Master's thesis

2021

Farsana Karadan

Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Information Security and Communication Technology

NTNU
Norwegian University of
Science and Technology

Farsana Karadan

Implications of Service Orchestration in 5G Networks

A use-case driven approach to identify and mitigate service orchestration challenges

April 2021

NTNU
Norwegian University of
Science and Technology

Norwegian University of
Science and Technology

Implications of Service Orchestration in 5G Networks

A use-case driven approach to identify and mitigate service orchestration challenges

Farsana Karadan

MSc in Communication Technology
Submission date: April 2021
Supervisor: Kornschnok Dittawit
Co-supervisor: Mazen Malek Shiaa

Norwegian University of Science and Technology
Department of Information Security and Communication Technology

Title: Implications of Service Orchestration in 5G Networks
Student: Farsana Karadan

Problem description:

5G mobile networks are expected to revolutionize the telecom industry by offering a wider range of services both improved in terms of quality and reliability. These services can be customised to a broader range of needs of the users cutting across industrial verticals. However, the research in the field is still nascent and has a significant amount of unaddressed challenges. This thesis aims at furthering the ongoing research on 5G service orchestration by understanding the different challenges that can come up during 5G network slicing. Further, this thesis aims at exploring the possible solutions to these challenges through the following tasks:

- Review 5G network and slicing capabilities for providing heterogeneous services
- Study existing service orchestration framework in 5G network and find out the limitation
- Define use case to validate the proposed solutions for service orchestration challenges
- Selection of tools and methods for the implementation of created testbed based on the use case
- Solve the challenges by using policy intervention

Responsible professor: Kornschnok Dittawit, NTNU
Supervisor: Kornschnok Dittawit, NTNU
Industrial supervisor: Mazen Malek Shiaa, Gintel AS

Abstract

The cellular network has been evolving over the past years and the next generation of cellular network termed as Fifth Generation (5G) network, promises new improved characteristics such as latency (<10 ms), throughput (Multi-Gbps) and connectivity. This enhancement brings new business opportunities for the network operators by providing customized services to the verticals through slicing. Slicing, multi-requirement tailored services with assurance of distinct Quality of Service (QoS), are deployed in the same physical infrastructure. An appropriate level of management and orchestration of these slices is essential for the realization of slicing.

The topic of service orchestration, on the other hand, is to address the problem in which multiple services need to be delivered in parallel, e.g. between legacy and new 5G networks. Processing services independently is one thing, but when multiple services have to interact, it can create conflicts. Ensuring the desired Quality of Service for each services is important when user is accessing multiple services.

This thesis looks into the various advancements in the field of 5G service orchestration and by considering the case of telephony slice from the manufacturing industry further explores various scenarios that can arise while realizing 5G service orchestration. The thesis identifies the potential challenges in this scenarios and concludes by proposing a policy framework for service orchestration.

Further, the thesis also outlines the possible implications of the findings on 5G service orchestration, both for academia and industry.

Preface

This master thesis is a representation of my work conducted in Autumn 2020 as part of my MSc degree in Communication Technology from the Department of Information Security and Communication Technology at the Norwegian University of Science and Technology (NTNU). The work presented here is an extension of the work done in the specialization project carried out in Spring 2020.

It was my personal goal to get familiar with the practicalities of 5G, 5G slicing and services orchestration challenges related to them as I find the technologies and services related to the telecommunication field very interesting and intriguing. The research work carried out in thesis and troubleshooting measures that were taken along the way have been both a great learning experience and interesting research journey.

Even though the work has been challenging at times, it has taught me a lot about the scope of 5G and technologies that realizes it. In addition, the dissemination of my research results in a comprehensive yet easily understandable way was a learning in itself.

I hope that the reader of this thesis finds the findings and discussions presented in here as interesting and valuable as I experienced it while writing this thesis.

Acknowledgements

I would like to take this opportunity to express my deepest gratitude to my supervisor, Kornschnok Dittawit at NTNU and my co-supervisor Mazen Malek Shiaa at Gintel AS. They made this project possible for me and it was an amazing experience to learn on this topic since the specialization project in Spring 2020. I want to specially thank my other co-supervisor, Guy Redmill at Gintel AS who has helped me through the entire year with his guidance on the topic.

I am grateful to the staff at the Department of the Information Security and Communication Technology, for supporting me with administrative things during the course of my master programe and specially this thesis work.

Last, but not the least, I want to thank my friends and family especially my husband Faheem and my son Hadin for their continuous support and encouragement throughout the year.

Contents

List of Tables

List of Figures

List of Algorithms

List of Acronyms

2G Second Generation network.

3G Third generation network.

3GPP The Third Generation Partnership Project.

4G Fourth Generation network.

5G Fifth Generation network.

5GC 5G Core.

5GS 5G System.

5QI 5G QoS Indicator.

AI Artificial Intelligence.

AMBR Aggregate Maximum Bit Rate.

AMF Access and Mobility Function.

AN Access Network.

API Application Programming Interface.

ARP Allocation Retention Priority.

AS Application Server.

ATCF Access Transfer Control Function.

AUSF Authentication Server Function.

BGCF Breakout Gateway Control Function.

CHF Charging Function.

CN Core Network.

CS Circuit Switched.

CSCF Call Session Control Function.

DAPS Dual Active Protocol Stack solution.

DCN Dedicated Core Network.

DN Data Network.

ECA Event Condition Action.

EDGE Enhanced Data GSM Evolution.

EM Element Manager.

eMBB enhanced Mobile Brodband.

eNB Evolved Node B.

ENI Experiential Networked Intelligence.

EPC Evolved Packet Core.

EPS Evolved Packet System.

ETSI European Telecommunications Standards Institute.

GBR Guaranteed Bit Rate.

GFBR Guaranteed Flow Bit Rate.

gNB Next Generation NodeB.

GPRS General Packet Radio Service.

GSM Global System for Mobile communication.

GSMA GSM Association.

GSt General Slice Template.

GUI Graphical User Interface.

HD High Defnition.

HO Handover.

HSDPA High-Speed Downlink Packet Access.

HSS Home Subscriber Server.

I-CSCF Interrogating- CSCF.

IETF Internet Engineering Task Force.

iFC Initial Filter Criteria.

IMS IP Multimedia Subsytem.

IP Internet Protocol.

IPSec IP Security.

ITU-T International Telecommunication Union- Telecommunication.

KPI Key Performance Indicators.

LO Lifecycle Orchestration.

LTE Long Term Evolution.

M2M Machine to Machine.

MANO MAnagement and Orchestration.

MEC Mobile Edge Computing.

MFBR Maximum Flow Bit Rate.

MGCF Media Gateway Control Function.

MGW IMS-Media Gateway.

ML Machine Learning.

MME Mobility Management Entity.

mMTC massive Machine Type Communications.

MOS Mean Opinion Score.

MRF Multimedia Resource Function.

MRFC Multimedia Resource Function Controller.

MSC Mobile Switching Center.

NaaS Network-as-a-service.

NB Node B.

NEF Network Exposure Function.

NEST Network Slice Type.

NF Network Function.

NFV Network Function Virtualization.

NFVI NFV Infrastructure.

NFVO NFV Orchestrator.

NR New Radio.

NRF Network Repository Function.

NSA Non standalone.

NSI Network Slice/Service Instance.

NSO Network Service Orchestrator.

NSP Network Service Provider.

NSS Network Slice Subnet.

NSSF Network Slice Selection Function.

NTNU Norwegian University of Science and Technology.

OSS/BSS Operation/Business Support System.

OTT Over-The-Top.

PBM Policy Based Management.

PCF Policy Control Function.

PCRF Policy and Charging Rules Function.

P-CSCF Proxy- CSCF.

PDN Packet Data Network.

PLMN Public Land Mobile Network.

PNF Physical Network Function.

PS Packet Switched.

PSTN Public Switched Telephone Network.

QFI QoS Flow Identifier.

QoE Quality of Experience.

QOS Quality of Service.

RAN Radio Access Technology.

RCS Rich Communication Service.

RO Resource Orchestration.

RQA Reflective QoS Attribute.

RTP Real-Time Transfer Protocol.

SA Standalone.

SBA Service Based Architecture.

SCC Service Centralization and Continuity.

SCEF Service Exposure Capability Function.

SCP Service Communication Proxy.

S-CSCF Serving- CSCF.

SD Slice Differentiator.

SDN Software Defined Network.

SGSN Serving GPRS Support Node.

SGW Serving Gateway.

SIP Session Initiation Protocol.

SLA Service Level Agreement.

SLF Subscriber Location Function.

SLR Subscriber Locator Function.

SMF Session Management Function.

S-NSSI Single Network Selection Assistance Information.

SO Service Orchestration.

SRVCC Single Radio Voice Call Connectivity.

SST Slice/Service Type.

UDM Unified Data Management.

UDR Unified Data Repository.

UE User Equipment.

UPF User Plane Functions.

URLLC Ultra Reliable Low Latency Communications.

UTRAN UMTS Terrestrial Radio Access Network.

V2X Vehicle-to-Everything.

VIM Virtualized Infrastructure Manager.

VNF Virtual Network Function.

VNFM VNF Manager.

Vo5G Voice over 5G.

vSRVCC video Single Radio Voice Call Connectivity.

ZSM Zero-touch-network and Service Management.

Chapter 1
Introduction

The cellular network has been evolving over the past years and the next generation of cellular network termed as Fifth Generation (5G) network promises new improved characteristics such as ultra-low latency (1ms), throughput, and connectivity. In contrast to the previous generations of cellular networks, 5G networks are targeted to expand the capabilities of a network to support a various range of services that are tailored to meet different requirements.

A wide range of vertical industries, such as manufacturing, automotive, health, energy, media, and entertainment are going to be some of the main beneficiaries of these changes in the 5G era. Such verticals demand different use cases imposing a new set of requirements in terms of scalability, latency, reliability and availability. However, the “one-fit-all” concept of the legacy network will not be able to meet these demands [ATS$^+$18]. So, 5G has proposed a new concept, known as Network slicing, to support these heterogeneous services.

Network slicing is the splitting of physical network infrastructure into multiple logical networks known as slices. These are composed of different network functionalities required to support various use cases that are derived from the shared network infrastructure [ETS18b]. The slices can be customized to a particular vertical industry or can be a public one. Such slices are controlled and managed by slice owners such as Network Slice Providers (NSP) or Mobile Virtual Network Operators (MVNO).

5G development is still in process and it is mainly driven by the different use cases it supports. However, major drivers (service types) in the development of 5G are enhanced Mobile Broadband (eMBB), massive Machine Type Communications (mMTC), Ultra Reliable Low Latency Communications (URLLC) or a combination of these [Moh17, 3GP17].

Despite all these new enhancements 5G will potentially suffer from complex network management challenges due to the variety of vertical services they are promising and the diversity and heterogeneity of underlying infrastructure. Service orchestration in 5G slicing is a hot topic and ongoing research aims at addressing potential problems where these services or a composition of such services need to be delivered in parallel. Ensuring that desired QoS is at all times maintained while delivering these services is also an integral part of service orchestration.

Maintaining service continuity is another identified challenge and issues with service continuity can arise when these services interact between 5G and legacy networks such as 4G, 3G and 2G. While processing the services independently is one thing, conflicts that arise when these services interact with each other needs an entirely different research focus and attention. The latter forms the focus of this thesis.

As service orchestration is a recurring term in this research work, I have considered the following definition of Service Orchestration (inspired from [SdSLPR+19] for the rest of this thesis

> *Service orchestration entails managing lifecycle of services that were split into various domains based on requirements and availability in order to maintain service assurance and service continuity.*

Therefore, this thesis investigates the service orchestration challenges in providing service assurance and continuity when multiple service interactions occur in different slice realizations. For the purpose of investigation, this thesis takes a use-case driven approach i.e. the whole concept of service orchestration challenges and corresponding solutions are studied using examples of different possible scenarios that can arise during service interaction.

Further, this research work is an extension of the specialization project done in Spring 2020 and it carries forward the same use cases that were studied during the specialization project [Kar20].

Finally, this thesis proposes a policy framework for identified service orchestration challenges, additionally a few complicated scenarios that can lead to service conflicts are identified along with proposing a path for potential future work in this field.

The following sections of this chapter give an overview of the thesis background and is followed by motivation for the thesis. This chapter also covers the research questions that were framed to solve the challenges briefly presented in this section and it ends by presenting the methodology used to solve the research questions.

1.1 Background

As briefly discussed above, over the last few years, mobile networks have evolved significantly and wireless communication standards have become a vital part of our life. Particularly in cellular network, we have witnessed significant enhancements to support the growing number of users and growing level of traffic [HZK+20]. The field of cellular networks is still evolving and there is ongoing research in both academia and industry to support the thirst for high bandwidth, more speed, connectivity, service quality along with an ever-increasing number of connections and users.

1.1.1 A brief history of cellular network evolution

Cellular networks were introduced to provide public voice services and the first entrant was First Generation (1G). According to a comparative study on evolution of cellular network [SG14], 1G was based on an analog system with a speed of up to 2kbps [SG14]. Later on, Second Generation (2G) / Global System for Mobile communication (GSM) was proposed with support for text messaging services along with voice services and its successors also supported data services such as General Packet Radio Service (GPRS) and Enhanced Data GSM Evolution (EDGE). In this thesis, 2G refers to the EDGE network.

Data services have ever since led to a series of positive changes in the society and demand for improved data services has increased day by day and continues to do so till date. Upon the arrival of Third Generation (3G), speed had improved up to 200 Kbps and it partially supported packet switching. 3G also brought new services such as video calling and seamless streaming of video with a download speed up to 3 Mbps. Its successors also introduced improved data rates such as High-Speed Downlink Packet Access (HSDPA) and Evolved High-Speed Packet Access (HSPA+). In this thesis, the terminology 3G refers to HSDA+.

The need for a full packet-switched network led to the development of Fourth Generation (4G) network based on Long Term Evolution (LTE) standard, it brought enhancements such as support for High Definition (HD) audio and video quality and support for Machine to Machine (M2M) communication with a download speed of up to 200 Mbps. In this thesis, 4G is referred to as the LTE standard. Need for better connectivity and data rate acted as a catalyst for the deployment of 5G. 5G will also be a fully packet switched one with a speed of up to 1 Gbps.

1.1.2 The potential of 5G cellular network

An attractive proposition of 5G is the business potential it promises. Ericsson's report on 5G business potential [Eri19] shows that while the enhancement of network capabilities is still ongoing, mobile operators face several challenges in monetizing the advancement and face tough pricing competition from other service providers such as Over-The-Top (OTT) services and this can potentially lead to market stagnation for the industry. Despite high growth in mobile subscription, devices, and mobile data traffic, overall service revenue growth has flattened out compared to the growth in 2010 [Eri19, Eri15]. In effect, mobile operators are struggling to find more ways to monetize the enhancement in voice services and improved mobile data rate services.

In the era of industrial digitalization, mobile operators have recognized the potential of 5G capabilities to solve certain challenges in industrial digitalization. This includes automation in the manufacturing industry, autonomous driving and eHealth among others. The proposed solution to these challenges is of offering customized services with guaranteed Quality of Service (QoS). Further, Ericsson's report[Eri19] also predicts a huge growth in the revenue opportunity, in tune of 582 billion USD, for telecom operators from industrial digitalization with the help 5G technology.

1.1.3 Network Slicing – an enabler for monetizing 5G technology

Each end-to-end connected network slice is designed in a way to serve a specific service or composed services with all the necessary network resources, physical or virtual network elements, and functions with proper isolation between them [ETS18a]. Network slicing allows the network operators to efficiently use and manage network resources, create differentiated services and generate revenue out of it [Eri18]. Software-Defined Networking (SDN) and Network Function Virtualization (NFV) are the key enabling technologies of network slicing which gives flexibility and programmability [ETS18a].

Further, network slicing has a stringent goal to provide absolute resource guaranty over latency, bandwidth, and reliability among others. Hence there are ongoing research works that aim at addressing the challenges associated with Network Slicing at different levels of slicing such as end-to-end slicing architecture [ETS18a], service assurance in slicing [3GP20d], security [ON20], service continuity, management, and orchestration of slice [3GP21c]. This thesis only covers the ongoing research on the management and orchestration of network slices in a dynamic network environment

In certain cases, services provided through slicing possess strict priority treatment even though they share common QoS characteristics. For example, a slice for eHealth and a slice for autonomous cars share the same QoS restriction on latency. However, during a medical emergency such as to save a critically ill patient, some data traffic in the eHealth services must have to be prioritized to achieve ultra-low latency communication. In another scenario, consider a road accident occurs simultaneously as the medical emergency, in such a scenario data traffic in the autonomous car slice must also be prioritized at the same time to notify the ambulance service. What are the likely consequences if there is not enough bandwidth available?

Such complex service interaction challenges demand an efficient automated network that must be able to quickly adapt to the network condition and eliminate service conflicts. One of the proposed solutions for this is Policy-Based Management (PBM) of the network. PBM has been proven to be a good choice to handle network operation and management in complex distributed systems and it aids the process of automation [VCC+17]. Given the heterogenous requirements of 5G slicing, the suitability of PBM for managing QoS enabled networks makes it an appropriate choice to enhance the research work on the topic of challenges in 5G slicing [CPC+18].

Hence, this thesis also takes a similar approach of making use of policy-based management framework to find out the potential solutions for network slicing challenges experienced during composed service orchestration.

1.2 Motivation

Network Slicing enables providers to offer their services in the form of Network-as-a-service (NaaS), this will eventually enhance the operational efficiency of the network and in effect reduce the time-to-market for new services [HNW+18].

However, the current scenario of networks slicing also poses certain challenges. For example, since the offered services differ in their requirements, assuring these services based on a guaranteed Service Level Agreement (SLA) will be challenging in a dynamically changing network environment. Additionally, challenges associated with following have also been identified in the literature:

- orchestrating the service request
- processing the request and
- dynamically allocating radio resource to the request to ensure the desired Quality of Service (QoS)

The complexity of network management further increases as these services need to be tailored to meet multiple requirements with sufficient service guarantees on a common infrastructure. Since 5G describes services as an End-to-End (E2E) concept, it also needs to be supported by all the network components, from the user equipment to the cloud and service application. Therefore, the operators should be capable of orchestrating the different technologies and resources available in modern network infrastructure.

Further, to ensure the flexibility and scalability of services, the management and orchestrating system should be capable of enabling autonomous and automated service deployment and adaption. Policies and protocols will play a very important role in the creation of customized services to reduce conflicts. Another potential challenge during service orchestration is the inter-operability of delivering these services along with the other legacy networks [AD20].

Additionally, the strict resource assurance demand makes network slicing expensive, which means the tenant who is subscribed to the particular slice has to pay more money for using the services [ETS18a]. This puts pressure on the network slice providers to monitor and manage the network slice all the time in a way to make sure the availability of the services and meet the desired QoS described in SLA. The tough market competition from Over-The-Top (OTT) services and other value-added service providers further aggravates the situation.

Therefore, the commercial success of 5G networks or monetizing the full potential of 5G - all lies on the 5G network quality and efficient slice management. Hence, it is vital to address the challenges that impede the success of 5G network and this need for pragmatic solutions to solve critical challenges forms the underlying motivation for this thesis.

Through the research work presented in the following Chapters of this thesis, I would like to contribute to the development of the field of network slicing and service orchestration in 5G.

1.3 Research questions

To address the problems presented in the preceding section I have formulated the following research questions. The research questions from the project are maintained [Kar20].

1. How can the request and access to common and customized services across 5G and other networks be realized?
2. What are the current challenges in this process?
3. What potential measures can mitigate these challenges to ensure the resolution of service conflicts?
4. How can these potential solutions be applied to several services during service orchestration?

1.4 Methodology

To analyze existing service orchestration and service continuity framework, and propose possible enhancement, a combination of recognized research methods was used. Tools used and steps taken during this work are briefly summarized below.

1.4.1 Literature Review

Literature review looked into different standards on 5G, industrial publications, white papers, blogs and research papers. Since the development of 5G and network slicing is still evolving and related works are continuously updated or newly published, I have considered only the latest version of standards and most recently published research on the topic. The criteria chosen to select the appropriate source of information was:

1. Research results from recognized standard bodies in this field (e.g.: Third Generation Partnership Project (3GPP), International Telecommunication Union (ITU), GSM Association (GSMA), European Telecommunications Standards Institute (ETSI)). Additionally, inputs from academic and industrial white papers and books were also considered.
2. Date of publication – most recent publications on the topic were prioritized.
3. Other supplementary fields of research relevant to the thesis topic were also considered.

Literature review focused on enlisting relevant pieces of literature in the field of 5G slicing and service orchestration and identifying the major research trends in the field. Identification of the relevant case studies for detailed analysis of service continuity and service orchestration in 5G slicing were also carried out.

1.4.2 Case study analysis

Challenges in a practical case study on 5G

As mentioned earlier, the uses cases from the Specialization Project (Spring 2020) were used as a basis for the empirical study in this thesis [Kar20]. During the starting phase of the thesis, I had checked the possibilities of setting up a testbed of use case with the help of a major network provider in Norway. The plan was to use their 5G network Lab. Unfortunately, due to the lack of necessary Application Programming Interface (API) needed to customize my use case requirements this collaboration was found to be impossible and the idea was dropped.

Further research on the practical possibilities of creating a testbed using open-source project was carried out. I used Kamailio IMS with Open5GCore, an open-source project for this purpose. After 2 months of working on the same platform, lack of available resources posed a challenge. This too had to be dropped. Due to the pandemic, getting a 5G smartphone and specifically a 5G Subscriber Information Module (5G SIM) were additional practical challenges. Lack of 5G softphones interoperable with Kamailio IMS and Open5GCore was also an issue at the time of this thesis.

Alternative solution to case-study analysis

A possible solution was identified in the second part of this research work. In consultation with my Supervisors, I decided to collect data from recent literature on the same services and use that as input to analyze service interaction challenges. To this purpose, I created two scenarios with three possible telephony services (composed service) under the telephony use case.

The reason for using a scenario-based approach was that it helped us to narrow down the vast topic of slicing and overcome time frame limitations. Key Performance Indicators (KPI) values chosen for describing the use case were collected from different test cases held recently and matched with the value range specified in the SLA[1] of the slice. Based on these, challenges in ensuring QoS[2] while providing service continuity were identified and studied in detail. A policy-based management approach was applied to solve these challenges through policy intervention. This is further detailed in the following Chapters.

Due to the time limitation and complexity of this work, this thesis only suggests general ideas and recommendations based on the analysis and discussion to limit service orchestration challenges in the network slicing. Finally, this thesis also proposes a policy framework for service orchestration. This thesis does not include any simulation work to validate the practical limitations and implications of the proposed solutions.

1.5 Thesis Outline

This thesis is outlined in the following way.

- **Chapter 2:** The standard architecture of the 5G and 5G IMS
- **Chapter 3:** Illustration of use case and scenarios considered for this thesis
- **Chapter 4:** The identified challenges in service orchestration based on the scenarios

[1]SLA: Service Level Agreement
[2]QoS: Quality of Service

- **Chapter 5:** The proposals of necessary policy rules and policy framework
- **Chapter 6:** Discussion and conclusion of the thesis

Chapter 2
Fifth Generation Mobile Networks (5G)

An overview of the State-of the art in 5G network and major findings from the most recent literature was presented in this Chapter. Additionally, a detailed discussion on the various terminologies associated with 5G network and slicing was also presented. The chapter starts with details on 5G and its architecture and key components inside 5G Core (5GC) network, and concludes with 5G IP Multimedia Subsystem (5G IMS) architecture.

Previous generations of cellular network mainly focused on communication services between people and on improving its efficiency. While on the other hand, 5G is envisioned to support a wider array of services and connecting everything to the network with key features such as high speed, low latency and improved flexibility. The development of 5G is focused on three main service types [Moh17, 3GP17], namely,

1. **Enhanced Mobile Broad Band (eMBB):** supports a stable connection with high peak data rate
2. **Massive Machine Type Communication (mMTC):** supports a huge number of Internet of Things (IoT) devices
3. **Ultra Reliable Low Latency Communication (URLLC):** supports ultra-low latency communication with high reliability

Based on these fundamental use cases 5G has opened up an array of services in various industries, some examples are smart city, smart home, eHealth, autonomous car and manufacturing industry.

5G holds a Service-Based Architecture (SBA), which differs from older generation network architecture in many ways [V1518]. Such as;

- 5G supports a new radio access technology called 5G New Radio (NR), it demands new devices with capabilities to support NR and its frequencies.
- The SBA considers that the services are provided by a common framework to Network Functions (NF) that are tailored with different requirements to provide various services [V1518] . The NFs in 5G control plane uses service-based interfaces to interact with other NFs.
- SBA is also capable of virtualizing the deployment of theses NFs and thus improve the scalability and resource utilization as the traffic through the network grows.

Since this thesis focuses on service orchestration in 5G network, it is vital to know the 5G architecture and interaction between different entities in the architecture.

2.1 5G Network Architecture

The key enabling technologies of 5G are Software Defined Networking (SDN) and Network Function Virtualization (NFV) and these are adapted to support various services with different data service requirements provided by 5G network. In order to provide the different data services, 3GPP has proposed a flat architecture for 5G [V1518], in which they have separated the control plane and data plane functions. Thus improving the scalability and flexibility of the network through dynamic resource allocation.

5G network architecture is composed of three components [V1518], namely,

1. 5G Access Network (5G-AN)
2. 5G Core Network (5GC)
3. UE

In the 5GC architecture, network elements are defined to be Network Functions (NF) and it can be either hardware or virtualized network functions. The architecture in Figure 2.1 shows the virtualized network functions present in the 5GC. Their roles in the network as follows [V1518, Dre, Gar19, Eve18]:

1. **User Plane Functions (UPF)** UPF acts as a mobility anchor point during handover procedures and enables connectivity with the core network. It is responsible for packet routing and forwarding, packet analysis, and applying necessary QoS parameters to the packets. It also represents an external PDU

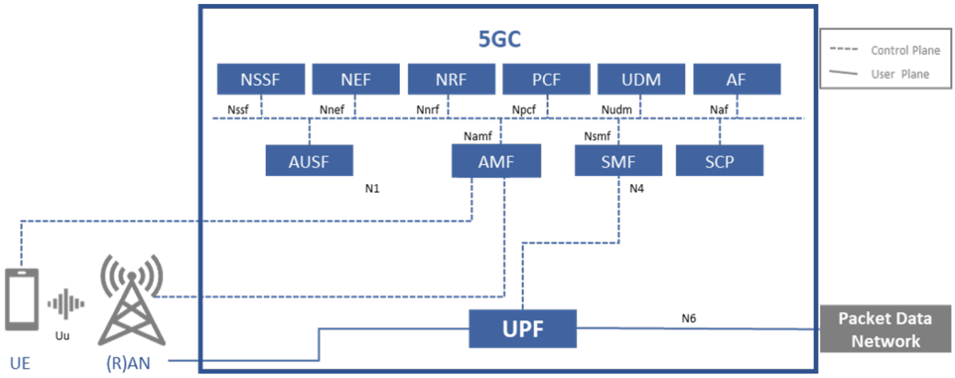

Figure 2.1: An illustration of 5GC Architecture, adapted from [V1518]

session point to connect traffic flow to the data network. It maintains and reports traffic statistics.

2. **Packet Data Network (PDN)** This represents a network for internet access or third-party services such as IMS for accessing multimedia services.

3. **Authentication Server Function (AUSF)** AUSF acts as same as the authentication function performed by 4G Home Subscriber Server (4G HSS). It implements the Extensible Authentication Protocol (EAP) authentication server and stores authentication keys. AUSF supports the authentication for 3GPP access and untrusted non-3GPP access.

4. **Access and Mobility Function (AMF)** AMF corresponds to the Mobility Management Entity (MME) in LTE and is responsible for the management of registration and authentication of subscribers. It is also responsible for the mobility management of subscriber (inter or intra) and take part in registration and connection management. Sometimes it also takes part in applying policies from PCF while a user is moving from one region to another.

5. **Session Management Function (SMF)** SMF is responsible for handling session management functions such as session establishment, modification, and release. SMF functionalities also include IP address (IPv4 / IPv6) allocation to UEs and enable packet routing. SMF determines the policy and charging implementation for every service.

6. **Service Communication Proxy (SCP)** As the name suggests it act as a proxy in 5GC to help in signaling when requesting a new service. It facilitates the communication between two NFs[1] over the service interface. Some functions

[1]NF: Network Function

include signaling aggregation and routing, load balancing and load distribution, overload handling, packet prioritization in network congestion situations, and signaling peak protection. Since SCP functions are complemented by NRF, SCP is a good choice to be used to provide network resiliency in case failure occurs.

7. **Network Slice Selection Function (NSSF)** NSSF will help in the selection of network slice which is available to the requesting UE. NSSF is used to uniquely identify a network slice which is also identified by Network Slice Selection Assistance Information (NSSAI). NSSAI is explained later in this thesis. NSSF is also responsible to determine which AMF should serve the UE.

8. **Network Exposure Function (NEF)** NEF provides border security to the 5G network, which means a secure exposure of capabilities and events of network services towards external applications over APIs. NEF also act as a translator when communicating with external application.

9. **Network Repository Function (NRF)** NRF is a repository that holds the information about available Network Functions in the network and services provided by them. It sends the notification to the user who is subscribed to the NFs when the status of NF changes (modified/updated/new service added etc.).

10. **Policy Charging Function (PCF)** PCF acts as same as the PCF component as part of PCRF in LTE. PCF implements a unified policy framework that handles the policy-related decision according to the network condition. PCF provides policy rules to the control plane functions to enforce them. PCF is also connected to the external data network and imposes policy control for them as well. It can access Unified Data Repository (UDR) to know about the user's service subscription details and apply policies accordingly.

11. **Unified Data Management (UDM)** UDM acts as a central repository of subscriber information and takes part in the authorization of users while accessing the services since it holds the authentication keys. It is responsible for subscription and SMS management. UDM involves in UE serving NF registration process and mobility management and thus support session/service continuity.

12. **Application Function (AF)** AF interacts with 5GC to provide services. It supports accessing NEF and efficient traffic routing. It interacts with the policy framework (PCF) for policy control and support external network (e.g. IMS) interaction with 5GC.

2.2 Deployment options of 5G network

Unlike other generations of cellular network, 5G allows integration with different access technologies and core network. This means that while being deployed the access network and core network of the 5G system can comprise of the same generation or it can be from other generations such as 4G (EPC and EUTRAN). Different configurations are possible with 5G and it is categorized into two, namely [GSM18b]:

1. Standalone (SA) network
2. Non-Standalone (NSA) network

SA uses only one access technology such as NR with 5GC as the core network while NSA combines multiple access technologies, such as 4G LTE or NR with EPC as core network. Three options of SA configuration are defined in 3GPP, namely[GSM18b][2]:

– Option 1: Composed of LTE evolved Node B (LTE eNB) as access network and EPC as the core network.

– Option 2: Composed of NR gNodeB (gNB) as access and 5GC as the core network.

– Option 5: Composed of LTE ng-eNB as access and 5GC as the core network.

NSA configuration also contains three other options, namely:

– Option 3: Represents the dual connectivity capability of the 5G network. In this, both LTE eNB and gNB can act as an access network, in which LTE eNB act as a master node while NR en-gNB act as a secondary node.

– Option 4: Uses NR gNB and LTE ng-eNB as access, in which former one act as a master node and later act as a secondary node.

– Option 7: Uses LTE ng-eNB as a master node and NR gNB as a secondary node and is connected to 5GC.

All these different configuration options allow the operators to adopt different strategies for 5G deployment according to their business model and competition needs in the market [GSM18b].

[2]The numbering for the options are in accordance with the source article[GSM18b]

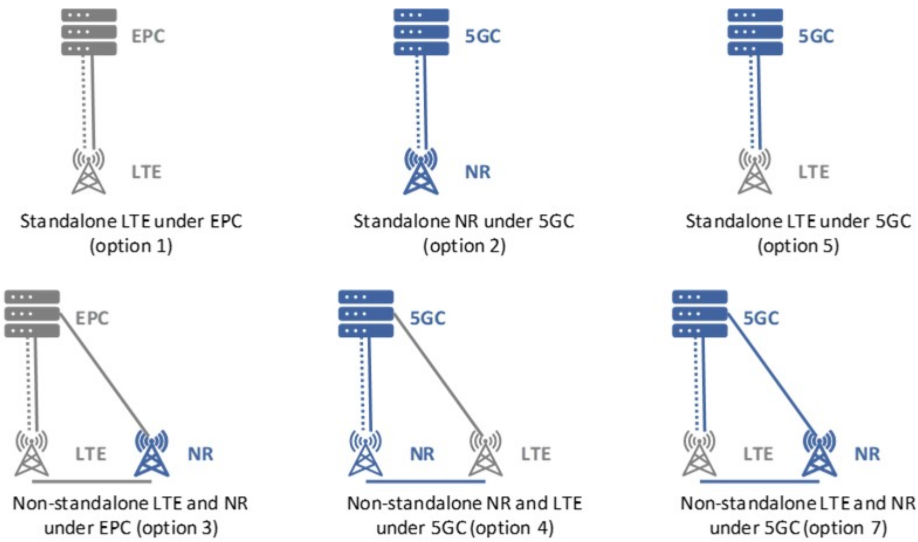

Figure 2.2: An illustration of 5G deployment options, adapted from [GSM18b]

SA option 2 with NR as access network and 5G as core brings a number of advantages when comparing it with other options with LTE access and core network. SA option 2 is capable of providing all the advancement of 5G such as flexibility, scalability and programmability features mentioned previously in this thesis.

However, before the 5G network is widely deployed and available everywhere it demands some level of inter-working with EPC and eNB to maintain the service continuity. During the early phase roll out of 5G it is also important that the inter-working with LTE to support dual connectivity feature of 5G. This to ensure that the UE can separately receive both signals and aggregate them to get a better data rate [Sea19].

At the same time inter-working of gNB and eNB is vital for supporting Inter Radio Access Technology (Inter-RAT) mobility which allows the UE to switch between 5G and LTE in case of signal degradation. Within the configuration options, options 2 and 7 can support network slicing. Interworking between 5G and legacy was part of 3GPP release 16 and details can be found in TS 23.216 [3GP20b]. SA option 3 offers a dual connectivity feature, it must be noted that dual connectivity demands that UE must be compatible with both LTE and NR radio access networks.

Since this thesis focuses only on telephony services, it is important to consider the possibilities of deployment options that are suitable for providing better service continuity. Among the proposed deployment options, all configurations support IMS

Table 2.1: Deployment options for IMS Service, adapted from [GSM18b]

Solution 1: IMS media and signaling via 5GC	**Solution 2: IMS media and signaling via EPC**	**Solution 3: CS voice over MSC**
Option 2: NR via 5GC	Option 3: LTE and/or NR via EPC	Option 3: CSFB to 2G/3G CS from EPC
Option 4: NR and/or LTE via 5GC	If IMS is not available, Circuit Switched Fall Back (CSFB) can be preferred.	Option 3: CSFB to 2G/3G Cs from 5GC.
Option 5: LTE via 5GC	none	none
Option 7: LTE and/or NR via 5GC	none	none

call and video services [GSM18b]. However, there are differences in choosing which Radio Access Technology (RAN) will be suitable for media, and Session Initiation Protocol (SIP) has some significance in terms of QoS and Quality of Experience (QoE). Previous research on this topic has identified different deployment option solutions for IMS voice and video and the details are summarized in the table2.1 [GSM18b].

The applicability of these deployment options for telephony services will be discussed in detail in the coming chapters.

2.3 5G Network slice (NS)

Network slicing is an end-to-end paradigm proposed by 5G to support new kind of applications that need resource guarantees in terms of latencies, bandwidth, jitter, reliability, throughput and privacy [ETS18a]. A Network Slice (NS) can be solely composed of Physical Network Function (PNF) or Virtual Network Function (VNF) or a combination of both. 3GPP has defined NS as [ETS18a]:

"a description of a service aware logical network that is composed of different physical or virtual network elements, resources and functions"

Network slice is aimed at providing tailored services with multiple requirements to the vertical industries to meet their need for more advanced technologies and thereby to support and accelerate their journey towards digitalization. NS can be considered as an independent managed instance of a logical network which shares the underlying infrastructure with other independent managed instances [ETS18a]. A three-layer model of slicing is shown in Figure 2.3.

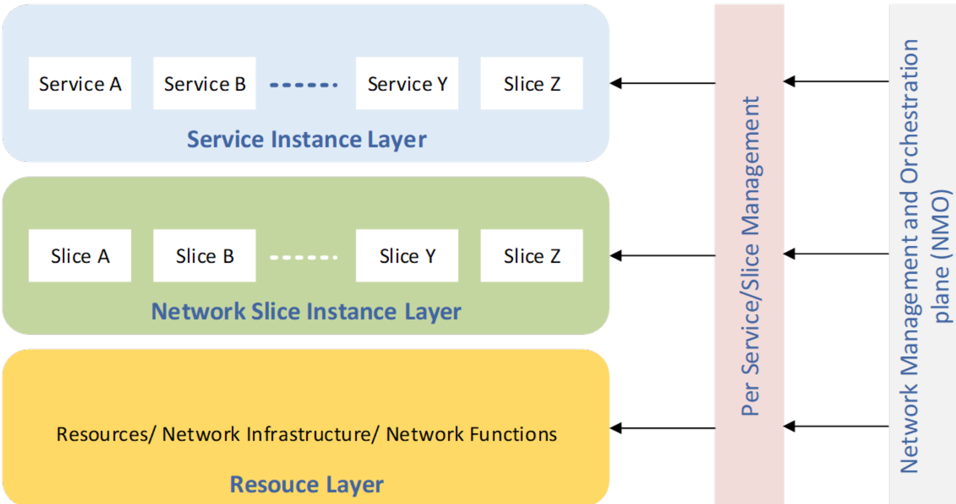

Figure 2.3: Three layer model of slicing, adapted from [HHS17, AP16]

The three-layer approach of network slicing framework is composed of [HHS17, AP16]:

1. **Service Instance Layer:** This layer represents the services provided by the slice to end user and verticals. Each service in this layer is represented by a Service Instance (SI). These services can either be provided by the network operator or by a third party such as Network Service Providers.

2. **Network Slice Instance Layer:** This layer consists of corresponding Slice instance for the services present in the service instance layer. A network operator can use a General Slice Template (GST) or network slice blueprint to map the Service Instance to the Network Slice Instance. The network slice instance provides the network characteristics required by a service instance. A network slice instance can be solely reserved for providing a service instance or it can serve multiple service instance.

3. **Resource Layer:** This layer is composed of all the physical and virtual network functions that are used to implement a slice instance.

Network Management and Orchestration (NMO) layer is responsible for providing management and orchestration functions of the three above mentioned layers. NMO functions can be taken as an entity which is needed to provide the orchestration and management of each slice and for different slices implemented for a particular vertical.

Some terminologies used in this thesis with regards to network slice is described below [3GP21b]:

- **Network Service Instance (NSI)** A logical network composed of a chain of network functions to provide a specific service. It can be composed of either VNF or PNF.
- **Network Slice Instance (NSI)** A service instance or a set of service instances of network functions and the required cloud resources needed to implement a network slice.
- **Single Network Selection Assistance Information (S-NSSI)** This component is used for the identification of a network slice in the network. NSSI is defined as the collection of S-NSSIs. Currently 3GPP only allows a maximum of eight slice S-NSSI that can be grouped under a NSSI, that means a UE can only access services from eight slices [SX17].

  S-NSSI will help UE to access the services selecting appropriate slice instance to provide requested service. NSSI is often associated with PLMN (PLMN ID) and have network specific values. S-NSSI is composed of:

  - A slice/ service type (SST), this will provide the information about which service type is supported by specific S-NSSI, for example to identify eMBB, uRLLC, uMTC supported slices.
  - A Slice Differentiator (SD) is used to provide additional information about the slice features other than the service type which allows the network to uniquely identify the different services in a slice.

2.4 Network Service Orchestration

According to 3GPP, network service orchestration is defined as a subset of NFV Orchestrator functions that are responsible for Network Service lifecycle management [ETS14].

Network Service Orchestrator (NSO) is a term used to convey the concept of different service orchestration techniques that rely on multiple technologies and paradigms to achieve a goal of user satisfied service [SdSLPR+19]. NSO comprises the semantics of requested service, and control and manage deployment to fulfill the service requirements and to realize end-to-end service lifecycle management. For getting a clear picture about NSO functionality, NSO has been divided into three subcategories, namely [SdSLPR+19],

1. **Service Orchestration (SO)** This orchestrator is responsible for service composition and service decomposition upon a service request. This can be considered as an application layer entity, in which it interacts directly with Operations Support System (OSS)/ Business Support System (BSS) and with the marketplace.

2. **Resource Orchestration (RO)** This orchestrator is responsible for efficient resource allocation either physical or virtual among services in a way to match service requirements. Efficient RO can be achieved with the help of Network Function Virtualization Orchestrator (NFVO) and different SDN controllers.

3. **Lifecycle Orchestration (LO)** LO is responsible for overall service orchestration from service request to service termination and deals with the management of workflows, processes, and dependencies across service elements. LO is also responsible for ensuring that the services are delivered according to the contracted SLA.

For the scope of this thesis, we will be mainly focusing on Lifecycle Orchestrator and others will be considered only for other supportive orchestration functions. Different bodies of standardization have proposed frameworks to overcome the challenges identified in service orchestration and management. A short review of such proposals will be discussed in the coming sections.

2.4.1 NFV MANO framework

As mentioned earlier, SDN[3] and NFV[4] is going to be the key enabling technology of network slicing. By virtualizing the NF[5] capabilities using NFV, NF will achieve flexibility in instantiating NFs anywhere in the network or data centers in very little time. This will also enable elasticity in dynamically allocating resources over the network when it is needed.

These paradigm shifts in the network bring a need for novel management and orchestration framework to create, control and manage the resources according to the requested services. Therefore, ETSI has introduced an NFV MANO (Management

[3]SDN: Software Defined Network
[4]NFV: Network Function Virtualization
[5]NF: Network Function

and Orchestration) architecture framework to realize management of Virtual Network Function (VNF) lifecycle and resource allocation [ETS14].The illustration of NFV MANO architecture is shown in figure2.4.

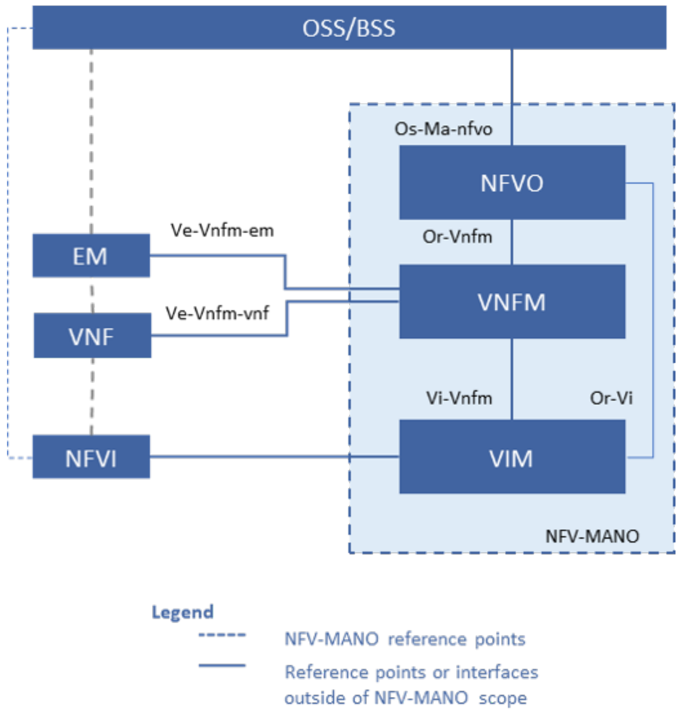

Figure 2.4: An illustration of NFV MANO architecture [ETS14]

NFV MANO consists of three main functional blocks:

1. **NFV Orchestrator (NFVO)** The main functionalities of NFVO can be categorized into two, one is network service life cycle management orchestration and the other one is NFVI resource orchestration across multiple VIMs according to the network load and service requirement. In-network orchestration, the main responsibilities of NFVO include [ETS21]:

   – Handling life cycle management of network services.

   – Monitoring and controling NS Performance Measurement (PM) and NS Fault Management (FM).

   – Management of software images.

   – On-boarding and management of Network Service Descriptor (NSD)

  - On-boarding and management of Physical Network Function descriptor (PNFD)

  Resource orchestration responsibilities of NFVO include [ETS21]:

  - Ensuring the availability and allocation of VIM resources.
  - Permitted allowance management.
  - Resource performance and fault management.
  - Resource reservation management.

2. **VNF Manager (VNFM)** The decoupling of NFs[6] from the physical infrastructure to virtual infrastructure demanded a management entity to manage the creation and lifecycle management of VNFs[7], and this is handled by VNFM. It acts as a management entity and it can access and request information on VNFs behavior from VNFI and EM[8] via appropriate interfaces. The Functionalities of VNFM comprised of:

   - VNF lifecycle management (LCM).
   - VNF configuration parameters.
   - VNF information management

3. **Virtualized Infrastructure Manager (VIM)** This functional block has the responsibility of controlling and managing the NFVI virtual resources (compute and storage) that are used by VNFs and Virtual Link (VL). VIM is capable of managing multiple types of NFVI resources such as compute-only, storage-only and network-only.

Some additional functional blocks are also present in NFV MANO architecture to support the interaction between main functional blocks. They are listed below:

- **Element Management (EM)** EM is responsible for managing the management entities performing the Fault, Configuration, Accounting, Performance, and Security (FCAPS) for the application functions of VNF. The EM interacts with NFV-MANO by communicating with the VNFM, via the Ve-Vnfm-em reference point [ETS21].
- **Virtualized Network Function (VNF)** The VNF is an entity managed by the VNFM and has an associated VNFD which provides deployment and operational information to manage its lifecycle [ETS21].

[6]NF:Network Function
[7]VNF: Virtual Network Function
[8]EM: Element Manager

- **OSS/BSS** This functional block represents the operator's operation and business support functions which are used for system and management application of other services other than NFVO and VNFM. NFV MANO provides an interactive functional block for service providers through OSS/BSS[9] to manage and operate their other businesses. The OSS/BSS interacts with NFV-MANO by communicating with the NFVO, via the Os-Ma-nfvo reference point [ETS21].
- **NFV Infrastructure (NFVI)** The NFVI encompasses all the underlying components of the infrastructure, the hardware, and the software, which are used to host VNFs. The VIM is the NFV-MANO entity that manages the NFVI, resources used by the VNFs and NSs, via the Nf-Vi reference point.

Even though NFV MANO covers most of the challenges in the service orchestration such as availability challenges, reliability, and security, some areas like service interaction with legacy networks and setting priority to the services are not covered [ETS21]. Most importantly NFV MANO inter-operability with the legacy network such as 3G or 2G is not covered in the architecture framework. The goal of this thesis is to use a possible scenario to validate the existing solutions that are proposed for mitigating service orchestration challenges. The non-mitigated challenges and problems are further analyzed and a proposal for mitigating these through policies is proposed in this thesis. This is further detailed in Chapter 5.

2.5 IP Multimedia Subsystem (IMS)

The use cases considered in this thesis are based on providing telephony services using a 5G slice. Therefore, it is vital to spend some time discussing 5G IMS. This section will present an overview of 5G IMS, its architecture, and IMS components.

IMS, standardized by the 3GPP is an architectural framework that allows access to multimedia services by using any device or IP network connection [3GP15]. The primary goal of IMS development was to provide a generic architecture for offering multimedia services, which means these services can be accessible from the wired or wireless terminal. These multimedia services include voice calls, video calls, conference calls, messaging, data, and web services to users.

The access independent nature of IMS can be useful in connecting IMS with 5G core for delivering voice over 5G (Vo5G /VoNR) and other multimedia services. One other reason for selecting IMS for 5G is its capability to provide guaranteed certain QoS for the IMS offered services. Session Initiation Protocol (SIP) is used as the core signaling protocol in IMS and it is an application layer protocol [IET02]. SIP

[9]OSS/BSS: :Operations Support System / Business Support System

can establish, modify, manage and terminate multimedia sessions and also provide multicast conferences.

2.5.1 IMS layered architecture

For a simpler representation of IMS architecture, I have adopted layered IMS architecture and it is shown in figure 2.5. The three-layer model of IMS architecture is consists of the Transport layer, Control layer, and Application layer.

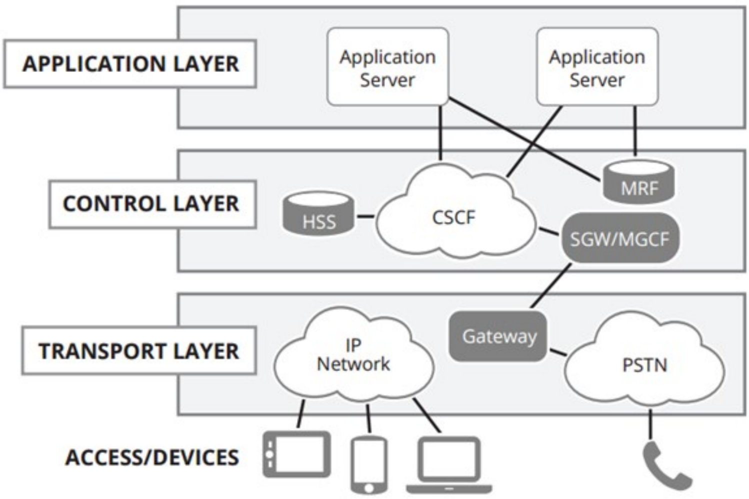

Figure 2.5: IMS architecture [Com21]

1. **Transport layer**

   This layer contains several access networks such as GPRS, LTE and PSTN, and is responsible to support the core network architecture of the cellular network. It is the connecting point for users to IMS infrastructure, connected either via circuit-switched or packet-switched. This layer includes routers, switches, firewalls along with gateways for translating protocols between new and legacy networks [LL07].

2. **Control layer** This layer is intended for session control, modification, termination, and management. It also facilitates the handover process between networks. It contains the central database of IMS (HSS), which stores the user's authentication and service authorization and Call Session Control Function (CSCF). There are three CSCF in IMS architecture, they are Proxy Call Session

Control Function (P-CSCF), Serving Call Session Control Function (S-CSCF), and Interrogating Call Session Control Function (I-CSCF).

a) **Call Session Control Function (CSCF)**

This central component of IMS is responsible for managing all signaling from end-users to services and other networks[Han09]. This component is further distributed according to its functionality into three components namely, PCSCF, SCSCF, and ICSCF. Which are defined in detail below:

i. **Proxy Call Session Control Function (P-CSCF)** This is the first contact point of User Equipment (UE) within the IMS network to request a service. It acts as a signaling proxy server between the UE and the IMS core network. The signaling proxy server will receive SIP requests from UE and either serve the request on its own or route the request to other servers.

P-CSCF can be found in both the home network and visited IMS network. During the registration process, a proxy server is assigned to an IMS terminal and it forwards the SIP REGISTER request from UE to the home network [3GP15]. P-CSCF also forwards SIP request/response messages for the multimedia session set up from UE to a SIP server and vice versa.

Since it is the first contact point for UE, P-CSCF also takes part to provide a secure connection between UE and IMS network. It also applies some compression processes on SIP messages to reduce the latency over the air interface [Han09]. P-CSCF includes a Policy Control Function (PCF), which is responsible for IP flow control and authorization of traffic bearer resources. PCSCF is also capable of taking necessary actions during emergency call sessions.

ii. **Interrogating Call Session Control Function (I-CSCF)** This entity is the contact point in the operator network which helps to determines which serving C-SCF should be assigned for handling the session upon the SIP request [3GP15]. The request received by I-CSCF can either from the home network or visited network through corresponding proxies. Upon the request, I-CSCF query the HSS about the address of the S-CSCF and provide it back to the P-CSCF for further handling of the multimedia request. An operator network may contain multiple I-CSCF for providing better scalability and minimize latency.

iii. **Serving Call Session Control Function (S-CSCF)** S-CSCF act as a central node for SIP signaling and is responsible for providing session control services for a user. S-CSCF is capable of taking different roles corresponds to the tasks it is doing such as it can act as a SIP registrar server, it can perform Session Control functions (session origination and termination), it can act as an interface to application servers, and interconnecting with other legacy service platforms [ZD05]. S-CSCF can also be responsible for providing notification messages to the users subscribed to a particular service and generates Call Detail Records (CDR) for accounting and billing purposes. There is another CSCF that is responsible for handling emergency call session control known as Emergency CSCF (E-CSCF) which routes the emergency requests to appropriate emergency centers.

b) **Home Subscriber Server (HSS)** HSS is a central database of IMS and located in the operator's network, which handles and maintains different service profiles of users who are subscribed to the IMS services. A service profile consists of all the information related to the service, service preferences, location information stored during the registration process of the IMS terminal, security-related information (for authentication and authorization), instant message service information, voice mailbox, etc. [ZD05].

   It is common to have a single HSS in the network for storing user information, but in some cases, if the number of users is high it may contain multiple HSS. In that case, the network contains another entity named Subscriber Locator Function (SLR), which contains the record of all IMS subscribers and corresponding HSS. SLR helps to locate the subscriber information in the correct HSS when getting a request from I-CSCF.

c) **Media Resource Function (MRF)** MRF entity is responsible for processing various tasks on media streams (over Real-Time Transfer Protocol (RTP)) associated with services. Media processing includes functions such as playing announcements, voice mail recording and playback, audio and video conferencing, speech recognition and video processing, etc. [ZD05]. The MRF is comprised of Media Resource Function Controller (MRFC) and Media Resource Function Processors (MRFP) [KG08].

d) **Breakout Gateway Control Function (BGCF)** This network element is responsible for determining the routing of SIP messages to the circuit-switched network such as PSTN or PLMN in a situation like S-CSCF is not able to route this message. BGCF selects a Media Gateway Control Function (MGCF), acts as an interface between the IMS network and PSTN and that will further route the call to the PSTN media gateway.

e) **IMS-Media Gateway (MGW)** The key functionalities of MGW include conversion of IP-based media streams from the IMS network to TDM based media streams on the PSTN side. MGW is controlled by MGCF [ZD05].

f) **Media Gateway Controller Function (MGCF)** The key functionalities of MGCF includes determining next hop for SIP routing, protocol conversion and CDR generation [3GP15]. MGCF also communicates with CSCF, BGCF, and legacy network entities.

3. **Service layer** This layer is composed of different application servers and content servers to provide value-added services to the subscribers. Application servers can operate in various modes such as Sip Proxy Mode, SIP User Agent Mode, or SIP Back-to-Back User Agent mode [LL07]. There are three types of application servers namely, SIP Application Server (SIP AS), Open Service Access – Gateway (OSA-GW), and IP Multimedia Service Switching Function (IM-SSF) [LL07].

2.5.2 IMS architecture in 5G network

Since IMS is an access independent technology it is compatible with 5G as well. Most of the IMS architecture remain as same as LTE, even though it needs some changes in the interfaces and components to interwork with 5GS. One case of exception from 4G IMS architecture is with the location of HSS, in 5G IMS architecture, the HSS can be co-located or implemented with UDM. However, it does not change the functionality of the HSS [Kar17].

Another change will be the interface that connects 5GS and IMS, now it will be connected via UPF and P-CSCF. As 5G SA deployment is still underway, multimedia services over 5G will be tightly coupled with existing LTE deployment. A 5G IMS architecture is shown in the Figure 2.6.

2.6 Quality of Service (QoS)

3GPP has defined QoS as:

"The collective effect of service performances which determine the degree of satisfaction of a user of a service" [3GP20a]

There are certain factors that can likely influence the performance of QoS, namely, service operability performance, service accessibility performance, service retainability performance, service integrity performance, and other factors specific to each service [3GP20a]. This section briefly describes about how QoS is assured in 5G network.

The 5G QoS model is based on QoS Flows while 4G is based on EPC bearer [V1518]. Each QoS packet flow is classified and allocated with the appropriate QoS Flow Identifier (QFI). The 5G QoS model supports three different QoS Flow Types, namely, Guaranteed Bit Rate QoS (GBR QoS) flow which requires guaranteed flow

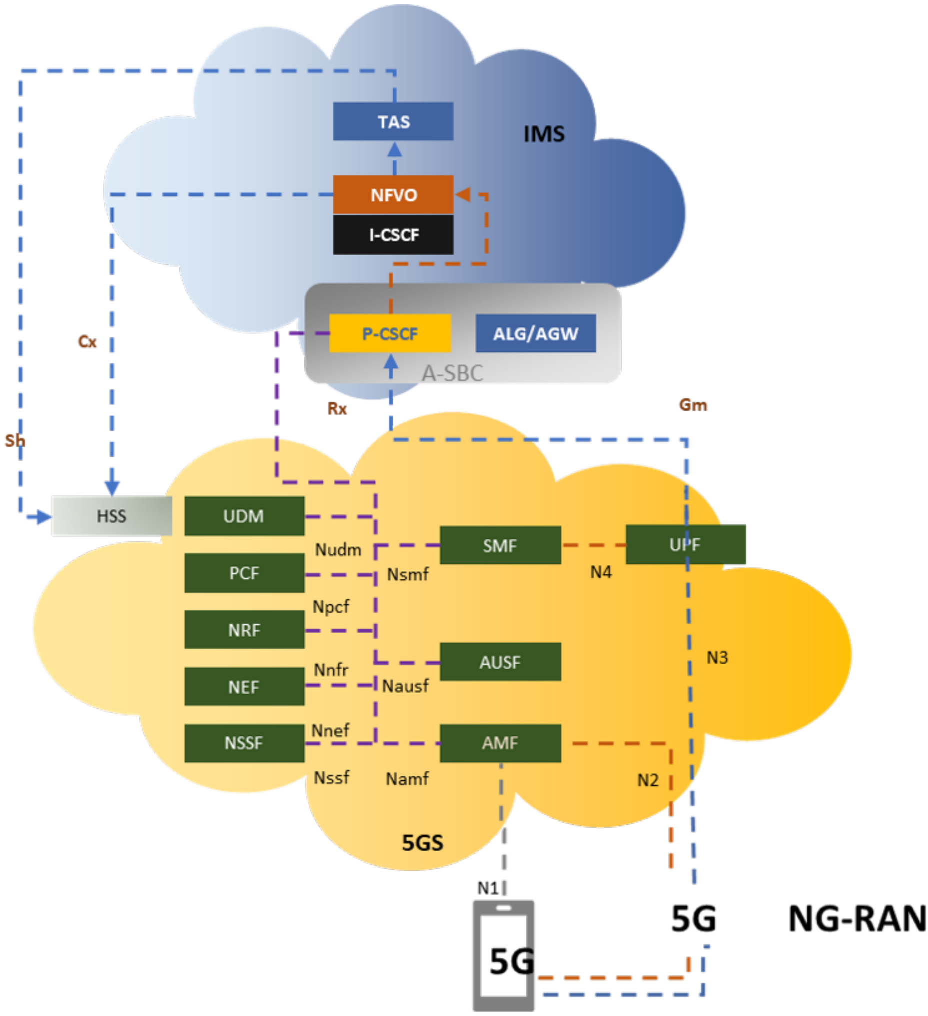

Figure 2.6: 5G IMS architecture [Kar17]

bit rate, Non-GBR QoS flow that does not require guaranteed flow bit rate, and delay critical QoS flow.

During PDU session establishment each flow is differentiated from one other using a QoS Flow ID (QFI). User plane traffic that comes under the same QFI will get the same traffic forwarding treatment. Each PDU session shall contain a unique QFI, and it can either dynamically be assigned or by using a standardized 5G QoS Indicator (5QI).

In 5GS, QoS flow control and management are the responsibilities of SMF. During a PDU session or a PDU session modification, SMF assigns the QFI for the flow and derives its QoS profile, UPF instructions, QoS rules from PCC rules, and other policy and charging related rules from PCF corresponding to the flow. A single QoS flow can have multiple QoS rules with the same QFI.

Like in LTE, 5GS also provides a default QoS rule to the QoS flow which comes under the category of Non-GBR during the PDU session establishment and this will be maintained until the termination of the PDU session. Reader is referred to 3GPP TS 123.501 [V1518] for a detailed illustration of 5G QoS.

2.6.1 5G QoS Parameters

QoS in 5G is characterized by the following parameters [V1518]:

- **5G QoS Indicator (5QI):** This is a standardized scalar value that is configured by the network operator. It is used as a reference to 5G QoS characteristics that means, to provide necessary QoS parameters when forwarding QoS flow on the network.
- **Allocation Retention Priority (ARP):** This entity is used when allocating priority to the QoS flow during congestion. The priority level defines the relative importance of a resource request among other requests.

  ARP allows the network to decide whether a new QoS flow should be added or removed or modified in the ongoing flow according to the operator policy rules. ARP is also represented in the form of scalar value ranging from 1 to 15, in which 1 has high priority.
- **Reflective QoS Attribute (RQA):** This is an optional parameter to indicate which QoS flow is subjected to Reflective QoS.
- **Notification control:** This parameter indicates whether SMF will get notified when GFBR can no longer provide the necessary QoS for a particular QoS flow or when GBR QoS flow is not able to adapt the QoS parameter according to the changing network conditions [V1518].
- **Flow Bit Rate/ Aggregate Maximum Bit Rate (AMBR):** This parameter is used to specify the maximum bit rate supported by QoS flow. For example, for GBR QoS flow includes two additional parameters they are: Guaranteed Flow Bit Rate (GFBR) and Maximum Flow Bit Rate (MFBR).

  GFBR is used to give a guaranteed QoS characteristic to the QoS flow by operators while MFBR is used to indicate the highest bit rate that can be allocated for a QoS flow.

– **Maximum Packet Loss Rate:** This parameter indicates the maximum rate of packet loss that can be tolerated in both uplink and downlink direction. This parameter is only applicable to the GFBR.

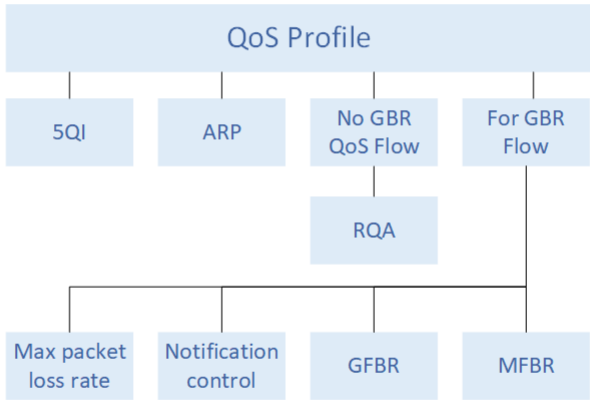

Figure 2.7: Content of QoS profile [V1518]

The content of a QoS profile for each service flow has shown in figure 2.7. 3GPP has achieved the goal to cover most of their QoS-related issues in the 5G network for example by introducing reflective QoS, limiting packet loss rate and delay among others. However, there are very few practical studies out there on how well this solution is perfectly solving the QoS issues and the topic continues to be an open and unanswered question. To support E2E QoS support in the 5G network, a harmonized QoS and policy framework is needed for achieving a satisfactory user experience.

Additionally, understanding the gap between how well the QoS will be guaranteed in the legacy network when accessing services that have E2E QoS support is a question. One way to mitigate this challenge is by adopting an efficient priority allocation to the QoS flow. Supportive technologies such as machine learning and big data analysis for enhancing QoS capabilities in the 5G network are also part of research.

2.7 Related Work

As mentioned in Section 1.4, this thesis uses data from use-cases that have been recently studied in the literature. Hence a brief description of the different identified related works is presented below.

In December 2017 ETSI launched a Zero touch network Service Management (ZSM) group which aims to accelerate the automation of 5G network [V1.20]. ZSM group proposed a reference architecture for satisfying the architectural, functional and operational requirements for the realization of end to end network and service automation.

A few research works were also proposed based on policy approaches to achieve service orchestration in 5G network. Some of this include a proposed dynamic spectrum slicing algorithm using a non-uniform source sharing agreement in the form of policy between different slices [GCU+18]. This work aims to provide the dynamic resource allocation among slices with accordance to the changing network behavior.

Shrivastava et al. [SSB18] propose a similar approach to solve the same dynamic resource allocation problem. The paper proposed a dynamic policy algorithm which allows the dynamic resource allocation by mapping the SLA of the service in to the QoS of the service occurs during slice realization.

Another standard architecture proposed for Network service management and orchestration is NFV MANO which has been detailed in 2.4.1. Inclusion of Artificial Intelligence (AI) for network automation has been studied in Experimental Network Intelligence (ENI) by ETSI [EG18]. This standard document defines how AI can be efficiently applied in the telecommunication network to support complex network management challenges such as adapting the network condition by taking necessary actions in accordance with the changing context, such as environment, resource demand and varying service requirements. However, an architecture to support ENI system is still is in progress.

SLA based approach for the network service orchestration was also part of the reviewed research [TKS+19]. Touloupou et al. proposed an integrated SLA management framework for the provisioning of network slice life cycle and this allows the operators to provide better QoE to the users of the slice.

These aforementioned works and other relevant research articles are mentioned at in their respective sections of this thesis as well.

Chapter 3
Industrial Use case

Taking a use case driven approach warrants this thesis to validate the proposed solutions for service orchestration challenges in network slicing by making a testbed for possible scenario. And then to study how well the current approaches solve the problem.

Many use cases have already been formulated to support 5G's goal to provide heterogeneous services to vertical industries. Hence, given the diversity in the service requirements for different use cases, it will be a challenge to validate different findings using a general use case. Covering all such possible needs and requirements of different services will require significant amount of time and effort, and is likely beyond the scope of this thesis.

As part of the specialization project in Spring 2020, I had considered manufacturing and health care and accompanying services as two use cases that will be utilizing the 5G network and its capabilities [Kar20]. Given the aforementioned limitations, for this thesis I have chosen the manufacturing industry for in depth study in this thesis.

3.1 Manufacturing industry

3.1.1 Choice of Manufacturing industry as use-case

Manufacturing industry was chosen as an application area because of both the potential for 5G in the manufacturing industry and the breadth of existing research on 5G in the manufacturing industry and accompanying challenges. Additionally, in the manufacturing industry we can also find the applicability of three main service types proposed by 5G - eMBB, mMTC and URLLC.

For example, we can make a slice for the telephony services for the employees under the eMBB service type, IoT application for smart manufacturing under mMTC service type, and finally for handling the emergencies in the manufacturing facility under URLLC service type. Nevertheless, due to the time limitation, this thesis will only focus on the eMBB service type for the manufacturing industry.

Additionally, telephony services were considered as my customized service set because it is a very common service and is easily relatable to any reader of this thesis. The recent explosion in the number of OTT services such as WhatsApp, Zoom and Skype and that they also offer similar telephony services at a cheaper rate compared to the cellular telephony services, makes it an even more interesting field to study. This brings a tight competition in this service set and how 5G can help network operators to compete with other OTT service providers will be undoubtedly an interesting topic of study. The blog post by Ericsson also motivated me to choose telephony services as my use case [ZYR+18].

3.1.2 About the example company

Consider a manufacturing company named *Company A* who has an office in Trondheim and is on its journey to digitalize its business. Some of the identified services for this company include telephony services, performance and monitoring services and emergency services. They have approached a Network Slice Provider (NSP) in their region named *Company B* to consult how the 5G network slicing can be beneficial in their case. Since the requested services differ in their requirement and need for a different QoS treatment, *Company B* has proposed a solution to use a separate slice for each service under the category of Network Slice-as-a-Service (NSaaS) [Tov18].

NSaaS is a customized cellular network with all the necessary physical or virtual network functions to fulfill the SLA made between NSP and tenant (network slice customer). Additionally, NSaaS also gives the right to slice customers to manage the slice and customize the communication services.

Scenarios in this case company is explained in Sections 4.1 and 4.2.

3.1.3 Telephony service as a slice

This section covers the telephony service as a slice and their requirements. Telephony service is referred to as the development and/or operation of a telephone or telephonic system which are part of a telecommunication system for transmitting voice and data over long distances with or without the use of wire [GSM18a].

Company B proposed a slice to provide telephony services for the employees of *Company A* by considering the prescribed service requirements. From this section onwards we call the slice a Telephony slice for the easiness to describe its functionalities. Telephony slice is composed of different other services such as basic voice call, call forwarding service, call barring service, SMS, voicemail, real-time content sharing, and video call services. Telephony slice intends to provide enhanced communication services to the employees at *Company A*.

Following terminologies will be used in this section to elaborate the use case [3GP21b].

- **Slice template:** A slice template is defined as a description of various capabilities of a slice that is needed to fulfill the requirement of different services deployed under the slice. The slice template is provided by an NSP and it can be either a general one or a customized one.
- **Network Slice Provider(NSP):** They are the one who owns a collection of such network slice templates and provide different slices.
- **Tenant (Network slice customer):** The tenant is referred to as the one who buys the network slice from the network slice provider and makes use of the slice description as a reference for SLA/contractual agreement with the network slice provider.

In our case network slice provider (NSP) is referred to as *Company B* while the tenant is referred to as *Company A*. The first step for creating a slice is to make a slice template for the Telephony slice by mapping the requirement of the slice to the capabilities of the network. GSM Association (GSMA) [GSM18a] has proposed a General Slice Template (GST) and Network Slice Type (NEST) to represent tenant's service requirements in a common language that is understandable by both NSP and tenant. A representation of telephony slice in three layer model can be found figure 3.1.

GST is comprised of different general attributes made by NSPs to characterize any slice and its diverse requirements. NEST is also a template with the attributes of GST filled by values according to the specific needs of the slice and it can also be used as an input to create Network Slice Template (NST) defined by 3GPP [GSM18a]. From the NST template, we can create Network Slice Instances (NSIs) which comprises of different NFs either physical or virtual needed for the deployment of the slice.

So, I have adopted the GST template to make NST for analyzing and describing the needs of the tenant and capabilities of NSP and further to make SLA out of it. Figure 3.2 illustrates the process of creating NST using GST template. However, as the scope of this thesis is on service continuity and service orchestration, the slice instantiation process will not be discussed in detail.

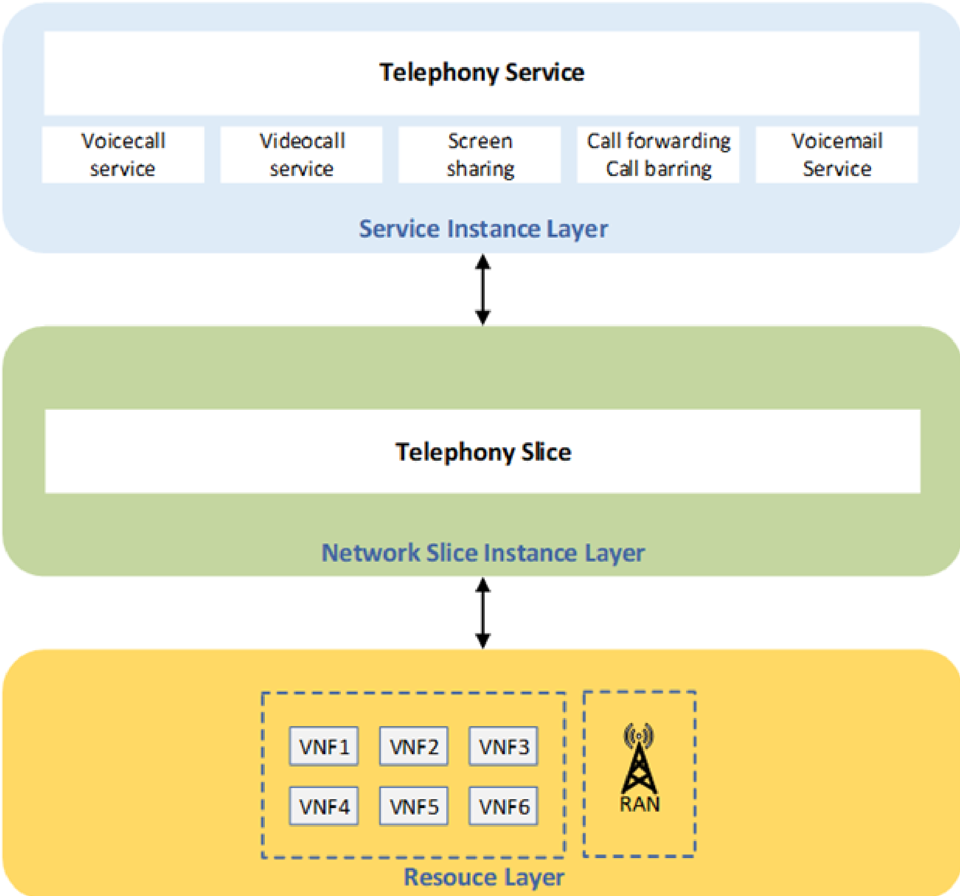

Figure 3.1: A representation of Telephony slice

A GST or slice blueprint is composed of three main parts [OLTR20], Slice Service Type (SST), Slice topology, and Slice Attributes. An illustration of this can be found in figure 3.3.

1. **Slice Service Type (SST)**

   It is a 3GPP parameter that specifies the 5G service category supported by the slice [vR18]. SST has been assigned the numerical values as follows "SST=1" for representing eMBB, "SST=22" for uRLLC, "SST=3" for mIoT, and "SST=4" for V2X. In our case, the telephony slice will come under the eMBB.

2. **Slice topology**

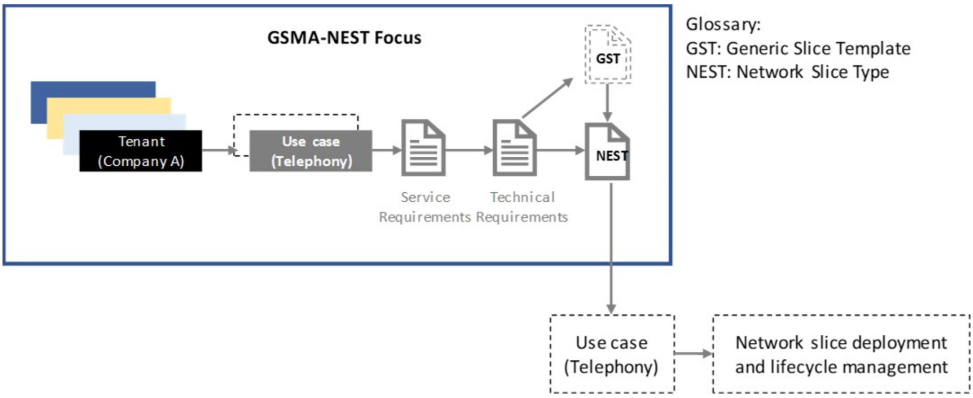

Figure 3.2: An illustration of Network slice description, adapted from [GSM18a]

This provides information about how a slice will be realized in the network from a logical point of view. It will focus on two things:

- Information on the nodes and their functionalities in the network is needed to establish end-to-end slice capabilities.
- Information on how these nodes will be connected and the connectivity type used.

Slice topology of the telephony slice includes nodes such as UE, the access point (gNodeB), the core network (AMF, SMF, UDM, NSSF and UPF), 3rd party VNF and MEC. Since this thesis does not contain any practical work involved, details on slice topology will not be studie further.

3. **Slice attributes**

   This is the set of attributes that are defined in GST. Slice attributes are determined based on both the performance and functional level requirements of the service. For example, functional requirements can be related to the attribute which specifies the functionality provided by the slice while performance attributes specify Key Performance Indicators (KPI) supported by that slice.

   Tables 3.1 and 3.2 summarize the functional and performance attributes of the telephony slice and it can be considered as a NEST for the Telephony slice. I have used the same NEST to define SLA between NSP and tenant.

GSMA has classified the status of each attribute into three “attribute presence”[GSM20], namely,

1. Mandatory – The attribute’s value must be present.

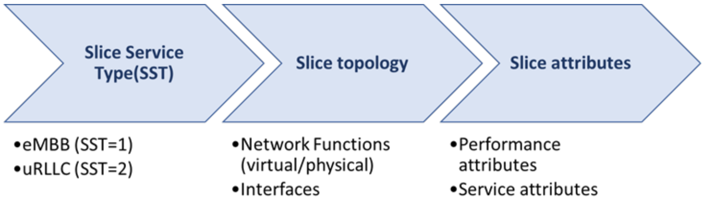

Figure 3.3: Illustration of a telephony slice template

2. Conditional – The attribute's value is mandatory if a certain condition exists.
3. Optional – The attribute's value doesn't need to be present.

Some attributes used in the table are elaborated in detail below [GSM20]:

- **Area of service:** This attribute indicates the area in which UE can access this Telephony slice. I have used the information provided by ISO 3166-phase-2. Since this slice is for *Company A*'s Trondheim office, I have selected the region of Trøndelag as the area of service.
- **Isolation:** This attribute indicates the degree of isolation that is required by the slice. More specifically, isolation can be achieved in 3 ways;
  - fully/partly isolation
  - physical isolation
  - logical isolation.

  A detailed discussion on this is provided elsewhere [GSM20]. Since Telephony slice demands a secure slice, tenant/isolation is very important to provide data integrity.
- **Mission Critical Support:** This indicates that this slice has priority over other network slices for control plane and user plane decisions. When a slice has been assigned mission-critical support, priority can be applied in two forms,
  - mission-critical capability support (inter-user prioritizations, pre-emption, local control (dynamic and temporary user prioritization))
  - mission-critical service support (different priority allocated to services in a slice)

  In certain situations telephony slice demands prioritization over a few other services. Some examples are;

Table 3.1: Functional attributes for telephony slice

Functional attributes	**Specification**	**Attribute presence**
Geographically limited services	1880 km2	Mandatory
Access	Indoor, outdoor coverage expected	NA
User density	1 user/5m^2	NA
Access restriction	All employees have access to every service except consultants and students.	Mandatory
Authorized access	All permanent employees should be authorized	Mandatory
Prioritization of services	Prioritizing some services during network traffic congestion or according to policies.	Mandatory
Area of service	According to ISO 3166-phase 2 Country code = NO Region (county) = NO-50 (trøndelag)	Conditional
Isolation (Physical isolation)	• Process and thread isolation • Physical memory allocation • Physical network isolation	Mandatory
Isolation (Logical isolation)	• Virtual resource isolation • Tenant/service isolation	Mandatory
Mission-critical support (user prioritization in a slice)	• Inter user prioritization • Pre-emtion • Local control	Conditional
Mission Critical (MC) service support (service prioritization in a slice)	• MC audio and video • MC interworking	Mandatory
User preferred service priority	• Supported	Mandatory
MMTel support (IMS support)	• Supported	Mandatory
Performance monitoring	• Active monitoring is needed	Mandatory

	• Passive monitoring supported for tenant	
Performance prediction subscription for tenant	Active prediction supported for all services (Request via API)	Optional
Root cause analysis data subscription	Passive investigation (request via API)	Optional
Simultaneous use of other network slices	Can be used with any slice that has the same SST value	Mandatory
Multimedia Priority Service (MPS) capability support (MPS=yes)	• User prioritization • Pre-emption	Conditional
Multimedia Priority Service support (multimedia service priority =yes)	• MPS for MMTel voice (priority =1) • MPS for MMTel video (priority =2) • MPS for data	Conditional
Supported device velocity	• Stationary 0 km/h • Pedestrians 10 km/h • Vehicular 120 km/h	Conditional
Group communication support	Broadcast/Multicast	Mandatory
Call forwarding service	Call forwarding is active when the callee is not available and the call route to a different number.	Mandatory
Call barring service	Call barring is active, cannot call someone who has in the call barred list.	Mandatory
Videocall service	Employees can call each other with video.	Mandatory
Voicemail service	Service is activated in case of the callee is not available and an announcement will be played.	Mandatory
Real-time screen sharing	Screen sharing service can be enabled during an ongoing voice or video call.	Mandatory
Voice call	Employees can use basic voice service	Mandatory

Table 3.2: Performance attributes for telephony slice

Performance attributes	Specification	Attribute presence
Bandwidth	High bandwidth is required in case of video calling is requested. Low bandwidth for voice calls.	Mandatory
E2E latency	<10ms is preferred for all services	Mandatory
Jitter	<30 ms, to avoid packet loss	Mandatory
Packet loss	Less than 5%	Mandatory
Availability	99.99%	Mandatory
Reliability	99.99%	Mandatory
Customized 5QI	Supported	Conditional
Connection density	Minimum of 1000 devices including cellphone, laptop, tablets etc.	Optional
Coverage	Mostly indoor coverage, outdoor coverage expected	Conditional
Mobility	Seamless handover between legacy networks.	Conditional
E2E encryption	At least 256 bit or higher.	Mandatory

  - HD quality voice and video will always have higher priority over other services
  - Prioritization of some users over others as there is a hierarchy of different employees within the tenant. High prioritized employees will have different privileges to access the service especially when the user is moving.

  More on this will be discussed in the SLA section.

- **User preferred service priority:** This attribute is proposed by this thesis and refers to the dynamic change in the service prioritization list when user preference over some services have been registered.

  This can improve the QoE of the service in a way to dynamically allocate the radio resources according to the user preferred service priority list. Most importantly, this attribute cannot be fully predicted because it usually happens when the user experiences some degradation in quality of some services and then decides to terminate other services to save the bandwidth for the most critical ones.

All other functional attributes are self-explanatory and hence are not explained in the above list.

Once the slice template is ready, slice instantiation configuration and activation will occur under the supervision of the slice manager. The slice instantiation process and life cycle management of slices are out of the scope of this thesis and is not further elaborated on.

3.1.4 Service Level Agreement (SLA)

According to International Tele-Communication Union (ITU-T) SLA is defined as [ITU02]

"a formal agreement between two or more entities that is reached after a negotiating activity with the scope to assess service characteristics, responsibilities, and priorities of every part"

The requirements and characteristics of different service types in legacy cellular networks are almost the same because services are mostly related to basic communication services such as voice and SMS. Therefore, most SLAs between the service provider and tenant contain the same metrics and threshold values [HHNS18].

On the other hand, slice-based 5G networks proposed a concept of multi-requirement tailored services over a common infrastructure. This creates a need for separate individual SLA solely for each slice which has different combinations of VNFs, metrics, and performance requirements within the same network [PFFOA+20].

After receiving the requirements of a slice from the tenant, the NSP will offer a defined level of service in the form of SLA. Here I have created this SLA from the NEST template (Figure 3.2). The SLA will ensure that it contains all the information regarding the contracted services, performance metrics, and their expected performance level, and it makes sure that both parties have the same understanding of the requirements.

Since 5G is envisioned to provide an end-to-end (E2E) network slice as a service to enterprises, the importance of an SLA is very high. An E2E slice process involves a lot of complexity in terms of the composition of network devices and orchestration of different services. So, real-time service monitoring is necessary for supporting a successful SLA.

Consistent real-time monitoring of the network will help to immediately identify network issues and aids in saving money and troubleshooting time. It is difficult to predict network behavior in real-time. The trustworthiness towards the SLA will improve if the operator can provide the status of the quality of the service and check how much it satisfies the SLA requirement while customers are using the services.

The level of mobility of the user (assuming that only indoor application is expected) is considered both while proposing the SLA for the Telephony slice and while adopting KPI values for the service performance. The value range in the SLA agreement is taken from different 5G recent test cases in literature [Dat20, ALA+19]. These values will further be used to analyze the need for policy making on the occasion of any SLA violation. I have defined an SLA template for the Telephony slice and it is described in Figure 3.4.

The metrics used to measure service performance in SLA are

- **Availability:** is defined as the proportion of time over a given period in which the slice is functional. This will depend on the availability of the infrastructure resources allocated and the services deployed over the slice. 5G has offered to provide 99.999% of availability.
- **Bandwidth (Mbps)**[1]
- **E2E latency (ms):** this metric shows the time a packet takes to reach its destination. Since 5G is an end-to-end service providing concept, I have adopted the metric E2E latency to analyze the delay. That means the latency is the sum of the delay experienced by the packet when transmitting from the UE to the destination UE (consists of 5G RAN, 5GC, and transport network).

  5G promises ultra-low latency of approximately 1ms when using 5G SA network while 4G provides 50ms [BILP20]. This extent of mismatch in latency can lead to serious issues in service continuity.
- **Jitter:** is defined as the variation in the delay of packets sent and received between the same set of receiver and sender. This primarily happens due to packet loss or packets arriving in incorrect order. This is caused by varying delays faced by different packets due to interference during air transmission. When the jitter value is high in a voice call, callee experiences a degradation in call quality.

  Experiments [Dat20] suggest that <30 ms jitter value is acceptable and a jitter value higher than that can impact call quality and lead to packet loss.
- **Packet loss:** is another metric, which is a very crucial parameter since it represents the user experience. Packet loss is defined as the number of packets lost during transmission between sender and receiver. It can be either represented as a number or a percentage of packets lost. It can be calculated as

  Experiments show that [Dat20] packet loss should be <5% to get a decent call quality and >10% packet loss can significantly impact the call quality.

[1] These metrics are not explained as those have been used in this thesis in accordance with commonly understood definitions.

SLA ID	xxxx
Slice owner	Company A
Network operator	Company B
Slice purpose	To provide telephony services
Slice name	Telephony
Slice ID	xxxx
Valid from	dd:mm:yy
Valid til	dd:mm:yy
User density (avg)	1 user/5 m2
Device density (avg)	100

KPI		Services				
		Voicecall	Call forwarding	Call barring	Videocall	Voicemail
Bandwidth	10-20					
	20-30	✓	✓	✓		✓
	>50				✓	
E2E Latency	<10ms	✓	✓	✓	✓	✓
	15-30ms					
	>30ms					
Jitter	<10ms	✓	✓	✓	✓	✓
	15-30ms					
	>30ms					
Packet Loss	<2%	✓	✓	✓	✓	✓
	<5%					
	>10%					
Availability	99.999%	✓	✓	✓	✓	✓
	99.99%					
	<99%					
Reliability	99.999%	✓	✓	✓	✓	✓
	99.99%					
	<99%					
Security	Physical isolation	✓	✓	✓	✓	✓
	Logical Isolation					
Voice Quality	EVS codec(HD voice +)	✓				
	HD AMR - WB					
Video Quality	H.266				✓	
	H.265					

Figure 3.4: Example SLA for Telephony slice

$$Packet\ loss(\%) = \frac{(No.\ of packets\ sent) - (No.\ of\ packets\ received)}{Total\ number\ of\ packets} \times 100$$

- **Security**[1]
- **Reliability**[1]

- **Voice quality:** this metric is equivalent to the speech quality and it provides the quality of the voice signal as it may be perceived by the user of the service.
- **Video quality:** this metric is used to indicate the quality of the video signal. According to GSMA and 3GPP, EVS codec and H.265 are mandatory for voice and video communication services over Vo5G [Hua18], and also a new codec H.266 for voice has been launched recently. So, I have used the latest codec in the SLA.

In addition to the proposed SLA Table I have made some other rules to the SLA and they are listed below:

- Employees in *Company A* are categorized into Gold, Silver and Bronze for the efficient allocation of the user priority while accessing the service. The employees under the category of Gold should be given HD[2] voice and video quality no matter from where they are accessing the service within the Trøndelag region.
- All the employees will be assured HD voice and video quality within the premises of *Company A*.
- When the services are accessed from a NSA deployment of 5G or a LTE network, this SLA will be applicable and it will have higher priority.
- In case user-preferred services are registered during an ongoing session the radio resources must be allocated accordingly.
- Best effort traffic will be offered when services under Telephony slice are accessed from the 3G/2G network.
- Real-time screen sharing service will not be supported if the services are requesting from 3G network.
- Network resources allocated to the Telephony slice will be shared among other slices serving *Company A*.

3.2 Scenarios

To analyze and point out the challenges in service assurance according to SLA for telephony slice is another major aspect in the case analysis. To propose a general policy that covers all the services under the slice will be a time-consuming task and may contain a lot of policy rules. So, for practical reasons, I decided to select three

[2]HD: High Definition

of the services from the telephony slice service list namely, voice call, video call, and real-time screen sharing to study the service continuity and service assurance challenges. Hereinafter, these services will be called composed service set.

The reason for selecting these services is that voice is the basic service and still the primary source of revenue for mobile operators and enhancement of voice quality and other features using the 5G capabilities is important to compete with other OTT services.

Even though the video call service has been offered from the 3G era, it was not a huge success for network operators in terms of revenue. So, a QoS assured video call may have a chance to change the revenue out of it. While on the other hand screen sharing is a new service proposed by 5G network and has a lot of potential to contribute to the revenue of the network operators.

Additionally, the composed services can also be mapped into the new 5G voice service proposed by Ericsson [Eri20] called interactive calling and also can be viewed as an application of Rich Communication Service (RCS) proposed by GSMA [GSM17]. RCS will eliminate the need for different apps for different services which we are using today like video service, messaging app, content sharing and group chat. Instead, RCS will bring this all under a single umbrella. Hence this set of services are the face of enhanced telephony services, selecting the same for identifying the challenges will be a real contribution to the current 5G deployment development.

By using the same composed service set I have created two scenarios and their description can be found in the following sections.

Before explaining the scenarios, I have listed some assumptions below for explaining the slice mobility capabilities for better illustration and these apply to both considered scenarios:

- **Assumption 1:** Mostly indoor applications will be expected.
- **Assumption 2:** It is assumed that the UEs are capable of dual-radio connectivity (UE can transmit/ receive data from multiple access networks).
- **Assumption 3:** All permanent employees in Company A have access to the services under the telephony slice (students and consultants excluded).
- **Assumption 4:** Seamless handover between 5G network.
- **Assumption 5:** All users in the scenarios are subscribed to the same IMS services (all UEs belong to the same subscription).

- **Assumption 6:** Best effort Service Assurance (SA) will be offered when the user mobility rate is high.
- **Assumption 7:** Telephony services over other wireless networks such as WiFi calling are not supported (handover to WiFi is not supported).

Scenario 1

Scenario 1 will be considered for explaining service assurance and this act as a reference to the second scenario. For ease of explanation, the scenarios are defined considering two employees Alice and Bob of Company A.

Consider an employee Alice from Company A at her office and is connected to the Telephony slice with the expectation of zero mobility in an ongoing video call session with screen sharing enabled with another employee Bob. Both Alice and Bob are connected to the same AN at the office and no change in IP address allocated to the UE and P-CSCF of IMS is expected.

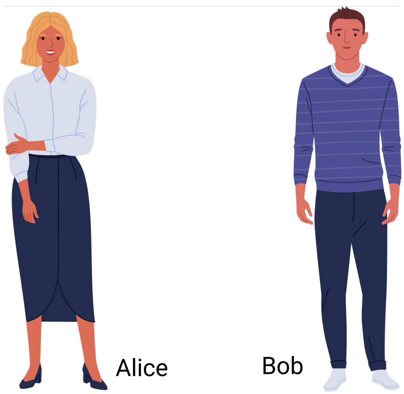

Figure 3.5: An illustration of Alice and Bob[3]who are employees at *Company A*. Both Alice's and Bob's UE are authorized and connected to the access point (5G RAN) at their office to the Telephony slice. Being in the Bronze category, neither Alice nor Bob has any special privileges given to Gold category employees

Both Alice and Bob come under the Bronze category of users, this will help to filter them out of the inter-user prioritization list. Illustration of this scenario is shown in figure 3.6

Scenario 2

This scenario is trying to find out the service continuity and assurance challenges when different service interaction happens while the user is mobile.

[3]Images are licensed from Adobe Stock

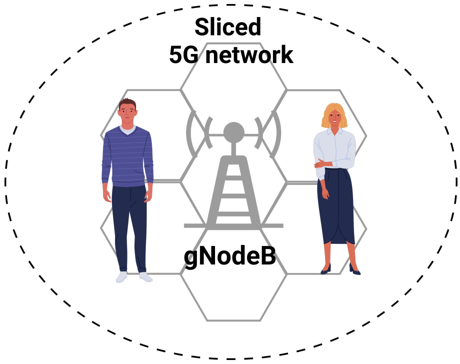

Figure 3.6: Illustration of Scenario 1

For this scenario, I have to consider the same employee Alice from *Company A* who is traveling and firstly she was attached to 5G RAN (5G SA) at her office and then moves to 5G public then to LTE network then to 3G/2G network. This Scenario contains the involvement of legacy networks and will help to understand the 5G network inter-working with the legacy networks. My scope of work looks into how the service continuity will be maintained and what factors influence the Quality of Services. Illustration of this Scenario is shown in figure 3.7

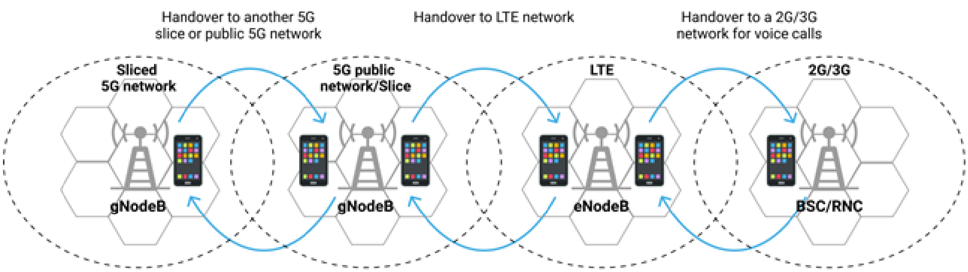

Figure 3.7: Illustration of Scenario 2

Considering these two scenarios as the point of departure, the case study analysis looks into the possible challenges and problems that may arise in different situations that these Scenarios can lead to. These are further studied in detail in the following Chapters.

Chapter 4
Case study analysis

As mentioned in Section 1.4.2, this chapter will focus on testbed implementation based on findings from literature where researchers had recently used a similar set of services in 5G use cases. Unfortunately, I have not found any research work that uses the same set of services that I had considered for my testbed. This is further addressed in Chapter 5.

Instead, a few hypothetical situations from both scenarios are taken and mapped on to recent research dealing with similar situations to analyze the possible outcomes and thereby answer my research questions. This chapter starts with Scenario 1 and its different KPI[1] values and checks the scope for any SLA violation and does the same for Scenario 2.

4.1 Scenario 1: Implementation challenges

When considering the first scenario where zero mobility is considered, we can elaborate the scenario with two employees working at *Company A*, Alice, and Bob. Both Alice's and Bob's UE are authorized and connected to the access point (5G RAN) at their office to the Telephony slice. Since they are connected to the slice and zero mobility is expected, according to the SLA[2] they are supposed to get all the benefits of 5G slice's E2E service assurance. Additionally, it is also important that they are getting the service performance level for different KPIs mentioned in the SLA. Being in the Bronze category we are assuming that both Alice and Bob have no special privileges to get a user priority and service priority that is assured for the employees in the gold category.

[1] Key Performance Indicators
[2] SLA: Service Level Agreement

4.1.1 Challenges with service orchestration

This section investigates the situations when the service orchestration fails. Even though the ongoing video session is giving all the required performance levels, the QoS[3] might change due to the state of change of any network infrastructure entity or monitored network parameters.

Two such examples where QoS might change are:

1. Consider an employee Charlie (Gold category user) of *Company A* who has started using a 4K video streaming service simultaneously as there is an ongoing video session between Alice and Bob. In such a situation according to SLA we must ensure two things:

    a) ensure that the gold category employee is getting the user and service priority

    b) to ensure all the employees in *Company A* will get HD [4] quality on voice and video services when the services are accessing within the premises of *company A*

    However, this both cannot be achieved at the same time due to the lack of radio resources. Scaling down the resources allocated for ongoing video call sessions for allocating resources for video streaming can cause service degradation in the video call. It will be interesting to understand how the network will cope with this situation.

2. If the service orchestrator decides to scale down the resource allocation for a particular service to save the resources, what will happen if that service is a user-preferred one. In our case network has sent a notification to Alice's UE that the HD quality of the video call will not available and instructed her to turned off the video, but she did not do that since it was needed. How will the system cope with the challenge when user preference and changes made due to network congestion conflicts with each other?

Other factors that may affect the service assurance can be security violations and the unavailability of infrastructure entities. Security isolation can be achieved by allocating dedicated private cloud and this can be maintained as long as there is no security attack. Unavailability of infrastructure entities-related faults and error management can be handled by the NFV-MANO framework. However, taking necessary action when the service is not restored within a certain time interval should be included in the policy.

[3] QoS: Quality of Service

[4] HD: High Defnition

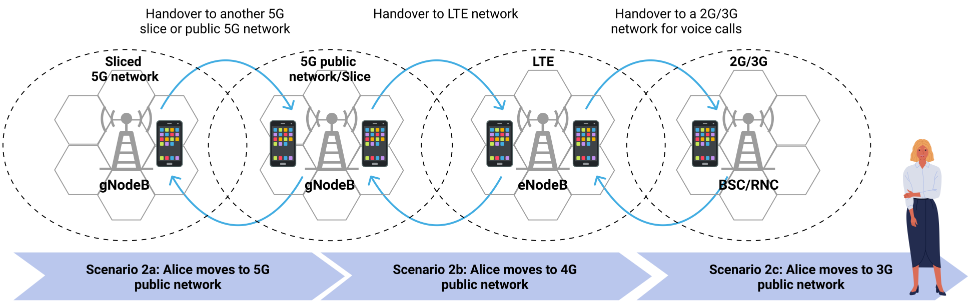

Figure 4.1: Three sub-scenarios in Scenario 2

According to NFV-MANO[5] architecture, service orchestration and network resources allocation are part of the responsibilities of NFVO[6], NFVI[7], and VIM[8]. Even though they can efficiently manage the resource allocation according to priorities assigned in the SLA and deliver the user desired QoS, the conflicts created by the above-described situation need the assistance of policy for ensuring the QoE for the user.

An efficient resource allocation and orchestration are vital for providing a desired level of QoS. Finding out all the possible service conflict situations that can arise during service execution especially when executing composed services are very important to avoid conflicts. The solution to these challenges will be presented in Chapter 5.

4.2 Scenario 2: Implementation challenges

This section is an elaborated version of Scenario 2 explained in Chapter 3. As illustrated in Figure 4.1 three sub scenarios of Scenario 2 are considered in this section, namely,

1. **Scenario 2a**: Alice moves to 5G public network
2. **Scenario 2b**: Alice moves to 4G public network
3. **Scenario 2c**: Alice moves to 3G public network

[5]MANO: Management and Operation
[6]NFVO: Network Function Virtualization Orchestrator
[7]NFVI: NFV Infrastructure
[8]VIM: Virtualized Infrastructure Manager

Further, potential challenges relating to mobility and security in all three sub scenarios are discussed with examples.

4.2.1 Scenario 2a: Alice moves to 5G public network

Alice has started moving and is out of coverage of Telephony slice network and now she is in the coverage area of SA[9] 5G network with 5G radio directly connected to 5G core. This network will be addressed as a 5G public network in the coming session. Alice is still in the video call session with Bob, and we are assuming that Bob is at the office. We have assumed that the NSP[10] has another SLA for ensuring service continuity while accessing Telephony slice services from outside the slice coverage area to ensure service continuity. This section will investigate the scope and challenges for maintaining the same QoS of the slice while Alice is mobile/moving.

Challenges with mobility management in Scenario 2a

Upon the agreed threshold for the signal strength of the sliced network, the Handover (HO) request must be initiated by Alice's UE to maintain the service continuity. Since she is still connected to a 5G network, there will be a chance for inter-slice handover in addition to the traditional way of horizontal or vertical handover. Unlike the other HO mechanisms inter slice HO always need not be an event triggered one and this HO can happen in two ways [SBJT20]:

1. **Network triggered**: i.e, according to the changing service requirements, service degradation or upon policy intervention, or;
2. It can be an **UE initiated** one.

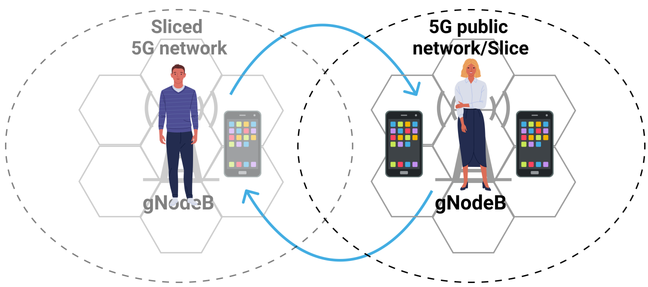

Figure 4.2: Scenario 2: Alice has started moving out of the telephony slice and is currently in the coverage of SA 5G network with with 5G radio directly connected to 5G core. Bob is still in the 5G Telephony Slice

[9]SA: Standalone
[10]NSP:Network Slice Provider

In our scenario, Alice may have two options for proceeding with the handover process, namely,

1. **Inter-slice HO**: if other slices are available which can ensure offered QoS in the SLA, HO request will be sent to the SMF of that slice (inter-slice HO) and the slice must belong to the same service type (in our case it is eMBB)
2. **Horizontal or vertical HO**: handover request will be sent to the public 5G network.

Both options have their own challenges and advantages. The causes that can lead to inter-slice handover are determined by several factors and they are listed in Table 4.1. Even though the selected slice for HO belongs to the same service type, there is a chance that it may not satisfy all the requirements specified in SLA since that slice may serve other services.

As mentioned, a slice is identified in the network using an identifier called S-NSSAI[11], a collection of S-NSSAI forms NSSAI. A user can only trigger an HO to its subscribed S-NSSAI and a set of subscribed S-NSSAI is referred to as Allowed NSSAI. This Allowed NSSAI list can contain all the NSSAI that the user can access including the user subscribed one (default configured NSSAI) and the one allocated by the NSPs (configured by serving PLMN).

In our scenario, when Alice is moving out of coverage of Telephony slice, according to 3GPP TS 23.502 [3GP21b] modification of Allowed NSSAI followed by PDU[12] session management process will be required since Telephony slice is a geographically limited one. PDU session management process is comprised of two processes,

1. PDU session release from the current slice
2. Establishment of another PDU session with the target cell. Release of PDU session with current slice is carried out according to the instructions from AMF, SMF[13], and PCF and this also results in the release of all QoS configuration, RAN resources, allocated IP address, etc.

After the PDU session release this step is followed by registration process in which a new AMF[14] selection will be carried out and all the user context information will be transferred to the new AMF. After completing the session registration to

[11] S-NSSAI:Single-Network Slice Selector Assistance Information
[12] PDU: Packet Data Unit
[13] SMF: Session Management Function
[14] AMF:Access and Mobility Function

Table 4.1: Causes of inter-slice handover, adapted from [SBJT20]

Main Causes	**Potential Inter-Slice Handover Triggers**	**Description**	**HO Initiation Point**
Slice-specific Condition	1. Access network condition	Received signal strength (RSS), Bit Error Rate (BER), Noise, interference, link capacity deterioration in the current slice	UE initiated/Network initiated
	2. Slice delay	Due to high end to end delay causing degradation of QoS	UE initiated/Network initiated
	3. Slice bandwidth	Bandwidth deterioration due to various reasons such as high traffic load.	UE initiated/Network initiated
	4. Reliability	The error rate from a slice might increase due to radio link failure or node failure.	UE initiated/Network initiated
	5. Audio/video quality	Due to the lack of enough bandwidth or noise, audio/video quality is deteriorating.	UE initiated/Network initiated
Service requirements	QoS requirements for each service or combined service in the SLA doesn't meet	Telephony slice doesn't meet the desired QoS for services in terms of jitter, packet loss, availability, latency, etc.	UE initiated/Network initiated
Tenant's preferences	1. Slice stress/load	Due to the overload traffic on T*elephony slice* for ensuring service assurance inter-slice handover will be triggered.	UE initiated
	2. Subscription policies	Since the Telephony slice is a geographically limited one, once the user is out of that region an inter-slice HO will be triggered.	UE initiated
	3. Pricing/billing	Telephony Slices may discontinue its services to a user if a user runs out of its available credit.	Network triggered
	4. Upon user priority	If one employee under the gold category requested service and due to lack of adequate resources in the slice HO to another slice will be triggered to ensure service assurance.	Network triggered
User preferences	Slice policies	Policies regarding user preference will be applicable here.	Network triggered
	Degradation of QoS	Degradation of QoS such as audio/ video quality, unavailability of screen sharing can cause HO.	UE initiated/ network triggered
	Security isolation	For getting a proper security assurance	Network triggered
Intra/inter Technology HO	Horizontal HO	When Alice is moving into new subnet may cause her to move out of coverage of *Telephony slice,* this can lead to inter-slice HO	UE initiated
	Vertical HO	When the Alice is moving from one access technogy to other, for example from 5G to 4G, an inter slice HO can be triggered	UE initiated

the target AMF, the UE will receive the allowed NSSAI for this access-type from AMF by consulting with PCF[15]. A detailed signal flow for this HO process can be found in [3GP21b]. Policy intervention has a crucial role in selecting an appropriate slice for HO, for example, some slices may be expensive, even though it satisfies all the QoS requirements of Telephony slice. HO to the expensive slice can cause extra pay to the employees, at the same time this extra money will not be a problem for employees in the gold category. So, a clear distinction for selecting an appropriate slice is necessary.

One another example is the rate of isolation provided by the slice. If the slice is not providing efficient security isolation specified in the SLA, HO to that slice must be avoided.

Inter-slice mobility to ensure service assurance and service continuity is challenging due to the complexity brought about by different slice requirements and diversity in the rate of slice isolation. However, very little research is going on this matter and the ones identified are based on V2X[16] communication due to the nature of stringent latency requirements [MNT19] [ZLC+17]. This thesis hypotheses that this challenges in inter-slice mobility can be solved thorough a study and research on every service type of slice (eMBB, URLLC, mMTC) and a general Policy algorithm for the inter-slice mobility management. In the following sections I delve into the details of such a general policy algorithm.

Mobility management in Scenario 2a

One of the main features of the 5G network is "network densification" which is to enlarge the cellular coverage and/or augment their cellular capacity by adding more small cells into the network [ETS15]. Even though network densification serves the purpose it also demands the network to perform HO more frequently to ensure service continuity [HQ20]. As this is the case, more efficient HO processes are needed to attain the full benefit of the 5G network to ensure the desired QoS offered by the network. Some of the techniques for maintaining better service continuity between 5GS are:

1. to apply make-before-break connectivity where the UE set up the connection with the targeted cell before losing connection from the serving cell [SGJ+18]
2. Synchronized handover mechanism

An earlier 5G network trial test as part of RAPID 5G PROJECT (intended at delivering broadband 5G services within 60 GHz band in a user dense area) shows

[15] PCF: Policy and Charging function
[16] V2X: Vehicle to everything

that most of the 5G promised KPI values can be achieved by using an efficient handover procedure named as a make-before-break mechanism and an optimized RSSI threshold level [SGJ+18]. These promised KPI values are similar to the ones we have in our SLA, which means KPI values in the SLA can be assured to the user when the user is accessing telephony services from the 5G public network.

Further, 3GPP release 16 [3GP21d] has been specified as an improved handover procedure known as enhanced make-before-break handover or Dual active protocol stack solution (DAPS) to reduce handover interruption time. DAPS reduces interruption during handover close to zero ms, by utilizing both radio resources from source and target cell while establishing the target cell radio link. DAPS also allows the reception of user data from source and target cell simultaneously.

A conditional handover procedure has also been specified as part of Release 16 to improve handover robustness [3GP21d] . At conditional handover, the handover request is transmitted with a triggering condition in advance to avoid poor cell-edge radio conditions and this request may include more than one potential target cell. This handover mechanism helps to achieve a more reliable handover process by proceeding with the HO only after all the configuration set up is completed in the source gNB [NHK20].

However, since the user is mobile, several factors can pose challenges in maintaining QoS of the services while the handover process occurs. Some of these factors are:

- User location to the base station antenna
- Traffic load at the moment
- Physical barriers related to the network densification technique including the difficulty to penetrate a high-frequency signal through walls
- Other competing signals such as OTT[17] signal
- Existing weather conditions

To ensure better service continuity intelligent real-time monitoring and analysis of the network is necessary. Xie et al. [XZG+18] state that the need for intelligent monitoring in a 5G network is to improve the accuracy of analytical data that is required to achieve a fully automated network and to minimize monitoring cost. Analytics by Artificial Intelligence (AI) and machine learning (ML) in networking is a hot research area [KFMJM18] and also ETSI Experiential Networked Intelligence

[17] OTT: Over The Top

(ENI) WorkGroup (WG) recommends applying ML and AI in SA[18] network for a better service assurance [EG18]. According to ETSI [EG18], the ENI system delivers enhanced customer experience by real-time monitoring which allows operators to understand the real-time status of their network at the moment and to monitor whether the network is providing desired QoS and apply reconfiguration if necessary.

The ENI system automatically collects network status and associated metrics, faults, and errors, and then uses AI and ML to ensure that network performance and quality of service are met. Because of the nature of user and expected mobility rate, our scenario also demands an ENI system for getting better service assurance. So, getting accurate real time network parameters and need for an on-demand network adaptation according to changing contextual information also underlines the importance of necessary policy management to avoid any conflicts.

Security challenges in Scenario 2a

Security threat poses an important challenge in accessing telephony slice services from the 5G public network. You [You20] discusses the security vulnerabilities related to the SIP[19]/RTP protocol and misuse of IP-sec encryption on IMS-5G[20] interfaces. The SIP protocol has been already facing security issues due to the textual message format. Since the SIP header and its values are exchanged between interfaces as in plaintext, it is possible for the attacker to manipulate several header values and it can lead to a negative impact on the QoE[21] of the user. This can also affect the authorization to the network as well.

Another possible attack is on Real-time Transport Protocol (RTP), RTP is a data transmission protocol used in VoIP. Traffic of RTP protocol can be reproduced and the attacker can have access to see whole packets that are exchanged between terminals. It is even possible to have an audio playback of the call and it is a security threat while using VoIP[22] service.

A third identified threat is related to the 3GPP standardized encryption algorithm and IP Security (IPSec), choice of selecting IPSec will depend on the manufacturer or carrier. 3GPP has made it mandatory to use IPSec on Z interface (between IMS and other network domain entities) but leaves it to the operator to choose IPSec on Gm interface (between UE and P-CSCF). So, if the UE is connected to another network other than the sliced network and has an operator policy that does not require use of IPSec, then the data will not be encrypted, and it poses a serious security threat.

[18]SA: Standalone
[19]SIP: Session Initiation Protocol
[20]IMS: IP Multimedia Subsystem
[21]QoE: Quality of Experience
[22]VoIP: Voice over IP

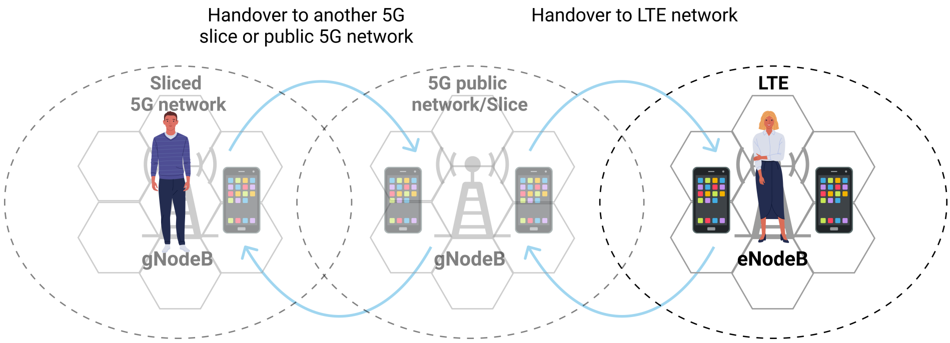

Figure 4.3: Scenario 2b: Alice moves to 4G public network

Earlier research proposed a handover authentication and key management protocol for eliminating replay and de-synchronization attacks in the 5G network[HQ20]. The authors also claim that there is a lack of forwarding secrecy in 5G. This can potentially be mitigated with the aid of security-related policies in 5G.

4.2.2 Scenario 2b: Alice moves to 4G public network

As illustrated in Figure 4.3 Alice has moved out of the 5G public network coverage and according to the signal strength threshold, UE has requested HO to the 4G/LTE/Evolved Packet System (EPS) network. Hereinafter in this report, 4G or LTE or EPS will be used interchangeably.

In this situation three types of HO can take place:

1. **LTE Slice (DCN):** If Alice is connected to another 5G slice, HO to an LTE slice can be initiated.
2. **LTE Network:** If Alice is connected to another 5G slice, HO to an LTE network can be initiated.
3. **LTE Network (from 5G Public):** If Alice is connected to the 5G public network, HO to LTE network can be initiated.

LTE Slice

For the first case, Alice's UE is connected to a sliced network and when the signal strength threshold is approached, HO will be triggered to EPS[23]. The process of interworking between slicing and EPS will be achieved via one of the EPS's features called Dedicated Core Network (DCN) [3GP21a]. DCN feature enables an

[23] EPS: Evolved Packet System

operator to deploy multiple DCN within the PLMN and each of them consists of one or more Core Network (CN) node. DCN can be considered as the first form of slicing in EPS and is also dedicated to serving a specific type of users with specific characteristics or functions, it provides isolation among users or services. DCN can contain multiple Mobility Management Entity (MME)/ Serving GPRS Support Node (SGSN)/ Serving Gateway (SGW), PDN Gateway (PDN GW), and PCRF to meet the specific requirements of services.

The selection of DCN in EPS is determined by an optional subscription information parameter called UE usage type, a value assigned to this parameter is configured by the operator and it will be stored in the HSS along with other subscription data. During HO serving network's gNB can select the DCN based on the UE Usage Type value configured by the operator and it is usually bound with the operator's policy.

In our case, a policy rule regarding the selection of appropriate DCN via NSP configured DCN-ID can help to provide service assurance and continuity when the services are accessed from the LTE network. Detailed interworking between sliced network and EPS can be found in other research works [vR18].

LTE Network (from 5G slice)

For the second case when Alice's UE is connected to the sliced network and HO to EPS is triggered, it also follows the same handover process described above. However, in this case, we assume that there is no available DCN. So the ongoing services will not get assured QoS due to a lack of dedicated network resources. And further, it follows the same HO process that applies to the interworking between 5G public networks and EPS.

LTE Network (from 5G Public)

For the third case where Alice's UE is connected to the 5G public network, HO is triggered to the EPS. Handover from 5G to 4G is supported through combo-nodes or combo-gateways (Combo-GW), which provide both 5G core network and EPC functionality [AM20]. Since Alice is connected to the same IMS core network, combined SMF (PGW-C+SMF) or Packet Data Network Gateway (PGW-U+UPF) provides IP address continuity during the handover between 5G and 4G network [3GP21b].

The HO process in this case will be as follows:

- When the signal strength is approaching the threshold mentioned in the policy, the handover decision is made by the gNB and notifies this to AMF by consulting with PCF.

- The AMF then informs the MME about the handover request, which triggers the setup of a 4G bearer for the IMS Signaling traffic for continuing the service and to set up a dedicated bearer for voice and video media.
- Further this can also allocate the required 4G radio resources upon handover request. Once this is completed, the AMF sends a handover command to gNB which triggers the UE to move from 5G to 4G radio.
- The handover process will be completed when UE sends a confirmation on handover completion to the eNB and this is passed to the MME.
- The MME then starts to complete the bearer setup, by using a bearer modification request sent towards the PGW-C+SMF. This step completes the video call handover process and all IMS signaling, and voice traffic is afterward carried over 4G radio and core.

However, according to 3GPP, Mobility between 5GC[24] to EPC does not guarantee that all active PDU Session(s) can be transferred to the EPC [3GP21a]. This means that only QoS supported by the EPC will be transferred to the EPC. For example, notification control feature of 5G QoS will not be transferred. This may affect the service quality. If the handover process is not successful or if it fails to complete due to various reasons a policy rule can be considered since this situation is violating the SLA. If the SLA violation for a particular service or combined services is not restored in a certain time interval or the handover interruption time exceeds a certain time a policy rule will be needed to give necessary compensation to the users of the slice.

Another scenario can arise when Alice turns off any of the sessions for example video/voice/content sharing. When any one of this action is carried out, this will then notify the MME and which will then allocate radio resources to the services accordingly. Before the session transfer, MME requests from ATCF[25] the codec choice that is to be used on the target RAN[26] [3GP21b]. Codec negotiation may or may not result in a degradation in the quality of audio and video. The same quality of audio and video may not be assured in the LTE network and also E2E latency is also different in LTE. 4G test cases [CTML16, O'D] show that the KPI values assured in the SLA have a variation of about double the KPI value in LTE. This indicates that guaranteed QoS in LTE will be poorer when compared to the 5GS.

4.2.3 Scenario 2c: Alice moves to 3G public network

In this scenario, Alice who is in the coverage area of the 4G network is in an ongoing video call session with Bob who is connected to the 5G Telephony slice. When Alice

[24] 5GC: 5G Core network
[25] ATCF: Access Transfer Control Function
[26] RAN: Radio Access Network

moves from LTE coverage to legacy 3G network coverage, the ongoing video session transfer from a PS[27] domain to CS[28] domain is realized via video Single Radio Voice Call Connectivity (vSRVCC) [23.16]. To support service continuity in both CS and PS domains both originated and terminated IMS sessions must be anchored in a Service Centralization and Continuity Application Server (SCC AS) in the IMS.

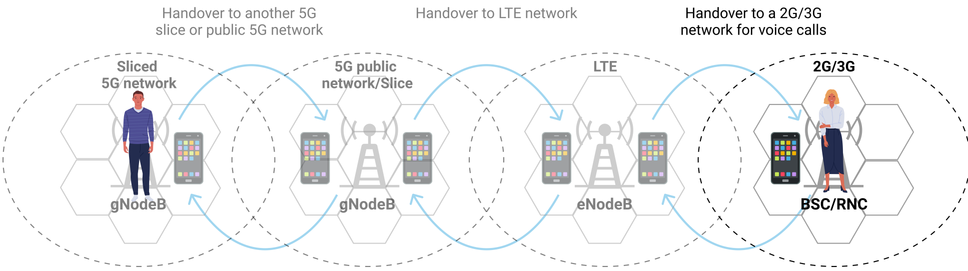

Figure 4.4: Scenario 2c: Alice moves to 3G public network

The SCC AS is also responsible for generating charging information for all session transfers for an IMS session for billing and charging. SCC AS is inserted into each session using Initial Filter Criteria (iFC) [ZD05]. For using SRVCC[29] between CS and PS domain (in our scenario 4G to 3G), it must be supported by UE, eNB, MME, PGW, P-CSCF, ATCF in the LTE network and MSC[30] server in the VPLMN, SCC AS, and HSS in HPLMN [ZD05].

Steps under go in the HO from LTE to 3G as follows [23.16]:

1. The handover process is triggered from MME when UE[31] is approaching the edge of the 4G network. When the configured signal strength threshold is exceeded, and MME will receives the handover request from eNB.
2. The MME will trigger the vSRVCC procedure with MSC server enhanced for vSRVCC via the Sv reference point with vSRVCC related information.
3. MSC server further interacts with the IMS and initiates the session transfer procedure to IMS along with the CS handover procedure to the target cell.
4. The MSC server that is enhanced for vSRVCC[32] sends a PS to CS HO response to the MME with all the necessary HO command information needed for the UE to access the UTRAN.

[27] PS: Packet Switched
[28] CS: Circuit Switched
[29] SRVCC: Single Radio Voice Call Connectivity
[30] MSC: Mobile switching center
[31] UE: User Equipment
[32] vSRVCC: video SIngle Voice Call Connectivity

5. If the SCC AS indicates that the ongoing session is with video, then the MSC server requests for radio resources for the BS30 bearer to support video and further proceeds with the vSRVCC procedure.
6. The UE receives the HO command with the BS30 bearer. This step completes the HO process after NodeB sends HO complete message to the UE.

If due to some reason, for example, if the current RAN type does not support video due to lack of resources or the user turned off the video, then video media cannot be supported in the CS network for the ongoing video session. In that case, the MSC server can set the port to zero in the "m=" line for the video media in the SDP[33] offer of the SIP INVITE request. At the same time, MSC server will inform the SCC AS that the video media is deleted and only the audio media of the speech and video session will be transferred to CS.

When the session is transferred to the 3G network, the real-time content sharing will not be available and the user will be notified about this, but voice and video calls will be continued if there are adequate resources available. In the case of screen sharing, it is a new feature of 5G, there is no evidence in the research that this new feature will also be available in the 3G network.

Regarding the QoS management in the 3G network [ZG15], UTRAN services are categorized into four;

1. Interactive service
2. Conversational service
3. Streaming
4. Background service.

Each of the service provide different functions and their requirements are also different. Resource utilization is also based on these different service types and their priority (user/service type can be prioritized). By using efficient resource scheduling algorithms 3G network can provide guaranteed QoS to high priority users.

In a 3G network, the Radio Network Controller (RNC) has responsibility for efficient resource allocation according to the service type. Three main parameters of 3G QoS are ARP[34], Scheduling Priority Indication (SPI), Traffic Handling Priority (THP). It is evident that when comparing to the 5G QoS and 4G QoS parameters

[33]SDP:Session description protocol
[34]ARP: Allocation Retension Priority

3G QoS seriously lack many of these parameters, and this will pose an impact on the assured QoS when Telephony slice services are accessed from the 3G network. There is little research on this field to provide necessary test results on the impact on 5G services when accessing from the 3G network.

4.2.4 Summary of Chapter 4

This Chapter presented and discussed in detail the various sub-scenarios that can occur for both Alice and Bob. A set of different cases that can occur under both Scenario 1 and Scenario 2 were identified and their possible implications, challenges and impact were discussed. These are further taken as learning points in formulating the Policy Engine for service orchestration in the next Chapter.

Chapter 5
Policy Engine for service orchestration

This chapter discusses how a policy management approach is suitable for service orchestration and can ensure service continuity and service assurance in 5G network. As a measure to mitigate the challenges identified in Chapter 4, the following sections map a set of policy rules that are potentially applicable to both Scenarios 1 and 2. State of the art of policy management is discussed, and this chapter ends with a proposal for a policy engine framework for the use case explained in Chapter 3.

5.1 SLA and Policy Based Management

Earlier in this thesis I had presented an SLA for telephony slice with all necessary KPIs to ensure the tenant's desired QoS level. In this chapter we go into further detail on SLA and Policy based management of network to guarantee the desired QoS. This chapter further finds out the service interaction challenges in the network management and propose a future work scope for reducing policy conflicts.

Effective and efficient network planning and management is necessary to deal with the constantly increasing number of mobile users and bandwidth-intensive services. With the emergence of different set of new and yet unforeseen services, users and applications, novel and less predictable mobile traffic patterns are also expected to arise with 5G deployment [HB15] . 3GPP release 16 proposed a Zero -touch-network and Service Management (ZSM) to automate network management and service delivery as a solution to the aforementioned problems [V1.20].

Due to the complexity of 5G and emerging services, 5G demands a highly automated network management to provide high performance, robustness, reliability and to ensure self-healing properties of a network. The ongoing research in the field has proposed a few comprehensive automation solution for 5G network management and policy-based network management is one of them [MAPS19].

Policy-Based Network Management (PBM) is where the network management is accomplished based on policy. Actuation Framework [MAPS19] and SDN[1] based management [HL12] are two examples of proposed PBM frameworks that uses the automation capabilities of 5G network for 5G slicing. For this thesis, I have adopted an SLA-driven policy management to describe my use case.

A policy is the combination of rules and services where rules define the criteria for resource access and usage and how these rules can be utilized to give better QoS and QoE to the user. A policy can be described as an aggregation of policy rules and each policy rule is composed of a set of conditions and a corresponding set of actions [V1.20]. The condition defines when the policy rule is applicable, and when it can get triggered on a particular event. Once a policy rule is activated when a condition is met, one or more actions contained by that policy rule may be executed.

5.2 ECA logic for network management

Two paradigms have been considered for describing the policy engine for the use case telephony services. They are:

1. Event-Condition-Action paradigm (ECA) (policy rules are composed of set of events, conditions and actions)
2. Condition-Action paradigm (policy rules are composed of a set of conditions and actions)

A few of the proposed policy languages for network management based on these paradigms are [HL12]:

- Policy Framework Definition Language (PFDL) - based on condition-action paradigm
- Policy Description Language (PDL) - based on Event Condition Action (ECA)
- Ponder - based on Event Condition Action (ECA)

ECA[2] policy enables actions to be automatically triggered based on the certain Event-Condition pair. Source of event that triggers the actions is fetched from different YANG data stores by subscribing to the continuous data store updates, this eliminates the need for periodic poll for the status of network parameters. YANG-Push mechanism is used to choose which network parameters are of interest and to

[1] SDN: Software Defined Network

[2] Event-Condition-Action - a language used to express complex events and actions based on one or more conditions [Din10]

get update on their state in changing network environment, thus providing a better view of network and its behavior for ensuring desired QoS for the network services [Bou19].

ECA also allows to directly translate some of the interested KPIs in SLA to filter in the YANG push subscription. When an event is activated, both network management and network devices check for the conditions that triggered an activated event, once the condition is satisfied, the ECA policy rule will be executed [Bou19]. Network management may rely upon one or multiple policies to ensure that all the possible events are encountered and make sure policies are enforced correctly, thus reducing the possibility of policy conflict and ensuring that the expected behavior is observed.

The suitability of ECA paradigm to design reactive systems and its ability to express complex actions and events and a declarative semantics makes it suitable for analysis [PPW06]. ECA has been chosen as a policy language for the use case in this thesis. There are several advantages in using ECA rules to implement an application's reactive functionality, rather than encoding it in application code. Some of these advantages are enhancing the modularity, maintainability, and extensibility of application [PPW06]. Hereinafter we call ECA paradigm for our scenario as Policy engine.

For the future scope of this thesis, proposed policy engine can be easily translated to any of the languages that is based on ECA. Moreover, policy engine evaluation using simulation or any other technique is out of scope of this thesis. So, ECA provides a better illustration of policy rules in a user centric way.

5.3 Architecture Concept of ECA policy

ECA Condition is evaluated as a TRUE or FALSE logical expression [V1.20]. The condition is specified as a hierarchy of comparisons and logical combinations of events, this allows for configuring logical hierarchies. In this thesis I have adopted a simple if and else condition check describing in natural language to map the correct event-condition pair of my scenarios. One example such of ECA policy execution is described below.

Algorithm 5.1 ECA policy algorithm

Result: Trigger policy rules
initialization;
Event A occurred:
Condition A;
Condition B;
if *condition A* **then**
  Execute Action A
else
  Execute Action B
end

Events

For the proposed policy engine of our scenarios, I have categorized the events in to two;

1. **Internal events:** are triggered when network preference over some changing network parameters is registered, such events can be a changing bandwidth, packet loss, jitter, fault etc. These parameters will be customized for each slice by NSP[3] and in our case these parameters are same as the KPIs specified in the SLA. This internal event can be received from the Network Data Analytics Function (NWDAF) in 5G network or from different SDN controllers or from Service Exposure Capability Function (SCEF) (EPS entity of the 3GPP network system for monitoring) and other operator configured monitoring system.
2. **External events:** are triggered when user preference is registered over certain network parameters or services. Such events can happen when a user prefer some services over other (eg. On an ongoing session of video call with audio, user turned off the video). User preferred events can also trigger same actions as triggered by network preferred events. External events are registered via Graphical User Interface (GUI) of the UE in which user is notified by the options via notification message or via specific icons on UE.

Table 5.1 summarizes the event attributes used in this policy engine and these are the main attributes specified in the SLA.

Table 5.1: A summary of internal and external events in the Policy Engine

Internal Events	External events

[3]NSP: Network Service Provider

BandwidthValue	UserVideoOFF
JitterValue	UserVideoON
PacketlossValue	UserAudioOFF
LatencyValue	UserAudioON
HORequest & HORequestFailed	ScreenShareOFF
AudioQuality & VideoQuality (5G or 3G)	ScreenShareON

Composition of these internal and external events are represented by logical operations [LLL07]. These logical operations are as follows:

- Conjunction (Event A & Event B & Event C): This refers to the composite event of internal and external event of type “AND” has more than one sub event. All the sub events must be true within a specific time interval to trigger an action.
- Disjunction (Event A | Event B| Event C): This refers to the composite event of internal and external event of type “OR” has more than one sub event. In this case any one of the sub events must be true within a specific time interval to trigger an action.
- Sequence (Event A, Event B, Event C): This refers to the composite events of internal and external of type “SEQ” has more than one event. In this case all the sub events must be occur sequentially to trigger an action.
- Negation (! Event A): This refers to the composite events of internal and external of type “NOT” has only one event.

Condition

The condition part of ECA policy is a statement that must be executed when the condition is true. In this thesis, “condition” is represented using a simple if and else statement in natural language. Condition statement contains certain network parameter values and is compared with the threshold for that value as described in SLA.

Most of the conditions in proposed ECA policy engine contain same event attributes as mentioned in Tabel 5.1. In addition to the above mentioned logical expressions (conjunction, disjunction etc.) there are a few additional ones such as "greater than (>)", "less than (<)", "greater than or equal to (>=)" and "less than or equal to (<=)" that are also used to represent composed conditions.

Actions/Policy rules

The action part of the ECA policy engine is the instruction set which contains pre-defined set of rules that must be executed when a condition is triggered by internal or external events. Composition of different actions are expressed using the same logical expression used for the events, such as conjunction, disjunction, sequence, and negation.

In the action set (policy rule), dynamically allocated resources refers to scaling down or up of radio resources by considering the network environment condition such as time of the day, network load, fault etc. This resource scheduling algorithms can be facilitated by automated network monitoring or by use of Artificial Intelligence (AI) and Machine Learning (ML) techniques. We are assuming that an efficient resource scheduling scheme has been adopted for this network, further discussion on the details of this out of scope of this thesis.

Whenever an event is registered, its value (Current_Value) will be compared with the previously registered value (Previous_Value) of the same event to check the credibility of the events. This helps to reduce the network overhead by not taking all events for consideration. If any condition is triggered on by the same event the corresponding action will be executed.

Table 5.2 summarizes the possible Event-Condition-Action pair of possible events of both Scenario 1 and Scenario 2.

A few of the terminologies used in Table 5.2 are explained below:

RLF recovery procedure: Since Scenario 2 considers a highly mobile user, it is very important to ensure the service continuity by triggering a HO request at the right time. If the HO request is sent too early or too late, it can affect the QoE of the service and can also lead to a temporary link disconnection, called Radio Link Failure (RLF). This is triggered by UE when a RLF is detected, in that case UE will select another cell and attempt RRC re-establishment procedure with the selected cell.

Mean Opinion Score (MOS): MOS is defined as "The value on a predefined scale that a subject assign to his opinion of the performance of the telephone transmission system used either for conversation or for listening to spoken material" [IT16]. Audio quality analysis in the VoLTE call is done by measuring the statistical and distribution information for Uplink and Downlink MOS.

MOS Call Quality (MOS CQ): Metric used to evaluate the audio quality over a conversational speech and it is represented as single number typically on a scale of 1-5, in which 1 is bad and 5 is best.

MOS Video Quality (MOS VQ): another metric used to evaluate the video quality over a conversational video service [IT16]. It uses the same scale as MOS CQ.

Table 5.2: Summary of Event-Condition-Action pairs for possible events for both Scenario 1 and Scenario 2

Event	Condition	Action (Policy rules)
Event attribute: BandwidthValue		
• Event A: Bandwidth deterioration detected (Current_BandWidthValue < previous_BandWidthValue) • Event B: User preference on bandwidth usage not met. • Event C: User preference over a service registered (UserVideoOff) • Event D: Bandwidth improvement detected after HO (Current_BandWidthValue > previous_BandWidthValue)	• Condition A: Triggered on by Event A, If Bandwidth <= 3 Mbps • Condition B: Triggered on Event C, Check the bandwidth sufficient for remaining services. • Condition C: Triggered on by Event D, If Bandwidth >= 1.8 Mbps	• Action A: Triggered by both Condition A&B, Allocate additional radio resource in the same network (Priority 1). • Action B: Triggered by both Condition A,B&C, HO triggered to any Slice /other network that has enough bandwidth to meet SLA (Priority 2). • Action C: if Action A&B is not possible, turn off the video (Priority 3) & send notification to the user that "video is not available".
Event attribute: PacketLossValue		
• Event A: Packet loss change detected (Current_packetLoss > Previous_PacketLoss)	• Condition A: If packet loss <= 5%	• Action A: Restoration of lost packet will be triggered. • Action B: Root cause analysis for packet loss will be triggered.
Event attribute: LatencyValue		
• Event A: E2E Latency value registered (Current_LatencyValue > Previous_LatencyValue)	• Condition A: If the UE is connected to 5G network/ sliced network & Current_LatencyValue > 5ms • Condition B: if the UE is connected to 4G network & LatencyValue >50 ms • Condition C: if the UE is connected to 3G network & LatencyValue >70 ms	• Action A: Triggered by condition A, HO request should send to any slice that provide desired QoS in the SLA. • Action B: Triggered by Condition A,B & C, Check the user status, if user is eligible for getting priority, prioritize the user traffic over others. • Action C: Triggered by Condition A &B, If Edge Computing (EC) is available for the network then priority will be given to process user traffic at network edge. • Action D: Triggered by all Conditions, a root cause analysis will be triggered to find out the delay contribution of each node and reconfiguration of parameters will be initiated.

Event attribute: JitterValue		
• Event A: Jitter value has registered (Current_JitterValue > Previous_JitterValue)	• Condition A: If the UE is connected to 5G network/ sliced network & JitterValue > 30 ms. • Condition B: if the UE is connected to 4G/3G network & JitterValue >50 ms	• Action A: Triggered by condition A, HO request should send to any slice that provide desired QoS in the SLA. • Action B: Triggered by condition B, user traffic will be prioritized over other traffic if the jitter is due to congestion. • Action C: Triggered by all Conditions, a root cause analysis will be triggered to find out the delay contribution of each node and dynamic resource allocation will be triggered.
Event attribute: HORequest		
• Event A: Handover request (HORequest) received from 5G sliced network	• If RSRP > -95 dB and for Reference Signal Received Quality (RSRQ) is >=-10 dB	• Action A: Initiate inter-slice HO (Priority 1) if the slice satisfies desired QoS in the SLA. • Action B: Interslice handover to the premium slice will be triggered for gold category employees. • Action B: Initiate HO to public 5G network (Priority 2).
Event attribute: UserPreference		
• Event A: User preference registered over services during handover.	• Condition A: If HO request satisfies user preferred QoS requirement in the target cell/network. • Condition B: If HO request doesn't satisfies user preferred QoS requirement.	• Action A: Triggered by Condition A, Continue HO request process to the target cell. • Action B: Triggered by Condition B, HO request will be dropped, and another HO request will be triggered by considering QoS requirement.

<table>
<tr><th colspan="3">Event attribute: VideoQuality_5G</th></tr>
<tr>
<td>• Event A: Video Quality deterioration detected in 5G slice/5G public
• (Current_VideoQuality_5G < previous_VideoQuality_5G)
• Event B: User preference over video services (UserVideoOFF) has registered (eg: video turned off by the user).
• Event C: Video Quality improvement detected (Current_VideoQuality_5G > previous_VideoQuality_5G)</td>
<td>• Condition A: If VideoQuality_5G != HD quality.
• Condition B: Dynamic resource allocation failed.
• Condition C: If AudioQuality_5G = HD voice + & Screen Sharing = ON
• Condition D: If AudioQuality_5G != HD voice + & Screen Sharing != ON
• Condition E: If VideoQuality = HD quality is restored.</td>
<td>• Action A: Triggered by Condition A, Dynamic resource allocation (either via HO or allocating different radio channel) will be triggered to maintain HD quality (high priority).
• Action B: Triggered by Condition B, notify the user with “video service not accessible”.
• Action C: Triggered by condition C, dynamic resource allocation triggered to full fill desired QoS in the SLA for remaining services.
• Action D: Triggered by Condition D, send notification “service is not available” to the user UE.
• Action E: Triggered by Condition E, send notification to the user about the event.</td>
</tr>
<tr><th colspan="3">Event attribute: AudioQuality_5G</th></tr>
<tr>
<td>• Event A: Audio Quality deterioration detected in 5G slice/5G public
• (Current_AudioQuality_5G < previous_AudioQuality_5G)
• Event B: User preference over services (UserAudioOFF) has registered (eg: audio turned off by the user).
• Event C: Audio Quality improvement detected (Current_AudioQuality_5G > previous_AudioQuality_5G)</td>
<td>• Condition A: Triggered on Event A, If MOS CQ < 4.5 & AudioQuality_5G != HD voice +
• Condition B: Triggered on Event B,If VideoQuality_5G = HD | ScreenSharing = ON
• Condition C: Triggered on Event C, If AudioQuality_5G = HD voice + quality is restored.</td>
<td>• Action A: Triggered by all condition on audio quality, maintaining voice call continuity has higher priority over other services.
• Action A: Triggered by Condition A, Dynamic resource allocation (either via HO or allocating different radio channel) will be triggered to maintain HD voice + quality (high priority).
• Action B: Triggered by condition B, service will continue on same radio channel if desired QoS met otherwise dynamic resource allocation triggered to full fill desired QoS in the SLA for remaining services.
• Action C: Triggered by Condition C, send notification to the user about the event.</td>
</tr>
</table>

Event attribute: VideoQuality_3G		
• Event A: Video quality deterioration detected in 3G network (Current_VideoQuality_3G < Previous_VideoQuality_3G) • Event B: User preference over services has registered (UserVideoOFF) • Event C: Video quality improvement registered (Current_VideoQuality_3G > Previous_VideoQuality_3G).	• Condition A: If VideoQuality_3G < 64 CSD & MOS VQ < 3. • Condition B: Dynamic resource allocation failed. • Condition C: Triggered on Event B, If MOS CQ >3.8 • Condition D: If AudioQuality_3G != HD voice + & Screen Sharing != ON • Condition E: If : If VideoQuality_3G > 64 CSD & MOS VQ > 3.	• Action A: Triggered by Condition A, Dynamic resource allocation (either via HO or allocating different radio channel) will be triggered to maintain HD quality (high priority). • Action B: Triggered by Condition B, notifies the user that "video is not available". • Action C: Triggered by Condition C, continue the user traffic on same radio channel otherwise allocate different radio channel to meet desired QoS. • Action D: Triggered by Condition E, notifies the user about video quality has restored
Event attribute: AudioQuality_3G		
• Event A: Audio quality deterioration detected in 3G network (Current_AudioQuality_3G < Previous_AudioQuality_3G) • Event B: User preference over services has registered (UserAudioOFF) • Event C: Audio quality improvement registered (Current_AudioQuality_3G > Previous_AudioQuality_3G).	• Condition A: Triggered on Event A ,If MOS CQ < 3.8 • Condition B: Triggered on Event B, If VideoQuality = HD • Condition C: Triggered on Event C, if If MOS CQ > 3.8	• Action A: Triggered by Condition A, Dynamic resource allocation will be triggered to maintain user preferred QoS on audio quality (Priority 1). • Action B: Triggered by Condition A, maintain HD quality / HD voice + if adequate resource available. • Action C: Triggered by Condition B, dynamic allocation of resources will be initiated to maintain QoS of video service. • Action D: Triggered by Condition C, Notify the user about that audio quality is improved.
Event attribute: HORequestFailed		
• HO request failed to proceed	• If it is due to the Radio Link Failure (RLF)	• Action A: UE will trigger a RFL recovery by re-establishing the connection. • Action B: Interruption time should be as low as possible.

5.4 General policies

This section presents some general policies that is applicable to the scenarios under telephony use case. This includes the general policy rules and threshold needed to ensure the QoE of the service.

1. Service unavailability, reliability and connectivity related issues due to server failure, maintenance of infrastructure, power failure, software or hardware upgrade in the network will be considered as a failure from NSP side except downtime caused by natural disasters. In case of such situations necessary penalty can be applied according to the severity of the problem.

2. If the audio quality is distorted due to high noise or Signal Noise Ratio (SNR), a root cause analysis to find out the origin of noise will be initiated.

3. High priority will be given to minimize Speech Path Delay (SPD) to those who are subscribed to the slice. A threshold for SPD will be less than 100 ms. In case threshold is not met by the service root cause analysis will be performed.

4. Call set up time threshold will be $<$2 ms in sliced network, in LTE it is $>$2 ms and in legacy network it is 5 ms. These thresholds must be taken care during session set-up to ensure the QoE of the service.

5. When the services are handed over to the other IP network, it is mandatory to use IPsec in Gm interface for ensuring the security.

6. An adaptive threshold mechanism for checking signal strength should be used for making handover decision for maintaining the quality of service.

7. In case of call drop due to network error, and if the service is not restored within X time, penalty can be applied. Penalty will be decided according to the severity of the problem.

8. If the service operation behaves well for a sufficiently long period without failure, then monitoring frequency will be dynamically reduced to save resources.

9. If KPI of the service is not met, due to the unavailability of infrastructure components, penalty can be applied.

10. Maintaining HD quality for voice call and video call has priority in every composed service which includes voice or video call service. If user preferences over voice call or video call have been registered, this priority will change accordingly.

5.5 Service interaction challenges

This section focuses on the possible service interaction challenges that can happen during the execution of the scenarios. Given the policy-based management approach of this thesis, it is very important to design the policy engine so as to reduce policy conflicts as much as possible. Nevertheless, some of the identified policy conflict situations will be elaborated in this section and possible interventions in such conflict situations are also put forward. These possible interventions can be further studied in depth as a future work of this thesis and is not tested or evaluated in depth beyond what is presented in the following paragraphs.

After defining and registering ECA policy rules, the pre-defined rules in the policy will be executed whenever a corresponding event is activated. However, due to the heterogeneous nature of network and services, policy rules can conflict with each other. Therefore, a key challenge in a policy-based management system is to detect and eliminate the possibility of such policy conflicts. These conflicts occurs when one or a group of event-condition pair triggers a conflicting policy rule/actions and that situation leads the system to a state in which it cannot decide which action should be executed.

5.5.1 Examples of possible policy conflicts

Example 1: User preferred events

Consider the following sequence of events in Scenario 2:

1. During an ongoing video session with screen sharing enabled, Alice has turned ON the video service
2. Eventually due to the bandwidth scarcity, policy rule has been triggered and video was turned OFF by the network.
3. Later a post-event happened leading to a bandwidth improvement in another cell
4. However, in accordance with the policy rule HO request was already initiated while at the same time a user preference over video service event was also registered, meaning video was turned off by Alice.

As seen, even though the session had video preference, the policy management could not guarantee the QoS during the session and failed to apply improvement in QoS[4] before Alice turned off the video. This is a clear conflict where user preferred events

[4] QoS: Quality of Service

did not get the priority and it affected the QoE[5] of the service. In this case if the network could have predicted the post events and notified the user about bandwidth improvement earlier or may be within a particular time frame, this conflict could have been avoided.

Example 2: Additional SLAs

The proposed SLA[6] in this thesis is applicable only for composed services under the telephony use case, however there can also be separate SLAs for each service from the same service provider or different operator. Sometimes the policy rules for the composed services can be in conflict with the policy rules for the individual services. Consider the same Scenario 2, suppose there are two additional SLAs;

1. For voice call quality saying that no matter where you are in the network you are guaranteed a HD quality voice, and;
2. Another SLA saying that wherever you are in the network you are guaranteed that you can enable real time screen sharing anytime.

Thus in our scenario, with the ongoing session, an event had registered that as per the SLA for composed services and policy rules the bandwidth was deteriorating. Thus the video and audio gets turned off, and the bandwidth gets reassigned solely for screen sharing. This situation can also lead to a conflict as QoS guaranteed as per SLA for voice call cannot be delivered. This is because the service in accordance with composed service SLA is only allowing screen sharing and no audio or video.

These two aforementioned situations are possible occasions where conflicts can arise in policy management. Mitigation of such conflicts calls for a highly automated and predictive network environment. This should be investigated in depth as a future work of this thesis.

5.6 Implementation of Policy Engine

This section discusses how the implementation of the policy engine presented above. A key point in service orchestration or any other network management is to efficiently sense any change in the state of network behavior or any change in user preferences for taking necessary action at right time. It is also important to carefully identify and configure the network parameters accordingly to ensure the QoE. Further, given the increasing relevance of APIs[7] and GUI[8], user preferences over any services

[5] QoE: Quality of Experience
[6] SLA: Service Level Agreement
[7] API: Application Programming Interface
[8] GUI: Graphical User Interface

demands more importance than ever before. So, the event categories I have chosen for this policy engine have a strong potential and can eventually contribute to further development of policy-based management of the 5G network.

5.6.1 Policy management proposed by 3GPP

To set the relevant context for the reader a brief presentation of the State of the art of policy management framework as proposed by 3GPP is discussed here. This policy and charging control framework has policy controls for [3GP20c]:

- Session management: for the ongoing session management of a service flow
- Access and mobility management: for necessary policy interventions during access and mobility management of a service flow

This framework is also used to explain the implementation of the policy engine proposed in this thesis. Further, 3GPP has categorized this framework into two based on how it is represented in the network, namely:

1. Service-based representation framework
2. Reference point representation framework

The service-based representation of policy framework has been adopted for this thesis. It enables the network functions to share their services with other authorized network functions through service interfaces. The reference architecture of policy and charging control framework (PCF) proposed by 3GPP is shown in Figure 5.1 [3GP20c].

PCF can be considered as the main network function that has responsibilities of policy management in the network. The framework provides authorized QoS for each service data flow and also the session management for services by applying appropriate policy and charging control for the service data flow. Inside PCF, the policy rules are stored in a database and the management of policy rules is carried out via a Policy and Charging Control rule (PCC). PCC helps the SMF to enforce the authorized QoS control for each specific service flow. PCC is comprised of two rules:

1. Dynamic rules: rules are created on-demand and it is provisioned by PCF to SMF
2. Pre-defined rules: are pre-configured on SMF

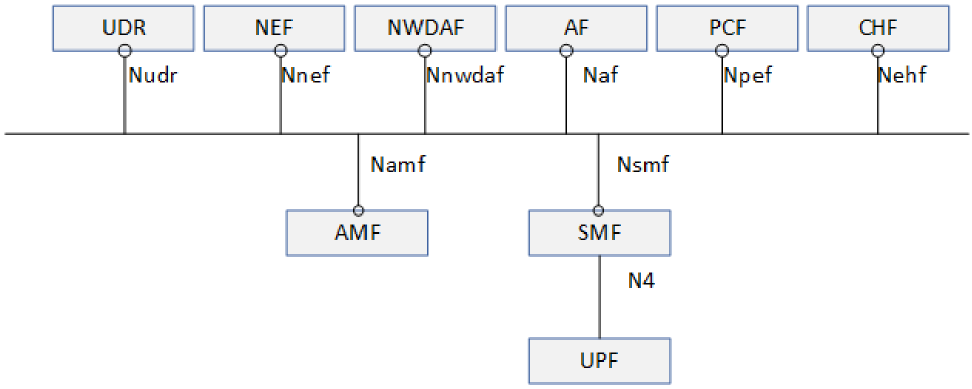

Figure 5.1: An illustration of Policy and charging control framework [3GP20c]

The proposed policy rules for our scenario can be considered as pre-defined rules since our policy rules are only triggered when the events from the SMF/NWDAF/AMF matches with the events in our Policy Engine.

Event handling for triggering policy rule in 5GS is carried out via NWDAF, it is in charge of monitoring usage of of network resources per user and session. NWDAF will provide the network status analytics of specific KPIs subscribed by PCF both for a slice or a user or session-based. PCF shall be able to collect slice specific network status analytic (NSA) information directly from NWDAF by subscribing to the notification on network status analytics of KPIs applicable to that slice.

SMF will notify the PCF when any network monitoring parameter exceeds the threshold as part of session management. PCF interacts with AF via Rx interface for ensuring interconnection with IMS network. This is so that PCF can report to the AF if any service flow no longer gets the guaranteed QoS in the SLA. Moreover, PCF accepts input from SMF, AMF, CHF, NWDAF, UDR and AF to make policy decisions in accordance with the policy rules.

The different steps in policy execution in 5G network can be summarized as:

- During PDU session establishment, PCF discovery and selection is carried out by AMF and SMF and its procedure can be found in TS 23.501 [3GP19].
- The NWDAF will notify the PCF in case of any event has registered which needs a further policy enforcement. The PCF can set the event filter as " telephony slice instance" to get only notified by the events relevant to the subscribed QoS parameters of services under telephony slice.

- The SMF and AMF also interact with PCF in case of violation in any of the configured parameters of the QoS during slice runtime. Policy rules related to session and mobility management are requested from SMF and AMF.
- Before enforcing the policy rule PCF will interact with CHF for getting information about the subscriber spending limitation.

NFV MANO architecture is also responsible for policy management in 5G slicing, it orchestrates the management of network resources according to the policy rule. Moreover, it has the responsibility for the management of policies, which means it will get involved when there is a need to modify the policies, deleting any policies, transferring policies, and activating new policies. During HO to other networks such as 4G or 3G these policy rules and QoS requirements will be exchanged with respective network entities.

5.7 Proposed policy framework

This section summarizes the details about the proposed policy framework for service orchestration of 5G slicing that has been presented so far. As established in this thesis, my research identified that a comprehensive solution for service orchestration for slicing and a clearly stated policy management of service orchestration is missing in the most common standard body of documentation available. For example, the different components that can promote an efficient service orchestration in NFV MANO architecture could not be found in the reviewed literature. Further, PCF[9] by 3GPP also lack information about how to efficiently manage the notification received from NWDAF.

In this section I further the ongoing discussion and propose a policy framework for service orchestration in a way to solve all the challenges this thesis has identified so far. As presented in detail above, the case study in this thesis had taken a slice with expected mobility of users and studied it's inter working with legacy network. Hence, this policy framework will be more suitable for a slice with the similar characteristics (high mobility users). This framework can be either implemented as part of a three layer model of slicing presented in Figure 2.3 or in NFV MANO architecture.

5.7.1 Architecture of the Policy Framework

Figure 5.2 presents the architecture of the proposed policy framework. The major architectural components of the framework are:

[9] PCF - Policy and Charging Framework

- **Service Orchestrator Layer:** As the name indicate this layer is responsible for control and management of all the services under a slice. Further it is composed of three subcomponents such as: SLA management, Policy management and multiple operator management. I have selected these three since I found these have an important role to perform in the mobility management and service assurance as explained in this thesis. Each management component is composed of another four entities and they are shown in Figure 5.3.
  - **SLA/Policy/Multiple operator Database:** This entity is acting as a central repository for storing slice specific SLA/policy/name of supported multiple operators. A policy repository will contain all the policy rules triggered when any SLA violation is met or any network parameter threshold has passed over. Multiple operator repository will contain the name of operator domain which have contract with the NSP to assist when the user is mobile (interacting with legacy networks).
  - **Evaluator:** This entity is responsible for make sure the events registered from monitoring tools are valid by mapping the events with triggering actions in the database.
  - **Monitoring tools/function:** This receives all the real-time monitored performance data from the other two layer.
  - **Decision point:** This entity is responsible for mapping corresponding action (policy rule) from database to the registered events.
- **Slice orchestrator layer:** This layer is responsible for control and management of network slice lifecycle management. Service orchestrator layer needs an interface with the slice orchestrator to allocate sufficient radio resources according to the input from policy decision point.
- **Infrastructure layer:** This layer can be related to as resource layer of network slicing. Upon the instruction from slice orchestrator, this layer allocate necessary VNF to the slice.

5.7.2 Policy Framework Workflow

This section will describe how the workflow of this framework will look based on our scenario. The critical aim of this framework is to monitor if any SLA violation that can affect the QoS and user satisfaction of a service occurs. Further, the following assumptions are made:

1. the telephony slice is created
2. all SLA requirements are translated to the network parameter

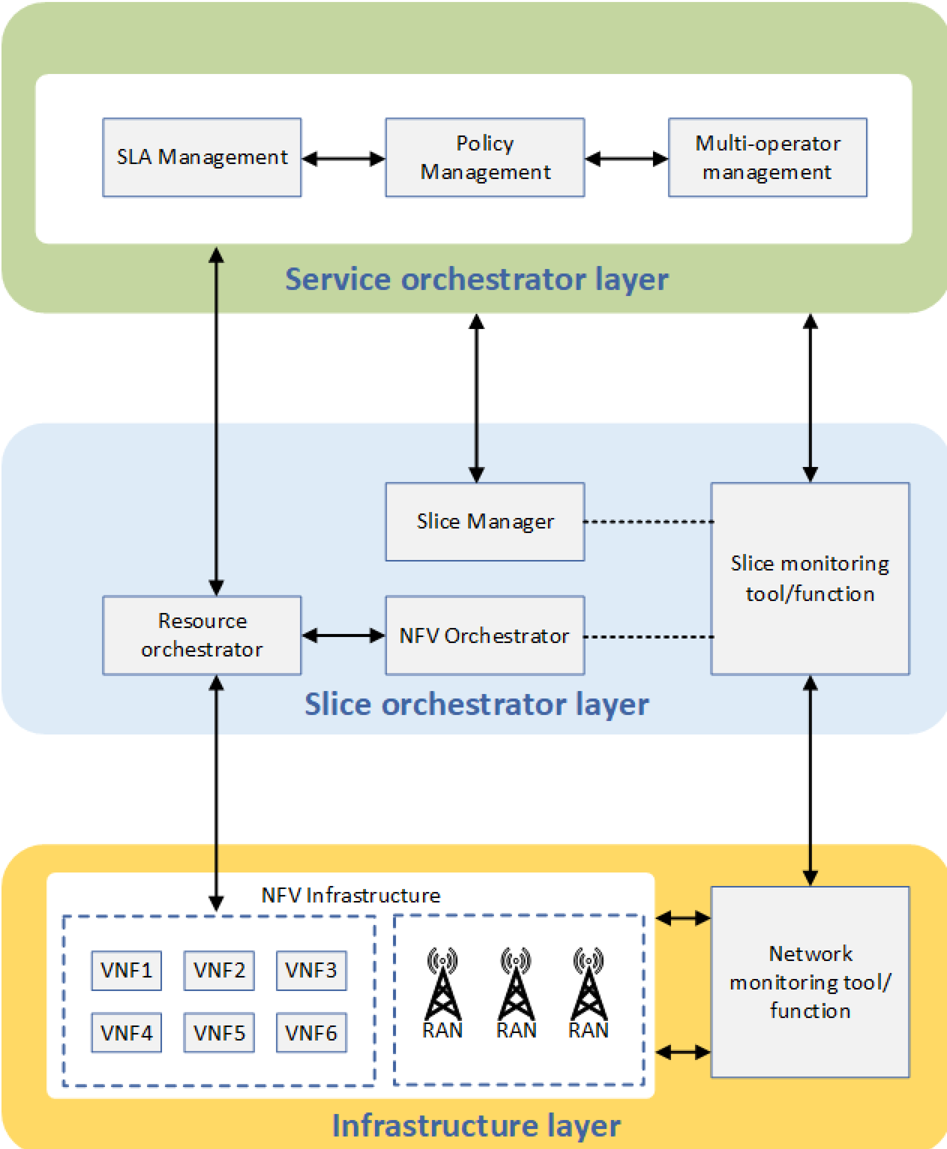

Figure 5.2: Architecture of the Policy Framework

3. virtual or physical resources necessary to support the slice requirements are allocated.

As shown in Figure 5.3 it is the responsibility of SLA/Policy/Multiple operator's management entity to constantly check the validity of the events it receives from the

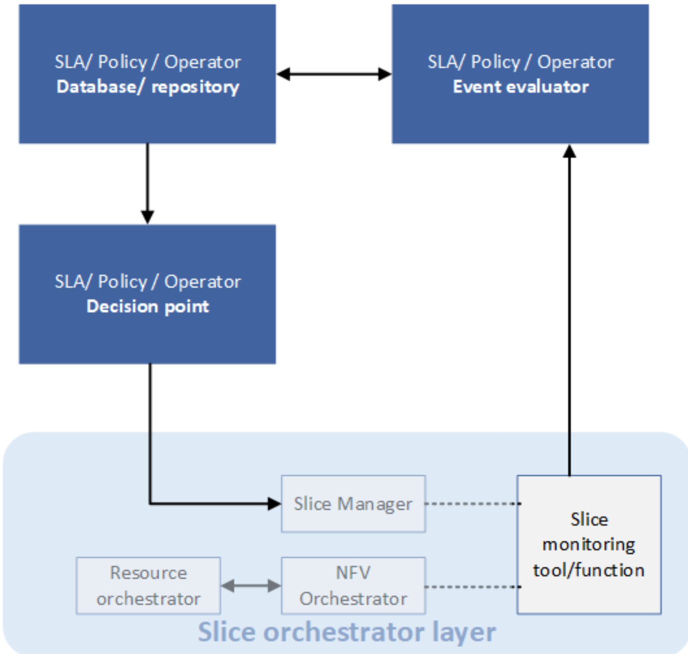

Figure 5.3: An illustration of four entities in each management component

monitoring function of slice orchestration layer. Further, with the help of Evaluator, it then compares these events with events identified from corresponding database and execute the correct policy action. Meanwhile multiple operator management entity will assist in the process of providing service continuity if the service is accessed from outside the boundary of a slice. This entity can assure the authorized access to the other network slices and HO to other network by the support of multiple operator. management Upon the right policy enforcement, scaling down or up of the infrastructure resources will be carried out with the help of slice manager.

This can be illustrated using a real example, lets consider that the bandwidth of a service is deteriorating and the UE is connected to the 5G public network. NWDAF in the Infrastructure layer will notify this to the monitoring function in the slice orchestration layer. Hereafter this event of deteriorating bandwidth will be compared with the threshold specified in the SLA and if it violates any SLA rule this will be forwarded to the policy management entity. Policy management entity will trigger the corresponding policy rule and this process will be assisted with multiple operator management. If the policy rule states the need of an additional resource from other operator domain, this will be decided by multiple operator management.

5.8 Summary of Chapter 5

This chapter summarizes the suitability of using policy based management for service orchestration. It also gives an overview of how the service orchestration challenges identified in Chapter 4 can potentially be solved using policy intervention.

Finally this chapter also proposes a policy framework for orchestrating services involving user mobility.

Chapter 6 Discussion

This thesis had taken a use case-driven approach to solve the research questions that were formulated in Section 1.3. As established through the preceding Chapters of this thesis, we all know that the potential of 5G is very vast and research is still ongoing to unfold more use cases to utilize the capabilities of the 5G network. Proposing a general solution that applies to all the use cases to answer my research questions is very challenging since the field of 5G is still developing and standardization of several things associated with 5G networks has not yet been achieved.

In my specialization project that was carried out as precursor to this thesis, I had considered only two use cases along with several possible services within each use case. Nevertheless, the complexity of analyzing all these associated services was both demanding and challenging due to several reasons identified earlier in this thesis. Hence, this demanded a narrowing down of the initially conceived scope.

The main challenge I faced was the abundance of research material on this topic which often lacked coherence with each other and presented varying approaches to the identified challenges. It is always challenging to have a clear orientation in a dynamically evolving field of research that lacks standardization or established frames of reference for different challenges. Nevertheless, the depth and diversity of research material available has also contributed to bringing in new ideas to the case study research carried out in the thesis.

6.1 Revisiting the research questions

For reasons elaborated earlier in the thesis, it was judicial to take a literature review-based approach to the case study and the empirical findings that followed gives sufficient basis for answering the research questions presented in Section 1.3.

My findings and answers to the formulated research questions are as follows.

RQ1 **How can the request and access to common and customized services across 5G and other networks be realized?**

For reasons outlined in Section 3.1.1, telephony services in the manufacturing industry was chosen as the use-case for this thesis. The process to request and access this telephony service was researched, analyzed and discussed in detail in Chapters 4 and 5.

Requesting and accessing services across 5G and other networks require multiple factors to be in place. Which involves among others, access to slices, inter-operability between slices and legacy networks, inter-operability between different components and streamlined handover processes.

This is explained in detail in the thesis with the help of two hypothetical scenarios (Sections 4.1 and 4.2). A number of possible situations of different HO processes were studied and presented in detail. From the discussion that followed, it was evident that the existing standards and procedures for 5G Network Slicing is not able to streamline the request and access of different services to 5G. Further, these can also pose several practical challenges while being implemented.

Additionally, different procedures for accessing and maintaining the service continuity have been illustrated in Chapter 4. The detailed description on the realization of the handover (HO) process and associated network components are also presented in the same Chapter.

Even though manufacturing industrial application was taken as an example in this thesis, the applicability of the findings presented in preceding chapters can be extended to other industries with similar challenges.

RQ2 **What are the current challenges in this process?**

A number of potential challenges likely to be encountered while requesting and accessing 5G networks were identified in this thesis. These are summarized as follows:

1. **Choosing appropriate HO mechanism:** A frequent handover will be needed for assuring service quality when the services are accessed from 5G network because of the network densification feature in the 5G network. 3GPP has proposed two main handover mechanisms for achieving zero handover failure rate. The dynamic selection of appropriate HO mechanism will be required to ensure desired QoS level specified in the SLA. This selection of appropriate HO mechanism is a challenge.

2. **Slice selection during inter-slice HO:** The condition proposed by 3GPP for inter-slice handover is that both the source and target slice should serve the same service type. However, the research in this thesis has shown that inter-slice handover should also consider other factors in addition to the ones identified by 3GPP, namely, price for the services, isolation, dedicated/shared resources between other slice are also must consider.

3. **Security threat under HO:** HO to the public 5G network or any other slice possesses a serious security threat to the information exchanged between two terminals.

4. **Performance monitoring and real-time notification:** There is a clear need for highly efficient performance monitoring analysis tools and real-time event notification to the SMF[1]. Both these are found to be challenging in the current situation and needs further development.

   Additionally, performance monitoring in legacy networks still use physical components and can pose challenges while interacting with SDN controllers in 5G network. Existing literature on this topic was found to be lacking.

5. **QoS parameter mismatch:** The mismatch between QoS parameters of the different network may lead to the network guaranteeing the same QoS when the services are being accessed from another network. For example, 5G QoS parameter "notification control" is not supported by the 4G QoS parameter. During HO, only QoS parameter supported by the target cell will be transferred.

6. **Service Continuity in Legacy networks:** Based on the literature reviewed some enhanced services such as real-time screen sharing offered by 5G may not be supported by the legacy network. In that case continuity of these services will not be met.

7. **Inter-operabilty between network providers:** In certain use cases, there will be a need for inter-operability between network providers, as in Scenario 2 with a highly mobile user. For the sake of simplicity, this thesis has only considered the NSP[2] by assuming that this use case is happening inside a single domain.

   But in a real scenario, there will be different type of network providers such as network infrastructure operator, mobile virtual network operator and other domain network operators. Literature on how an agreement between appropriate providers will be reached to ensure service assurance was found to be lacking.

8. **Interoperability with legacy networks:** As the 5G technology stands today, a complete roll out of 5G technology across the globe is not realistic in

[1] SMF: Session Management Function
[2] NSP: Network Slice Provider

the near future. Thus it can be assumed that legacy networks will continue to be the primary technology during this period. This also warrants inter-operability with the legacy network for ensuring service continuity. This variation in service quality between 5G and legacy networks can pose challenges to network providers.

9. **Lack of field tests:** The number of field tests on inter-operability between 5G and legacy network is very limited, the literature reviewed in this thesis primarily focuses on PS network. For reasons identified in the preceding point, neglecting legacy networks can pose practical challenges while rolling out 5G at a larger scale.

10. **Identifying and customizing network parameters:** Identifying and customizing specific network parameters to monitor for getting real-time performance data according to each service set is important. Previously in the legacy network, SLA for provisioning of QoS mainly contains the same type of network parameters (KPI[3]), and the need for customizability was less. But now for the enhanced services and the different combination these services clearly demands customized network parameters (KPIs).

11. **Service conflicts:**During the service orchestration of composed services, I have identified the chances of possible service conflicts when the network has to ensure that the services are delivered in accordance with both the SLA for the composed services and the SLA for specific services.

12. **Inter-operability between different network components:** This can be a challenge when considering use cases in which mobility of the user is involved. For example, 5G network contains SDN controllers which brings scalability and flexibility to the 5G network while at the same time the legacy networks are still composed of physical components. Even though an upgrade from physical to virtual components is possible, this can take considerable time leading to mismatch in the service quality.

13. **Vulnerability of network virtualization:** 5G aims to fully automatize network through their the concept of ZSM[4] [V1.20] and this automation is also applicable to the orchestration and management techniques which in turn makes it fully virtualized. Vulnerabilities of SDN and other virtualization techniques in terms of their performance, reliability, security, etc. is a persistent challenge and should addressed through additional research.

14. **Global monitoring for ensuring QoS:** Effective and efficient real-time performance monitoring will need a global view of each network resource.

[3]KPI: Key Performance Indicator
[4]ZSM: Zero-touch-network and Service Management

The status of each network resource and centralized control and management of this are very crucial for maintaining desired QoS of service flow. Service orchestration can only be achieved when we can ensure that the services we are getting has necessary QoS assurance in each layer. Even though this may be possible for 5G, there is no such global monitoring for legacy networks.

RQ3 **What potential measures can mitigate these challenges to ensure the resolution of service conflicts?**

The primary contribution of this thesis is a Policy Engine that proposes a set of policy rules for mitigating the challenges mentioned in reply to Research Question 1 and 2. The research in this thesis work establishes that most of the service orchestration challenges can be solved through policy-based management and most importantly give better performance and ensure high QoS. The discussion also follows that the policy-based management of services must be assisted with techniques such as AI[5] and ML[6]. I have used the ECA[7] policy paradigm to describe my policy rules in Chapter 5.

To support service orchestration for the scenarios considered in this thesis, three policy algorithms have been proposed based on the policy engine presented in Table 5.2, namely:

1. **Mobility management:** In the mobility management policy algorithm, I have considered all possible HO[8] scenarios that can arise during the execution of Scenario 2 and accompanying challenges relating to the inter-slice HO and traditional HO. The research presented in this thesis establishes that inter-slice HO needs much more care since choosing an optimal target slice for HO can be influenced by different factors based on the requirements of the source slice. The algorithm also covers most of the factors that can affect the quality of services that were chosen for the Scenarios presented in this thesis. Further, HO to the public network such as 5G, 4G, and 3G is also covered in this policy algorithm.
2. **User Preference management:** For the user preference management policy algorithm, the focus was on changing the priority assigned to the services after registration of user preference to ensure the best QoE to the user.

[5]AI: Artificial Intelligence
[6]ML: Machine Learning
[7]ECA: Event Condition Action
[8]HO: Handover

3. **QoS management:** For the QoS management policy algorithm, the focus was to provide QoS treatment to the composed service set specified in the SLA at all times. This thesis also identified that QoS and QoE can be largely impacted without adequate policy support.

These algorithms are presented in the AppendixA of this thesis.

The proposed policy rules addresses most of the identified challenges. Nevertheless, some of the identified challenges presented in Chapters 4 and 5 will need additional studies and research. This can be part of the future work based on this thesis. As mentioned earlier, this thesis had considered a static policy approach. However, for further improvement of service orchestration in the 5G network enforcing dynamic policies can also be considered. One way to create dynamic policy rules is to use technologies such as Big Data Analytics and Robotic Process Automation (RPA) for preparing the network in advance to cope with any change in network behavior [ZYR+18].

RQ4 **How can these potential solutions be applied to several services during service orchestration?**

The policy framework for service orchestration proposed in this thesis (Section 5.7.1)can potentially facilitate the application of the above proposed solutions. The framework is composed of three components in the Service Orchestrator Layer (Figure 5.2) that are important in the service orchestration for the discussed scenarios. They are namely, SLA management, policy management and multiple-operator management.

Inter-working of these three components can facilitate the service orchestration during the interaction of several services. This proposed framework can contribute to the faster-growing 5G deployment all over the world through streamlined approach to service orchestration. Even though this thesis had considered only one use-case, the preceding Chapters discusses the different service interactions in the selected use-case in detail. Hence, the findings from this use case can be applied to other use-cases as well.

6.2 Contributions of this thesis

The contributions of this thesis to the field of 5G Service Orchestration are trifold:

1. This thesis studies, analyses and presents the nuances of 5G service orchestration for a telephony slice in a detailed manner. Such an approach is not common in

the identified literature. This relatively novel approach gives the reader and researchers in this topic an holistic picture of slice creation and execution with the help of concrete examples.

2. The second contribution of this thesis is mapping of the policy rules that are needed to support Quality of the Services during a slice execution. The thesis also proposes Policy Based Management as a possible solution to Service Orchestration challenges.

3. Thirdly, the thesis has formulated a policy framework with a set of components that are identified to be crucial to support service orchestration while 5G enhanced services are accessed from other networks such as 4G, 3G and 2G.

6.3 Implications of this thesis

Further, this thesis is primarily targeted at two set of readers, researchers working on 5G service orchestration and industrial actors looking to realize the benefits offered by 5G Network slicing. Hence, it is imperative to outline the implications of the findings of this thesis for these two reader groups. These are:

- **Researchers:** The overall literature discussed in this thesis contributes to the ongoing scientific discourse on the topic. Additionally, this thesis also identifies certain practical challenges that still impede development and testing of use case and network slices within a research setting (Section 6.5). The experiences presented in this thesis can be used as a point of departure while designing testbeds in similar research projects.

- **Industrial actors:** This thesis has mapped the requirements for the telephony slice from the vertical industries and presented it as a template (Tables 3.1 and 3.2). This template can be used as a reference while designing telephony slice for other industries.

6.4 Future work

A few areas for potential future work based on this thesis are identified below.

1. **Policy language that supports the ECA paradigm:** One of the next steps for the continuation of this thesis of course will be to map the policy rules into any policy language that supports the ECA paradigm. One of the possible policy languages that can be considered is Ponder 2. Once the mapping to any policy language is done, a practical implementation of this policy framework

can be tested with the same composed services and with an adequate amount of data to analyze the behavior of the network using any simulation tool.

Such a study can better answer questions on the efficiency, time required to enforce the policy and mapping of the correct policy among others. Results from such a study can also shed more light on possible policy conflicts that were not identified in this thesis due to time and resource limitations.

2. **Field Study:** The enhancement suggested in this thesis is based on a hypothetical situation and potential interaction challenges associtated with the situations. Moreover, the KPI values in SLA can also change while implementing the policy rule. This can also be another area of future work. Analyzing the network behavior with a different threshold can also be interesting.

3. **Investigate different use-cases:** One or two students can also work on the service orchestration challenges between different use cases. A suggestion on that will be to analyze two use cases that come under different service types such as eMBB and URLLC. This makes the scenarios much more complex and needs more time to investigate. So, this work should be carried out with two or three students.

6.5 Practical experimentation and lessons learned

Though the findings presented in this thesis were based on a use-case driven approach, the initial plan was to develop a test-bed and test these use cases in a 5G network lab. However, this practical experimentation was not realized due to a few limitations. These limitations and lessons learned are summarized below for the readers of this thesis interested in taking this research further.

- **Lack of API:** At the time of this thesis work, availability of Application Programming Interface (API) for customizing the scenarios considered in this thesis using the infrastructure of a commercial 5G network operator was limited. This made it difficult to develop a test-bed and collect data through simulations in a 5G network lab. Availability and access to such resources should be ensured by researchers planning to work on these topics.

- **Remote test-bed is time consuming:** Additionally, two commercial testing tools were considered for developing remote test-beds. But initial dialog with one of these tool providers showed that using these tools will be both time consuming and beyond the time scope of such a master thesis.

– **Resource challenges while using Oper Source:** Kamailio IMS[9] and Open5GCore[10] were considered for developing 5G testbeds. However this required either both a physical 5G handset and 5G SIM or a 5G-enabled softphone. While the former was expensive and very early in its roll-out in Fall 2020, the softphone was not yet available. Additionally, finding the right guidelines for using this was also challenging. Discussions in different channels on Discord and guidelines on www.open5gs.org can be an inspiration for a few.

These were some of the major challenges that impeded the practical experimentation in this thesis. Additionally, as mentioned earlier, this research field is still quite nascent and any research is both time and resource intensive. Hence it is advised that students planning to a right a master thesis on this topic set aside additional time for these experiments and explore the possibility to collaborate for addressing the resource challenge.

6.6 Conclusion

5G network will suffer from complex network management challenges due to the assurance of guaranteed QoS for a variety of network services. Further, diversity and heterogeneity of underlying infrastructure will aggravate the situation. One of the major findings of this thesis is that only a fully automated 5G network can be utilized in all capabilities of the 5G network.

This thesis shows that policy-based management will have an important role in the automated network framework of 5G, particularly for orchestrating tailored services. Detection of any events that affect the performance and quality of the services and taking appropriate actions to solve the problem in a very short time demands highly efficient infrastructure components and optimization techniques. These are presented and discussed in this thesis. Additionally, this thesis also concludes the need for more research and field test cases on 5G service orchestration and its inter-operabi lity with legacy networks.

To summarize, this research work aims at contributing to the ongoing 5G research through a set of concrete examples and solutions on 5G service orchestration.

[9] An Open Source IMS core proposed by Fraunhofer FOKUS in 2006. It provides an open source implementation of CSCF and lightweight HSS

[10] An open source standard based implementation of a 5G core network that is suitable for creating a 5G testbed

References

[23.16] 3GPP TS 23.216. Digital cellular telecommunications system (Phase 2+) (GSM); UMTS; LTE; Single Radio Voice Call Continuity (SRVCC) stage 2, (3GPP TS 23.216 version 13.2.0 Release 13), October 2016. URL: https://www.etsi.org/deliver/etsi_ts/123200_123299/123216/13.02.00_60/ts_123216v130200p.pdf.

[3GP15] 3GPP TS 23.228 version 10.9.0 Release 10. Digital cellular telecommunications system (Phase 2); Universal Mobile Telecommunication System (UMTS); LTE; IP Multiumedia Subsystem (IMS) satge 2, July 2015. URL: https://www.etsi.org/deliver/etsi_ts/123200_123299/123228/10.09.00_60/ts_123228v100900p.pdf.

[3GP17] 3GPP TR 38.913 version 14.3.0 Release 14. 5G; Study on scenarios and requirements for next generation access technologies, October 2017. URL: https://www.etsi.org/deliver/etsi_tr/138900_138999/138913/14.03.00_60/tr_138913v140300p.pdf.

[3GP19] 3GPP TS 23.501 version 15.4.0 Release 15. 5G; System Architecture for the 5G System, March 2019. URL: https://www.etsi.org/deliver/etsi_ts/123500_123599/123501/15.04.00_60/ts_123501v150400p.pdf.

[3GP20a] 3GPP TR 21.905 version 16.0.0 Release 16. Digital cellular telecommunications system (Phase 2+) (GSM); Universal Mobile Telecommunications System (UMTS); LTE; 5G; Vocabulary for 3GPP Specifications, November 2020. URL: https://www.etsi.org/deliver/etsi_tr/121900_121999/121905/16.00.00_60/tr_121905v160000p.pdf.

[3GP20b] 3GPP TS 23.216 version 16.4.0 Release 16. Digital cellular telecommunications system (Phase 2+) (GSM); Universal Mobile Telecommunications System (UMTS); LTE; Single Radio Voice Call Continuity (SRVCC); Stage 2, November 2020. URL: https://www.etsi.org/deliver/etsi_ts/123200_123299/123216/16.04.00_60/ts_123216v160400p.pdf.

[3GP20c] 3GPP TS 23.503 version 15.8.0 Release 15. 5G; Policy and charging control framework for the 5G System (5GS); Stage 2, January 2020. URL: https://www.etsi.org/deliver/etsi_ts/123500_123599/123503/15.08.00_60/ts_123503v150800p.pdf.

[3GP20d] 3GPP TS 28.536 version 16.0.0 Release 16. LTE; 5G; Management and orchestration; Management services for communication service assurance; Stage 2 and stage 3, July 2020. URL: https://www.etsi.org/deliver/etsi_ts/128500_128599/128536/16.00.00_60/ts_128536v160000p.pdf.

[3GP21a] 3GPP TS 23.401 version 16.9.0 Release 16. LTE; General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access, January 2021. URL: https://www.etsi.org/deliver/etsi_ts/123400_123499/123401/16.09.00_60/ts_123401v160900p.pdf.

[3GP21b] 3GPP TS 23.502 version 15.12.0 Release 15. 5G; Procedures for the 5G System (5GS), January 2021. URL: https://www.etsi.org/deliver/etsi_ts/123500_123599/123502/15.12.00_60/ts_123502v151200p.pdf.

[3GP21c] 3GPP TS 28.531 version 16.8.0 Release 16. 5G; Management and orchestration; Provisioning, January 2021. URL: https://www.etsi.org/deliver/etsi_ts/128500_128599/128531/16.08.00_60/ts_128531v160800p.pdf.

[3GP21d] 3GPP TS 38.331 version 16.3.1 Release 16. 5G; NR; Radio Resource Control (RRC); Protocol specification, January 2021. URL: https://www.etsi.org/deliver/etsi_ts/138300_138399/138331/16.03.01_60/ts_138331v160301p.pdf.

[AD20] Nadine Akkari and Nikos Dimitriou. Mobility management solutions for 5G networks: Architecture and services. *Computer Networks*, 169:107082, March 2020. URL: https://linkinghub.elsevier.com/retrieve/pii/S1389128619306346, doi:10.1016/j.comnet.2019.107082.

[ALA+19] Muhammad Arif, Olli Liinamaa, Ijaz Ahmad, Ari Pouttu, and Mika Ylianttila. On the Demonstration and Evaluation of Service-Based Slices in 5G Test Network using NFV. In *2019 IEEE Wireless Communications and Networking Conference Workshop (WCNCW)*, page 6, Marrakech, Morocco, 2019. URL: https://ieeexplore.ieee.org/document/8902745CA, doi:10.1109/WCNCW.2019.8902745.

[AM20] Mohamed Abdel Monem. How does 5G do Voice Handover to 4G?, March 2020. URL: https://moniem-tech.com/questions/how-does-5g-do-voice-handover-to-4g/.

[AP16] NGMN Alliance and Hedman Peter. Description of Network Slicing Concept, January 2016. URL: https://www.ngmn.org/wp-content/uploads/160113_NGMN_Network_Slicing_v1_0.pdf.

[ATS+18] Ibrahim Afolabi, Tarik Taleb, Konstantinos Samdanis, Adlen Ksentini, and Hannu Flinck. Network Slicing and Softwarization: A Survey on Principles, Enabling Technologies, and Solutions. *IEEE Communications Surveys & Tutorials*, 20(3):2429–2453, 2018. URL: https://ieeexplore.ieee.org/document/8320765/, doi:10.1109/COMST.2018.2815638.

[BILP20] H. R. Barzegar, N. E. Ioini, V. T. Le, and C. Pahl. Wireless Network Evolution Towards Service Continuity in 5G enabled Mobile Edge Computing. In *2020 Fifth International Conference on Fog and Mobile Edge Computing (FMEC)*, pages 78–85, April 2020. doi:10.1109/FMEC49853.2020.9144852.

[Bou19] M. Boucadair. Framework for Use of ECA (Event Condition Action) in Network Self Management, March 2019. URL: https://tools.ietf.org/id/draft-bwd-netmod-eca-framework-00.html.

[Com21] Ribbon Communications. What is an IP Multimedia Subsystem (IMS)? | Ribbon Communications, 2021. URL: https://ribboncommunications.com/company/get-help/glossary/ip-multimedia-subsystem-ims.

[CPC+18] A. H. Celdrán, M. G. Pérez, F. J. García Clemente, F. Ippoliti, and G. M. Pérez. Policy-based network slicing management for future mobile communications. In *2018 Fifth International Conference on Software Defined Systems (SDS)*, pages 153–159, April 2018. doi:10.1109/SDS.2018.8370437.

[CTML16] G. Carullo, M. Tambasco, M. D. Mauro, and M. Longo. A performance evaluation of WebRTC over LTE. In *2016 12th Annual Conference on Wireless On-demand Network Systems and Services (WONS)*, pages 1–6, January 2016.

[Dat20] Manasa Dattatreya. ThesisReport_manasadattatreya_4749634_replaced.pdf, August 2020. URL: http://resolver.tudelft.nl/uuid:14b46f31-7312-4af6-825f-c148d262f5d3.

[Din10] Jianguo Ding. Policy-Based Network Management, 2010. URL: http://www.ittoday.info/Articles/Policy-Based_Network_Management/Policy-Based_Network_Management.htm.

[Dre] Simon Dredge. What is the 5G Service-Based Architecture (SBA)? URL: https://www.metaswitch.com/knowledge-center/reference/what-is-the-5g-service-based-architecture-sba.

[EG18] ENI 001 V1.1.1 ETSI GR. Experiential Networked Intelligence (ENI); ENI use cases, 2018. URL: https://www.etsi.org/deliver/etsi_gr/ENI/001_099/001/01.01.01_60/gr_ENI001v010101p.pdf.

[Eri15] Ericsson AB. Ericsson Mobile Business Trends, 2015. Technical report, Ericsson AB, November 2015. URL: https://docplayer.net/10429937-Mobile-business-trends-ericsson-mobility-report-appendix.html.

[Eri18] Ericsson. Network slicing – Give a slice of 5G, July 2018. Last Modified: 2021-02-22T09:15:52+00:00. URL: https://www.ericsson.com/en/network-slicing.

[Eri19] Ericsson AB. 5G Business Potential beyond mobile broadband. Technical report, Ericsson AB, 2019. URL: https://www.ericsson.com/4a83c4/assets/content/7339a7861a2547ffaebd676f56904e67/mea-5g-business-potential-beyond-mobile-broadband-screen-v02.pdf.

[Eri20] Ericsson. Voice Over 5G, February 2020. URL: https://the5gzone.com/index.php/voice-over-5g/.

[ETS14] ETSI GS NFV-MAN 001 V1.1.1. Network functions virtualisation (NFV); terminology for main concepts in NFV, December 2014. URL: https://www.etsi.org/deliver/etsi_gs/nfv-man/001_099/001/01.01.01_60/gs_nfv-man001v010101p.pdf.

[ETS15] ETSI GS mWT 002 V1.1.1. millimetre Wave Transmission (mWT); Applications and use cases of millimetre wave transmission, August 2015. URL: https://www.etsi.org/deliver/etsi_gs/mWT/001_099/002/01.01.01_60/gs_mWT002v010101p.pdf.

[ETS18a] ETSI GR NGP 011 V1.1.1. Next Generation Protocols (NGP); E2E Network Slicing Reference Framework and Information Model, September 2018. URL: https://www.etsi.org/deliver/etsi_gr/NGP/001_099/011/01.01.01_60/gr_NGP011v010101p.pdf.

[ETS18b] ETSI TS 128 530 V15.0.0. 5G; Management and orchestration; Concepts, use cases and requirements (3GPP TS 28.530 version 15.0.0 Release 15), October 2018. URL: https://www.etsi.org/deliver/etsi_ts/128500_128599/128530/15.00.00_60/ts_128530v150000p.pdf.

[ETS21] ETSI GS NFV 006 V2.1.1. Network Functions Virtualisation (NFV) Release 2; Management and Orchestration; Architectural Framework Specification, January 2021. URL: https://www.etsi.org/deliver/etsi_gs/NFV/001_099/006/02.01.01_60/gs_NFV006v020101p.pdf.

[Eve18] EventHelix. 5G Service-Based Architecture (SBA), December 2018. URL: https://medium.com/5g-nr/5g-service-based-architecture-sba-47900b0ded0a.

[Gar19] Kumar Garg. You need a robust signaling solution in 5G too!, October 2019. Last Modified: 2020-09-02T16:33:19+00:00. URL: https://www.ericsson.com/en/blog/2019/10/you-need-a-robust-signaling-solution-in-5g-too.

[GCU+18] A. A. Gebremariam, M. Chowdhury, M. Usman, A. Goldsmith, and F. Granelli. SoftSLICE: Policy-Based Dynamic Spectrum Slicing in 5G Cellular Networks. In *2018 IEEE International Conference on Communications (ICC)*, pages 1–6, May 2018. ISSN: 1938-1883. URL: https://ieeexplore.ieee.org/abstract/document/8422148/?casa_token=zg7zILTIccQAAAAA:pAG3Lil79h3UiHErkEJRPqDJpC4pnnIWoL9BQQOz2EmZLxmWcOZ51z9BDDJTccNE70vqnl8, doi:10.1109/ICC.2018.8422148.

[GSM17] GSMA. RCS: A Formidable New Entrant to the Market for B2C Campaigns, November 2017. URL: https://www.gsma.com/futurenetworks/digest/rcs-b2c-campaigns/.

[GSM18a] GSMA. 5G-Network-Slicing-Report-From-Vertical-Industry-Requirements-to-Network-Slice-Characteristics_gsma, August 2018. URL: https://www.gsma.com/futurenetworks/wp-content/uploads/2018/09/5G-Network-Slicing-Report-From-Vertical-Industry-Requirements-to-Network-Slice-Characteristic pdf.

[GSM18b] GSMA. Road to 5G: Introduction and Migration, April 2018. URL: https://www.gsma.com/futurenetworks/wp-content/uploads/2018/04/Road-to-5G-Introduction-and-Migration_FINAL.pdf.

[GSM20] GSM Association. Generic Network Slice Template Version 4.0, November 2020. URL: https://www.gsma.com/newsroom/wp-content/uploads//NG.116-v4.0-2.pdf.

[Han09] Arun Handa. Session Control - an overview | ScienceDirect Topics, 2009. URL: https://www.sciencedirect.com/topics/computer-science/session-control.

[HB15] Akram Hakiri and Pascal Berthou. Leveraging SDN for The 5G Networks: Trends, Prospects and Challenges. In *Software Defined Mobile Networks (SDMN): Beyond LTE Network Architecture*, Mobile & Wireless Communications, pages 61–80. arXiv preprint arXiv:1506.02876, June 2015. URL: https://hal.archives-ouvertes.fr/hal-01161569.

[HHNS18] Mohammad Asif Habibi, Bin Han, Meysam Nasimi, and Hans D. Schotten. The Structure of Service Level Agreement of Slice-based 5G Network. *arXiv:1806.10426 [cs]*, June 2018. arXiv: 1806.10426. URL: http://arxiv.org/abs/1806.10426.

[HHS17] Mohammad Asif Habibi, Bin Han, and Hans D. Schotten. Network Slicing in 5G Mobile Communication Architecture, Profit Modeling, and Challenges. *arXiv:1707.00852 [cs]*, July 2017. arXiv: 1707.00852. URL: http://arxiv.org/abs/1707.00852.

[HL12] Weili Han and Chang Lei. A survey on policy languages in network and security management. *Computer Networks*, 56(1):477–489, January 2012. URL: https://www.sciencedirect.com/science/article/pii/S1389128611003562, `doi:10.1016/j.comnet.2011.09.014`.

[HNW+18] D. T. Hoang, D. Niyato, P. Wang, A. De Domenico, and E. C. Strinati. Optimal Cross Slice Orchestration for 5G Mobile Services. In *2018 IEEE 88th Vehicular Technology Conference (VTC-Fall)*, pages 1–5, August 2018. ISSN: 2577-2465. `doi:10.1109/VTCFall.2018.8690608`.

[HQ20] J. Huang and Y. Qian. A Secure and Efficient Handover Authentication and Key Management Protocol for 5G Networks. *Journal of Communications and Information Networks*, 5(1):40–49, March 2020. Conference Name: Journal of Communications and Information Networks. `doi:10.23919/JCIN.2020.9055109`.

[Hua18] Huawei. vo5g-technical-white-paper-en-v2.pdf, July 2018. URL: https://www-file.huawei.com/-/media/corporate/pdf/white%20paper/2018/vo5g-technical-white-paper-en-v2.pdf.

[HZK+20] E. Hajlaoui, A. Zaier, A. Khlifi, J. Ghodhbane, M. B. Hamed, and L. Sbita. 4G and 5G technologies: A Comparative Study. In *2020 5th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*, pages 1–6, September 2020. ISSN: 2687-878X. `doi:10.1109/ATSIP49331.2020.9231605`.

[IET02] IETF. IETF RFC 3261: "SIP: Session Initiation Protocol", 2002.

[IT16] ITU-T. Mean opinion score (MOS) terminology, February 2016. URL: https://www.itu.int/rec/T-REC-P.800.1/en.

[ITU02] ITU-T. Framework of a service level agreement, June 2002. URL: https://www.itu.int/rec/T-REC-E.860-200206-I/en.

[Kar17] Karel Berkovec. VoIMS – IMS and 5G, October 2017. URL: https://realtimecommunication.wordpress.com/2017/10/31/ims-and-5g/.

[Kar20] Farsana Karadan. Implications of service orchestration in 5G networks. Technical report, Department of Information Security and Communication Technology, NTNU – Norwegian University of Science and Technology, May 2020.

[KFMJM18] Ved P. Kafle, Yusuke Fukushima, Pedro Martinez-Julia, and Takaya Miyazawa. Consideration On Automation of 5G Network Slicing with Machine Learning. In *2018 ITU Kaleidoscope: Machine Learning for a 5G Future (ITU K)*, pages 1–8, Santa Fe, November 2018. IEEE. URL: https://ieeexplore.ieee.org/document/8597639/, doi:10.23919/ITU-WT.2018.8597639.

[KG08] H. Khlifi and J. Grégoire. IMS Application Servers: Roles, Requirements, and Implementation Technologies. *IEEE Internet Computing*, 12(3):40–51, May 2008. Conference Name: IEEE Internet Computing. doi:10.1109/MIC.2008.57.

[LL07] Ling Lin and Antonio Liotta. Presence in the IP Multimedia Subsystem. *Mobile Information Systems*, 3(3-4):187–202, 2007. URL: http://www.hindawi.com/journals/misy/2007/758738/abs/, doi:10.1155/2007/758738.

[LLL07] W. Lee, S. Lee, and K. Lee. Conflict Detection and Resolution method in WS-ECA framework. In *The 9th International Conference on Advanced Communication Technology*, volume 1, pages 786–791, February 2007. ISSN: 1738-9445. doi:10.1109/ICACT.2007.358468.

[MAPS19] Rafael Montero, Fernando Agraz, Albert Pages, and Salvatore Spadaro. Actuation Framework for 5G-Enabled Network Slices with QoE/QoS Guarantees. In *2019 21st International Conference on Transparent Optical Networks (ICTON)*, pages 1–4, Angers, France, July 2019. IEEE. URL: https://ieeexplore.ieee.org/document/8840548/, doi:10.1109/ICTON.2019.8840548.

[MNT19] N. Mouawad, R. Naja, and S. Tohme. Inter-Slice Mobility Management Solution In V2X Environment. In *2019 International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, pages 1–6, October 2019. ISSN: 2160-4894. doi:10.1109/WiMOB.2019.8923342.

[Moh17] Eiman Mohyeldin. Minimum Technical Performance Requirements for IMT-2020 radio interface(s), October 2017. URL: https://www.itu.int/en/ITU-R/study-groups/rsg5/rwp5d/imt-2020/Pages/ws-20171004.aspx.

[NHK20] G. Noh, B. Hui, and I. Kim. High Speed Train Communications in 5G: Design Elements to Mitigate the Impact of Very High Mobility. *IEEE Wireless Communications*, 27(6):98–106, December 2020. Conference Name: IEEE Wireless Communications. doi:10.1109/MWC.001.2000034.

[O'D] Bob O'Donnell. 5G Latency Improvements Are Still Lagging. Section: 5G. URL: https://www.forbes.com/sites/bobodonnell/2020/02/18/5g-latency-improvements-are-still-lagging/.

[OLTR20] Jose Ordonez-Lucena, Christos Tranoris, and Joao Rodrigues. Modeling Network Slice as a Service in a Multi-Vendor 5G Experimentation Ecosystem. In *2020 IEEE International Conference on Communications Workshops (ICC Workshops)*, pages 1–6, Dublin, Ireland, June 2020. IEEE. URL: https://ieeexplore.ieee.org/document/9145225/, doi:10.1109/ICCWorkshops49005.2020.9145225.

[ON20] Ruxandra F. Olimid and Gianfranco Nencioni. 5G Network Slicing: A Security Overview. *IEEE Access*, 8:99999–100009, 2020. URL: https://ieeexplore.ieee.org/document/9099823/, doi:10.1109/ACCESS.2020.2997702.

[PFFOA+20] A. Papageorgiou, A. Fernández-Fernández, L. Ochoa-Aday, M. S. Peláez, and M. Shuaib Siddiqui. SLA Management Procedures in 5G Slicing-based Systems. In *2020 European Conference on Networks and Communications (EuCNC)*, pages 7–11, June 2020. ISSN: 2575-4912. doi:10.1109/EuCNC48522.2020.9200904.

[PPW06] George Papamarkos, Alexandra Poulovassilis, and Peter T. Wood. Event-condition-action rules on RDF metadata in P2P environments. *Computer Networks*, 50(10):1513–1532, July 2006. URL: https://linkinghub.elsevier.com/retrieve/pii/S1389128605003683, doi:10.1016/j.comnet.2005.10.022.

[SBJT20] Muhammad Mohtasim Sajjad, Carlos J. Bernardos, Dhammika Jayalath, and Yu-Chu Tian. Inter-Slice Mobility Management in 5G: Motivations, Standard Principles, Challenges and Research Directions. *arXiv:2003.11343 [cs]*, August 2020. arXiv: 2003.11343. URL: http://arxiv.org/abs/2003.11343.

[SdSLPR+19] Nathan F. Saraiva de Sousa, Danny A. Lachos Perez, Raphael V. Rosa, Mateus A. S. Santos, and Christian Esteve Rothenberg. Network Service Orchestration: A survey. *Computer Communications*, 142-143:69–94, June 2019. URL: https://www.sciencedirect.com/science/article/pii/S0140366418309502, doi:10.1016/j.comcom.2019.04.008.

[Sea19] Sean Kinney. What is dual connectivity and why is it important for 5G?, August 2019. URL: https://www.rcrwireless.com/20190812/5g/dual-connectivity-5g.

[SG14] Roopali Sood and Atul Garg. Digital Society from 1G to 5G: A Comparative Study. *International Journal of Application or Innovation in Engineering & Management (IJAIEM)*, 3(2):186–193, 2014.

[SGJ+18] Matthias Steeg, Nathan J. Gomes, Adrian A. Juarez, Michal Kosciesza, Mason Lange, Yigal Leiba, Hiroshi Mano, Hiroshi Murata, Michal Szczesny, and Andreas

Stohr. Public Field Trial of a Multi-RAT (60 GHz 5G/ LTE/WiFi) Mobile Network. *IEEE Wireless Communications*, 25(5):38–46, October 2018. URL: https://ieeexplore.ieee.org/document/8524892/, doi:10.1109/MWC.2018.1800052.

[SSB18] R. Shrivastava, K. Samdanis, and A. Bakry. On Policy Based RAN Slicing for Emerging 5G TDD Networks. In *2018 IEEE Global Communications Conference (GLOBECOM)*, pages 1–6, December 2018. ISSN: 2576-6813. URL: https://ieeexplore.ieee.org/abstract/document/8647358, doi:10.1109/GLOCOM.2018.8647358.

[SX17] Tony Saboorian and Amanda Xiang. Network Slicing and 3GPP Service and Systems Aspects (SA) Standard - IEEE Software Defined Networks, December 2017. URL: https://sdn.ieee.org/newsletter/december-2017/network-slicing-and-3gpp-service-and-systems-aspects-sa-standard.

[TKS+19] M. Touloupou, E. Kapassa, C. Symvoulidis, P. Stavrianos, and D. Kyriazis. An Integrated SLA Management Framework in a 5G Environment. In *2019 22nd Conference on Innovation in Clouds, Internet and Networks and Workshops (ICIN)*, pages 233–235, February 2019. ISSN: 2472-8144. doi:10.1109/ICIN.2019.8685916.

[Tov18] Thomas Tovinger. Management, Orchestration and Charging for 5G networks, March 2018. URL: https://www.3gpp.org/news-events/1951-sa5_5g.

[V1.20] ETSI GR ZSM 005 V1.1.1. Zero-touch network and Service Management (ZSM); Means of Automation, May 2020.

[V1518] ETSI TS 123 501 V15.2.0. System Architecture for the 5G System, June 2018. URL: https://www.etsi.org/deliver/etsi_ts/123500_123599/123501/15.02.00_60/ts_123501v150200p.pdf.

[VCC+17] D. Verma, S. Calo, S. Chakraborty, E. Bertino, C. Williams, J. Tucker, and B. Rivera. Generative policy model for autonomic management. In *2017 IEEE SmartWorld, Ubiquitous Intelligence Computing, Advanced Trusted Computed, Scalable Computing Communications, Cloud Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, pages 1–6, August 2017. doi:10.1109/UIC-ATC.2017.8397410.

[vR18] 3GPP TS 23.501 version 15.2.0 Release 15. System Architecture for the 5G System, June 2018. URL: https://www.etsi.org/deliver/etsi_ts/123500_123599/123501/15.02.00_60/ts_123501v150200p.pdf.

[XZG+18] Min Xie, Qiong Zhang, Pal Gronsund, Paparao Palacharla, and Andres J. Gonzalez. Joint Monitoring and Analytics for Service Assurance of Network Slicing. In *2018 IEEE Conference on Network Function Virtualization and Software Defined Networks (NFV-SDN)*, pages 1–6, Verona, Italy, November 2018. IEEE. URL: https://ieeexplore.ieee.org/document/8725792/, doi:10.1109/NFV-SDN.2018.8725792.

[You20] Ilsun You, editor. *Information Security Applications: 21st International Conference, WISA 2020, Jeju Island, South Korea, August 26–28, 2020, Revised Selected Papers*, volume 12583 of *Lecture Notes in Computer Science*. Springer International Publishing, Cham, 2020. URL: http://link.springer.com/10.1007/978-3-030-65299-9, doi:10.1007/978-3-030-65299-9.

[ZD05] Simon Znaty and Jean-Louis Dauphin. IP Multimedia Subsystem : Principles and Architecture. *Tutorial of EFORT (Etudes et FORmations en Télécommunications)*, page 11, 2005. URL: https://www.academia.edu/download/43066670/IMS_EFORT.pdf.

[ZG15] K. Zvinys and D. Gursnys. QoS efficiency for individual user in 3G networks. In *2015 Open Conference of Electrical, Electronic and Information Sciences (eStream)*, pages 1–4, April 2015. doi:10.1109/eStream.2015.7119502.

[ZLC+17] Haijun Zhang, Na Liu, Xiaoli Chu, Keping Long, Abdol-Hamid Aghvami, and Victor C. M. Leung. Network Slicing Based 5G and Future Mobile Networks: Mobility, Resource Management, and Challenges. *IEEE Communications Magazine*, 55(8):138–145, August 2017. URL: https://ieeexplore.ieee.org/document/8004168/, doi:10.1109/MCOM.2017.1600940.

[ZYR+18] N. Zhang, P. Yang, J. Ren, D. Chen, L. Yu, and X. Shen. Synergy of Big Data and 5G Wireless Networks: Opportunities, Approaches, and Challenges. *IEEE Wireless Communications*, 25(1):12–18, February 2018. Conference Name: IEEE Wireless Communications. doi:10.1109/MWC.2018.1700193.

Appendix

Appendix A

A.1 Policy algorithm for mobility management

Algorithm A.1 Policy algorithm for mobility management

```
Result: To Provide service continuity
initialization;
 Get signal strength change event;
 if Current_SignalStrength < SignalStrength_Threshold then
    Search for target cell;
    if target cell is a slice then
        if target_slice_ServiceType ≡ source_slice_SliceType then
            if target_slice_QoS_parameters ≡ source_slice_QoS_parameters
             then
                if target_slice_charging ≡ source_slice_charging then
                    if target_slice_Security ≡ source_slice_Security then
                        Trigger HO to target_slice
                    else
                        Check another target cell in 5G public network;
                         Check another target cell in 4G/3G public network;
                    end
                end
            end
        end
    end
    if target cell is in 5G public network then
        if target_cell_QoS_parameters ≡ source_cell_QoS_parameters then
            trigger HO to 5G network
        else
            HO failed
        end
    end
end
```

Algorithm A.2 Policy algorithm for mobility management - Part 2

```
if target cell is in 4G public network then
    if target_cell_QoS_parameters ≡ source_cell_QoS_parameters then
        if DCN_Available then
            trigger HO to DCN
        else
            trigger HO to 4G public network
        end
    end
    if 4G network is not available then
        search for other target cell;
    else
        HO failed;
    end
end
if target cell is in 3G public network then
    if target__cell__QoS__parameters ≡ source__cell__QoS__parameters then
        trigger HO to 3G network & notify the use about un-supported services
    else
        HO failed;
    end
end
```

A.2 Policy algorithm for user preference management

Algorithm A.3 User preference management

```
Result: Trigger policy rules related to user preference
initialization;
 Get notification through GUI of UE;
 if UserVideoOFF||UserAudioOFF||ScreenSharingOFF then
    if QoS of other services are met then
        change priority of the services
    else
        trigger HO to assure desired QoS;
    end
end
```

A.3 Policy algorithm for user preference management

Algorithm A.4 QoS management

Result: Trigger policy rules related to KPI parametrer changes;
initialization;
Get internal event change on KPI parameters;
if $Current\_BandWidthValue < previous\_BandWidthValue$ **then**
| scale up or down radio resources on same cell;
else
| trigger HO to assure desired QoS;
end
if $Current\_packetLoss > Previous\_PacketLoss$ **then**
| Restoration of lost packet will be triggered;
else
| Root cause analysis for packet loss will be triggered
end
if $Current\_LatencyValue > Previous\_LatencyValue$ **then**
| trigger HO to assure desired LatencyValue apply priority if applicable to either user or service or both
else
| Root cause analysis for change in LatencyValue will be triggered;
end
if $Current\_JitterValue > Previous\_JitterValue$ **then**
| trigger HO to assure desired JitterValue;
| apply priority if applicable to either user or service or both;
else
| Root cause analysis for change in JitterValue will be triggered;
end
if $Current\_VideoQuality\_5G < previous_VideoQuality\_5G$ **then**
| trigger dynamic resource allocation;
else
| send notification to UE about service unavailability
end
if $Current\_AudioQuality\_5G < previous\_AudioQuality\_5G$ **then**
| trigger dynamic resource allocation within 5G network;
else
| trigger HO to other network for service assurance;
end
if $Current\_VideoQuality\_3G < Previous_VideoQuality\_3G$ **then**
| trigger dynamic resource allocation within 3G network
else
| trigger HO to other network for service assurance
end
if $Current\_AudioQuality\_3G < Previous\_AudioQuality\_3G$ **then**
| trigger dynamic resource allocation within 3G network;
else
| trigger HO to other network for service assurance
end