PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA

MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH

YAHIA FARES UNIVERSITY OF MEDEA



Faculty of **T**echnology

Electrical Engeneering Departement

Master's degree dissertation

Domain : Telecommunication

Speciality : Telecommunication systems

Network slicing in the 5G virtual RANs

Analysis, model and proposals for effective approaches

Presented by:

OULDKHAOUA Soufiane

BOUGUERRA Riadh

Supervised by:

Dr. KHEDDAR Hamza

2022 promotion

ACKNOWLEDGEMENTS

First of all, we would like to thank Allah the almighty and merciful, who gave us the strength and the patience to do this work.

Secondly, with a lot of gratitude and sincerity, we warmly thank **Dr. Hamza KHEDDAR** for his suggestions, criticism and guidance throughout this thesis' work.

Our sincere thanks also go to the jury members for the interest they have shown in our work by agreeing to examine it and to enrich it with their expertise.

Much love and appreciation to our parents, families and friends who supported us.

Finally, a big thank you to everyone who contributed directly or indirectly to the accomplishment of this work.

DEDICATIONS

I dedicate this dissertation to my mother, the best support system anyone can ask as I owe all my achievements to her.

To my brothers **Fayçal** and **Moncef**,

To my aunt **Wahiba** for all the help she gave me

To my partner **Soufiane** in this work.

To **Dr.KHEDDAR** for all his guidance and insightful feedback.

To all my friends and family.

Riadh Bouguerra

DEDICATIONS

To the women without whom I would have never been where I am today, my mother.

Before all, praise be to Allah for helping us accomplish our work.

We wish to express our sincere gratitude to professor **Hamza KHEDDAR** for his guidance, valuable advices and providing us with insightful feedback.

I dedicate my dissertation work to my parents, my brother and my sister.

I also dedicate this dissertation to my friends who have supported me throughout the process. I will always appreciate all they have done.

To my partner **Riadh** in this work, thank you for sharing this thesis with me, I would have never made it without you and your help.

Soufiane OULDKHAOUA

الملخص

إن مجال شبكات الاتصالات في العالم شهد نموا غير مسبوق في العقد الأخير أدى الى إصدار عدة اجيال حرصت على إرضاء المستخدمين من خلال الزيادة في سرعة التدفق و التقليل من وقت الاستجابة خاصة في إصدار الجيل الخامس 5G.

مع التسويق المستمر لشبكة الجيل الخامس في العالم و التي قد يتم العمل بها في الجزائر في السنوات القليلة المقبلة فإن الشبكة المعرفة بالبرمجيات (SDN) و وظيفة الشبكة الافتراضية (NFV) ستستهلا عملية تقطيع الشبكة لعدة شرائح مما يسمح ذلك بالاستخدام الفعال للموارد المادية و الافتراضية المتوفرة. يمكن لهذا المشروع ان يحلل و ان يقترح نموذجا فعالا لتحقيق أقصى استفادة من الموارد المتاحة للشبكة الجيل الخامس 5G.

الكلمات المفتاحية: الجيل الخامس 5G، الشبكة المعرفة بالبرمجيات (SDN)، وظيفة الشبكة الافتراضية (NFV)، شبكات الاتصال، تقطيع الشبكة.

Abstract

The telecommunication field has seen unprecedented growth in the last decade that has led to the release of several generations that have been committed to satisfy users by increasing the data rate and reducing the latency, especially in the 5G network.

With fully commercialized 5G networks that may be launched in the next few years in Algeria, Software-defined network (SDN) and network function virtualization (NFV) will facilitate the implementation of network slicing. SDN and NFV will serve as the basis for network slicing, allowing efficient use of both physical and virtual resources. This project makes it possible to analyze, propose an efficient model, and utilize all of the available resources of the 5G network.

Keywords: 5G network, Software-defined network (SDN), Network function virtualization (NFV), Telecommunications, Network slicing.

Résumé

Le domaine de télécommunications a connu une croissance au cours de la dernière décennie qui a conduit à l'apparence de plusieurs générations qui ont été engagés à satisfaire les utilisateurs, en augmentant le débit de données et en réduisant la latence, en spécifiquement dans le réseau 5G.

Avec des réseaux 5G entièrement commercialisés qui pourraient être lancés dans les prochaines années en Algérie, Réseau défini par logiciel (SDN) et virtualisation des fonctions réseau (NFV) peut facilite l'implémentation de network slicing. SDN et NFV serviront de base pour le découpage du réseau, permettant une utilisation efficace des ressources physiques et virtuelles. Ce projet permet d'analyser, proposer un modèle efficace et d'utiliser toutes les ressources disponibles du réseau 5G.

Mots clés: Réseau 5G, Réseau défini par logiciel (SDN), virtualisation des fonctions réseau (NFV), télécommunications, découpage du reseau.

Table of contents

ABSTRACT ..... **V**

TABLE OF CONTENTS ..... **VI**

LIST OF TABLES ..... **IX**

LIST OF FIGURES ..... **X**

LIST OF ABBREVIATIONS ..... **XII**

GENERAL INTRODUCTION **1**

CHAPTER I: THE SECOND-GENERATION MOBILE NETWORK SYSTEM

I.1 INTRODUCTION 3

I.2 GSM COMMUNICATION SYSTEM 3

I.2.1 GSM Architecture 4

I.3 GSM INTERFACES 7

I.3.1 The handover in GSM 8

I.3.2 Roaming 10

I.4 GSM RADIO TRANSMISSION 10

I.4.1 Frequency bands in GSM 10

I.4.2 Physical channels 11

I.4.3 Logical channels 12

I.5 THE GPRS 13

I.5.1 Network architecture upgrade 13

I.6 THE EDGE 14

I.7 NETWORK OPTIMIZATION PARAMETERS 14

I.7.1 KPIs 14

I.7.2 Major KPIs 15

I.7.3 TCH drop rate 16

I.7.4 TCH congestion 16

I.7.5 SDCCH drop rate 17

I.7.6 SDCCH congestion 17

I.7.7 Handover 17

I.8 CONCLUSION 18

CHAPTER II: THE THIRD AND THE FOURTH-GENERATION MOBILE NETWORK SYSTEM

II.1 INTRODUCTION 19
II.2 UMTS NETWORK 19
II.2.1 UMTS architecture 20
II.2.2 The radio access network 20
II.2.3 The radio access 21
II.2.4 Spreading 21
II.2.5 The handover in UMTS 23
II.2.6 Cell selection 24
II.2.7 Major parameters and spectrum allocations in the UMTS network 25
II.2.8 UMTS channels 25
II.2.9 RRC states 25
II.2.10 UMTS network optimization 27
II.3 LTE NETWORK 29
II.3.1 Overview of the LTE network 29
II.3.2 LTE network architecture 30
II.3.3 The radio access network 31
II.3.4 The core access network 31
II.3.5 LTE Interfaces 33
II.3.6 Handover in LTE 34
II.3.7 LTE radio access type 36
II.3.8 LTE optimization and KPI analysis 37
II.4 CONCLUSION 38

CHAPTER III: HORIZONTAL SLICING IN 5G NETWORK

III.1 INTRODUCTION 39
III.2 5G SERVICES AND REQUIREMENTS 39
III.3 NETWORK ARCHITECTURE IN 5G 41
III.3.1 New generation radio access network 43
III.3.2 5G core network 44
III.4 THE ROLE OF NETWORK FUNCTION VIRTUALIZATION AND SOFTWARE DEFINED NETWORKING IN 5G 46
III.4.1 Software defined networking 46
III.4.2 Network function virtualization 47

III.5 NETWORK SLICING 47
III.6 THE END-TO-END ARCHITECTURE OF NETWORK SLICING 48
III.7 TYPES AND SCENARIOS OF NETWORK SLICING 50
III.7.1 Network slicing scenarios 50
III.7.2 Network slicing types 52
III.8 NETWORK SLICING AT THE RAN LEVEL 53
III.8.1 Resource management 54
III.8.2 Control plane and user plane 55
III.8.3 Control of specific slice admission 55
III.8.4 UE comprehension of RAN configurations 56
III.9 NETWORK SLICING MANAGEMENT FRAMEWORK 56
III.10 OPPORTUNITIES INTRODUCED BY THE NETWORK SLICING 59
III.11 DESCRIPTION OF THE PROPOSED MODEL 60
III.11.1 Network architecture of the model 61
III.11.2 Model objective 63
III.11.3 CVX optimization tool 65
III.12 CASE SCENARIO STUDY 66
III.13 THE INFLUENCE OF THE QUALITY OF THE VNO RADIO LINK 68
III.14 THE INFLUENCE OF THE NUMBER USERS 71
III.15 CONCLUSION 74
GENERAL CONCLUSION 75
BIBLIOGRAPHY 76

List of tables

Table I. 1.GSM network interfaces. 8
Table I. 2. Characteristics technique of GSM Network 10
Table I. 3. Summarizes the existing logical channels in GSM system. 13
Table II. 1. WCDMA's key parameters. 25
Table II. 2. Summarizes the existing logical channels in UMTS system. 26
Table II. 3. CS call drop optimization. 28
Table II. 4. PS call drop optimization. 28
Table II. 5. RRC connection failure optimization. 28
Table II. 6. HO failure optimization. 28
Table II. 7. IRAT HO failure optimization. 29
Table II. 8. Power congestion optimization. 29
Table II. 9. The difference between LTE-FDD and LTE-TDD [18]. 36
Table II. 10. KPIs of LTE RAN. 37
Table III. 1. The 5G mobile network requirements. 40
Table III. 2. VNO’s parameters and priorities. 66
Table III. 3. Network parameters. 66
Table III. 4. Results obtained from $w_{vs,i}^{usr}$ with service flow rates s. 67
Table III. 5. The influence of quality on the VNO 1 radio link. 68
Table III. 6. The influence of quality on the VNO 2 radio link. 69
Table III. 7. The influence of quality on the VNO 2 radio link. 69
Table III. 8. The influence of the number of users on VNO 1. 71
Table III. 9.The influence of the number of users on VNO 2. 72
Table III. 10. The influence of the number of users on VNO 3. 72

List of figures

Figure I. 1. GSM globale architecture. 4
Figure I. 2. Structure of the IMEI number. 5
Figure I. 3. Handover in GSM network. 9
Figure I. 4. Time/frequency diagram of FDMA systems. 11
Figure I. 5. Time/frequency diagram of TDMA systems. 12
Figure II. 1. UMTS network architecture connected with legacy network. 20
Figure II. 2. OVSF Code tree. 22
Figure II. 3. Handover types in UMTS network. 23
Figure II. 4. RRC connection establishment. 27
Figure II. 5. LTE Network architecture. 30
Figure II. 6. LTE user plane protocol stacks. 34
Figure II. 7. Inter-RAT mobility procedures. 35
Figure II. 8. TDD mode. 36
Figure II. 9. FDD mode. 36
Figure III. 1. Services categories provided in 5G. 40
Figure III. 2. 5G subsystems network architecture [20]. 41
Figure III. 3. PDU sessions and QoS flows for user plane [20]. 42
Figure III. 4. Available functionalities in NG-RAN and 5GC. 43
Figure III. 5. Overall NG-RAN architecture. 43
Figure III. 6. 5G system architecture. 44
Figure III. 7. SDN slicing network architecture. 46
Figure III. 8. NFV architecture. 47
Figure III. 9. 5G network slicing concept. 48
Figure III. 10. Network slicing architecture. 49
Figure III. 11. UE connection with the network slices. 49
Figure III. 12. Network slice examples. 50
Figure III. 13. Illustration of vertical and horizontal network slicing. 53
Figure III. 14. Multiple slices in FDM sharing a radio resource such as slice#1 and slice#2 or #3, and TDM such as slice #2 and #3. 55
Figure III. 15. Network slice management. 57
Figure III. 16. Service to network slice pairing. 57
Figure III. 17. Service instance to network slice instance. 58
Figure III. 18. Management and resources for network slices. 59

Figure III. 19. Network architecture of the virtual RAN [29]. 62
Figure III. 20. The impact of VNOs' weights on the capacity share. 68
Figure III. 21.The influence of the quality of the VNO1 radio link. 69
Figure III. 22. The influence of the quality of the VNO2 radio link. 70
Figure III. 23. The influence of the quality of the VNO3 radio link. 70
Figure III. 24. The influence of the number of users on the allocated data rate in VNO1. 72
Figure III. 25. The influence of the number of users on the allocated data rate in VNO2. 73
Figure III. 26. The influence of the number of users on the allocated data rate in VNO3. 73

List of abbreviations

2G	Second Generation.
3G	Third Generation.
3GPP	Third-Generation Partnership Project.
4G	Fourth Generation.
5G	Fifth-Generation.
AKA	Authentication and Key Agreement.
AuC	Authentication Center.
AUSF	Authentication Server Function.
BCH	Broadcast Channel.
BE	Best Effort.
BG	Best Effort with Minimum Guaranteed.
BSC	Base Station Controller.
BSS	Base Station Subsystem.
BTS	Base Transceiver Station.
CaaS	Capacity-as-a-Service.
CDMA	Code Devision Multiple Access.
CEPT	European Conference of Postal and Telecommunications Administrations.
CRRM	Common Radio Resource Management.
CSFB	Circuit Switch Full Back.
EDGE	Enhanced Data for GSM Evolution.
EIR	Equipment Identification Registry.
eMBB	enhanced Mobile Broadband.
EPC	Evolved Packet Core.
ETSI	European Telecommunications Standards Institute.
FDD	Frequency Division Duplex.
FDMA	Frequency Division Multiple Access.
GB	Guaranteed Bitrate.
GBR	Guaranteed Bit Rate.
GGSN	Gateway Service Node.
GMSC	Gateway Mobile Switching.
GMSK	Gaussian Minimum Shift Keying.
GPRS	General Packet Radio Service.
GSM	Global System for Mobile
GSMA	GSM-Association.
Het-Nets	Heterogeneous Networks.
HLR	Home Location Registry.
HSPA	High Speed Packet Access.
IMEI	International Mobile Station Equipment Identity.
IMS	IP Multimedia Subsystem.
IMSI	International Mobile Subscriber Identity.
IoT	Internet-of-Things.
ISDN	Integrated Service Digital Network.

ITU	International Telecommunication Union.
KPI	Key Performance Indicators.
LTE	Long-Term Evolution.
MAC	Medium Access Control.
MIMO	Multiple Input Multiple Output.
MME	Mobility Management Entity.
mMTC	massive Machine Type Communication.
MS	Mobile Station.
MSC	Mobile Switching Center.
MSISDN	Mobile Subscriber ISDN Number.
MSRN	Mobile Station Roaming Number.
NFV	Network Function Virtualization.
NGMN	Next Generation Mobile Network.
NMO	Network Management and Orchestration.
NSS	Network Switching Subsystem.
OFDM	Orthogonal Frequency Division Multiplexing.
OFDMA	Orthogonal Frequency Division Multiple Access.
OVSF	Orthogonal Variable Spreading Factor.
PCCH	Paging Control Channel
PCH	Paging Channel
PCU	The Packet Control Unit.
PDCP	Packet Data Convergence Protocol.
PDN	Packet Data Network.
PDU	Packet Data Unit.
P-GW	Packet Data Network Gateway.
PIN	Personal Identification Number.
PLMN	Public Land Mobile Network.
PSK	Phase-Shift Keying.
PSTN	Public Switched Telephone Network.
QoE	Quality of Experience.
QoS	Quality of Service.
RACH	Random-Access Channel.
RAKE	Radio Activated Key Entry.
RAT	Radio Access Technics.
REST	Representational State Transfer.
RLC	Radio Link Control.
RNS	Radio Network Subsystem.
RRM	Radio Resource Management.
RRU	Radio Resource Unit.
SAE	System Architecture Evolution.
SC-FDMA	Single Carrier Frequency Division Multiple Access.
SDN	Software Defined Networking.
SGSN	Serving GPRS Support Node.
SGW	Serving Gateway.
SIM:	Subscriber Identity Modul.

SIR	Signal-to-Interference Ratio.
SLA	Service Layer Agreements.
TCH	Traffic Channel.
TDD	Time Division Duplex.
UE	User Equipment.
UHD	Ultra-High Definition.
UMTS	Universal Mobile Telecommunications Radio Network System.
URLLC	Ultra-Reliable Low Latency Communication.
USIM	Universal Subscriber Identity Module.
UTRAN	UMTS Terrestrial Radio Access Network.
VLR	Visitor Location Registry.
VNO	Virtual Network Operator.
VoIP	Voice over IP.
VRRM	Virtual Radio Resource Management.
WCDMA	Wideband Code Division Multiple Access.

General introduction

Mobile network evolution has been categorized into ‘generations’ as a differentiating factor between different technologies and their capabilities, it all started with the first generation (1G) in the 1980s, it based on analogue transmission techniques. At that time, there was no worldwide (not even Europe-wide) coordination for the development of the technical standards for the system. Nordic counties deployed Nordic mobile telephones (NMTs). while UK and Ireland went for a total access communication system (TACS). It offered mainly speech and related services and were highly incompatible with each other. Thus, their main limitations were the limited number of services offered and their incompatible nature.

An increase in the necessity for a system that catered to mobile communication needs and also offered increased compatibility with other systems, resulted in the birth of the second-generation (2G) mobile systems. International bodies played a key role in evolving a system that would provide better services and was more transparent and compatible with networks globally. But unfortunately, the second-generation network standards could not fulfil the dream of having just one set of standards for networks globally. The standards in Europe differed from the standards in Japan and that of the Americas and so on. Of all the standards, the global system for mobile communication (GSM) went all the way in fulfilling the technical and the commercial expectations. GSM experienced changes in the forms of general packet radio services (GPRS) and enhanced data rates in GSM environment (EDGE).

In EDGE, very high movement of the data was possible, but still, the packet transfer on the air-interface behaves like a circuit switch call. Thus, part of this packet connection efficiency is lost in the circuit switch environment. Moreover, the standards for developing the networks till the second generation were different for different parts of the world. Hence, it was decided to have a network that provides services independent of the technology platform and whose network design standards are the same globally. Thus, 3G was born. In Europe, it was called universal terrestrial mobile system (UMTS), which is European telecommunications standards institute (ETSI) driven. IMT-2000 is the international telecommunication union-telecommunication standardization sector (ITU-T) name for the third-generation system while CDMA2000 is the name of the American 3G variant. Wideband code division multiple access (WCDMA) is the air interface technology for the UMTS. This platform offers many services that are based on the Internet, along with video calling, imaging, etc.

Further advancements to mobile networks technology led to long term evolution (LTE), a technology referred to as 4G. the main aim of LTE was to increase the speed and capacity while reducing latency of the mobile networks. The mobile networks become simpler in architecture

while moving towards an ALL-IP system. The air-interface used in LTE is orthogonal frequency division multiplexing (OFDM).

Fifth-generation or 5G networks will probably be the networks for which technology evolution will occur. The new elements in 5G radio would be new radio (NR) and a network that will be a truly convergent network that would include 4G LTE, Non-3GPP access technologies, Wi-Fi and core including network functions virtualization (NFV), software-defined networking (SDN), Internet of things (IoT), cloud. 5G networks are expected to connect billions of devices, a multifold increase in data, a latency as low as 1 ms and data rates as high as 100 Mbps.

In this theme, we proposed a 5G RAN model to simulate multiple scenarios on it by varying different factors such as the radio link quality, with the goal of further understanding the optimization process for a satisfactory experience of all the users.

This document is divided into 3 main chapters:

The first chapter describes the 2G network's architecture and its different generation.

The second chapter discuses both of the UMTS and LTE networks.

The third chapter is divided into two parts, the first one describes the 5G network's architecture and main functions and features, the second part represents the results of multiple simulations on proposed 5G core network model in the purpose of further understanding the optimization prosses of the network.

CHAPTER I: The second-generation mobile network system

I.1 Introduction

In the early 1980s, analog systems were introduced as the first generation of mobile phones. The second generation (2G) of digital mobile phones, as well as the first digital mobile networks, arrived in the 1990s. The mobile telecommunications sector grew exponentially in terms of customers and value-added services (VAS) throughout the second generation. Limited data support is available on second generation networks, rising from 9.6 kbps to 19.2 kbps [1].

From 1982 to 1985, members of the global system for mobile (GSM) group, which initially hosted by European conference of postal and telecommunications administrations (CEPT), debated whether to construct an analog or digital system. It was determined to develop a digital system after several field testing. A narrowband time division multiple access (TDMA) solution was chosen, and gaussian minimum shift keying (GMSK) has been selected as modulation strategy. By 1987, the technological foundations had been established, and the first specification had been written by 1990. With Radiolinja in Finland, GSM became the first commercially functioning digital cellular system in 1991. With over a billion people utilizing the system, GSM is the most popular and commonly used cellular system by 2005. The system's popularity was enhanced by features like prepaid calling, international roaming, and so on. Naturally, this resulted in the creation of smaller, lighter handsets with a greater number of functions. Apart from making phone calls, the system grew more user-friendly by adding a variety of services. Voice mail, short message service SMS, call waiting, and other services were available. By the year 2000, about 15 billion SMS had been sent every month, indicating that SMS had been a huge success. Higher quality of digital voice and low-cost options to communicate, such as text messaging, have been the primary advantages of GSM networks. Because the open standard enables easy interoperability, network operators have been able to install equipment from many providers.

In this chapter, an overview of the GSM network’s architecture with its different subsystems and the functions of each one, the different processes that happens within the network such as handover and roaming, all the iterations of 2G such as: general packet radio service (GPRS), enhanced data for GSM evolution (EDGE). The upgrades that were made to the architecture, and most importantly the process of the network optimization are thoroughly taken into consideration.

I.2 GSM communication system

The GSM standard was invented by the european telecommunications standards institute (ETSI) to outline the 2G digital networks protocols employed by terminals such as mobile phones. In December 1991, the first usage was in Finland, by the mid-2010s, it had become a

worldwide standard for communication mobile, spreading over of more than 90% and established in 193 countries operators [1].

I.2.1 GSM Architecture

The fixed network, which consist of the installed infrastructure and mobile subscribes, are the two key components of the network mobile system. The fixed installed network can be further broken down into three subnetworks, or subsystems which are: radio networks, mobile switching networks, and management networks. The main respective subsystems are:

- The mobile station (MS).
- The base station subsystem (BSS).
- The network switching subsystem (NSS).
- The operation support subsystem (OSS).

Figure I.1 depicts the architecture of GSM and its entities.

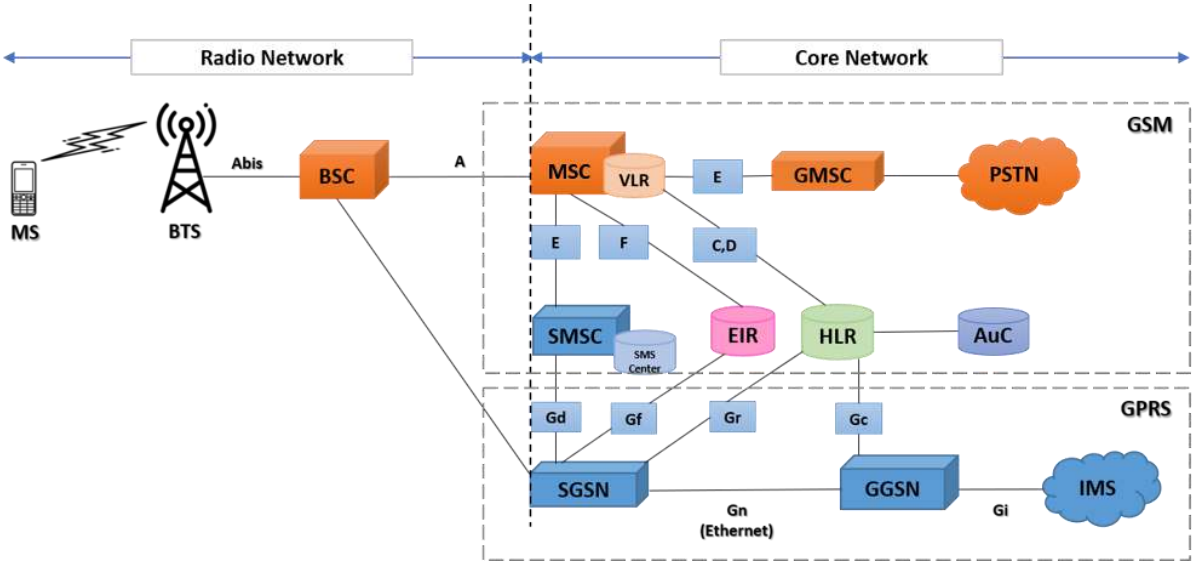


Figure I. 1. GSM globale architecture.

I.2.1.1 The mobile station

There are specific entities that used by mobile users to access many services. This latter are made up of a number of important components:

- ***Mobile equipment (ME):*** Or, as they're more often known, cell phones, which are the only GSM terminals that the user enables users to access the GSM network.
- ***Subscriber identity module (SIM):*** The subscriber identification module (SIM) gives a unique identity to mobile equipment. Certain subscriber characteristics, as well as the subscriber's personal data, are saved on the SIM card. The subscribers are identified by the SIM card in the network. Before using the phone, customers must

enter a 4-bit personal identification number (PIN) to secure the SIM card from unauthorized usage.

- ***International mobile station equipment identity (IMEI):*** It is a unique serial number used to identify mobile stations all around the world. The equipment manufacturer assigns it, and network operators record it in the equipment identification registry (EIR). The IMEI is a hierarchical address that consists of the following components:
  - Type approval code (TAC): Six digits, it is typically associated with the cellular modem or module.
  - Final assembly code (FAC): Identifies the location of the last production stage and consists of two digits.
  - Serial number (SNR): Consists of six digits, assigned by the manufacturer.
  - Spare (SP): Consist of one digit.

Figure I.2 depicts the main components of the IMEI number.

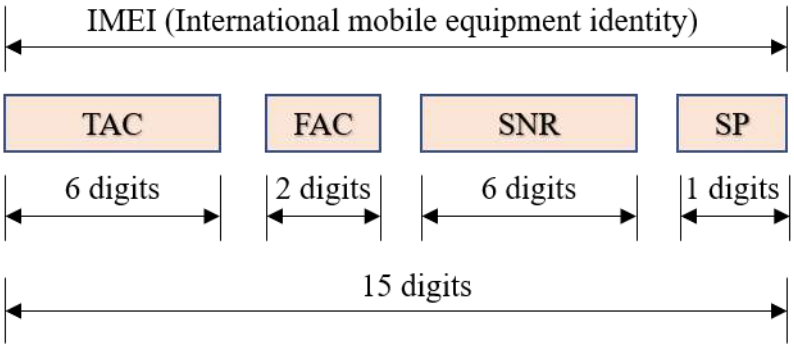


Figure I. 2. Structure of the IMEI number.

- ***International mobile subscriber identity (IMSI):*** Each subscriber obtains a unique identification called IMSI upon registering for service with a network operator, which is registered in the SIM. The mobile equipment with a valid IMEI works properly if the SIM card with a valid IMSI is plugged in it.
- ***Mobile subscriber ISDN number (MSISDN):*** The MSISDN is the MS's actual phone number. It is assigned to the subscriber; hence one MS is able to have many MSISDNs based on the SIM card. The MSISDN and IMSI association registered in the home location registry (HLR) must be known before the subscriber identification can be derived from the MSISDN.

- ***Mobile station roaming number (MSRN):*** The MSRN is a temporary integrated service digital network (ISDN) number that changes based on the location. It is assigned to each MS in its territory by visitor location registry (VLR), the MSRN is used to route MS calls. On request, the MSRN is sent from the HLR to the gateway mobile switching (GMSC).

I.2.1.2 The base station subsystem

The BSS is primarily responsible for connecting the mobile devices to the network. It is made up of two parts:

- ***The base transceiver station (BTS):*** The radio transmitter receivers, as well as their accompanying antennas, are part of the BTS in a GSM network, they broadcast and receives signal to connect to the mobile directly. Each cell is defined in a specific BTS. The BTS exchange communication with the mobile phones through the Um interface, which has its own set of protocols.
- ***The base station controller (BSC):*** The BSC controls the radio subsystem, specifically the BTSs. The main roles of the BSC are the radio resources and handover management. It is also responsible for managing the power control. It controls the signaling, and the security of the operation and maintenance (OM) subsystem. The BSC uses the Abis interface to connect with the BTSs [2].

I.2.1.3 The network switching subsystem

The mobile switching center (MSC), which is the most important element of the NSS, handles call switching between mobile and other fixed or mobile network users, as well as mobile service administration such as authentication.

The NSS system includes the following functional elements:

- ***Mobile switching center (MSC):*** The MSC is the most important component of the entire GSM network architecture's core network area. It functions similarly to a standard switching node in a public switched telephone network (PSTN) or an ISDN, but adds capabilities to satisfy the needs of mobile users. such as authentication call location, call routing and subscriber location, inter-MSC handover, and registration. It also plays the role of a gateway to the PSTN, enabling mobile communications calls to be routed from the public land mobile network (PLMN) to a PSTN. It connects to other MSCs, allowing calls to be established to mobile phones on various operators.
- ***Home location register (HLR):*** It's a database that stores all the subscriber's administrative information as well as their last known location. The MSC may then route calls to the appropriate BTS where the MS is located. When a user turns on their

mobile, it registers with the network, allowing the BTS that it connects, to be identified and the incoming calls is then routed correctly to the appropriate subscriber. Even when the phone is turned off (or idle mode), it registers the new state to guarantee that the HLR is up to date on its location. Although it may be deployed over several sub-centers for operational reasons, each network has only one HLR [1].

- ***Visitor location register (VLR):*** It comprises information from the HLR that permits and provides the visitor subscriber's services. Although the VLR can be designed as a distinct entity, it is more implemented done as an integrated element of the MSC, this improves access by making it faster and more efficient.
- ***Equipment identity register (EIR):*** The EIR is the authority that decides whether to allow certain mobile devices into the network. Every mobile device has a number called IMEI. As described previously, this number is installed in the device and checked by the network during registration. Based on the information stored in the EIR, the phone can be assigned one of three states - allowing access to the network, blocking access, or monitoring if something goes wrong.
- ***Authentication center (AuC):*** The AuC is a database that stores the secret key, which may also be found on the user's SIM card. On the radio channel, it's utilized for ciphering and authentication.
- ***Gateway mobile switching center (GMSC):*** The GMSC is the first point to which a ME terminating call is routed, even if the MS's location is unknown. As a result, the GMSC is responsible for getting the MSRN from the HLR based on the MSISDN and routing the call to the proper visiting MSC.

I.2.1.4 Operation and support subsystem

The OSS, is a supervision component of the overall GSM mobile communications network architecture, it is linked to NSS and BSS. It is utilized to regulate and monitor the whole GSM network, as well as supervise the BSS's traffic load. The three major functions of OSS are:

- Maintain all telecommunication hardware and network operations with a particular market.
- Manage all charging and billing procedures.
- Manage all mobile equipment in the system.

I.3 GSM Interfaces

Standard interfaces were created to guarantee that networks could be constructed using components from various manufacturers. This ensured that they communicated in the same way regardless of the manufacturers. Table I.1 shows the interfaces in the GSM network [3].

Table I. 1.GSM network interfaces.

GSM interface	Link	Role
Um	MS and BTS	Carries the GSM bursts carrying data and control information. Also referred as Air interface.
Abis	BTS and BSC	Supports two types of communication links via traffic channel at 64 kbps and signaling channel at 16 kbps.
A	BSC and MSC/VLR	Supports 2Mbps standard digital connection as per CCITT.
B	MSC and VLR	Used whenever the MSC needs access to data regarding a MS located in its area.
C	HLR and GMSC. Also, between MSC and HLR.	Used to obtain the requiring information required to complete the call.
D	HLR and VLR	Used to exchange the data related to the location of the ME and to the management of the subscriber.
E	MSC and another MSC or G-MSC.	Exchanges data related to handover between the anchor and relay MSCs.
F	EIR and MSC and between EIR and G-MSC.	Used to confirm the status of the IMEI of the ME gaining access to the network.
G	VLR and another VLR.	Used to transfer subscriber information.

I.3.1 The handover in GSM

One of the most important features of a mobile phone or cellular telecommunications system is that it is divided into numerous small cells in order to maximize frequency reuse and coverage. However, it must be feasible to maintain the connection while the phone goes from one cell to the next. The procedure is called as handover or handoff. The handover process within any cellular system is very important. This is a critical process that can result in lost calls if not switched correctly. Calls drop are especially annoying to users, and when the number of Calls drop rise up, customer dissatisfaction increases and they may probably to switch to another network. Therefore, GSM handover is a field of special concern in the standard development process [4].

Figure I.2 shows the steps of handover process. When one subscriber moves from blue cell to the green cell, the handover process launched to maintain the call.

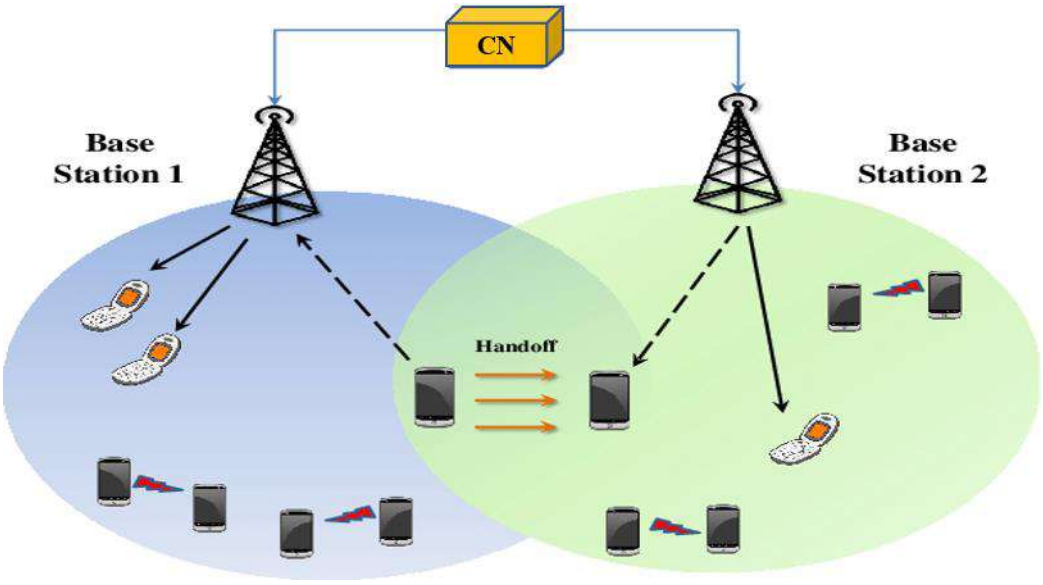


Figure I. 3. Handover in GSM network.

I.3.1.1 Different types of handover

For GSM-only systems, there are four forms of handover that may be conducted inside the GSM system:

- ***Intra-BTS handover:*** This type of handover happens when a mobile phone's frequency or slot needs to be changed due to interference. In this type of GSM handover, the mobile phone stays connected to the same BTS, but the channel or slot is switched.
- ***Inter-BTS intra BSC handover:*** It happens when a mobile phone leaves the area of one BTS and goes to another managed by the same BSC, this type of GSM handover happens. Before releasing the link between the former BTS and the MS, the BSC can launch the handover and allocate a new slot in a new channel [4].
- ***Inter-BSC handover:*** It happens when the mobile leaves the range of cells mastered by one BSC, the MS passed from one BTS to another BTS that belong to another BSC. The MSC is in charge of this handover in this case.
- ***Inter-MSC handover:*** This type of handover occurs when mobile leave MSC area to another MSC area, the two concerned MSC negotiate to master the handover.

I.3.2 Roaming

Roaming permits mobile users to automatically send and receive data, attempt to send and receive calls, or exploit other GSM services while traveling outside the geographic coverage area of their network via a visiting network. Authentication, billing, and mobility management procedures support roaming. This latter, called domestic, when visiting and home networks are in the same country, otherwise, it called international roaming. The inter-standard roaming happens when the visiting and home networks use different technical standard.

Roaming in GSM, allows mobile subscribers to utilize a single number, a single bill, and a single phone in as many as 219 countries.

I.4 GSM radio transmission

I.4.1 Frequency bands in GSM

GSM network occupies two frequency bands at around 900 MHz, which are:

- Uplink (890 to 915 MHz)
- Downlink (935 to 960 MHz)

Each bandwidth is divided to 125 channels (each channel is 200 KHz). These channels are not sufficient in urban areas, so it's necessary to attribute an additional band (DCS 1800) which has the same characteristics as the GSM 900 band in terms of protocols and services [7]. DCS occupies two frequency bands around 1800 MHz, which are:

- DCS uplink (1710 to 1785 MHz)
- DCS downlink (1805 to 1880 MHz)

Table I.2 summarizes most of signal characteristics of GSM-900 and GSM-1800 for both up-link and down-link.

Table I. 2. Characteristics technique of GSM Network

	GSM-900	**DCS-1800**
Downlink (MHz)	935 to 960	1805 to 1880
Uplink (MHz)	890 to 915	1710 to 1780
Links bandwidth (MHz)	45	95
Number of channels	125	375
Channels bandwidth (KHz)	200	200
Number of time slots (TDMA)	8	8
Logical channels number	992	2992

I.4.2 Physical channels

I.4.2.1 Access techniques

GSM networks use digital technologies to transmit data through radio waves. Frequency division multiple access (FDMA) and time division multiple access (TDMA) are the two technologies used in GSM.

- ***FDMA:*** is a technique for extracting, from transmission bearer, two or more continuous and simultaneous channels by allocating a different piece of the available frequency spectrum to each channel. It's used to send several broadcasts to a single transponder at the same time. The term FDMA refers to the fact that each BTS has its own radio frequency channel. MS in neighboring cells (or the same cell) can work at the same time, although they are separated by frequency. Multiple carrier frequencies, 125 in GSM 900 and 375 in GSM 1800, are used in the FDMA technology. Figure I.3 illustrates the principle of FDMA.

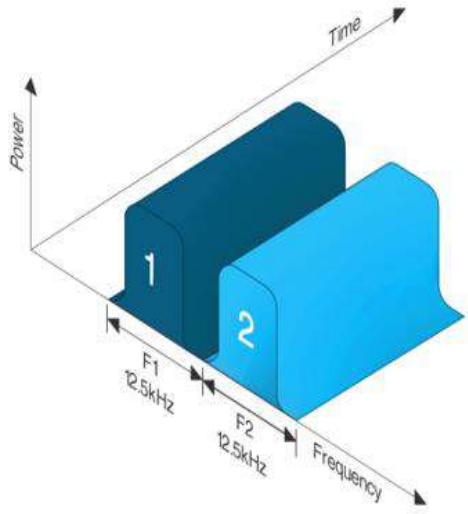


Figure I. 4. Time/frequency diagram of FDMA systems.

- ***TDMA:*** is a communications technique that allocates distinct time slots to various users on a shared channel (multipoint of broadcast) enabling conversations among several users. GSM employs both TDMA and FDMA, as well as a spread spectrum technology known as frequency hopping. Figure I.4 illustrates the principle of TDMA.

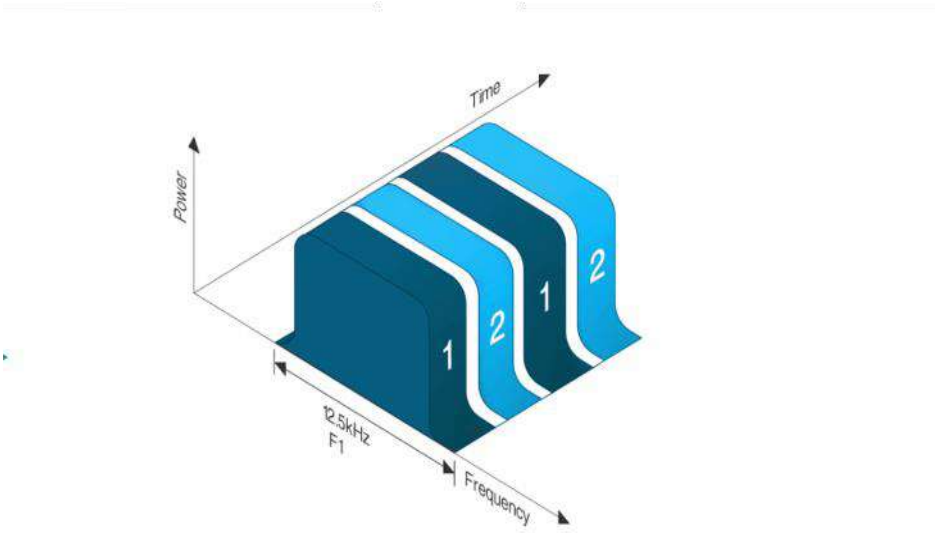


Figure I. 5. Time/frequency diagram of TDMA systems.

I.4.3 Logical channels

In GSM, logical channels are used to transport traffic and control data. The traffic channels are used to send information to users (speech or data). Control and signaling information are transmitted over the control channels (CCHs), and used for short-message services. The cell broadcast channel (CBCH) is used to send user information from a service center to mobile stations in a particular cell area.

Common control channel (CCCH), dedicated control channel (DCCH), and broadcast channel (BCH) are the control channels. The BCHs are used for frame synchronization, frequency correction, and control channel structure. They are downlink-only with point-to-multipoint channels. The broadcast control channel (BCCH), frequency correction channel (FCCH), and synchronization channel (SCH) are among these channels.

The BCCH is used to send cell identities, information about CCCHs, available cell service, and so on. Frequency correction data bursts containing all "0" are transmitted over the frequency correction channel. This results in a constant frequency shift in the RF carrier, which the mobile station can use for frequency correction. The SCH is employed to keep mobile stations time synchronized. Frame number and the BTS identity code (BSIC), which is required by MSs when measuring base station signal strength, are among the data on this channel.

Paging channel (PCH), random-access channel (RACH), and access grant channel (AGCH) are examples of CCCHs. The mobile stations are paged using PCHs. During certain times, the mobile stations must listen for paging. The AGCHs are used to assign mobiles to stand-alone dedicated control channels (SDCCHs). The RACHs are uplink-only channels that MSs use to transmit requests for dedicated network connections.

SDCCH and associated control channel (ACCH) are the two types of DCCHs. SDCCHs are point-to-point bidirectional channels that are used for ciphering initiation, service requests, subscriber authentication, and traffic channel (TCH) assignment, and equipment validation. The ACCHs are point-to-point bidirectional channels that are linked to a specific TCH and SDCCH. Out-of-band signaling and control data are sent between the MS and the BTS over these channels. Signal strength measurements from a MS to the BTS and transmission timing information from the BTS to the MS are two examples of their use. Table I.3 summarized the existing logical channels and their role in GSM network [5].

Table I. 3. Summarizes the existing logical channels in GSM system.

Logical channel	**Role of the logical channel**
TCH/F TCH/H BCCH	Full rate traffic channel to carry payload at 22.8 kbps Half rate traffic channel to carry payload at 11.4 kbps Broadcast network information.
SCH	Synchronization of the mobile stations.
FCCH	Used for frequency correction.
AGCH	Accepts a channel request from a mobile device and assigns an SDCCH to it.
FACCH	Traffic burst is stolen for a full signaling burst TCH in-band signaling, e.g., for link monitoring, for time critical signaling over TCH (e.g., for handoff signaling).
SACCH	TCH in-band signaling, e.g., for link monitoring
SDCCH	Registration/location updates FACCH for SDCCH during signaling exchange, such as during call setup. For a full signaling burst, the SDCCH burst is stolen.
FACCHs	For SDCCH, FACCH is the acronym. For a full signaling burst, the SDCCH burst is stolen.
SACCHs	SDCCH in-band signaling, e.g., for link monitoring.

I.5 The GPRS

For 2G cellular systems, GPRS provided a packet data capability, allowing GSM to evolve to include a data capability. Additional network entities must be added to the overall architecture for the latter purpose; two of the most important are the GPRS gateway service node (GGSN) and the serving GPRS support node (SGSN).

I.5.1 Network architecture upgrade

There are a number of network architectural changes that must be made as a result of GPRS to provide additional connectivity in terms of packet data (Figure I.1).

- ***SGSN:*** element provides a variety of services focused on the IP parts of the overall system. It offers a wide range of mobile services, including: Packet routing and transfer, mobility management, authentication, charging data [6].
- ***GGSN:*** is a critical component of the GPRS network design. It can be thought of as a hybrid of a gateway, router, and firewall. When the GGSN receives data addressed to a specific user, it first verifies that the user is online before sending the data.
- ***PCU:*** The packet control unit (PCU), which is connected to the BSC, is a hardware router. It converts data for the normal GSM network (circuit switched data) to data for the GPRS network (GPRS data) vise-versa.

I.6 The EDGE

With the introduction of EDGE technology, some of the limitations of the GPRS network were alleviated. Both TDMA and GSM networks are compatible with EDGE. Because it can be installed on any machine that supports GPRS, it is considered a subset of GPRS. The theoretical data rate for EDGE is 500 kbps. phase-shift keying (PSK) modulation technique, is used to achieve the theorical throughput. EDGE features increased the throughput per timeslot (8.8–59.2 kbps/timeslot), modulation changes from GMSK in GPRS to 8-PSK, decreased the signal sensitivity, and increased capacity and coverage. With the exception of some BTS hardware modifications and network software upgrades, EDGE does not necessitate significant hardware changes [1].

I.7 Network optimization parameters

I.7.1 KPIs

Key performance indicators (KPI) are a measure of successful network performance and its quality. KPIs can be used for the following functions:

- To keep track of and improve the performance of the radio network in order to provide better service to subscribers. To detect unacceptable performance related issues in a cellular network immediately.
- To provide detailed information to radio frequency planners.

Typically, KPI can be categorized into the following:

Accessibility: Measure the ability to set up a call.

Retainability: Measure the ability to maintain the existing connection [7].

Mobility: Measure the ability to maintain user connection while moving in network.

Integrity: Measure the degree of service after the service has been accessed by user.

I.7.2 Major KPIs

Include all the following parameters:

A. **Call set up success rate (CCSR):** CCSR is a measure of successful call attempts over the total number of calls attempts. A call is affirmed to be successful if SDCCH is accessed, occupied then TCH is accessed successfully. CCSR is usually affected due to blocking (SDCCH/TCH) or due to operational issues.

B. **Dropped call rate (DCR):** DCR is a metric that compares the number of dropped calls to the number of calls that were successfully set up. If TCH is dropped from the ager assignment, a call is confirmed to be dropped. From an RF standpoint, DCR is dependent on radio link timeout (RLT), and handover is reduced. RLT is a counter that keeps track of how many SACCH frames have been lost. When one SACCH is lost, it decreases by one, and when one SACCH is received, it increases by two. The RLT value is influenced by the RF environment (interference, low coverage). The probability of an increase in handovers is directly proportional to the probability of an increase in handover failures. These drops can be caused by both RF and operational issues. From an RF standpoint, the neighbor plan should be optimal, frequency clashes in all neighboring cells should be avoided, and all handover control and adjacency permeameters should be set to their maximum value. DCR may also rise as a result of operational issues [7].

C. **Handover successes rate (HSR/HOSR):** HSR is a metric that compares the number of successful handover attempts to the total number of attempts. Outgoing HSR (out HSR) and incoming HSR (in HSR) are the two types of HSR.

   Outgoing HSR is dependent on the adjacency and handover control parameters. It can also be affected due to TCH Blocking Issues in the adjacent cells leading to blocked handovers. Incoming HSR is also dependent on the adjacency and handover control parameters. It is affected by TCH blocking on source cell as well.

D. **Minutes per drop (MPD):** MPD is the average number of minutes after which a call gets dropped from the network. For example, if a certain cell has and MPD of 60, it would imply than a call strays on that cell for 60 minutes on average, before getting dropped. MPD is calculated by dividing the minutes of voice traffic with the total of TCH drops.

E. **TCH raw blocking:** it is a measure of TCH blocking during call set up. It does not include TCH blocking due to incoming handovers attempts, or TCH blocking due to incoming directed retry TCH seizure attempts.

F. **SDCCH blocking:** it is a measure of attempts for SDCCH seizure that could not be facilitated due to non-availability of SDCCH.

I.7.3 TCH drop rate

It's the difference between the total number of calls attempted and the number of calls dropped. The following are the major causes of the TCH drop rate:

A. Low signal strength downlink and uplink

If the signal strength of the uplink (UL) and the downlink (DL) is below the threshold the call gets dropped. To optimize the signal, the output power needs to be checked and the coverage plan.

The expected solution: Add a signal repeater to increase the coverage area, increase the emission power output, exchange the BTS antenna, add more BTSs to the network, or lock/unlock TRX.

B. Bad quality UL and DL

If the quality of the UL and DL is below the acceptable threshold, the call get dropped caused by the not sufficient signal quality. To solve this issue interference on the BCCH and TCH, frequency hop (FHOP) configuration needs to be verified.

The expected solution: Frequency change for BCCG, TCH, change BSIC, mobile allocation index offset (MAIO), FHOP, hopping sequency number (HSN), use a spectrum analyzer to identify the interference source, or resolve the quality problem.

C. Sudden connection lost

It’s the case were the call gets dropped suddenly. To resolve this problem the BTS’s error logs needs to be consulted.

The expected solution: Hardware failures needs to be fixed, reset TRX, ensure that the synchronization and the A-bis connection are stable, increase the transmission capacity, or investigate a handover problem.

D. Excessive timing advanced (ETA)

TCH drops caused an excessive TA. To resolve this issue TA max needs to be checked.

The expected solution: Adjust TA max to a value close to 63, adjust the antenna tilt and reduce its height, or adjust the output power.

I.7.4 TCH congestion

MSC requests a signaling assignment to the BSC to confirm the MS request for a TCH, if none are available, the BSC refuse the request. The main causer if the TCH congestion is the lack of the TCH resources. To resolve the TCH congestion all the TRXs in the cell needs to be checked, verify if the issue is linked to the occurrence of interferences or handover drops [7].

The expected solution: Half rate: It uses half the bandwidth of the full rate codec and operates at 5.6 kbit/s. As a result, network capacity for voice traffic is doubled at the expense of audio quality, or assignment to another cell: immediately after the detection of an affected SDCCH the call would be transferred to TCH of another adjacent cell.

I.7.5 SDCCH drop rate

It is the number of SDCCH channels blocked compared to the number of channels allocated. Causes of SDCCH drop rate are the following:

A. Low signal strength UL and DL: The expected solution is to plan a new site to cover for the low coverage area, increase power output, repair the faulty hardware, or adjust the antenna's tilt and height.

B. Poor quality DL and UL: The expected solution is to isolate the interfering frequency and change it, or do a coverage optimization of either the service cell or the jammer.

I.7.6 SDCCH congestion

It occurs once the BSC receives an SD request from the MS and all SDCCH resources are saturated at the time. The causes the SDCCG congestion are the flowing:

A. Insufficient system capacity: The expected solution is to properly plan the location area, configure more SDCCHs, utilize the SDCCH dynamic allocation, or add more TRXs to the system.

B. Poor LAC planning

If the contours of different location areas are at a high traffic area an SDCCH congestion occurs.

The expected solution: Resizing the location area codes (LAC) to avoided the highly saturated areas, or configure the reselection condition.

C. Interference: The expected solution is to adjust correctly the minimal threshold of the RACH access, or eliminate the interferences by adjusting the frequency.

I.7.7 Handover

Cause of handover problems:

A. Improper handover settings: The expected solution is to adjust the cell coverage parameters, adjust the cell reselect offset (CRO), or expand or adjust the TRX configuration between high and low traffic cells.

B. Faulty hardware: The expected solution to monitor transmission and advisory alarms, check if a transmission disconnection occurred and for faulty advisory cards, or adjust the synchronization.

C. Other causes: The expected solution is to change the frequencies to avoid signaling failures due to high interferences on the TCH, adjust antennas (tilt and height), or rewiring the cables between the TRX and antenna [7].

I.8 Conclusion

This chapter describes the main aspects of a GSM network, it's architecture with all the different subsystems and entities, all the releases of the 2G network (GPRS, EDGE) with the upgrades that were made to the network. Also defined some of the network key functions such as the handover and roaming. Moreover, the procedures and methods that can be utilized to optimize mainly coverage and capacity of the cellular network. It has been shown that the optimization process requires a network analysis with the use of different types of KPIs and troubleshooting and fine-tuning parameters to maintain the reliability and quality of the network.

CHAPTER II: The third and the fourth-generation mobile network system

II.1 Introduction

From the second generation, which is entirely digital and includes features such as SMS. The third generation (3G) of mobile technology was introduced to handle faster data rates and services such as mobile television, video calls, and so on. Finally, the fourth generation (4G) is about to take over the world. Long-term evolution (LTE) was created to meet the growing need for faster data speeds at reasonable prices.

The universal mobile telecommunications radio network system (UMTS) has been significantly enhanced and offers broadband speeds far beyond the original design. These high-speed enhancements are referred to as high-speed packet access (HSPA). The totally redesigned radio access network (AN) in UMTS standard, was the primary enhancement of UMTS in this initial stage compared to GSM. A new technology known as wideband code division multiple access (WCDMA) was invented instead of using Um interface's time and frequency multiplexing capabilities.

The eighth release of the third-generation partnership project (3GPP) specification series included the first version of LTE. It was able to take advantage of the most up-to-date knowledge and technological developments from HSPA and HSPA+. Consumers can perform tasks on a smart device as efficiently as if they were connected to broadband in their home or office, thanks to LTE's advanced data services. LTE, as a whole, improves the efficiency of using a smart device.

In this chapter, you'll learn about the architecture of UMTS and LTE networks, including all of their subsystems and their functions, as well as the various processes that occur within the network, such as handovers, and the WCDMA concept used to develop the networks. The architecture enhancements and changes, and most importantly, the network optimization process that brought UMTS and LTE together.

II.2 UMTS Network

The radio transmission standard for UMTS is the wideband code WCDMA. Many of the concepts used in GSM have been carried over and improved for UMTS. The SIM card has been changed into a global SIM card (USIM). Furthermore, the network has been constructed such that the GPRS and EDGE advancements may be applied to UMTS. As a result, the needed investment is maintained to a minimum [8].

The GSM-inspired core network (CN) topology composed of two user-traffic-dependent domains which are packet switch (PS) and circuit switch (CS) domains.

For both the GSM and UMTS access networks, the CS and PS domains manage their relevant traffic types in parallel. All circuit-switched traffic for the UMTS and GSM access

networks is handled by the CS domain, which consists of MSC and GMSC entities. Similarly, the PS domain, which is made up of the GPRS access network, is responsible for all PS traffic for both GPRS and UMTS, which consist of SGSSN and GGSN entities. In terms of data flow and connection, the UMTS also has a structure that separates the protocol stack from the relevant network entities [9].

For subscriber administration, mobile station roaming and identification, and managing diverse services, both PS and CS domains utilize the features of the common legacy network entities such as the HLR in conjunction with the EIR and AuC. As a result, the HLR contains subscriber information for GSM, GPRS, and UMTS.

II.2.1 UMTS architecture

With GPRS, the inclusion of packet data necessitated the establishment of more network entities. The UMTS network design was built on the merging of these two network pieces [10]. Figure II.1 depicts the total UMTS architecture with GPRS network. There are three basic parts to the UMTS network design:

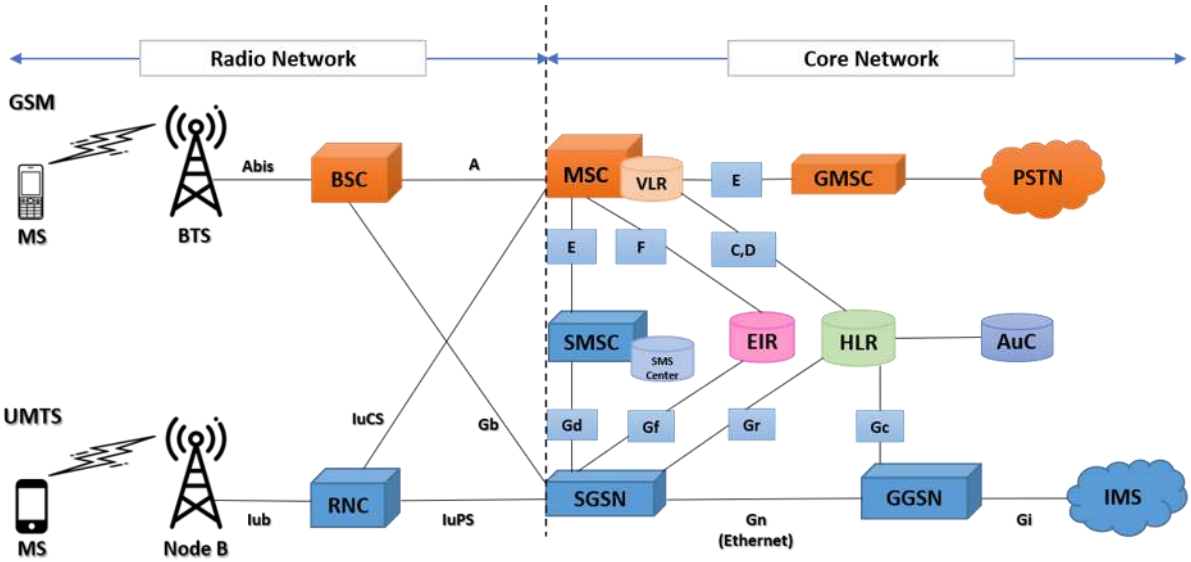


Figure II. 1. UMTS network architecture connected with legacy network.

II.2.2 The radio access network

II.2.2.1 User equipment

The user equipment (UE) is a crucial component of the 3G network's overall design. It serves as the user's last interface. It is made up of:

- ***RF circuitry:*** The RF regions are in charge of all signal components, both for the receiver and the transmitter. Reduced power consumption was one of the primary problems for the RF power amplifier. WCDMA modulation necessitates the usage of a linear amplifier.
- ***Baseband processing:*** Digital circuitry makes up the majority of base-band signal processing.
- ***Battery:*** The battery must be considered during circuit conception to get a low reduced consumption. Lithium ion (Li-ion) batteries are now commonly used to keep battery is small and light while maintaining their overall life charge.
- ***Universal subscriber identity module (USIM):*** A USIM is also included in the UE. It is an evolved version of the SIM card employed in GSM, yet it still holds a similar user information, such as MSISDN and IMSI.

II.2.2.2 UTRAN

The UTRAN, also known as the radio network subsystem (RNS), is made up of two main components:

- ***The node-B:*** It is the UMTS network's base station, and its primary function is to process signal processing in the Uu interface (Interleaving, adaptation rate, spreading, channel coding, radio resource management (RRM), power control) [9].
- ***The radio network controller (RNC):*** This element manages the radio resources as well as the NodeBs that are connected to it. Call admission control, RRM, code allocation, power control, packet scheduling, handover management, and CS domain encryption are all handled by it.

II.2.3 The radio access

WCDMA, which is a variants of time division duplex (TDD) and frequency division duplex (FDD), were the spread spectrum forms the underlying technique for WCDMA which is the main radio access technology used in UMTS. The following are the basic concepts used in the WCDMA system: channelization and scrambling, channel coding, power control, and handover [9].

II.2.4 Spreading

The rate of user data bits is R, assuming values of ± 1. The spreading technique involves multiplying each user data bit by a four-bit code sequence known as chips (as shown in Figure II.2). The resulting spread data has the same pseudo noise-like (random appearance) as the spreading code and is generated at a rate of $4 \times F$. The spread user data signal's occupied

spectrum bandwidth widens by a factor of four when the signaling rate is increased by a factor of four. Dispreading restores a bandwidth proportional to the signal's intended rate R. The spreading sequence is encoded in a particular manner, respecting an orthogonality factor between data channels. The advantage of using a long orthogonal variable spreading factor (OVSF) is that it adds redundancy to the transmitted data. The OVSF code length is proportional to the user's bit rate signal represented in the chip sequence [9]. The principle of OVSF code generation is depicted in Figure II.2.

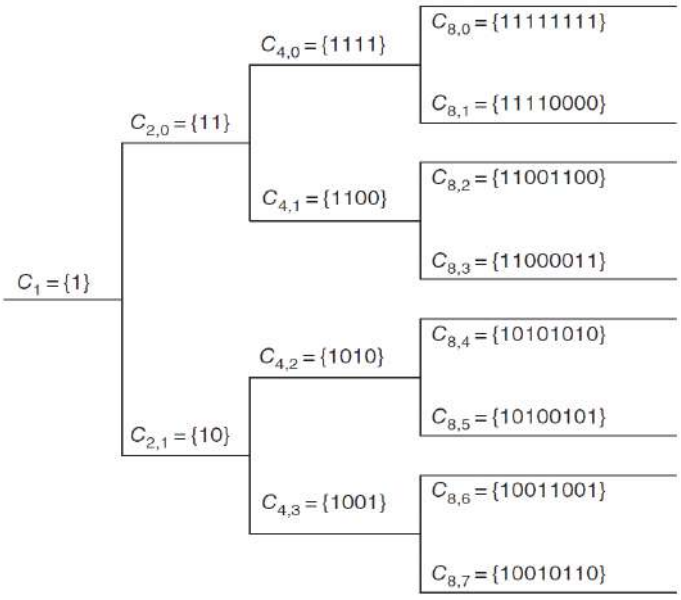


Figure II. 2. OVSF Code tree.

II.2.4.1 Power control

The power control is instrumental for WCDMA technique and used basically to:

- The minimum required power for transmission is calculated to get an adequate performance of radio link.
- Keep a high signal-to-interference ratio (SIR) throughout the conversation. At three levels of control speed, power control is implemented, namely:

- ***Open-loop:*** Both downlink and uplink communications use the open-loop power control mechanism. It provides an initial power setting of the MS for the uplink channel. The signal level calculated by the MS in the down-link channel cannot be used by the FDD WCDMA system**.**
- ***Open-loop:*** Both downlink and uplink communications use the open-loop power control mechanism. It provides an initial power setting of the MS for the uplink channel. The signal level calculated by the MS in the down-link channel cannot be used by the FDD WCDMA system [9].

- ***Closed-loop:*** Power control using closed-loop maintains the received signal level of each mobile's uplink signal threshold, reducing interference within the cell.

II.2.5 The handover in UMTS

There are different types of handover in WCDMA:

- **Hard handover**

The term "hard handover" refers to a "hard" change that occurs during subscriber movement from one area to another. The radio links are ruptured from the last cell to the current cell for hard handover.

- **Soft handover**

The term " soft handover" happens when the UE is in overlapping coverage area of two cells. soft handover occurs. The UE can establish links to both base stations at the same time, allowing it to communicate with both and combine them using the UE's signal processing module's radio activated key entry (RAKE) receiver capability. The latter process is only possible in downlink; in uplink, only the best signal is chosen. Having multiple links active during the handover process makes the process more reliable and smoother [11].

- **Softer handover**

The term " softer handover" happens when the UE can receive signals from two sectors mastered by one Node B. This can happen because the sectors overlap, or more commonly because of multipath propagation caused by reflections from buildings, etc.

The signals received by Node B, as well as the signals from the two sectors, can be routed to the same RAKE receiver in the uplink, where they can be combined to provide an enhanced signal. The different types of handover in UMTS are illustrated and compared in Figure II.3.

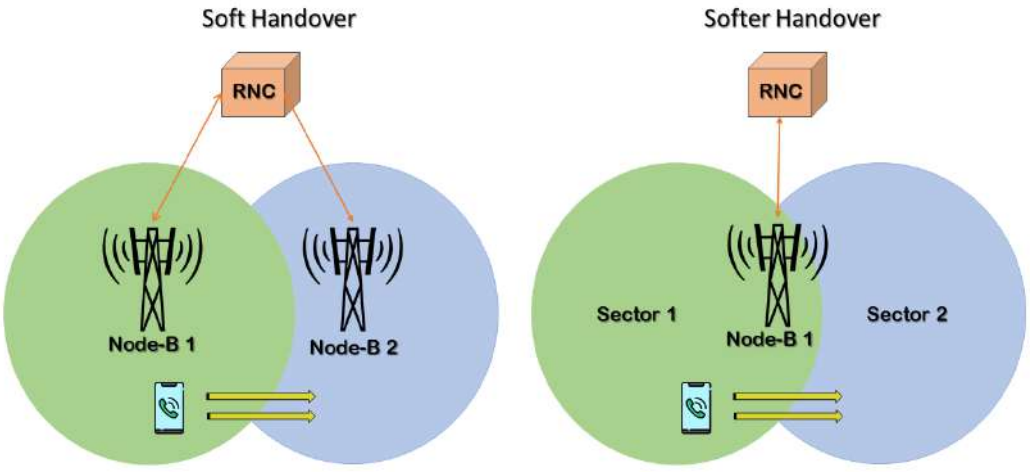


Figure II. 3. Handover types in UMTS network.

- **Inter-RAT handover**

In many cases, switching from the UMTS radio AN to the 2G GSM network is required. Because they are handing over between different types of RAT, these handovers are referred to as Inter-RAT (I-RAT) handovers. I-RAT handover can be divided into two categories.:

- ***UMTS to GSM handover:*** This category of handover is divided into two subcategories.:
  - Compressed mode handover: The UE uses the gaps in transmission that occur when using compressed mode handover to analyze the reception of local GSM base stations [11].
  - Blind handover: This sort of handover happens when the Node B hands off the UE by giving the details of the new cell to it without establishing the link of the mobile for the new cell [11].
- ***Handover from GSM to UMTS:*** The transition from GSM to UMTS occurs to improve performance and is only possible when the conditions are ideal. When this occurs, the UE will be notified by the neighbor list.

II.2.6 Cell selection

The UMTS network's cells can be classified into three different states:

- ***Active set***: All of the cells in the active set receive user information. The active set cells in FDD are involved in soft handover. The active set in TDD always consists of one cell. Only active set cells in dynamically variable cell information list will be considered by the UE [12].
- ***Monitored set:*** Cells that are not part of the active set but are part of the cell information list.
- ***Detected set:*** Includes all cells detected by the UE that are neither in the cell information list nor in the active set. Only intra-frequency measurements taken by UEs are eligible for reporting of the detected set's measurements [12].

There are three scenarios in which the state of any given cell can change:

- When a cell's signal is strong enough to activate it.
- When a cell in the active set becomes weak and needs to be removed.
- When the signal of a non-active cell improves to the point where it is superior to that of an active cell, the new stronger cell replaces the weaker one in the active set.

II.2.7 Major parameters and spectrum allocations in the UMTS network

The majority of WCDMA spectrum is around 2 GHz, with uplink frequencies ranging from 1920–1980 MHz and downlink frequencies ranging from 2110–2170 MHz. This spectrum is a key component of the IMT-2000 standard, which is used in Europe, Asia and Brazil. Re-farming refers to the process of deploying WCDMA in an existing GSM or CDMA frequency band [10]. The main parameters of WCDMA technology are summarized in Table II.1.

Table II. 1. WCDMA's key parameters.

Access method	**Direct sequence CDMA**
Duplex	FDD/TDD
Synchronization in radio access	Asynchronous operation
Chip rate	3.84 Mbps
Frame length	10 ms
Multi service capability	Multiple services with different QoS in concurrent operation
Detection method	Coherent detection using pilot symbols

II.2.8 UMTS channels

Based on the specifics of the service being offered, any service provided to UE is associated with a radio bearer that defines the setting and configuration of radio link control (RLC), medium access control (MAC), and the physical layer. At different protocol layers, the information flow associated with a radio access beacon (RAB) is mapped into different types of channels [10].

The RLC layer creates logical channels, which are then mapped to transport channels in the MAC layer. Control plane and user plane bearers are provided by two types of logical channels: control channels for signaling and traffic channels for bearer. The existing logical channels are summarized in Table II.2.

II.2.9 RRC states

The RRC state machine has two modes: RRC idle mode and RRC connected mode.

II.2.9.1 RRC idle mode

The UE, in idle mode, is identified by non-access stratum (NAS) identities such as TMSI and IMSI rather than by UTRAN. The network can page the MS based on the terminal's position in the registration area. As long as the mobile remains in the same registration area, it is not necessary for the terminal to initiate an RRC connection. When the mobile enters a new registration area, it must switch to RRC connected mode to make a location update. [10].

Table II. 2. Summarizes the existing logical channels in UMTS system.

Channels	Use	Direction
Control channels		
Broadcast control channel (BCCH)	Broadcast of the information system message that contains code sequences, cell identifier, timers,...etc.	DL
Paging control channel (PCCH)	Incoming calls or other messages should be announced to users in the location area	DL
Dedicated control channel (DCCH)	Every user with an RRC connection to the RNC has a bidirectional signaling channel. It sends out measurement reports as well as RRC control messages	DL/UL
Common control channel (CCCH)	Setup of the connection, channel assignment, and cell reselection	DL and UL
Traffic channels		
Dedicated traffic channel (DTCH)	Information for a single user's specific service is transferred. Multiple services may be provided to a single user on multiple coexisting DTCHs at the same time	DL and UL
Common traffic channel (CTCH)	SMS cell broadcast message, for example, is a point to multipoint channel that carries information for a group of users	DL
Transport channels		
Broadcast channel (BCH)	The BCCH logical channel's transport is provided by this component	DL
Paging channel (PCH)	The PCCH logical channel's transport is provided by this component	DL
Random access channel (RACH)	Before a traffic channel can be assigned, short signaling information must be provided during the initial access to the system	UL
Forward access channel (FACH)	A logical channel is carried for a specific UE. RACH's reply	DL
Common packet channel (CPCH)	The RACH channel has been extended to allow for the transmission of larger data packets	UL
Downlink shared channel (DSCH)	Similar to GPRS, a pool of physical resources is allocated on a time transmission interval (TTI) basis to different users based on packet scheduling policy	DL

II.2.9.2 RRC connected mode

The UE can transmit data using allocated radio resources in RRC connected mode. So, each UE has a temporary radio network identity (RNTI) that identifies the mobile phone, in common channels such as RACH or CPCH.

Figure II.4 depicts the message flow for establishing an RRC connection. On the uplink CCCH mapped to RACH, the MS sends an RRC connection request message. The UTRAN, like any other request from NAS, must establish a signaling radio bearer (SRB) between the UTRAN and the UE. The mobile is identified by its 'initial UE ID,' which is a NAS identifier similar to the IMSI or TMSI. The RRC connection request content includes the initial UE ID.

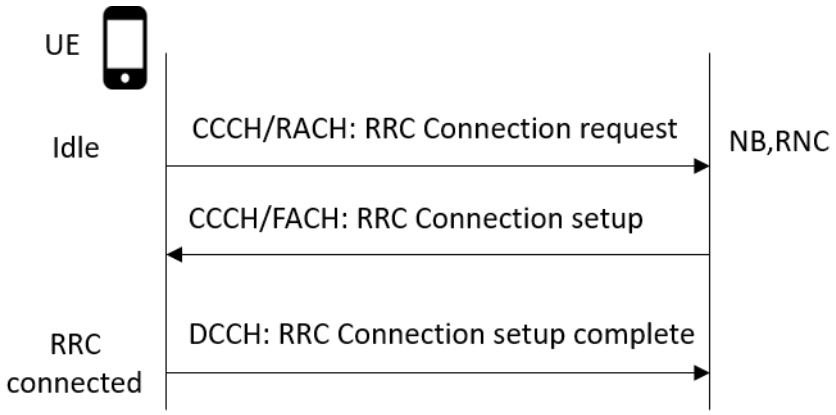


Figure II. 4. RRC connection establishment.

II.2.10 UMTS network optimization

II.2.10.1 KPIs

The KPIs of the UMTS network can be divided into the following subcategories:

- ***Accessibility:*** It is a user's ability to successfully connect to the UMTS network and use the available services.
- ***Retainability:*** It is defined as a user's ability to keep a requested service for the specified amount of time.
- ***Mobility:*** It refers to a user's ability to move between neighboring UMTS and GSM cells while maintaining their requested service (PS or CS).
- ***Integrity:*** represent the user's perception of quality.

CS call drop, PS call drop, RRC connection setup failure, HO failure (3G-3G), IRAT HO failure (3G-2G), and power congestion issue are the major UMTS network KPIs.

II.2.10.2 KPIs optimization

All of the KPIs mentioned above aid in the continuous optimization of the network, which is an essential part of its operations. By monitoring them, several issues within the network can be identified and resolved as needed, as detailed in tables II.3, tables II.4, tables II.5, tables II.6, tables II.7, and tables II.8.

Table II. 3. CS call drop optimization.

Possible reasons	Resolutions
Poor coverage	Adjust the antenna's tilt and azimuth.
Neighbor planning issues	Check for neighbors planning of the cells in the area
Received total wideband power (RTWP) issue	Tracing the source of the external interference by using a spectrum analyzer or check the related hardware
Alarms	UTRAN rectification
Overshooting issue	RF conditions needs to be checked
Scrambling code planning (SCP) clash (cells coverage)	Re-plan the PSC of the cells

Table II. 4. PS call drop optimization.

Possible reasons	Resolutions
Poor RF conditions	Antenna adjustments
Missing neighbor issue	Check for neighbors planning of the cells in the area
RTWP issue	Tracing the source of the external interference by using a spectrum analyzer or the hardware needs to be checked
Alarms	UTRAN rectification
Tracing the particular UE	Check the call history recordS (CHR)logs

Table II. 5. RRC connection failure optimization.

Possible reasons	Resolutions
RRC failure	Apply a RRC reject analysis
Power congestion	Increase the power output or check for RTWP issues
CE congestion	Load balancing
Code congestion	Check the HSDPA codes
No reply	UE issue
UE issue	Cell scenario analysis

Table II. 6. HO failure optimization.

Possible reasons	Resolutions
Overshooting issue	Check RF conditions
Missing neighbor issue	Check cells planning in the area
Handover event threshold	Retune handover event threshold
PSC clash	Replan the cells PSC
Poor coverage	Antenna adjustments

Table II. 7. IRAT HO failure optimization.

Possible reasons	Resolutions
GSM cell discrepancies	Remove all GSM cell discrepancies
Interference on the GSM cell	Trace and resolve the interferences
IRAT parameters	Retune IRAT parameters
Particular cell issue	Filter the cells with continuous IRAT failures

Table II. 8. Power congestion optimization.

Possible reasons	Resolutions
RTWP issue	Check hardware
Power output issue	Increase TCP and CPICH power to increase coverage and capacity
Overshooting cell	Check RF conditions

II.3 LTE Network

II.3.1 Overview of the LTE network

LTE's main goal is to provide a higher data rate than UMTS, as well as low latency and packet-optimized radio access technology that allows flexible bandwidth deployments. At the same time, its network architecture was created with the goal of supporting packet-switched traffic while maintaining high quality of service and mobility. When compared to previous cellular systems, LTE has introduced many new technologies. They allow LTE to operate more efficiently in terms of spectrum utilization while also delivering higher data rates that are requires [13]:

- Orthogonal frequency division multiplexing (OFDM): It is the signal format for LTE because it allows a high data bandwidth to be transmitted efficiently while remaining resistant to interferences and reflections. Since the traffic was carried on a several number of carriers, the system was able to cope even if some were lost due to interference from reflections, the uplink and downlink had different access schemes [13].
- Orthogonal frequency division multiple access (OFDMA): It has many advantages including its robustness in the face of multipath fading and interference. It is used in the downlink, but in the uplink, a single carrier frequency division multiple access (SC-FDMA) instead because the peak of average power ratio of SC-FDMA is lower than that of OFDMA, which will extend battery life.
- Multiple input multiple output (MIMO): The multi-path signals could be used to advantage and increase throughput by using MIMO. When using MIMO, multiple antennas are required to distinguish between the different paths [13].

- System architecture evolution (SAE): While the term "LTE" refers to the evolution of UMTS radio access via the evolved UTRAN (E-UTRAN), the term "SAE" refers to the evolution of non-radio aspects such as the evolved packet core (EPC) network. The evolved packet system (EPS) is made up of LTE and SAE. To route IP traffic from a gateway in the packet data networks (PDN) to the UE, EPS employs the concept of EPS bearers. Between the gateway and the UE, a bearer is an IP packet flow with a defined quality of service (QoS). Together, the E-UTRAN and EPC set up and release bearers as needed by applications.
- IP data: 4G LTE is a completely IP-based data system. Circuit switched voice was included in UMTS, but it is not included in LTE. Originally, it was thought that operators would provide data and that voice would be delivered via over-the-top (OTT) applications. To address this, the GSM-Association (GSMA) established the voice over LTE (VoLTE) scheme as the industry standard. The construction of an IP multimedia subsystem (IMS) core was necessary for VoLTE, but it cost delayed the rollout of this functionality [13].

II.3.2 LTE network architecture

LTE was created to support only PS services, with the goal of providing seamless IP connectivity between the UE and the PDN without causing any interruption to the end users' services when moving.

The uper-level of LTE network is consisted of following three main entities (Figure II.5): The EPC, the UE, and the evolved UTRAN.

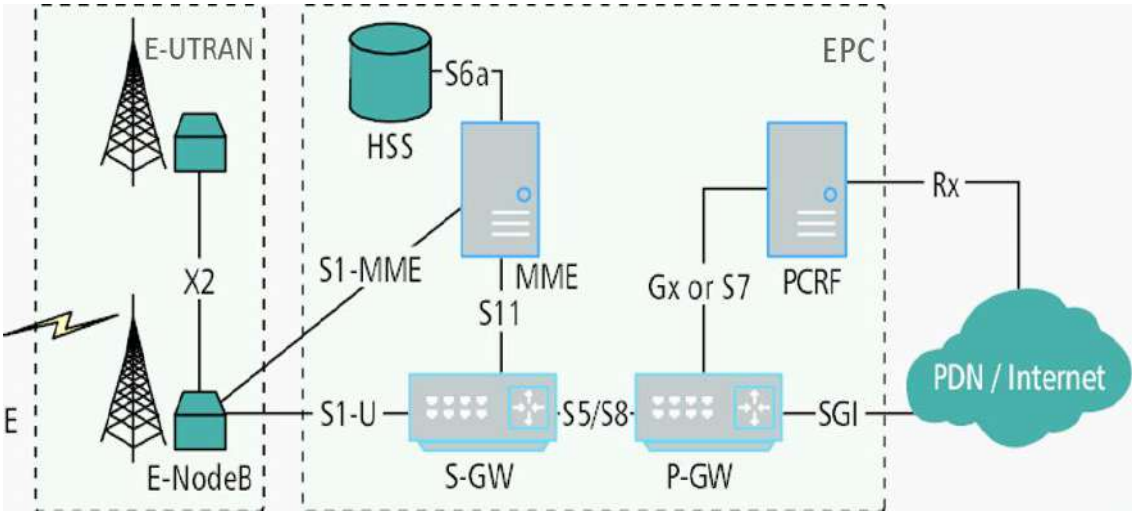


Figure II. 5. LTE Network architecture.

II.3.3 The radio access network

II.3.3.1 The user equipment

The internal architecture of LTE user equipment is similar to that of GSM and UMTS. The following important modules were included in the ME:

- ***Mobile termination (MT):*** It manages all aspects of communication.
- ***Terminal equipment (TE):*** It disconnects the communication stream.
- ***Universal integrated circuit card (UICC):*** For LTE equipment, this is referred to the USIM card. This latter is similar to a 3G SIM card in that it stores user-specific data, such as the user's phone number, home network identity and security keys, ...etc.

II.3.3.2 The E-UTRAN

The eNodeB or eNB, are the single component of the E-UTRAN, and they manage radio communications between the UE and the EPC. Each eNB controls mobile devices as a base station for one or more cells. The base station that communicates with the mobile is known as the serving eNB. LTE Mobile only connects with one base station and one cell at a time, and the eNB provides two basic roles:

- Using the LTE air interface's digital signal and analogue processing functionality, the eNB sends and receives radio transmissions to all mobiles.
- The eNB manages all the low-level operations of the mobiles by sending them signaling messages such as handover commands.

The S1 interface connects each EPC to the eNB, and the X2 interface connects it to nearby base stations, which is primarily used for signaling and packet forwarding during handover.

II.3.4 The core access network

It is made up of components that perform tasks such as mobility authentication, management, quality of service, download and upload IP packet routing, and IP address allocation, among others. The EPC's flat IP architecture enables the network to efficiently and cost-effectively handle large amounts of data traffic [14]. and it's made up of:

- ***The mobility management entity (MME):*** It manages all signaling between the UEs and the EPC, as well as between the eNodeBs and the EPC. The MME's signaling is also known as NAS signaling because it is done using the NAS protocol. Through the S1-AP interface, the MME connects to the eNodeB and performs authentication. It establishes a connection with the HSS and orders the authentication information for the user attempting to join the network. The functions of the MME are as follows:

- Authentication: After exchange the authentication information between the HSS and the UE, UEs can authenticate to the network.
- Mobility management: Allows subscribers to move around the network or between networks.
- Location update: tracks the subscriber's current location and state within the network [14].
- Bearer establishment: If more gateways are available, bearers are established by determining on a gateway routes to the Internet.
- Handover support: Enables inter-eNodeB handover (for handover on the S1 interface).
- While the eNodeB has handover capabilities, when the X2 interface is unavailable, the MME sends handover messages between eNodeBs [14].

The MME is also in charge of generating and allocating temporary identities to UEs, as well as terminating NAS signaling. It verifies the UE's permission to camp on the PLMN of the service provider and enforces UE roaming restrictions. The MME is the network's ciphering/integrity protection and security key management termination point for NAS signaling. It also advocates for the legal interception of signaling. With the S3 interface terminating at the MME from the SGSN, it provides the control plane function for mobility between 2G/3G access and LTE ANs. For roaming UEs, the MME also terminates the S6a interface to the home HSS [14].

- ***The home subscriber server (HSS)*:** Stores the SAE subscription of the user, including the EPS-subscribed QoS profile and any roaming access restrictions. It also contains data on the PDNs to which the user has access. An access point name (APN) or a PDN address could be used. Furthermore, the HSS stores dynamic data such as the MME's identity to which the user is currently connected or registered. The HSS can also include the AuC, which generates authentication and security key vectors.
- ***The serving gateway (SGW)*:** When the UE moves between eNodeBs, the S-GW serves as the local mobility anchor for the data bearers, transferring all user IP packets. When the UE is idle, it saves information about the bearers and temporarily buffers downlink data while the MME initiates UE paging to re-establish the bearers. The S-GW collects data for legal interception and charging. It also acts as a mobility anchor for other technologies like UMTS and GPRS, allowing them to communicate with each other.
- ***The packet data network gateway (P-GW)*:** The P-GW is in charge of assigning IP addresses to UEs, charging and QoS enforcement based on PCRF rules. The P-GW

is in charge of filtering downlink user IP packets and distributing them to the various QoS-based bearers. This is done using traffic flow templates (TFTs). For guaranteed bit rate (GBR) bearers, the P-GW performs QoS enforcement. It also acts as a mobility anchor for non-3GPP technologies like CDMA2000 and WiMAX networks, allowing them to communicate with each other.

- ***The policy and charging rules function (PCRF):*** In the policy control enforcement function (PCEF), which is housed in the P-GW, it is responsible for policy control decision-making as well as controlling the flow-based charging functionalities. The PCRF provides the QoS authorization that determines how a particular traffic data will be handled in the PCEF and guarantees that it is consistent with the subscription profile of the user.

II.3.5 LTE Interfaces

- ***LTE-Uu:*** It connects the UEs to the eNodeBs. It is responsible for all signaling messages between the MME and the eNodeB, and the data traffic between the S-GW and UE.
- ***S1:*** For both the user and control planes, the S1 interface connects the E-UTRAN and the EPC. It is divided into two parts: the S1-AP (GTP-AP) which belongs to the control plane, and the S1-U (GTP-U), which belongs to the user plane [14].
- ***X2:*** Allows two or more eNodeBs to interconnect with each other. The X2 interface is divided into two parts: the X2-C which connects the control planes of eNodeBs, and the X2-U, which connects the user planes of eNodeBs. The structure of the X2-C and X2-U interfaces is identical to that of the S1 interface [14].
- ***S5:*** Connects S-GW and P-GW, it provides user-plane tunneling and tunnel management. It is based on the GTP protocol and is used to connect a non-collocated P-GW for PDN connectivity, and S-GW relocation due to UE mobility [15].
- ***SGi:*** It is responsible for connecting the P-GW to the packet data network. As a provision for IMS services, a packet data network can be an operator's external public or private packet data network or an intra-operator packet data network [15].

Figure II.6 shows the user plane protocol stacks for the LTE network.

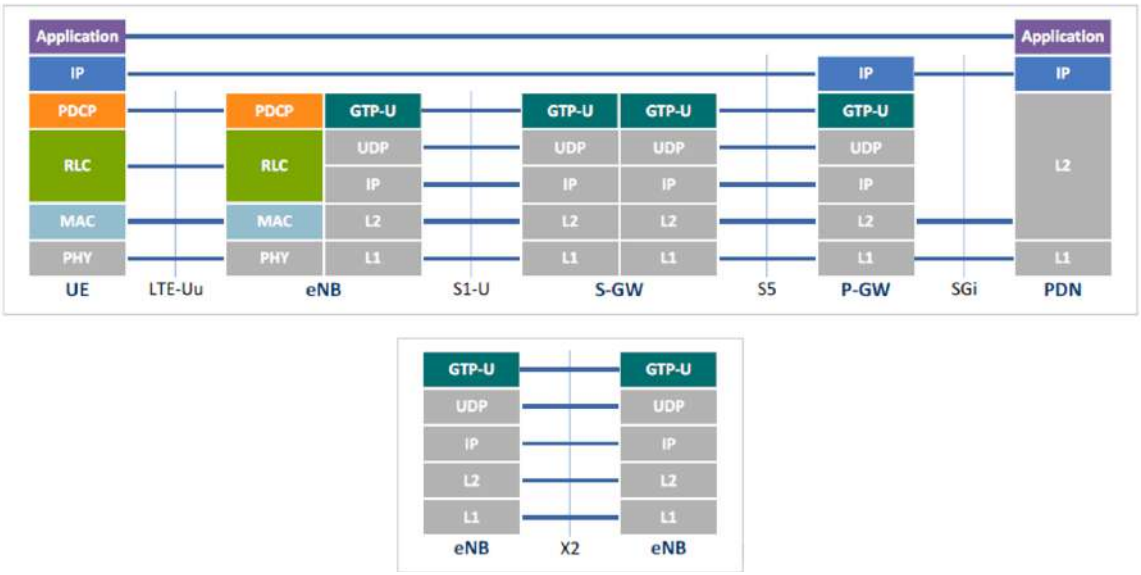


Figure II. 6. LTE user plane protocol stacks.

II.3.6 Handover in LTE

When it comes to mobility, such as handover and reselection, LTE faces significant challenges. In comparison to UMTS, LTE technology makes handover more difficult. Intra-LTE handovers, as well as handovers between LTE and UTRAN, LTE and GSM-EDGE radio access network (GERAN), and other mobile networks, are all handled by LTE.

- **Inter-RAT mobility**

When a UE is turned on, it conducts a cell search to locate and chose an appropriate cell to lodge on. Cell reselection is an idle mode state and the mechanism for switching cells after a UE has been camped on one [16]. Handover is a connected mode state that occurs when two connected states of the different or the same radio access technologies are switched over. Redirection is a process that changes the state of the UE from connected to idle, often for circuit switch full back (CSFB) purposes.

Because a handover is not required and may not be supported by the network or the UE in the case of CSFB, redirection is the default technique to use. Moreover, due to factors like inter-RAT measurement cell/delay detection, 10 time-to-trigger, network handover initiation, and where the call originated or terminated, redirection can happen before handover in some cases [16]. LTE signaling and state transitions are depicted in Figure II.7.

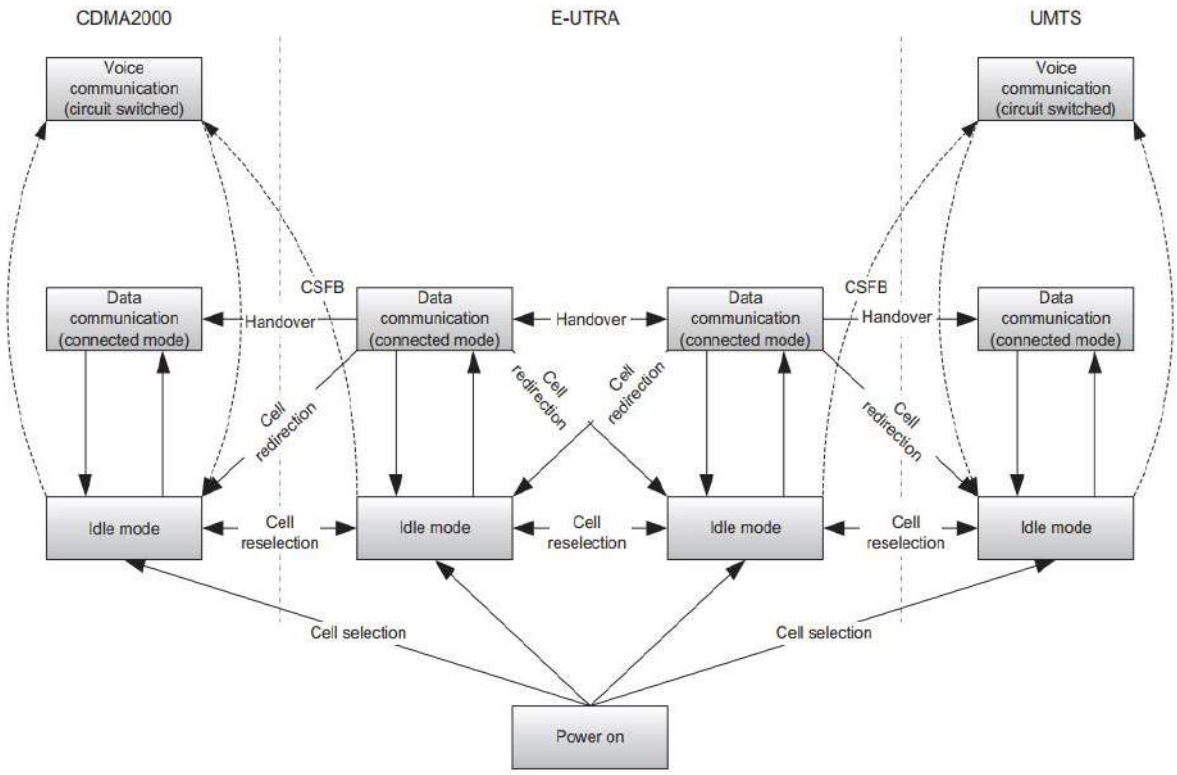


Figure II. 7. Inter-RAT mobility procedures.

- **Mobility towards Home eNodeBs**

Many functions from the source LTE RAN node and the MME are required for mobility towards home eNodeBs (HeNBs). The source RAN node should include the closed subscriber group identity (CSG ID) and the access mode of the target HeNB in the *handover required* message to the MME in addition to the E-UTRAN cell global identifier (ECGI) so that the MME can execute access control to that HeNB. When MME rejects the handover, it sends back *handover preparation failure* message if MME fails the access control and the target HeNB is in closed access mode.

- **Seamless handover**

If the source eNodeB selects seamless handover mode for a given bearer, it suggests to the target eNodeB that a GTP tunnel must be established to operate the downlink data forwarding.

- **Lossless handover**

If the source eNodeB selects the lossless mode for a given bearer, those user plane downlink packets that have been processed by the packet data convergence protocol (PDCP) are still locally buffered because it’s have not been neither sent nor acknowledged by the UE but it will be forwarded over the X2 interface.

II.3.7 LTE radio access type

On both the FDD and TDD fields, LTE is being installed around the world (Table II.9). In FDD mode, two frequency channels with a guard band are used for communication, while in TDD mode, a single frequency is employed to transmit and receive data using different timeslots. FDD employs a paired spectrum, whereas TDD employs an unpaired spectrum [17]. Table II.9 summarizes the differences between LTE-FDD and LTE-TDD.

Table II. 9. The difference between LTE-FDD and LTE-TDD [18].

Parameter	FDD	TDD
Spectrum	Paired spectrum	Unpaired spectrum
Traffic	The asymmetry depends on available spectrum	The asymmetry is dynamically adjustable
Frequency/Time separation	Need a guard band	Need a guard period
Interferences intra-system	Unexpected to occur	eNodeBs synchronization is required
Size of cell	Fit large or small cells	Fit small cells because of guard period
Costs of hardware	Complicate duplexer causes a high cost	There is no significant cost

Figures II.7 and II.8 depict in details the principle of TDD and the FDD modes.

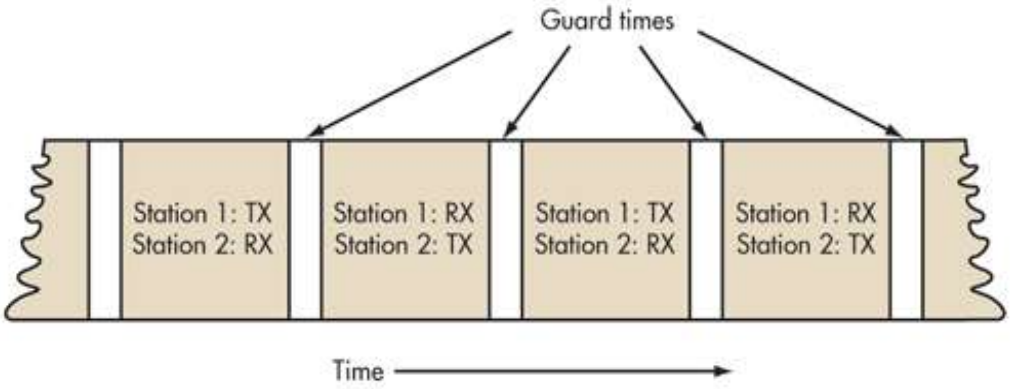


Figure II. 8. TDD mode.

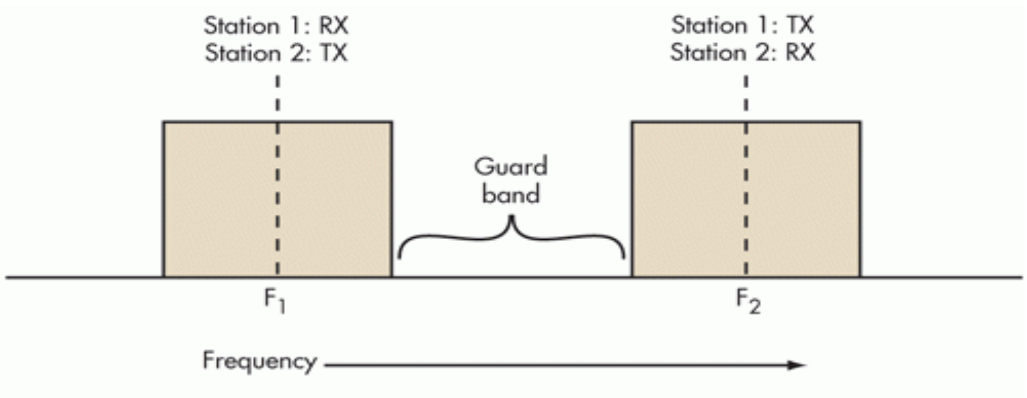


Figure II. 9. FDD mode.

II.3.8 LTE optimization and KPI analysis

Typically, KPI in LTE can be splitted into many subcategories:

- **Accessibility KPI:** Used to determine if the users' requested services can be accessed in a specific condition, as well as the quality of availability when users require it [19]. For example, a user may request access to the network, a voice call, or exchange, ...etc.
- **Retainability KPI:** Used to determine how well a network can keep a user's possession or provide services to them.
- **Mobility KPI:** Used to assess the performance of a network's ability to handle user movement while maintaining service to the user.
- **Integrity KPI:** Used to assess a network's character or honesty to its users, such as throughput, latency, and the number of users served.
- **Availability KPI:** Used to determine whether a network is available, ready, or suitable for users to use services.
- **Utilization KPI:** Used to determine whether a network's capacity has been reached and whether its resource has been depleted [19].

Table II.10 summarizes in details the LTE KPI and its indicators.

Table II. 10. KPIs of LTE RAN.

LTE KPI	Indicators
Accessibility	• RRC setup success rate. • ERAB setup success rate. • Call setup success rate.
Retainability	• Call drop rate. • Service call drop rate.
Mobility	• Intra-Frequency handover out success rate. • Inter-Frequency handover out success rate. • Inter-RAT handover out success rate (LTE to WCDMA).
Integrity	• E-UTRAN IP Throughput. • IP Throughput in DL. • E-UTRAN IP Latency.
Availability	• E-UTRAN Cell availability.
Utilization	• Mean active dedicated EPS bearer utilization.

II.4 Conclusion

This chapter described the main aspects of the LTE and UMTS networks, its architecture with all the different subsystems and entities, all the enhancements and changes that were made to the network compared to the GSM. Also, some of the network's key functions was defined such as all the different types of handovers and all the new functions and concepts that the WDCMA brought, all the logical channels that are used for the communication between the different entities.

CHAPTER III: Horizontal slicing in 5G network

III.1 Introduction

Telecommunications companies and technology businesses all around the world have been collaborating to explore and create innovative technological solutions in order to address rising consumer and industrial demand for mobile data. Fifth-generation (5G) mobile technologies are the next generation of mobile communications technologies aimed at improving current mobile networks.

Faster speeds, more capacity, and the ability to support new features and services are all expected from 5G networks. The latter technologies were created in response to the growing need for mobile data (i.e., more people using more data on more devices). 5G cellular networks are intended to provide throughput of up to 10 Gbps, ten times energy efficiency, and high reliability, and latency up to 1 ms, one thousand greater than usual number of devices.

5G networks are aimed for providing three types of services which are: massive machine type communication (mMTC), ultra-reliable low latency communication (URLLC), and enhanced mobile broadband (eMBB). Network slicing is one of the key technologies in the 5G architecture that has the ability to divide the physical network into multiple logical networks, such as slices, with different network characteristics. It's a useful option for enabling 5G networks to provide a wide range of services and applications, including augmented reality, e-health care, mobile gaming, smart banking, smart farming, and smart transportation systems.

This chapter contains an overview on the 5G network with its architecture and different characteristics, the new features and function that differentiate from the previous generations. A practical part about 5G horizontal slicing will be simulated, to extract users' weights for each service, in order to satisfy service layer agreements (SLAs) between the tenants and infrastructure providers.

III.2 5G services and requirements

The international telecommunication union (ITU) has classified 5G mobile network services into three categories [32]:

- ***eMBB:*** Refers to a straighter forward progression of today's mobile broadband services, allowing huge data amounts and a better user experience, such as supporting even higher end-user data rates, ultra-high definition (UHD) videos, virtual reality (VR), and augmented reality (AR).
- ***mMTC:*** This term refers to services with a large number of devices, such as remote sensors, equipment monitoring, and controllers. Low device cost and low device energy consumption are key needs for such services, allowing for at least several

years of device battery life. Because each device typically consumes low energy and produces a limited quantity of data, high data rate support is less important.

- ***URLLC:*** Aims to meet expectations for the demanding digital industry and focuses on low latency and extremely high reliability services, such as assisted and automated driving, and factory automation.

Figure III.1 depicts the different services that the mMTC, URLLC and eMBB offers in the 5G network.

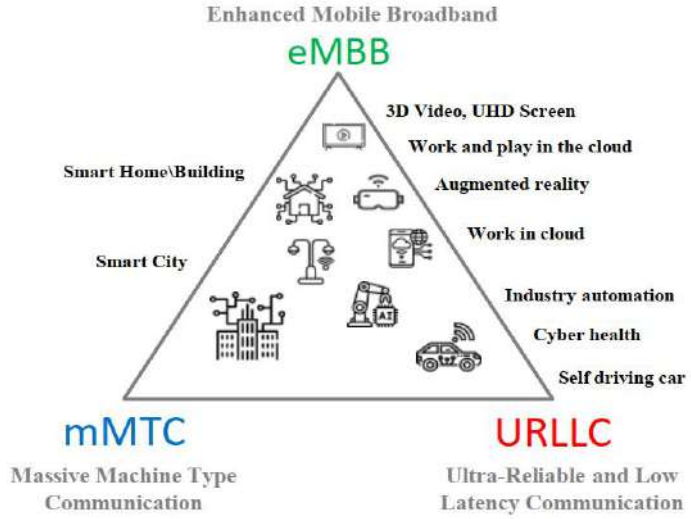


Figure III. 1. Services categories provided in 5G.

By accounting for the majority of needs, the following set of 5G requirements is gaining industry acceptance. Table III.1 summarrizes the quantitative terms of the qualitative requirements of a 5G mobile network.

Table III. 1. The 5G mobile network requirements.

Capability	Description	5G requirement
Downlink peak data rate	Minimum maximum data rate technology must support	20 Gbps
Uplink peak data rate		10 Gbps
Data rate in up-link	In a densely populated urban, test setting for data rate is 95% of the time.	50 Mbps
Data rate in down-link		100 Mbps
Latency	Contribution of the radio network to packet travel time	1 ms (uplink)
		4 ms (Downlink)
Mobility	Maximum speed for handoff and QoS requirements	500 km/h
Density of connections	Total number of devices per unit area	$10^6/km^2$
Efficiencies in energy	Per unit of energy consumption, data is received or sent (by device or network)	Same as 4G
Traffic capacity in area	Traffic totals for the whole service area	10 $Mbps/m^2$
Peak downlink spectrum efficiency	Throughput per unit wireless bandwidth and per network cell	30 bps/Hz

III.3 Network architecture in 5G

A 5G access network (AN) and a 5G core network (5GC) constitute the 5G network (Figure III.2). The 5G new radio (NR) interfaces connect both the 5G new-generation radio access network (NG-RAN), and a non-3GPP AN (e.g. WiFi, xDSL, etc.) to a 5G core network. An underlying TCP/IP transport network connects the various network elements and offers different services (diff-serv) QoS [20].

Figure III.2 illustrates the overall architecture of the 5G network with its different connected entities.

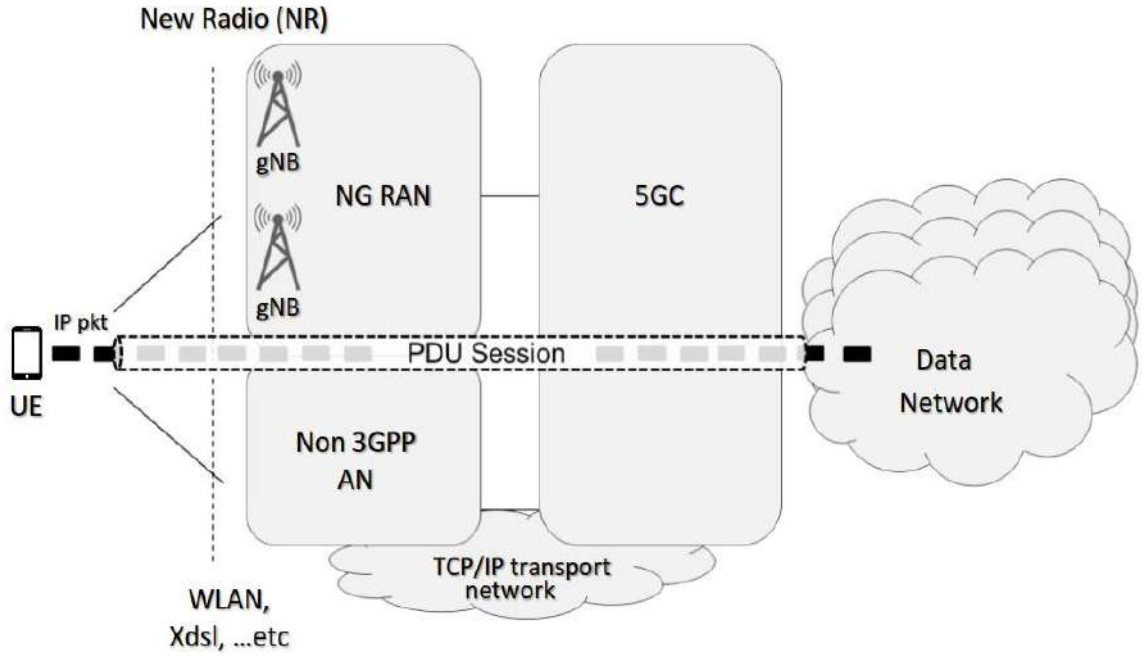


Figure III. 2. 5G subsystems network architecture [20].

The 5G network, like previous generations, links UEs to external data networks. Packet data unit (PDU) session is the name of the 5G connection service. A PDU session is made up of a sequence of new generation tunnels in 5GC and one or more radio bearers on the radio interface. The UE's control functions and the external data network for user traffic exchange are finally connected by this collection of "pipes" (Figure III.3) [20]. The mobile network's primary purpose is to dynamically construct and release tunnels and bearers in order to track user movements and statuses such as idle, connected, and so on.

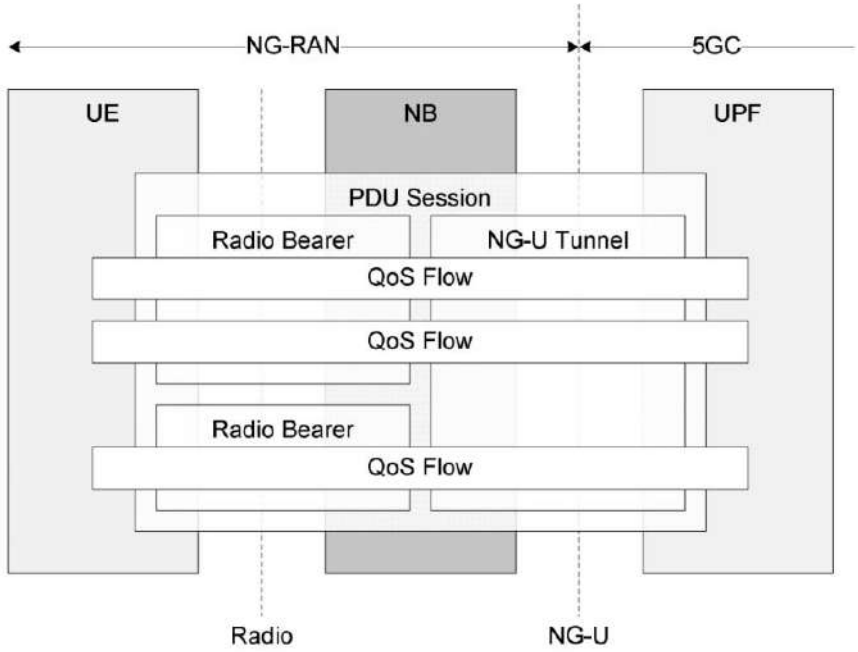


Figure III. 3. PDU sessions and QoS flows for user plane [20].

Except for the QoS scheme and permitted user data units, a PDU session is virtually identical to an EPS bearer in LTE. A PDU session can transmit not only user plane IP packets but also ethernet or unstructured frames, facilitating layer-2 communication among groups of UE. The 5G model is built on a new notion known as QoS flow, which refers to the finer points of QoS differentiation. A single PDU session may have many QoS flows.

The present 5G functionalities in the NG-RAN and the 5GC are shown in Figure III.4. In general, the NG-RAN is responsible for creating, maintaining, and releasing the PDU sessions that cross the radio interface. It handles physical impairments (e.g., fading, interference, power reduction); inter-gNodeB (gNB) handovers (cf. section III.3.1), and scheduling (session multiplexing) [20]. The 5GC functions are in charge of the remaining PDU sessions as well as all other non-radio access operations such as IP address allocation, security, mobility management, and so on.

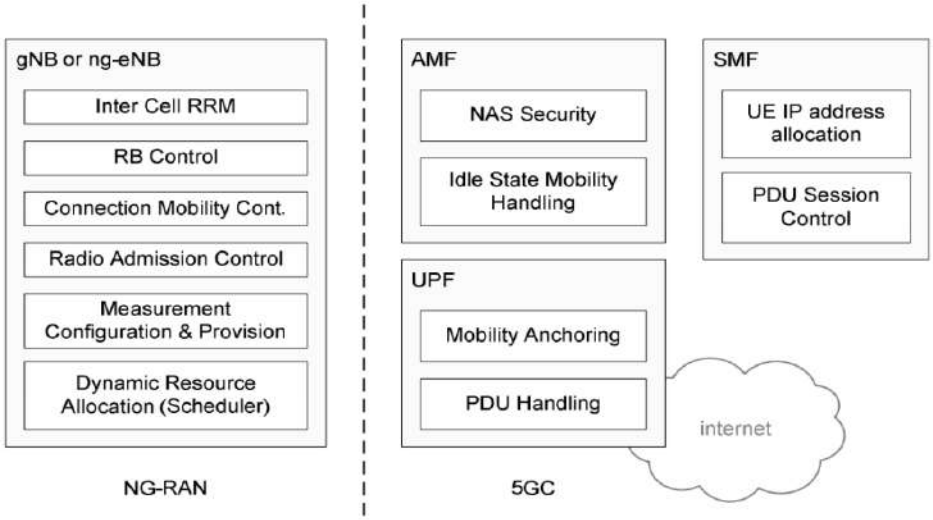


Figure III. 4. Available functionalities in NG-RAN and 5GC.

III.3.1 New generation radio access network

The NG-RAN is made up of a number of 5G base stations (gNBs) that are linked to the 5GC via a series of logical interfaces. gNBs may be linked through the Xn interface, much as LTE, to improve mobility and management capabilities like inter-cell interference cooperation. A gNB's functionality is occasionally distributed. In such situation, the resultant architecture is made up of a central unit (gNB-CU) that uses the F1 interface to operate one or more distributed units (gNB-DU). A remote radio head (RRH), such as the real radio transceiver, is attached to a dispersed unit. The central unit is divided into two sections, one for control plane operations (gNB-CU-CP) and the other for user plane functions (gNB-CU-UP). Figure III.5 shows the overall NG-RAN network architectures and its entities [20].

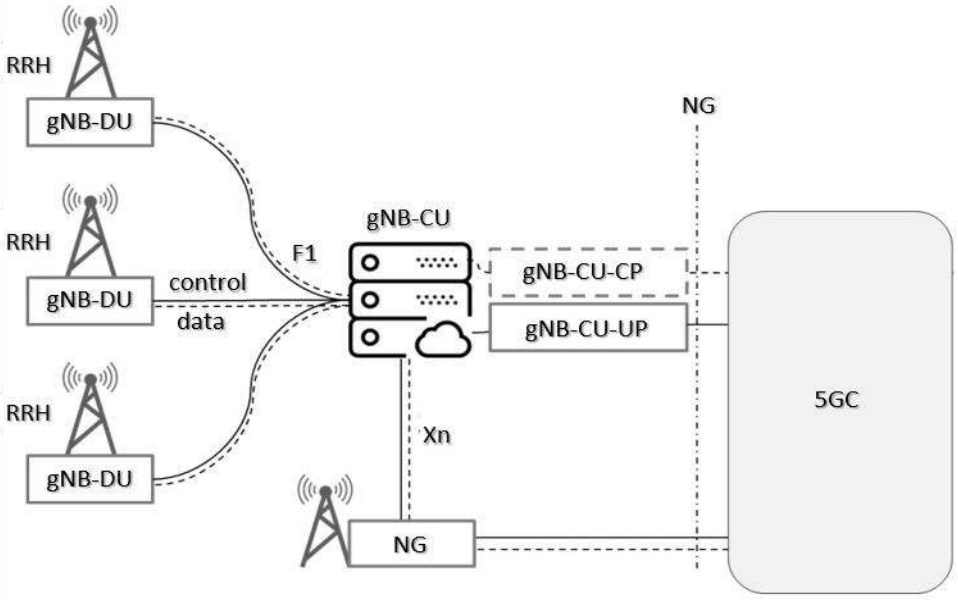


Figure III. 5. Overall NG-RAN architecture.

III.3.2 5G core network

The architecture of the 5G core network is unique compared to legacy networks. The breakdown of the functions performed by previous generations' network nodes resulted in a 5G architecture described entirely in terms of network functions (NF) that are offered as services [20].

In 5G, there exist a control and user plane separation, just like the NG-RAN. There are one or more user plane functions (UPFs) on the user plane, which are responsible for packet forwarding across the many NG-U tunnels that make up the PDU session. The control plane is responsible for all other network functions. The interface modeling has also changed dramatically from previous mobile generations, moving from "bit-oriented point-to-point" to "web-oriented service-based" (from simple bit flows to using APIs). Any pair of interacting 2G, 3G, or 4G network entities has a standardized point-to-point interface (physical or logical), and this interface employs a specified bit-oriented protocol. The interactions between control plane elements in the 5GC are facilitated through service-based interfaces and web-based tools such as they hypertext transfer protocol 2.0 (HTTP/2), representational state transfer (REST) and JavaScript object notation (JSON) [20]. Figure III.6 describes the architecture system of the 5G network.

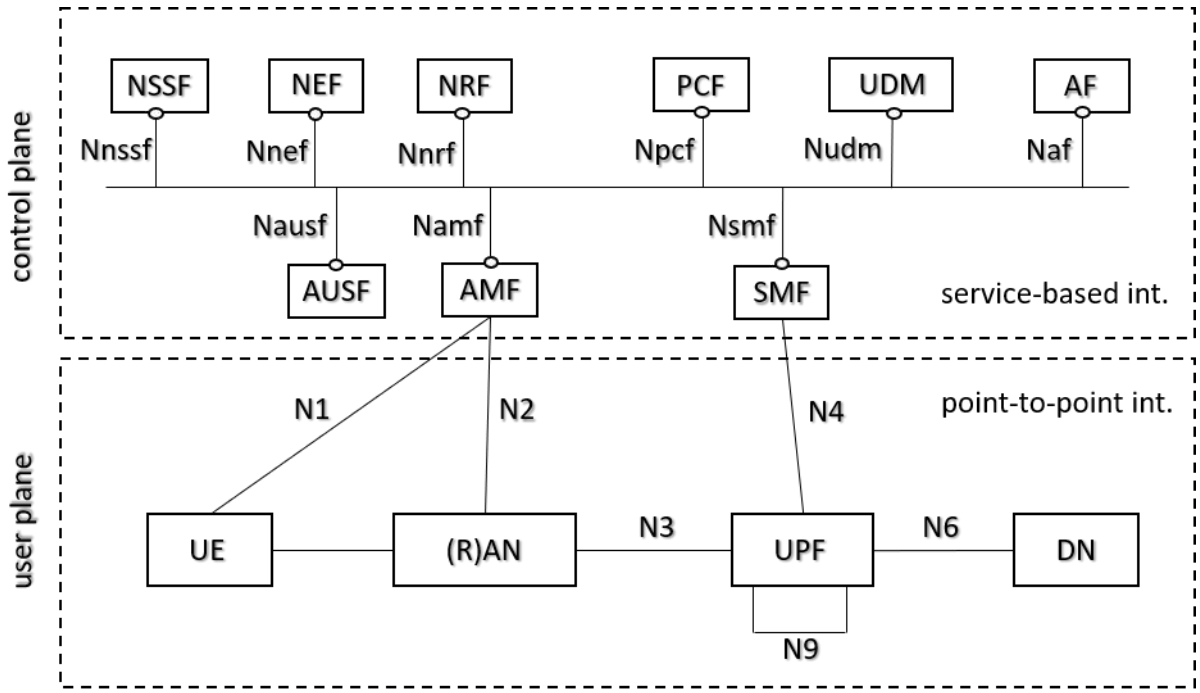


Figure III. 6. 5G system architecture.

The main roles of 5G NFs are the following:

- ***Access and mobility management function (AMF):*** NAS signaling termination, registration management, NAS ciphering and integrity protection, access authentication and authorization, connection management, mobility management, and security context management are all supported by AMF [21].
- ***Session management function (SMF):*** Supports session administration (session creation, modification, and release), DHCP operations, UE IP address allocation and management, DL data notification, NAS signaling termination for session management, and UPF traffic steering setup for optimal traffic routing.
- ***User plane function (UPF):*** Supports packet routing and forwarding, QoS management, and packet inspection, as well as acting as an external PDU session point of connection to DN and an anchor point for intra- and inter-RAT mobility.
- ***Policy control function (PCF):*** Supports the unified policy framework, which provides policy rules for control plane (CP) operations and access subscription information for UDR policy decisions.
- ***Authentication server function (AUSF):*** Serves as an authentication server. It provides UE authentication service to the requester NF.
- ***Unified data management (UDM):*** Authentication and key agreement (AKA) credentials creation, access authorization, user identification, and subscription management are all supported by UDM [21].
- ***Application function (AF):*** Supports application effect on NEF access, traffic routing, and policy control interaction with the policy framework.
- ***Network exposure function (NEF):*** Supports the translation of internal and external data, the public exposure of capabilities and events, the secure transmission of data from external applications to the 3GPP network.
- ***NF repository function (NRF):*** Maintains the NF profile and available NF instances, as well as supporting the service discovery function.
- ***Network slice selection function (NSSF):*** Allows users to choose which network slice instances will serve the UE, determining the AMF set and allowed network slice selection assistance information (NSSAI) to be used to serve the UE [21].

III.4 The role of network function virtualization and software defined networking in 5G

III.4.1 Software defined networking

The goal of software defined networking (SDN) is to make networks more adaptable and flexible. SDN was initially characterized as a method of planning, constructing, and administering networks that separates the control and forwarding planes of the network, allowing network control to be directly programmable and the underlying infrastructure to be abstracted for applications and network services.

SDN's main idea is to decouple the control plane from network hardware and enable external data control via a logical software object known as a controller [22]. The controller is located between network devices and applications and controls packet-flow control to allow intelligent networking (Figure III.7). Network control becomes programmable in this design. Administrators will be able to simply manage the 5G network and deploy new services or adjustments using the controller [22].

Providers will be able to innovate, both in their operations and in their service offerings, thanks to a 5G network that is intelligent, virtualized, and programmable. They will be able to deploy new services on demand, increasing their total efficiency.

The programmability of 5G SDN networks allows new business models and accelerates revenue growth. The architecture will thus be dynamic, controllable, and cost-effective, making it ideal for the high-bandwidth, dynamic nature of 5G use cases.

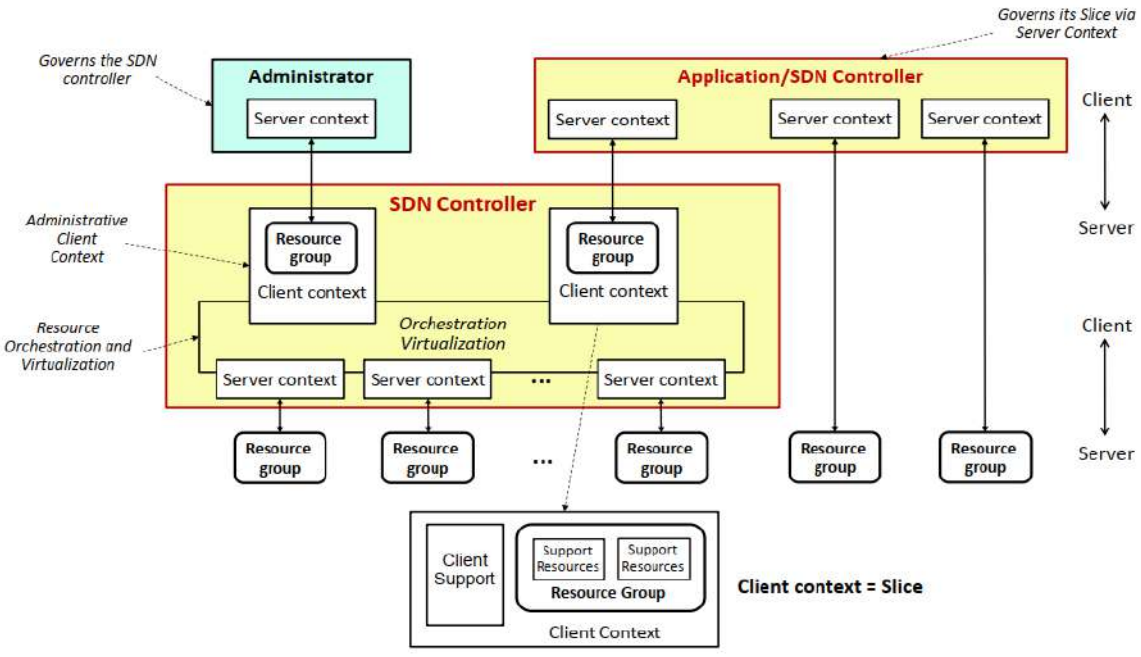


Figure III. 7. SDN slicing network architecture.

III.4.2 Network function virtualization

The replacement of network functions on specialized appliances such as routers and firewalls with virtualized instances running as software on commercial off-the-shelf (COTS) hardware is known as network function virtualization (NFV). The goal of NFV is to change how networks are created and services are delivered. Any business may use NFV to simplify a wide range of network tasks, increase efficiency, and deploy new revenue-generating services more quickly and easily than ever before [22].

NFV is a critical enabler of the next 5G infrastructure, since it helps to virtualize all of the network's numerous equipment. Network slicing, allows virtual network architectural component of 5G to be constructed on top of a shared physical infrastructure. Applications, devices, services, operators, and consumers may all benefit from customizing virtual networks. NFV will also permit the distributed cloud, allowing operators to build more flexible and programmable networks in the future. Figure III.8 illustrates the architecture of the NFV with its different entities.

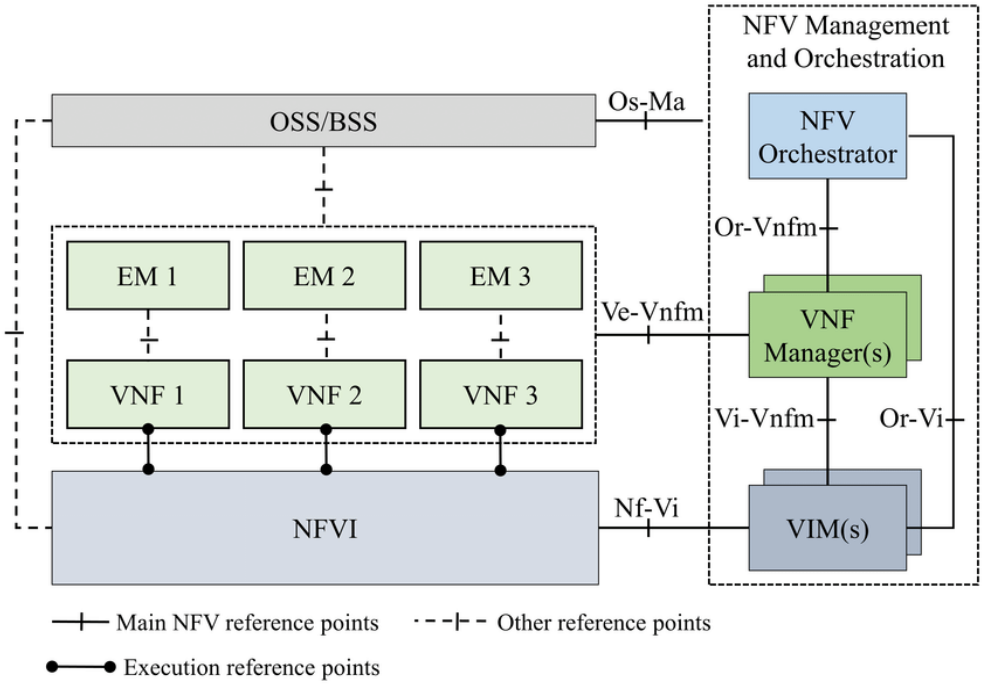


Figure III. 8. NFV architecture.

III.5 Network slicing

The term "5G network slicing" was invented and used for the first time by the next-generation mobile network (NGMN). A network slice, according to the NGMN, is an end-to-end logical network or cloud that is mutually separated and has independent control and management plans that may be generated on demand. A network slice might include cross-domain components from multiple domains in distinct or the same administrations, as well as

components for the transport network, access network, edge network, and core network. Multiple logical and self-contained networks may be built on top of a shared physical infrastructure platform using the network slicing principle.

Network slicing can offer radio, cloud and networking resources to application providers or different virtual segments that have no physical network infrastructure. That way, it enables service differentiation by customizing the network operation to meet the requirements of customers based on the type of service. Figure III.9 represents the NGMN slice capabilities, where each slice represents a single or multiple service offered by a 5G network [23].

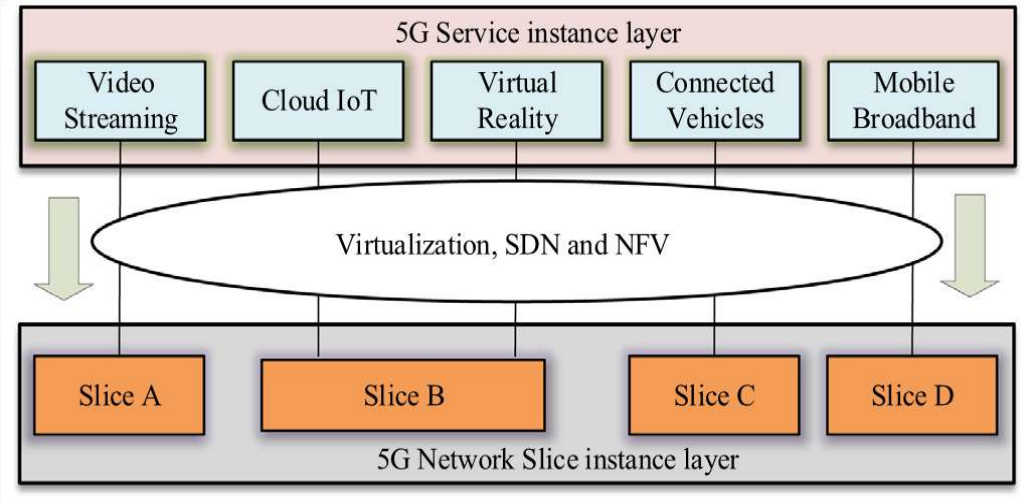


Figure III. 9. 5G network slicing concept.

III.6 The end-to-end architecture of network slicing

Access slices, 5GCN slices, and the selection function that joins these slices into a full network slice make up the network slicing architecture. The selection function directs communications to a selected 5GCN slice that provides certain services. The necessity to accommodate multiple service/application needs is one of the criteria for establishing access slices and 5GCN slices. Each 5GCN slice is made up of a collection of NFs. The system architecture employed in network slicing is shown in Figure III.10 [24].

Some NFs may be utilized across numerous slices, while others are particular to a single slice. This is an essential consideration when slicing. The mapping between 5GCN slices, slices, and access slices can be 1:1:1 (single slice per service) or 1:M:N (multiple slices per service), Figure III.11 illustrates these two factors, for example, multiple access slices can be utilized by a single a device as device B,C and D, and multiple CN slices could connect to an access slice like RAN slice #3 and RAN slice #4.

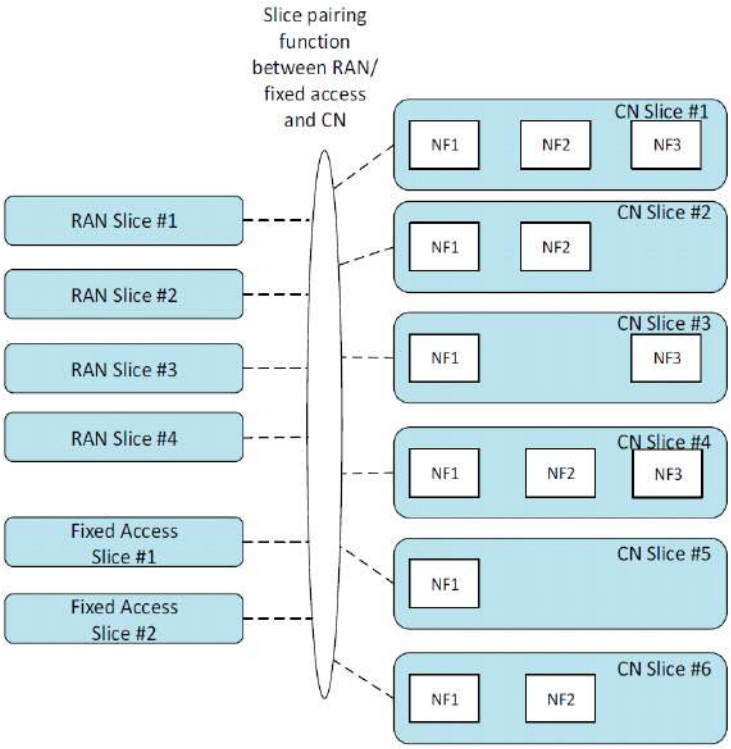


Figure III. 10. Network slicing architecture.

To attain the required network function and communication demands, the pairing of access slices and CN slices might be static or semi-dynamic. A network slice would offer complete NF support to the devices attached to it for the duration of the targeted service lifetime [24].

Examples of network slices include: slices that remote control a factory, a virtual operator and slices that are optimized for streaming video. Figure III.12 illustrates some examples.

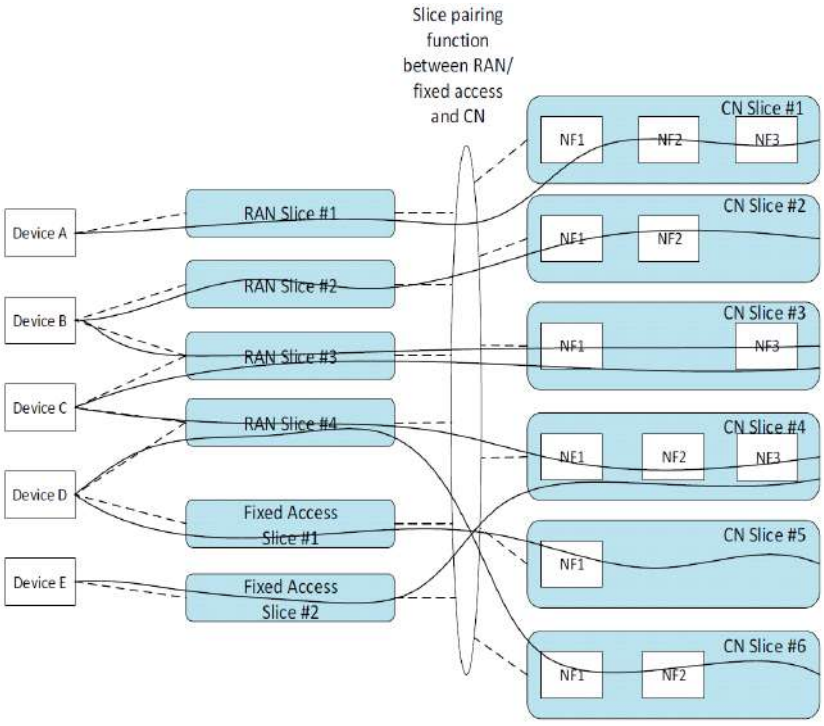


Figure III. 11. UE connection with the network slices.

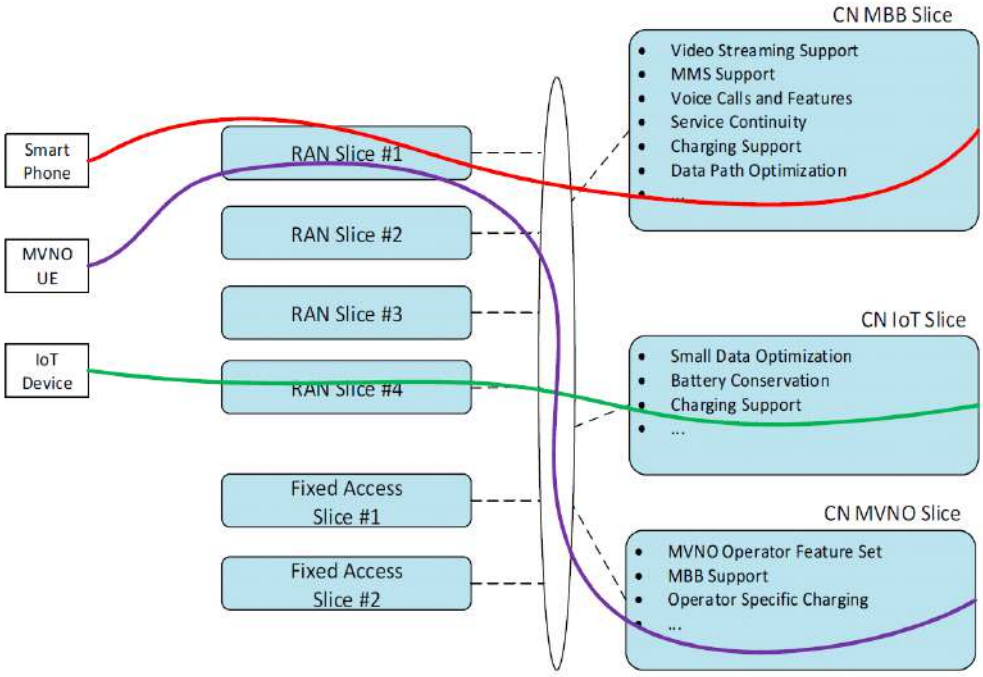


Figure III. 12. Network slice examples.

Some slices, such as the "Smart Phone" and "MVNO UE" devices served by RAN slice #1, can share an access slice, as illustrated in Figure III.12. Slices servicing various utility companies, for example, can all be linked to the access slice designed to serve mMTC, while the slice optimized for streaming video can share the mobile broadband access slice (MBB). A dedicated access slice and a 5GCN slice would be included in some slices. A slice for factory remote control, for example, may require exclusive access and 5GCN operations to provide assured latency and dependability. A specialized 5GCN slice suited for fixed broadband access might be added to a fixed broadband access slice [24].

III.7 Types and scenarios of network slicing

In this section we mention the different scenarios that slicing can enable and the different types of network slicing.

III.7.1 Network slicing scenarios

III.7.1.1 Slicing for QoS

Diverse applications with varying QoS needs must be served on the same infrastructure in 5G networks at the same time. However, satisfying the QoS needs of individual applications such as automobiles, smartphones, and huge internet-of-things (IoT) applications on a single network infrastructure is extremely difficult in general. Slices have their own customized topologies and dedicated resources and perform network functions and management policies such as routing, resource scheduling, and admission control. With the network slicing, each type of application can be served on a different slice. A slice's setup or operation may be

adjusted to meet the QoS requirement since it focuses on servicing the same kind of applications with similar QoS requirements. SDN can implement this adaptable setup [25]. Slices can be allocated to individual users (Tenants) based on user demand. Various tenants might own distinct slices that service the same applications. Slices may have varying priorities depending on tenant precedence or QoS limits imposed by SLAs between tenants and infrastructure providers. Slices held by tenants that pay more for premium services, or slices with strong QoS limits on SLAs, for example, might be given more priority [25].

The resources of each slice must be separated in order for the slices to provide guaranteed QoS to apps. Allocating dedicated bandwidths to the slices allows for resource separation. By controlling queues in wired networks, bandwidths may be reserved and allocated to slices. At each router, separate slice queues are configured and their output rates can be changed based on the requests from the slices. However, resource isolation in wireless networks may be difficult owing to radio interference across connections. A transmission of a slice interferes with other broadcasts of other slices due to the omnidirectional signal propagation property of wireless channels. This interference not only impairs the QoS of the other slices but also modifies the available real bandwidths of slices over time. As a result, it is crucial to reduce wireless interference among slices so that slices formed in a wireless network may effectively meet application QoS requirements. To minimize interference, it may be feasible to use wireless resource management technologies that assign and arrange orthogonal wireless channels among connections, such as frequency or time slot. If the number of orthogonal channels is less than the number of slices, however, interference between the slices cannot be avoided enough.

III.7.1.2 Slicing for infrastructure sharing

Future 5G mobile networks will face many problems in terms of network design and essential technologies as the mobile Internet, IoT, and many new network applications and services, such as virtual reality, high-definition video, cloud services, online gaming, and so on, continue to emerge. Network slicing has been presented as a way to deliver tailored services for various business situations in a 5G network in a flexible, efficient, and cost-effective manner. The subject of resource sharing for network slicing in a multi-tenant situation has been actively investigated in recent years utilizing virtualization technology, which is based on the advantages of network slicing in 5G networks. In-network caching, on the other hand, has been considered a promising technology in 5G networks to reduce duplicate content transmission in networks and improve end-user quality of experience (QoE) following extensive research into another new technology known as information-centric networking (ICN) [26].

Although there have been some good publications on network slicing and in-network caching in 5G networks, these two critical challenges have generally been handled

independently in the literature, to the best of our knowledge. Physical storage may be abstracted as a virtual cache resource pool using virtualization technologies. Furthermore, the infrastructure provider (InP) can provide cache resource slicing based on the tenant's needs. As a result, combining these two advanced strategies to increase end-user QoE and cache resource usage in 5G networks is important [26].

The basic idea behind this network slicing scenario is to virtualize a wireless network's RAN domain and then distribute it across many operators. There are two types of slice tenants: slice owners and slice tenants. Based on an agreement, the owner delivers the slice to a tenant. The tenant has complete control over the slice's services and infrastructure. This notion of network slicing allows for the optimization of network cost models in order to increase total revenue while maintaining network scalability [26].

III.7.2 Network slicing types

Vertical and horizontal slicing are two types of slicing that can be used in 5G communication networks. Vertical slicing, which focuses on 5GCN, enables vertical industries and services. Horizontal slicing, on the other hand, enhances system performance and improves end-user experience, and is primarily concerned with RAN infrastructure [28].

III.7.2.1 Vertical slicing

Vertical slicing aims to promote vertical markets and industries. and it is mostly focusing on the core domain of mobile networks. It allows applications and services to share resources and avoids or simplifies a classic QoS engineering challenge. In a vertical slice, end-to-end traffic flows between the 5GCN and the terminal devices, and each of the network nodes performs similar tasks across slices [27,28].

III.7.2.2 Horizontal slicing

Horizontal slicing is a step forward in terms of increasing mobile device capabilities and improving user experiences. Horizontal slicing goes over and beyond the physical limits of platforms. Horizontal slicing allows a network's nodes and devices to share resources. Highly competent network devices/nodes, for example, share their resources like communication, processing, and storage with less capable network devices/nodes, resulting in improved overall network throughput [27,28].

The notion of vertical and horizontal network slicing is depicted in Figure III.13. In the vertical domain, properly designed slice pairing functions slice the physical computation, storage, or radio processing resources in the network infrastructure (as denoted by base stations and servers) and the physical radio resources (in terms of frequency, space, and time,) to form end-to-end vertical slices. When slicing the radio, the RAN, and the 5GCN, the criterion might

be different. To construct end-to-end slices for various services and applications, slice pairing functions are provided to couple the radio, RAN, and CN slices.

Physical resources (in terms of compute, storage, and radio) at nearby tiers of the network hierarchy are sliced to form horizontal slices in the horizontal domain. A machine might work on numerous slices at the same time. A smart phone, for example, can function in a vertical slice on MBB service, a vertical slice on health care service, and a horizontal slice supporting wearable devices [27].

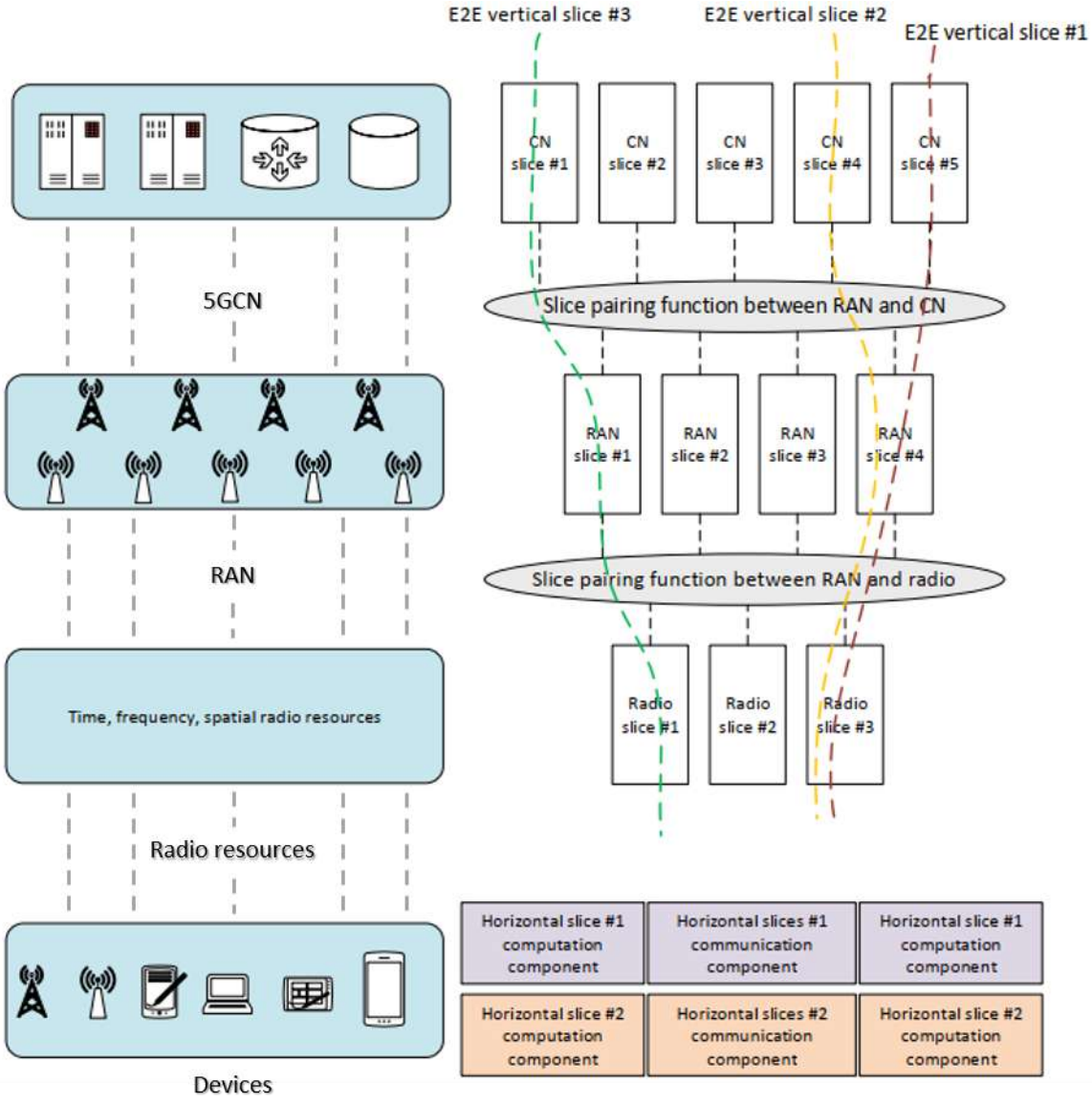


Figure III. 13. Illustration of vertical and horizontal network slicing.

III.8 Network slicing at the RAN level

It's necessary to be aware about two things when analyzing how network slicing is enabled in the RAN [24]:

- The slice's RAT that supports the slice's network services.

- The setting of RAN resources to interact with and support the network slice in the most efficient way possible.

5G must achieve multi-dimensional KPI objectives. Some of the objectives cannot be achieved at the same time. Low latency and dependability, for example, frequently come at the expense of spectral efficiency, huge connection optimization, optimizations for low latency and high reliability, and high spectrum efficiency [24].

Some design and operational requirements on RAN slicing, which are linked with each slice to execute the network services, are:

- A set of configuration rules is applied to the RAN control and user plane operations to support each slice in the RAN.
- Several slices share certain network operations (e.g., mobility management).
- The use of RAN resources among the slices is coordinated by common control functions. This is to boost the overall performance and efficiency of the 5G network.

The following design aspects can be considered in RAN slicing:

III.8.1 Resource management

Radio slices in the RAN can share the radio resources (time, frequency, space) and the corresponding communication hardware (digital baseband processing components, analog radio components) in a dynamic or static manner according to the configuration rules for the network slice. Each slice has access to resources depending on its demand and priority through dynamic resource sharing. Scheduling or contention can be used to share radio resources among the slices. If scheduling is used, each slice sends resource requests to a central scheduler, such as a base station scheduler or a central RAN controller. The scheduler then assigns radio resources to the slices depending on the number of resources requested, the priority of the task being performed by the slice, and the overall traffic load. In a contention-based system, each slice accesses radio resources on its own, following pre-defined criteria. A slice is pre-configured to run in a dedicated resource during its operating time with static resource assignment. The allocation of resources to the slices is assured with static resource sharing. Overall resource utilization can be optimized via dynamic If scheduling is used, each slice sends resource requests to a central scheduler, such as a base station scheduler or a central RAN controller. The scheduler then assigns radio resources to the slices depending on the number of resources requested, the priority of the task being performed by the slice, and the overall traffic load. In a contention-based system, each slice accesses radio resources on its own, following pre-defined criteria. A slice is pre-configured to run in a dedicated resource during its operating time with static resource assignment. The allocation of resources to the slices is assured with

static resource sharing. Overall resource utilization can be optimized via dynamic resource sharing [24].

Figure III.14 depicts a scenario in which each slice is assigned dedicated radio resources. FDM and TDM technologies are used to divide radio resources among the slices. The appearance of RAN configuration to radio access types may change due to differing dynamic resource configuration for distinct radio access types.

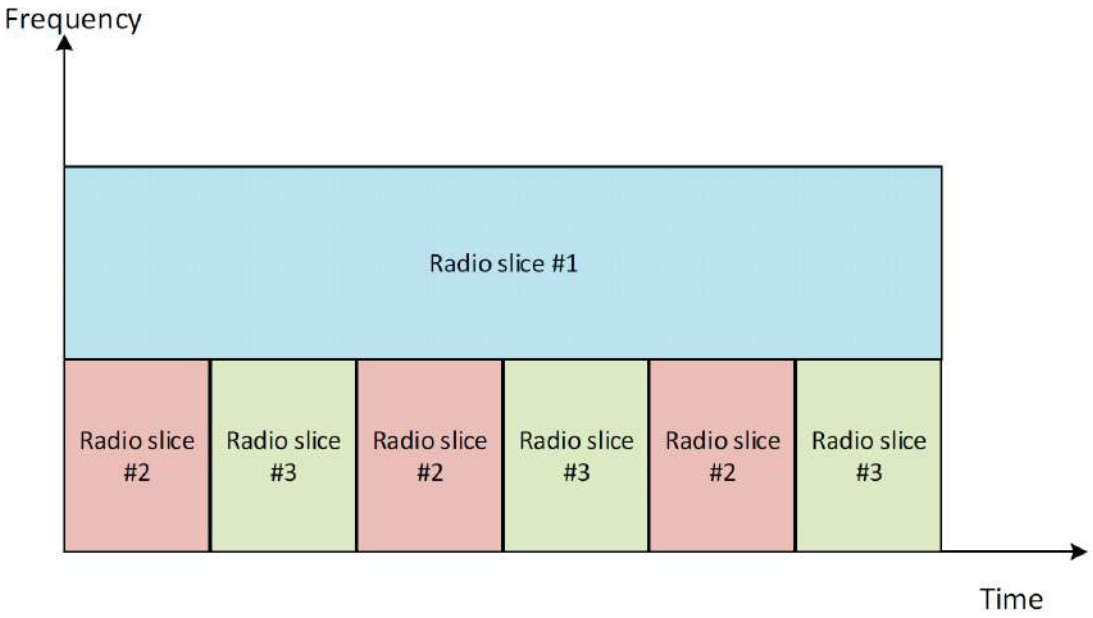


Figure III. 14. Multiple slices in FDM sharing a radio resource such as slice#1 and slice#2 or #3, and TDM such as slice #2 and #3.

III.8.2 Control plane and user plane

In the layer 2 CP, common CP functions are defined, such as slice specific configuration rules, and device idle mode for each of the slices. It's worth noting that not all slices will require the use of all CP functions, and their RAN setting rules will specify which ones are required. In the layer 2 UP, each slice could have a specific configured UP protocol stack. A slice that supports several QoS classes, for example, might be set to employ distinct packet segmentation and scheduling layers (similar to the MAC and RLC layers in LTE). To decrease the size of the packet header, slices with a single QoS class and small packets might be set to employ a single L2 UP. An outer layer MAC for intra-slice scheduling and an inner layer MAC for inter-slice scheduling can be set as part of a hierarchical MAC layer [24].

III.8.3 Control of specific slice admission

To fulfill the initial access needs of distinct network slices, admission control is required. The capacity to accommodate the access demands of a range of network slices is provided by RAN configuration rules. A network slice that serves mission-critical services, for example,

must have assured low-latency access. A UE operating in one slice may be denied admittance to an access point if that slice is not active in that access point, according to slice-specific admission control configuration rules. The common admission control can be used to activate slices in the RAN that have not been activated at an access point or slices that do not require a slice-specific admission control [24].

III.8.4 UE comprehension of RAN configurations

The RAN settings for the various services may or may not be known to the UE. The UE seeks access to a service specified by one core network slice in the first case, and the RAN assigns appropriate radio resources to provide the service [23]. The UE might be unaware of the RAN resource configuration. In the second situation, the UE requires defined RAN configuration and explicit system information signaling before being able to access the service, such as non-default services provided by the RAN.

III.9 Network slicing management framework

A network slice model is defined by NGMN as a detailed description of the configuration, structure, and work flows for how to construct and govern a network slice instance throughout its life cycle. The end-to-end slicing architecture presented in section III.6 is a logical dissection of the network slice instance layer that considers particular network domain functionalities like 5GCN and radio network domains [24].

From an operational standpoint, network operators must meet the following fundamental conditions for network slicing activities, as defined by the 3GPP:

- Create and control services for the pairing of slices feature.
- Create blueprints for network slices.
- Deploy and maintain fault, configuration, accounting, performance, and security (FCAPS) functionalities for the network slices described. Per slice, performance management (PM), fault management (FM), and configuration management (CM) should be available. As a result, operators may track end-to-end service, including important information on the slice's performance in relation to the needed QoS as well as the performance of the NFs.
- Create and maintain the pairing function for slice components (e.g., CN slice to RAN slice pairing rules).
- Network slices may be modified, including adding, removing, and altering network slices.

- Because resource management for a network slice may span operator domains, collaborating resource management domains may be required.
- The slice-as-a-service idea enables a third party to design and maintain network slice configurations (e.g., scale slices) using appropriate application programming interfaces (APIs) while staying within the operator's constraints.

The network management and orchestration (NMO) plane, seen in Figure III.15, offers management and orchestration services for the three levels that make up the sliced network architecture. Each slice is orchestrated and managed individually by NMO functions.

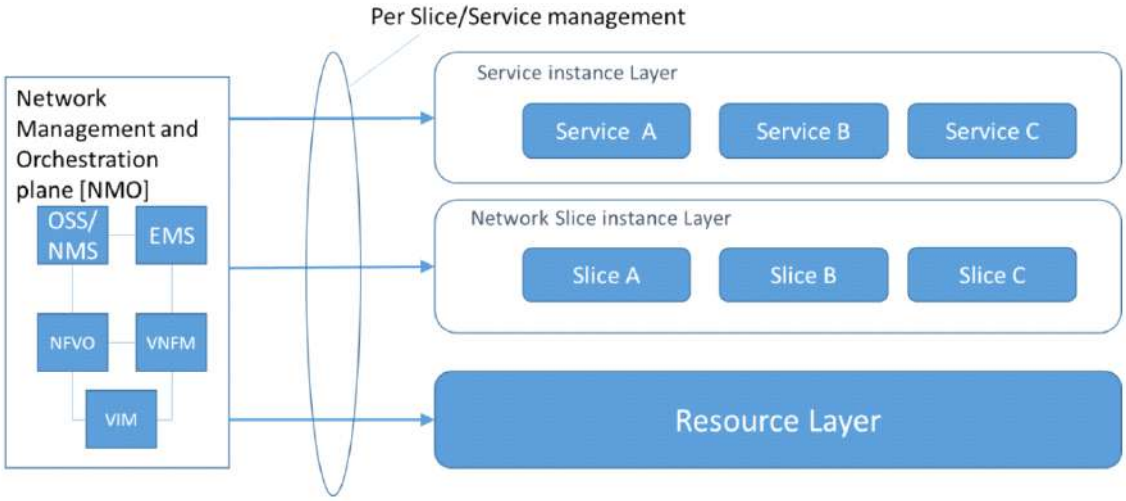


Figure III. 15. Network slice management.

As mentioned in Figure III.15 network slicing management needs to consider the relevant mapping functions used to build a slice. It must also take into consideration the criteria that the service provider employs to establish the mapping's initial state and life cycle decisions depending on network and use dynamics [24]. The necessary pairing functions that must be maintained for network slicing are shown in Figure III.16.

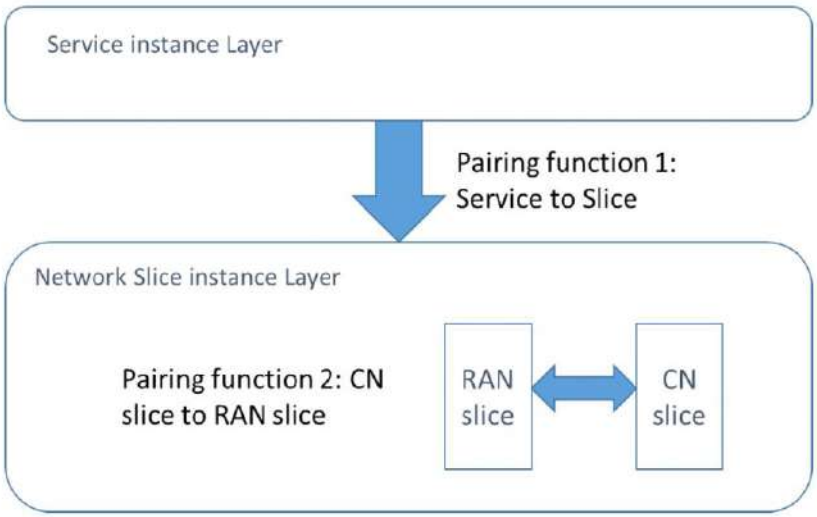


Figure III. 16. Service to network slice pairing.

Figure III.17 depicts how different slices of services can be mapped according to operator policies and business plans. Each slice has a set of characteristics, such as end-to-end low latency or high bandwidth, that make it more suitable for specific types of applications.

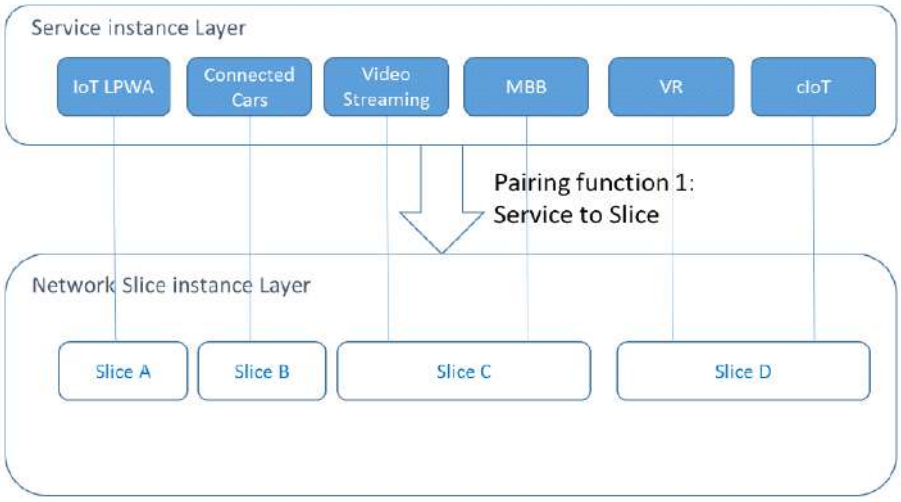


Figure III. 17. Service instance to network slice instance.

From an operational standpoint, the management system must enable operators to map services to slices, such as:

- A provided service must be mapped to a specific network slice by the operator
- A network slice is a collection of network capabilities and SLAs that are suited for a variety of service kinds. As a result, a single network slice may be mapped to several services. An eMBB-oriented slice, for example, could be able to support music downloads, video streaming, and huge file transfers. Another low-latency slice, meanwhile, may handle real-time communications and VoIP.

Different slices may be assigned to users that have the same service type. When tenant isolation is required, each slice is allocated to a certain group of users (for example, local users and roaming users). Another example is geographic separation, which directs users to the slice nearest to them in order to decrease latency and backhaul expenses. A network slice might be allocated to a certain type of service. IoT communications, for example, may be directed to a slice with minimal latency and minimum features, which is optimized for IoT devices. Multiple administrative domains may be involved due to the nature of a network slice. This includes virtualized and non-virtualized wired and wireless network resources. Roaming, for example, is a circumstance in which a network slice instance travels through many administrative domains. Support for business verticals is another example, where different administrative domains may be able to give extra capabilities. Services, network service instances, and network slice instances must thus be managed across various administrative domains. In this

case, the service will require a network slicing orchestration function to manage network slicing. The resource orchestrator, which is made up of network functions virtualization orchestration (NFVO) and 3GPP application resource configurators, is responsible for network slice management at the resource layer [24].

In figure III.18, the OS-Ma interface is used to demonstrate slicing governance.

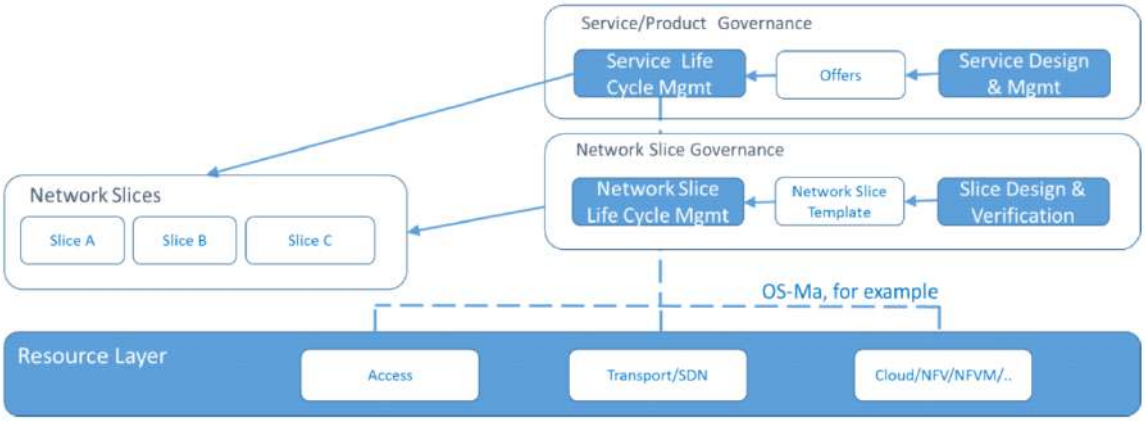


Figure III. 18. Management and resources for network slices.

III.10 Opportunities introduced by the network slicing

Operators may now differentiate themselves through service personalization thanks to the advent of slicing. Various automobile sellers, for example, may operate in different slices when it comes to linked cars. Operators will be able to assign specialized virtual resources to various automobile manufacturers and run them independently. Different services, such as traffic safety, traffic efficiency, news and entertainment, emergency help and security, may be available from different automobile manufacturers. Mobile carriers can offer distinct pricing strategies by differentiating their services. Another area where network slicing might help operators is in monetizing QoE. In 5G, QoS may be defined based on the specific requirements of each application. For example, an operator may construct several slices for different MVNOs, each with a different QoE value, allowing them to monetize their services.

One of the most significant difficulties that today's operators face is that a few misbehaving devices may clog the network, reducing the availability and dependability of network resources for other lawful devices. Network slicing gives operators another option in which particular types of traffic are confined within a slice and the behavior of that network segment has no effect on the other slices. A faulty sensor or actuator in one network slice will have no effect on a key service, such as public safety, in another.

Network slicing might be a best way to resolve one of the wireless industry's problems: how to expand the market to serve vertical services and sectors. The wireless sector will need to continue to improve user experience, expand device capabilities, and handle traffic scalability

in the future [24]. Slicing technology can be used to overcome these issues. Through the air interface, computing resources from network infrastructure and mobile devices may be sliced and combined to construct virtual computation platforms. Communication and computing work together to meet capacity scalability and device capability augmentation goals, thanks to the virtual computation platform. Techniques like edge computation, for example, can terminate traffic at the network edge, easing capacity scaling demands in the deeper network architecture. Higher capacity scaling is easier to achieve at the network edge through densification and radio resource reuse [23]. Communication also permits compute offloading from low-capability devices, such as wearables, to high-capability devices, such as portable devices, or network infrastructure, allowing the low-capable device's potential to be augmented beyond its physical limitations.

III.11 Description of the proposed model

From the perspective of operators, the evolution of mobile networks is a never-ending quest to find a balance between capacity expansion and cost reduction. NFV enables providing customized services to users, since each different virtual network operator (VNO) can define its own protocols over the common set of radio resources. Wireless network virtualization, on the other hand, allows for the deployment of convergent resource management mechanisms, which is critical for constantly evolving heterogeneous networks (Het-nets) [29].

Orchestration, which is the process of service coordination among multiple VNOs to provide each one's desired functionality, it is one of the major challenges in NVF. Each service must be managed in isolation because vertical applications have diverse and often orthogonal service requirements. To cater application-specific customization while ensuring an effective way of using radio resources across slices, convergence among VNOs must be achieved. Those having a network-wide approach for radio access slicing are more effective ways for RRM specifically in Het-Nets. In the mobile network architecture, such models often propose a logical management entity on top of base stations [29]. RRM models that reach entire network not only allow dynamic resource sharing, but they also need minimum adjustments to conventional base stations.

A network-wide RRM approach for virtual RANs is presented in this section of Chapter III. The main contribution is the proposal of a model that promotes multi-tenancy and can realize the idea of service-oriented separation between the roles of InPs and VNOs. In this case, various parties request Capacity-as-a-Service (CaaS) from a centralized virtualization platform in order to meet various contractual SLAs. The fundamental purpose is to maximize the utilization of the entire available capacity given by the aggregation of radio resources from multiple RATs

by serving users according to a proportional fairness framework with pre-defined serving weights and priority.

III.11.1 Network architecture of the model

In this part the conceptual architecture of the network with the functionalities of the components involved. In order to highlight the main difference between the performance of traditional Het-Nets and the model proposed for virtual RANs and to explain the functionalities of the parties involved in the model, Figure III.19 proposes a hierarchical network architecture based on softwarization, which is a conceptual architecture consistent with the business model suggested by 3GPP, in which InPs do not offer services to end users. Each VNO requests capacity as a service from the centralized virtualization platform virtual radio resource management (VRRM). The VRRM is in charge of managing the total available capacity provided by the common radio resource management (CRRM), aggregating all radio resource units (RRUs) from different RATs. These RRUs can be OFDM (carriers for Wi-Fi), OFDMA (resource-blocks for LTE), CDMA (codes for UMTS) and TDMA (time-slots for GSM). The VRRM allows each VNO to deploy its own protocol stack on the same type of RRUs per RAT. The required capacity must be delivered in compliance with the SLA agreements between the VNOs and the InP [29,30].

In network slicing, the low-level physical RRUs are aggregated, abstracted, and virtualized to provide high-level resources, allowing multiple VNOs to share a BS or access point (AP). This allows VNOs to use the same radio resources to deploy different protocol stacks. As a result, the entire available radio spectrum is treated as a single resource that must be divided into virtual resource slices. Furthermore, the process for resource allocation can be summarized as follows: users ask for services from the VNO that are connected to CRRM, informs VRRM about the total capacity that can be available for it based on this information, then VRRM allocates the required data rates to each service.

Three VNOs are proposed and which correspond mainly to the 3 major areas of application of the next fifth-generation (5G) mobile telephone network, in particular eMBB, mMTC, and uRLLC. These three aforementioned VNOs represent three (03) different categories of SLA service contracts depending on the data rate perceived by the user (User Data Rate), distributed as follows:

- Guaranteed bitrate (GB), Guaranteed bitrate (GB), in which the RAN provider promises the VNO a minimum and maximum data rate despite of network state. The VNO is completely satisfied when the maximum promised data rate is assigned to it.

VNOs can have complete control over their networks thanks to the upper barrier in this sort of SLA [29,30].

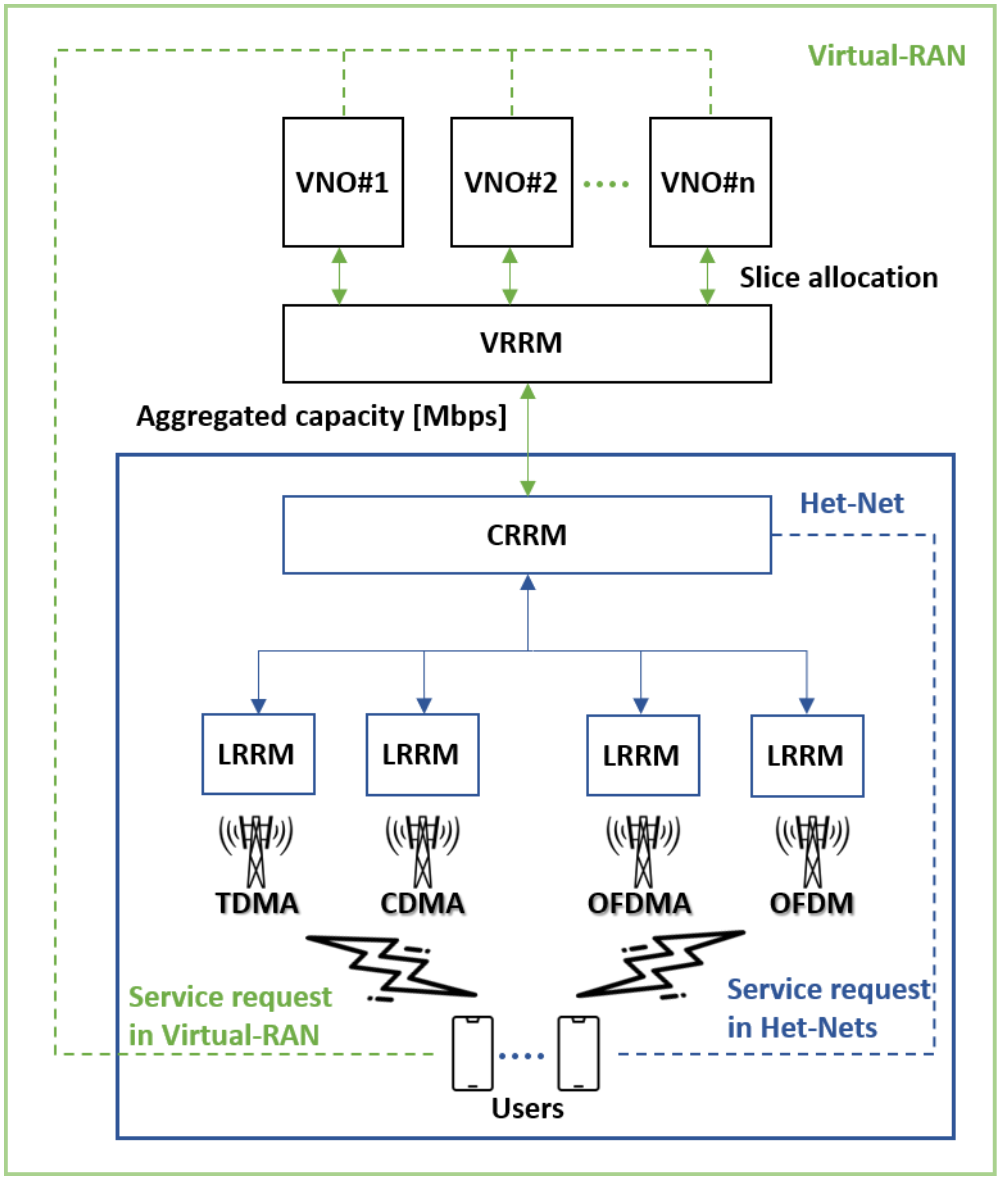


Figure III. 19. Network architecture of the virtual RAN [29].

- Best effort with minimum guaranteed (BG), where a minimum level of service is guaranteed to the VNO. Requests for data rates greater than the specified level are handled with the utmost care; as a result, the minimum guaranteed data rate is the one obtained during peak hours. In this circumstance, although not investing as much as in the past, VNOs may nevertheless guarantee a minimal level of QoS to their subscribers. From the perspective of the subscribers, the adequate service (but not as good as the former ones) is provided at a reduced cost.

- Best effort (BE), in which the VNO is delivered with the pure best effort approach. Operators and consumers may experience reduced QoS and resource scarcity during peak hours in this instance, but the accompanying cost will be lower as well.

III.11.2 Model objective

The main objectives of this model are:

- Promote the coexistence of a multiple VNO of the same InP.
- Separation between the role of the InPs and VNOs.
- Assurance of the principle of CaaS to satisfy the different SLAs.
- Maximize the utilization of the total capacity provided by resource aggregation.

This model is initially based on the following two concepts:

III.11.2.1 Resource allocation

When allocating resources in Het-Nets, the VRRM model is designed to strike a balance between efficiency and fairness. Sm detailed in equation 1, Maps a part of network bandwidth that users employ it a measure of efficiency, quantifying the presumed satisfaction of users given in the allocated resources [29].

$$Sm = \max_{\boldsymbol{w^{usr}}} \sum_{vs=1}^{N^{srv}} \lambda_{vs} \log\left(\sum_{i=1}^{N_{vs}^{usr}} w_{vs,i}^{usr} \frac{R_{vs\,[Mbps]}^{srv_{max}}}{R_{[Mbps]}^{CRRM}} \right) \quad (1)$$

w^{usr}: users weights vectors.

λ_{vs}: tuning weight of service s from VNOv.

N_{vs}^{usr}: number of users per service s.

N^{srv}: total number of services of all VNOs combined.

R^{CRRM}: total offered capacity from CRRM to VRRM.

$R_{vs}^{srv_{max}}$: maximum data rate that can be assigned to a VNOv service user.

Such that:

$$\lambda_{vs} = \gamma_v \ \delta_s \quad (2)$$

δ_s : service s weight from VNOv.

γ_v : VNOv weight.

By optimizing the allocation of radio resources for different applications, VRRM creates a bridge between the functionalities of the MAC and higher layers.

It is useful to point out that δ_s projects the internal policy of each VNO in the context of capacity distribution through the services provided and that its values are fixed for each VNO in order to facilitate the calculations. This nonlinear optimization function will make it possible

to find the optimal values of the vector relating to the weight of users in order to satisfy the various customers of the SLAs available, subject to the linear constraints cited in (III.11.3).

The allocation principle is based on priority based on weight, for example, a service with a higher priority receives a higher service weight and vice versa. The principle of weight assignment serves to prioritize the allocation of data rates to different services of different VNOs. Services with a higher weight are served with higher data rates. The choice of these weights is based on the level of service contracts between VNO and VRRM.

III.11.2.2 The constraints

There are constraints in the allocation of data rates that must not be violated and must be respected in order to achieve the aforementioned objectives while satisfying the optimization function put in place.

- ***Constraint with respect to the data rate of the service s:*** The level of service offered to be guaranteed imposes the next set of constraints. The data rates assigned to each service should be above the guaranteed minimum. For best effort and best effort with guaranteed minimum services, there is no upper limit; however, the data rate allocated for guaranteed services is limited to the maximum. The constraint in question can be expressed as follows:

VRRM's goal function must be solved while maintaining to certain limitations. Each user's average long-term data rate must fall within the VNOs-defined tailored range of data rate changes, which may be represented as detailed in equation 3:

$$R_{vs\,[Mbps]}^{srv_{min}} \leq w_{vs,i}^{usr}\, R_{vs\,[Mbps]}^{srv_{max}} \leq R_{vs\,[Mbps]}^{srv_{max}} \quad (3)$$

$R_{vs}^{srv_{min}}$: Minimum data rate that can be assigned to a user of service s, according to VNOv.

- ***Constraint against the total capacity allocated to each VNOv:*** Further constraint is the capacity between VNOs and InPs, which varies depending on the SLA type This latter can be expressed as the total throughput of all users using the different services of any VNO, in other word, it represents a range of data rates acceptable for each VNO varied between min and max. Formally, the total capacity constraint allocated in each VNOv is expressed in equation 4:

$$R_{v\,[Mbps]}^{vno_{min}} \leq \sum_{vs=1}^{N_v^{srv}} \sum_{i=1}^{N_{vs}^{usr}} w_{vs,i}^{usr}\, R_{vs\,[Mbps]}^{srv_{max}} \leq R_{v\,[Mbps]}^{vno_{max}} \quad (4)$$

Where:

$R_v^{vno_{max}}$, $R_v^{vno_{min}}$: The max, the min contracted data capacity of the VNOv with the InP.

N_v^{srv}: Number of services provided by the VNOv.

- ***Constraint in relation to the total aggregate maximum capacity of the network:*** The principle constraint is respecting the total network capacity provided by the CRRM. The sum of all data rates assigned to all services of all active VNOs must not exceed the total aggregate capacity of the network, and can be calculated according to equation 5.

$$\sum_{v_s=1}^{N_v^{srv}} \sum_{i=1}^{N_{v_s}^{usr}} W_{v_s}^{usr} \, R_{v_s[Mbps]}^{srv_{max}} \leq R_{[Mbps]}^{CRRM} \tag{5}$$

The proposed problems could be solved by CVX optimization tool.

III.11.3 CVX optimization tool

CVX is a MATLAB-based convex optimization modeling system [31]. CVX converts MATLAB into a modeling language, allowing users to specify restrictions and objectives using regular MATLAB expression syntax. CVX enables a disciplined convex programming approach to convex optimization in its default mode. Convex functions and sets are constructed with this method using a minimal number of principles from convex analysis and a base library of convex functions and sets. These rules automatically translate constraints and objectives into a canonical form and solve them. Consider the convex optimization model below:

$$Maximize \ \| Ax - b \|_2$$
$$Subject\ to \ Cx = d$$
$$\| x \|_\infty \leq e$$

The following MATLAB code generates and solves a random instance of this model:

```
m = 20; n = 10; p = 4; %initialization of the variables
A = randn(m,n); b = randn(m,1);
C = randn(p,n); d = randn(p,1); e = rand;
cvx_begin % creates a placeholder for the new CVX specification
    variable x(n) % declares x to be an optimization variable of dimension n
    minimize (norm(A * x - b, 2)) % specifies the objective function to be minimized
    subject to %adding the constraints
        C * x == d
        norm(x, Inf)<= e
```

The least-squares problem is solved when MATLAB reaches the cvx_end command, and the result is written to MATLAB variable [31].

III.12 Case scenario study

Table III.2 shows the different parameters and priorities of the three VNOs used for this study. In this scenario, there are three VNOs, each with a different SLA contract (GB, BG, BE).

Table III. 2. VNO's parameters and priorities.

VNO	SLA	Weight	$R_{[Mbps]}^{vno_{max}}$	$R_{[Mbps]}^{vno_{min}}$
1	GB	10	0.7 Rcrrm	0.4 Rcrrm
2	BG	5	Rcrrm	0.4 Rcrrm
3	BE	1	Rcrrm	0

The network parameters specifications are presented below (Table III.3). Each one of the three VNOs is in charge of providing three different types of service: voice, IoT and eMBB. As well as the user mix (% share) of each service, and their allocated weights as mentioned above. It also shows the different data rates based on the deferent SLAs contracts. Lastly the maximum data rate capacity of the totality of the network is set to 630Mbps (R_{CRRM}).

Table III. 3. Network parameters.

R_{CRRM}	VNO	Service	$R^{srv_{min}}$	$R^{srv_{max}}$	User %	$R_{[Mbps]}^{vno_{min}}$	$R_{[Mbps]}^{vno_{max}}$
630	GB	Voice	0.032	0.064	20	$0.4R_{CRRM}$ =252	$0.7R_{CRRM}$ =441
		IoT	0.5	1	50		
		eMBB	4	R_{CRRM}	30		
	BG	Voice	0.032	0.064	20	$0.4R_{CRRM}$ =252	R_{CRRM} =630
		IoT	0.5	1	50		
		eMBB	4	R_{CRRM}	30		
	BE	Voice	0	0.064	20	0	R_{CRRM} =630
		IoT	0	1	50		
		eMBB	0	R_{CRRM}	30		

With the help of CVX implemented in MATLAB, a simulation was ran to dictate the weights and data rates allocated to individual users of each of the 12 services. The preliminary results obtained from $w_{vs,i}^{usr}$ with the data rates provided and corresponding to each service are given in Table III.4.

Table III. 4. Results obtained from $w_{vs,i}^{usr}$ with service flow rates s.

R_{CRRM}	VNO	Service	$R^{srv_{min}}$	$R^{srv_{max}}$	Service weight	VNO weight	w^{usr}	R^{srv} provided	R^{vno} provided
630	GB	Voice	0.032	0.064	5	10	1	0.064	315.1509
		IoT	0.5	1	4		1	1	
		eMBB	4	R_{CRRM}	3		0.014	8.82	
	BG	Voice	0.032	0.064	5	5	1	0.064	252
		IoT	0.5	1	4		1	1	
		eMBB	4	R_{CRRM}	3		0.0106	6.678	
	BE	Voice	0	0.064	5	1	1	0 .064	62.8491
		IoT	0	1	4		0.7036	0.7037	
		eMBB	0	R_{CRRM}	3		0.0014	0.882	

R^{vno} provided is basically the data rate share of each VNO, and R^{srv} provided presents the data rate allocated to each service. The results shown in Table III.4 clearly states that the w^{usr} is highly dependent on the weights of both of the VNO and service however, the data rate depends on the $R^{srv_{max}}$.

By analyzing the results obtained previously, the following is observed:

- The weight values W for the users requesting both voice and IoT services are equal to the maximum value which is 1, and this depends on their priorities and the low speeds required,
- The data rate provided for each service is guaranteed for all users while respecting the constraint linked to the data rate of the service set up by the model,
- The total capacity allocated to each VNO is sufficient to serve all active users and also respects the second constraint, which perfectly reflects the reliability and performance of the model in terms of capacity sharing,
- It is also observed that both of the voice and IoT services don't use high data rate, as they just need a consistent data stream to function properly. As for the eMBB service it needs a marginally higher data rate, however, it differs for each VNO. (γ_{BG})

It is assumed that the weight of VNO BG (γ_{BG}) fluctuates between 10 and 1 and both of the GB and BE VNOs are fixed to their initial values to test the impact of weight tuning on the overall capacity share of the VNOs. Results for 100 users per VNO are illustrated in Figure III.20. For the lower values of γ_{BG}, although the first assumption is to have equivalent levels of capacity sharing with VNO BE (since the weights of the BE and BG VNOs are similar), it

can be seen that VRRM maintains the values of BG at 40%. This is VNO BG's guaranteed minimum capacity, which is substantially greater than VNO BE's shared capacity. By incrementing γ_{BG} the capacity share stays at 40% up γ_{BG} equals seven (7), where the capacity share of VNO GB and VNO BG get closer and finally converge at 10the capacity share of VNO BG and VNO GB converge to each other and finally are equal when γ_{BG} equals ten (10).

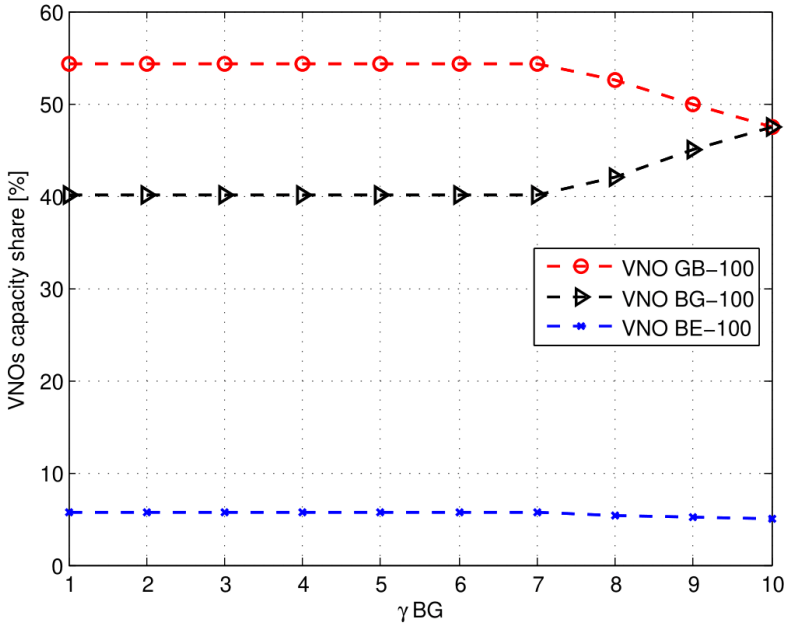


Figure III. 20. The impact of VNOs' weights on the capacity share.

These findings show that all VNOs are satisfied with their SLAs separately within the physical infrastructure of InP, whereas all available capacity is shared 100% between the three VNOs.

III.13 The influence of the quality of the VNO radio link

In this experiment we ran our model through different simulations where we fixed the total number of users to 300 user, and each VNO has the same amount of users (100 user per VNO) and we vary the max capacity of the radio link the network (R_{CRRM}) from -50% to +50% with adding 25% each time of its original value of R_{CRRM} which is set to 630Mbps. The aim is to study its influence on the overall data rate allocated to the individual users of the three VNOs. The results of the variation of R_{CRRM} on the weight of users w^{usr} are presented in Table III.5, Table III.6 and Table III.7 for VNO1, VNO2, and VNO3 respectively.

Table III. 5. The influence of quality on the VNO 1 radio link.

Service	-50% R_{CRRM}	-25% R_{CRRM}	R_{CRRM}	+25% R_{CRRM}	+50% R_{CRRM}
Voice	1	1	1	1	1
Iot	0.7857	1	1	1	1
eMBB	0.0127	0.0132	0.014	0.0144	0.0149

Table III. 6. The influence of quality on the VNO 2 radio link.

Service	-50% R_{CRRM}	-25% R_{CRRM}	R_{CRRM}	+25% R_{CRRM}	+50% R_{CRRM}
Voice	1	1	1	1	1
Iot	0.5	1	1	1	1
eMBB	0.0127	0.0097	0.0106	0.0112	0.0115

Table III. 7. The influence of quality on the VNO 2 radio link.

Service	-50% R_{CRRM}	-25% R_{CRRM}	R_{CRRM}	+25% R_{CRRM}	+50% R_{CRRM}
Voice	1	1	1	1	1
Iot	0.0786	0.4993	0.7037	0.908	1
eMBB	0.0003	0.0013	0.0014	0.0014	0.0015

Figure III.21, figure III.22 and figure III.23 represents the obtained results of the variation of the maximum capacity shared R_{CRRM} on the weight of service R^{srv} for VNO1, VNO2 and VNO3 respectively.

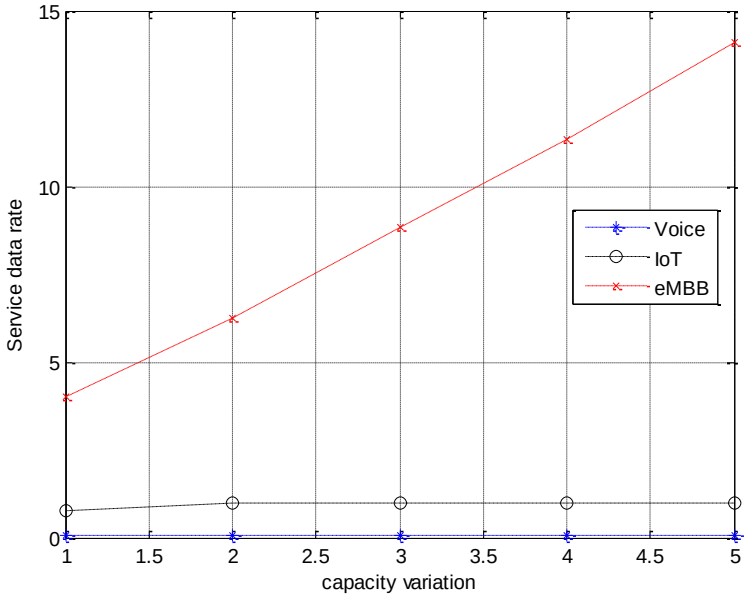


Figure III. 21.The influence of the quality of the VNO1 radio link.

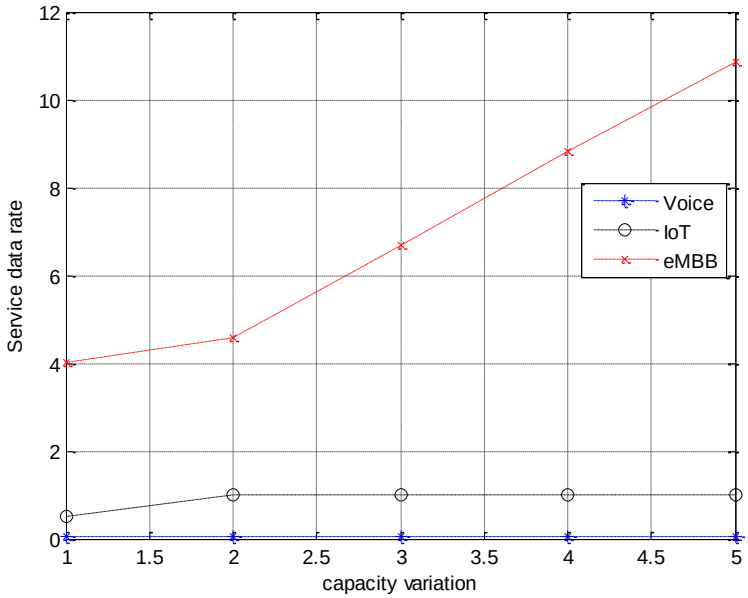


Figure III. 22. The influence of the quality of the VNO2 radio link.

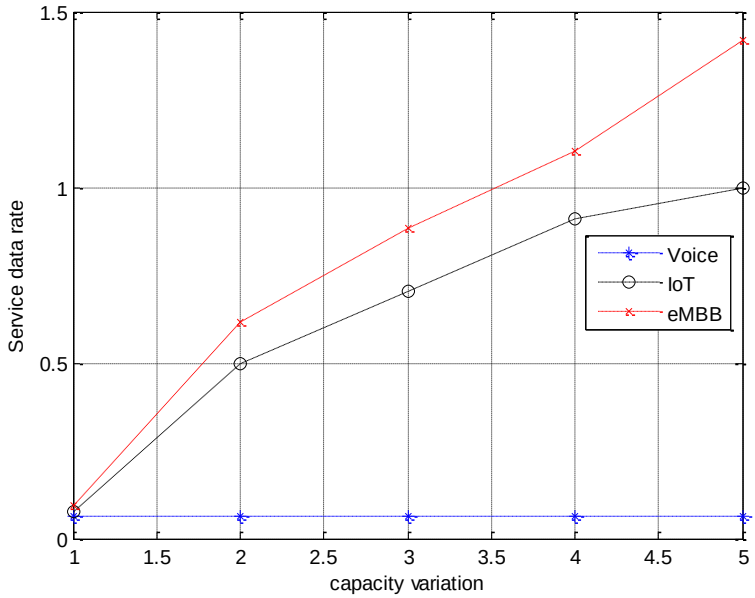


Figure III. 23. The influence of the quality of the VNO3 radio link.

We can observe from the results obtained above the following:

- The data rate allocated to each of the voice service for the all three VNOs stayed somewhat consistent throughout all the different values of R_{CRRM},
- As for the IoT services, same can be said when it comes to VNO 1 and 2 but in VNO 3 it is noticed that by increasing the $R_{CRRM,}$ the data rate increases to reach its maximum value at +50% R_{CRRM},
- Lastly, It is noticed that significant variance in the data rate of the eMBB services depending on the value of R_{CRRM} especially in the third VNO. Also, the data rate is

highly dependent on γ_v as we can see a decrease each time, we pass form a higher priority VNO to a lower one,

- The increase or decrease in the data rate assigned to each service follows a non-linear function which is justified by the use of the logarithm function present in the optimization model,
- For an increase or decrease of the overall capacity by 25% or 50%, the data rate of each service will be allocated according to the weight of the corresponding VNO, for example:

  - Increase by 25%: the eMBB service of VNO1 increases by 28%, VNO2 increases by 32% and increases approximately by 25% for VNO3,

  - Decrease of 50%: the eMBB service of VNO1 decreases by 54%, VNO2 decreases by 40% and by 89% for VNO3.

  - It is noted that all the VNOs are independently satisfied according to their SLA contracts and whatever the radio conditions,

  - The total available capacity is shared between the three VNOs.

III.14 The influence of the number users

We simulated different scenarios on our model similarly to the previous case however, this time the R_{CRRM} was fixed and we vary the number of users by adding 20% of the max capacity of network. As the maximum number of users can't go over 645 in totality, because the network is limited with the fixed value of the radio link (the R_{CRRM} needs to be increased to increase the capacity of supported users), where each VNO has the same number of users each time. We have studied the effects on the allocated date rate to each individual user in the different services offered in the three VNOs. The results of the variation of the user number on the weight of user w^{usr} are described in Table III. 8, Table III. 9, and Table III. 10 for VNO1, VNO2, and VNO3 respectively.

Table III. 8. The influence of the number of users on VNO 1.

Service	20%	40%	60%	80%	100%
Voice	1	1	1	1	1
Iot	1	1	1	1	0.5
eMBB	0.04	0.016	0.001	0.0063	0.0063

Table III. 9.The influence of the number of users on VNO 2.

Service	20%	40%	60%	80%	100%
Voice	1	1	1	1	1
Iot	1	1	1	1	0.5
eMBB	0.03	0.0126	0.0077	0.0063	0.0063

Table III. 10. The influence of the number of users on VNO 3.

Service	20%	40%	60%	80%	100%
Voice	1	1	1	1	0.832
Iot	1	0.8367	0.5150	0.2356	0.0171
eMBB	0.004	0.0016	0.001	0.0004	0.00003

Figure III.21, figure III.22, and figure III.23 represents the obtained results of the variation of the user number on the weight of service R^{srv} for VNO1, VNO2 and VNO3 respectively.

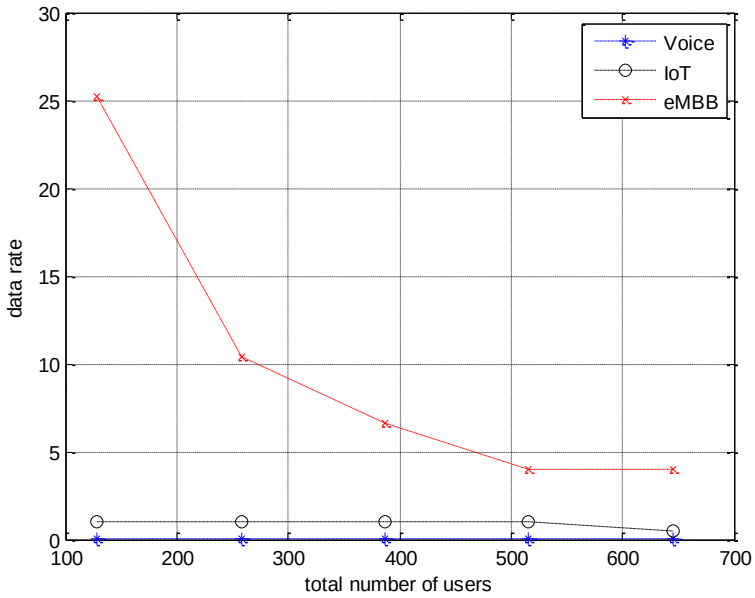


Figure III. 24. The influence of the number of users on the allocated data rate in VNO1.

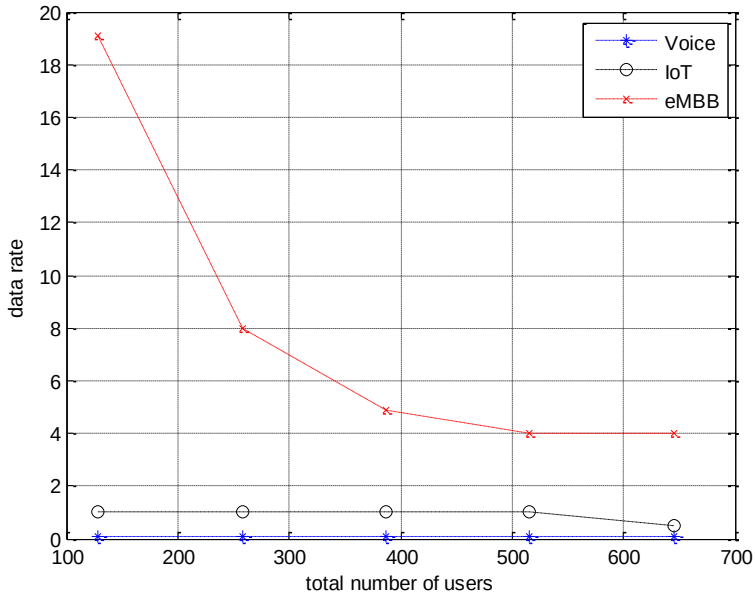


Figure III. 25. The influence of the number of users on the allocated data rate in VNO2.

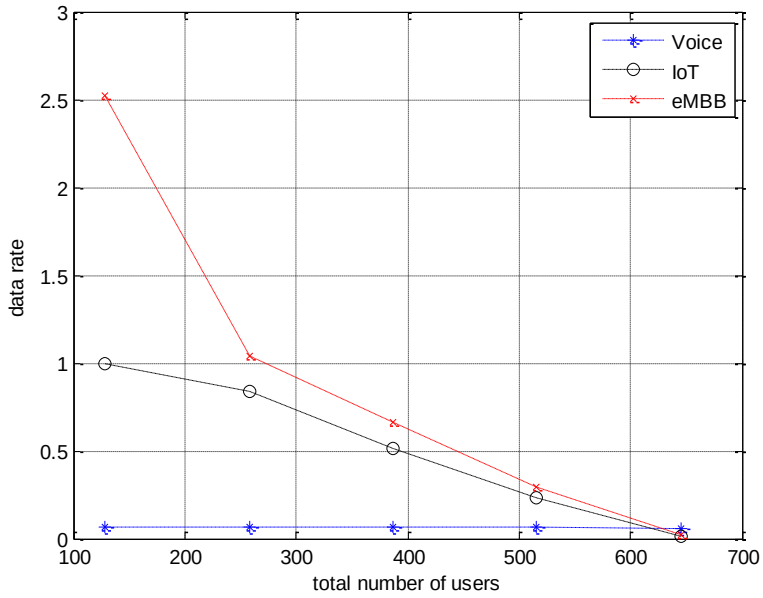


Figure III. 26. The influence of the number of users on the allocated data rate in VNO3.

It is noticed that a few changes occurred compared to the previous experiment:

- The maximum number of users can't go over R_{CRRM} value in totality as the network is limited with the fixed value of the radio link (the R_{CRRM} needs to be increased to increase the capacity of supported users)
- When it comes to the voice service, the data rate allocated to each of the three VNOs stayed consisting throughout. However, we noticed a minimal dip in the data rate in the case were the VNO3 was 100% occupied.

- For the first two VNOs, the data rate allocated to the IoT service stayed at its maximum value of 1 Mbps, except when the VNO 3 were at 100% of its user capacity it decreased to 0.5 Mbps.
- The data rate allocated to the eMBB in the three VNOs experienced a significant variance in values. The data rate decreases each the VNO has a lower weight, and the capacity of supported users get it maximum limit.

III.15 Conclusion

This chapter discussed the standardization roadmap towards 5G with its main features, an overviews the network with its structure, the roles of its different elements, the new services offered by the network and all its requirements. Two of the main features of the 5G are SDN and NFV which enabled the ability of hardware virtualization and brought many advanced key functions to 5G network such as network slicing orchestration for managing slices in the network and infrastructure sharing.

In the second part, a set of optimization has been done on the proposed network model where we experimented, with varying the radio link throughput quality and the number of the user. We fixed the throughput and vary the number of users and verse-versa, to study the influence of each one on the allocated data rate. We conclude that by increasing the value of R_{CRRM} it maximizes the data rate allocated to the users in each VNOs. To satisfy all the users of the network we need to set the R_{CRRM}. to an optimal value dictated by the use case.

It is noted that using the horizontal slicing, all the VNOs are independently satisfied according to their SLA contracts and whatever the radio conditions.

General conclusion

Network performance optimization is vital because the interconnected real-time world is fully dependent on the available, secure, and reliable transfer of data. With every year that passes, there are more and more demands that are being placed on networks. There is no denying that we live in a world that is driven by data. The efficient running of a network plays a role in every single aspect of our digital lives. This reason alone shows why network optimization is so critical. After all, there are many different reasons why there are growing strains placed on our networks today. Network optimization is critical for the end-user experience and cutting down business costs.

This dissertation was organized around three main chapters. The first overviewed the GSM networks architecture, and the changed that were made to it that brought both of the GPRS and EDGE, the procedures and methods that can be utilized to optimize mainly coverage and capacity of the cellular network in the GSM network. The second chapter discussed the architecture of both of the UMTS and LTE, also defined the WCDMA wireless standard and its implementation in the two networks. As for the third chapter, it contained the 5G network's architecture. The main focus of this study is the orchestration of the network slices in the horizontal domain.

For all the cellular network generation, the main focus when it comes to the network performance is the infrastructure planning and the optimization. This is more apparent in 5G because managing the network became more software oriented thanks to the virtualization technologies introduced in it. In order for the InPs to satisfy the VNOs' SLAs to the highest achievable level, while at the same time slicing the resources by considering the internal policy of each VNO in capacity sharing among different services, the radio link throughput for each VNO need to be optimized independently to ensure the CaaS, and depending on the number of users in each one. This was all simulated and experimented on in the second part of the third chapter using the proposed RAN model and determine the optimal case of optimization to respect the SLA contracts.

As perspective, we hope that the slicing process and orchestration would be treated in automated manner, this latter could be realized by the mean of the deep learning to make the hole system self-optimized.

Bibliography

[1] Mishra, Ajay R., ed. Advanced cellular network planning and optimisation: 2G/2.5G/3G... evolution to 4G. John Wiley & Sons, 2007.

[2] https://www.electronics-notes.com/articles/connectivity/2g-gsm/network-architecture.php. Available online on March 02, 2022.

[3] https://www.rfwireless-world.com/Tutorials/gsm-network-interfaces.html. Available online on March 04, 2022.

[4] https://www.electronics-notes.com/articles/connectivity/2g-gsm/handover-handoff.php. Available online on March 05, 2022.

[5] Garg, Vijay. Wireless communications & networking. Elsevier, 2010.

[6] https://www.electronics-notes.com/articles/connectivity/2g-gprs/network-architecture.php. Available online on March 05, 2022.

[7] M. KALACHE Mohamed amine, M. KOUADRI Oussama, ‘‘OPTIMISATION Radio 2G/3G’’ Universite Dr.Yahia Fares De MEDEA, 2016.

[8] https://www.electronics-notes.com/articles/connectivity/3g-umts/what-is-umts-wcdma-tutorial.php Available online on March 17, 2022.

[9] Kukushkin, Alexander. Introduction to mobile network engineering: GSM, 3G-WCDMA, LTE and the road to 5G. John Wiley & Sons, 2018.

[10] https://www.electronics-notes.com/articles/connectivity/3gumts/networkarchitecture.php. Available online on March 17, 2022.

[11] https://www.electronics-notes.com/articles/connectivity/3g-umts/handover-handoff-hardsoft-softer.php/ available online on March 18, 2022.

[12] https://www.3glteinfo.com/soft-softer-handover-in-umts-system/. Available online on March 21, 2022.

[13] https://www.electronics-notes.com/articles/connectivity/4g-lte-long-term evolution/what is-lte-basics-tutorial-overview.php Available online on March 13, 2022.

[14] https://yatebts.com/documentation/concepts/lte-concepts/ Available online on April 07, 2022.

[15] Ahmadi, Sassan. LTE-Advanced: a practical systems approach to understanding 3GPP LTE releases 10 and 11 radio access technologies. Academic Press, 2013.

[16] Sesia, Stefania, Issam Toufik, and Matthew Baker. LTE-the UMTS long term evolution: from theory to practice. John Wiley & Sons, 2011.

[17] Mishra, Ajay R. Fundamentals of Network Planning and Optimisation 2G/3G/4G: Evolution to 5G. John Wiley & Sons, 2018.

[18] Kibret, Nadew Sikuwaru. "Study of LTE and LTE-Advanced as a Low Cost Backhauling Solution for HSPA Small Cells." (2014).

[19] https://telecom-hyb.blogspot.com/2016/12/4g-optimization-and-kpi-analysis.html. Available online on April 13, 2022.

[20] Detti, Andrea. "Functional architecture." MA Marsan, NB Melazzi and S. Buzzi,"5g Italy White Book: From Research To Market (2018): 59-68.

[21] Rahnema, Moe, and Marcin Dryjanski. From LTE to LTE-Advanced Pro and 5G. Artech House, 2017.

[22] https://medium.com/@alifyahussain/5g-sdn-and-nfv-e411dbe927b1. Available online on April 11,2022.

[23] Barakabitze, Alcardo Alex, et al. "5G network slicing using SDN and NFV: A survey of taxonomy, architectures and future challenges." Computer Networks 167 (2020): 106984.

[24] 5G Americas. "Network Slicing for 5G Networks & Services." 5G Americas White PaperNetwork Slicing for 5G and Beyond (2016): 1-34.

[25] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6630757/. Available online on Mai 12,2022.

[26] Jia, Qingmin, et al. "Caching resource sharing for network slicing in 5g core network: A game theoretic approach." Journal of Organizational and End User Computing (JOEUC) 31.4 (2019): 1-18.

[27] Habibi, Mohammad Asif, Bin Han, and Hans D. Schotten. "Network slicing in 5G mobile communication architecture, profit modeling, and challenges." arXiv preprint arXiv:1707.00852 (2017).

[28] Li, Qian, et al. "An end-to-end network slicing framework for 5G wireless communication systems." arXiv preprint arXiv:1608.00572 (2016).

[29] Caeiro, Luísa, Behnam Rouzbehani, and Luis M. Correia. "A fair mechanism of virtual radio resource management in multi-RAT wireless het-nets." 2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC).IEEE, 2017.

[30] Rouzbehani, Behnam, Luis M. Correia, and Luísa Caeiro. "An optimised RRM approach with multi-tenant performance isolation in virtual RANs." 2018 IEEE 29th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC). IEEE, 2018.

[31] Grant, Michael, and Stephen Boyd. "CVX: Matlab software for disciplined convex programming, version 2.1." (2014).

[32] Marsh, P. "5G System Design-Architectural and Functional Considerations." (2018).