2022

Network Slicing in 5G: Admission, Scheduling, and Security

Raneem Jassim Alghawi
rja0007@mix.wvu.edu

Network Slicing in 5G: Admission, Scheduling, and Security

by

Raneem Jassim Alghawi

Thesis submitted to the
College of Engineering and Mineral Resources
at West Virginia University
in partial fulfillment of the requirements
for the degree of

Master of Science
in
Electrical Engineering

Brian Woerner, Ph.D.
Hany Ammar, Ph.D.
Matthew Valenti, Ph.D., Chair

Lane Department of Computer Science and Electrical Engineering

Morgantown, West Virginia
2022

Keywords: 5G , RAN , QoS

Abstract

Network Slicing in 5G:
Admission, Scheduling, and Security

by

Raneem Jassim Alghawi
Master of Science in Electrical Engineering

West Virginia University

Matthew Valenti, Ph.D., Chair

In the past few decades, there was an increase in the number of devices that have wireless capabilities such as phones, televisions, and home appliances. With the high demand for wireless networking, the fifth generation (5G) of mobile networks was designed to support the different services of new applications. In addition, one of the technical issues that 5G would evolve is the increase in traffic and the need to satisfy the user's experience. With the evolution of wireless networking and 5G, Network Slicing has been introduced to accommodate the diverse requirements of the applications. Thus, network slicing is the concept of partitioning the physical network infrastructure into multiple self-contained logical pieces which can be identified as slices. Each slice can be customized to serve and meet different network requirements and characteristics. In terms of security, network security has allowed for new security vulnerabilities such as Distributed Denial of Service (DDoS) and resource exhaustion. However, slices can be isolated to provide better resource isolation. In addition, each slice is considered an end-to-end virtual network, operators would be able to allocate resources to the tenants which are the service providers. The isolated resources are controlled by the tenants; each tenant has control over how to use them to meet the requirements of the clients.

One of the challenges in network slicing is RAN slicing. The target of RAN Slicing is to meet the QoS requirements of different services for each end-user. However, the coexistence of different services is challenging because each service has its requirements. Each slice must estimate its network demands based on the QoS requirements and control the admission to the slice. To solve this issue, we consider the scenario for the enhanced mobile broadband (eMBB) and the ultra-reliable-low-latency communication (URLLC) use cases' coexistence , and we slice the RAN based on the priority of the user application.

Keywords: 5G, QoS, RAN, eMBB, URLLC, DDoS.

Acknowledgements

First and foremost, I would like to acknowledge and give my warmest thanks to my supervisor, Dr. Matthew C. Valenti, who benefited me greatly from the wealth of his knowledge and meticulous experience. His guidance and continuous support carried me through all the stages of my thesis. I would also like to thank my committee members for letting my defense be a memorable and enjoyable moment. Thank you for your brilliant comments and suggestions.

I would also like present my sincerest gratitude to my mother Salwa and my deceased father Jassim, for taking 16-hour-flights to present to me their great role, and countless scarifies for me and my siblings.

Getting through my defense required more than academic support, and I have many. I thankfully acknowledge my family and friends who have been unwavering in their personal and professional support during the time I spent at West Virginia University.

Finally, I must express my very profound gratitude, for letting me through all the difficulties. I have experienced your guidance day by day. This accomplishment would have never been possible without you. Thank you.

Contents

List of Figures

List of Tables

Acronyms

1G First Generation.

2G Second Generation.

3G Third Generation.

3GPP 3rd Generation Partnership Project.

4G Fourth Generation.

5G Fifth Generation.

CN Core Network.

D2D Device-to-Device.

DDoS Distributed Denial-of-Service.

eMBB Enhanced Mobile Broadband.

gNB Next Generation Node B.

IMT-2020 International Mobile Telecommunications-2020.

ioT Internet of Things.

ITU International Telecommunications Union.

KPI Key Performance Indicator.

LTE Long Term Evolution.

M2M Machine-to-Machine.

MAC Medium Access Control.

MANO Management and Orchestration.

MIMO Multiple-Input/Multiple-Output.

mMTC Massive Machine Type Communications.

NFV Network Function Virtualization.

NFVO Network Functions Virtualization Orchestrator.

NR New Radio.

OFDM Orthogonal Frequency Division Multiplexing.

PER Packet Error Rate.

PKI Public Key Infrastructure.

PRR Prioritized Round Robin.

QoS Quality of Service.

RAN Radio Access Network.

RAT Radio Access Technology.

RB Resource Block.

SCS Subcarrier spacing.

SDN Software Defined Networking.

SLA Service Level Agreement.

SLSA Security Level Service Agreement.

TDD Time Division Duplex.

TTI Transmission Time Interval.

UE User Equipment.

UMTS Universal Mobile Telecommunications Service.

URLLC Ultra Reliable Low Latency Communications.

V2X Vehicle-to-Everything.

VLAN Virtual Local Area Network.

VMs Virtual Machines.

VNF Virtual Network Function.

VPN Virtual Private Network.

Chapter 1

Introduction

1.1 Fifth Generation 5G Wireless Technology

Growth in the mobile network and cellular industry is introducing a new generation of technology every 10 years: the 3G Universal Mobile Telecommunication Service (UMTS), the 4G Long-Term Evolution (LTE), and now the 5G New Radio (NR) that was released in 2020. With every new generation, there is an improvement to the services provided in terms of speed, connectivity, and performance. The focus of the technological improvements made in each generation is the Radio Access Network (RAN). As shown in Fig. 1.1, the RAN consists of base stations that cover a specific area and have a joint connection via network interfaces that are connected to the Core Network [1].

5G is an expansion of the 4G technology, to support lower latency, higher throughputs, improved efficiency, and the new broadband services that it enables is called Enhanced Mobile Broadband (eMBB). While some operators identify that the mmWave spectrum is the main improvement of the 5G technology, others think it is the concept of virtualization which would lower the operational costs [1]. 5G is viewed as a flexible technology that is designed to meet the requirements of the diverse 5G use cases. Each continent has different goals and motivations for advancing 5G technology. For instance, operators in Asia are hoping to have a higher throughput for the eMBB service. European operators are looking for Internet of Things (ioT) advancements, and in North America they want to increase the internet connectivity in the suburban areas. Therefore, 5G technology goals depend on the operators and their targets [1].

It is expected that there will be an increase in the growth of network traffic in general and

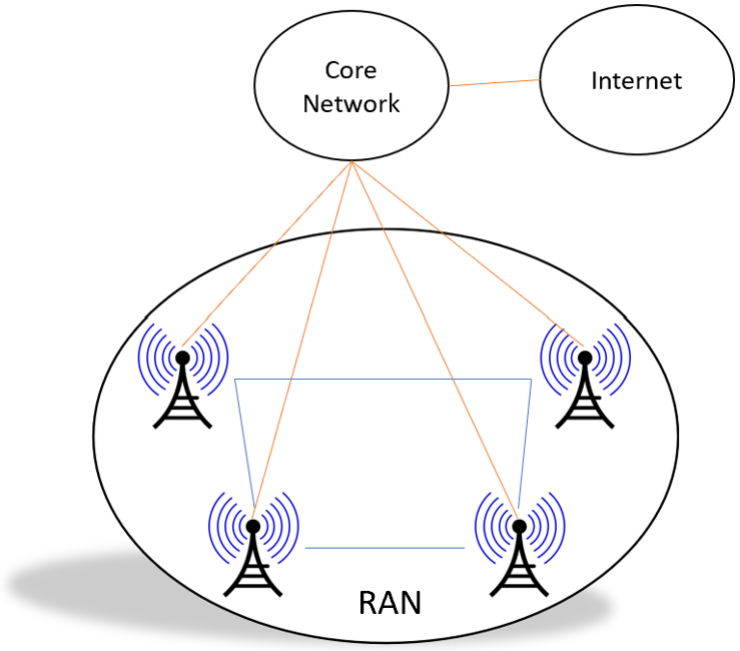

Figure 1.1: Radio Access Network and its relationship to the Core Network.

ioT in particular. The traffic will increase exponentially, therefore there is a high need for the 5G technology, because it will be able to support the millions of devices at high speeds. The RAN must be designed efficiently to be able to serve the traffic [1].

1.2 Use Cases and Requirements

1.2.1 Use Cases

5G is expected to satisfy a range of diverse goals such as having high speed, reliable and low latency communication [2]. 5G is intended to create a new user experience, unlike the previous generations which were focused on improving the radio technology and the spectrum bands. The different 5G system use cases can be classified as follows:

1. **Enhanced Mobile Broadband (eMBB):**

   The eMBB use case supports a steady connection with a high peak data rate, and it also supports cell-edge users to improve upon one of the main limitations of the 4G system. The traffic of eMBB is characterized by large payloads that are stable over long periods of time.

The target of eMBB is to boost the data rate and have a reliable connection with a packet error rate of 10^{-3} [3].

2. **Ultra-Reliable Low Latency Communications (URLLC):**

   The URLLC use case supports small payloads that require high reliability and a very low latency. The transmissions of URLLC are infrequent. Due to the reliability requirement, the transmissions require the flexibility of being either scheduled on a share channel or sent on demand via a random access channel. Scheduling guarantees that there are some available resources, while random access avoids having too many resources left idle due to the infrequent transmissions. Moreover, to satisfy the latency requirements the response to the transmissions has very strict timing constraints and restrictions. Due to the tight timing constraints, reliability cannot be boosted through the reactive retransmission of incorrectly received packets upon the receipt of a negative acknowledgement; i.e., by using protocols such as ARQ or hybrid-ARQ. Instead, the reliability objective is met by proactively sending multiple transmissions using multiple frequency or spatial resources without waiting for a negative acknowledgement [3]. This results in systems that might not be very spectrally efficient, but they have high reliability.

3. **Massive Machine-Type Communications (mMTC):**

   The mMTC use case supports a massive number of devices such as sensors and meters that are low-cost and low-power. The service results in a high device density [4]. The transmissions of mMTC send small payloads from devices that are only occasionally active [3]. The objective of this use case is to target the PER of the transmissions on the order of 10^{-1}. Another objective of mMTC is the ability to run a device from a battery that could go a long time without needing to have its battery replaced, perhaps as long as 10 years for some applications [3].

1.2.2 5G Requirements

Each radio interface, and its associated air interface, has to meet the requirements of International Mobile Telecommunications-2020 (IMT-2020) radio interfaces. As described earlier, the three major use cases in 5G are eMBB, URLLC, mMTC. Each use case imposes different requirements

on the air interface. For instance, eMBB requires a higher throughput with reasonable latency and energy efficiency. URLLC has a strong demand for low latency, low error rates, and high availability. Lastly, mMTC needs a higher connectivity density because of the large number of devices transmitting [1]. Here are some of the major requirements for 5G and some of the potential solutions:

1. **High data rates**

   One of the essential factors for advancing the generations of wireless communication networks is the data rate. There are some services that require high data rates such as video streaming and virtual reality. Video streaming requires 8-15 Mbps, and some gaming applications require about 25 Mbps. Also, in 5G, the data rate at the cell-edge should be increased to 100 Mbps, which is 100 times better than it was in the 4G system. Some of the potential solutions for increasing the data rate are the use of millimeter wave communications, massive Massive multiple-input/multiple-output (MIMO) systems, and software-defined networking [5].

2. **Latency**

   Some applications in 5G require a very low latency such as tactile internet service. This service requires a very low latency communication system and demands a low round-trip latency in the data plane. The latency that is required by this service should be around 1 ms, to match the time scales associated with the sense of touch. Also, the URLLC use case requires low latency because of the real-time interactions. Latency reduction would help improve the user experience, and that could be achieved by having a flexible architecture with the help of the SDN concept [5].

3. **Scalability**

   With the anticipated increase of network traffic, network scalability is one of the major advancements needed in the next generation of wireless communication network technology. High scalability of a network would be able support the high traffic density of IoT services and autonomous vehicles. However, high scalability would in fact require an upgrade to all network layers. A sufficient amount of frequency spectrum resources is needed on the physical layer to support the increasing workload. The network infrastructure should reduce

Capability	Description
Peak data rate	10-20 Gbps
User-experienced data rate	50 Mbps - 100 Mbps
Latency	1 ms
Mobility	500 km/h
Connection density	$10^6/km^2$

Table 1.1: IMT-2020 requirements

the interference by controlling the transmitted power for channels. The media access control (MAC) should also be designed to handle the traffic. However, scheduling has a big impact on minimizing the latency. For the network and transport layers, a better routing algorithm should be arranged for a very large group of devices such as user equipments (UEs). For large-scale mobility, efficient methods are for accomplishing reliable handoffs for the many devices that are expected and their associated mobility patterns. These many scalability requirements can be accomplished using SDN and NFV [5].

4. **Connectivity and reliability**

   Some use cases, such as mMTC, will result in a very high density. The base stations should be able to handle all of the handovers. The network has to deploy handover routing algorithms that would support coverage at the cell-edge area. Handovers make it hard to be connected because of the authentication process with each handover, which would result in delay [5].

The eMBB use case supports connections that require a high peak data rate of 10-20 Gbps. However, 5G is known to provide the fastest connection with user-experienced data rate of 100 Mbps- 1 Gbps. Use cases that require real-time interaction, such as URLLC require an immediate reaction accomplished with a 1 ms latency. As users are expected to change their location within the network, 5G is expected to support users' mobility at speed of up to 500 km/h. The mMTC use case is about connecting a massive number of devices, which would cause a high traffic density that would require the network to enable connections up to 1 million devices per square kilometer. Table 1.1 is a summary of the IMT-2020 [6].

1.3 5G Enabling Technologies

There are several key enabling technologies that have been studied, developed, and proposed for advancing wireless network technology to the fifth generation and beyond [7]. A summary of these key technologies is given below.

1. **Massive MIMO**

   Massive MIMO is a key technology that has been introduced in the 5G technology to handle the growth in the network traffic. The MIMO concept, which involves the use of multiple antennas at the base station and/or the user device, has been deployed for a while, for instance as part of the 4G LTE system. Massive MIMO builds upon MIMO by supplying each base station with a very large array containing many antennas, thereby allowing it to reach a large group of users through spatial multiplexing technologies. This technology typically operates in time-division duplex (TDD) mode, where both the uplink and downlink transmissions use the same frequency band, but transmit and receive at different times. TDD systems are particularly well suited to massive MIMO due to the reciprocity of the uplink and downlink signals, since they are at the same frequency. That reciprocity avoids the need for channel information to be fed back from each user device to the base station [8].

2. **Beamforming**

   Beamforming technology is the concept of having the base station finding the most efficient direction to transmit signals to each user. Like Massive MIMO, beamforming requires the use of large antenna arrays. However, beamforming helps to reduce the interference issue caused by the massive MIMO technology by managing the arrival time of packets to allow users to send data at the same time. It can also help with increasing data rate for the millimeter waves. Due to their short wavelength, they can not serve long-range applications, beamforming would help by sending beams to users [9].

3. **Full Duplex**

   Full duplex technology is the concept of transmitting and receiving in a single time and frequency channel. Full duplex would improve the spectral efficiency of a network by a

factor of two, but can create a self-interference issue by not having the uplink and downlink channels separated in the time or frequency domain [10]. Managing such self-interference typically requires some kind of interference canceling process.

4. **Small Cells**

   Small cells are base stations that use low power and are known to be energy efficient. Due to their lower power, small cells provide coverage to small areas, hence the name. This technology has been introduced to improve the connectivity for cell edge users. However, having a large number of small cells can cause interference issues, which can be solved by having a small cells management scheme [11].

5. **Millimeter Waves**

   Due to their limited bandwidths, frequencies below 6 GHz are no longer able to seriously accommodate the increase in the network traffic due primarily to users with new applications requiring very high data rates. Increasing the frequency for the wireless communication is a viable solution to this issue. Millimeter waves range from 30 GHz to 300 GHz and their wave length can vary from 1 to 10 mm. These frequency ranges have very high bandwidth and are presently not very crowded with other users, and these properties will allow increased capacity for connecting more users [12].

6. **Device-to-Device Communication (D2D)**

   Device-to-device (D2D) technology allows devices to communicate directly without their signals needing to traverse any infrastructure nodes. This technology will support use cases that require low latency due to the shorter and more direct travel path between devices, and it will increase the network capacity by reducing the load on the infrastructure [13].

1.4 Classification of 5G Components

5G enables a significantly improved network connection, greater capacity, higher speed, and lower latency than the previous networks. There are several major components and technologies required to enable 5G networks, as described below:

1. **Core Network:**

   A core network (CN) is an essential component of any wireless network. The main task of the core network is to connect the RAN with the third-party network or the tenant that is providing the end-to-end connection to the client. There are three distinct planes used in core network functionality: service management, session management, and mobility management. In the previous fourth generation (4G) system, the core network was responsible for supplying the data pipe. However, in the fifth-generation (5G) the core network architecture is known to be a service-oriented architecture, which provides the network as a service that is broken down into functions and sub-functions. Functions contain the session management, mobility management, access management and user plane functions [14].

2. **Radio Access Network:**

   The radio access network (RAN) is a sub-component of the network that has been used in previous generations of technology. The radio access network (RAN) is a collection of base stations and antennas that enables wireless communication between devices. The RAN components are responsible for providing network coverage. Radio sites are responsible for radio access and resource management at various radio sites, transmissions to the network occur by RAN, and RAN is responsible for delivering the signal to the endpoint. Another component of the RAN is the RAN controller, which is responsible for providing functionality, radio resource management, and UE messaging through SDN switches [14].

3. **Software-defined Network:**

   Software defined network (SDN) is an essential component in the 5G network architecture. SDN is a way to describe network components and their functionality. It also helps with network management and changing any of the network configurations. The goal of using SDN is to separate the control plane outside the switches and control the data using the SDN controller, and it also helps to deploy new services and applications by enhancing the network [14].

4. **Network Function Virtualization:**

   Network function virtualization (NFV) is about virtualizing all of the network node functions

that consist of virtual machines with different software running on on the top of servers, switches, storage devices and cloud infrastructure. All of the virtualized network functions are connected to deliver the requirements of each use case [14]. NFV technology would enable the concept of network slicing discussed in chapter 2.

5. **Orchestration:**

   Orchestration is the concept of organizing, connecting, managing, and scheduling tasks to deliver an end-to-end services. The Management and Orchestration (MANO) framework provides the orchestration. Service orchestration can help with resource allocation, workflow execution, management of network topology, and lastly configuring and activating. Orchestration plays a big role in the NFV architecture, such that the Network Functions Virtualization Orchestrator (NFVO) is in charge of controlling the network services and creating the end-to-end servies across the Virtual network functions (VNFs) [15].

1.5 Overview and Contributions of Thesis

The purpose of this thesis is to provide a comprehensive look into how network slicing works. A key focus is on security issues related to network slicing and scheduling algorithms. The remainder of the thesis is organized as follows.

In chapter 2, we review the concept of network slicing and its key concepts. Network slicing is considered to be a solution that has been introduced with the evolution of wireless networking to accommodate the diverse requirements of the different applications. Some of the key concepts that are introduced are the network slicing architecture, slice life cycle, and the network slicing types which may occur in the core network or in the radio access network.

Chapter 3 considers how network slicing allows the creation of multiple virtual networks that share the same core infrastructure. Due to enabling the resource sharing, the shared network will be exposed to some security concerns such as Directed Denial of Service (DDOs). We review the security challenges of the network slicing. We also analyze the life cycle of security threats, with a focus on intra-slice and inter-slice security concerns. Lastly, we introduce some mitigation strategies and technical solutions to help guarantee a secure network.

In chapter 4, we review the key concepts of a 5G network, such as the 5G NR frame structure,

numerology and subcarrier spacing, slot length, and resource blocks. Then, we introduce the different types of RAN scheduling that can support the 5G traffic. We present a prioritized round robin scheduling algorithm, which gives priority to Ultra-reliable-low-latency (URLLC) traffic while being fair and efficient with its resource allocation to other use cases in an effort to avoid starvation.

Finally in chapter 5, we review the findings of the thesis, and suggest ideas for future research.

Chapter 2

Network Slicing

2.1 Overview

In the past few decades, there has been an increase in the number of devices that have wireless capabilities such as phones, televisions, and home appliances. With the high demand for wireless networking, the fifth generation (5G) of mobile networks is designed to support the different services of new applications. In addition, one of the technical issues that 5G needs to accommodate is the increase in traffic and the need to satisfy the user's experience. In comparison to the fourth generation (4G), the goals for 5G are significantly faster data rates, reduced latency, and support for an increased number of connected devices, and all of these objectives need to be met. [16].

In addition, in the last few years there has been an increased diversity in the kinds of wireless access networks with new types of communications, distinct access technologies, and services. Over time, 5G will need to cope with the diverse requirements by offering new management and control techniques. Each application has different requirements and characteristics that may include high-speed data rates, low latency, better connectivity, availability, or reliability. However, the network infrastructure would be responsible for handling all the diverse traffic patterns, service requirements, and capabilities of the devices [16].

With the evolution of wireless networking and 5G, network slicing has been introduced to accommodate the diverse requirements of the applications. Network slicing is the concept of partitioning the physical network infrastructure into multiple self-contained logical pieces which can be identified as slices. Each slice can be customized to serve and meet different network requirements and characteristics. In terms of security, slices are isolated to provide better resource isolation.

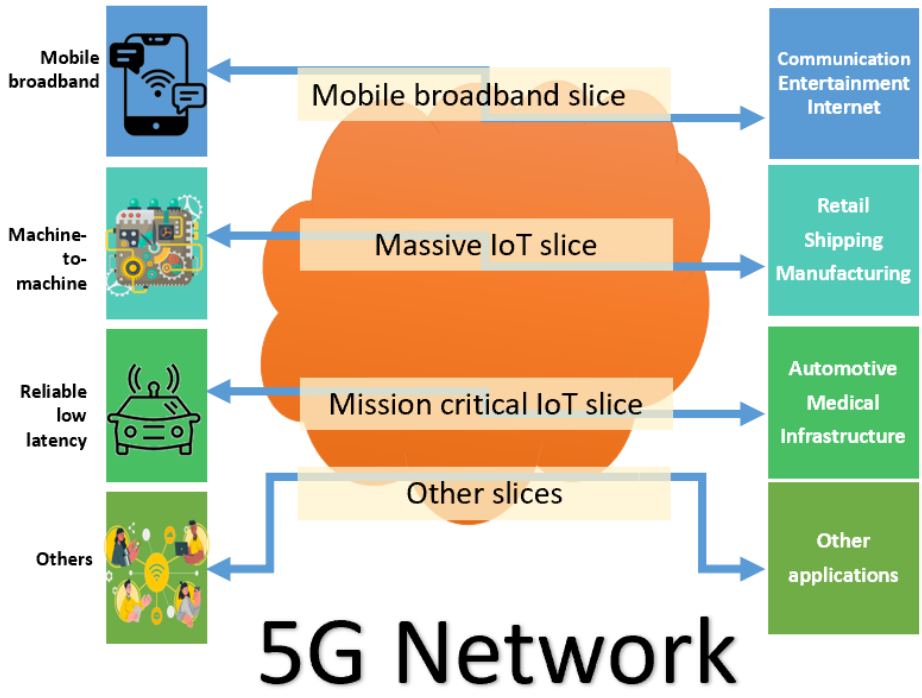

Figure 2.1: Network Slicing concept

In addition, each slice is considered an end-to-end virtual network, and operators can allocate resources to the tenants which are the service providers. The isolated resources are controlled by the tenants, each tenant has control over how to use them to meet the requirements of the clients [16].

Network slicing is enabled by the capabilities of software-defined networking (SDN) and network function virtualization (NFV). It can be made into an end-to-end network by implementing the network slicing from the core through the Radio access network (RAN). Virtualization technologies such as SDN and NFV can virtualize the network functions in the slices to satisfy their characteristics [17].

2.2 Network Slicing Architecture

Network slicing allows the deployment of virtualization technology by using both concepts of SDN and NFV in order to deploy the communication system of the physical infrastructure to grant diverse 5G services to be delivered using a common infrastructure. Operators are allowed to integrate the slicing infrastructure adeptly and manage the resources of the network to handle the

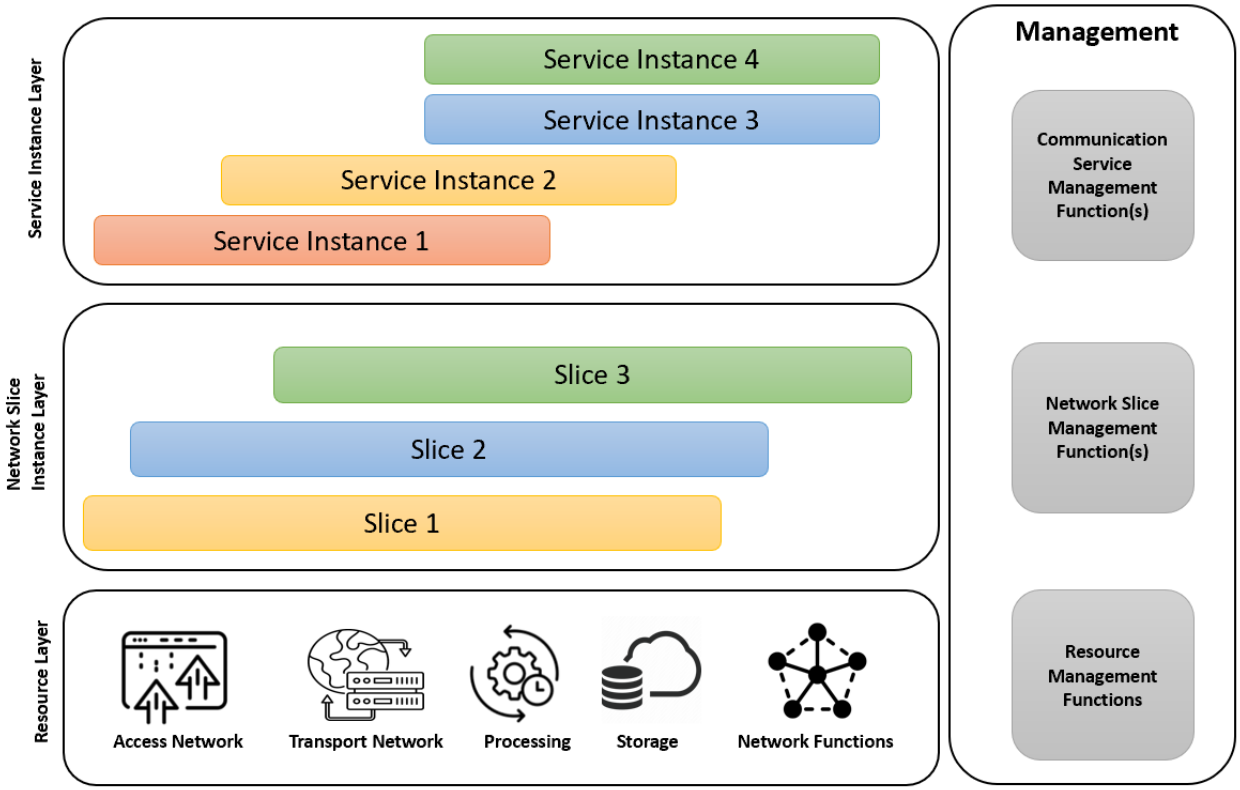

Figure 2.2: Overall Architecture in Network Slicing

increase in network traffic effectively. However, the network slice instance is a group of resources from the virtualized infrastructure platform that provides virtual resources such as computing, storage, and computing. Those virtual resources are organized to form a network slice [18]. A network slice instance is an independent virtualized instance and is made by grouping the available network resources and is considered to be an end-to-end logical network that can be designed to serve a specific set of requirements for a use case, such as voice communication, vehicular communication, and video streaming. The slices are differentiated based on the functionality expected and the performance required. However, Satisfying a large set of QoS requirements for diverse use cases is considered to be difficult [18].

A logical network can be defined as a set of network function instances on top of the physical and virtual resources. Each slice instance is considered to be a logical network. The RAN and the Core Network (CN) are the network domains that make up the 5G network architecture [19] [20].

As shown in Fig. 2.2, the network slicing architecture consists of three layers.

1. **Resource Layer**

   The bottom layer of the network slicing architecture, the resource layer is made up of network functions and resources to serve the end-user. Some examples of network functions are slice selection, switching, and routing functions. However, storage, processing, and the transmission of nodes are network resource examples. Both the resources and functions can serve one or more slices if requested [19].

2. **Network Slice Instance Layer**

   The network slice instance layer consists of slices, each slice is designed to satisfy the network capabilities that were requested by the service instance. A slice can serve one or several service instances which are the services provided to the customers and can either run across or openly over another slice or network resources respectively. In addition, any two separate slices may not integrate on the same physical architecture [19].

3. **Service Instance Layer**

   The service instance layer is made up of service instances that are ready and available for clients. Each of the resource managing functions is linked to the core network functions and resources, and both can be assigned distinct management domains. However, the life cycle of a slice is managed by the network slice management function. The network slice management function oversees the slices entire process and integrates with the rest of the administration functions. If a slice is made up of sub-slices, they each have their own management function. The slice manager communicates with the network services management function, which handles the entire service process [19].

2.3 Slice Life Cycle

The life cycle of a slice consists of four phases as shown in Fig. 2.3 and described below:

1. **Slice Commissioning**

   During the commissioning phase, the slice does not exist yet, and the environment of the network is getting prepared by identifying the services that an end-user is expecting by

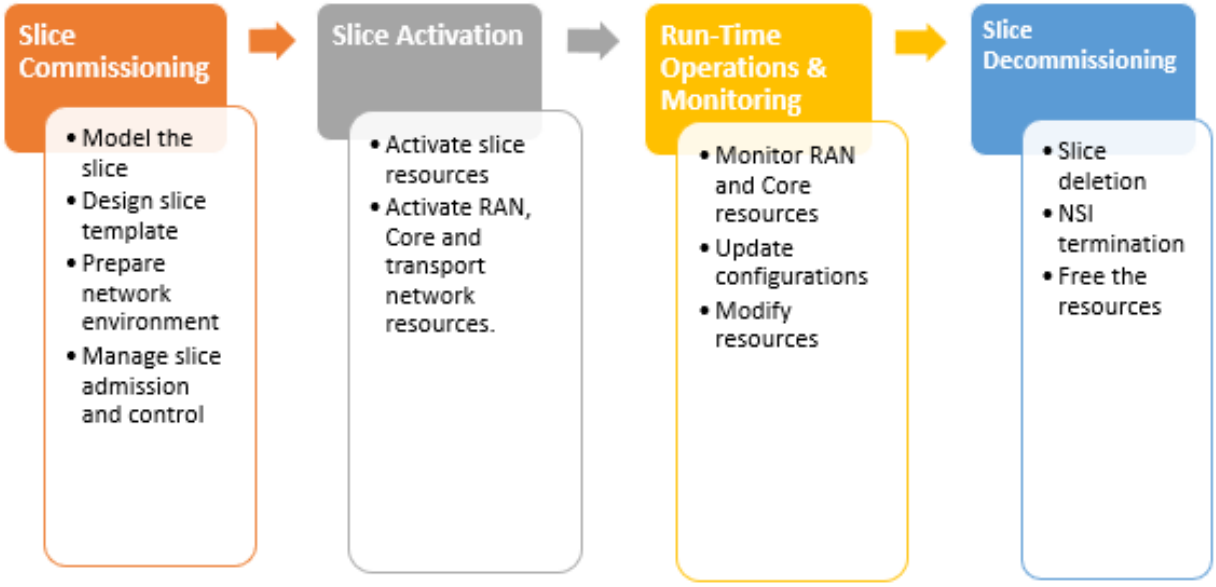

Figure 2.3: Life Cycle of a Network Slice

gathering the information from each client in order to create the slice template. A slice template has information about each component, structure, and configuration of a slice [19].

2. **Slice Activation**

   In the slice activation phase, all of the network resources and functions are installed and set up. Slices are created using the slice template that has been built in the first stage. After the slices are created, they are ready to be activated [19].

3. **Run Time Operations and Monitoring**

   In the run-time operations and monitoring phase, the slices are monitored and reporting is performed. Slice now are operating and providing the requested services from the client. However, slices can be modified, for instance, by changing some configurations and links to both the network functions and resources [19].

4. **Slice Decommissioning**

   In the slice decommissioning phase, which occurs during the last stage of the life cycle, all slices get decommissioned. All of the network functions and resources are freed and slices do

not exist any longer [19].

2.4 Network Slicing Types

Each network slice instance consists of network resources and functions, which are configured to build a complete network to provide the demands of the service instance. In brief, each network slice is considered to be a separate part of the RAN, transport, and CN. Using the NFV technology, each slice can be set up to be efficient and flexible. Using that technology, network components can be virtualized across the CN and the RAN. Slicing can happen in the Core Network or in the Radio Access Network [21].

2.4.1 Network Slicing in CN

Core Networks serve various purposes as single network architecture. However, using virtualization technology has helped to design networks that are more efficient and flexible. Having a flexible system has made it easier for network designers to customize solutions for clients. Network slicing in the core networks allows the separation of each core network. Slices can run on the same infrastructure or a different one based on the operator, that way we do not have a single physical network layer. In order to have an efficient system, each network slice would have an access to the infrastructure, cloud services, and the VNFs. Using the SDN technology, we would be able to separate the control plane from the user plan which would improve the efficiency of 5G core networks [17].

Separation of the planes has many effects on the 5G core- network architecture. Such separation helps the migration to cloud-based deployments, moves the user plane closer to the user device, and allows for the functionality of the user plane to be matched to the needs of the various slices. With the diverse use cases, and as each of them needs different connectivity requirements, user-plane functionality can be customized to meet the most cost-efficient scheme for each use case. For instance, the mMTC use case which results in a low device density and low mobility, has low connectivity requirements compared to the eMBB use case, which results in a high payload volume. eMBB can have various services, such as web browsing and gaming. For each sub-service of the use case, the requirements of that service can be implemented within the network slice. One of

the benefits of plane separation is that planes may be placed in different locations. For instance, having the control plane located in the center makes the execution more efficient. While the user plane can be allocated over the sites that are closer to the user, it would help decrease the round trip between the user and the service requested from the network [17].

2.4.2 Network Slicing in RAN

There are two important prospects in the concept of network slicing in the RAN; The radio access technology (RAT) that actually supports the network service that is provided by the slice, and the structure of RAN resources to support the network slice. However, 5G is expected to meet certain Key Performance Indicators (KPIs) targets. The fourth generation (4G) was mainly focused on supporting eMBB services. However, in later releases, 4G improved to support various use cases such as machine-type communications (MTC) and direct device-to-device (D2D) communications [17]. Furthermore, RAN slicing is the concept of aligning a slice-ID to a set of configuration rules for the RAN. The planes do not get separated, but the RAN configuration rules get assigned for each slice to meet the requirements of the network services [17].
Some of the requirements of designing a RAN Slicing are:

1. Each slice must be configured by applying the rules to the RAN control and user plane functions to be supported in the RAN.

2. Functionality of slices may be similar to several slices.

3. RAN resource usage is controlled by the common control functions.

While addressing some of the configuration aspects in RAN slicing:

1. **Resource management:**

   Slices that are supported in the RAN may share the same radio resources (time, frequency, space) and the communication hardware (analog radio components, digital baseband processing components) based on the configuration rules applied to the network slice. Each slice attains the resources needed based on their request and priority. Slices get their requested resources by scheduling or contention. By scheduling, each slice proposes a request to the scheduler that can be in the base station or the central RAN controller. The scheduler's

responsibility is to allocate radio resources to each slice based on their request. On the other hand, it can happen by the contention, that a set of pre-defined configuration rules are allocated to the slices [17].

2. **Slice-specific admission control:**

   Admission control is essential to manage the access to the network slices. slices are configured based on their demands. For instance, a URLLC use case must get a guaranteed low latency with high-reliability access. With the admission control, a UE sends a request to the scheduler to get admission but they may not get access if the slice is not activated in the access point [17].

3. **UE awareness on the RAN configurations:**

   When UEs get admitted to a network slice, the UE is unaware of the RAN configurations that are applied to the slice. Therefore, the UE needs to receive the RAN configuration before getting admitted to the slice. Due to some services requiring advanced RAN configurations [17].

Chapter 3

Security and Challenges in Network Slicing

3.1 Overview

Unlike the previous generation networks, 5G supports network slicing within its architecture. Network slicing is considered to be the distinguishing key in 5G that would help to achieve the enhanced capacity to serve more clients. With the existence of different use cases, 5G technology satisfies these diverse requirements by using the concept of network slicing. A slice is a logical network that can be optimized to meet a wide range of heterogeneous requirements [22].

Network slicing provides a fragmented architecture where multiple logical networks are embedded within a shared infrastructure. Each logical network has its own logical topology, security structure, performance requirements, network functions, network resources, and end-to-end connections. That being said, each logical network performs its designated tasks because each segment of the network is designed by the network characteristics to have the capability of delivering a service to an end-user. For instance, autonomous vehicles require low latency and high reliability. On the other hand, HD streaming requires high bandwidth [22].

By partitioning the network, each virtual network will be designed for specific purposes, such as communication, storage, and reliability. The ability to efficiently use network resources to provide different levels of service is a major advantage of this technology. The platform is scalable with service requirements according to the design load, ensuring greater reliability [19].

Businesses are innovating their business models more easily now and that is due to the differ-

entiated architecture. From a business view, this has helped to diversify revenue opportunities and improve communication channels. For instance, the automated operation processes have made it easier to order products, process payments, and deliver items. However, these types of processes can be performed now easier and quicker, with high security to allow service providers to serve the clients more efficiently Businesses are innovating their business models more easily now and that is due to the differentiated architecture. From a business view, this has helped to diversify revenue opportunities and improve communication channels. For instance, the automated operation processes have made it easier to order products, process payments, and deliver items. However, these types of processes can be performed now easier and quicker, with high security to allow service providers to serve the clients more efficiently [19].

The concept of slicing in Ethernet networks can be traced back to Virtual Local Area Networks (VLANs), which 5G network slicing borrows. Network slicing particularly relies on Software Defined Networking (SDN), and Network Function Virtualization (NFV) [22]. NFV allows the creation of network slices through interconnected virtual machines (VMs). These virtual machines are coordinated through SDN orchestration, which flexibly configures network slices and reserves resources for the heterogeneity that end users may use. SDN is used to control encryption within each network slice. This helps to ensure security and privacy for data passing between networks [19].

3.2 Challenges in RAN Slicing

As discussed in the second chapter, one of the main targets in RAN slicing is to satisfy the QoS requirements of the diverse requested services. However, designing a RAN slicing system that would support the various use cases is complicated, the issue can be illustrated as follows according to [21] and [23]:

As specified by the 3GPP, the three major use cases in 5G can be categorized into URLLC, mMTC, and eMBB. Each of the use cases demands a different data rate, reliability, and latency from the network. For example, eMBB requires a high data rate and can endure high latency and low reliability. In contrast, the URLLC use case focuses on low latency and very high reliability. For example, autonomous driving cars use cases need an immediate reaction and high reliability

to avoid car accidents. However, for the mMTC use case, it needs high connectivity since we have a massive number of requests. Hence, satisfying each use case is hard and the framework of RAN slicing needs to be designed reasonably to satisfy the diverse QoS requirements for the use cases. Additionally, according to [23] the resource allocation for a large number of slices can cause many difficulties. Each slice must satisfy the QoS targets and the fairness between each slice must be guaranteed.

While there are also some challenges in terms of the user's mobility, each user can change their location due to their life and workstyle, which affects the service traffic among the cells. In addition, the dynamics of a cellular network may have two patterns, a long-term and a short-term pattern. An example of a long-term pattern is service traffic. In contrast, the dynamics of a wireless channel are considered a short-term pattern. The pattern for the traffic of service over a period of the network frame will differ from time to time. Control over the RAN slicing multiple times and spatial is true.

Interference is considered to be an issue in the wireless network environment. Slices will be completely isolated, considering a single-cell scenario. On the other hand, the spectrum and frequency will be reused in the multi-cell scenario which will cause interference concerns. Radio resources should be multiplexed for a better RAN slicing operation.

In addition, the network controller can not have the competence to enhance the QoS overall performance service for each use case. For example, the Vehicle-to-Everything (V2X) use case request, which requires low latency and high reliability. Its service requirements are hard to achieve because of the costed latency. Mobile data will be needed by the network controller to be able to satisfy the users by signaling overhead. So, a control at different times over the network is needed to avoid signaling overhead which would increase the cost of advancing the RAN slicing control.

Moreover, each service has its associated Service Level Agreement (SLA). However, SLA defines the QoS requirements for each slice and use case. For example, the bandwidth, the throughput, and the latency that the network must deliver to each service. The SLA has to be monitored in the case that the QoS targets can not be met for a specific service, to inform the customer so they can readjust their security measurements. Monitoring can also help to alert if there are any flows with the functionality. Hence, monitoring can be done individually for each user associated with a different slice. Additionally, virtual network monitoring requires monitoring of the physical and

virtual resources. After each SLA monitoring, the key performance indicators (KPI) have to be reported with every user and slice details. For instance, throughput, reliability, delays, efficiency, and lastly the slice load.

The network in the slicing concept accepts third parties such as tenants that act as the virtual network operator. However, this can cause some challenges depending on the level of control given to the tenant.

In the lowest level of control, the operator has full control over the SLA, it's job is to report the slice and use case operation and is responsible for managing the network. However, the third party can have an access to monitor the KPI.

In the case of giving the third party full control over the network slice, the operator would just provide the infrastructure to the third party, which operates the slice. In this case, if the slices were designed and pre-configured by the operator, or it's the third party's job to design a slice by the slice templates that were created by the operator or the tenant can design a new slice with specific functionality.

Lastly, if the third party has partial control over the network slice. The third-party can partially control some functionalities of the network and maybe change some configurations as well. For example, if the tenant wants to expand the coverage for a specific area. Some configurations can be accepted as long as it does not affect the isolation of the slice. However, this could raise some difficulties in the global performance of the network.

3.3 Security Challenges in Network Slicing

Network slicing by using a fragmented architecture with different partitions for facilitative infrastructure and resources opens up 5G technology to a variety of security vulnerabilities. One of the biggest security challenges for network slicing in 5G is Distributed Denial of Service (DDoS) attacks [24]. These attacks involve targeting services with the goal of overloading them with a large amount of traffic till they are inaccessible to other end-users.

DDoS attacks can also take the form of depriving a target of resources it shares with other hosts. The misappropriation of necessary features in 5G technology, such as overload control metrics, could facilitate these types of attacks. Overload control metrics in 5G design, such as

the 3rd Generation Partnership Project (3GPP), are used to limit communication across networks under the following conditions: Overload within a network. However, at this time, this feature does not incorporate checks to ensure that overload header indicators are placed by the intended users. Within the virtual networking infrastructure offered by network slicing, this lack of control opens up vulnerabilities for users. This overload control indicator may be misused by others, which could lead to a denial of service [24].

Moreover, network slicing has another security concern which is sharing and depletion of resources. Network slicing allows dynamic resource sharing among slice tenants, in a way that allows for better resource efficiency. Managing slices is considered to be a challenge, allocating resources to each slice to be capable of delivering the requested services and to ensure that each slice is maximally optimized. Controlling the slices is also a challenge, since the network operator has the maximum control, there is sufficient incentive to allow partial permission. In order to fully take advantage of 5G networking, it is necessary to address the challenge of sharing resources and allocation among different devices. Further sharing of resources between network slices creates security vulnerabilities. With a shared core architecture, network slices are designed to have different security protocols in order to optimize performance. However, this creates vulnerabilities if other slices on the network share the same security protocol [25].

Security protocols that can be adjusted on a slice-by-slice basis can be easily exploited to harm other slices. Therefore, it is important to take into account security protocols among the multiplicity of slices when dealing with concerns about resource sharing. The proliferation of security challenges posed by network slicing in 5G technology creates new challenges for computer technology. With previous wireless networks, such concerns tended to be more limited in scope. However, with resource sharing becoming an essential part of network slicing, such security breaches now have a wider capacity and therefore a greater impact. For example, DDoS attacks targeting individual slices now present a significant risk of starvation of host resources, as slices are tenants in a larger shared virtualization infrastructure [24]. However, resource sharing in 5G technology is an important fundamental feature that explains many of the benefits of 5G technology.

Infrastructure sharing has many advantages in cost and resource consumption. However, sharing resources causes a challenge in securing 5G networks and mitigating risks associated with a network slice architecture. However, resource sharing and network slicing are essential in 5G technology

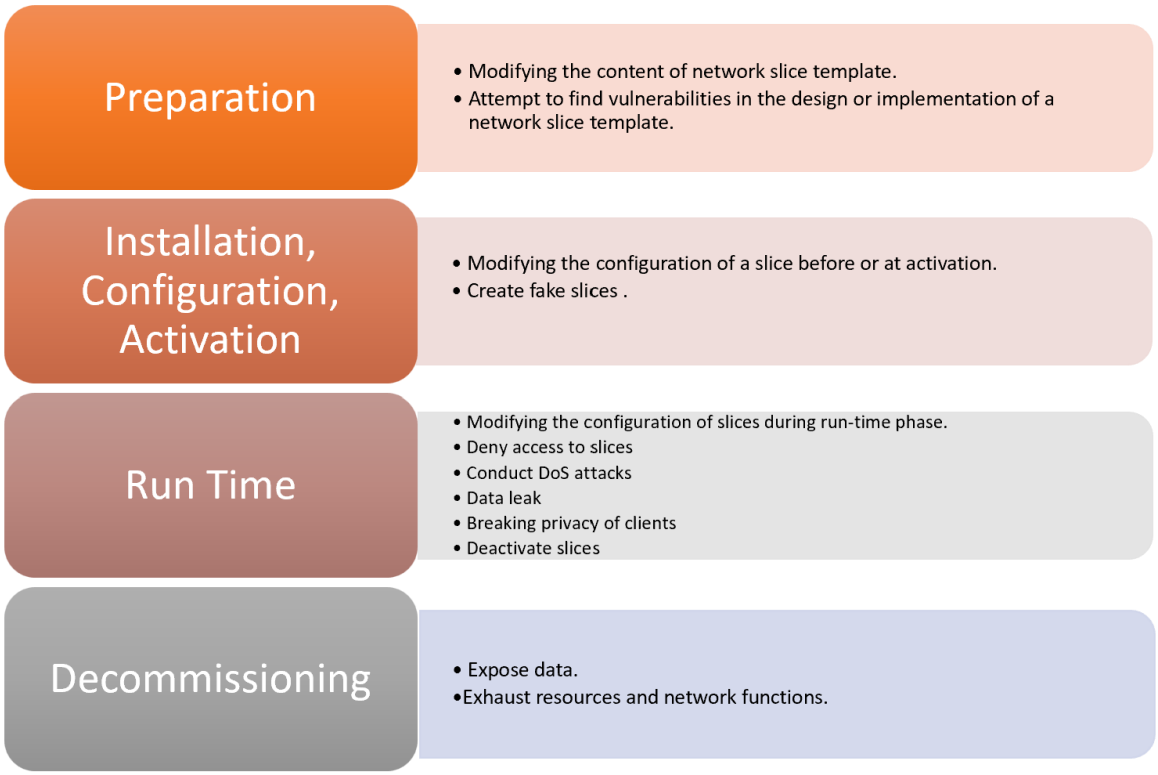

Figure 3.1: Threats for each phase of the slice life cycle

and can not be eliminated to avoid such concerns. There is a need to find a solution to manage the security challenges posed by multi-tenancy [19].

The various security concerns related to network slicing in 5G technology have been highlighted in the section above. This section analyzes these security concerns with a much more in-depth approach based on the nature of the vulnerability. There are three primary categories of security vulnerabilities:

1. Network slice life-cycle security concerns
2. Intra-slice security concerns
3. Inter-slice security concerns

3.3.1 Slice Life-Cycle Security Challenges

In chapter 2 section 2.3, we discuss that a slice consists of four phases: slice commissioning, activation, run-time operations, monitoring, and lastly slice decommissioning. However, there are

some security concerns within the slice life-cycle. At various points during this life cycle, various security concerns are raised with the network slice. These security concerns can be attributed to both design vulnerabilities in the preparation, installation configuration, and activation phases, as well as functional vulnerabilities in the run-time phase and the decommissioning.

In the commissioning stage also known as preparation, poorly designed network slice templates may expose the slice to vulnerabilities that could be exploited in the future. This would be the main point of attack. Design flaws that were found during the development stage of this system once implemented can lead to serious security concerns during run-time and during decommissioning [19].

The second stage of a slice's life cycle is installation, activation, and configuration. There may be certain security concerns that could result from this. The implementation of the design template to a functional network slice marks the beginning of actual functionality in the network. This design template may lead to errors that open up the network to security vulnerabilities such as creating fake slices or changing the configuration of slices before or during this phase [19].

The run-time phase of the slice's life cycle is characterized by many vulnerabilities. If The first two phases are potentially flawed, then this phase will display all of the flaws. These flaws can result in vulnerability to DDoS attacks, data privacy breaches, deletions, or denial of access to slices, thus actually impacting slice functionality. In addition, any changes made to the resource configuration during this phase may expose slices to vulnerabilities [19].

The main concerns in the last phase of the slice's life cycle around data privacy and resource appropriation. Inefficient decommissioning protocols may allow sensitive data to be exposed or use up previously allotted resources more than necessary. Therefore, it is clear that throughout the life cycle of a network slice, various security concerns are always present. To this end, it is important to take appropriate mitigation measures at all stages of the slice's life cycle [19].

3.3.2 Intra-slice Security

The network itself presents multiple security vulnerabilities, even while ignoring the other points of attack. 5G technology is being largely diverted to non-technical users who are not aware of security, unlike organizations that are considered to be technical users. For shared infrastructure, this is a serious attack point for unauthorized users. If access to the slices is available, this opens

up the possibility of DoS attacks that can have a significant impact on the functionality of network slices. In addition, unauthorized access to slices raises concerns about slice functionality, where the identity of slices is a major factor limiting legitimate users' access to slices, in order to reduce the risk of unauthorized access to 5G devices. It is important to implement those strong authentication and control devices [19].

The slice service interface presents another potential security vulnerability in network slicing. The interface between the slice and the services it facilitates may be vulnerable to exploitation. Attacks on the service affect the functionality of the slice, which consequently affects other services running on the same slice. Addressing these concerns requires implementing robust security controls and improving service configurations to limit their vulnerability to attack. More importantly, it may be effective to achieve partial isolation between services provided by the same slice, effectively limiting the chain reaction of attacks on a single service being transmitted to other services. All communication interfaces within a slice, i.e. layers, sub-slices, slices, services, etc. between slices and resources. To ensure that the target level of security is maintained, the system should have robust mechanisms [19].

Depending on the configuration of a network slice, sub slices to the slice might present feasible attack points that jeopardize the security of the slice. A slice may include multiple interconnected sub-slices with differing security protocols, in which case an attacker may easily exploit the weakest link in the chain of sub-slices to gain access to the entire slice. Therefore, it is essential to prevent this possibility by implementing controls that address the risk of sub-slice interconnection, such as end-to-end isolation of sub-slices. One way to address this issue is through mutual authentication requirements whereby both tenant and the host are involved in authentication in the case of a single slice manager. Alternatively, if multiple tenants are slice managers, they must authenticate each other to reduce the risk of unauthorized access to the slice. In addition, strict legal restrictions must be imposed and implemented to prevent tenants from accessing requests, data sources, and features. Access is legally permitted. Finally, the resources and network function that the slice relies on can be attacked in various ways, including physical damage, cyber-attack, or software attack to damage the functionality of the slice. To ensure reliable authentication, secure booting, credential access, integrity verification, and physical security, all of these must be given high priority [19].

3.3.3 Inter-slice security

The third set of security concerns in network slicing arises from the potential for network slicing to compromise information and resources within a shared core infrastructure. With 5G consumer devices becoming more common, hackers may find them to be a valuable target. Specifically, a consumer device that is granted access to one site may exploit inter-linkages with other slices to gain unauthorized access to other slices. This has been a fairly high level of success, especially since the adversarial device is not a stranger. A chain of slices in the network. It is easier to execute a DoS attack against a network slice that shares resources than against a network slice that does not share resources. This risk increases disproportionately if the access technology is different. In addition, if a single device is allowed multiple access to different slices with different levels of security protocols, there is an increased risk of data breaches from more secure sites to less secure sites. To protect against security concerns, the most effective strategy is to set up strict isolation parameters between slices sharing the same network resources [19].

service-to-service communication can also be a point of vulnerability on a network slice. In networks where different slices of the infrastructure offer different services, vulnerabilities in one service may allow attacks on other services. However, in 5G networks, where services are run on independent slices, this risk is low. Intra-slice communication can present a security risk for network slices in 5G networking. Less secure slices may be exploited as avenues to access more secure slices through communication interfaces between these slices. Strengthening slice isolation will help protect against vulnerabilities. This keeps the slices from being affected by any compromise [19].

Management systems may present inter-slice security concerns in network slicing. Sharing management of slices makes it possible to execute attacks on other slices. To protect the network from attacks, proper isolation is necessary to configure the shared network infrastructure. In addition, proper restrictions must be put in place to make changes to slices within the slice manager. The shared resource infrastructure is a key vulnerability in network slicing. One way hackers can damage a network is by using attacks that consume resources, such as those that exhaust computer resources. This can have a ripple effect on other parts of the network, exacerbating the damage. The best way to mitigate the risk of this happening lies in preferential isolation and resource allotment

[19].

3.4 Technological Solutions Towards 5G Network Security

The security vulnerabilities presented by network slicing have been analyzed in depth above based on the type of vulnerability. This section will assess the various solutions that can be implemented to combat these vulnerabilities. In particular, end-to-end security, isolation, secure management, and orchestration of network slices within a shared architecture will be explored [19].

3.4.1 End-to-End Security

One of the most effective strategies to address the 5G security concerns posed by network slicing technology is to manipulate the system architecture to achieve end-to-end security. It is important to note that these security concerns may come from two different types of adversaries [24]:

1. Adversaries with administrative control: someone who manages the resources of the slice such as the tenant, the mobile operator, or an external attacker who compromised the slice.

2. External adversaries: a user of the slice, or someone who could attack the slice.

End-to-end security can manage security concerns from external adversaries, but it can also provide significant insulation against adversaries who have administrative control. The architecture model of network slicing features multiple virtual networks that are isolated from each other and from the shared network infrastructure core [25]. These virtual networks have the necessary characteristics to enable them to perform a specific function or to be similar to other networks that are already in use. The overall security of these networks reduces the concerns of vulnerabilities that result from communication between host resource interfaces and slices, between slices themselves, and between sub-slices or slice components of a network slice to limit the threats of the network at endpoints including access, permission [24]. External adversaries often try to attack the interface between communication lines in order to get into the network. End-to-end security in encryption and decryption to and from the network slice significantly reduces this risk making it harder for external attackers to infiltrate the network. This ensures that the data is adequately protected from potential intrusion and unauthorized access and is stored in a format that bets ensure its security.

3.4.2 Isolation

Resource sharing in network slicing presents significant security vulnerabilities. Multiple tenants sharing a shared infrastructure can create new security challenges in 5G, which need to be addressed. To address security concerns arising from network slicing, isolation has been sought as an efficient means to minimize the impact on host resources and network communication. This research studies intra-slice and inter-slice security vulnerabilities on various levels, aiming to help protect devices from 5G communication and sharing resources. Addressing these risks allows for the full exploitation of 5G network capacity. Two tiers of slice isolation have been proposed and explored in minimizing network slicing-associated vulnerabilities [24]:

1. **Inter-slice isolation:**

   These features separate the host hardware resources from the network slice as a whole. This severely restricts communication between slices since they do not share hardware resources. [24]

2. **Intra-slice isolation:**

   The separation of host hardware resources between slice components allows each slice to focus on its own tasks, eliminating the sharing of resources between slices. The goal of intra-slice isolation is to create different hosts for slice components so that they can share resources more efficiently, reducing the need for defense against all slices [24].

Slice isolation has emerged as a more efficient solution for mitigating DDOs attacks than target network upscaling and traffic blocking. One way to reduce the risk of a large-scale cyber-attack is to share common infrastructure. This way, if one slice of the network is attacked, the others can remain unaffected. Reducing the number of communication links between the various slices reduces the risk of a widespread attack. Experimental simulations have shown that slice isolation can help to minimize the impact of DDoS attacks, which can then be contained and prevented from spreading to other slices [24].

Slicing through isolation has to be done efficiently in order to ensure that users receive the desired quality of service (QoS). Inter-slice isolation provides strong resource isolation, limiting the possibility of an attack. However, it also limits the efficiency of the use of resources, counters

to optimizing the benefits of 5G technology. [24]. Therefore, a number of algorithms and block diagrams have been explored to determine isolation models that retain the desired QoS level [26].

3.4.3 Secure Management and Orchestration

Network slicing-associated security concerns can be managed by enhancing the management of network slices. While end-to-end security is best for external attackers, it's worth noting that there are also significant risks to attackers with administrative control. As such, effective management and orchestration of security risks are essential in mitigating this type of risk. A variety of actions can be taken to ensure the security of slices. These include tighter access control regimes, firewalls, virtual private networks (VPNs), and authentication capabilities designed to mitigate security vulnerabilities created by resource sharing in network slices [26].

For example, to limit security breaches caused by multiple tenants, mutual authentication can be implemented to limit unauthorized access to other slices. This should be done in conjunction with the implementation of the Security Level Service Agreement (SLSA), which will allow for enhanced and efficient management of network slices [27]. The SLSAs require security service providers to guarantee that specific security protocols are in place to address security concerns within a shared infrastructure. This vertical integration management in security provision will tackle security concerns from both internal and external adversaries, as well as address mobility-related challenges in network slicing [27].

Some scholars have proposed that slice management allows manual distribution of functions between slices, which improves the security of the management and limits the network slice's vulnerability to DDoS attacks. Resource allocation while maximizing benefits from network slicing is a challenging issue in network slicing management [19]. A reliable system is required to guarantee the accommodation of maximal diversified service requests with limited resource allotment.

This requires the establishment of a Public Key Infrastructure (PKI) that further improves slice management and optimizes security for the slices. To effectively manage and orchestrate security measures to mitigate network slicing vulnerabilities, organizations will need to adopt standardized standards for cybersecurity that legally protect network slices. This will be essential in closing loopholes in slice management and promoting slice security [26].

Chapter 4

Comparison of Different RAN Scheduling Methods

4.1 Overview

Designing a fifth-generation (5G) mobile network to support the various use cases with different requirements is a very difficult challenge. However, 5G is considered to have a better performance in terms of efficiency and flexibility compared to the past generations. As discussed in Chapter 1, the eMBB service is suitable for high data rate applications such as video streaming and is able to satisfy the need for moderate reliability with a packet error rate (PER) of 10^{-3}. On the other hand, the URLLC use case is the service that has a short packet size and is in need of a PER with 10^{-5}, which corresponds to very high-reliability [21].

The concept of network slicing is about sharing the network services between different use cases. Since there are multiple waveform configurations in 5G NR, radio frames can be customized based on these needs. In addition, the 5G NR standard requires that the sub-carrier spacing (SCS) increases when higher frequency bands are used. The radio frame in the 5G NR is set to 10 ms, while each sub-frame is always 1 ms, disregarding the numerology of the network system. The number of symbols within a slot and the number of the slots within a sub-frame can be changed by changing some parameters of the network such as numerology [21].

For the URLLC service, as it needs an immediate reaction to satisfy the low latency requirement, one way to do that is by changing the SCS. Having a higher SCS would affect the length of the time slot. Another way is to reduce the symbols in the packet Transmission Time Interval (TTI)

by using the concept of mini slots. Studying the scenario for the coexistence of the eMBB and URLLC services over the same resources introduces problems for the eMBB traffic, and contrarily the latency requirements of the URLLC transmissions can not be disregarded [21]. There are two proposed scheduling solutions by the Third Generation Partnership Project (3GPP):

1. **Orthogonal Scheduling:**

   In orthogonal scheduling, some of the frequency channels are reserved for the upcoming URLLC traffic. There are two types of reservations: semi-static and dynamic reservations. When the gNB transmits the configuration of the frame structure at irregular times, this is known as semi-static. On the other hand, dynamic reservation is when the frame structure is modified with each scheduled user by the downlink control channel. The disadvantage of orthogonal scheduling is that when there is no upcoming traffic for the URLLC, all of the reserved resources will be wasted [28].

2. **Preemptive Scheduling:**

   In preemptive scheduling, the gNB would pause the transmission of the eMBB service if there is upcoming URLLC traffic. The traffic for the URLLC would be scheduled in a short TTI in an effort to satisfy the latency restraint. However, the reliability of the eMBB transmission would be affected, so a solution for the services' coexistence is needed [28].

4.2 5G NR Frame Structure

The Third Generation Partnership Project (3GPP) has defined the 5G NR frame structure specifications and here we present details of the frame structure in regards to numerologies, subcarrier spacing, and slots [29].

4.2.1 5G NR Numerology and Subcarrier Spacing

In the 4G system (LTE), there is only one type of numerology/subcarrier spacing considered in the network which is a subcarrier spacing of 15 kHz. However, the major improvement in 5G NR is having multiple Orthogonal Frequency Division Multiplexing (OFDM) numerologies to support diverse services as given by Table 4.1 where μ is the numerology for the subcarrier spacing and

μ	$\delta f = 2^{\mu} * 15[kHZ]$	Cyclic prefix
0	15	Normal
1	30	Normal
2	60	Normal, Extended
3	120	Normal
4	240	Normal

Table 4.1: Transmission numerologies.

μ	N_{symb}^{slot}	$N_{slot}^{frame,\mu}$	$N_{slot}^{subframe,\mu}$
0	14	10	1
1	14	20	2
2	14	40	4
3	14	80	8
4	14	160	16

Table 4.2: Number of OFDM symbols per slot, number of slots per frame, and number of slots per subframe for the normal cyclic prefix.

μ	N_{symb}^{slot}	$N_{slot}^{frame,\mu}$	$N_{slot}^{subframe,\mu}$
2	12	40	4

Table 4.3: Number of OFDM symbols per slot, number of slots per frame, and number of slots per subframe for the extended cyclic prefix.

δf is the subcarrier spacing. There are two types of cyclic prefixes it can be normal or extended according to 3GPP [29]. A cyclic prefix is a prefix of a symbol or the guard interval between each symbol to protect the signals from intersymbol interference.

4.2.2 Slot Length

Depending on the μ numerology chosen for the network configuration, slots within a subframe are numbered $n_s^{\mu} \in \{0, ..., \mathrm{N}_{slot}^{subframe,\mu} \text{ -1}\}$. While slots within a subframe can be numbered, $n_{s,f}^{\mu}$ $\in \{0, ..., \mathrm{N}_{slot}^{frame,\mu} \text{ -1}\}$. Th number of the consecutive OFDM symbols in a slot depends on the cyclic prefix of a frame, and is given by Tables 4.3, 4.5.

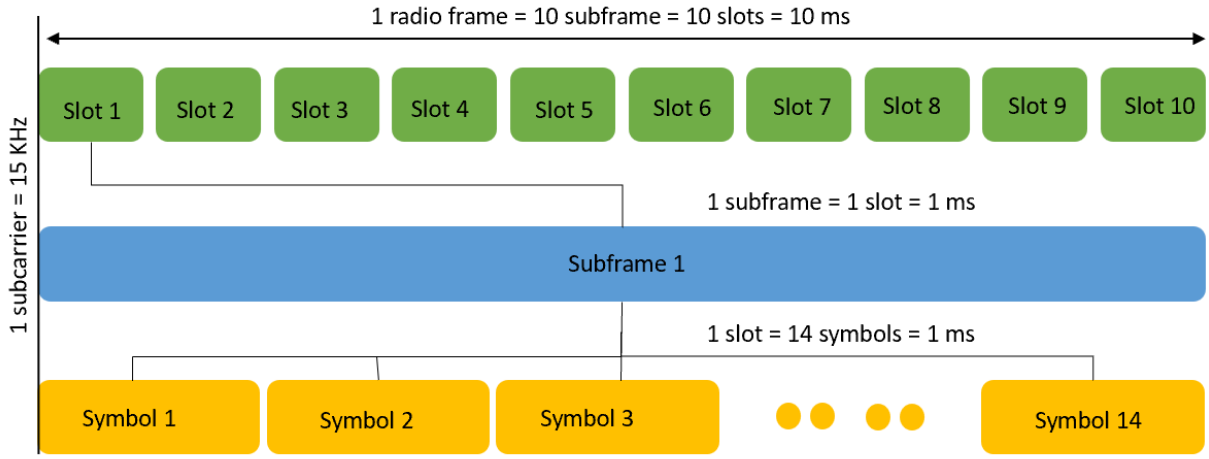

Figure 4.1: Frame structure of a subcarrier spacing of 15 kHZ

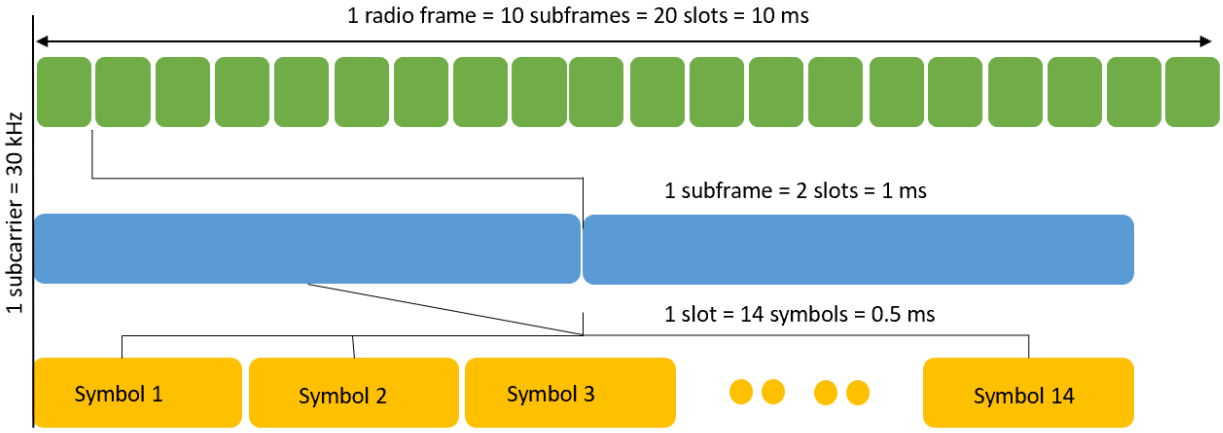

Figure 4.2: Frame structure of a subcarrier spacing of 30 kHZ

4.2.3 Mini Slots

As mentioned earlier, LTE uses fixed numerology of 15 kHz SCS, unlike 5G which supports a SCS of 15 - 240 kHz as referenced in Table 4.3. 5G NR supports having fewer OFDM symbols in a slot. A slot originally consists of 14 OFDM symbols. The mini slot concept can support a slot with only 2, 4, or 7 OFDM symbols per slot. These shorter slots help to decrease the transmission time as shown in Table 4.5 [30].

SCS [kHZ]	7 symbols	4 symbols	2 symbols
15	500 μs	286 μs	143 μs
30	250 μs	143 μs	71 μs
60	125 μs	71 μs	36 μs

Table 4.4: The Time Duration of a Mini-Slot as a Function of the Number of Symbols It Contains

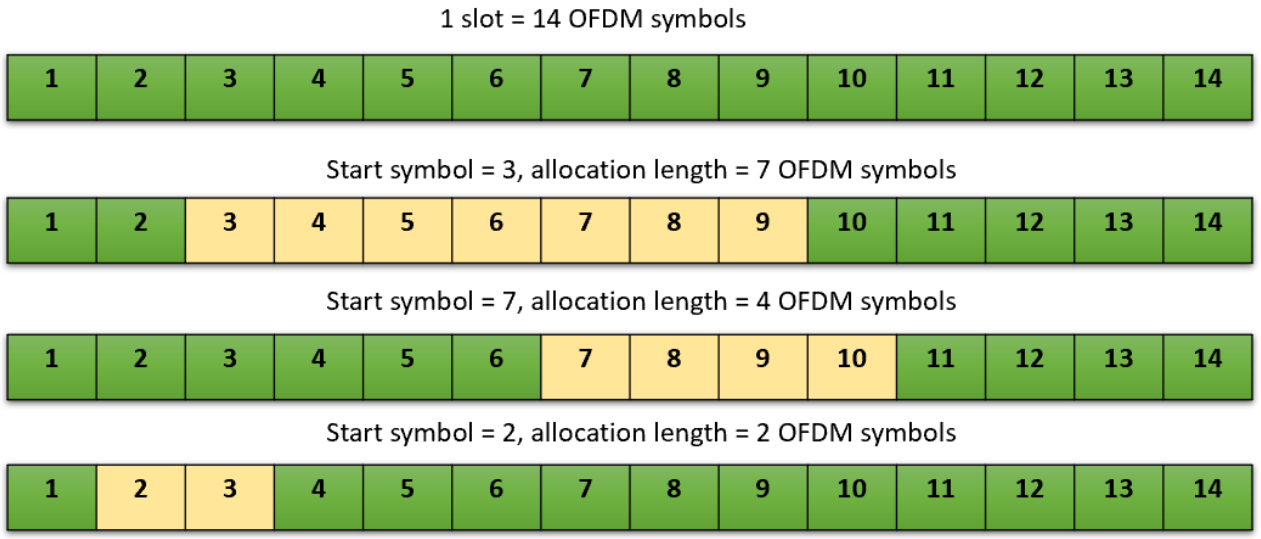

Figure 4.3: Mini Slots

μ	SCS [kHZ]	Bandwidth in kHZ per RB = $12 \times SCS$
0	15	180
1	30	360
2	60	720
3	120	1440
4	240	2880

Table 4.5: bandwidth occupied by a Resource Block.

4.2.4 Resource Blocks

A Resource Block (RB) consists of 12 consecutive subcarriers in the frequency domain. In LTE, a RB is defined to occupy one slot in a time domain and 12 subcarriers in the frequency domain. However, in 5G, a RB is only defined in the frequency domain. As given in Table 4.6, the bandwidth per RB can vary depending on the numerology.

Bandwidth [MHz]	15 kHZ	30 kHZ	60 kHZ
5	25	11	NA
10	52	24	11
15	79	38	18
20	106	51	24
25	133	65	31
30	160	78	38
40	216	106	51
50	270	133	65
60	NA	162	79
80	NA	217	107
90	NA	245	121
100	NA	273	135

Table 4.6: Maximum number of RBs for each channel bandwidth .

However, there is a maximum number of RBs for each channel bandwidth and subcarrier spacing as specified by 3GPP [29] and can be summarized in Table 4.7.

4.3 Prioritized Scheduling with Round Robin

Round robin scheduling is an algorithm that schedules the processes fairly by distributing all of the available resources among the processes, giving each process an equal, fixed amount of time, known as a time quantum. It is considered to be an efficient algorithm to avoid starvation since all of the processes in the ready queue will be executed in the same amount of time [31]. A flow chart for the round robin scheduling algorithm is as shown in Fig. 4.4.

On the other hand, priority scheduling is an algorithm that schedules the processes depending on their priority. A process with a higher priority would be admitted to the network first. A flow chart of the Prioritized Round Robin (PRR) scheduling algorithm is shown in Fig. 4.5.

Scheduling algorithms may be compared on the basis of different times, including the turn around time and the waiting time. The **turn around time** is the total amount of time spent by the process from coming in the ready state for the first time to its completion. It can be expressed by the following equation:

$$\text{Turn Around Time} = \text{Burst time} + \text{Waiting time}$$

The **waiting time** is the total time spent by the process in the ready state waiting to be

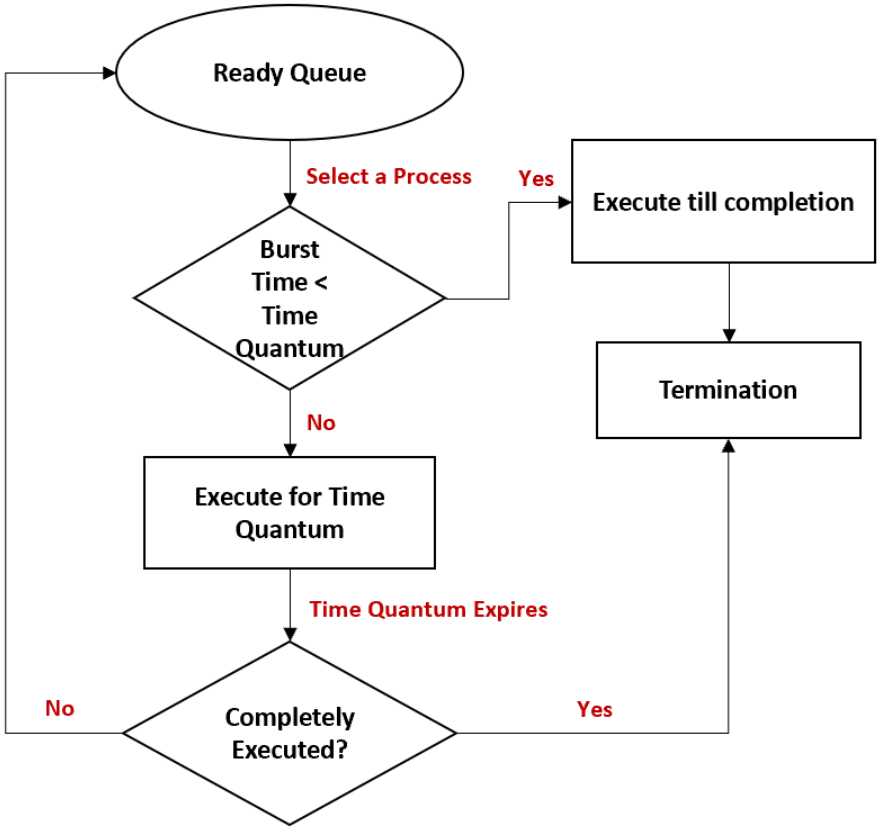

Figure 4.4: Round Robin Flow Chart

admitted. It can be expressed by the following equation:

$$\text{Waiting Time} = \text{Turn around time} - \text{Burst time}$$

Prioritized scheduling with round robin is a modified scheduling algorithm that combines the two types of scheduling which are: round robin, and priority scheduling. URLLC traffic is assigned with the highest priority, while we use round robin to avoid starvation for eMBB traffic that has low priority.

4.3.1 Considerations in Prioritized Round Robin Simulation

To provide a concrete illustration of how prioritized round robin works and give numerical results, we have engaged in a simulation study. In the simulation study, we consider the following parameters:

1. **Subcarrier Spacing:** The subcarrier spacing considered in the simulation is 30 kHz.

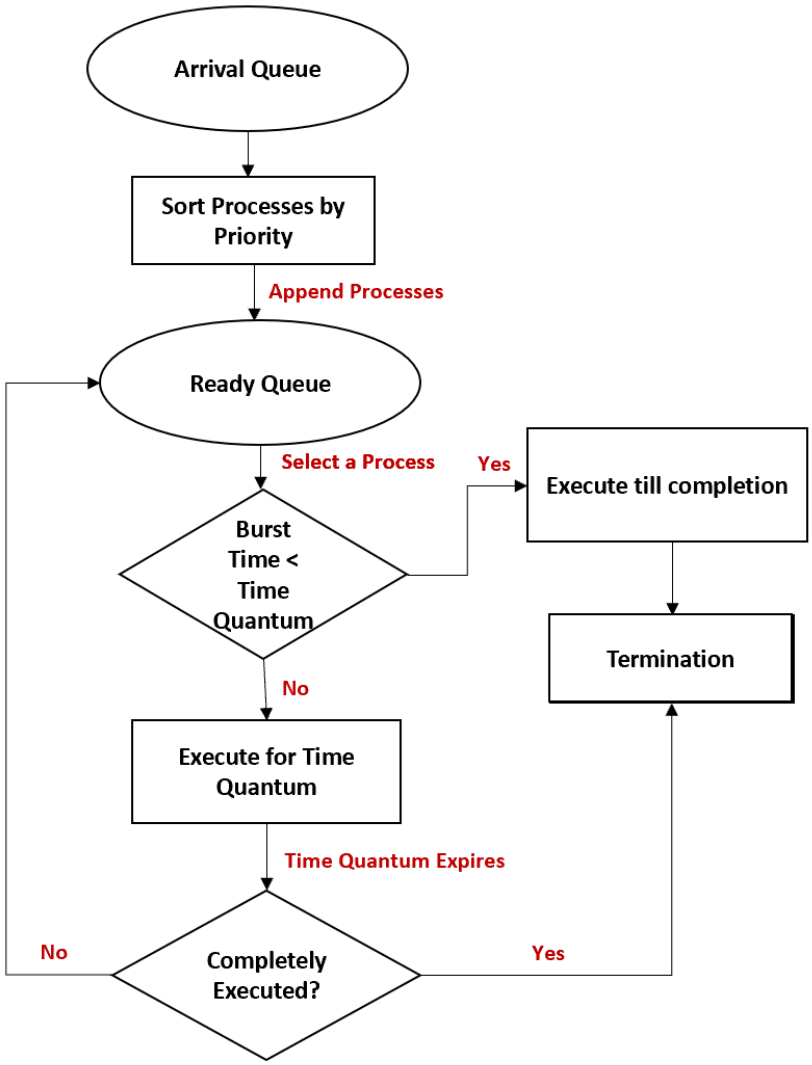

Figure 4.5: Prioritized Round Robin Flow Chart

2. **Time quantum:** For the simulation, the time quantum used in the system is the slot duration for each use case in the prioritized static round robin. While in the prioritized dynamic round robin, the time quantum is the burst time for the use case.

$$\text{slot Duration} = \frac{1}{\frac{\text{SCS}}{15kHZ}} = \frac{1}{\frac{30\text{kHZ}}{15\text{kHZ}}} = 0.5 \text{ ms}$$

3. **Bandwidth:** 20 MHz

4. **Resource Blocks:** For the subcarrier spacing of 30 kHz and using 20 MHz for the bandwidth, referring to Table 4.6, the number of resource blocks is 51. However, the resource block

Simulation Parameter	Value
eMBB Users	4
URLLC Users	2
Subcarrier Spacing	30 kHZ
Frequency	20 MHz
Number of Frames	20
Resource Blocks	51
Slot Duration	0.5 ms

Table 4.7: Simulation Parameters.

requests are generated randomly for each user, eMBB would require more resource blocks than the URLLC use cases.

5. **Burst Time:** The burst time in round-robin scheduling is the time that each process needs to finish its execution. The burst time for each process can be calculated based on the needs of the resource blocks.

$$\text{Burst Time} = \frac{\text{slot Duration} \times \text{RB request}}{\text{RB available}}$$

6. **Arrival Times:** between every two eMBB processes, one URLLC process is generated.

7. **Number of Frames:** Number of frames in the simulation are 20. Each radio frame is 10 ms.

8. **Number of Slots in the Frame:**

$$N_{slots}^{frame} = \frac{10}{\text{slot Duration}} = \frac{10}{0.5} = 20 \text{ slots}$$

9. **Number of Slots in the Simulation:**

$$N_{slots}^{Sim} = \text{Number of Frames} \times \text{Number of Slots in Frames} = 20 \times 20 = 400 \text{ slots.}$$

10. **Priority:** URLLC use cases are assigned to have a higher priority than eMBB use cases.

However, there are many different ways to improve the performance of the round robin scheduling algorithm.

4.3.2 Prioritized Round Robin Algorithm with Static Time Quantum

A round robin scheduling algorithm with a static time quantum uses a fixed value for the time quantum for each process execution. In this case, we consider the coexistence scenario of the two different use cases, URLLC, and eMBB. We prioritize the URLLC use cases by assigning them a higher priority than the eMBB traffic. The time quantum value that is used is the actual value of the slot Duration which is 0.5 ms.

4.3.3 Prioritized Round Robin Algorithm with Dynamic Time Quantum

To optimize the round robin algorithm scheduling, we use a dynamic time quantum. The value that is used for the time quantum will not be a fixed value, it will change depending on each use case. The time quantum will be the same as the burst time of each use case, which is the total time that a use case needs to complete its execution.

4.4 Comparing Results

After conducting 3 trials of the simulation, the results are shown in Figs. 4.9 - 4.11 and Tables 4.8, 4.9. We compare the results based on the waiting and turn around times.

4.4.1 Waiting Times Static vs Dynamic

For the waiting time, It is shown that the PRR using a dynamic time quantum shows better results than using a static time quantum, and that is because varying the time quantum based on the UE needs would help avoid having idle resources and that would decrease the transmission time of the next available use case that is in the ready queue waiting to be admitted as shown in Fig. 4.9, 4.7 and 4.8. Moreover, looking at the average waiting time and average turnaround time that is shown in Table 4.8, the average waiting time for the overall simulation is improved, so that it does not cause the eMBB traffic which has a low priority, to get jammed and affect their latency requirements.

Trial	PRR Static Time Quantum	PRR Dynamic Time Quantum
1	0.47385	0.4084
2	1.1535	1.0441
3	1.0571	0.8921

Table 4.8: Comparison of Average Waiting Times in ms

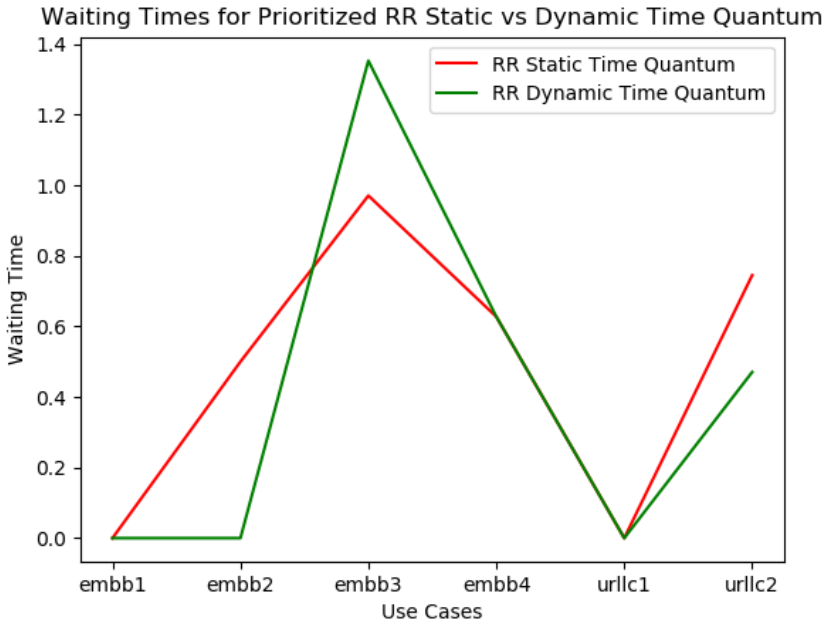

Figure 4.6: Waiting Times for Prioritized Round Robin Static vs Dynamic Time Quantum (trial 1)

4.4.2 Turn Around Times Static vs Dynamic

Turn around time is the amount of time for each use case from the arrival to the ready queue till their completion of execution. The goal is to minimize the turnaround time for URLLC use cases, as seen in the figures 4.9 - 4.8, using PRR with a dynamic time quantum is showing better results that are due to giving each use case their resource needs, without having to assign each process an equal amount of time in a circular order. Resource block needs for each UE can be calculated from the burst time and resource block request for each use case. The eMBB use cases usually have a higher turnaround time because they usually request more resources than the URLLC use cases.

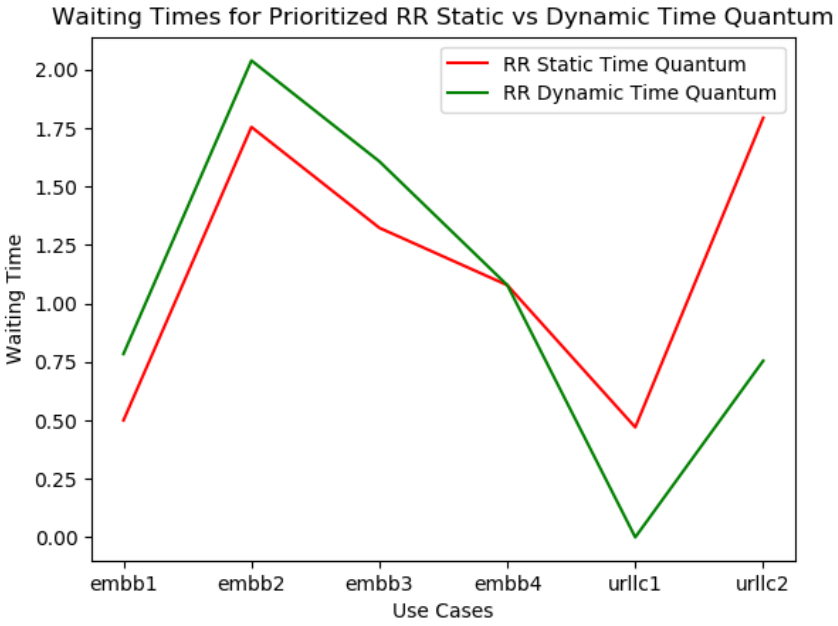

Figure 4.7: Waiting Times for Prioritized Round Robin Static vs Dynamic Time Quantum (trial 2)

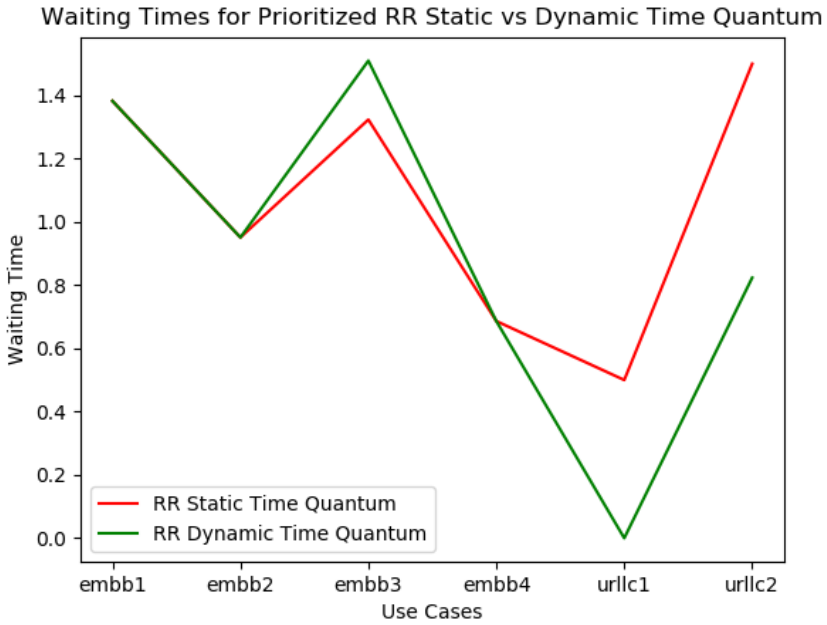

Figure 4.8: Waiting Times for Prioritized Round Robin Static vs Dynamic Time Quantum (trial 3)

Trial	PRR Static Time Quantum	PRR Dynamic Time Quantum
1	1.6503	1.5849
2	2.4771	2.3676
3	2.2663	2.1013

Table 4.9: Comparison of Average Turn Around Times in ms

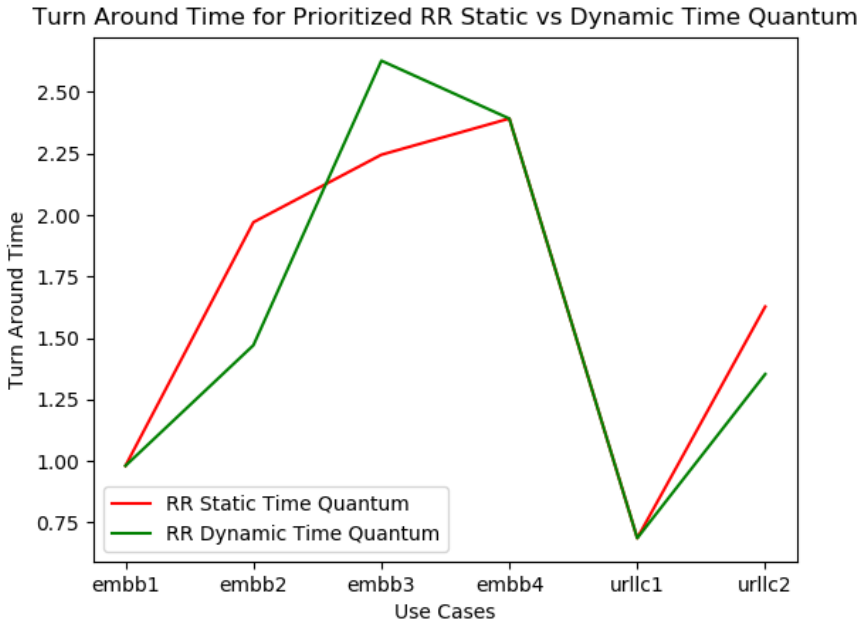

Figure 4.9: Turn Around Times for Prioritized Round Robin Static vs Dynamic Time Quantum (trial 1)

4.5 Secured RAN Slicing

Network slicing is considered to be a new technology and we have identified some of the security threats in network slicing or RAN specifically, as discussed in chapter 3. There are many significant challenges within the implementation of RAN. One of the main security threats is the Denial of Service (DoS) and the exhaustion of resources. To avoid such a vulnerability, isolation technology must be used to ensure the isolation of resources, traffic, and users. However, with the small cells enabled technology in 5G, millimeter waves should be Incorporated to avoid interference or RAN technologies [26]. Ensuring security in RAN slicing is essential, due to having to meet the Quality of Service (QoS) and the Quality of Experience (QoE). Every slice would have a tenant

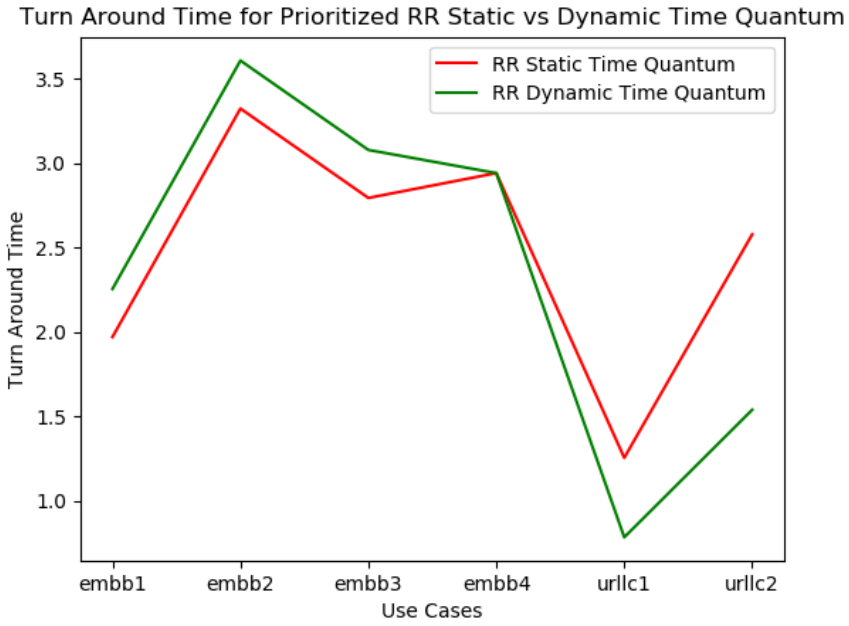

Figure 4.10: Turn Around Times for Prioritized Round Robin Static vs Dynamic Time Quantum (trial 2)

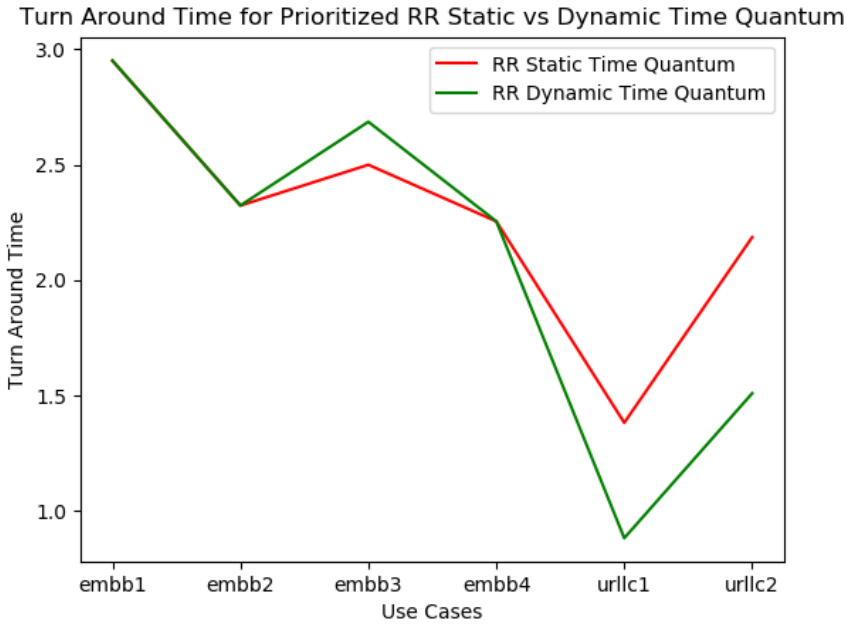

Figure 4.11: Turn Around Times for Prioritized Round Robin Static vs Dynamic Time Quantum (trial 3)

that is responsible for the management of the radio resources in the slice, some of the management functions should be isolated from the other slices such as authentication and authorization. In addition, operators and tenants are responsible to manage the resource to assure that each slice is equipped with the maximum amount of resources that it needs and all resources must be isolated. RAN must be configured properly for each slice based on the slice's demands. Some parameters must be isolated to achieve high security and meet the demands of each use case. First, isolating the network traffic so that each slice has its network traffic flow and wouldn't exhaust the shared network resources. Second, setting a bandwidth for each slice based on their needs and priority. Isolating the bandwidth would avoid utilizing the assigned bandwidth for the other slices [32].

Chapter 5

Conclusion

5.1 Summary

Network slicing is a feature of 5G that distinguishes it from the previous network generations. It is considered to be the backbone behind 5G technology [27]. Network slicing is one of the technologies that will lead to the success of the fifth-generation mobile network. This is because it allows the network to achieve different capabilities. For instance, it enables different kinds of 5G connectivity in terms of high data speeds and low latency.

In Chapter 2, we define the concept of network slicing, which is the creation of multiple virtual networks on top of a shared core infrastructure network. Moreover, network slices are end-to-end logical networks that are designed to satisfy use case requirements such as high throughput, low latency, and high reliability.

The architecture of network slicing consists of three layers. The resource layer contains the network functions and resources to deliver a specific service for an end-user. The network slice instance layer contains slices that are configured and designed by the request of the end-user. While the service instance layer has the services that are ready and available for UEs. Each slice undergoes four stages: commissioning, activation, run-time, and decommissioning. Network slicing can happen in either the core network or in the Radio Access Network (RAN).

The network slicing concept deploys the network function virtualization and orchestration that assigns a tenant that is responsible for managing the resources for each slice. However, that has allowed for new security vulnerabilities such as Distributed Denial of Service (DDoS) and resource exhaustion. There are some representative security threats in the network slicing architecture, and

some can happen within the life-cycle of a network slice. In addition, some security concerns can be analyzed concerning the network slice giving rise to the inter-slice and intra-slice security concerns. Understanding these risks in depth helps to explore mitigation strategies that can be employed to avoid such risks. In Chapter 3 we have represented some of the challenges in the 5G network, network slicing, and slicing in the Radio Access Network (RAN). While we have also represented some of the technological solutions to minimize the security concerns raised by network slicing. One of the solutions is the isolation of network slices by limiting the communication between slices, sub-slices, and resources host. We have included some other solutions, such as securing the management and orchestration and designing an end-to-end security system.

Lastly, in Chapter 4 we have designed an algorithm that combines the two scheduling algorithms, round-robin scheduling with priority scheduling. The main goal of the algorithm is to give the Ultra-reliable-low-latency (URLLC) a high priority since it is sensitive in terms of latency and reliability and gives the enhanced mobile broadband (eMBB) a lower priority. We used a round-robin to avoid the starvation that is caused by admitting the use cases with higher priority to the network, while the other use cases with low priority such as eMBB get jammed in the ready queue waiting to be admitted to the network. There were two proposed scheduling algorithms, the prioritized round-robin (PRR) which uses a static time quantum, and the prioritized round-robin which uses a dynamic time quantum.

5.2 Future work

The simulation presented in this thesis is based on the scenario of eMBB-URLLC coexistence. We considered implementing the scheduling algorithm for use cases admission, we compared waiting times and turnaround times for each use case for both of the cases, PRR that uses a dynamic vs static time quantum. A great extension to the implementation would be simulating the given scenario in a network simulator to measure the latency for the URLLC use cases, the throughput for the eMBB use cases, and creating a traffic generator by instantiating packets objects to keep on track with the reliability, throughput, and latency constraints. In addition to looking at the priority levels of the security control tasks in scheduling simulations to compare the overall performance.

References

[1] S. Sirotkin, *5G Radio Access Network Architecture: The Dark Side of 5G*. John Wiley & Sons, 2020.

[2] "Network slicing and 5g future shock."

[3] P. Popovski, K. F. Trillingsgaard, O. Simeone, and G. Durisi, "5g wireless network slicing for embb, urllc, and mmtc: A communication-theoretic view," *Ieee Access*, vol. 6, pp. 55 765–55 779, 2018.

[4] S. Zhang, "An overview of network slicing for 5g," *IEEE Wireless Communications*, vol. 26, no. 3, pp. 111–117, 2019.

[5] I. F. Akyildiz, S. Nie, S.-C. Lin, and M. Chandrasekaran, "5g roadmap: 10 key enabling technologies," *Computer Networks*, vol. 106, pp. 17–48, 2016.

[6] ITU, "Minimum requirements related to technical performance for IMT-2020 radio interface(s)," International Telecommunication Union, Report, 2017, iTU-R M.2410-0. [Online]. Available: https://www.itu.int/dms$_pub/itu-r/opb/rep/R-REP-M.2410-2017-PDF-E.pdf$

[7] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. Soong, and J. C. Zhang, "What will 5g be?" *IEEE Journal on selected areas in communications*, vol. 32, no. 6, pp. 1065–1082, 2014.

[8] E. Björnson, E. G. Larsson, and T. L. Marzetta, "Massive mimo: Ten myths and one critical question," *IEEE Communications Magazine*, vol. 54, no. 2, pp. 114–123, 2016.

[9] R. Chataut and R. Akl, "Massive mimo systems for 5g and beyond networks—overview, recent trends, challenges, and future research direction," *Sensors*, vol. 20, no. 10, p. 2753, 2020.

[10] Z. Zhang, K. Long, A. V. Vasilakos, and L. Hanzo, "Full-duplex wireless communications: Challenges, solutions, and future research directions," *Proceedings of the IEEE*, vol. 104, no. 7, pp. 1369–1409, 2016.

[11] L.-P. Tung, L.-C. Wang, and K.-S. Chen, "An interference-aware small cell on/off mechanism in hyper dense small cell networks," in *2017 International Conference on Computing, Networking and Communications (ICNC)*. IEEE, 2017, pp. 767–771.

[12] T. Kebede, Y. Wondie, J. Steinbrunn, H. B. Kassa, and K. T. Kornegay, "Precoding and beamforming techniques in mmwave-massive mimo: Performance assessment," *IEEE Access*, vol. 10, pp. 16 365–16 387, 2022.

[13] S. Hakola, T. Chen, J. Lehtomäki, and T. Koskela, "Device-to-device (d2d) communication in cellular network-performance analysis of optimum and practical communication mode selection," in *2010 IEEE wireless communication and networking conference*. IEEE, 2010, pp. 1–6.

[14] P. Subedi, A. Alsadoon, P. Prasad, S. Rehman, N. Giweli, M. Imran, and S. Arif, "Network slicing: A next generation 5g perspective," *EURASIP Journal on Wireless Communications and Networking*, vol. 2021, no. 1, pp. 1–26, 2021.

[15] K. Sienkiewicz, W. Latoszek, and P. Krawiec, "Services orchestration within 5g networks—challenges and solutions," in *2018 Baltic URSI Symposium (URSI)*. IEEE, 2018, pp. 265–268.

[16] M. Richart, J. Baliosian, J. Serrat, and J.-L. Gorricho, "Resource allocation and management techniques for network slicing in wifi networks," in *NOMS 2020-2020 IEEE/IFIP Network Operations and Management Symposium*. IEEE, 2020, pp. 1–6.

[17] 5GAmericas, "Network Slicing for 5G Networks Services," *Network Slicing for 5G and Beyond*, 11 2016. [Online]. Available: https://www.5gamericas.org/wp-content/uploads/2019/07/5G$_{A}mericas_{N}etwork_{S}licing_{1}1.21_{F}inal.pdf$

[18] F. Granelli, "Network slicing," in *Computing in Communication Networks*. Elsevier, 2020, pp. 63–76.

[19] R. F. Olimid and G. Nencioni, "5g network slicing: A security overview," *IEEE Access*, vol. 8, pp. 99 999–100 009, 2020.

[20] M. A. Habibi, B. Han, and H. D. Schotten, "Network slicing in 5g mobile communication architecture, profit modeling, and challenges," *arXiv preprint arXiv:1707.00852*, 2017.

[21] J. Mei, X. Wang, and K. Zheng, "An intelligent self-sustained ran slicing framework for diverse service provisioning in 5g-beyond and 6g networks," *Intelligent and Converged Networks*, vol. 1, no. 3, pp. 281–294, 2020.

[22] P. Wright, C. White, R. C. Parker, J.-S. Pegon, M. Menchetti, J. Pearse, A. Bahrami, A. Moroz, A. Wonfor, R. V. Penty *et al.*, "5g network slicing with qkd and quantum-safe security," *Journal of Optical Communications and Networking*, vol. 13, no. 3, pp. 33–40, 2021.

[23] S. E. Elayoubi, S. B. Jemaa, Z. Altman, and A. Galindo-Serrano, "5g ran slicing for verticals: Enablers and challenges," *IEEE Communications Magazine*, vol. 57, no. 1, pp. 28–34, 2019.

[24] D. Sattar and A. Matrawy, "Towards secure slicing: Using slice isolation to mitigate ddos attacks on 5g core network slices," in *2019 IEEE Conference on Communications and Network Security (CNS)*. IEEE, 2019, pp. 82–90.

[25] X. Li, M. Samaka, H. A. Chan, D. Bhamare, L. Gupta, C. Guo, and R. Jain, "Network slicing for 5g: Challenges and opportunities," *IEEE Internet Computing*, vol. 21, no. 5, pp. 20–27, 2017.

[26] A. Mathew, "Network slicing in 5g and the security concerns," in *2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC)*. IEEE, 2020, pp. 75–78.

[27] P. Alemany, D. Ayed, R. Vilalta, R. Muñoz, P. Bisson, R. Casellas, and R. Martínez, "Transport network slices with security service level agreements," in *2020 22nd International Conference on Transparent Optical Networks (ICTON)*. IEEE, 2020, pp. 1–4.

[28] M. Alsenwi, N. H. Tran, M. Bennis, S. R. Pandey, A. K. Bairagi, and C. S. Hong, "Intelligent resource slicing for embb and urllc coexistence in 5g and beyond: A deep reinforcement learning based approach," *IEEE Transactions on Wireless Communications*, vol. 20, no. 7, pp. 4585–4600, 2021.

[29] 3GPP, "5G; NR; Physical channels and modulation ," 3rd Generation Partnership Project (3GPP), Technical Specification (TS), 04 2022, version 17.1.0. [Online]. Available: https://www.etsi.org/deliver/etsi$_t s$/138200$_1$38299/138211/17.01.00$_6$0/$ts_1$38211v170100$p.pdf$

[30] J. Zhang, X. Liu, M. Liang, H. Yu, and Y. Ji, "Low Latency DWBA Scheme for Mini-Slot Based 5G new Radio in a Fixed and Mobile Converged TWDM-PON," *Journal of Lightwave Technology*, vol. 40, no. 1, pp. 3–13, 2021.

[31] A. A. Alsulami, Q. A. Al-Haija, M. I. Thanoon, and Q. Mao, "Performance evaluation of dynamic round robin algorithms for cpu scheduling," in *2019 SoutheastCon*. IEEE, 2019, pp. 1–5.

[32] Z. Kotulski, T. Nowak, M. Sepczuk, M. Tunia, R. Artych, K. Bocianiak, T. Osko, and J.-P. Wary, "On end-to-end approach for slice isolation in 5g networks. fundamental challenges," in *2017 Federated conference on computer science and information systems (FedCSIS)*. IEEE, 2017, pp. 783–792.