred to as Run 1) per parameter is depicted in the graphs that follow. The chapter is categorized according to the different parameters measured. The results of each of the IM services are grouped based on the parameter under consideration. Section 7.1 presents the results for the bitrate for voice and video calls. Section 7.2 explains the packet loss parameter. Section 7.3 presents the results for jitter followed by Section 7.4 which presents the latency results.

7.1. Bitrate

Bitrate stands for the number of bits transmitted or received over a given interval.

7.1.1. Voice Call

The trace of the voice call is plotted in the graph depicted in Figure 7.1. As soon as the call is established and the voice packets are being sent, there is a spike in the graph which continues for the duration of the call. When saying voice packets, it means that the voice data is being transported over the IP network. Wireshark is used to collect the packet trace. The media stream (voice in this case) is represented by the RTP data that is transmitted between the two users and this is indicated in Figure 7.1. At the beginning of a call, the number of RTP packets sent are expected to increase as the voice data is being carried forward. This behaviour is observed in the Figure 7.1. When the call is terminated, there is a drop in the number of RTP packets sent/received per second indicating that the call has ended. RTCP packets convey information about the on-going session and the participants involved in it. The different

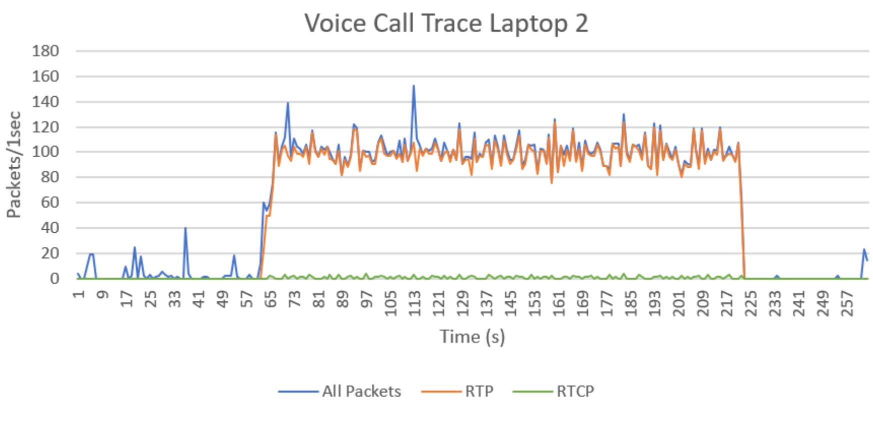

Figure 7.1: Voice Call Trace

QoS parameters are recorded and plotted using the dump file generated by the webrtc-internals page.

All values in Kbps	Run 1	Run 2	Run 3
Incoming bitrate	61	62	63
Outgoing bitrate	63	63	63

Table 7.1: Bitrate values for voice call over three tests

The graph for the outgoing and incoming audio bitrate is indicated in Figure 7.2. The incoming and outgoing values are almost the same. G722 Audio codec is used in the call. The codec offers high quality audio but requires an average bandwidth of 80kbps. This codec has a bitrate value ranging between 48-64 kbps. The additional bits are for overhead (to include information like headers). In practice, data is encoded at a bitrate of 64 kbps. The values obtained during the measurement are between 62-64 kbps which is within the acceptable range [55]. 5G NSA architecture is deployed. These values therefore, do not reflect the performance of a system with a full 5GC deployed. Voice in case of a 5G SA has a predefined set of parameters to measure their performance. These Key Performance Indicator (KPI) values are defined in [1].

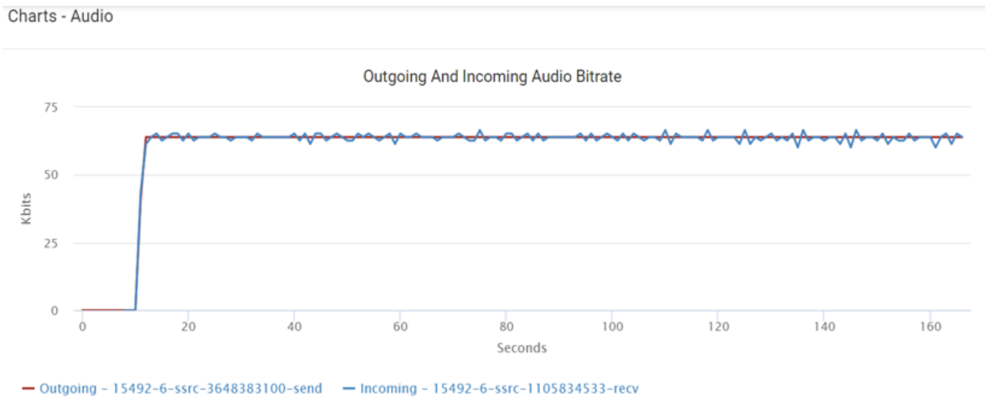

Figure 7.2: Outgoing and Incoming Audio Bitrate - Run 1

7.1.2. Video call

Placing a video call is quite similar to that of a voice call except for the fact that you have a visual of the person you are talking to. This involves the negotiation of some additional parameters when compared to a voice call. Codecs that will be used will need to be agreed upon. WebRTC acts as a facilitator here for the negotiation of the media plane. A default bearer is established which will carry all the signaling information in a similar fashion as that explained in a VoLTE IMS call. A user plane is setup next to carry the voice and video information between the two users. A bearer carrying voice and another carrying video will be established to allow both types of data to be transmitted/received. In this case, the agreed upon codecs were G722 for audio and VP8 for video. WebRTC technology aims at utilising browsers to place call to another user. The impact of this also falls on the codec used. IMS prefers H.264 codec for video services in Video over LTE (ViLTE). H.264 is however licensed and is not available freely. As a result of this, VP8 codec has been the chosen video codec for WebRTC in general. VP8 is an open source codec and is available freely.

All values in Kbps	Run 1	Run 2	Run 3
Incoming bitrate Audio	62	60	62
Outgoing bitrate Audio	63	63	62
Incoming bitrate Video	524	22	1255
Outgoing bitrate Video	925	1103	1020

Table 7.2: Average bitrate values during video call over three tests

The call trace for the video call is shown in Figure 7.3. Similar to the case of an audio call, the graph clearly depicts the start of the call with an increase in the number of RTP packets. The number of RTP packets transmitted is dropped when the call ends. RTCP packets carry information regarding the user plane and the users involved in it.

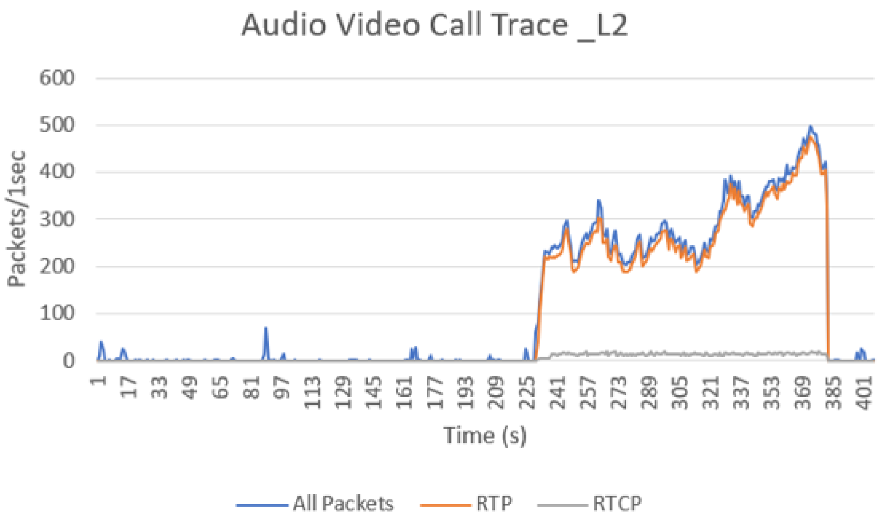

Figure 7.3: Video Call trace

There will be two values for bitrate here one each for voice and video. The graphical representation of the variation of audio bitrate over the duration of a video call is represented in Figure 7.4 followed by the graph for the varying video bitrate over the call duration in Figure 7.5.

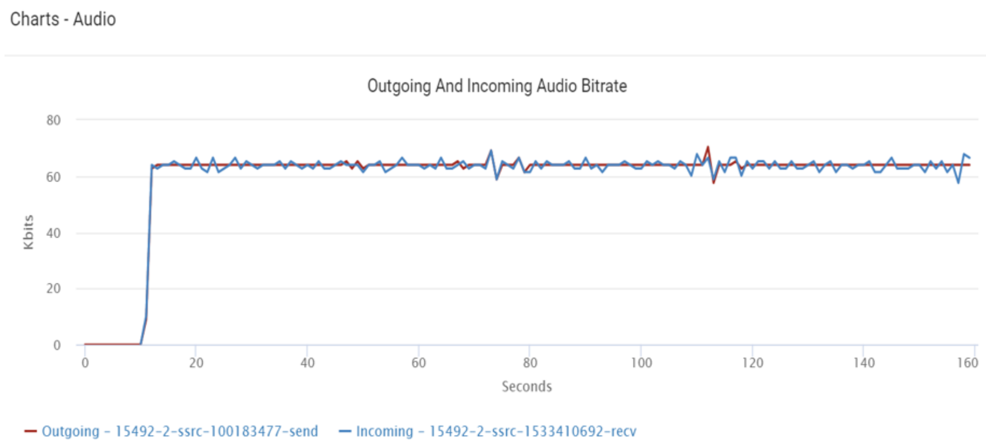

Figure 7.4: Outgoing and Incoming Audio Bitrate

It can be observed that the audio bitrate is fairly constant around the values of 60 kbps, which is similar to what was observed in the voice call. The video bitrate appears to be increasing with time. Video bitrate affects the quality of the video. Higher bitrate usually means higher quality video, but it would also consume more bandwidth. In Figure 7.4, it can be observed that the bitrate is gradually increasing which means that the video quality improved over the duration of the call. There are constraints that can be set regarding the bitrate usage in WebRTC providing flexibility in the applications.

7.2. Packet Loss

Packet loss represents the number of packets lost between the sender and the receiver. The measurement for this needs to be done on the receiver laptop (user 2). Packet loss can be represented in percentage values or as number of packets lost during the transmission. The general formula for the

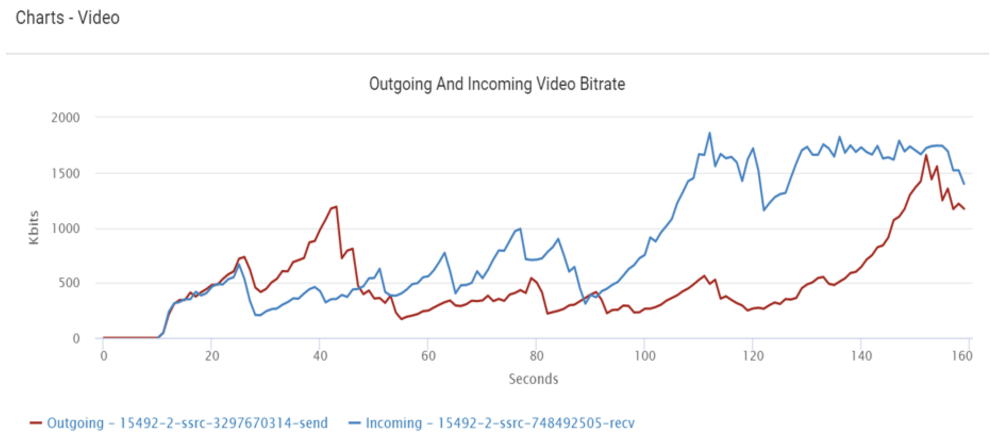

Figure 7.5: Outgoing and Incoming Video Bitrate

calculation of the packet loss percentage is as follows:

$$Packet\ loss\ percent = \frac{(No.\ of\ Packets\ sent - No.\ of\ Packets\ received)}{Total\ Number\ of\ packets} * 100 \tag{7.1}$$

There are two kinds of RTCP packets called Sender Report (SR) and Receiver Report (RR). The number of packets lost between two consecutive RR's gives the packet loss count in that interval. The expected number of packets is given by the difference in the last sequence numbers received. The ratio of these values gives the packet loss fraction over that particular interval [56].

7.2.1. Voice Call

The incoming and outgoing packet loss percentage values over three tests are mentioned in Table 7.3.

All values in percentages	Run 1	Run 2	Run 3
Incoming Packet Loss	2.7	0.0	0.0
Outgoing Packet Loss	0.2	0.0	0.0

Table 7.3: Packet loss percentage for audio call over three tests

The graphical representation of the packet loss percentage variation over the duration of the voice call is obtained from the analysis of the webrtc-internals dump file. The graph from one run is presented in Figure 7.6. In order to attain a decent call quality, the packet loss percentage has to be lower than

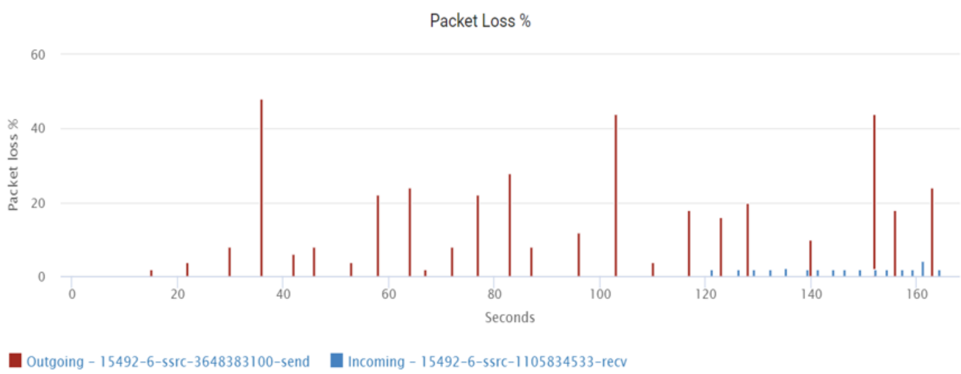

Figure 7.6: Outgoing and Incoming Audio Packet Loss

5%. That value is much lower in all the three tests contributing to the factors that make the call quality is pretty decent. Packet loss is an important parameter for the experiment because it represents the user experience. Any loss value above 10% would result in significant distortion and the user would notice the voice gaps in the stream being heard. Figure 7.7 is the Packet capture (pcap trace) of the call highlighting the instances where missing packets can be observed. These are traced by following the packet sequence numbers in the RTP stream. The particular portion of the trace constitutes the filtered RTP and RTCP packets between two specified IP addresses on Laptop 1. The IP addresses were identified by comparing wireshark traces on both the laptops. IP address of the laptop itself was found by using the command prompt. When observing the trace files, it was noted that both laptops had one common IP address when sending the data indicating that it was the common gateway. As a result, the results were filtered. In Figure 7.7, it can be observed that the sequence numbers in this particular part of the trace are not missing any packet. It is to be noted that the direction of packet flow influences the sequence number. A packet travelling from Laptop 1 to Laptop 2 will have a sequence number different from a packet travelling from Laptop 2 to Laptop 1. Figure 7.8 shows the packets received on Laptop 2. The sequence numbers appear to be jumbled solely because of the direction of the packet traversal.

ip.addr==192.36.162.55 && ip.addr==192.168.43.192

	Time	Source	Destination	Protocol	Length	DSCP Value	Sequence number	Info
10906	105.697255	192.36.162.55	192.168.43.192	RTP	228	Default	2045	PT=ITU-T G.722, SSRC=0x43A40BE1, Seq=2045, Time=188221
10925	105.726175	192.36.162.55	192.168.43.192	RTP	228	Default	2046	PT=ITU-T G.722, SSRC=0x43A40BE1, Seq=2046, Time=188221
10929	105.733934	192.36.162.55	192.168.43.192	RTP	228	Default	2047	PT=ITU-T G.722, SSRC=0x43A40BE1, Seq=2047, Time=188221
10973	105.796024	192.36.162.55	192.168.43.192	RTP	228	Default	2048	PT=ITU-T G.722, SSRC=0x43A40BE1, Seq=2048, Time=188221
10975	105.799872	192.36.162.55	192.168.43.192	RTP	228	Default	2049	PT=ITU-T G.722, SSRC=0x43A40BE1, Seq=2049, Time=18822
10982	105.808285	192.36.162.55	192.168.43.192	RTP	228	Default	2050	PT=ITU-T G.722, SSRC=0x43A40BE1, Seq=2050, Time=18822
10989	105.822458	192.36.162.55	192.168.43.192	RTP	228	Default	2051	PT=ITU-T G.722, SSRC=0x43A40BE1, Seq=2051, Time=18822
11011	105.846682	192.36.162.55	192.168.43.192	RTP	228	Default	2052	PT=ITU-T G.722, SSRC=0x43A40BE1, Seq=2052, Time=18822
11034	105.879570	192.36.162.55	192.168.43.192	RTP	228	Default	2053	PT=ITU-T G.722, SSRC=0x43A40BE1, Seq=2053, Time=18822
11042	105.887818	192.36.162.55	192.168.43.192	RTP	228	Default	2054	PT=ITU-T G.722, SSRC=0x43A40BE1, Seq=2054, Time=18822
11064	105.929502	192.36.162.55	192.168.43.192	RTP	228	Default	2055	PT=ITU-T G.722, SSRC=0x43A40BE1, Seq=2055, Time=18822
11065	105.929599	192.36.162.55	192.168.43.192	RTP	228	Default	2056	PT=ITU-T G.722, SSRC=0x43A40BE1, Seq=2056, Time=18822
11084	105.953397	192.36.162.55	192.168.43.192	RTP	228	Default	2057	PT=ITU-T G.722, SSRC=0x43A40BE1, Seq=2057, Time=18822
11100	105.974158	192.36.162.55	192.168.43.192	RTP	228	Default	2058	PT=ITU-T G.722, SSRC=0x43A40BE1, Seq=2058, Time=18822
11175	106.087382	192.36.162.55	192.168.43.192	RTP	228	Default	2059	PT=ITU-T G.722, SSRC=0x43A40BE1, Seq=2059, Time=18822
11176	106.087478	192.36.162.55	192.168.43.192	RTP	228	Default	2060	PT=ITU-T G.722, SSRC=0x43A40BE1, Seq=2060, Time=18822
11177	106.087529	192.36.162.55	192.168.43.192	RTP	228	Default	2061	PT=ITU-T G.722, SSRC=0x43A40BE1, Seq=2061, Time=18822

Figure 7.7: Wireshark trace and packet loss - Laptop 1

ip.addr==192.36.162.55 && ip.addr==192.168.43.28

No.	Time	Source	Destination	Protocol	Length	Sequence number	Info
4528	105.664537	192.36.162.55	192.168.43.28	RTP	228	17617	PT=ITU-T G.722, SSRC=0xD975E87C, Seq=17617, Time=289781532
4529	105.682759	192.36.162.55	192.168.43.28	RTP	228	17618	PT=ITU-T G.722, SSRC=0xD975E87C, Seq=17618, Time=289781548
4530	105.685932	192.168.43.28	192.36.162.55	RTP	228	19271	PT=ITU-T G.722, SSRC=0x41E9B225, Seq=19271, Time=242212537
4531	105.698819	192.36.162.55	192.168.43.28	RTP	228	17619	PT=ITU-T G.722, SSRC=0xD975E87C, Seq=17619, Time=289781564
4532	105.705156	192.168.43.28	192.36.162.55	RTP	228	19272	PT=ITU-T G.722, SSRC=0x41E9B225, Seq=19272, Time=242212553
4533	105.718736	192.36.162.55	192.168.43.28	RTP	228	17620	PT=ITU-T G.722, SSRC=0xD975E87C, Seq=17620, Time=289781580
4534	105.723232	192.168.43.28	192.36.162.55	RTP	228	19273	PT=ITU-T G.722, SSRC=0x41E9B225, Seq=19273, Time=242212569
4535	105.740076	192.36.162.55	192.168.43.28	RTP	228	17621	PT=ITU-T G.722, SSRC=0xD975E87C, Seq=17621, Time=289781596
4536	105.745575	192.168.43.28	192.36.162.55	RTP	228	19274	PT=ITU-T G.722, SSRC=0x41E9B225, Seq=19274, Time=242212585
4537	105.752779	192.36.162.55	192.168.43.28	RTP	228	17622	PT=ITU-T G.722, SSRC=0xD975E87C, Seq=17622, Time=289781612
4538	105.762761	192.168.43.28	192.36.162.55	RTP	228	19275	PT=ITU-T G.722, SSRC=0x41E9B225, Seq=19275, Time=242212601
4539	105.764604	192.36.162.55	192.168.43.28	RTP	228	17623	PT=ITU-T G.722, SSRC=0xD975E87C, Seq=17623, Time=289781628
4540	105.785603	192.168.43.28	192.36.162.55	RTP	228	19276	PT=ITU-T G.722, SSRC=0x41E9B225, Seq=19276, Time=242212617
4541	105.787582	192.36.162.55	192.168.43.28	RTP	228	17624	PT=ITU-T G.722, SSRC=0xD975E87C, Seq=17624, Time=289781644
4542	105.802695	192.168.43.28	192.36.162.55	RTP	228	19277	PT=ITU-T G.722, SSRC=0x41E9B225, Seq=19277, Time=242212633
4543	105.826566	192.168.43.28	192.36.162.55	RTP	228	19278	PT=ITU-T G.722, SSRC=0x41E9B225, Seq=19278, Time=242212649
4544	105.837662	192.36.162.55	192.168.43.28	RTP	228	17625	PT=ITU-T G.722, SSRC=0xD975E87C, Seq=17625, Time=289781660
4545	105.843324	192.168.43.28	192.36.162.55	RTP	228	19279	PT=ITU-T G.722, SSRC=0x41E9B225, Seq=19279, Time=242212665
4546	105.843522	192.36.162.55	192.168.43.28	RTP	228	17626	PT=ITU-T G.722, SSRC=0xD975E87C, Seq=17626, Time=289781676
4547	105.861228	192.168.43.28	192.36.162.55	RTP	228	19280	PT=ITU-T G.722, SSRC=0x41E9B225, Seq=19280, Time=242212681
4548	105.862578	192.36.162.55	192.168.43.28	RTP	228	17627	PT=ITU-T G.722, SSRC=0xD975E87C, Seq=17627, Time=289781692
4549	105.884535	192.168.43.28	192.36.162.55	RTP	228	19281	PT=ITU-T G.722, SSRC=0x41E9B225, Seq=19281, Time=242212697

Figure 7.8: Wireshark trace and packet loss - Laptop 2

7.2.2. Video Call

The percentage of packets lost in a video call is calculated in this section. Just like the case of voice call, the packet loss is impacted by different conditions in the network. As mentioned in the previous section, the connection to the IMS network is over a standard PDN (internet) session. This factor impacts the different parameters drastically. Testing methodology is similar to that mentioned in the previous sections. The wireshark trace recorded during the call produced the graph shown in Figure 7.3 mentioned in section 7.1.

All values in percentages	Run 1	Run 2	Run 3
Incoming Audio	0.3	3.8	0
Outgoing Audio	0.3	0.3	0
Incoming Video	0	0	0
Outgoing Video	0	0	0

Table 7.4: Packet loss percentages for Video call over three tests

The average packet loss values over three tests are tabulated in the Table 7.4. Figure 7.9 and 7.10 are the graphs depicting the voice and video packet loss percentages separately varying over the duration of a call. The average values depicted in Table 7.4 reflect that while some of the voice data

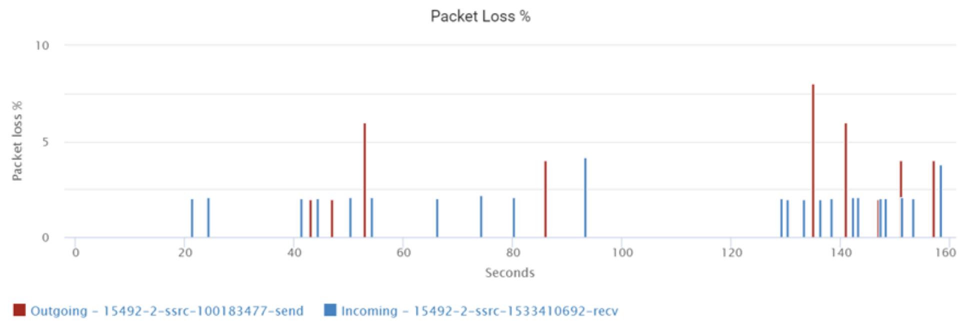

Figure 7.9: Outgoing and Incoming Audio Packet Loss

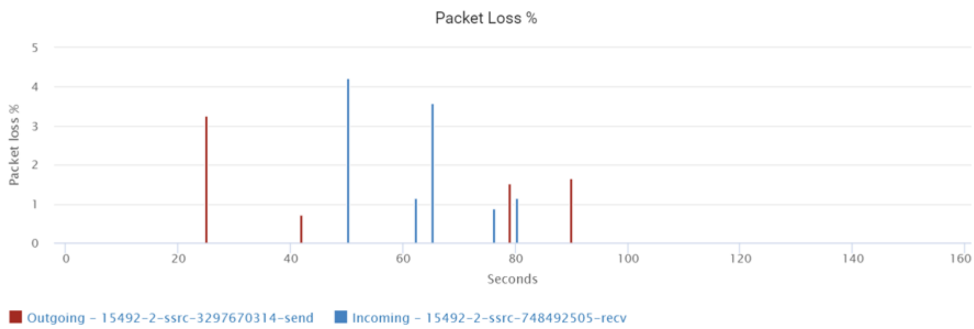

Figure 7.10: Outgoing and Incoming Video Packet Loss

was lost, the video data performed pretty well. It goes on to explain further that the video frame rate was fairly constant.

7.2.3. Additional Results

Table 7.4 indicates that the packet loss corresponding to the video stream in a video call is zero. The testrtc tool seemed to approximate the number of packets lost to zero as there weren't many packets lost. In order to inspect this and to also see if the duration of the calls had any impact over the results, the experiments were repeated. The results are grouped into two groups: Calls with duration upto 200s (on the left) and calls with duration upto 500s (on the right). Local end refers to the device on which the measurements are taken. Remote end refers to the device on the receiver side of the call. Furthermore, the calls have also been tested between a laptop and a mobile phone.

Consider the case of a voice call whose packet loss information is presented in Figures 7.11 - 7.14. AudIn represents the Incoming Voice RTP stream. The outgoing audio RTP stream with respect to one source is same as the Remote Incoming RTP audio stream (RemAudIn). RemAudIn represents the stats relevant to the remote receiving end. Data from the remote end is received via RTCP Receiver Report (RR) [57]. Local and remote data are combined to compute the stats. R1-R12 represents the Run/ trial number. L1/L2 represents the laptop on which the measurement was taken. M1 represents the mobile device on which the measurements are taken.

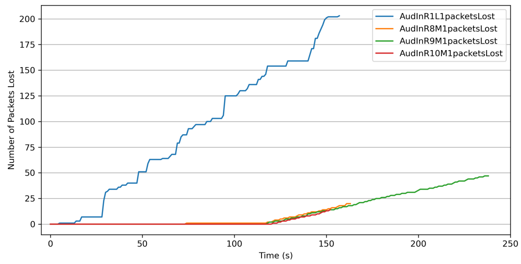

Figure 7.11: Packets lost - voice call duration < 200s

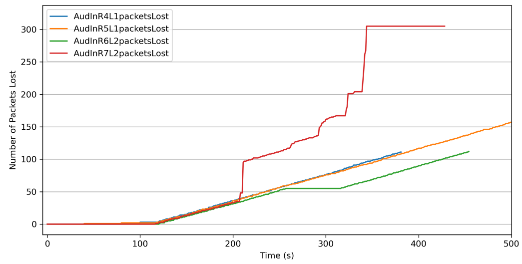

Figure 7.12: Packets lost - voice call duration < 500s

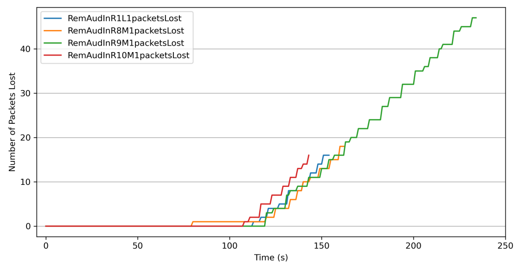

Figure 7.13: Packets lost - remote end

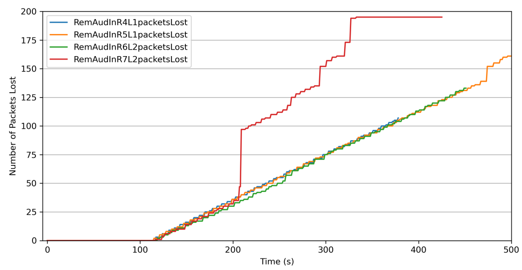

Figure 7.14: Packets lost - remote end

Figure 7.11 and Figure 7.12 indicate that fewer packets are lost during the first 100 seconds of the call and gradually begins to increase after it. Similar trend is also presented in Figure 7.13 and Figure 7.14. As the time increases, the number of RTP packets lost also increases. Ideally, packet loss value should be as low as possible. However, the WNC used for testing acts like a hotspot. As a result, the laptop/mobile phone is connected over the Wifi. This could be one of the reasons that the number of packets lost increases. During the testing, the WNC device sometimes showed no connectivity and sometimes full strength. This kind of fluctuation might have also impacted the call. A wired connection or improved hardware might solve the problem or improve the performance. In Chapter 6, it is already explained that the PDN connection used to transport the RTP data is over a standard internet connection. With longer duration of the call, the congestion would probably be more resulting in more number of packets being lost.

Next, consider the case of a video call. Video call includes voice and video RTP streams. We have AudIn and RemAudIn streams representing the voice RTP data. VidIn and RemVidIn represent the video RTP stream. Results are grouped into calls with duration upto 200 seconds and calls with duration upto 500 seconds. In Figures 7.19 - 7.22, the voice stream in the video call follows a similar trend as that in a normal voice call. The packet loss value increases with time. The value is not zero but still lower than the voice streams. The number of video packets lost during longer calls is less than the calls with shorter time duration. Similar observation can be made with the data recorded at the remote end.

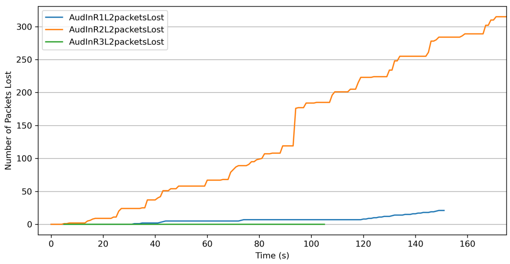

Figure 7.15: Voice Packets lost - video call duration < 200s

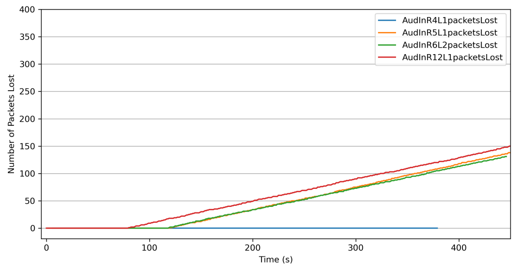

Figure 7.16: Voice Packets lost - video call duration < 500s

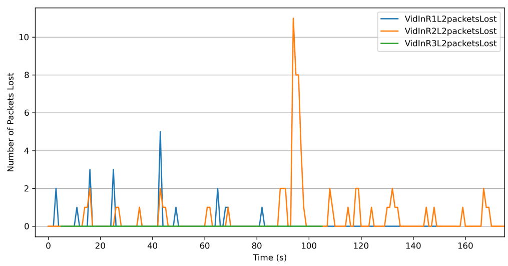

Figure 7.17: Video Packets lost - video call duration < 200s

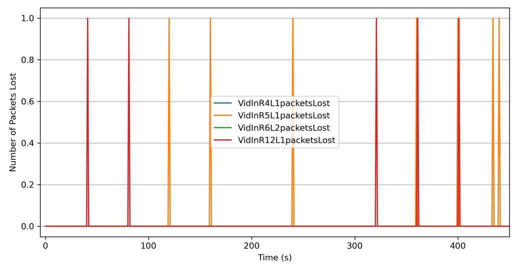

Figure 7.18: Video Packets lost - video call duration < 500s

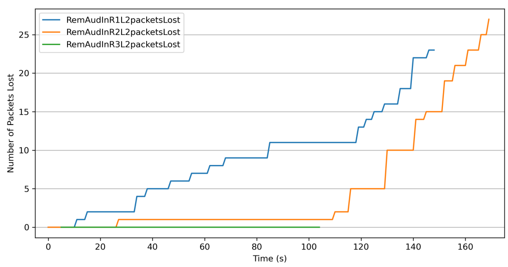

Figure 7.19: Voice Packets lost - remote end

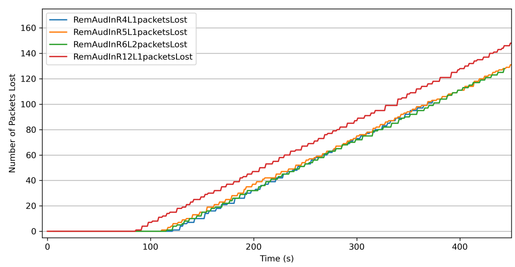

Figure 7.20: Voice Packets lost - remote end

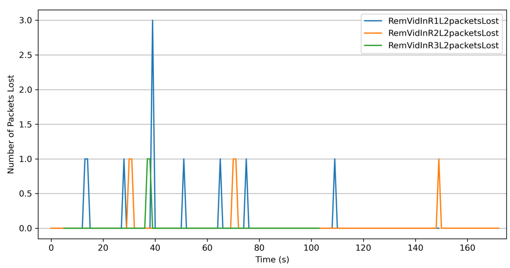

Figure 7.21: Voice Packets lost - remote end

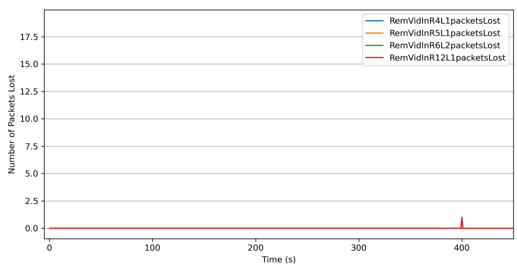

Figure 7.22: Video Packets lost - remote end

7.3. Jitter

During a voice call, the transmission of packets might result in some packets being lost or some packets arriving in an incorrect order. The result is a jumbled conversation to the user who is listening. Jitter is the variation in the delay of the packets sent between the same set of sender and receiver [58]. As the definition suggests, jitter is dependent on latency (explained in Section 7.4) [58].

If the jitter value is low, the user might not even notice the jumbled packets in the conversation. Call quality will degrade as the jitter value increases. Jitter is impacted by a number of factors, some of them being the hardware used, mode of connectivity (wireline or wireless) and congestion. High jitter values often results in Packet loss (refer Section 7.2) [59].

Calculation of Jitter is explained as follows:
Si is the RTP timestamp from packet i and Ri is the time of arrival in RTP Timestamp units for packet i, then for two packets i and j, the difference, D is calculated as [56]:

$$D(i,j) = (R_j - R_i) - (S_j - S_i) = (R_j - S_j) - (R_i - S_i) \tag{7.2}$$

The inter-arrival jitter, J, should be calculated continuously as each data packet i is received from source SSRC_n, using this difference D for that packet and the previous packet i-1 in order of arrival (not necessarily in sequence), according to the formula :

$$J(i) = J(i-1) + (|D(i-1,i)| - J(i-1))/16 \tag{7.3}$$

7.3.1. Voice Call

In the setup used for the experiment, the values of jitter are quite low indicating a good quality of the voice call. This low jitter value could be because there is not a lot of traffic on the network. The only two users connected to the network are the two laptops used to make the call. So essentially there is the data traffic of just one conversation in the whole network at least on the RAN side. Between the EPC and WCG, some traffic could be expected because the voice packets are being transmitted over a PDU connection meant for general internet services called "Internet PDN" (presented in Figure 7.23). This is because the WCG is located remotely and this is the only available way to access it from the internet. As a result, there is no dedicated PDU session (with a signalling bearer and a voice bearer) for IMS related communication. The voice data traverses through Internet PDN connection. So the jitter values recorded could represent the effect over the voice traffic in the link between the EPC and the WCG. Table 7.5 presents the average values of jitter over three trials of the voice call over the same network.

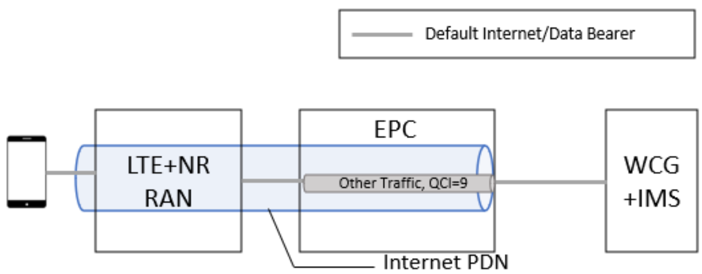

Figure 7.23: Internet PDN connecting the EPC and WCG

All values in ms	Run 1	Run 2	Run 3
Incoming jitter	14	8	7
Outgoing jitter	11	7	8

Table 7.5: Average jitter values for voice call over three tests

The graphical representation of the jitter variation over the duration of the voice call is represented in Figure 7.24.

7.3.2. Video Call

The jitter calculation for video call is similar to that of a voice call. Here, the packets carrying the encoded video also contribute to the results along with voice packets. RTP is the protocol that carries this information just as explained in the previous sections. Table 7.6 represents the average jitter values obtained over the duration of the video call in three trials. It is notable that audio packets experience

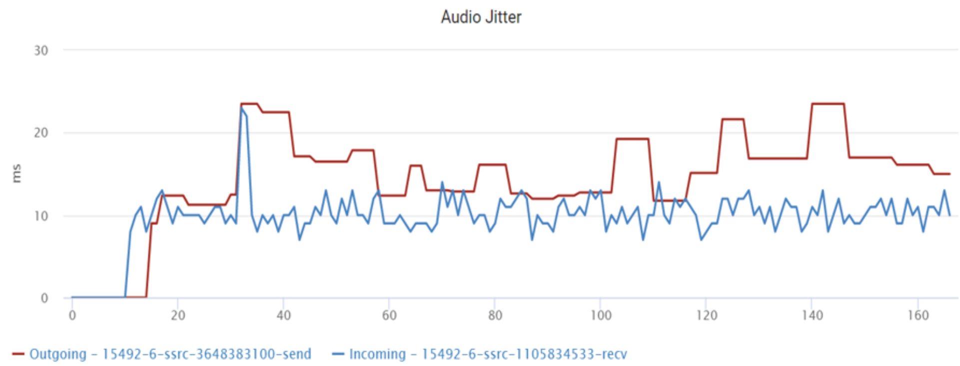

Figure 7.24: Outgoing and Incoming Voice Jitter

more jitter in comparison to the video stream. Jitter is measured at the receiver end. In general, the occurrence of jitter is likely due to a congested network or some form of (electromagnetic) interference or the synchronization signals being corrupted.

All values in ms	Run 1	Run 2	Run 3
Incoming Audio	10	16	15
Outgoing Audio	12	13	15
Incoming Video	0	0	0
Outgoing Video	17	32	47

Table 7.6: Average jitter values for Video call over three tests

The values recorded above are the average values over the duration of the call. As mentioned in the previous sections, an Internet PDN is established between the 5G EPC to the IMS core network. This could be the reasons for these values of jitter in case of a video call. Similar to the case of the voice call, jitter values below 30ms is acceptable. Any values greater than this will be noticeable by the end user. The video might appear choppy or you might receive a sudden burst of frames all at once which might make things illegible. The graphs for the voice and video jitter variation over the duration of a call is shown in Figures 7.25 and 7.26 respectively. The values for an incoming video jitter being zero was traced back to a bug reported in Google Chrome. It causes incorrect reporting of the value.

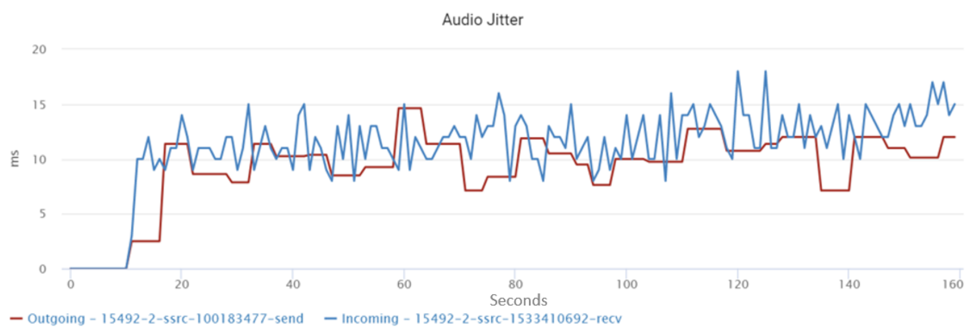

Figure 7.25: Outgoing and Incoming Voice Jitter of one call

Jitter can be reduced to some extent by the use of buffers at the receiver. Buffers tend to smooth out the missing packets in order to make the video call smooth. Modifying the buffer size plays a crucial role in handling both jitter and end-to-end delay values. High jitter will demand a longer buffer size to reorder the packets. Low jitter will require a smaller buffer size which will in turn, reduce the E2E delay. Congestion is another factor that impacts the latency and in turn jitter. So reducing the traffic over the same network by reducing the number of devices using the network can improve the jitter during the call.

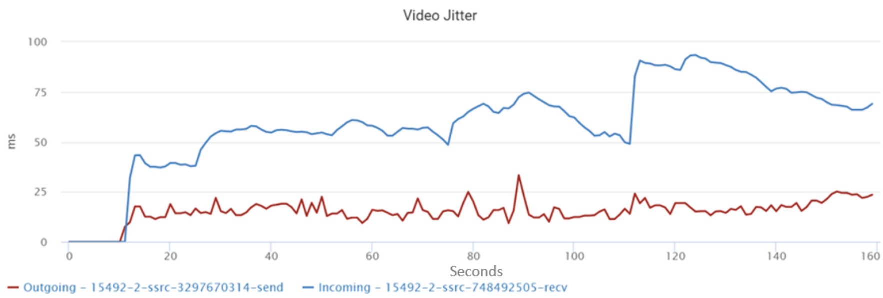

Figure 7.26: Outgoing and Incoming Video Jitter of one call

7.3.3. Additional Results

The voice and video call tests were repeated to see if the duration of the call had an impact on the parameters that are measured. The jitter values for call durations are presented similar to Section 7.2.3. The terminology remains the same. Figures 7.27 - 7.30 represent the jitter for voice call. The variation in the delay is approximately between 5ms and 20ms. For longer duration of the call, shows a higher value. The delay variation is between 5ms and 50ms. At the remote end, the results show a similar trend. The voice call was still quite clear.

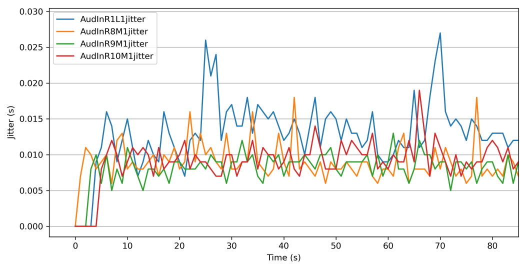

Figure 7.27: Jitter - voice call duration < 200s

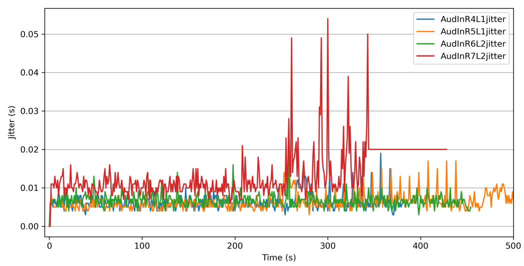

Figure 7.28: Jitter - voice call duration < 500s

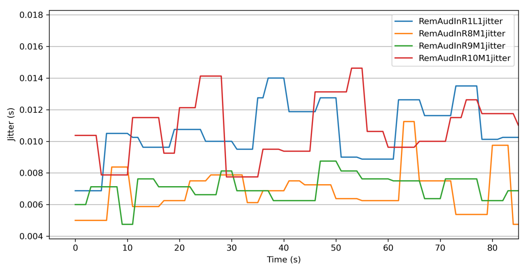

Figure 7.29: Jitter - voice call remote end

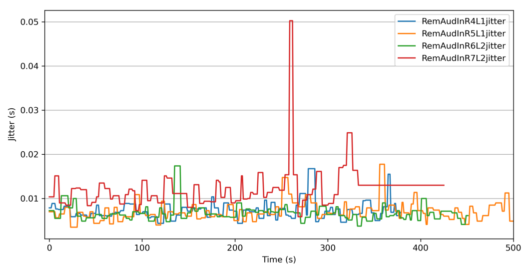

Figure 7.30: Jitter - voice call remote end

Figures 7.31 - 7.36 represent the delay variation /jitter for the video call. The voice streams follow a similar trend as a normal voice call. Figure 7.32 presents a different trend. One of the runs (AudInR4L1jitter) depicts an increasing variation in the delay while the other runs seem to perform similar to the calls of shorter duration. This was not expected. It could be caused by faulty connectivity to the NSA network. The VidIn jitter values are not presented here due to a bug as mentioned in Section 7.3.2. On the remote end, jitter seems to be fairly constant for both shorter and longer duration of calls.

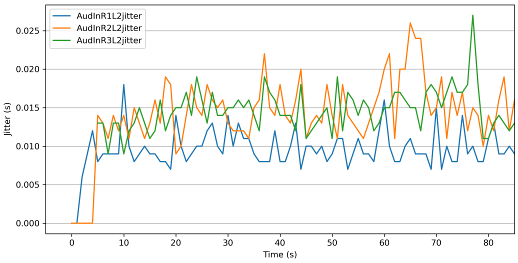

Figure 7.31: Voice Jitter - video call duration < 200s

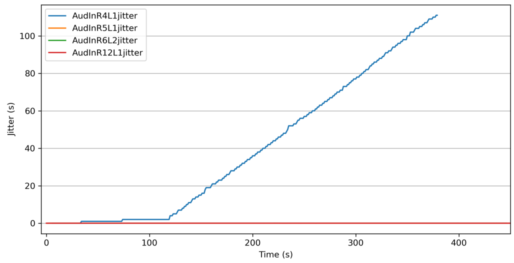

Figure 7.32: Voice Jitter - video call duration < 500s

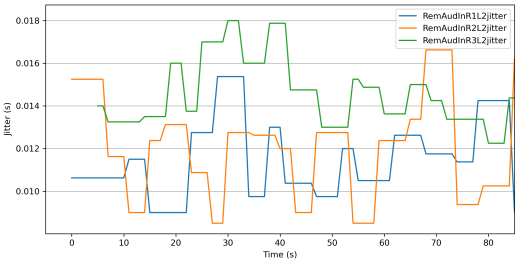

Figure 7.33: Voice Jitter - remote end

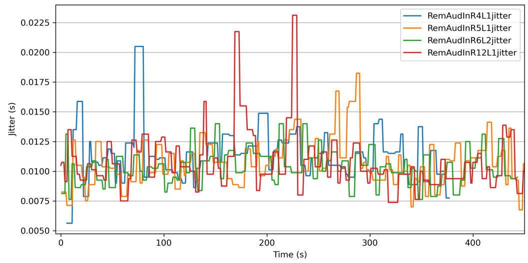

Figure 7.34: Voice Jitter - remote end

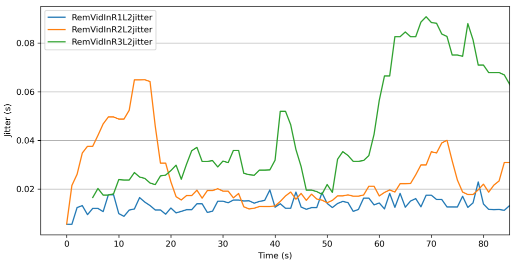

Figure 7.35: Video Jitter - remote end

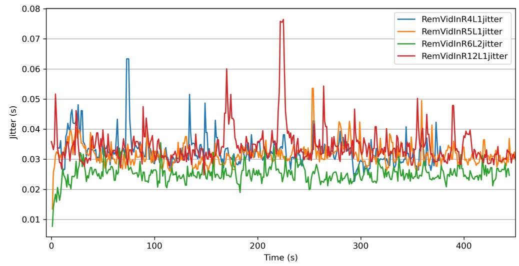

Figure 7.36: Video Jitter - remote end

7.4. Latency

Latency in networks is defined as the time taken by the data packet to reach its destination. Latency can be categorized into different types [60]:

- Round Trip Latency : Time taken by a packet to reach the destination and traverse back to the source.
- End-to-End (E2E) Latency: Time taken by the packet to reach the destination from the source.

Latency is also measured in different parts of the network. RAN latency is the time taken for the packet to reach the gNodeB/eNodeB from the device. Core network latency is time taken by the packet to travel from the nodeB to the core network. In the case of this experiment, there is also some delay expected between the 5G EPC and the IMS core. Different command line tools are used to test the network connectivity. They are listed as follows:

- Ping: This command allows us to calculate the latency in the network. Depending on the destination, ping can be used to measure latency in different parts of the network. Figure 7.38 depicts the results for the ping tests between the device and an iperf server running at the EPC core. Whenever it is mentioned that the EPC is pinged, it is implied that the iperf server running in the core is pinged. The network overview is depicted in Figure 7.37. Ping test was also carried out

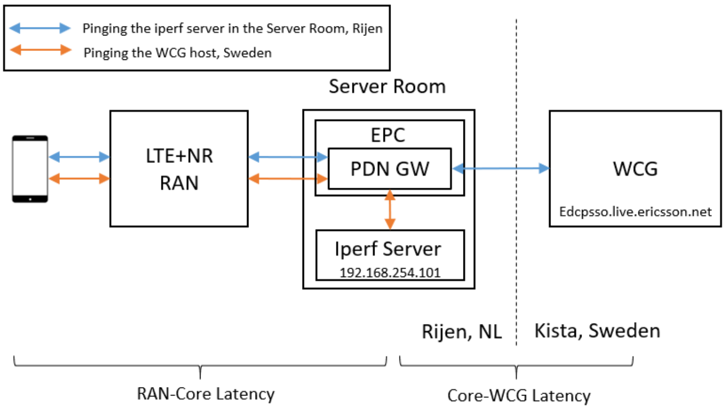

Figure 7.37: Network overview for command line tests

```
msmnile:/data # ping 192.168.254.101
ping 192.168.254.101
PING 192.168.254.101 (192.168.254.101) 56(84) bytes of data.
64 bytes from 192.168.254.101: icmp_seq=1 ttl=62 time=11.5 ms
64 bytes from 192.168.254.101: icmp_seq=2 ttl=62 time=18.7 ms
64 bytes from 192.168.254.101: icmp_seq=3 ttl=62 time=17.3 ms
64 bytes from 192.168.254.101: icmp_seq=4 ttl=62 time=8.90 ms
64 bytes from 192.168.254.101: icmp_seq=5 ttl=62 time=11.1 ms
64 bytes from 192.168.254.101: icmp_seq=6 ttl=62 time=10.8 ms
64 bytes from 192.168.254.101: icmp_seq=7 ttl=62 time=10.9 ms
64 bytes from 192.168.254.101: icmp_seq=8 ttl=62 time=10.8 ms
64 bytes from 192.168.254.101: icmp_seq=9 ttl=62 time=9.49 ms
64 bytes from 192.168.254.101: icmp_seq=10 ttl=62 time=18.9 ms
64 bytes from 192.168.254.101: icmp_seq=11 ttl=62 time=20.4 ms
64 bytes from 192.168.254.101: icmp_seq=12 ttl=62 time=17.3 ms
64 bytes from 192.168.254.101: icmp_seq=13 ttl=62 time=17.0 ms
64 bytes from 192.168.254.101: icmp_seq=14 ttl=62 time=16.3 ms
64 bytes from 192.168.254.101: icmp_seq=15 ttl=62 time=18.2 ms
64 bytes from 192.168.254.101: icmp_seq=16 ttl=62 time=11.0 ms
64 bytes from 192.168.254.101: icmp_seq=17 ttl=62 time=18.3 ms
64 bytes from 192.168.254.101: icmp_seq=18 ttl=62 time=12.3 ms
64 bytes from 192.168.254.101: icmp_seq=19 ttl=62 time=18.8 ms
64 bytes from 192.168.254.101: icmp_seq=20 ttl=62 time=17.1 ms
64 bytes from 192.168.254.101: icmp_seq=21 ttl=62 time=18.3 ms
64 bytes from 192.168.254.101: icmp_seq=22 ttl=62 time=10.8 ms
64 bytes from 192.168.254.101: icmp_seq=23 ttl=62 time=9.14 ms
64 bytes from 192.168.254.101: icmp_seq=24 ttl=62 time=16.4 ms
64 bytes from 192.168.254.101: icmp_seq=25 ttl=62 time=17.0 ms
64 bytes from 192.168.254.101: icmp_seq=26 ttl=62 time=17.2 ms
64 bytes from 192.168.254.101: icmp_seq=27 ttl=62 time=11.4 ms
64 bytes from 192.168.254.101: icmp_seq=28 ttl=62 time=17.2 ms
64 bytes from 192.168.254.101: icmp_seq=29 ttl=62 time=15.4 ms
64 bytes from 192.168.254.101: icmp_seq=30 ttl=62 time=17.9 ms
64 bytes from 192.168.254.101: icmp_seq=31 ttl=62 time=18.8 ms
64 bytes from 192.168.254.101: icmp_seq=32 ttl=62 time=18.3 ms
64 bytes from 192.168.254.101: icmp_seq=33 ttl=62 time=17.7 ms
64 bytes from 192.168.254.101: icmp_seq=34 ttl=62 time=24.5 ms
64 bytes from 192.168.254.101: icmp_seq=35 ttl=62 time=18.4 ms
64 bytes from 192.168.254.101: icmp_seq=36 ttl=62 time=17.8 ms
64 bytes from 192.168.254.101: icmp_seq=37 ttl=62 time=17.3 ms
64 bytes from 192.168.254.101: icmp_seq=38 ttl=62 time=18.2 ms
64 bytes from 192.168.254.101: icmp_seq=39 ttl=62 time=19.3 ms
64 bytes from 192.168.254.101: icmp_seq=40 ttl=62 time=18.4 ms
```

Figure 7.38: Ping test from device to EPC

to see the delay between the device and the IMS core network. vWA is the entry point for IMS core network in our setup, so the ping packets were sent out to the virtual Web Access (vWA)/ Web Communication Gateway (WCG) to see the one way delay. By default, the ping command sends out 4 data packets to the destination and consolidates the results. This is exactly what is observed in the Figure 7.39. This command also presents the Round Trip Times and the packet

```
C:\Users\edtaamn>ping edcpsso.live.ericsson.net

Pinging edcpsso.live.ericsson.net [153.88.74.15] with 32 bytes of data:
Reply from 153.88.74.15: bytes=32 time=108ms TTL=235
Reply from 153.88.74.15: bytes=32 time=866ms TTL=235
Reply from 153.88.74.15: bytes=32 time=667ms TTL=235
Reply from 153.88.74.15: bytes=32 time=104ms TTL=235

Ping statistics for 153.88.74.15:
    Packets: Sent = 4, Received = 4, Lost = 0 (0% loss),
Approximate round trip times in milli-seconds:
    Minimum = 104ms, Maximum = 866ms, Average = 436ms
```

Figure 7.39: Ping test from device to the WCG

loss percentages for the connection. The average RTT value recorded is 436ms with a zero percent packet loss.

- Iperf: This tool returns the maximum bandwidth that can be achieved over a link that is sending/receiving TCP or UDP packets [61]. Iperf command was run for two instances, device to EPC core (Figure 7.40) and device to device (Figure 7.41) The average throughput achieved in these

```
msmnile:/data # ./iperf3 -c 192.168.254.101 -p 5201 -R -t 30
./iperf3 -c 192.168.254.101 -p 5201 -R -t 30
Connecting to host 192.168.254.101, port 5201
Reverse mode, remote host 192.168.254.101 is sending
[ 4] local 10.31.0.26 port 41028 connected to 192.168.254.101 port 5201
[ ID] Interval        Transfer     Bandwidth
[ 4]  0.00-1.00   sec  67.5 MBytes  566 Mbits/sec
[ 4]  1.00-2.00   sec  82.3 MBytes  690 Mbits/sec
[ 4]  2.00-3.00   sec  81.5 MBytes  682 Mbits/sec
[ 4]  3.00-4.00   sec  81.8 MBytes  689 Mbits/sec
[ 4]  4.00-5.00   sec  81.7 MBytes  686 Mbits/sec
[ 4]  5.00-6.00   sec  80.8 MBytes  678 Mbits/sec
[ 4]  6.00-7.00   sec  81.1 MBytes  679 Mbits/sec
[ 4]  7.00-8.00   sec  82.3 MBytes  691 Mbits/sec
[ 4]  8.00-9.00   sec  82.4 MBytes  691 Mbits/sec
[ 4]  9.00-10.00  sec  82.3 MBytes  690 Mbits/sec
[ 4] 10.00-11.00  sec  82.3 MBytes  690 Mbits/sec
[ 4] 11.00-12.00  sec  82.0 MBytes  687 Mbits/sec
[ 4] 12.00-13.00  sec  82.1 MBytes  689 Mbits/sec
[ 4] 13.00-14.00  sec  82.6 MBytes  691 Mbits/sec
[ 4] 14.00-15.00  sec  82.0 MBytes  689 Mbits/sec
[ 4] 15.00-16.00  sec  81.9 MBytes  688 Mbits/sec
[ 4] 16.00-17.00  sec  82.6 MBytes  693 Mbits/sec
[ 4] 17.00-18.00  sec  82.5 MBytes  692 Mbits/sec
[ 4] 18.00-19.00  sec  82.2 MBytes  688 Mbits/sec
[ 4] 19.00-20.00  sec  81.4 MBytes  684 Mbits/sec
[ 4] 20.00-21.00  sec  80.9 MBytes  677 Mbits/sec
[ 4] 21.00-22.00  sec  80.4 MBytes  676 Mbits/sec
[ 4] 22.00-23.00  sec  81.7 MBytes  686 Mbits/sec
[ 4] 23.00-24.00  sec  81.2 MBytes  682 Mbits/sec
[ 4] 24.00-25.00  sec  82.5 MBytes  691 Mbits/sec
[ 4] 25.00-26.00  sec  81.9 MBytes  688 Mbits/sec
[ 4] 26.00-27.00  sec  80.1 MBytes  672 Mbits/sec
[ 4] 27.00-28.00  sec  82.3 MBytes  691 Mbits/sec
[ 4] 28.00-29.00  sec  81.9 MBytes  687 Mbits/sec
[ 4] 29.00-30.00  sec  82.6 MBytes  693 Mbits/sec
- - - - - - - - - - - - - - - - - - - - - - - - -
[ ID] Interval        Transfer     Bandwidth       Retr
[ 4]  0.00-30.00  sec  2.39 GBytes  683 Mbits/sec    0          sender
[ 4]  0.00-30.00  sec  2.39 GBytes  683 Mbits/sec               receiver

iperf Done.
```

Figure 7.40: Iperf from device to EPC (iperf server)

iperf tests between the device and the EPC located in Building 2 is approximately 683Mbps. The cable connecting the eNodeB to the core is a 1 Gbps link. Observing the results in Figure 7.40, the value is significantly lower. Upon further investigation, it was found that the internet bearer creates a bottleneck because of which the throughput is lower than the expected value.

```
C:\Users\edtaamn\Downloads\iperf-3.1.3-win64>iperf3.exe -c 192.168.43.28 -p 5201
Connecting to host 192.168.43.28, port 5201
[  4] local 192.168.43.192 port 64111 connected to 192.168.43.28 port 5201
[ ID] Interval           Transfer     Bandwidth
[  4]   0.00-1.00   sec  4.50 MBytes  37.7 Mbits/sec
[  4]   1.00-2.00   sec  4.25 MBytes  35.6 Mbits/sec
[  4]   2.00-3.00   sec  4.38 MBytes  36.6 Mbits/sec
[  4]   3.00-4.00   sec  4.38 MBytes  36.8 Mbits/sec
[  4]   4.00-5.00   sec  4.25 MBytes  35.7 Mbits/sec
[  4]   5.00-6.00   sec  4.38 MBytes  36.5 Mbits/sec
[  4]   6.00-7.00   sec  4.38 MBytes  36.8 Mbits/sec
[  4]   7.00-8.00   sec  4.38 MBytes  36.7 Mbits/sec
[  4]   8.00-9.00   sec  4.25 MBytes  35.7 Mbits/sec
[  4]   9.00-10.00  sec  4.50 MBytes  37.7 Mbits/sec
- - - - - - - - - - - - - - - - - - - - - - - - -
[ ID] Interval           Transfer     Bandwidth
[  4]   0.00-10.00  sec  43.6 MBytes  36.6 Mbits/sec                  sender
[  4]   0.00-10.00  sec  43.6 MBytes  36.6 Mbits/sec                  receiver

iperf Done.
```

Figure 7.41: Iperf-device to device

Figure 7.41 gives the average throughput of about 36Mbps between the two devices that are connected to the network and are used to make the voice and video calls at a later stage. The E2E includes the connectivity between the device (making the call later), through the EPC, IMS core and again through the EPC towards the second device (receiving the call). Although, given that the RAN side is 5G NSA, the value is expected to be much higher and the latency is expected to be lower than that of a normal E-UTRA network. The connectivity from the EPC and IMS core is a standard internet PDN connection, which as mentioned previously influences the network performance and results in a much lower throughput.

7.4.1. Voice Call

Voice call setup is the same as that explained in the previous sections. Round trip latency graphs are obtained by the analysis of the WebRTC internals dump. Figure 7.42 shows an average value of approximately 150ms over the duration of the call.

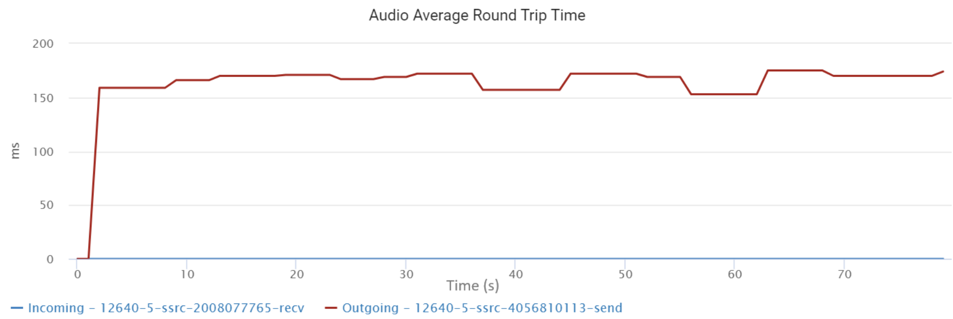

Figure 7.42: Outgoing and Incoming Audio RTT latency

For VoLTE, the maximum value of RTT can be about 300ms. 5G promises Ultra low latency (approximately 1ms latency) for different applications. This could be possible with a 5G Core at the center. In the experiment, we have a modified 4G EPC which is why lower values can be expected. Even so, the network seems to perform nicely with the achieved values [62].

7.4.2. Video Call

The results are obtained and presented in a similar fashion as the previous Sections 7.3 and 7.2. Video call involves both voice and video streams. Therefore there are two sets of graphs presenting the performance over the duration of a call in Figures 7.43 and 7.44. The values of RTT are approximately 200ms which is again very good. These values could be further improved by applying QCI restrictions over the PDN bearer connecting the EPC and the WCG. This would will improve the way the traffic is handled. As explained in Section 7.3.1, with the setup used for the experiments, it wouldn't make

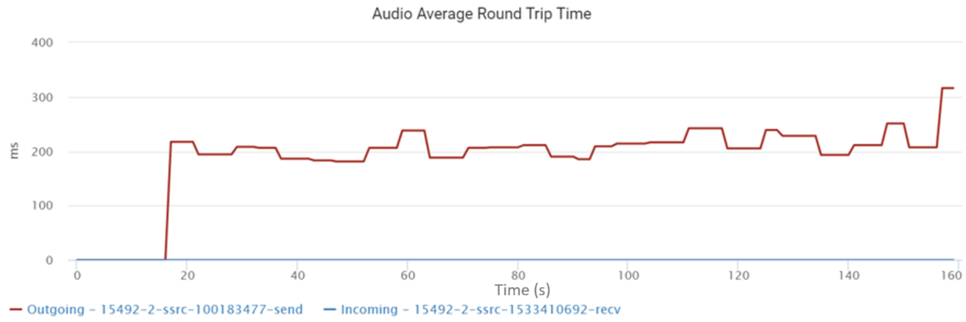

Figure 7.43: Outgoing and Incoming Audio Average RTT

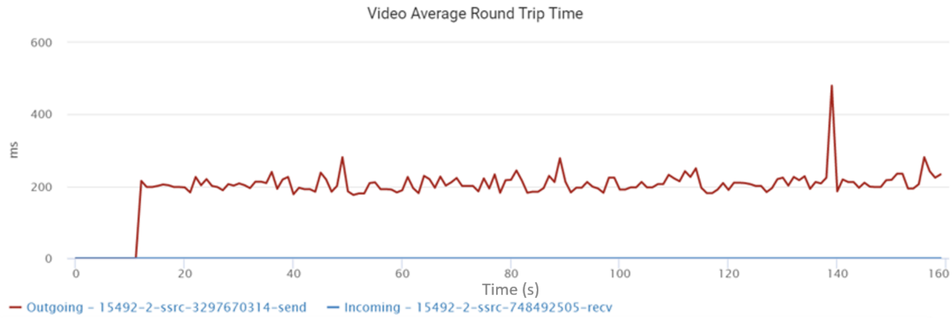

Figure 7.44: Outgoing and Incoming Video Average RTT

sense to enforce any QCI handling due to the presence of just two users in the whole network. With an increased number of connected devices, more traffic and higher congestion, enforcing QoS would improve the performance. A 5G Core and a dedicated connection to IMS (like in the case of VoLTE) would help enforce QoS restrictions on the call flows. Having the IMS located locally would also result a noticeable improvement in the performance during these calls. Implementing a 5G SA network is likely to meet the 5G expectations.

7.4.3. Additional results

Figure 7.45 and Figure 7.46 represent the latency in terms of RTT for one source. This value is dependent on the RTCP timestamps in the RR. It is presented in seconds (s) [56][57]. RTT is always measured over the Remote Inbound stream as per the definition. The average RTT is about 200ms. A lower value would be preferred. Testing over 5G SA core might improve these results.

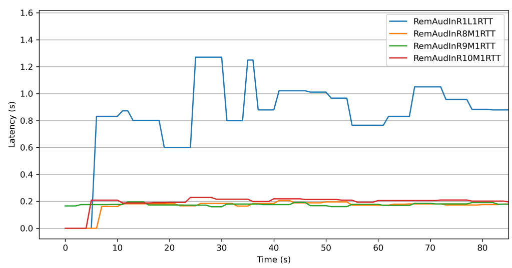

Figure 7.45: Voice RTT - video call duration < 200s

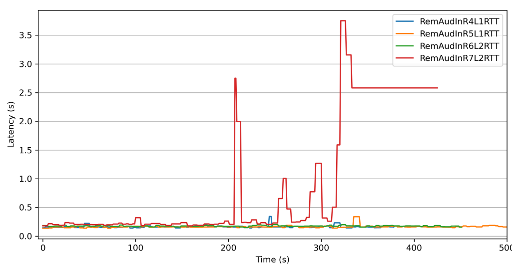

Figure 7.46: Voice RTT - video call duration < 500s

Figures 7.47 - 7.50 represent the variation in the RTT during a video call. In Figure 7.47, one of the tests indicate a much higher value than the rest. This is also reflected in Figure 7.49. This indicates that it is probably an issue with the bearer used for the call. The RTT values for the voice streams are below 40ms in both shorter and longer call duration. Video RTT in Figure 7.50 appears to be fluctuating a lot. This could be because of a large amount of data being transmitted over a long time.

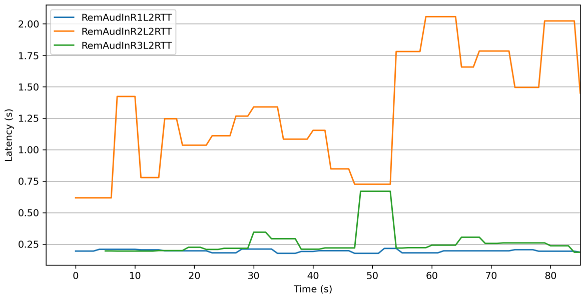

Figure 7.47: Voice RTT - remote end

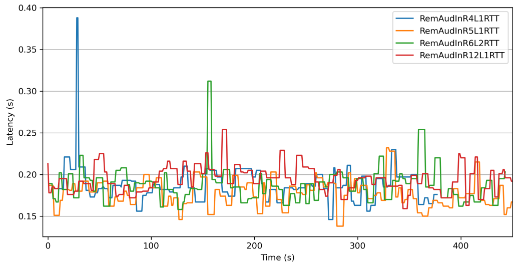

Figure 7.48: Voice RTT - remote end

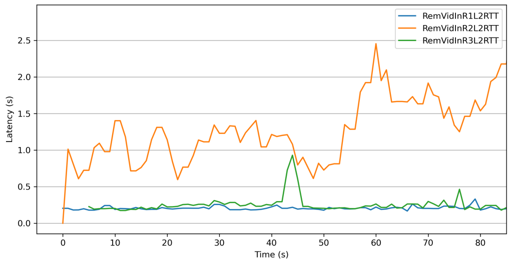

Figure 7.49: Video RTT - remote end

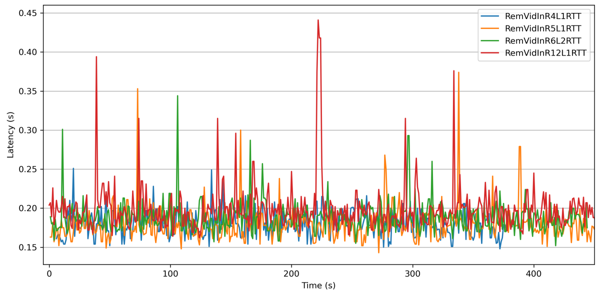

Figure 7.50: Video RTT - remote end

7.5. Instant Messaging

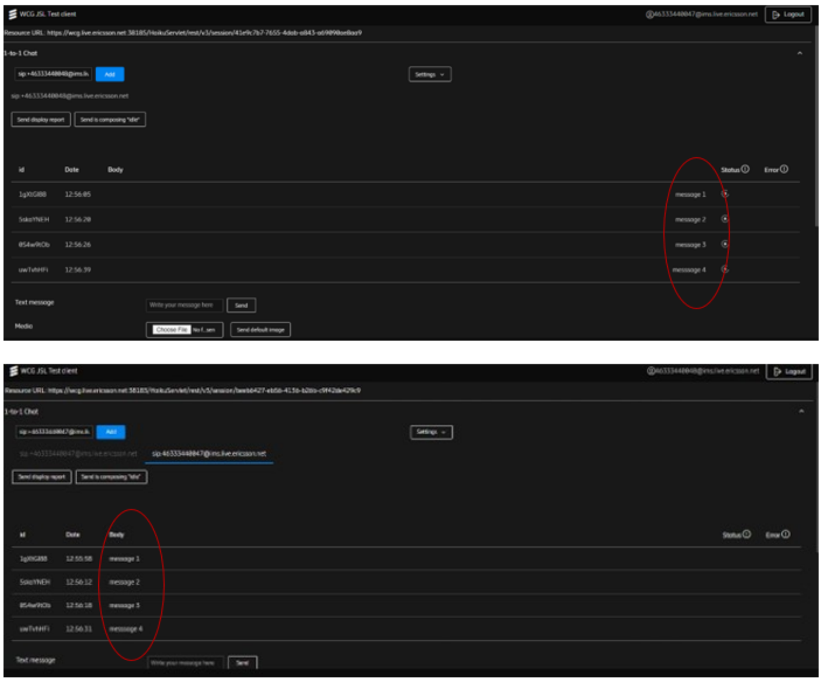

Figure 7.51: Snapshot of Instant Messaging service

Apart from voice and video calling, the WCG supports 1-to-1 chat option as well. Web Clients can send or receive messages to each other. A snap shot of the messaging between two users is shown in Figure 7.51. It can be noticed with the timestamp on the messages that the messages are instantaneous. Sometimes third party apps experience a delay while receiving messages from another user. IMS support in this situation would serve two objectives. It would eliminate the need for a third party app (a browser could be used with WebRTC) and IMS core would be able to ensure some QoS guarantees. Messaging doesn't always require a session to be set up explicitly. If the messages being sent occasionally, the message could be tagged along with the SIP Invite request, without the need to create a PDU session. In case of an extensive messaging application like live chat, would require a SIP session setup to serve the purpose. This messaging feature could also be improved to send a media message or a small file.

8 Conclusions and Future Work

- The path to implementation of a full 5G network is paved by the implementation of a 5G network that is dependent on an enhanced 4G EPC. While the implementation of a fully independent 5G SA network is underway,the 5G NSA will continue to cater to the demands of the industry in the mean time. This is what makes this thesis so relevant.
- Further more, with VoLTE, IMS has already proven its reliability in offering guaranteed QoS to the users. Adapting IMS to the 5G context would be the next step to aid the voice and video capabilities over 5G.
- IMS was initially designed to be flexible and be adaptable to the evolving mobile network. From the studies undertaken in this thesis, it can be concluded that IMS can be easily integrated with the 5G network whether it is NSA or SA. The major difference would be in the connectivity between the packet core network and the IMS core. In VoLTE, this is the link between the S-GW and P-CSCF. In 5G NSA, it would still be between S-Gw and P-CSCF, but the functions of the P-GW would be adapted to support 5G RAN. In 5G SA, it would be between the UPF and the P-CSCF.
- There are a lot of similarities between the 4G and 5G architecture. The functions remain the same in principle, but the elements carrying out these functions are new with new names. A comparative study of both these architectures puts both the technologies into perspective. It provides more clarity in a system where both these technologies co-exist.
- Technologies like WebRTC, SDN and Virtualisation will dictate how the 5G capabilities are utilised to the maximum extent. Voice and Video would be the most widely used services. With the promise of low latency and higher throughput, combining these technologies will enable the creation of a wide number of services.
- During the experiments, WebRTC proved to be a good technology to be combined with 5G. Enabling a simple browser to browser communication would completely eliminate the need for a third party app. Anybody with a browser installed on their phone would be able to place a voice or video call. The APIs are well documented and constantly updated.
- Voice, Video and Instant messaging were the IMS services that were tested during this thesis. The results from these are also be representative of the other IM services like Presence and file sharing. The experiment results prove that the IMS services can be integrated well with WebRTC and 5G.
- The number of devices connected over the network was just two. The goal of the thesis was to see how the IMS services would fare over the given network. As a result, the traffic conditions were far from practical values. Different QoS parameters were measured for the test. The biggest limiting factor was the link between the packet core network and IMS core. The standard PDN (internet) bearer was used which impacted the values of the parameters. Incorporating a dedicated connection between the mobile core and the packet core could improve the values achieved during the experiments.

- The bitrate values are in accordance with the codecs used indicating a good performance. Even the video call performs pretty well and the video bitrate is fairly constant. G722 and VP8 codecs are used because they are open source, backward compatible and approved by the IETF. The use of improvised codecs would definitely improve these values.
- Packet loss values obtained during the tests are very low indicating that the 5G NSA will be capable of providing a good call quality. When a video call was tested over the same network, the voice packet loss values increased slightly while the video packet loss values were nil. Such a situation with high packet loss for audio and low video packet loss would present a good video call with bad/ jumbled audio. It might not result in a pleasant experience for the users. Thus, it is important to keep these values as low as possible.
- Jitter values obtained are acceptable. If the value of jitter is very high, it can lead to a jumbled conversation as the arriving packets may not be arranged in the correct order.
- Latency is another important parameter that contributes to the call quality. Through the experiments, the values of latency recorded were lower than the maximum allowable range. As a result, the call quality was better than that of VoLTE. However, the ultra low latency promised by the 5G networks involving a 5G core which would further improve the values recorded here.
- The different parameter values achieved during the calls indicate that the 5G NSA conforms to the theoretical expectation. 5G NSA indeed performs better than the predecessor 4G network. IMS services do perform well in terms of QoS while using WebRTC for media plane negotiation. The quality can be further improved by upgrading to 5G core, using improved codecs, adaptable buffers and dedicated link between the mobile core and the IMS core. Implementing an IMS core network locally would be able to contribute to an improved performance of the IMS services.

A

Messaging Sequence for Sessions

A.1. IMS registration for an unregistered user

IMS registration procedures differ depending on the status of the user (registered/not registered). The Public User Identity (PUI) plays an important role in the registration procedure. Detailed explanation about the Registration process is given in [10].

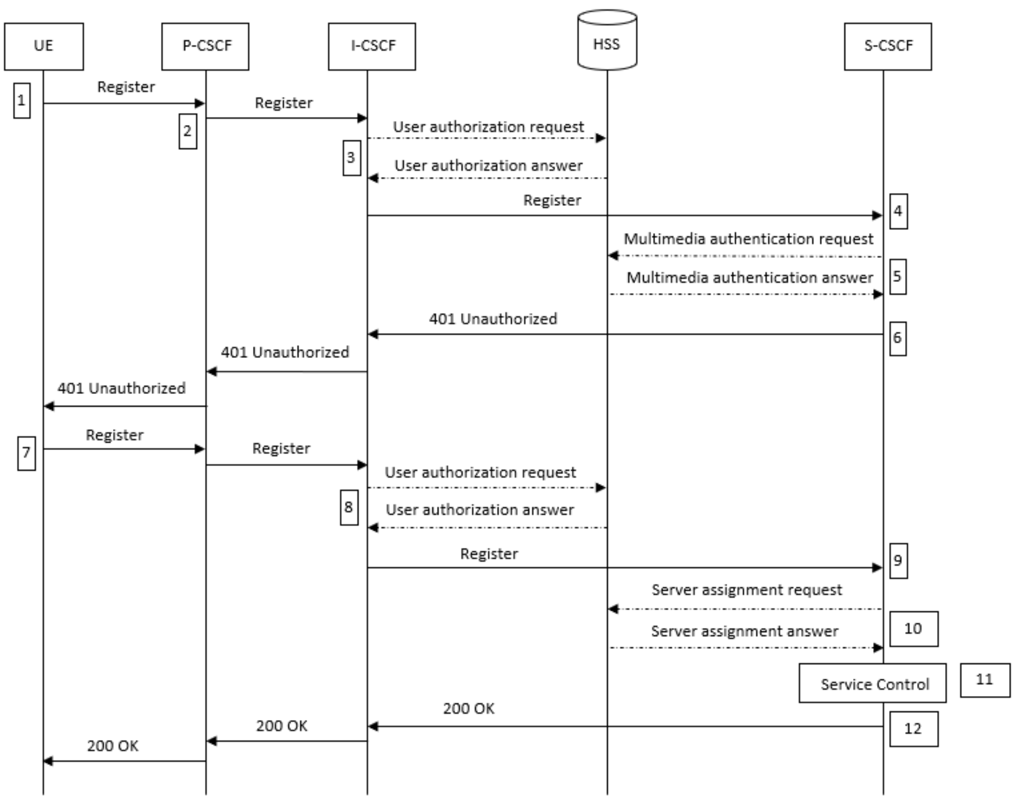

Figure A.1: IMS registration message sequence [9]

Assuming that the UE is already EPS attached, the sequence is presented in Figure A.1. The following are the different steps in the IMS registration process:

- 1 UE obtains the IP address of the P-CSCF either by DHCP or from the mobile network. IMPU, IMPI and the address of the home IMS is stored on the ISIM. SIP register request is constructed using the information obtained previously and sent to the P-CSCF. A Route header in the Register request determines the path.

- 2 P-CSCF forwards the SIP request to the I-CSCF which will choose the correct S-CSCF to forward this request to.
- 3 The I-CSCF contacts the HSS which verifies if the UE attempting to register is a subscriber or not. The HSS also checks the user subscription and also indicates if the user is allowed to register in that P-CSCF network. I-CSCF receives the capabilities from the HSS in response.
- 4 I-CSCF will choose the correct S-CSCF and forward information like Public User Identity, Private User Identity and UE-IP address. The S-CSCF stores the P-CSCF network ID information.
- 5 The S-CSCF then sends a multimedia authentication request to the HSS. The HSS stores the address of the S-CSCF and respond with an answer.
- 6 If the MAR in Step 5 fails, then a 401 UNAUTHORIZED messaged is relayed back to the UE.
- 7-9 The UE re-initiates the registration process.
- 10 The S-CSCF will send a server assignment request to the HSS which will respond with an server assignment answer.
- 11 The S-CSCF shall send the register information to the service control platform and apply the relevant service control procedures based on the filter criteria.
- 12 The S-CSCF message sends a 200 OK indicating to the UE that the IMS registration process is complete.

A.2. IMS De-registration initiated by the UE

IMS de-registration can be either network-initiated or UE-initiated. Network-initiated de-registration could be due to network maintenance, network/traffic flow, application layer parameters or subscription management. The procedures vary depending on the reason for de-registration. Detailed procedures are explained in [10]. UE-initiated de-registration occurs when the UE wants to de-register from IMS. The message sequence followed is similar to the IMS registration procedure. Figure A.2 presents the de-registration flow. UE-initiated de-registration flow has the following steps:

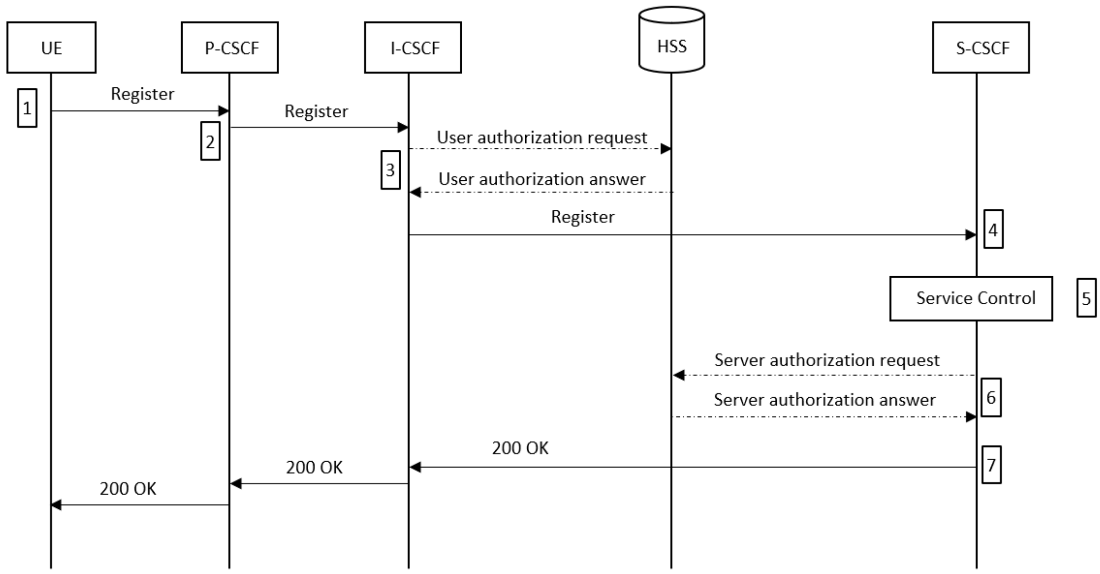

Figure A.2: IMS de-registration message sequence [9]

- 1-2 The de-registration process is initiated by the UE. A Register request with an expiration value of seconds is sent by the UE to the P-CSCF and then forwarded to the I-CSCF.

- 3 I-CSCF queries the HSS using the information it receives from the P-CSCF. The HSS determines that the Public User Identity (PUI) is registered sends a response to the I-CSCF.
- 4 The I-CSCF will forward the de-register information flow to the S-CSCF.
- 5 Depending on the filter criteria, the S-CSCF will apply service control procedures using the de-registration information received from the I-CSCF. All the subscription-related information relating to the particular PUI is removed.
- 6 S-CSCF notifies the HSS that the PUI is no longer registered with an SAR (Type:de-register) and receives a response, SAA (type:de-register). The HSS then clears or keeps the S-CSCF name for the PUI depending on the request.
- 7 S-CSCF sends a 200 OK message to the I-CSCF and may later release all the registration information regarding the specific registration of the PUI. I-CSCF forwards the 200 OK message to the P-CSCF, which also releases all the registration information with respect to the PUI. P-CSCF then forwards the 200 OK message to the UE, completing the de-registration process.

B

QCI Values and Mapping

B.1. QCI Class Identifiers -VoLTE

QCI	Resource Type	Priority Level	Packet Delay Budget (NOTE 13)	Packet Error Loss Rate (NOTE 2)	Example Services
1 (NOTE 3)	GBR	2	100 ms (NOTE 1, NOTE 11)	10^{-2}	Conversational Voice
2 (NOTE 3)		4	150 ms (NOTE 1, NOTE 11)	10^{-3}	Conversational Video (Live Streaming)
3 (NOTE 3, NOTE 14)		3	50 ms (NOTE 1, NOTE 11)	10^{-3}	Real Time Gaming, V2X messages Electricity distribution - medium voltage (e.g. TS 22.261 [51] clause 7.2.2) Process automation - monitoring (e.g. TS 22.261 [51] clause 7.2.2)
4 (NOTE 3)		5	300 ms (NOTE 1, NOTE 11)	10^{-6}	Non-Conversational Video (Buffered Streaming)
65 (NOTE 3, NOTE 9, NOTE 12)		0.7	75 ms (NOTE 7, NOTE 8)	10^{-2}	Mission Critical user plane Push To Talk voice (e.g., MCPTT)
66 (NOTE 3, NOTE 12)		2	100 ms (NOTE 1, NOTE 10)	10^{-2}	Non-Mission-Critical user plane Push To Talk voice
67 (NOTE 3, NOTE 12)		1.5	100 ms (NOTE 1, NOTE 10)	10^{-3}	Mission Critical Video user plane
75 (NOTE 14)		2.5	50 ms (NOTE 1)	10^{-2}	V2X messages
71		5.6	150ms (NOTE 1, NOTE 16)	10^{-6}	"Live" Uplink Streaming (e.g. TS 26.238 [53])
72		5.6	300ms (NOTE 1, NOTE 16)	10^{-4}	"Live" Uplink Streaming (e.g. TS 26.238 [53])
73		5.6	300ms (NOTE 1, NOTE 16)	10^{-8}	"Live" Uplink Streaming (e.g. TS 26.238 [53])
74		5.6	500ms (NOTE 1, NOTE 16)	10^{-8}	"Live" Uplink Streaming (e.g. TS 26.238 [53])
76		5.6	500ms (NOTE 1, NOTE 16)	10^{-4}	"Live" Uplink Streaming (e.g. TS 26.238 [53])
5 (NOTE 3)	Non-GBR	1	100 ms (NOTE 1, NOTE 10)	10^{-6}	IMS Signalling
6 (NOTE 4)		6	300 ms (NOTE 1, NOTE 10)	10^{-6}	Video (Buffered Streaming) TCP-based (e.g., www, e-mail, chat, ftp, p2p file sharing, progressive video, etc.)
7 (NOTE 3)		7	100 ms (NOTE 1, NOTE 10)	10^{-3}	Voice, Video (Live Streaming) Interactive Gaming
8 (NOTE 5)		8	300 ms (NOTE 1)	10^{-6}	Video (Buffered Streaming) TCP-based (e.g., www, e-mail, chat, ftp, p2p file sharing, progressive video, etc.)
9 (NOTE 6)		9			
69 (NOTE 3,		0.5	60 ms (NOTE 7,	10^{-6}	Mission Critical delay sensitive signalling (e.g., MC-PTT signalling, MC Video signalling)

Figure B.1: QCI VoLTE- part 1 [40]

NOTE 9, NOTE 12)			NOTE 8)		
70 (NOTE 4, NOTE 12)		5.5	200 ms (NOTE 7, NOTE 10)	10^{-6}	Mission Critical Data (e.g. example services are the same as QCI 6/8/9)
79 (NOTE 14)		6.5	50 ms (NOTE 1, NOTE 10)	10^{-2}	V2X messages
80 (NOTE 3)		6.8	10 ms (NOTE 10, NOTE 15)	10^{-6}	Low latency eMBB applications (TCP/UDP-based); Augmented Reality

NOTE 1: A delay of 20 ms for the delay between a PCEF and a radio base station should be subtracted from a given PDB to derive the packet delay budget that applies to the radio interface. This delay is the average between the case where the PCEF is located "close" to the radio base station (roughly 10 ms) and the case where the PCEF is located "far" from the radio base station, e.g. in case of roaming with home routed traffic (the one-way packet delay between Europe and the US west coast is roughly 50 ms). The average takes into account that roaming is a less typical scenario. It is expected that subtracting this average delay of 20 ms from a given PDB will lead to desired end-to-end performance in most typical cases. Also, note that the PDB defines an upper bound. Actual packet delays - in particular for GBR traffic - should typically be lower than the PDB specified for a QCI as long as the UE has sufficient radio channel quality.

NOTE 2: The rate of non congestion related packet losses that may occur between a radio base station and a PCEF should be regarded to be negligible. A PELR value specified for a standardized QCI therefore applies completely to the radio interface between a UE and radio base station.

NOTE 3: This QCI is typically associated with an operator controlled service, i.e., a service where the SDF aggregate's uplink / downlink packet filters are known at the point in time when the SDF aggregate is authorized. In case of E-UTRAN this is the point in time when a corresponding dedicated EPS bearer is established / modified.

NOTE 4: If the network supports Multimedia Priority Services (MPS) then this QCI could be used for the prioritization of non real-time data (i.e. most typically TCP-based services/applications) of MPS subscribers.

NOTE 5: This QCI could be used for a dedicated "premium bearer" (e.g. associated with premium content) for any subscriber / subscriber group. Also in this case, the SDF aggregate's uplink / downlink packet filters are known at the point in time when the SDF aggregate is authorized. Alternatively, this QCI could be used for the default bearer of a UE/PDN for "premium subscribers".

NOTE 6: This QCI is typically used for the default bearer of a UE/PDN for non privileged subscribers. Note that AMBR can be used as a "tool" to provide subscriber differentiation between subscriber groups connected to the same PDN with the same QCI on the default bearer.

NOTE 7: For Mission Critical services, it may be assumed that the PCEF is located "close" to the radio base station (roughly 10 ms) and is not normally used in a long distance, home routed roaming situation. Hence delay of 10 ms for the delay between a PCEF and a radio base station should be subtracted from this PDB to derive the packet delay budget that applies to the radio interface.

NOTE 8: In both RRC Idle and RRC Connected mode, the PDB requirement for these QCIs can be relaxed (but not to a value greater than 320 ms) for the first packet(s) in a downlink data or signalling burst in order to permit reasonable battery saving (DRX) techniques.

NOTE 9: It is expected that QCI-65 and QCI-69 are used together to provide Mission Critical Push to Talk service (e.g., QCI-5 is not used for signalling for the bearer that utilizes QCI-65 as user plane bearer). It is expected that the amount of traffic per UE will be similar or less compared to the IMS signalling.

NOTE 10: In both RRC Idle and RRC Connected mode, the PDB requirement for these QCIs can be relaxed for the first packet(s) in a downlink data or signalling burst in order to permit battery saving (DRX) techniques.

NOTE 11: In RRC Idle mode, the PDB requirement for these QCIs can be relaxed for the first packet(s) in a downlink data or signalling burst in order to permit battery saving (DRX) techniques.

NOTE 12: This QCI value can only be assigned upon request from the network side. The UE and any application running on the UE is not allowed to request this QCI value.

NOTE 13: Packet delay budget is not applicable on NB-IoT or when Enhanced Coverage is used for WB-E-UTRAN (see TS 36.300 [19]).

NOTE 14: This QCI could be used for transmission of V2X messages as defined in TS 23.285 [48].

NOTE 15: A delay of 2 ms for the delay between a PCEF and a radio base station should be subtracted from the given PDB to derive the packet delay budget that applies to the radio interface.

NOTE 16: For "live" uplink streaming (see TS 26.238 [53]), guidelines for PDB values of the different QCIs correspond to the latency configurations defined in TR 26.939 [54]. In order to support higher latency reliable streaming services (above 500ms PDB), if different PDB and PELR combinations are needed these configurations will have to use non-standardised QCIs.

Figure B.2: QCI VoLTE - part 2 [40]

B.2. QoS Class Identifiers for 5G - 5QI

5g QoS model is based on QoS Flows. Similar to VoLTE, 5G QoS model supports both Guaranteed Bit Rate (GBR) and non-GBR flows. QoS flow ID is used to identify a flow in the 5GS. With a 5G core deployed, the SMF will control the QoS flow. Detailed explanation for the QoS model is explained in [6]. Figure B.3 and Figure B.4 list the one-to-one mapping of the standardized 5QI values to 5G QoS characteristics.

5QI Value	Resource Type	Default Priority Level	Packet Delay Budget	Packet Error Rate	Default Maximum Data Burst Volume (NOTE 2)	Default Averaging Window	Example Services
1	GBR (NOTE 1)	20	100 ms (NOTE 11, NOTE 13)	10^{-2}	N/A	2000 ms	Conversational Voice
2		40	150 ms (NOTE 11, NOTE 13)	10^{-3}	N/A	2000 ms	Conversational Video (Live Streaming)
3 (NOTE 14)		30	50 ms (NOTE 11, NOTE 13)	10^{-3}	N/A	2000 ms	Real Time Gaming, V2X messages Electricity distribution – medium voltage, Process automation - monitoring
4		50	300 ms (NOTE 11, NOTE 13)	10^{-6}	N/A	2000 ms	Non-Conversational Video (Buffered Streaming)
65 (NOTE 9, NOTE 12)		7	75 ms (NOTE 7, NOTE 8)	10^{-2}	N/A	2000 ms	Mission Critical user plane Push To Talk voice (e.g., MCPTT)
66 (NOTE 12)		20	100 ms (NOTE 10, NOTE 13)	10^{-2}	N/A	2000 ms	Non-Mission-Critical user plane Push To Talk voice
67 (NOTE 12)		15	100 ms (NOTE 10, NOTE 13)	10^{-3}	N/A	2000 ms	Mission Critical Video user plane
75 (NOTE 14)							
71		56	150 ms (NOTE 11, NOTE 13, NOTE 15)	10^{-6}	N/A	2000 ms	"Live" Uplink Streaming (e.g. TS 26.238 [76])
72		56	300 ms (NOTE 11, NOTE 13, NOTE 15)	10^{-4}	N/A	2000 ms	"Live" Uplink Streaming (e.g. TS 26.238 [76])
73		56	300 ms (NOTE 11, NOTE 13, NOTE 15)	10^{-8}	N/A	2000 ms	"Live" Uplink Streaming (e.g. TS 26.238 [76])
74		56	500 ms (NOTE 11, NOTE 15)	10^{-8}	N/A	2000 ms	"Live" Uplink Streaming (e.g. TS 26.238 [76])
76		56	500 ms (NOTE 11, NOTE 13, NOTE 15)	10^{-4}	N/A	2000 ms	"Live" Uplink Streaming (e.g. TS 26.238 [76])
5	Non-GBR (NOTE 1)	10	100 ms NOTE 10, NOTE 13)	10^{-6}	N/A	N/A	IMS Signalling
6		60	300 ms (NOTE 10, NOTE 13)	10^{-6}	N/A	N/A	Video (Buffered Streaming) TCP-based (e.g., www, e-mail, chat, ftp, p2p file sharing, progressive video, etc.)
7		70	100 ms (NOTE 10, NOTE 13)	10^{-3}	N/A	N/A	Voice, Video (Live Streaming) Interactive Gaming
8							

Figure B.3: 5G QCI categories- part 1 [6]

		80	300 ms (NOTE 13)	10^{-6}	N/A	N/A	Video (Buffered Streaming) TCP-based (e.g., www, e-mail, chat, ftp, p2p file sharing, progressive video, etc.)
9		90					
69 (NOTE 9, NOTE 12)		5	60 ms (NOTE 7, NOTE 8)	10^{-6}	N/A	N/A	Mission Critical delay sensitive signalling (e.g., MC-PTT signalling)
70 (NOTE 12)		55	200 ms (NOTE 7, NOTE 10)	10^{-6}	N/A	N/A	Mission Critical Data (e.g. example services are the same as 5QI 6/8/9)
79		65	50 ms (NOTE 10, NOTE 13)	10^{-2}	N/A	N/A	V2X messages
80		68	10 ms (NOTE 5, NOTE 10)	10^{-6}	N/A	N/A	Low Latency eMBB applications Augmented Reality
82	Delay Critical GBR	19	10 ms (NOTE 4)	10^{-4}	255 bytes	2000 ms	Discrete Automation (see TS 22.261 [2])
83		22	10 ms (NOTE 4)	10^{-4}	1354 bytes (NOTE 3)	2000 ms	Discrete Automation (see TS 22.261 [2])
84		24	30 ms (NOTE 6)	10^{-5}	1354 bytes (NOTE 3)	2000 ms	Intelligent transport systems (see TS 22.261 [2])
85		21	5 ms (NOTE 5)	10^{-5}	255 bytes	2000 ms	Electricity Distribution-high voltage (see TS 22.261 [2])

NOTE 1: A packet which is delayed more than PDB is not counted as lost, thus not included in the PER.
NOTE 2: It is required that default MDBV is supported by a PLMN supporting the related 5QIs.
NOTE 3: This MDBV value is set to 1354 bytes to avoid IP fragmentation for the IPv6 based, IPSec protected GTP tunnel to the 5G-AN node (the value is calculated as in Annex C of TS 23.060 [56] and further reduced by 4 bytes to allow for the usage of a GTP-U extension header).
NOTE 4: A static value for the CN PDB of 1 ms for the delay between a UPF terminating N6 and a 5G-AN should be subtracted from a given PDB to derive the packet delay budget that applies to the radio interface. When a dynamic CN PDB is used, see clause 5.7.3.4.
NOTE 5: A static value for the CN PDB of 2 ms for the delay between a UPF terminating N6 and a 5G-AN should be subtracted from a given PDB to derive the packet delay budget that applies to the radio interface. When a dynamic CN PDB is used, see clause 5.7.3.4.
NOTE 6: A static value for the CN PDB of 5 ms for the delay between a UPF terminating N6 and a 5G-AN should be subtracted from a given PDB to derive the packet delay budget that applies to the radio interface. When a dynamic CN PDB is used, see clause 5.7.3.4.
NOTE 7: For Mission Critical services, it may be assumed that the UPF terminating N6 is located "close" to the 5G_AN (roughly 10 ms) and is not normally used in a long distance, home routed roaming situation. Hence a static value for the CN PDBof 10 ms for the delay between a UPF terminating N6 and a 5G_AN should be subtracted from this PDB to derive the packet delay budget that applies to the radio interface.
NOTE 8: In both RRC Idle and RRC Connected mode, the PDB requirement for these 5QIs can be relaxed (but not to a value greater than 320 ms) for the first packet(s) in a downlink data or signalling burst in order to permit reasonable battery saving (DRX) techniques.
NOTE 9: It is expected that 5QI-65 and 5QI-69 are used together to provide Mission Critical Push to Talk service (e.g., 5QI-5 is not used for signalling). It is expected that the amount of traffic per UE will be similar or less compared to the IMS signalling.
NOTE 10: In both RRC Idle and RRC Connected mode, the PDB requirement for these 5QIs can be relaxed for the first packet(s) in a downlink data or signalling burst in order to permit battery saving (DRX) techniques.
NOTE 11: In RRC Idle mode, the PDB requirement for these 5QIs can be relaxed for the first packet(s) in a downlink data or signalling burst in order to permit battery saving (DRX) techniques.
NOTE 12: This 5QI value can only be assigned upon request from the network side. The UE and any application running on the UE is not allowed to request this 5QI value.
NOTE 13: A static value for the CN PDB of 20 ms for the delay between a UPF terminating N6 and a 5G-AN should be subtracted from a given PDB to derive the packet delay budget that applies to the radio interface.
NOTE 14: This 5QI is not supported in this Release of the specification as it is only used for transmission of V2X messages over MBMS bearers as defined in TS 23.285 [72] but the value is reserved for future use.
NOTE 15: For "live" uplink streaming (see TS 26.238 [76]), guidelines for PDB values of the different 5QIs correspond to the latency configurations defined in TR 26.939 [77]. In order to support higher latency reliable streaming

Figure B.4: 5G QCI Categories - part 2 [6]

C

Network Functions Virtualization

There is an underlying NFV infrastructure that contains the compute, storage and network resources to the software processes. This infrastructure is shared between different NFs. The hardware is available as Commercial Off The Shelf (COTS). All of the NFs and the NFV infrastructure require a Management and Orchestration layer (MANO) to facilitate all the other processes.

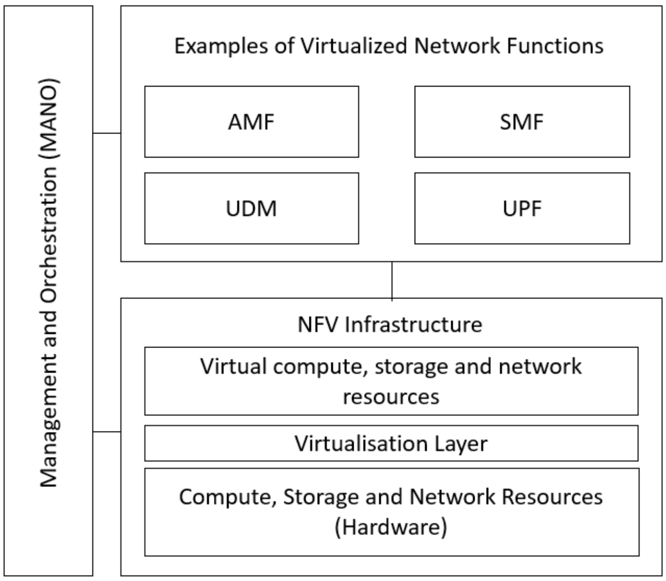

Figure C.1: NFV in 5G

NFV enables network slices. Allows the creation of different logical network (slices) over the same physical infrastructure maximizing the utilization of available resources. Service providers have a lot of flexibility about the way the services are provided to the customers. For example, there can be one slice across the 5G RAN and 5G core network that is dedicated to voice services that has low latency voice and signalling requirements with guaranteed QoS. Another for Vehicle to Anything communication which is designed for low latency and critical communication. A third one for smart metering with large number of connected devices having low data requirements and minimal network activity. All these environments can co-exist with a maximum of 8 network slices.

D

Key Performance Indicators

5G is expected to provide better User Experience in comparison with the previous generations of mobile network. This user experience depends on factors like data rate, latency, mobility and are associated with some KPIs. Figure D.1 presents the requirements for the User Experience when utilizing a 5GS. More explanation is provided in [1].

Use case category	User Experienced Data Rate	E2E Latency	Mobility
Broadband access in dense areas	DL: 300 Mbps UL: 50 Mbps	10 ms	On demand, 0-100 km/h
Indoor ultra-high broadband access	DL: 1 Gbps, UL: 500 Mbps	10 ms	Pedestrian
Broadband access in a crowd	DL: 25 Mbps UL: 50 Mbps	10 ms	Pedestrian
50+ Mbps everywhere	DL: 50 Mbps UL: 25 Mbps	10 ms	0-120 km/h
Ultra-low cost broadband access for low ARPU areas	DL: 10 Mbps UL: 10 Mbps	50 ms	on demand: 0-50 km/h
Mobile broadband in vehicles (cars, trains)	DL: 50 Mbps UL: 25 Mbps	10 ms	On demand, up to 500 km/h
Airplanes connectivity	DL: 15 Mbps per user UL: 7.5 Mbps per user	10 ms	Up to 1000 km/h
Massive low-cost/long-range/low-power MTC	Low (typically 1-100 kbps)	Seconds to hours	on demand: 0-500 km/h
Broadband MTC	See the requirements for the Broadband access in dense areas and 50+Mbps everywhere categories		
Ultra-low latency	DL: 50 Mbps UL: 25 Mbps	<1 ms	Pedestrian
Resilience and traffic surge	DL: 0.1-1 Mbps UL: 0.1-1 Mbps	Regular communication: not critical	0-120 km/h
Ultra-high reliability & Ultra-low latency	DL: From 50 kbps to 10 Mbps; UL: From a few bps to 10 Mbps	1 ms	on demand: 0-500 km/h
Ultra-high availability & reliability	DL: 10 Mbps UL: 10 Mbps	10 ms	On demand, 0-500 km/h
Broadcast like services	DL: Up to 200 Mbps UL: Modest (e.g. 500 kbps)	<100 ms	on demand: 0-500 km/h

Figure D.1: User Requirements and KPI’s in 5G [1]

E

RTP and RTCP

RTP provides end-to-end delivery services for data with real-time characteristics such as interactive audio and video. RTP runs on top of User Datagram Protocol(UDP) to utilise capabilities like multiplexing and checksum services.This occurs at the transport level. There is no guarantee of timely-delivery or quality-of-service (QoS). The packets are tagged with a sequence number which allows the receiver to reconstruct the sender's packet sequence. RTCP is the protocol used to monitor the QoS along with conveying information about the ongoing session.

E.1. Port Usage

RTP and RTCP port numbers must never be the same since demultiplexing in RTP data streams and RTCP control streams relies on the port numbers. There are certain restrictions to allocate the ports for sending/receiving streams. The restrictions for allocating port numbers are as follows:

- **UDP and similar protocols:** RTP uses even destination port number and corresponding RTCP stream will be assigned the next higher (Odd) destination port number.
- **Single port number as a parameter:** RTP and RTCP will need to derive the RTP/RTCP port pair from this number. If an odd number is supplied, then the application should replace the number with the next lower (even) number to use as the base of the port pair.
- **Applications defining RTP/RTCP port numbers explicitly:** The application may disregard the restrictions that the port numbers must be even/odd.
- **Unicast session:** Both participants need to choose a port pair for receiving RTP/RTCP packets.

E.2. Sender/Receiver Reports (SR and RR)

RTCP involves the periodic transmission of control packets between different participants involved in a session. The RTCP carrying the the transmission and reception statistics from active senders involved in the session is referred to as the Sender Report (SR). When the same statistics are from participants that involve senders that are not active is referred to as Receiver Report (RR). The SR and RR have different packet codes. The SR includes a 20-byte sender information section to be used by the active senders (highlighted in Figure E.1). SR is sent if there has been any data packet has been transmitted since the previous report, else an RR is issued [56]. Figure E.1 presents the packet formats for the SR and RR. Detailed explanation about SR and RR is provided in [56].

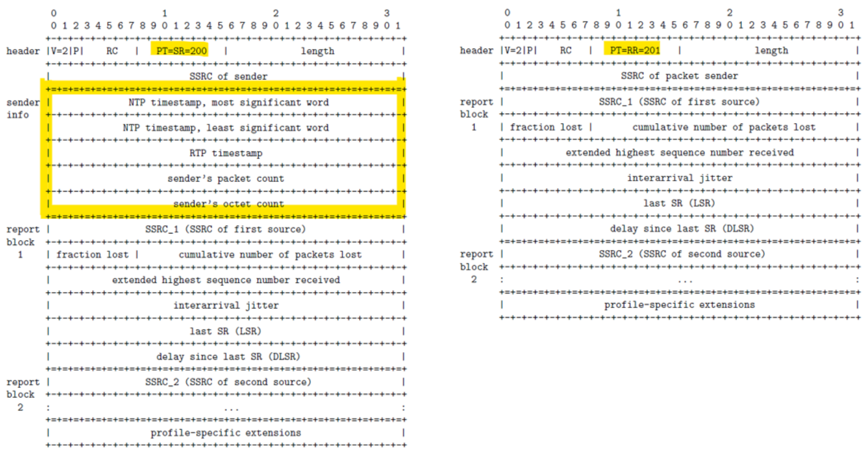

Figure E.1: SR (left) and RR (right) packet formats [56]

F

3GPP Technical Specifications

3GPP Technical Report (TR) / Technical Specification (TS) Number	Title	Release
TR 21.915	Release Description	Release 15
TR 21.916	Release Description	Release 16
TS 23.002	Network Architecture	Release 15
TS 23.203	Policy and Charging Control Architecture	Release 16
TS 23.401	GPRS Enhancements for E-UTRAN Access	Release 16
TS 23.501	System Architecture for the 5G System	Release 16
TS 23.502	Procedures for the 5G System	Release 16
TS 23.503	Policy and charging control framework for the 5G System (5GS)	Release 16
TS 23.228	IP Multimedia Subsystem (IMS)	Release 16
TS 24.229	IP multimedia call control protocol based on SIP and SDP	Release 16
TS 29.229	Cx and Dx interfaces based on Diameter protocol details	Release 15
TS 37.340	E-UTRA and NR; Multi-connectivity	Release 16

Figure F.1: 3GPP technical specifications frequently used in the thesis

Bibliography

[1] R.E Hattachi and J Erfanian. *5G White Paper - NGMN Alliance*. Tech. rep. v1.0. Feb. 2015.

[2] Ericsson Technology Review. *5G and Evolution of LTE*. Issue 1. n.d 2017.

[3] K.J Salchow and F5 Networks Inc. *Introduction to IP Multimedia Subsystem (IMS): IMS Basic Concepts and* . Tech. rep. White Paper. Aug. 2007.

[4] D Myers and IHS Markit Technology. *Cloud-Native IMS: Critical for Mobile Operators, White Paper*. Tech. rep. v1.0. Aug. 2019.

[5] 3GPP. *Technical Report (TR) 23.783 Study on Mission Critical MC Services support over 5GS Release 16*. Tech. rep. v0.6.0. May 2019.

[6] 3GPP. *Technical Specification (TS) 23.501 System Architecture for the 5G System Release 16*. Tech. rep. v16.2.0. Sept. 2019.

[7] Samsung. *5G Core Vision - Samsung 5G Core*. Tech. rep. Vol 1. n.d.

[8] Huawei. *5G Network Architecture - A High-Level Perspective*. Tech. rep. n.d n.d.

[9] 3GPP. *Technical Specification (TS) 23.502 Procedures for the 5G System Release 16*. Tech. rep. v16.1.1. June 2019.

[10] 3GPP. *Technical Specification (TS) 23.228 IP Multimedia Subsystem (IMS) Release 15*. Tech. rep. Version 15.3.0. Sept. 2018.

[11] 3GPP. *Technical Specification(TS) 24.229 IP Multimedia Call Control based on Session Initiation Protocol(SIP) and Session Description Protocol(SDP) Stage 3 Release 16*. Tech. rep. v16.2.0. June 2019.

[12] Spirent. *IMS Architecture The LTE User Equipment Perspective*. Tech. rep. Rev B. Apr. 2014.

[13] 3GPP. *Technical Report (TR) 23.794 Study on Enhanced IMS to 5GC Integration Release 16*. Tech. rep. v1.0.0. Mar. 2019.

[14] 3GPP. *Technical Report (TR) 23.756 Study for Single Radio Voice Continuity from 5GS to 3G Release 16*. Tech. rep. v16.0.0. Aug. 2019.

[15] Ericsson. *5G Voice - Network Evolution Aspects Prerequisites for voice services in 5G and 3GPP Option 3 Deployment*. Paper 1. n.d 2019.

[16] Ericsson. *5G Voice - Network Evolution Aspects Voice Services in a 5G System with 3GPP Option 2 Deployment*. Paper 2. n.d 2019.

[17] Ericsson. *5G Voice - Network Evolution Aspects Voice over NR in a 5G System and Migration from Evolved Packet System Fallback*. Paper 3. n.d 2019.

[18] Ericsson. *5G Voice - Network Evolution Aspects Evolution and Industry Options towards target solutions for 5G Voice*. Paper 4. n.d 2019.

[19] Ericsson. *5G Voice - Network Evolution Aspects SMS and Emergency Calls*. Paper 5. n.d 2019.

[20] Huawei. *Vo5G Technical White Paper*. Tech. rep. July 2018.

[21] Z. Zhang. "*Research on Handover Technologies in 5th Generation Wireless Communication System*". In: (June 2018).

[22] *Cellular Wireless 1G, 2G, 3G, 4G, 5G – Watch The Evolution*. URL: `%7Bhttps://blog.commscopetraining.com/cellular-wireless-watch-the-evolution/%7D`. (accessed: 11.03.2020).

[23] *Evolution of 4G: A Study*. May 2015.

[24] *1G*. URL: `%7Bhttps://en.wikipedia.org/wiki/1G%7D`. (accessed: 09.03.2020).

[25] *The future of Mobile Telecom*. URL: %7Bhttps://sunilmobiletelecom.blogspot.com/2013/02/network-architecture-evolution-1g-to-4g.html%7D. (accessed: 09.03.2020).

[26] *2nd Generation(GERAN)*. URL: %7Bhttps://www.etsi.org/technologies/mobile/2g%7D. (accessed: 13.03.2020).

[27] *2nd Generation(GERAN)*. URL: %7Bhttps://www.gta.ufrj.br/ensino/eel879/trabalhos_vf_2014_2/rafaelreis/background.html%7D. (accessed: 13.03.2020).

[28] *3rd Generation (UMTS)*. URL: %7Bhttps://www.etsi.org/technologies/mobile/3g%7D. (accessed: 20.03.2020).

[29] R. Noldus. *Cellular Architecture Services - Lecture 5 4G Mobile Communications Network*. Slide pack belonging to the course EE4630 at Technical University of Delft. n.d 2019.

[30] Ochang P.A and Irving P.J. "*Evolutionary Analysis of GSM, UMTS and LTE Mobile Network Architectures*". In: *World Scientfic News* 54.27-39 (2016), pp. 32–34. DOI: http://www.worldscientificnews.com/wp-content/uploads/2016/01/WSN-54-2016-27-39.pdf.

[31] Singh R.K and Singh.R. "*4G LTE Cellular Technology: Network Architecture and Mobile Standards*". In: *International Journal of Emerging Research in Management and Technology ISSN: 2278-9359* 5.Issue 12 (2016), p. 4. DOI: December.

[32] Ochang P.A and Irving P.J. "*Evolutionary Analysis of GSM, UMTS and LTE Mobile Network Architectures*". In: *World Scientfic News* 54.27-39 (2016), pp. 35–37. DOI: http://www.worldscientificnews.com/wp-content/uploads/2016/01/WSN-54-2016-27-39.pdf.

[33] *Policy and Charging Rules Function (PCRF) in LTE EPC Core Network Technology*. URL: %7Bhttps://www.netmanias.com/en/post/techdocs/10997/lte-pcrf/policy-and-charging-rules-function-pcrf-in-lte-epc-core-network-technology%7D. (accessed: 07.04.2020).

[34] R. Noldus et al. *IMS Application Developer's Handbook*. Academic Press, 2011. ISBN: 9780123821928.

[35] 3GPP. *Technical Specification (TS) 23.002 Network Architecture Release 15*. Tech. rep. v15.0.0. Mar. 2018.

[36] M. Lumumba et al. *IMS VoLTE Architecture - Voice Over LTE Tutorial*. URL: http://www.3glteinfo.com/ims-volte-architecture/. (Accessed 10.04.2020).

[37] GSM Association. *IMS Roaming , Interconnection and Interworking Guidelines*. 2018.

[38] 3GPP. *Technical Specification (TS) 23.401 GPRS Enhancements for E-UTRAN Access Release 16*. Tech. rep. v16.3.0. June 2019.

[39] GSM Association. *FCM.01 VoLTE Service Description and Implementation Guidelines*. Mar. 2014.

[40] 3GPP. *Technical Specification (TS) 23.203 Policy and Charging Control Architecture Release 16*. Tech. rep. v16.1.0. June 2019.

[41] GSM Association. *Road to 5G: Introduction and Migration*. n.d 2018.

[42] GSM Association. *5G Spectrum GSMA Public Policy Position*. Mar. 2020.

[43] Inc. Qualcomm Technologies. *What can we do with 5G NR Spectrum Sharing that isn't possible today?* Presentation notes. Dec. 2017.

[44] GSMA. *5G Implementation Guidelines*. v2.0. Mar. 2019.

[45] *5G Non Standalone Solution Guide, StarOS Release 21.11*. URL: https://www.cisco.com/c/en/us/td/docs/wireless/asr_5000/21-11_6-5/5G-NSA-Solution/21-11-5G-NSA-Solution-Guide/21-11-5G-NSA-Solution-Guide_chapter_01.html. (accessed: 04.12.2019).

[46] *Discover 5G Core Network Functions compared to 4G LTE*. URL: %7Bhttps://www.linkedin.com/pulse/discover-5g-core-network-functions-compared-4g-lte-paul-shepherd/%7D. (accessed: 20.11.2019).

[47] *Is the 5G PCF (Policy Control Function) Just a New Name for the PCRF?* URL: `%7Bhttps://apistraining.com/news/is-the-5g-pcf-policy-control-function-just-a-new-name-for-the-pcrf%7D`. (accessed: 20.07.2020).

[48] Xilinx. *Fronthaul, Midhaul and Backhaul*.

[49] Ericsson Technology Review. *Enabling Intelligent transport in 5G*. Issue 9-2017. Mar. 2018.

[50] P.Persson. *Voice in 5G - Speech on NR Steps to introduce voice in 5G, RAN perspective*. Slide pack belonging to Ericsson B.V. Nov. 2018.

[51] A.Ryde and R.Keller. *5G Voice Voice aspects of 5G evolution steps*. Slide pack belonging to Ericsson B.V. Nov. 2018.

[52] *WebRTC API*. URL: `%7Bhttps://developer.mozilla.org/en-US/docs/Web/API/WebRTC_API%7D`. (accessed: 12.05.2020).

[53] *WebRTC 1.0: Real-time Communication Between Browsers*. URL: `%7Bhttps://www.w3.org/TR/2019/CR-webrtc-20191213/%7D`. (accessed: 12.05.2020).

[54] *Getting started with WebRTC*. URL: `%7Bhttps://webrtc.org/getting-started/overview%7D`. (accessed: 12.05.2020).

[55] *How Much Bandwidth Is Needed for VoIP?* URL: `%7Bhttps://avoxi.com/blog/how-much-bandwidth-is-needed-for-voip/%7D`. (accessed: 20.06.2020).

[56] H. Schulzrinne et al. *RFC3550 RTP A Transport Protocol for Real-Time Applications*. July 2003.

[57] *Identifiers for WebRTC's Statistics API*. URL: `%7Bhttps://www.w3.org/TR/2020/CR-webrtc-stats-20200114/%7D`. (accessed: 05.08.2020).

[58] M.Abdurahman K.Sambath and V.Suryani. *High Quality of Service Video Conferencing over IMS*. International Journal of Information and Education Technology, Vol. 6, No. 6. June 2016.

[59] *Understanding Jitter in VoIP and How To Defeat It*. URL: `%7Bhttps://https://getvoip.com/blog/2018/04/02/understanding-jitter-voip/%7D`. (accessed: 22.06.2020).

[60] A.Kurian. *Latency Analysis and Reduction in a 4G Network*. Thesis report *http://resolver.tudelft.nl/uuid:e1badd8d-a384-49a1-b958-a0c1e499c539*. Feb. 2018.

[61] *iPerf - The ultimate speed test tool for TCP, UDP and SCTP*. URL: `%7Bhttps://iperf.fr/%7D`. (accessed: 24.06.2020).

[62] *Acceptable Jitter Latency for VoIP: Everything You Need to Know*. URL: `%7Bhttps://getvoip.com/blog/2018/12/20/acceptable-jitter-latency/%7D`. (accessed: 24.06.2020).