POLITECNICO DI TORINO
Department of Electronics and Telecommunications
M.Sc in Communications and Computer Networks Engineering

Master Thesis

Network slicing and QoS in 5G systems and their impact on the MAC layer

Advisors:
Dr. Carla Fabiana Chiasserini
Dr. Achim Nahler

Author:
Enida Mataj

July 10, 2020

Abstract

5G is the new generation of cellular communication systems that will provide service not only for mobile users, but also for business unites. Different use-cases have been introduced along with 5G such as Enhanced Mobile Broadband Communication (eMBB), Ultra Reliable Low Latency Communication (URLLC) and Massive Machine Type of Communication (mMTC). Compared to well-established 4G communication systems based on LTE technologies, the requirements for these services are higher in terms of low latency 1ms (e.g. for cellular communication support of autonomous driving, remote surgery and industrial automation), high bit rate in values of Gbps. Despite the enhancements introduced for 5G NR (New Radio) with Release 15 by 3GPP, it is quite challenging and inefficient to meet all the service requirements within the same network. In order to fulfill the requirements set by International Telecommunication Union (ITU), it has been proposed Network Slicing as a possibility to set up and configure logical networks per each use case. Network Slicing (NS) is thought as a network virtualization technique that will make the network more flexible by optimising the utilisation of the infrastructure and the allocation of resources.

This thesis will focus on analysing Network Slicing and Quality of Service (QoS) Framework from 3GPP Standards and not only. Then, Media Access Control (MAC) will be analyzed by pointing out the changes for Network Slicing and QoS deployments. Last, emulations will be carried out to show the impact of dedicated network slicing and QoS in the network performance.

Keywords: 5G, Network Slicing, eMBB, mMTC, URLLC, Quality of Service, MAC layer.

Context & Motivation

PriMO-5G Project

PriMO-5G is a joint project from Korean and European partners within the HORIZON2020 framework. This project is willing to develop 5G system based on mmWave for fire-fighting scenarios. The consortium believes that recent enhancements on mobile communication technology shall improve the rescue process in Public Protection and Disaster Relief (PPDR) scenarios.

The project is focused on fire-fighting use-case and the intention is to develop technologies of mmWave access, 5G core networks and AI-assisted communications fulfilling network requirements. The project use-cases and the scenarios are described in the deliverable [38] which is a public document and can be found on the website: www.primo-5g.eu

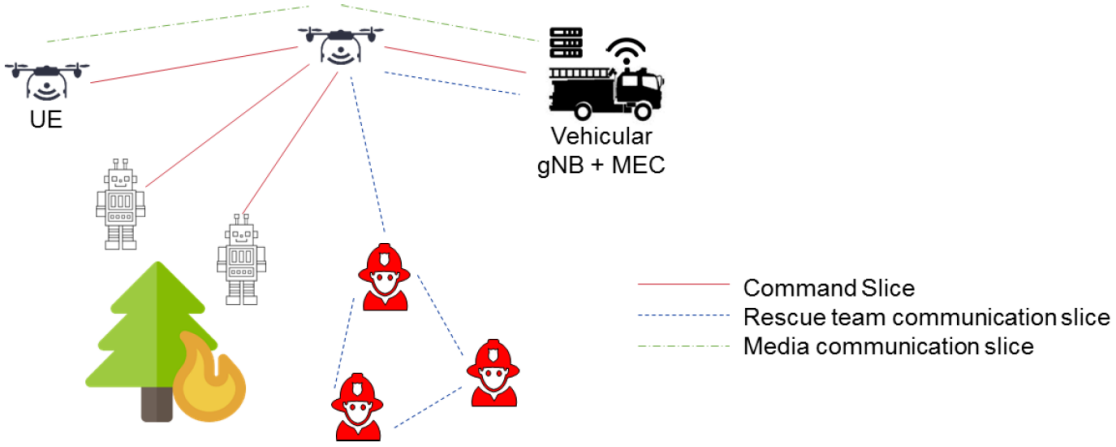

Figure 1: Fire-fighting scenario PriMO-5G. Figure 4.4 from [38]

This thesis is based on the scenario as shown in the figure 1 which represents a fire-fighting scenario in a forest where the mobile network coverage is not feasible. Therefore, vehicular gNB shall provide network connectivity for the rescue team, drones and robots. Furthermore, it shall contain computing resources or the so called Multi-access Edge Computing (MEC) in order process the traffic. The drones will receive the control commands from gNB and will send live streaming traffic towards gNB. In this way it will become easier to manage the situation and instruct the rescue team and the robots.

A brief description of each components shown in the Figure 1 is given:

Firefighters are equipped with UEs for voice connection with the gNB

Robots will be simple UEs and can be used for management or data transmission. The robots that are equipped with moving gNBs will be managed remotely to allocate network and radio slices required for other devices with low latency requirements such as UAVs.

UAVs, similar to robots, can be simple UEs or Unmanned Aerial Vehicles (UAVs) that carry aerial gNB which then should be managed from URLLC controller.

Incident commander is assumed to locate at a fire engine which is capable of gNB, MEC and URLLC network slice management functions. If the fire break-out is in a large scale, the incident commander will need connectivity to the control centre. Thus, the fire truck should also have wireless backhaul to the closest available infrastructure.

From the above scenario are driven 2 main use-cases enhanced Mobile Broadband Communication (eMBB) and Ultra Reliable Low Latency Communication (URLLC). The mentioned use-cases have different network requirements in terms of delay, packet loss and throughput which makes it difficult to meet the service requirements if they run in one network. Therefore, it was needed to analyze the performance of Network Slicing and / or Quality of Service on fulfilling the use-case KPIs.

Based from the described scenario, the research work of this thesis has been motivated.

Dedikuar prindërve të mi...

Contents

Abbreviation

ITU	International Telecommunication Union
3GPP	The 3rd Generation Partnership Project
NGMN	Next Generation Mobile Networks
O-RAN	Open-Radio Access Network
5G NR	5th Generation New Radio
CN	Core Network
5G CN	5G Core Network
LTE	Long-Term Evolution
eMBB	enhanced Mobile Broadband Communication
mMTC	massive Machine Type Communication
URLLC	Ultra Reliable Low Latency Communication
KPI	Key Performance Indicator
CP	Control Plane
UP	User Plane
RAN	Radio Access Network
C-RAN	Cloud - Radio Access Network
NG-RAN	Next Generation - Radio Access Network
CU	Centralized Unit
DU	Distributed Unit
RU	Radio Unit
O-CU	O-RAN Centralized Unit
O-DU	O-RAN Distributed Unit
O-RU	O-RAN Radio Unit
PNF	Physical Network Function
VNF	Virtual Network Function
CNF	Core Network Function
ML	Machine Learning
AI	Artificial Intelligence
MEC	Multi-access Edge Computing
AMF	Access & Mobility Management Function
SMF	Session Management Function
UPF	User Plane Function
UDR	Unified Data Repository
UDM	Unified Data Management
PCF	Policy Control Function
NRF	Network Repository Function
NEF	Network Exposure Function
NSSF	Network Slice Selection Function
AUSF	Authentication Server Function
NSSAI	Network Slice Selection Assistance Information

S-NSSAI	Single-Network Slice Selection Assistance Information
SBA	Service-Based Architecture
NFV	Network Function Virtualization
SDN	Software Defined Networks
AN	Access Network
non-AN	non Access Network
RRC	Radio Resource Control
SDAP	Service Data Adaption Protocol
PDCP	Packet Data Convergence Protocol
SDAP	Service Data Convergence Protocol
RLC	Radio Link Control
RLC-AM	Radio Link Control Acknowledged Mode
RLC-UM	Radio Link Control Unacknowledged Mode
RLC-TM	Radio Link Control Transparent Mode
MAC	Media Access Control
RB	Radio Bearer
PRB	Physical Radio Bearer
DRB	Data Radio Bearer
SRB	Signaling Radio Bearer
RRM	Radio Resource Management
PDU	Protocol Data Unit
SDU	Service Data Unit
TA	Tracking Area
DCCH	Dedicated Control Channel
DTCH	Dedicated Traffic Channel
CCCH	Common Control Channel
PCCH	Paging Control Channel
BCCH	Broadcast Control Channel
PUCCH	Physical Uplink Control Channel
PUSCH	Physical Uplink Shared Channel
PDCCH	Physical Downlink Control Channel
PDSCH	Physical Downlink Shared Channel
PRACH	Physical Random Access Channel
PBCH	Physical Broadcast Channel
DPDK	Data Plane Development Kit
LC	Logical Channel
LCID	Logical Channel Identifier
LCP	Logical Channel Prioritization
ARQ	Automatic repeat request
HARQ	Hybrid Automatic Repeat Request
BSR	Buffer Status Report
PHR	Power Headroom Report
SPS	Semi-Persistent Scheduling
TDM	Time Division Multiplexing
FDM	Frequency Division Multiplexing
TDD	Time Division Duplexing
MCS	Modulation and Coding scheme
DCI	Downlink Control Information
TB	Transport Block
TBS	Transport Block Size
CQI	Channel Quality Indicator
RSRP	Reference Signal Receive Power

BLER	Block Error Rate
SINR	Signal to Noise Ratio
CA	Carrier Aggregation
C-RNTI	Cell-Radio Network Temporary Identifier
Int-RNTI	Interruption Network Temporary Identifier
DRX	Discontinuous Reception
RR	Round Robin
WRR	Weighted Round Robin
PFS	Proportional Fair Scheduling
EDF	Earliest Deadline First
QFI	Quality of Service Flow Identifier
5QI	5G Quality of Service Indicator
RQI	Reflective Quality of Service Indicator
RDI	Reflective Quality of Service to Data Radio Bearer Indicator
GFBR	Guaranteed Flow Bit Rate
non-GFBR	non-Guaranteed Flow Bit Rate
ARP	Allocation and Retention Priority
MDBV	Maximum Data Burst Volume
MFBR	Maximum Flow Bit Rate
MPLR	Maximum Packet Loss Rate
AMBR	Aggregate Maximum Bit Rate
PDB	Packet Delay Budget
PER	Packet Error Rate
PDR	Packet Detection Rule
PFD	Packet Flow Description
RQA	Reflective QoS Attribute
DSCP	Differentiated Services Code Point
NS	Network Slicing
NSSAI	Network Slice Selection Assistance Information
S-NSSAI	Single Network Slice Selection Assistance Information
SST	Slice/Service Type
SD	Slice Differentiator
TempID	Temporal Identifier
NSI	Network Slice Instance
NSSI	Network Slice Subnet Instance
TN	Transport Network
OAM	Orchestration and Management
CSMF	Communication Service Management Function
NSMF	Network Slice Communication Function
NSSMF	Network Slice Subnet Management Function
NFV-MANO	Network Function Virtualization Management and Orchestration
RIC	Radio Access Network Intelligence Controller
UAVs	Unmanned Aerial Vehicles
PPDR	Public Protection and Disaster Relief

List of Figures

List of Tables

Chapter 1

Introduction

Due to the huge success of of 4th generation of cellular communication systems based on LTE technologies, 5G is designed to not only provide higher data rate, but also to support a diverse set of new services coming from vertical industry. International Telecommunication Union (ITU) has set diverse application-specific requirements to be supported by 5th Generation Mobile Network.

Three main service categories have been defined as 5G Use-Cases as shown in Figure 1.1: Enhanced mobile broadband (eMBB), ultra-reliable and low-latency communication (URLLC) and massive machine-type communication (mMTC).

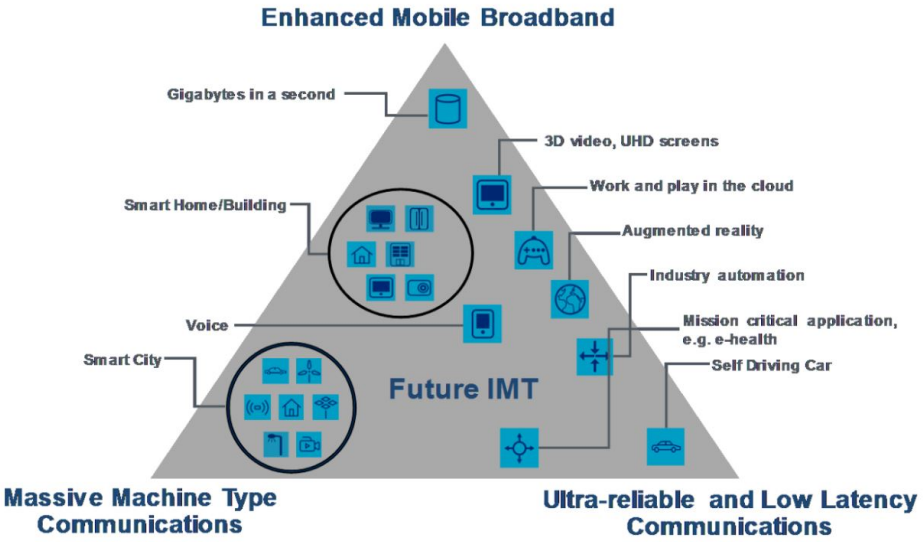

Figure 1.1: 5G Use-Cases defined by ITU.

Enhanced Mobile Broadband Communication (eMBB) is a data-driven use case for serving mobile users requiring high data rates.

Ultra Reliable Low Latency Communication (URLLC) is designed to serve for mission critical communications with strict latency and reliability requirements such as remote surgery or autonomous vehicles.

Massive Machine Type Communications (mMTC) is thought to serve a very

large number of devices such as Internet of Things (IoT) and sensors, which may only send data sporadically. These type of devices do not require high bit rate, but there are restrictions in terms of network coverage underground(e.g. basement), extended battery-life to several years.

Recent European research projects, stakeholders, as well as the 3GPP, have issued several 5G architectures that support network slicing.

Th 3rd Generation Partnership Project (3GPP) is a collaborative project between seven telecommunication organizations, focused on developing standards that are globally acceptable. The project is composed of 3 main work groups that aim researching on different areas: Technical Specification Group for Radio Access Networks (TSG RAN), for Core Network & Terminal (TSG CT) and for Service & System Aspects (TSG SA). The standards are published in form of Technical Specifications and Technical Reports. The latter contains proposed solutions for specific topic that need to be further discussed and analyzed. Whereas, the former provides the standardized solutions obtained after several group meetings and discussions. The documents are available online and free of charge. The first phase of 3GPP 5G specifications has been completed with Release 15, December 2017. The second phase in Release 16 is scheduled for completion in 2020.

Other stakeholders such as O-RAN Alliance and Next Generation Mobile Networks (NGMN) are as well working on this topic and have published their proposals and specifications on respective websites.
O-RAN Alliance is an organization driven mostly by network operators such as : such as China Mobile and AT&T aiming to standardize the protocol interfaces needed for real network implementation. Their solutions are adopted from 3GPP specifications and willing to facilitate open in a sense of exchangeable infrastructure components.

The services to enable eMBB, URLLC and mMTC use cases have different needs in terms of network performance, such as low latency access, high communication reliability and the support of massive number of devices. Therefore, the static point-to-point architecture currently in use by 4G is no more sufficient to fulfill the diverse set of requirements. This led to think of designing the network architecture in a more flexible way. In order to make the network more flexible and configurable, has been proposed the separation of Control Plane Functionalities from the User Plane Functionalities. This separation is possible by enabling network virtualization techniques: Network Function Virtualization (NFV), Software Defined Network (SDN) and cloud computing. SDN allows to have a centralized control plane that manages the user plane functions deployed on different scenarios. Through network virtualization it is possible to run the network functions such as routers, firewalls on a general-purpose server and configure the network based on the operator needs. Switching from a point-to-point network architecture towards a Service Based Architecture (SBA), makes easier the separation of Control Plane Network Functions from User Plane Network Functions. This configurable and flexible architecture allows enabling of logical networks through common network functions and improve the service. Network slicing is thought as a solution of creating virtual networks that share the network infrastructure. Hence guarantee the heterogeneous requirements of the vertical industry.

On the other hand, Radio Access Network Architecture has been defined by 3GPP in Release 15 and can be implemented as C-RAN (Cloud-RAN). It is an architecture where C-RAN is a centralized, cloud computing-based architecture for radio access networks. C-RAN, which is a composition of Centralized Unit and Distributed Unit, has managed to separate the control plane from the user plane also. The radio protocol stack between

the UE and gNB, has been defined for control plane and user plane. Based on the defined architecture, RRC, SDAP and PDCP sublayer will run at Centralized Unit (CU) component. Whereas the lower radio protocol stack layer will run at Distributed Unit (DU). Since 3GPP has not defined the so called Fronthaul Interface between the Distributed Unit and the Radio Unit, making it a vendor depended protocol interface, O-RAN Alliance is researching to set the standards.

The following characteristics highlight the enhancements of 5G System from Long Term Evolution (LTE): Service Based Architecture, Network Slicing, New Radio.

Service-based Architecture (SBA) is based on the premise that 5G will support a multitude of services and very different performance requirements. The difference compared to P2P architecture used in LTE, stands at the control plane. In LTE architecture, components were connected with predefined connections. Whereas in SBA has been applied another elements has been introduced: NF Repository Function (NRF) which connects network elements following the service model approach. Each network element query an NF Repository Function (NRF) to discover and communicate with each other. Hence, the network is more flexible to add new network element instances and services.

Software Defined Network (SDN) enables the separation of control plane from user plane. 5G aims to create a centralized control plane that manages and configures user plane network functions. In order to provide the means of isolated control functions, 5G architecture has been designed with a separate User Plane Function (UPF) , Access and Mobility Function (AMF) and Session Management Function (SMF).

Network Slicing is expected to be a key component of future 5G networks due to variety of services supported by 5G. Enabling configurable logical networks with functionality specific to the service or customer. Some scenarios that might need specific slices are: autonomous car communication require low latency whereas video - streaming requires high throughput.

New Radio (NR) has been introduced as a radio access technology able to meet the requirements of each aforementioned use-cases. It promises to provide better user experience through serving in ultra low latency and higher throughput. Several features have been enhanced comparing with LTE such as: utilization of higher frequency bands, massive MIMO, flexible numerology. Transmitting on higher frequency bands does not have good channel conditions. On the other hand, flexible numerology permits enabling variable duration slots. Doing so, is possible to assign short duration slots to User Equipment (UEs) serving URLLC traffic.

The thesis is organized as follows: Chapter 2 provides an overview of fundamental features of 5G System, covering the core network architecture, enhancements on New Radio compared to LTE. Chapter 3 described the functionalities and structure of radio protocol stack starting from Radio Resource Control Layer down to Physical Layer. Chapter 4 and 5 investigate on Network Slicing and Quality of Service (QoS) framework, respectively. Chapter 6 provides an analysis of required features by MAC Protocol Layer for supporting Network Slicing and Quality of Service. Chapter 7 describes the system environment used for testing, the emulation scenarios and the obtained results. Finally, chapter 8 describes conclusions on obtained results and future scope.

Chapter 2

5G System Overview

2.1 Introduction

5G is thought as a dynamic and flexible framework willing to serve diverse services with specific network requirement. 5G is designed to provide flexibility on Core network and RAN deployment such as Cloud-RAN and Service-Based Architecture. In the following subsections will be described aspects of 5G System that are relevant to understand Network Slicing and Quality of Service approach. The references for this chapter are mainly three technical specifications (TS) from 3GPP, as following: TS 23.501 [1], TS 38.300 [11] and TR 28.801 [4].

2.2 Service-Based Architecture

Core Network (CN) architecture used in LTE has some drawbacks in terms of complexity and scalability such as: a lot of interfaces to be established and diverse protocols used per connection. Hence, it is not possible to support diverse services required in 5G.

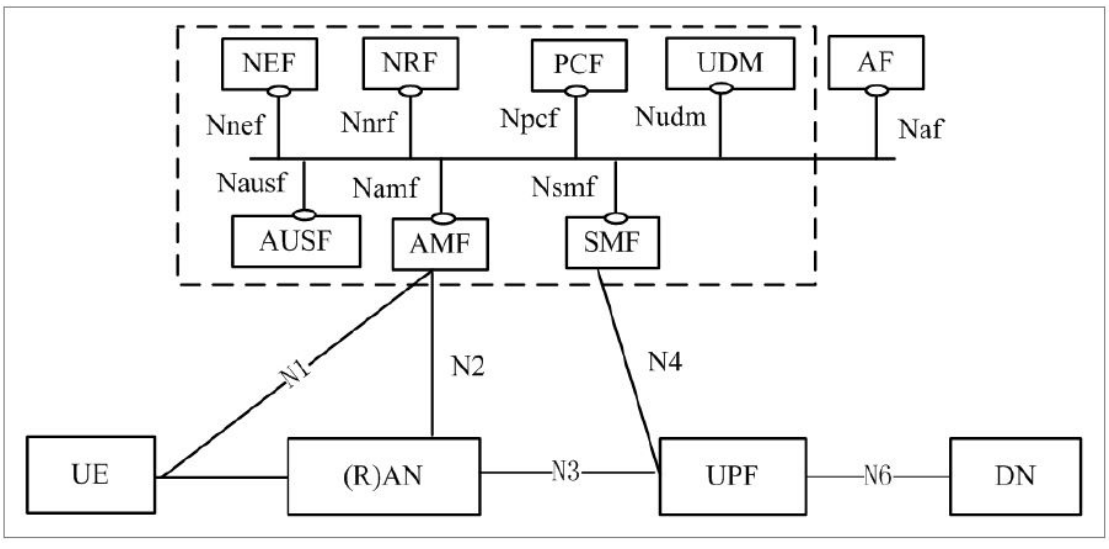

Figure 2.1: 5G - Service-Based Architecture. Figure 4.2.3-1 from [1]

Therefore, another flexible solution has been proposed by 3GPP (Figure 2.1), Service Based Architecture (SBA). It represents a cloud-networking approach where Core Network Functions (CNFs) that reside in Control Plane (CP), are connecting through a

service model interface. Each NF can behave as a service producer by offering services to other NFs, or service customer by requesting services from other NFs. The component that makes the communication possible is Network Repository Function (NRF). NFs from Control Plane register their services on NRF and query NRF for services from other network components [29].

The figure 2.1 shows the network elements composing 5G Core Network (5G CN) and some changes from LTE Network can be noticed. Hence, Access and Mobility Management Function (AMF) and Session Management Function (SMF) in LTE were combined in one component. Furthermore, Network Repository Function (NRF), as above mentioned, is introduced only on 5G.

2.2.1 Core Network Functions

New Core Network Function have been introduced along with SBA such as Network Repository Function (NRF) and Network Exposure Function (NEF). Other CNF have the same functionalities as in LTE but the names have been slightly updated. Below are described the most significant network elements: [1]:

Access and Mobility Management Function (AMF): Manages registration procedure of UEs, mobility procedure, access authentication and authorization. It terminates N1 and N2 interface from UE and NG-RAN by providing also the communication bridge between Access Stratum and Non Access Stratum.

Session Management Function (SMF) provides Session Management e.g. Session Establishment, Session Modification, release. It terminates N4 interface with UPF.

User Plane Function (UPF) is in charge of handling packet inspection, routing and forwarding. Furthermore, it support QoS rules per flow and reports to SMF traffic usage.

User Data Management (UDM) interacts with control plane functions for providing subscribed used data. The main functionality is handling of user identification.

Policy Control Function (PCF) provides policy rules to Control Plane function(s). Moreover, it interacts with UDR for retrieving structured data required to take policy decisions.

Network Repository Function (NRF) supports service discovery. It is in charge of updating the subscribed NFs with NF services newly registered or deregistered.

Network Exposure Function (NEF) is in charge of exposing capabilities and events for e.g. 3rd party, Application Functions, Edge Computing. Moreover, it interacts with UDR for retrieving structured data.

Authentication Server Function (AUSF) manages the authentication of connections arriving from 3GPP access or untrusted non-3GPP access.

2.2.2 Service Model Connection

Application Programmable Interface provides flexibility for enabling new services and to establish connections between network components. HTTP/2 JSON has been standardized by 3GPP as the format to be used in 5G. Consequently, HTTP methods such as: GET, POST etc will be used to register and/or request services from NRF. Huawei has described this new connection model in a simplified manner (refer Figure 2.2). Network Repository Function (NRF) is placed in the middle in order to maintain the communication with other Service Provider and Service Consumer NFs.

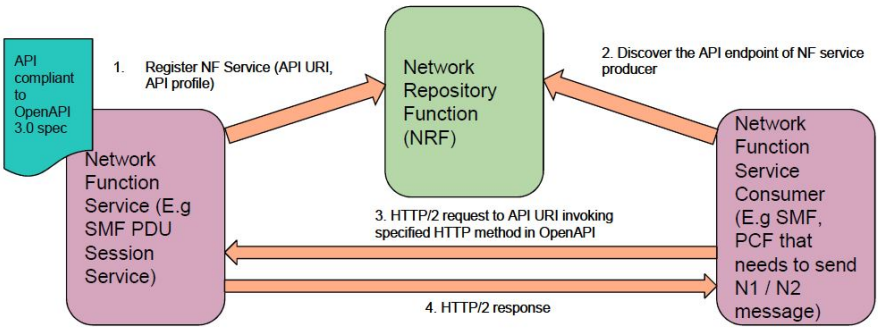

Figure 2.2: NRF-NF communication. Figure from [28]

2.3 Cloud RAN

Operators transform networks using a network architecture based on data center (DC) in which all functions and service applications are running on the cloud DC [30].

Plenty of solutions propose to run core network or access network on cloud instead of physical dedicated devices. Advantages are evaluated in terms of deployment flexibility, deployment cost, improvement of user experience etc. In order to select the best solution, is required an evaluation of split options proposed by 3GPP.

3GPP RAN is composed of 2 main components: Centralized Unit (CU) and Distributed Unit (DU). Higher layer functionalities which are non time sensitive such as: Radio Resource Control (RRC), Packet Data Convergence Protocol (PDCP) are located at Centralized Unit (CU) side. Whereas at Distributed Unit (DU), are located time sensitive functionalities such as scheduling, modulation and coding etc. Furthermore, based on 3GPP's architecture, Radio Unit is also included at DU. However, other stakeholders such as O-RAN, have also introduced a separate RU component. With this architecture, new interfaces are also requires for maintaining the connectivity between function components. F1, E1 and NG interfaces have already been defined by 3GPP.

2.3.1 RAN Split Options

In order to ease the fulfillment of the requirements for different use-cases such as eMBB, URLLC and mMTC, 3GPP has proposed several RAN protocol split options between CU and DU. The split option are analyzed in terms of time synchronization and Transport Network requirements. More details can be found in section 11 of [6]. Interfaces based on the RAN functional split options can be grouped as following:

Low Layer Split (LLS) between radio and central RAN functions: Being specified in various forums and standards bodies. The study done by 3GPP can be found on [8]. Low Layer Split (LLS) includes the Options 68 proposed on the figure 2.3

High Layer Split (HLS) between distributed and central RAN functions: Being specified in 3GPP as F1 for gNB and being studied in 3GPP as V1 for eNB. High Layer Split (HLS) includes the Options 15 proposed on the figure 2.3

CU-CP and CU-UP (control/user plane split within central RAN functions): Being specified in 3GPP as E1 for gNB.

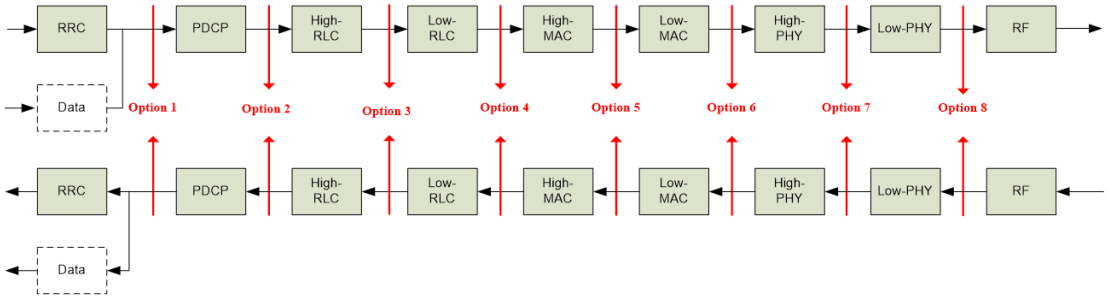

Figure 2.3: RAN Functional Split Options - 3GPP. Figure 11.1.1-1 from [6]

3GPP has specified the interfaces for RAN Split Option-2 by separating control from user plane (Figure 2.4). Split Option-2 or so called PDCP-RLC split, consists on implementing SDAP & PDCP sublayer at Centralized Unit. Whereas the lower layer of protocol stack shall be implemented at Distributed Unit (DU). The interfaces required for implementation of Split option 2 have been defined by 3GPP: E1, F1-Control, F1-User, NG-Control and NG-User [9]. F1 interface connects gNB Centralized Unit (gNB CU) with gNB Distributed Unit (gNB DU). Since the Control Plane and the User Plane at Centralized Unit (CU) can be further split, 3GPP has introduced also F1-C and F1-U interface. F1-Control (F1-C) interface connects the CU-CP with Distributed Unit of the same gNB. Whereas F1-U connects CU-UP with the Distributed Unit. Furthermore, the control plane & user plane split requires another interface to enable the connectivity between CU-CP & CU-UP. Therefore, E1 interface has been defined by 3GPP and is described on TS 38.460 & TS 38.463. Whereas F1 interface has been defined on 3GPP TS 38.470 & TS 38.473.

Lower layer protocols (Higher Physical Layer, MAC, RLC), have time-sensitive functionalities e.g. scheduling, retransmission of lost packets. Transport network requirements are strongly related with the split option. Hence, for split option 3,4,5 (refer to Figure 2.3) the latency requirements on transport network connecting CU-DU will be very critical. Since lower layer functionalities such as scheduler, modulation, coding, encoding/decoding, re-transmission are time sensitive, those layers shall be co-located for having low latency. Implementing Split option 6 or 7 will require another interface between DU and RU, which has not been standardized from 3GPP. In order to meet the low latency requirements when RU is located further from DU, fiber infrastructure is required at this interface.

The figure 2.4 shown gNB architecture. NG-RAN consists of a set of gNBs connected

to 5GC via the NG (Next Generation) interface which is . gNBs can be interconnected through the Xn interface. A gNB may consist of a gNB-CU and one or more gNB-DU(s) where one gNB-DU is connected to only one gNB-CU.

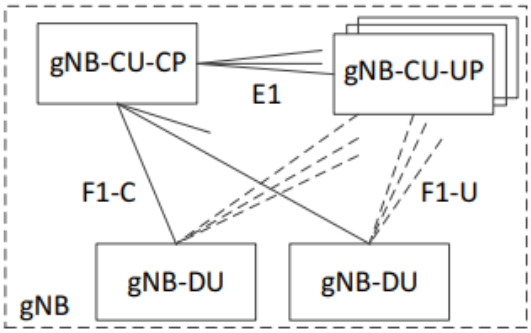

Figure 2.4: RAN Architecture Split Option 2 - 3GPP. Figure 6.1.2-1 from 3GPP TS 38.401 [9]

2.3.2 RAN deployment scenarios

NG-RAN is a logical architecture and the real deployment scenario is left to vendor decision based on service requirements and preferences. 3GPP has defined the framework with the components and the required interfaces. The individual functional entities Radio Unit (RU), Distributed Unit (DU), Centralized Unit User Plane (CU-UP) and Centralized Unit Control Plane (CU-CP) entities may be place at different physical locations according to operator requirements, physical site constraints and transport network topology, latency and capacity limitations. Hence, Centralized Unit (CU) & Distributed Unit (DU) can be co-located in the same Physical Network Function/Virtual Network Function (PNF/VNF) or not based on service requirements. Furthermore, depending on the selected placement, the transport network has different requirements in terms of bandwidth or latency. Placements of functional components is flexible and should be selected based on service requirements such as latency, bandwidth etc.
3GPP has studied NG-RAN deployment options for Split Option-2 and have proposed 3 implementation scenarios that are briefly described below (refer to section 6 of [7]).

Scenario 1: This scenario represents the basic case for CU-DU split with dedicated Centralized Unit - Control Plane (CU-CP) and Centralized Unit - User Plane (CU-UP) parts which may be located in one common or separated central entities. The Centralized Unit - Control Plane (CU-CP) is centralized to coordinate the operation of several Distributed Units (DUs). The CU-UP is centralized to provide a central termination point for UP traffic in dual-connectivity (DC) configurations (see Figure 2.5). This deployment solution may be applied for eMBB use case and dual-connectivity scenario.

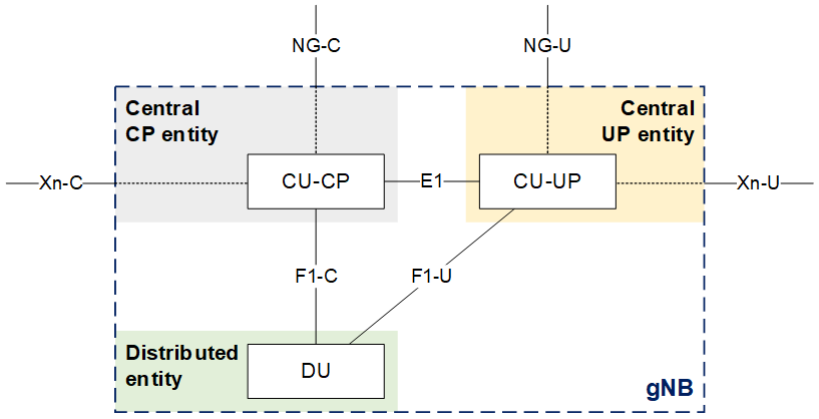

Figure 2.5: RAN Deployment Scenario 1 - 3GPP. Figure 6.2.2-1 from 3GPP TS 38.806 [7]

Scenario 2: In this second scenario the Centralized Unit - Control Plane (CU-CP) is deployed in a distributed manner and co-located with the DU. The Centralized Unit - User Plane (CU-UP) is centralized to provide a central termination point for User Plane (UP) traffic in dual-connectivity (DCs) configurations. In this scenario, the latency of the control signalling toward the UE and F1-C signalling is reduced as the Centralized Unit - Control Plane (CU-CP) is co-located with the Distributed Unit (DU) (see Figure 2.6).

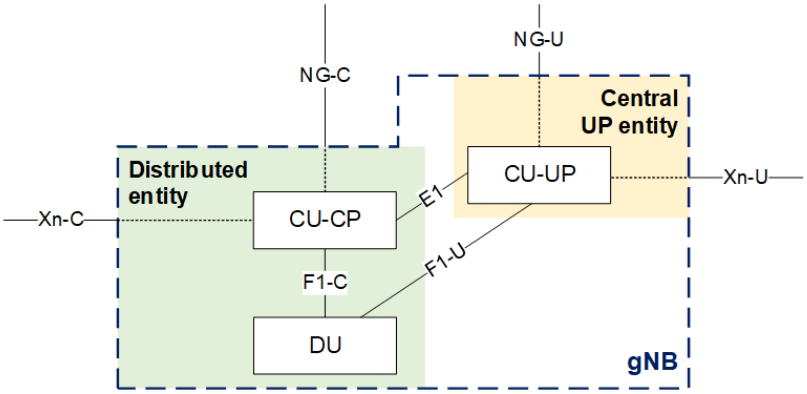

Figure 2.6: RAN Deployment Scenario 2 - 3GPP. Figure 6.2.2-2 from 3GPP TS 38.806 [7]

Scenario 3: Centralized Unit - Control Plane (CU-CP) is centralized to coordinate the operation of several Distributed Units (DUs). The Centralized Unit - User Plane (CU-UP) is distributed and co-located with a single Distributed Unit (DU). This scenario fits best to low latency use-case.

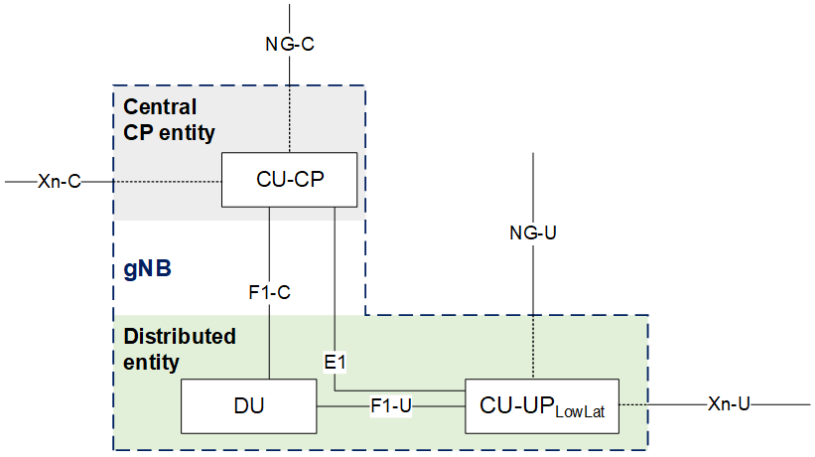

Figure 2.7: RAN Deployment Scenario 3 - 3GPP. Figure 6.2.2-3 from 3GPP TS 38.806 [7]

As a conclusion, the placement of NG-RAN components can be selected based on the slice type.

2.3.3 RAN architecture by O-RAN

Open-Radio Access Network (O-RAN) Alliance is another international organization working on defining the standards for NG-RAN and Network Slicing. While 3GPP organization is mainly driven by the activities of infrastructure vendors such as Ericsson, Nokia and Huawei, O-RAN is a consortium where the network operators search for solutions that fit their needs on top of 3GPP standards. They aim to set the standards for implementing NG-RAN by defining interfaces left open by 3GPP. The proposed architecture is compliant with split option 2 and split option 7-2x (Figure 2.8). Furthermore, they introduced a separate O-RAN Radio Unit (O-RU) component and the Open Fronthaul Interface for connecting it with O-RAN Distributed Unit (O-DU). The interfaces F1-C, F1-U and E1 have been defined by 3GPP in compliance with RAN Split Option-2. The O-RU component includes RF module, filters, power amplifier and digital-to-analog converter. Radio Protocol Layers such as Radio Link Control, Media Access Network, Higher Physical Layer are located at O-DU component. whereas at O-CU component are implemented the higher layer of radio protocol stack : Packet Data Convergence Protocol, Service Data Adaption Protocol and Radio Resource Control.

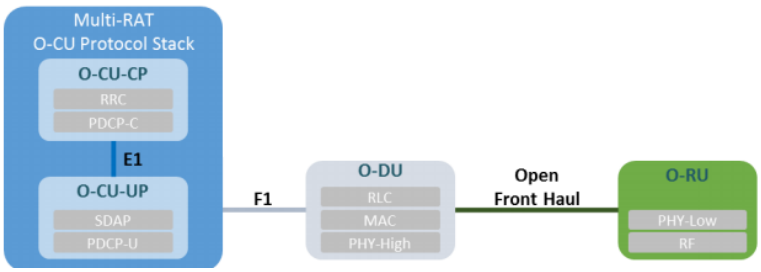

Figure 2.8: O-RAN Radio Access Network Architecture. Figure 4 from [27]

O-RAN has specified the interfaces for connecting the above mentioned RAN com-

ponents when those are implemented as separate PNF/VNF. However, when multiple Network Functions are running on one single PNF/VNF, it is up to the operator to decide whether to enforce the O-RAN interfaces between the embedded Network Functions.

Since the placement of entities is flexible, O-RAN has also analyzed the service impact of different deployment scenarios that can be found in the following technical report [27]. Furthermore, O-RAN has defined the components and interfaces for managing and orchestrating the radio resources and those are described in the following chapters 4.3.3.

2.4 Summary

The chapter describes the 5G System and the key features needed for Network Slicing. Decoupling the Control plane from User plane and enabling a service based model for network component interactions, transforms the static architecture into an extensible one. Usage of API interfaces and JSON data model converts it into a programmable network. On the other hand, separation of RAN functionalities in 2 components (CU & DU), gives more freedom to implementation. 3GPP has standardized the interfaces for Split Option 2 (so called PDCP-RLC split) and has separated the control plane from user plane. 3GPP RAN architecture is composed of Centralized Unit (CU), Distributed Unit (DU) and transport Network. Whereas O-RAN Alliance, has further extended 3GPP's work by introducing another component O-RU (O-RAN Radio Unit) and the interface connecting the later with O-DU (O-RAN Distributed Unit), Fronthaul Interface. However, the deployment of those components is left to vendor's choice.

Chapter 3

RAN Protocols

3.1 Introduction

3GPP has released specification about New Radio (NR) and New Generation - Radio Access Network (NG-RAN) Overall Description. In order to investigate the impact of Network Slicing and QoS on MAC Layer, it is necessary to analyze each Radio Protocol Layer. The following subsections describe the protocol structure and functionalities for every Radio Protocol Layers on 5G New Radio. In the figures 3.1 & 3.2 are represented the radio protocol stack for user plane and for control plane at UE and gNB side. The User Plane protocol stack handles data traffic where the Service Data Adaption Protocol translates the Application traffic into QoS Flows and Access Network resources. Whereas, the control plane stack handles the control traffic and signaling for establishing, managing, controlling the connection, session etc. Radio Resource Control is the main protocol of control stack since it configures the sublayers for every incoming connections.

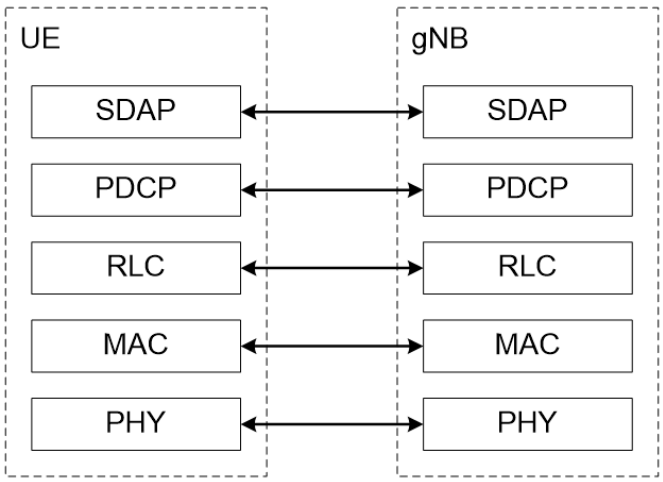

Figure 3.1: User Plane Protocol Stack. Figure 4.4.1-1 from 3GPP TS 38.300 [11]

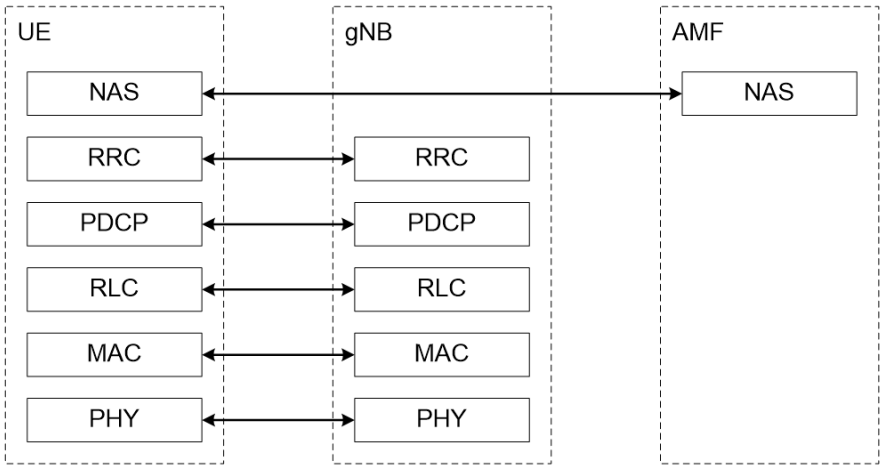

Figure 3.2: Control Plane Protocol Stack. Figure 4.4.2-1 from 3GPP TS 38.300 [11]

On this chapter will be described the Radio Protocol Layer structure and functionalities from Radio Resource Control (RRC) Layer down to Physical Layer.

3.2 Radio Resource Control - RRC

Radio Resource Control (RRC) Layer is the brain of Radio Protocol stack which is in charge of establishing the connections with the Core Network, configuring Lower Sublayer Entities for handling the upcoming connection.
The specification [5] describes the procedures and messages specified for the UE equally and apply to the RN for functionality necessary for the RN (see section 5 of [5]).

The UE can be either in RRC_CONNECTED state or in RRC_INACTIVE state when an RRC connection has been established (Figure 3.3). If this is not the case, i.e. no RRC connection is established, the UE is in RRC_IDLE state. The RRC_INACTIVE is a new state introduced recently in 5G for maintaining the connection established with Core Network when the UE is not sending/receiving data traffic. The RRC states can further be characterised as follows:

RRC_IDLE: UE does not have a connection established with the network

RRC_INACTIVE: UE has the connection active with Core Network (RRC of gNB) but it is in power save mode and cannot send or receive data traffic.

RRC_CONNECTED: UE has the connection active with Core Network and can send/receive data traffic.

Therefore, UEs that generate sporadic traffic could maintain active the connection with the Core Network and minimize the delay for sending data traffic. In figure 3.3 is shown the state transition diagram for the UE. From RRC_IDLE state the UE shall pass directly to RRC_Connected state where the session is established with the Core Network and UE can generate data. Furthermore, based on the RRC control messages (Release or Release with Suspend, see section 5.3 from [5]), the UE can change back to RRC_IDLE state or RRC_INACTIVE state.

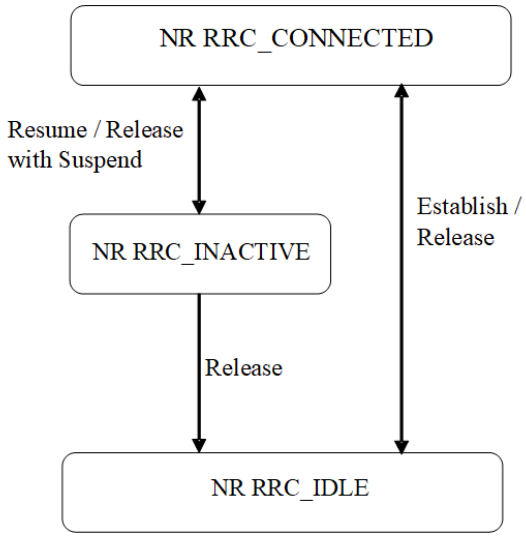

Figure 3.3: UE state machine and state transitions in NR. Figure 4.2.1-1 from [5]

The procedures that UE can handle on each state are further described in the section 4.2.1 from [5].

RRC controls the messages and Non-Access Stratum (NAS) messages are transmitted to lower sublayer through particular channels known as : Signaling Radio Bearers (SRBs). 4 types of Signaling Radio Bearers (SRBs) have been defined from 3GPP and each SRB handles specific RRC messages (see section 4.2.2 from [5]:

- **SRB0** is used for RRC messages using the Common Control Channel (CCCH) logical channel.
- **SRB1** is used for RRC messages (which may include a piggybacked Non-Access Stratum (NAS) message) as well as for Non-Access Stratum (NAS) messages prior to the establishment of SRB2, all using DCCH logical channel.
- **SRB2** is used for Non-Access Stratum (NAS) messages, all using Dedicated Control Channel (DCCH) logical channel.
- **SRB3** is used for specific RRC messages when UE is in E-UTRAN New Radio – Dual Connectivity(EN-DC) or New Radio - Dual Connectivity (NR-DC), all using Dedicated Control Channel (DCCH) logical channel.

RRC Layer has several functionalities which are listed below:

- Broadcast of system information (see section 5.2 of [5])
- RRC connection control which includes Paging procedure, Establishment/modification/suspension/resumption/release of RRC connection etc. The procedures and the messages exchanged can be found in the section 5.3 from [5]
- Inter-Radio Access Technology (RAT) mobility including transfer of RRC context information (see section 5.4 from [5]).
- Measurement configuration and reporting that includes establishment/modification/release of measurement configuration, measurement reporting etc. The above mentioned procedures are described in the section 5.5 of [5]

- Transfer of UE radio access capability information which is described in the section 5.6 of [5]).

Furthermore, 3GPP has also defined the Protocol Data Unit (PDU) format for every message that shall be generated from RRC Layer including the parameters that can be configured. The Protocol Data Unit (PDU) Formats are described in details in the section 6 of [5].

3.3 Service Data Adaption Protocol - SDAP

Service Data Adaption Protocol (SDAP) is a radio protocol introduced in 5G for handling QoS traffic used only for user traffic. Service Data Adaption Protocol (SDAP) performs mapping of Data traffic into QoS Flows in both directions, based on QoS Rules that have been signaled by Session Management Function (SMF) or derived by Reflective QoS.
The protocol structure is defined by 3GPP as a framework (see section 4.2 of [12]). Service Data Adaption Protocol (SDAP) Sublayer should be configured by Radio Resource Control (RRC) Layer. Furthermore, several SDAP Entities are located on Service Data Adaption Protocol (SDAP) Sublayer. One Service Data Adaption Protocol (SDAP) Entity is associated with one PDU Session. Hence, the UE can have configured more than one Service Data Adaption Protocol (SDAP) Entities and more than one PDU Session (see section 4.2.2 of [12]). One possible sublayer structure proposed by 3GPP is represented in figure 3.4. However, the above mentioned solution does not restrict the implementation decision.

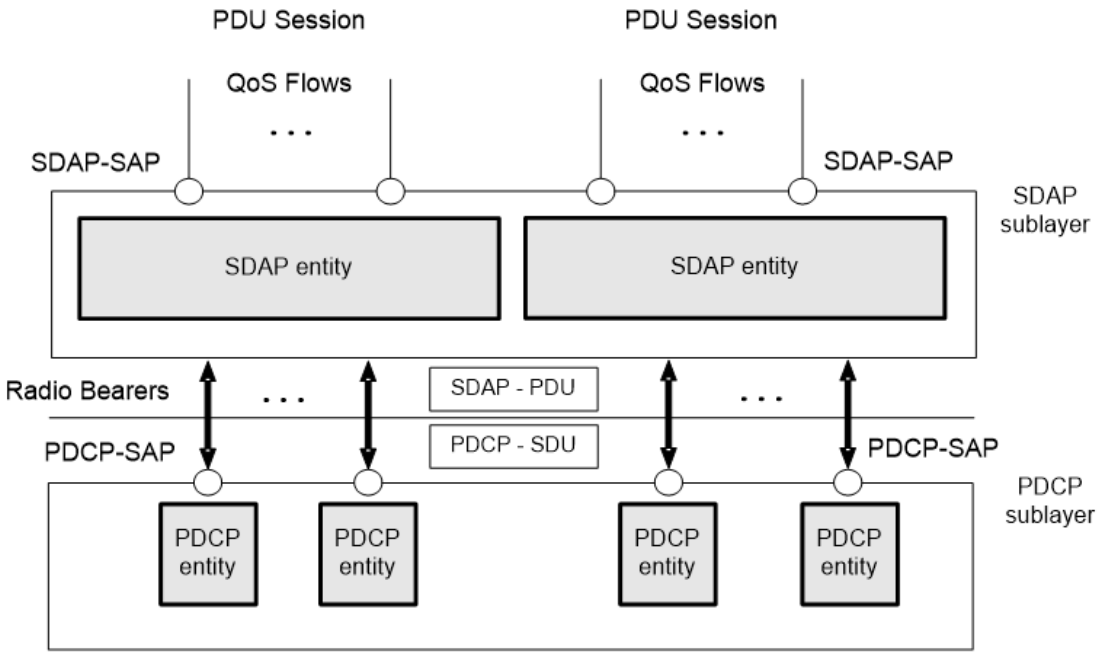

Figure 3.4: SDAP Sublayer - structure view. Figure 4.2.1-1 from [12]

Furthermore, it applies the mapping of QoS Flows into Data Radio Bearers (DRBs) (access-specific resources) based on rules configures by Radio Resource Control (RRC) Layer. When Service Data Adaption Protocol (SDAP) is processing the received traffic, it adds the protocol header by including the corresponding QFI and may optionally

include the RQI field (see section 5.2 of [12]).

A SDAP entity receives/delivers SDAP Service Data Units (SDUs) from/to upper layers and submits/receives SDAP data Packet Data Units (PDUs) to/from its peer SDAP entity via lower layers. As it is represented in the Figure 3.5, the SDAP Entity at the transmitter side, receives a SDAP SDU from upper layers, it maps the received SDUs into QoS Flows and further into Data Radio Bearers (DRBs) based on the rules configured by RRC Layer. Furthermore, the transmitting SDAP Entity constructs the corresponding SDAP data PDU by adding the header if the feature is configured by RRC Layer (see section 5.2.1 of [12]) and submits it to lower layers (PDCP Layer). Whereas, SDAP Entity at the Receiver side receives SDAP data PDU from lower layers (PDCP Layer), it retrieves the SDAP data SDU and delivers it to the upper layer (see section 5.2.2 of [12]).

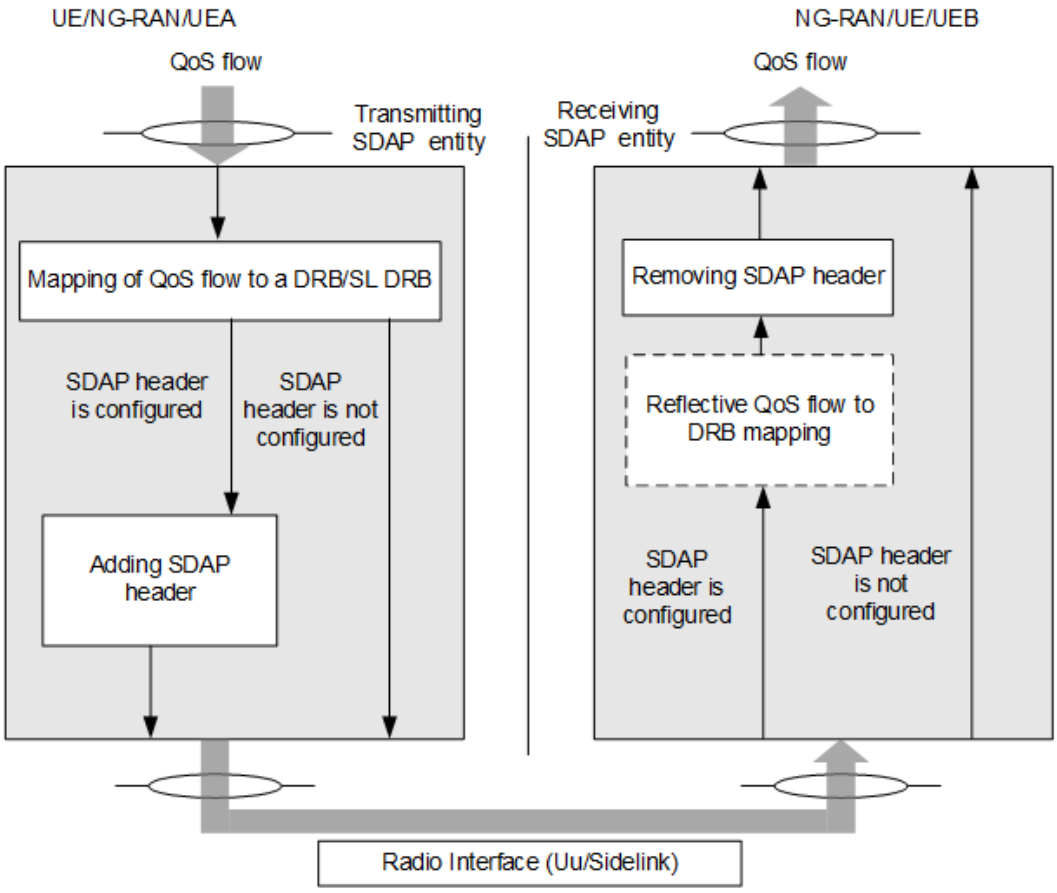

Figure 3.5: SDAP Layer - functional view. Figure 4.2.2-1 from [12]

SDAP PDU is categorized in three types: Control PDU, Data PDU and End-Marker control PDU (see section 6.1 of [12]). SDAP data PDU may contain the following fields on the protocol header:

- QoS Flow Identifier (QFI), indicating the ID of the QoS flow to which the SDAP PDU belongs. QFI field length is 6 bits.
- Reflective QoS Indicator (RQI), indicates whether Non-Access Stratum (NAS) should be informed of the updated of SDF to QoS flow mapping rules. RQI field

length is 1 bit.

- Reflective QoS flow to DRB mapping Indication (RDI),indicates whether QoS flow to DRB mapping rule should be updated at the UE side and its length is 1 bit.

3.4 Packet Data Convergence Protocol - PDCP

Packet Data Convergence Protocol (PDCP) Sublayer performs header compression for user data, ciphering/deciphering and integrity protection. Furthermore, it may perform packet duplication for reliable transmissions and in-order-delivery to higher layers. The Packet Data Convergence Protocol (PDCP) Sublayer is configured by Radio Resource Control (RRC) Layer. The Packet Data Convergence Protocol (PDCP) Sublayer is used for Radio Bearers mapped on the following type of logical channels:

- Dedicated Control Channel (DCCH) that serves Signaling Radio Bearers : SRB 1, SRB 2, SRB 3. SRB0 is not mapped to Dedicated Control Channel (DCCH), therefore is not handled by Packet Data Convergence Protocol (PDCP) Sublayer.
- Dedicated Traffic Channel that serves only data traffic for one PDU Session.

Each RB (except for Signaling Radio Bearer0 - SRB0) is associated with one Packet Data Convergence Protocol (PDCP) Entity. Each PDCP entity is associated with one, two, three, four, six, or eight Radio Link Control (RLC) entities depending on the RB characteristic (e.g uni-directional/bi-directional or split/non-split) or Radio Link Control (RLC) Transmission mode (see section 4.2.1 of [13]).

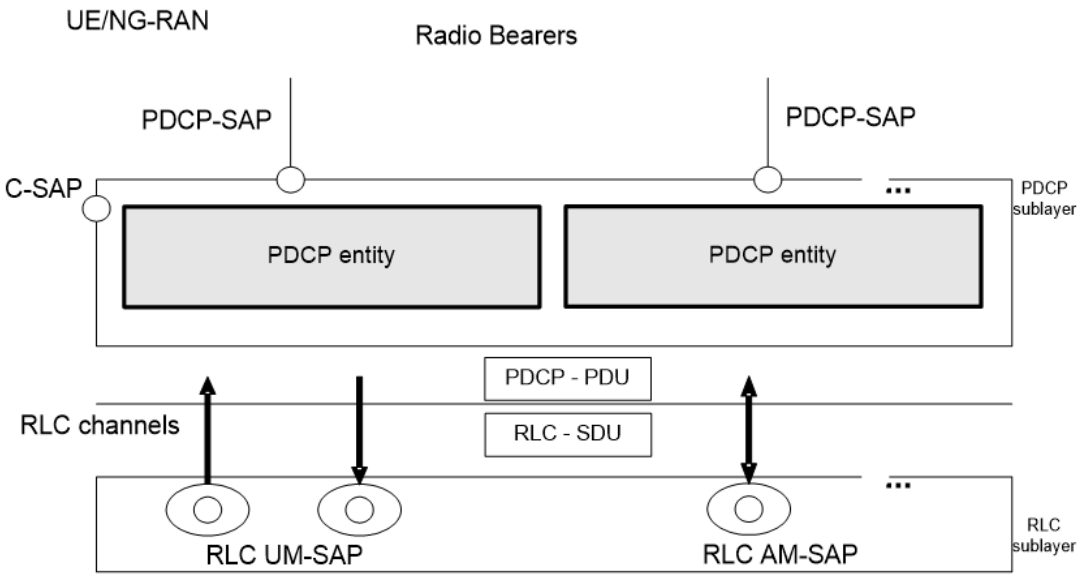

Figure 3.6: PDCP Layer - structure view. Figure 4.2.1-1 from [13]

The Packet Data Convergence Protocol (PDCP) entities are located in the Packet Data Convergence Protocol (PDCP) sublayer. Several PDCP entities may be defined for a UE. Each PDCP entity is carrying the data of one radio bearer (refer to Figure 3.6). A Packet Data Convergence Protocol (PDCP) entity is associated either to the control plane or the user plane depending on which radio bearer it is carrying data for.

Protocol structure is organized in Packet Data Convergence Protocol (PDCP) Entities and one PDCP Entity is associated with only one Data Radio Bearer (DRB). 3GPP has proposed a possible Layer Structure to be implemented, shown on Figure 3.4. One Packet Data Convergence Protocol (PDCP) Entity is associated with one RB and one or two Radio Link Control (RLC) Entities, depending on the Transmission Mode of Radio Link Control (RLC) Entity which will be covered on the next section.

A Packet Data Convergence Protocol (PDCP) entity associated with one Data Radio Bearer (DRB) can be configured by RRC Layer to use header compression, integrity protection, ciphering and deciphering, data duplication, in order delivery etc. When PDCP Entity at the transmitting side, received SDU from upper layers, it has to generate Packet Data Convergence Protocol (PDCP) PDU based on the features configured by RRC for RB, and send it to lower layers (RLC). Whereas, the Packet Data Convergence Protocol (PDCP) Entity at the receiving side, receives PDCP SDUs from lower layers (RLC), it has to retrieve the Packet Data Convergence Protocol (PDCP) PDU as it is shown on the Figure 3.7 and send it to upper layers (see section 5.2 of [13]).

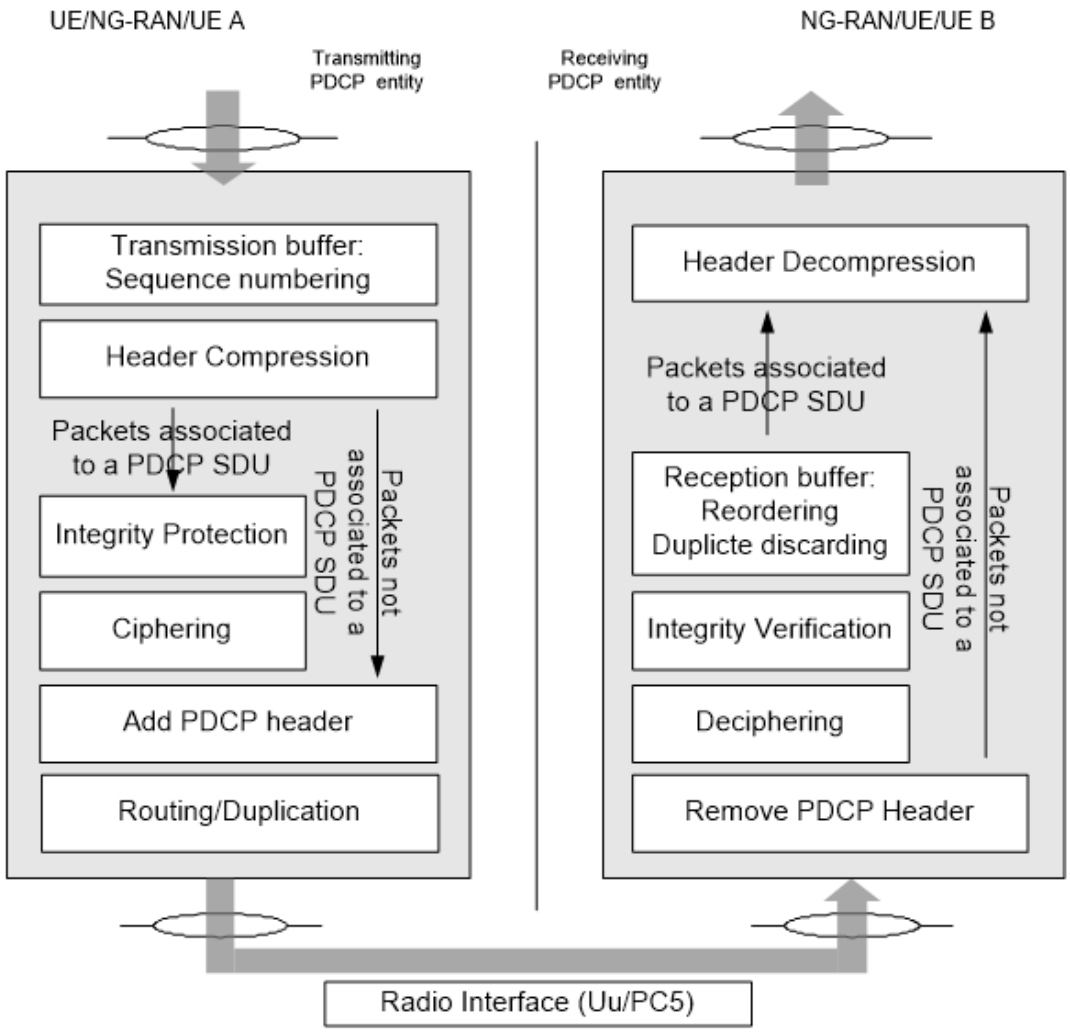

Figure 3.7: PDCP Layer - functional view. Figure 4.2.2-1 from [13]

PDCP Duplication is a new feature introduced in 5G NR consists of sending twice the same SDU (see section 5.11 of [13]). Duplication at Packet Data Convergence Protocol

(PDCP) consists in submitting the same PDCP PDUs twice: once to the primary Radio Link Control (RLC) entity and a second time to the secondary RLC entity (Figure 3.8). With two independent transmission paths, packet duplication therefore increases reliability and reduces latency and is especially beneficial for URLLC services. (see section 16.1.3 of [11]). Packet Duplication shall be configured independently for each Radio Bearer (RB) by Radio Resource Control (RRC) Layer.

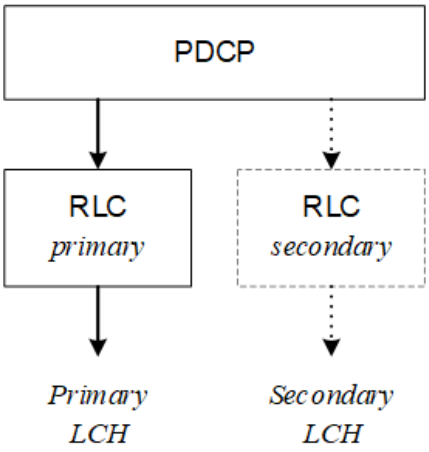

Figure 3.8: PDCP Layer - Packet Duplication. Figure 16.1.3-1 from [11]

3.5 Radio Link Control - RLC

Radio Link Control (RLC) Protocol handles data transmission to upper/lower layers. The structure of the protocol is organized in Radio Link Control (RLC) Entities and 1-4 Radio Link Control (RLC) Entities may be associated with one Data Radio Bearer (DRB). The number of entities depends on the configured transmission mode. The entities are configured by RRC Layer when establishing a Data Radio Bearer (DRB). The following transmission modes can be configured per each entity:

Transparent Mode (TM) is used only for control messages mapped to Broadcast Control Channel (BCCH), DL/UL Common Control Channel (CCCH) and Paging Control Channel (PCCH) (see section 4.2.1.1 of [14]).

Unacknowledged Mode (UM) is only for user data mapped to Dedicated Traffic Logical Channel (DTCH) (see section 4.2.1.2 of [14]).

Acknowledged Mode (AM) is used for user data mapped to Dedicated Traffic Logical Channel (DTCH) and for control data mapped to DCCH Logical Channel (see section 4.2.1.3 of [14]).

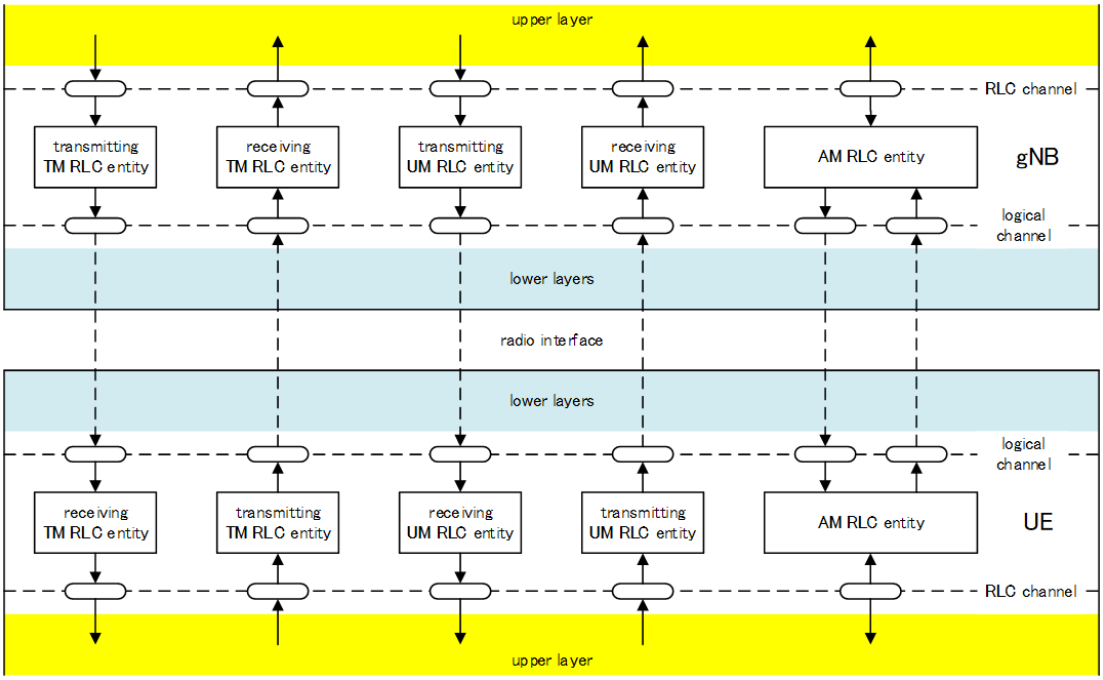

Figure 3.9: Overview model of the RLC Sublayer. Figure 4.2.1-1 from [14]

On Transparent and Unacknowledged transmission mode are required two Radio Link Control (RLC) entities to be associated with the Data Radio Bearer (DRB) where one entity is the receiving one and the other entity is the transmitting one. Whereas on Acknowledged transmission mode only one Radio Link Control (RLC) Entity is associated with the Data Radio Bearer (DRB) because that Entity has to process the received SDUs and send back the feedback (ACK).

Radio Link Control (RLC) Entities have different functionalities on each transmission mode. As an instance, AM performs error correction through Automatic Repeat Request (ARQ) technique, duplication detection and error detection. Therefore, AM may be used for reliable transmission whereas UM may be used for transmitting big payloads. Furthermore, Radio Link Control (RLC) protocol does not provide anymore in-order-delivery to higher layers. On 5G, this particular feature is supported only by Packet Data Convergence Protocol (PDCP) layer.

Each RLC SDU is used to construct an RLC PDU without waiting for notification from the lower layer (i.e. by MAC) of a transmission opportunity. In the case of UM and AM RLC entities, an Radio Link Control (RLC) SDU may be segmented and transported using two or more RLC PDUs based on the notification from the lower layer. RLC PDUs are submitted to lower layer only when a transmission opportunity has been notified by lower layer. Functionalities per each type of Radio Link Control (RLC) entity are listed in the Table 3.1 (see section 4.4 of [14]):

Functionalities	TM	UM	AM
Segmentation and reassembly of RLC SDUs	No	Yes	Yes
Re-segmentation of RLC SDU segments	No	No	Yes
Protocol error detection	No	No	Yes
Error correction through ARQ	No	No	Yes
Duplicate detection	No	No	Yes
RLC SDU discard	No	Yes	Yes

Table 3.1: Functions Summary - RLC Entities

The detailed description of each transmission mode can be found on section 5.2 of [14].

3.6 Media Access Control - MAC

MAC Architecture is based on MAC Entities. MAC sublayes at UE may be configured with one MAC Entity or two MAC Entities when the Secondary Cell Group (SCG). RRC Layer is in control of MAC sublayer and configures the MAC Entities. 3GPP has proposed in the one possible implementation solution for MAC Sublayer (Figure 6.1).

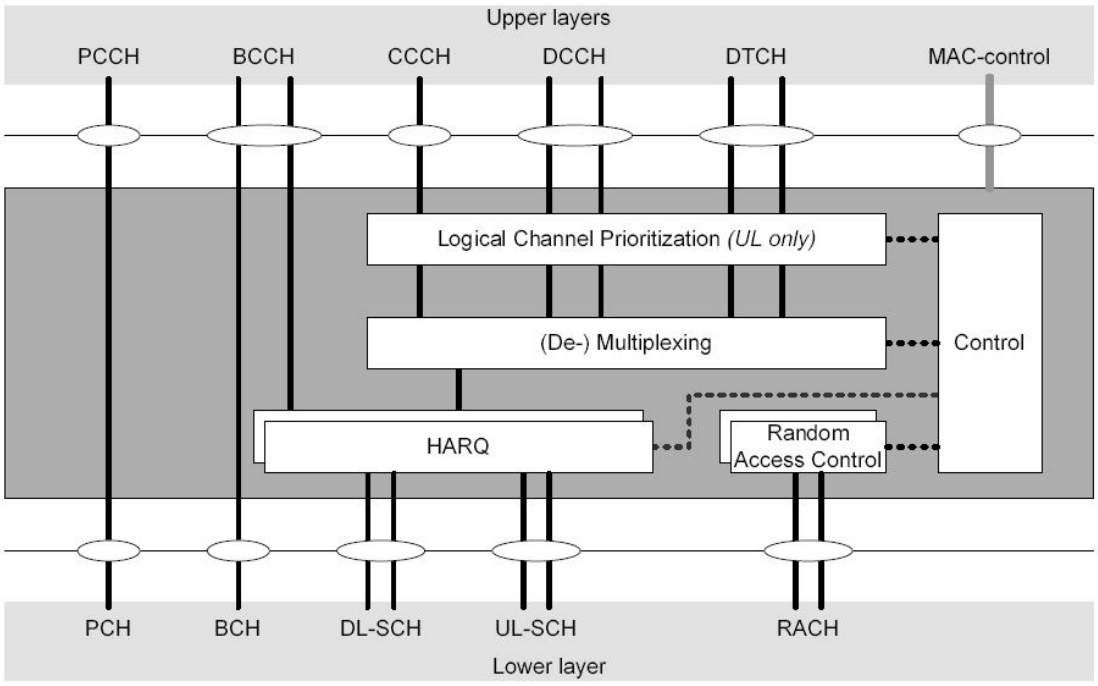

Figure 3.10: MAC Architecture. Figure 4.2.2-1 from [15]

The MAC sublayer operates on the transport channels and logical channels The MAC sublayer uses the transport channels to send/receive data from Layer 1 and are categorized as Downlink Channels or Uplink Channels (see table 4.5.2-1 [15]). Whereas on Logical channels it provides data transfer services to/from RLC Sublayer. Logical channels are categorized as Control Channels and Traffic Channels (see table 4.5.3-1 [15].

MAC Entity has several functionalities (see section 4.4 from [15]):

- Mapping between logical channels and transport channels. 3GPP has set the rules of mapping the traffic from logical channels to transport channel and the opposite and can be found on section 4.5.4 from [15].
- Scheduling information reporting
- Multiplexing of MAC SDUs from one or different logical channels onto transport blocks (TB) to be delivered to the physical layer on transport channels
- Demultiplexing of MAC SDUs to one or different logical channels from transport blocks (TB) delivered from the physical layer on transport channels;
- Error correction through HARQ
- Logical Channel prioritisation.

Media Access Control Layer is in charge of multiplexing and/or demultiplexing the SDUs from/to Logical Channel, scheduling the transmission grants and transmit the data to the Physical Layer. Scheduling process is done at MAC Layer of gNB. Depending on the scheduling algorithm which is implementation dependent, the resources and transmission grants are assigned to UEs in UL and DL. 3GPP has set the rules and requirements only for MAC Layer at UE side. Whereas the MAC Layer implementation at gNB side is completely left to implementation. The MAC entity of the UE handles the following transport channels:

- Broadcast Channel (BCH)
- Downlink Shared Channel(s) (DL-SCH)
- Paging Channel (PCH
- Uplink Shared Channel(s) (UL-SCH)
- Random Access Channel(s) (RACH)

The MAC Sublayer provides data transfer services on logical channels listed below:

- Broadcast Control Channel
- Paging Control Channel
- Common Control Channel
- Dedicated Control Channel
- Dedicated Traffic Channel

Both for uplink and downlink, the MAC entity is responsible for mapping logical channels onto transport channels. This mapping depends on the multiplexing that is configured by RRC. Uplink and Downlink mapping are represented in the table 4.5.4.2-1 and 4.5.4.3-1 of [15].

MAC Entity shall handle several procedures such as: Random Access Procedure, DL-Scheduling data transfer, UL -Scheduling data transfer, HARQ operation, Discontinuous Reception (DRX), MAC CE handling etc. (see section 5 of [15]). In the following sections is described in more details Scheduling procedure and Logical Channel Prioritization.

MAC Protocol handles each functions and procedure through the use of Control Elements (CEs) which are type of messages exchanged by MAC layer. Each CE has a specific format type, length and functionality defined by 3GPP (see section 6.1.3 from [15]).

The list of defined Control Elements are presented in table: Table 6.2.1-1 (LCID Values for DL-SCH) and Table 6.2.1-2 (LCID Values for UL-SCH) of [15].

3.7 Physical Layer

Physical Layer in 5G has came with new enhancements for fulfilling the requirements in terms of low latency and high throughput. Millimetric - wave (mm-wave) frequency bands are additionally considered for 5G NR, in contrast with LTE. Hence, there is a delimitation of frequency ranges (FR) arranged as shown in Table 3.2 The frequency range number 2 (FR2) corresponds to the operating mm-wave frequency range in Release 15.

Frequency range destination	**Corresponding frequency range**
FR 1	410 MHz - 7125 MHz
FR 2	24250 MHz – 52600 MHz

Table 3.2: Definition of frequency ranges. Table 5.1-1 from [19]

3GPP Release 15 ensures flexible resource allocation through a combination of various strategies. The first one to mention is the numerology concept, also called subcarrier configuration μ. The Subcarrier spacing defines the frame and lattice structure of a waveform. The variety of transmission numerologies (Table 3.3) aims at providing support for devices having different transmission capabilities. The subcarrier configuration μ= 0 corresponds to the subcarrier configuration mandated for LTE, with subcarrier spacing (SCS) 15 kHz. Whereas further numerologies configure SCSs that are multiples of LTE's SCS by a factor 2μ.

μ	$\Delta f = 2^{\mu} \cdot 15\,[kHz]$	**Cyclic Prefix**	**Supported for data**	**Supported for synch**
0	15	Normal	Yes	Yes
1	30	Normal	Yes	Yes
2	60	Normal, Extended	Yes	No
3	120	Normal	Yes	Yes
4	240	Normal	No	Yes

Table 3.3: Supported transmission numerologies. Table from 3GPP 38.300 [11]

The numerology configurations $\mu = 0, 1, 2$ are thought for FR1, whereas $\mu = 2, 3, 4$ are available for FR2.

The time structure gets also configured by the selected numerology. The duration of a radio frame is defined as $T_f = 10ms$. A subframe for NR corresponds to a time interval that lasts $T_{sf} = 1ms$, i.e. 10 subframes compose a frame. The number of OFDM symbols per slot remains 14 for normal cyclic prefix in all numerologies, whereas it changes to 12 symbols for an extended cyclic prefix. It is then clear that the number of OFDM symbols per slot is doubled with respect to that of LTE. The number of slots per subframe is accordingly halved with respect to LTE, to keep up with the ratio between the duration of a frame and a subframe. In the following table 3.4 it is represented how the subframe changes based on numerology.

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 3.4: Number of OFDM symbols per slot, slots per frame, and slots per subframe for normal cyclic prefix. Table 4.3.2-1 from 3GPP 38.211 [17]

Two multiplexing schemes to switch between UL and DL are recommended for 5G, namely Frequency Division Multiplexing (FDM) and Time Division Multiplexing (TDM). FDM lets both gNB and UE transmit and receive simultaneously by using one frequency band for UL and another frequency band for DL. This is referred to as paired spectrum operation in Release 15. TDD, in contrast, lets gNB and UE transmit/receive at different times using the same frequency band. This is referred to as unpaired spectrum operation in Release 15. The time distribution of DL or UL opportunities is configured as a pattern in a per-symbol basis by the slot format.

The slot format assigns one of 3 uses to an OFDM symbol index:

- 'D' to a symbol that shall be used for DL communication.
- 'U' to a symbol that shall be used for UL communication.
- 'F' to a symbol that has flexibility to be used for either DL or UL communication.

Notice that the slot format can be reconfigured through signaling. There are 256 slot formats present in Release 15, with 199 of them being currently reserved. For a detailed description of 5G NR slot configuration, refer to 38.213 Section 11.1 [**?**].

Furthermore, 3GPP has defined several Modulation & Coding Schemes that can be applied in different transmission scenarios. MCS value determines the Modulation scheme (Qm) and Coding Rate (R) to be used for a particular transmission. Those values are selected from MAC Scheduler at gNB Side based on the implemented algorithm and are signaled to the physical layer through Downlink Control Information (DCI). DCI provides the UE with the necessary information such as physical layer resource allocation, power control commands, HARQ information for both uplink and downlink. 3GPP has defined several DCI messages and the corresponding formats are explained in the specification related to Physical Layer [18].
In the specification [18], have been defined 3 tables for different use-cases where Table Table 5.1.3.1-1 includes the modulation scheme (Qm) till 64 QAM, table 5.1.3.1-2 provides MCS values for achieving high throughput (256 QAM is being used) and the last table, 5.1.3.1-3 shall be used for URLLC transmission.

Based on MSC value configured for each transmission and number of transmission layers, it is possible to calculate the Transmission Block Size per each time slot (time slot depends on the selected numerology). The data from the upper layer (or MAC) given to the physical layer in LTE system is basically referred as transport block size (TBS). 3GPP has determined the TBS value in 5G in the section 5.1.3.2 from specification [16].

3.8 Summary

On this chapter is provided an overview of New Radio (NR) Protocol Stack. The sublayer architecture and the functionalities of every sublayer have been briefly described. In the Control Plane, in the Radio Resource Control (RRC) Layer has been introduced RRC_INACTIVE State during which the UE does not receive/transmit data traffic but it maintains active the connection with the Core Network. New Radio (NR) has came with new sublayer on the User Plane: Service Data Adaption Protocol (SDAP) which handles the QoS Flows and maps those into Access Network (AN) resources. Furthermore, in sequence ordering is supported only by Packet Data Convergence Protocol (PDCP) layer and not from Radio Link Control (RLC) layer as it was done on Long-Term Evolution (LTE). Every sublayer has come with enhancement for improving the data transmission in terms of low latency and reliability.

Chapter 4

Network Slicing

4.1 Introduction

This chapter will provide a description of Network Slicing, deployment scenarios and implementation solutions proposed by stakeholders. Network slicing is a technology that enables multiple logical networks on the top of a common shared physical infrastructure. Its purpose is to allow service customers to program the network based on their needs and based on service requirements.

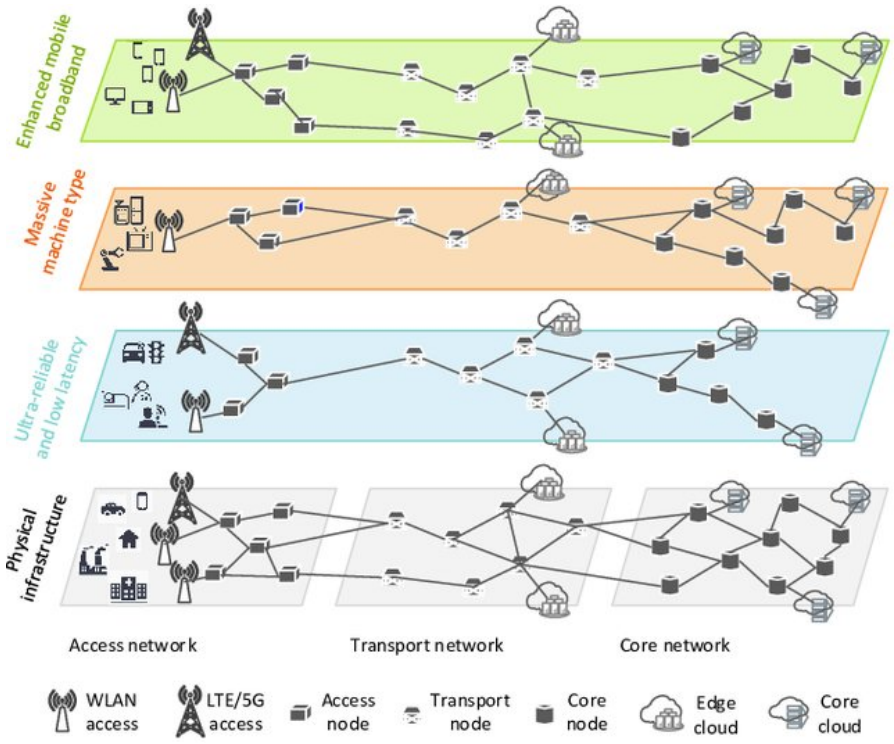

Figure 4.1: fig: 5G Network Slicing Architecture. Figure from [33]

In order to ensure an End-to-End (E2E) communication, the slice requires Core Networks parts, Radio Access Network (RAN) parts and Transport Network (TN) parts. Core Network Functions can be slice dedicated or shared between different slices and the decision is left to deployment [1]. Figure 4.1 demonstrates the architecture of End-to-End (E2E) Network Slicing for different use-cases. Network Functions can be deployed

in Physical Network Function (PNF) or Virtual Network Function (VNF), ensuring flexibility on implementation. Each network slice can be tailored to support specific applications and/or be operated by a communications provider other than the owner of the physical network infrastructure. Hence, deploying the network based on service requirements e.g. URLLC slice may require that most of Network Functions (NFs) placed at cell site and/or edge site for serving delay sensitive applications. On the other hand, mMTC slice will serve to many connected users with no latency or bandwidth requirements. Therefore, Network Functions (NFs) may be placed mostly on Core Network (CN) part.

4.2 Composition of Network Slicing

3GPP has defined the Network Slice as a composition of Network Slice Subnet Instance and Network Slice Instances (see section 4.4 [3]). These concepts are related to network slicing management part which is covered later on this chapter (see section 4.4).

4.2.1 3GPP Communication model

3GPP has defined the communication model for Network Slicing Management, introducing new concepts (refer to 3GPP specification [4]):

- *Communication Service Instance (CSI)*: represents the requirements from the communication service application.
- *Network Slice Instance (NSI)*: represents a group of NSSIs
- *Network Slice Subnet (NSS)*: a representation of the management aspects of a set of Managed Functions and the required resources (e.g. compute, storage and networking resources).
- *Network Slice Subnet Instance Core Network (NSSI)*: represents a group of network function instances that form part or complete constituents of a Network Slice Instance (NSI).

The composition of 3 Network Slice Instances (NSIs) based on 3GPP's model is represented in Figure 4.2. Each Network Slice Instance (NSI) is composed of Network Slice Subnet Instance (NSSI) from Core Network (CN), Network Slice Subnet Instance (NSSI) from Access Network (AN) and Transport Network (TN) which ensures the connectivity between NSSIs. The Network Slice Subnet Instance (NSSI) contain different Network Functions (NF) from Core Network (CN) or Access Network (AN) which are necessary for providing the requested service. Furthermore, 3GPP highlights that Network Slice Instances (NSIs) might share Access Network (AN) resources as it is shown for Network Slice Instance (NSI) B and C.

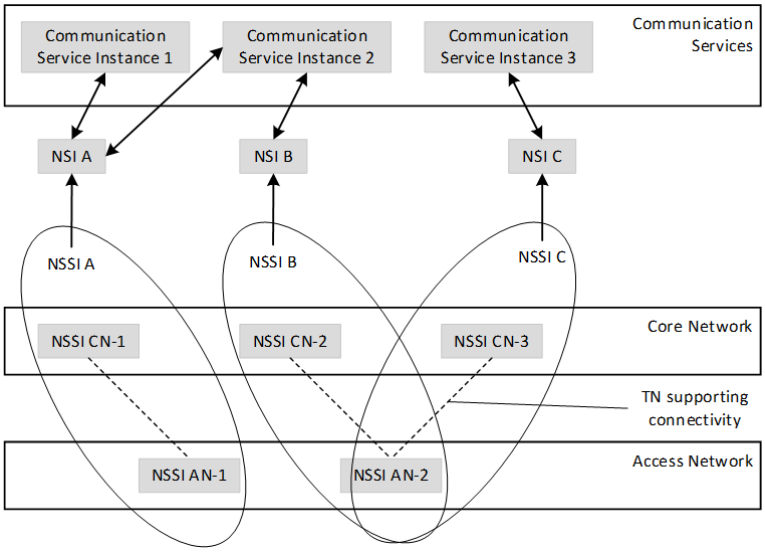

Figure 4.2: End to End services provided by NSI(s) - 3GPP. Figure 4.9.3.1 from [4]

4.2.2 Core Network Slicing

One network Slice is composed of a collection of Physical Network Functions (PNFs) and/or Virtual Network Functions (VNFs), transport network elements and Radio Access Network (RAN) resources. The deployment of Core Network Functions among the slices has been analyzed by 3GPP and other organizations. User Plane Function (UPF) is defined to be slice dedicated entity and it can be located at the Core Network or at the Access Network site, close to the Users (refer to section 6.3.3 of [1]). The deployment approach shall be selected accordingly with the use-case requirements. Hence, placing the functions at the edge would improve the latency. Therefore, making it the best solution for delay-sensitive applications.

Session Management Function (SMF) can be slice dedicated function, but it is left to implementation from network operators. Additionally, Access & Mobility Management Function (AMF) shall be shared between slice in order to manage user requests for service. 3GPP has described that users shall be able to connect with more slices simultaneously (maximum of 8 since Requested NSSAI can contain up to 8 S-NSSAIs), but only one signalling connection is maintained (see section 16.3.1 [11]). Therefore, AMF, being a control plane function, shall be in common to all network slices serving one UE (refer to section 5.15.1 of 3GPP TS 23.501 [1]).

Nokia's solution for network slice composition is shown on Figure 4.3 [31]. The approach introduces also slicing the resources for Network Exposure Function (NEF), Network Repository Function (NRF) and Policy Control Function (PCF). Furthermore, they suggest to have a dedicated Session Management Function (SMF), User Plane Function (UPF) and Application Function (AF) for handling the traffic of each slice. Whereas, Access & Mobility Management Function (AMF), Unified Data Management (UDM) and Network Slice Selection Function (NSSF) shall be shared between all active slices. However, the decision for deploying the network functions is left to implementation.

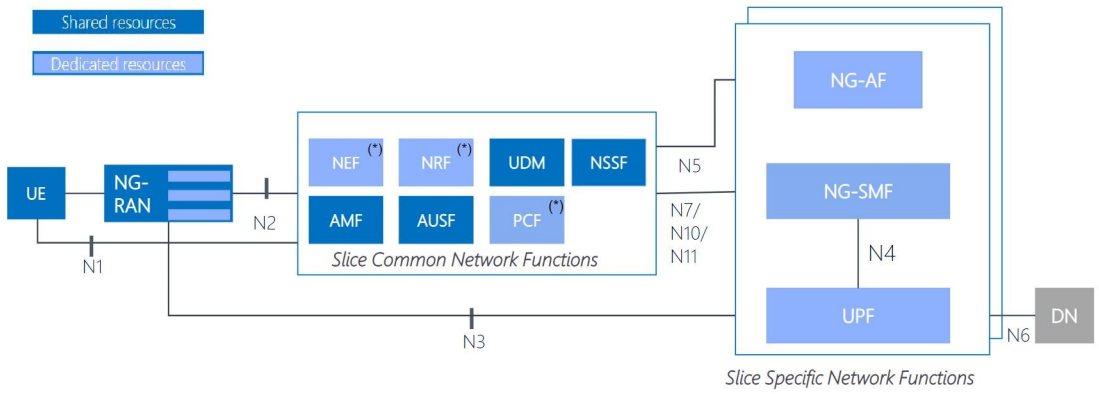

Figure 4.3: E2E Network Slicing Architecture. Figure from [31]

4.2.3 RAN Slicing

Radio Access Network (RAN) Slicing is still an on-going work for network service providers. On the other hand, stakeholders have defined key principles for supporting network slices at NG-RAN (see section 16.3.1 from [11]). Isolation among slices is a fundamental feature ensuring that the traffic of one slice does not negatively impact other slices. In a single-cell scenario, isolation can be achieved by assigning an orthogonal set of physical radio resources for a certain period to each tenant in accordance with its requirements. Nevertheless, when considering the slicing of a multi-cell RAN, isolation becomes more challenging due to interference among tenants of different cells. Therefore, flexibility is left to the tenant for ensuring slice isolation at RAN. RAN Slicing is on study item for Release 17 by 3GPP [21].

4.2.4 RAN Principles - 3GPP

3GPP has defined the key principles for supporting Network Slicing in NGRAN (see section 16.3.1 [11]). NGRAN shall be aware of slices being available in the network. Hence, a slice ID has been standardized by 3GPP, named SingleNetwork Slice Selection Assistance Information (SNSSAI), and identifies the slices within a Public Land Mobile Network (PLMN) and also in roaming case. Then, it shall be able to enable the set of network functions and features that comprise each slice. Furthermore, RAN shall select the Core Network Functions (CNFs) to serve the UE based on slice id. As aforementioned, slice isolation is a critic feature to be enabled and it can be provided through Radio Resource Management (RRM) Policies. NG-RAN shall support policy enforcement between slices and shall select the best policy for the Service Level Agreement (SLA) to each supported slice. RAN shall be able to differentiate handling for different slices for providing access control.

4.2.4.1 Slice Identification

One of the 3GPP's key requirement for supporting Network Slicing on Radio Access Network is the slide identification. Hence, a network slice ID is required to be enabled in the network. 3GPP has defined the slice ID as 32 bit variable Single-Network Slice Selection Identifier (S-NSSAI) composed of 2 fields: Slice/Service type (SST) and Slice Differentiator (SD) (Figure 4.4).

3GPP has defined 4 globally standardized Slice/Service type (SST) values for slicing in order to support interoperability between Public Land Mobile Networks (PLMNs).

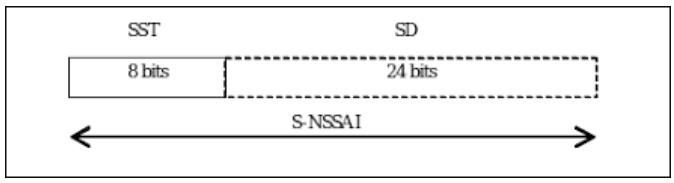

Figure 4.4: S-NSSAI structure. Figure from [11]

On the Table 4.1 are shown the standardized Slice/Service type (SST) values:

Slice/Service Type	**SST Value**	**Characteristics**
eMBB	1	Slice suitable for the handling of 5G enhanced Mobile Broadband
URLLC	2	Slice suitable for the handling of ultra-reliable low latency communication
MIoT	3	Slice suitable for the handling of massive IoT
V2X	4	Slice suitable for the handling of V2X services

Table 4.1: Standardized SST values. Table 5.15.2.2-1 from 3GPP TS 23.501 [1]

However, network operators can configure local Slice/Service Type (SST) values based on their service requirements. Slice Differentiator (SD) is applied by network vendor in order to distinguish active slices with the same Slice/Service Type (SST) (see section 16.3.1 of 3GPP TS 38.300 [11]).

Network Slice Selection Assistance Information (NSSAI) is a collection of Single-Network Slice Selection Identifiers (S-NSSAIs) and it is categorized in 3 groups:

- **Configured NSSAI**: set of S-NSSAI configured by default or by Serving PLMN at UE
- **Requested NSSAI**: set of S-NSSAI sent from UE to gNB at Initial Access.
- **Allowed NSSAI**: set of S-NSSAI allowed/provided by actual PLMN

S-NSSAI is provided by UE to gNB during Registration Procedure and is used by gNB to select Core Network elements serving the UE.

More details about Slice Identification, can be found at section 5.12.2 from 3GPP TS 23.501 [1].

4.2.5 Core Network selection

NG-RAN receives the UE's request for network connection and it may contain also a set of Requested Single-Network Slice Selection Identifier (S-NSSAI) which may be used for routing the request to the most appropriate AMF (see Figure 4.5). Before establishing the connection with the core network, NG Setup procedure takes places where gNB updates Access & Mobility Management Function (AMF) connected with it, the list of S-NSSAIs supported per Tracking Area (TA) and AMF updates with the list of S-NSSAIs supported per PLMN . Therefore, when gNB receives the requested S-NSSAIs from UE, will query the list of supported S-NSSAIs per each Access & Mobility Management Function (AMF) and select the the matching AMF. If no matching is found, gNB

shall route the request to a set of default AMFs that will further handle the connection request by assigning an allowed S-NSSAI.

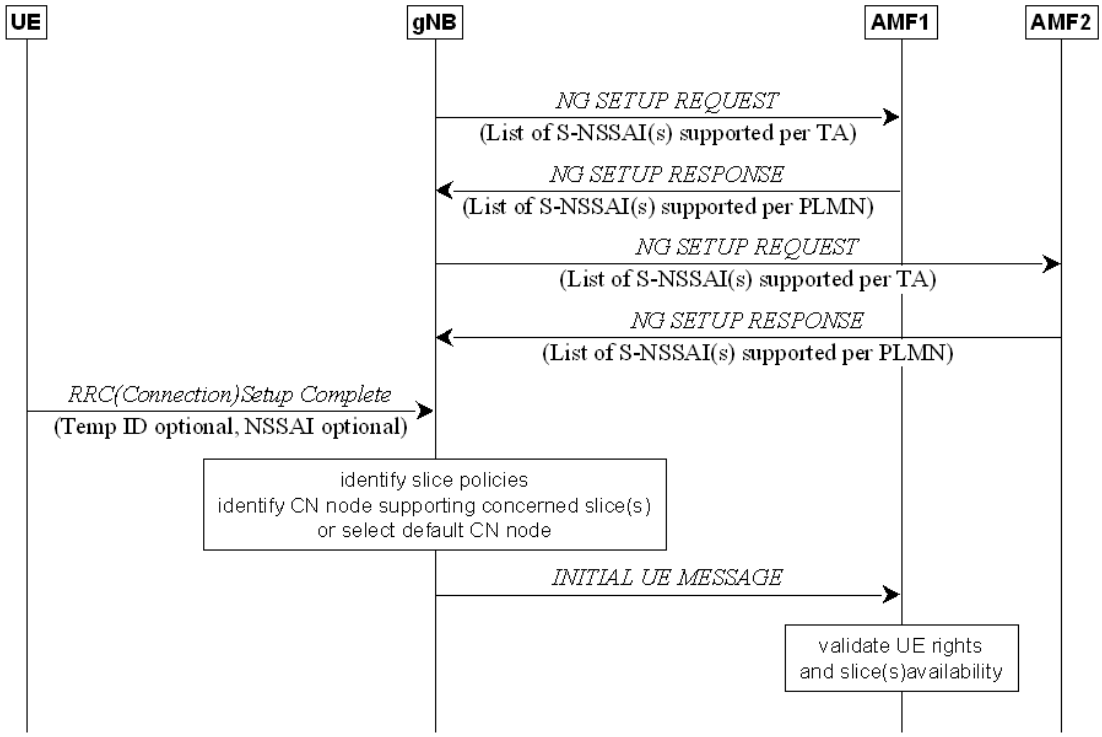

Figure 4.5: AMF Selection. Figure 16.3.4.2-1 from 3GPP TS 38.300 [11]

However, gNB may select the Access & Mobility Management Function (AMF) based on another parameter that can be provided by UE, Temporal Identifier (TempID). Temporal ID can be provided to the UE by the current network connected to and can be used for future connections with the same network. A summary of Access & Mobility Management Function (AMF) selection is provided below (Table 4.2):

TempID	**NSSAI**	**AMF Selection by NG-RAN**
not available or invalid	not available	One of the default AMFs is selected
not available or invalid	present	Selects AMF which supports UE requested slices
valid	not available or present	Selects AMF per CN identity information in Temp ID
NOTE:The set of default AMFs is configured in the NG-RAN nodes via OAM.		

Table 4.2: AMF Selection.

More details about Core Network Selection can be found at section 6.3 from 3GPP TS 23.501 [1] and section 4.2.2.2 from 3GPP TS 23.502 [2].

4.2.6 Radio Resource Management

Network Slices can be pre-configured in the network or can be enabled based on user's request on real time. Configuring the slices in advance it is an easier task since it can

be done during network configuration phase and does not have a time dependency. On the other hand, enabling on-request slices requires Orchestration and Management that can share the resources in real time and establish the Network Functions (NFs) required for the slice. Management policies are required for assigning the resources among slices based on the UE's demands. However, the Radio Resource Management (RRM) Policy is left to the service provider implementation. Therefore, several solutions have been proposed by research groups by introducing also Artificial Intelligence (AI) and Machine Learning (ML) techniques for improving the network performance [26]. Management and Orchestration solutions proposed by stakeholders, will be covered in the following section.

4.3 Orchestration of Network Slicing

Orchestration of Network Slices is an on-going research topic and the implementation of this technology is very important. Orchestration becomes a crucial part of the network when slices shall be enabled on-demand or shall be updated based on resource usage. On these cases, the Orchestrator shall have a global view of the network, in terms of resource usage, capabilities and also incoming requests from customers. Several solutions have been proposed by research groups and standardization alliances. Their solutions are concentrated in a hierarchical architecture consisting of several entities that deal with resource management, maintaining the update of network state (resource utilization, slice availability).

4.3.1 Network Slicing Management by 3GPP

3GPP has defined the requirements for Network Management system in [3] and has also proposed 3 entities for network slicing orchestration in [4]:

- *Communication Service Management Function (CSMF)*: Responsible for translating the communication service related requirement to network slice related requirements.

- *Network Slice Communication Function (NSMF)*: Responsible for management and orchestration of NSI.

- *Network Slice Subnet Management Function (NSSMF)*: Responsible for management and orchestration of NSSI.

The relation between the above-mentioned entities is represented in the following picture (Figure 4.6):

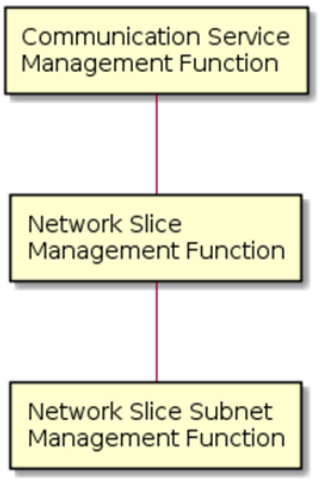

Figure 4.6: Network Slice related management functions - 3GPP. Figure 4.10.1 from 3GPP TR 28.801 .[4]

Network Slice Instances life-cycle management has been defined by ETSI & 3GPP. Network Slice Instance goes through phases shown in the Figure 4.7. In the **Preparation** phase, are performed network slice design, capacity planning, evaluation of network functions etc, required for Network Slice Instance (NSI) creation. In the second phase of **Commissioning**, the Network Slice Instance (NSI) is created including the resource allocation and configuration. When the resources are allocated and the Network Slice Instance (NSI) is enabled, is shall be activated and super visioned for ensuring the Key Performance Indicators (KPIs). These processes are included in the third phase of **Operation**. The last step, **Decommissioning**, consists of decommissioning of non-shared resources and removing the slice specific configurations for terminating the Network Slice Instance (NSI).

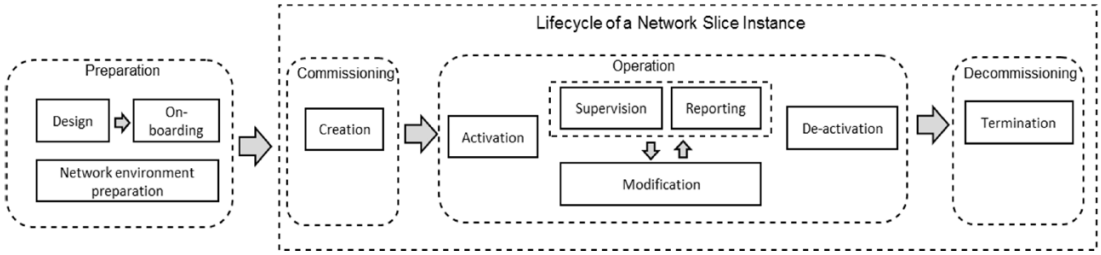

Figure 4.7: Management aspects of network slice instance (NSI) - 3GPP. Figure 4.3.1.1 from 3GPP TS 28.530 [3]

The detailed description of each procedure can be found in [20].

4.3.2 Network Slicing Orchestration by ETSI

European Telecommunications Standards Institute (ETSI), has defined the management and orchestration architecture for Network Function Virtualization (NFV-MANO), described in [22]. Since Network Slicing deployment is based on Software Defined Network (SDN) and Network Function Virtualization (NFV), NFV-MANO architecture has been enhanced by connecting 3GPP entities for Network Slicing management on Annex A.4 from [23].

3GPP's management solution consists of three management functions only without specifying the interfaces and integration with other part of the Network. Whereas, the above mentioned proposal is using the interfaces introduced by Network Function Virtualization - Management and Orchestration (NFV-MANO) architecture and is also integrating this solution with other part of 3GPP's management functions (see figure 4.8).

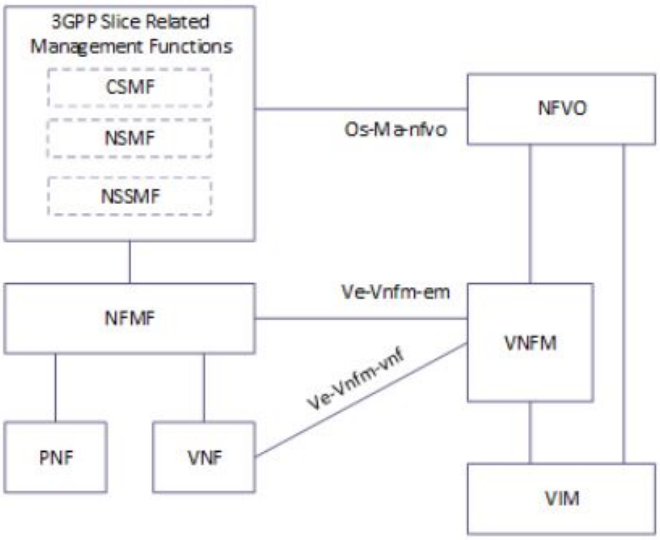

Figure 4.8: Slicing Orchestration in 3GPP and ETSI NFV MANO. Figure from [23]

4.3.3 Network Slicing Orchestration - O-RAN

Another Alliance has focused its attention on Slicing Orchestration and Management by defining the standards on the topic where 3GPP Alliance has not defined yet. O-RAN is working on defining the standards for Orchestration of Network Slicing following the communication model standardized by 3GPP. The Network Slicing Management framework is composed of 2 main functionalities: Non-Real Time RAN Intelligent Controller (RIC) and Near-Real time RAN Intelligent Controller (RIC). The above components shall optimize the orchestration of radio resources between the network slices. AI and ML techniques will be applied at Non-Real time RIC for efficient radio resource allocation in non real time. Data reported from RAN components shall be used as input for obtaining the ML model. O1 interface, shown in Figure 4.9, is introduced for exchanging Orchestration and Management (OAM) messages between the Orchestration Layer and other network components. Furthermore, through aforementioned interface, shall be sent the measurements required as input for Machine Learning (ML). Whereas Near-Real Time RAN Intelligent Controller (RIC) enabled control and optimization of RAN elements and resources. E2 interface enabled the connection between near-real time RIC and RAN components. Non-Real time RIC and Near-Real time RIC are connected with A1 interface which is used to send the ML models for orchestrating the radio resources.

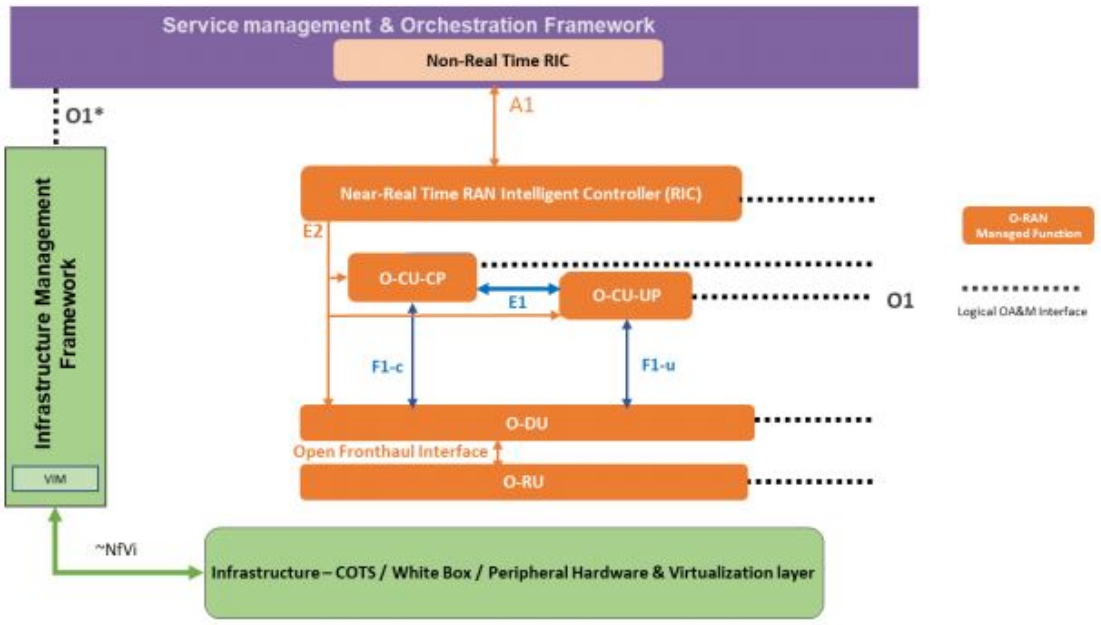

Figure 4.9: O-RAN Orchestration & Management Architecture . Figure from [26]

4.4 Impact on Radio Protocol Layers

Network slicing implementations implies also Radio Access Network slicing. The Radio Protocol Sublayer shall be configured based on the service requirements of every slice. Furthermore, RAN shall share the physical and computational resources between active slices. How the resources shall be assigned to every slice, is left to Radio Resource Management Policies which are implementation depended.

From 3GPP have been defined only the principles for implementing RAN Slicing (see section 16.3.1 from 3GPP TS 38.300 [11]) without defining any implementation framework. Furthermore, on Release 17 it is being discussed one study item on enhancement of RAN Slicing (see [21]).

Therefore, RAN Slicing is left to vendor's implementation. Several techniques are proposed by research groups such as:

- Configuring Slice Layer Descriptor which enables the parameters required for that service (e.g. URLLC slice may require Packet duplication enabled at Packet Data Convergence Protocol (PDCP) layer whereas eMBB may not necessary require it)

- Orchestration Layer configures directly Radio Protocol Layers through configuration messages on a specific interface.

- Therefore, implementation is flexible including also the configuration of Interfaces (how logical channel will be used, is left to implementation. depending on isolation level among slices, will be defined if the channels will be slice dedicated or shared).

- Nonetheless, is important to understand the required features on each layer for different use-cases.

On the Introduction chapter it is mentioned that Network slices have different service requirements, such as End-to-End Latency of 1 ms, high throughput or serving thousands of UEs (i.e. sensors or IoT devices). In order to achieve specific network performance, Radio Protocol Layers shall be configured differently for every slice.

4.4.1 Slice Descriptors

Configuring logical networks above one shared physical network requires management for distributing RAN resources properly. Network Slices can be configured on demand or dynamically upon UEs request. Furthermore, slices can be isolated (in terms of physical resources assigned) or share the physical resources. Several approaches have been proposed by 3GPP and other organizations. However, the configuration is left to implementation by considering service characteristics, network deployment and use-cases in [35].

Slice descriptor are introduced as a method for enabling the slices at NG-RAN. Slice Descriptors shall be configured in gNB by Network Orchestrator and shall configure each Radio Protocol Sublayer (Figure 4.10).

It is stated that slices can share some Layer Descriptors if those have the same requirements for that specific layer, e.g. 2 eMBB slices might have the same Physical Layer Descriptor but different MAC and above layer descriptor.Therefore, slice isolation is provided by assigning Layer specific descriptors per each slice. Nonetheless, this does not limit the implementation solution.

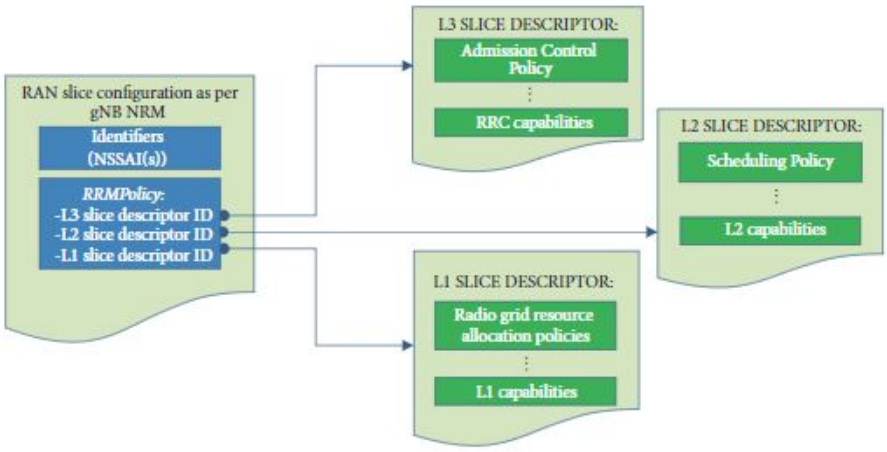

Figure 4.10: RAN Slice Descriptors. Figure from [35]

However, other project such as NGMN, are also working on Network Slicing Management following the management model defined by 3GPP. For NS creation are used NS Templates or Blueprints configured by Orchestration & Management (OAM) or Orchestration Layer (refer to [25]). The concept of *Network Slice Blueprint (NSB)* or *Network Slice Template (NST)* is similar to *Network Slice Instance* introduced by 3GPP, which is a composition of NSSIs (a group of Core Network Functions (CNFs), Access Network Functions (NFs) and Transport Network).

4.4.2 RAN Slicing Examples by NOKIA

As it is mentioned above, for enabling several slices with different service requirements, Radio Protocol Sublayers shall be configured accordingly for meeting the KPIs. One

eMBB slice demands high throughput due to video streaming or video traffic, URLLC slice requires low latency and high reliability on data transmission for maintaining the connection established. URLL communication is very important for factory use case and autonomous vehicles. For guaranteeing E2E latency of 1 ms and high reliability of 99.9999 %, radio sublayer shall be configured to process the packet faster and ensure the delivery (including re-transmissions). NOKIA has came with the following examples (see Table 4.3) for configuring the sublayers for every slice type .

	mMTC	**eMBB**	**URLLC**
RRC	Handover measurements omitted	State Handling Optimized for reduced RAN-CN signaling	State Handling Optimized for reduced state change latency
PDCP	Potential omitting of ciphering and header compression	Default	Potential omitting of ciphering and header compression
RLC	UM only	default	AM only
MAC	HARQ optimized for coverage	default	HARQ omitted for low latency, RACH prioritization
PHY	Coding optimized for coverage, energy efficiency	Coding optimized for very large payloads	Coding optimized for short payloads, low latency.

Table 4.3: RAN Slice Examples by Nokia [32]

In order to ensure packet delivery, Packet Duplication can be enabled at Packet Data Convergence Protocol (PDCP) layer which implies duplication of the packet and routing on different paths. At the receiver side the duplication detection is enabled in order to drop the second packet. Since Packet Data Convergence Protocol (PDCP) Sublayer may perform ciphering and header compression, it is possible to omit these procedures for decreasing the packet processing time. Re-transmission of the packet shall be enabled for ensuring the reliability. Therefore, the Radio Link Control (RLC) entity associated with URLLC Data Radio Bearers (DRBs) shall be Acknowledged Mode (AM) only. Uplink Scheduling is another procedure that requires more time before transmitting the data because of signaling messages. Configured Grant Type 1 and Configured Grant Type 2 allows UEs to transmit uplink data traffic on fixed periodic slots. More about configured grant procedures will be described on the chapter related to MAC Enhancements. 3GPP has standardized three Modulation & Coding Scheme (MCS) tables for achieving high throughput (see table 5.1.3.1-2 from [16]) but also for having a reliable transmission (see table 5.1.3.1-3 from [16]). Therefore, for achieving very High Throughput it is possible to configure at the physical layer the Modulation & Coding Scheme (MCS) corresponding to 64 Quadrature Amplitude Modulation (QAM) or 256 Quadrature Amplitude Modulation (QAM). Whereas in case of URLLC data traffic, it is suggested to configure one Modulation & Coding Scheme (MCS) value from the third table (see table 5.1.3.1-3 from [16]) which provides more robust transmission by applying lower coding rate.

To conclude, one possible implementation solution would to be configure RAN Slice Descriptors and for every slice type the radio sublayers configuration may be similar to Nokia's examples.

4.5 Network Slice Deployment Scenarios

Several researches have been done on deployment solutions that best fits the use-cases. Even if 3GPP has standardized the architecture of RAN and the interfaces for Split Option-2, the deployment is left to implementation. Network operators may implement all network functions in the same PNF/VNF, located on the same DC but they can also implement CU at centralized running in one DC and DU running close to the UE. Therefore, this solutions provide flexibility to the vendor on implementing the Network Functions (NFs) accordingly. Some deployment solutions have been proposed by NGMN Alliance [24], O-RAN Alliance [27] for Radio Access Network based on the use-cases. Both solutions highlight the level of flexibility and freedom on network operator's site. Furthermore, Low Layer Split (LLS) enables the possibility that network operators provide a propriety solution for Radio Unit (RU) and Fronthaul Interface. However, O-RAN is working to set standards for vendor inter-operable Fronthaul Interface.

The Next Generation Mobile Networks Alliance (NGMN) is founded by world-leading mobile network operators. Its goal is to ensure that the standards for next generation network infrastructure, service platforms and devices will meet the requirements of operators. The organization is researching on Network Slicing topic and deployment scenarios for implementing the slices for different use-cases such as URLCC, eMBB. Ensuring diverse service demands can be fulfilled by placing the network functions in different sited of the network.

URLLC slice composition is respresented in figure 4.11 and it requires the User Plane Functions (UPF) and Multi-access edge computing (MEC) components to be placed close to the cell site. In this way, latency requirements can be met and requirements for Transport Network can be fulfilled. Whereas, when the RAN components are placed on different sites, requirements for Transport Network will increase and will become more difficult to ensure low latency End-to-End (E2E) communication.

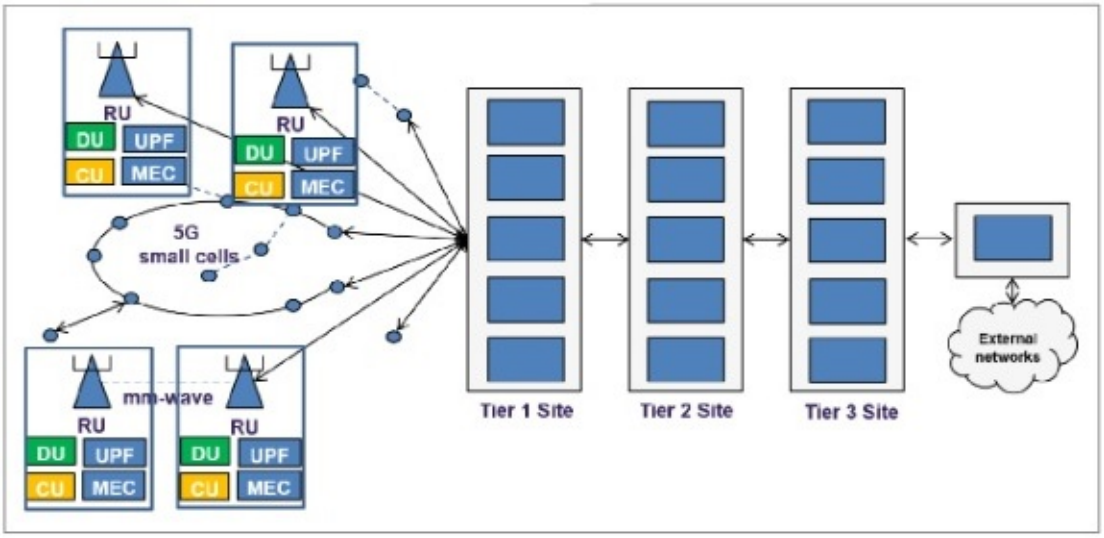

Figure 4.11: Ultra-Low-Latency (1ms) Deployment. Figure from NGMN [24]

On the other hand, this is not a must for eMBB use-case when the low latency is not crucial for ensuring the service connectivity (see Figure 4.12). Therefore, some

component might be placed on other sites of the Network and the components (e.g. CU, DU, RU) shall be connected with Transport Network. More details and analysis of each proposed option can be found on [24].

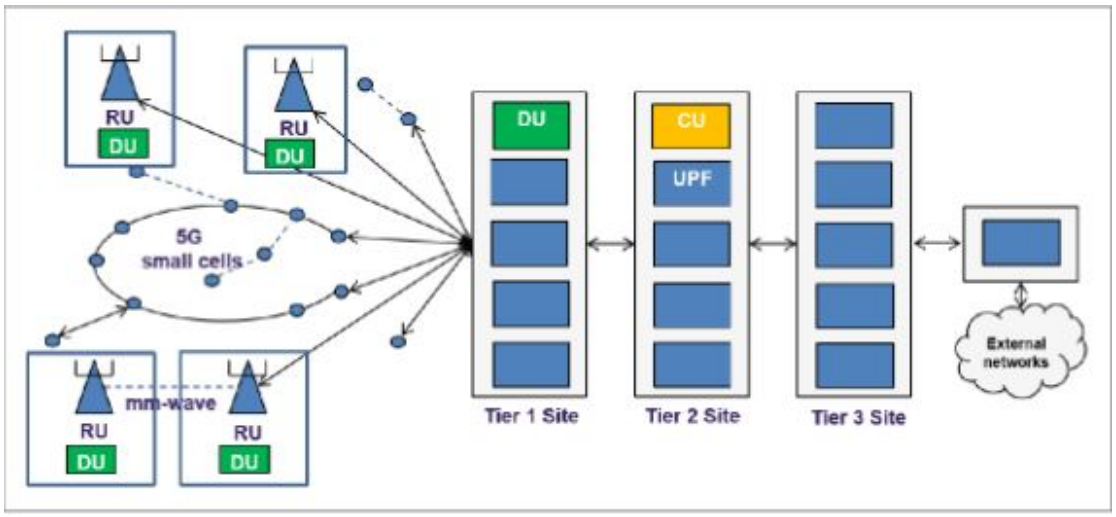

Figure 4.12: Deployment Topology for <10ms Latency. Figure from NGMN [24]

4.6 Summary

This chapter covers Network Slicing technique based on 3GPP perspective and not only. Determining the specified features and solutions, helps understanding the its evolution towards deploying in a physical network. 3GPP Alliance has defined the procedures, features, requirements and the framework for Slice Orchestration. However, that is not enough to implement Network Slicing on reality. Therefore, O-RAN Alliance shows up and proposes its solution for deploying and configuring Network Slicing in a physical network by standardizing the features and solutions not defined by 3GPP.

Chapter 5

QoS model

5.1 Introduction

5G promises a range of capabilities from high throughput to supporting real-time low-latency factory automation or self-driving cars. This range of capabilities requires that the Quality of Service (QoS) characteristics, such as delay, error rate, and priority be specified and enforced. In LTE the Quality of Service (QoS) is enabled and it is based on bearers. In order to better understand the enhancements introduced in 5G, a comparison with legacy mobile generation is necessary. EPS bearer/E-RAB established when UE connects to a PDN is called the default bearer, while any additional bearer is referred to as a dedicated bearer.Each bearer is characterized by the same packet forwarding treatment (e.g., scheduling policy, queue management policy, rate shaping policy, RLC configuration, etc.). A bearer is called guaranteed bit-rate (GBR) bearer if it is provided dedicated network resources, otherwise, it is called non-GBR bearer. Unlike LTE, where QoS model is based on Radio Bearers, in 5G it is flow based (Figure 5.1). Furthermore, in LTE there is a strict one-to-one mapping between Evolved Packet Switched System Bearers (EPS-bearers) and EPS radio access bearers (E-RAB). Whereas, in 5G it is possible to map several QoS Flows to one Data Radio Bearer (DRB).

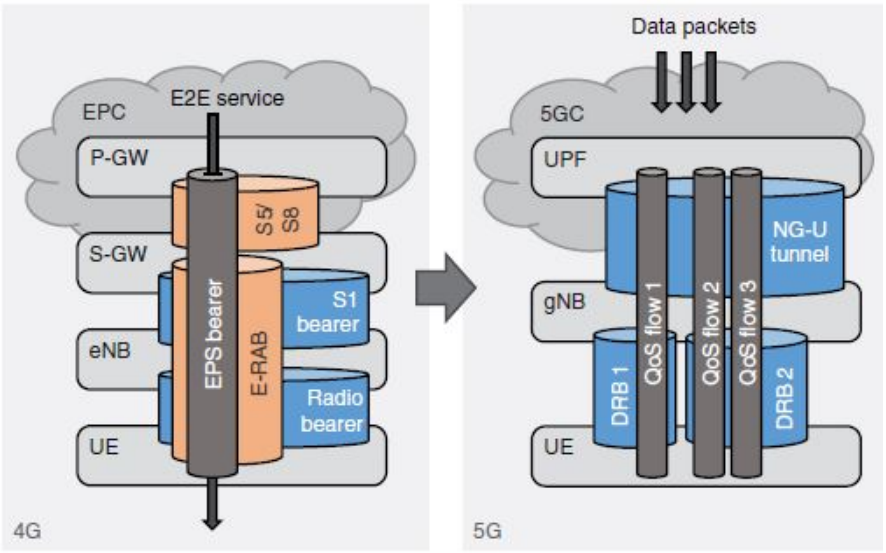

Figure 5.1: QoS Flows in 4G & 5G.

Network slicing provides a holistic end-to-end virtual network for a given tenant. Whereas, QoS model makes possible the differentiation of traffic coming from the same UE. Therefore, Network Slicing differentiates from QoS because it will enable end-to-end virtual networks encompassing compute, storage and networking functions. Contrarily, QoS cannot discriminate and differently treat the same type of traffic (e.g. VoIP traffic) coming from different tenants or perform traffic isolation. It is thought that Network Slicing in conjunction with QoS model will provide better service and improve user experience.[34].

This chapter is organized as follows: First will describe the QoS Flows, QoS Profile and the QoS Parameters in 5G. Then, it contains the rules for mapping the QoS Flows in Uplink and Downlink direction. Concluding with the impact of QoS method on RAN interfaces.

5.2 QoS Flows

3GPP has defined the QoS Flow as the finest granularity of QoS differentiation in the PDU Session. QoS Flows are identified by a QoS Flow ID (QFI) in the network which shall be unique within a Packet Data Unit (PDU) Session and is is carried in an encapsulation header on N3 tunnel (UPF-gNB Interface). On 5G, one QoS Flow can be mapped to one or more Data Radio Bearers (DRBs) that have the same QoS requirements (Figure 5.2.

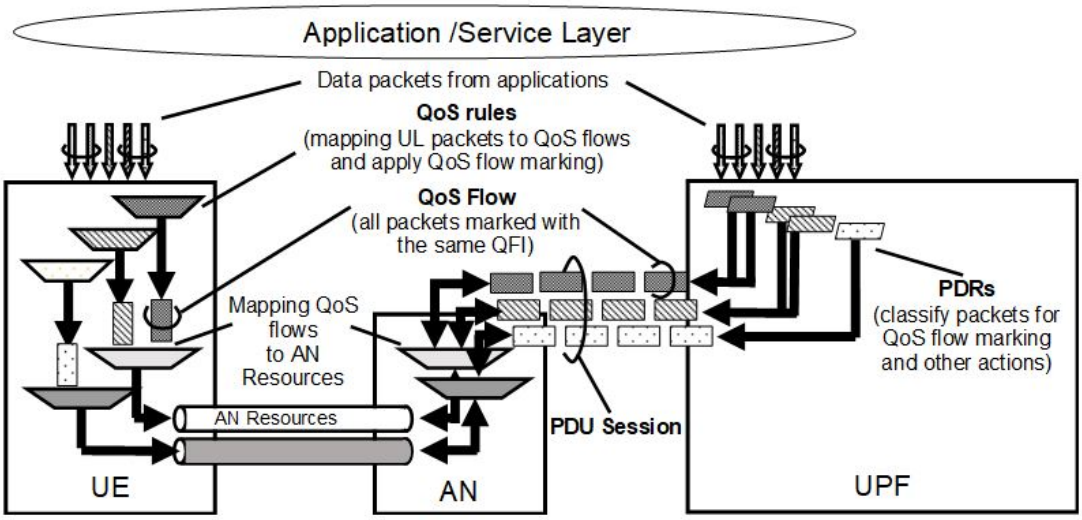

Figure 5.2: QoS Flows Mapping in 5G. Figure taken from [1]

Each QoS Flow is characterized by the **QoS Profile**, **QoS Rules** and **UL/DL Packet Detection Rules (PDRs)**.

QoS Profile contains the QoS parameters for the Physical Layer Requirements in terms of delay budget, packet loss etc.
QoS Rules contains the QFI of the associated QoS Flow, a Packet Filter Set used for mapping the IP traffic to QoS Flows and the opposite. Furthermore, it contains a precedence value for assigning the priority of QoS Rules at the UE side e.g. the order that QoS Rules shall be used for mapping one particular IP Flow.

UL/DL Packet Detection Rules (PDRs) contains packet filters required by User Plane Function (UPF) to map and mark the packets into QoS Flows in UL and DL direction. There rules are provided from Session Management Function (SMF) to User Plane Function (UPF) once the QoS Flow is enabled.
One QoS Flow has associated with it also another identifier, 5G QoS Identifier (5QI) which is included in the QoS Profile and is linked with the QoS Parameters for the Physical Layer. 5G QoS Identifier (5QI) parameter can be selected from the standardized values (see table 5.7.4-1 [1]) or dynamically assigned value. More details are provided in the following section. QoS Flows are categorized in three main groups: *Guaranteed Bit Rate (GBR)*, *non-Guaranteed Bit Rate (Non-GBR)* and *Delay Critical QoS Flow*. Each category has specific QoS Parameters to be assigned based on the requirements for physical layer performance, e.g. Guaranteed Bit Rate (GBR) QoS Flows have a QoS Parameter that signals the minimum bit rate to be ensured by the network. 3GPP has provided the list of standardizes QFI

5.2.1 QoS Profile

QoS profile contains the parameters that determine the QoS flow type as: GBR, non-GBR or Delay Critical. QoS Parameters signalled for a QoS Flow are the following:

- *5G QoS Identifier (5QI)* is a scalar that is used as a reference to 5G QoS characteristics for controlling QoS forwarding treatment for the QoS Flow . 5QI value may be assigned dynamically or from standardized set of values. Standardized 5QI values are specified for services that are assumed to be frequently used and thus benefit from optimized signalling by using standardized QoS characteristics (Found in [1] table 5.7.4-1). Dynamically assigned 5QI values (which require a signalling of QoS characteristics as part of the QoS profile) can be used for services for which standardized 5QI values are not defined.

- *Allocation and Retention Priority (ARP)* contains information about the priority level, the pre-emption capability and the pre-emption vulnerability. The ARP Priority parameters defines the priority level of allocated resources per each flow. Therefore, admission control can be performed for GBR Flows). Furthermore, it may used to select which existing QoS Flow to pre-empt during resource limitations. ARP value for each QoS Flow is selected and signaled by SMF. More details about the range of values can be found in [1] clause 5.7.2.2.

- *Reflective QoS Attribute (RQA)* is an optional parameter applied only for non-GBR Flows. It indicates that certain traffic carried on this QoS Flow is subject to Reflective QoS mapping (see section 5.3.1.1). It is used at UE side to map the UL traffic into QoS Flows. When RQA is enabled for a QoS Flow, reflective QoS Indicator (RQI) is signalled by NG-RAN.

- *Guaranteed Flow Bit Rate (GFBR)* is a parameter set only for GBR flows and shall be defined for both directions, UL and DL. Denotes the bit rate that is guaranteed to be provided by the network to the QoS Flow over the Averaging Time Window.

- *Maximum Flow Bit Rate (MFBR)* is a parameter configured only for GBR flows and shall be defined for both directions, UL and DL. Limits the bit rate to the

highest bit rate that is expected by the QoS Flow.

- *Notification control* is an optional parameter applied only for GBR Flow when GFBR can no longer be guaranteed. NG-RAN signals the Notification Control messages towards SMF. The parameter itself is configured by SMF for each GBR Flow, based on QoS Policy and Charging Control (PCC) rules.

- *Maximum Packet Loss Rate* indicates the maximum rate for lost packets of the QoS flow that can be tolerated in the uplink and downlink direction. It can be applied for GBR flows carrying voice traffic.

- *Session Aggregate Maximum Bit Rate (Session-AMBR)* limits the aggregate bit rate that can be expected to be provided across all Non-GBR QoS Flows for a specific PDU Session. It is calculated over the Averaging Window. It is signaled per each non-GBR QoS Flow to the UPF, NG-RAN and UE.

- *UE Aggregate Maximum Bit Rate (UE-AMBR)* AMBR limits the aggregate bit rate that can be expected to be provided across all Non-GBR QoS Flows of a UE. It is calculated over the Averaging Window. UE-AMBR is provided to NG-RAN by AMF and is applied only for Non-GBR Flows.

- *QoS Flow Identifier (QFI)* it is a QoS Flow identifier and is used to identify a QoS Flow in the 5G System. QFI is not carried in the QoS profile itself, but is carried in an encapsulation header on N3 (and N9) i.e. without any changes to the e2e packet header. The QFI shall be unique within a PDU Session. The QFI may be dynamically assigned or may be equal to the standardized 5QI value if it is less than 64. QFI can be in the range: 0-63.

Other QoS characteristic that can be configured per each QoS Flow are the following:

- *Resource type* defines if the resources are permanently allocated to a QoS Flows. GBR Flows have commonly 'on demand' allocated resources. Therefore, requires dynamic policy and charging control. It may use either the GBR resource type or the Delay-critical GBR resource type. Whereas, Non-GBR Flows may be pre-authorized through static policies and charging control. PER and PDB defined below differentiate depending on QoS resource type.

- *Packet Delay Budget* defines an upper bound for the time that a packet may be delayed between the UE and the UPF that terminates the N6 interface. For a certain 5G QoS Identifier (5QI) the value of the Packet Delay Budget (PDB) is the same in UL and DL. For GBR QoS Flows using the Delay-critical resource type, a packet delayed more than Packet Delay Budget (PDB) is counted as lost if the data burst is not exceeding the Maximum Data Burst Volume (MDBV) within the period of Packet Delay Budget (PDB) and the QoS Flow is not exceeding the GFBR. For GBR QoS Flows with GBR resource type not exceeding Guaranteed Flow Bit Rate (GFBR), 98 percent of the packets shall not experience a delay exceeding the 5QI's Packet Delay Budget (PDB).

- *Packet Error Rate* the PER defines an upper bound for a rate of non-congestion related packet losses. The purpose of the PER is to allow for appropriate link layer protocol configurations. For GBR QoS Flows with Delay critical GBR resource type, a packet which is delayed more than Packet Delay Budget (PDB) is counted as lost.
- *Averaging Window* it is a parameter assigned to each GBR QoS Flow. The Averaging window represents the duration over which the Guaranteed Flow Bit Rate (GFBR) and Maximum Flow Bit Rate (MFBR) shall be calculated.
- *Maximum Data Burst Volume* is applied only for GBR QoS Flow with Delay-critical resource type shall be associated with a Maximum Data Burst Volume (MDBV). MDBV denotes the largest amount of data that the 5G-AN is required to serve within a period of 5G-AN PDB.

The table 5.1 provides a summary of QoS Parameters and characteristics required per each QoS Flow type.

Flow Type	Non-GBR	GBR	
QoS Parameters Resource Type	**Non-GBR**	**GBR**	**Delay-Critical GBR**
5QI	M	M	M
QFI	M	M	M
ARP	M	M	M
RQA	O	NO	NO
GFBR	NO	M	M
Notification Control	NO	O	O
MPLR	NO	O	O
Session-AMBR	M	NO	NO
UE-AMBR	M	NO	NO
PDB	M	M	M
PER	M	M	M
Averaging Window	M	M	M
MDBV	NO	M	M
M: Mandatory, O: Optional			

Table 5.1: Qos Parameters - QoS Flow

5.2.2 QoS Rules

The UE performs the mapping and marking of UL traffic into QoS Flows based on QoS Rules. The QoS Rule contains the following parameters:

- *QFI* of the associated QoS Flow.
- *Packet Filter Set* is used to identify one or more packet (IP or Ethernet) flow(s). The Packet Filter Set may contain one or more Packet Filter(s). Every Packet Filter is applicable for the DL direction, the UL direction or both directions.There are two types of Packet Filter Set, i.e. IP Packet Filter Set, and Ethernet Packet Filter Set, corresponding to those PDU Session Types. More details can be found on [1] clause 5.7.6.1.

- *Precedence value* it is a parameter set per each QoS Rule to determine the order in which a QoS rule shall be evaluated. The evaluation is performed in increasing order of their precedence value.

- *QoS rule identifier* is assigned only for explicitly signaled QoS Rules. It shall be unique within the PDU Session and is generated by SMF.

QoS Rules are signalled from SMF to UE when the QoS Flow is enabled, .in the following 3 ways (see section 5.7.1.4 [1]):

- Explicitly signaled to the UE through PDU Session Establishment/Modification procedure.

- Pre-configured in the UE before the QoS Flow establishment.

- By applying Reflective QoS which is a method introduced in 5G for applying the same mapping rules for UL traffic as the rules applied for DL traffic. Therefore, SMF shall not signal explicitly the rules for mapping the UL traffic because the UE shall obtain those by monitoring the received DL traffic of the same QoS Flow. Reflective QoS Mapping is described in the following sections.

5.2.3 Packet Detection Rules (PDRs)

Packet Detection Rules (PDRs) are used by UPF to classify and mark the packet into QoS Flows. The PDRs are signaled from SMF to UPF through N4 interface. Every PDR is used to detect packets in a certain transmission direction, e.g. UL direction or DL direction. Packet Detection Rule table contains a set of attributes for identifying the rules within a PDU Session, to set the rules for monitoring the QoS. One of the Attributes is the *QoS Rule ID* used to uniquely identify this rule. *Precedence Value* is used to determines the order in which the detection information of all rules is applied. *QFI* is included in PDR message in order to identify the QoS Flow which it corresponds to and may contain standardized 5QI value or non-standardized QFI value. *CN tunnel info* identifies the N3 or N9 tunnel where the packets will be transferred. *Packet Filter Set* along with *Forwarding Action Rule ID* is used to identify packet (IP or Ethernet) flow(s) and the forwarding action that has to be applied. In order to monitor that QoS requirements are met for one flow, the following attributes are signaled: *QoS Monitoring Packet indicator* which identifies the packet is used for QoS Monitoring, *List of Usage Reporting Rule* identifies a measurement action that has to be applied and *List of QoS Enforcement Rule* identifies a QoS enforcement action that has to be applied.
More details about Packet Detection Rules (PDRs) and parameters exchanged on N4 interface can be found in section 5.8.2.11 from [1].

5.3 QoS Flow Mapping

QoS Flows shall be established and managed by Session Management Function (SMF). The latter shall provide the QoS requirements to gNB, UE and User Plane Function (UPF). Furthermore, it interacts with Unified Data Management (UDM) for retrieving UE subscribed data (which contains Allowed or Configured QoS Rules) and with Policy Control Function (PCF) for retrieving Authorized QoS Rules based. On the other hand, Session Management Function (SMF) signals to NG-RAN the QFI and QoS Profile. Moreover, User Plane Function (UPF) receives from Session Management Function

(SMF) Packet Detection Rules for DL/UL direction along with the corresponding Precedence Value, the corresponding transport packet marking information (QFI or DSCP value of outer IP header) and QoS related information such as GFBR, MFBR. The Figure 5.3 summarizes the signaling flows for establishing one QoS Flow.

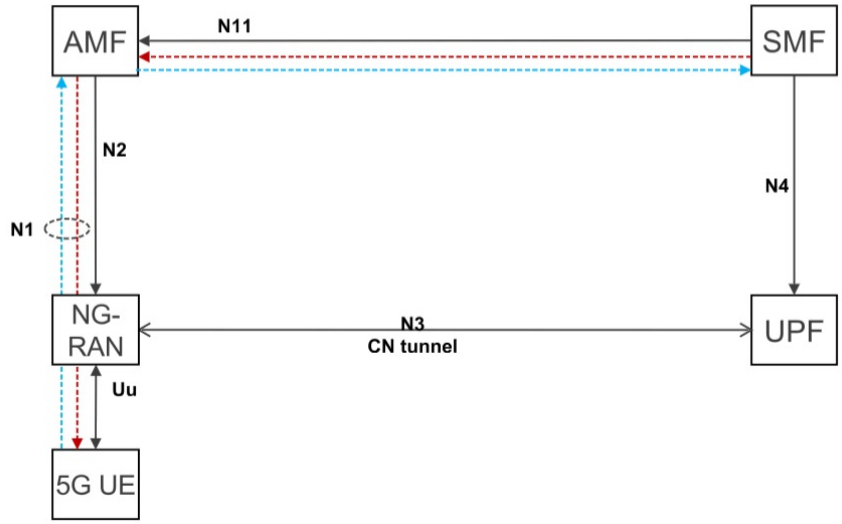

Figure 5.3: QoS Flows Signalling in 5G.

QoS Flow Mapping is done in 2 steps: first mapping the traffic flow coming from application layer into QoS Flows based on QoS Rules. Then, the QoS Flows are further mapped into Access Network Resources (DRBs) with no restriction of one-to-one mapping. UE applies QoS Rules to map the traffic flows into specific QoS Flows. Whereas at User Plane Function (UPF) side, Packet Detection Rules for UL and DL direction are applied.

5.3.1 Mapping procedure in Uplink

Uplink QoS Mapping is performed at UE side using QoS Rules which may be configured directly at UE, explicitly signaled by Session Management Function (SMF), or derived by UE when Reflective QoS Indicator (RQI) is enabled. SMF receives the authorized QoS Rules to be applied, from Unified Data Management component which contains the authorized information for every used. The QoS Rules and QoS Profile is signaled to gNB and UE from SMF through the N1 SM messages and N2 SM messages (refer to 5.5).

N1 Session Management (SM) signaling is sent from the network towards the UE (and vice-versa) during PDU Session Establishment and it contains QoS Rules etc. 3GPP has defined the message format and the parameters signaled (refer to section 8.3 [10]). Authorized Rules, Session Aggregated Maximum Bit Rate (AMBR) etc are signed to UE through the message shown in the figure 5.4. Authorized Rules are mandatory (M) parameter to be sent to the UE, whereas the parameters marked with O - are optional. Instead N2 Session Management (SM) signaling is send from the network to gNB during the PDU Session Establishment.

IEI	Information Element	Type/Reference	Presence	Format	Length
	Extended protocol discriminator	Extended protocol discriminator 9.2	M	V	1
	PDU session ID	PDU session identity 9.4	M	V	1
	PTI	Procedure transaction identity 9.6	M	V	1
	PDU SESSION ESTABLISHMENT ACCEPT message identity	Message type 9.7	M	V	1
	Selected PDU session type	PDU session type 9.11.4.11	M	V	1/2
	Selected SSC mode	SSC mode 9.11.4.16	M	V	1/2
	Authorized QoS rules	QoS rules 9.11.4.13	M	LV-E	6-65538
	Session AMBR	Session-AMBR 9.11.4.14	M	LV	7
59	5GSM cause	5GSM cause 9.11.4.2	O	TV	2
29	PDU address	PDU address 9.11.4.10	O	TLV	7, 11 or 15
56	RQ timer value	GPRS timer 9.11.2.3	O	TV	2
22	S-NSSAI	S-NSSAI 9.11.2.8	O	TLV	3-10
8-	Always-on PDU session indication	Always-on PDU session indication 9.11.4.3	O	TV	1
75	Mapped EPS bearer contexts	Mapped EPS bearer contexts 9.11.4.8	O	TLV-E	7-65538
78	EAP message	EAP message 9.11.2.2	O	TLV-E	7-1503
79	Authorized QoS flow descriptions	QoS flow descriptions 9.11.4.12	O	TLV-E	6-65538

Figure 5.4: N1 SM Signaling Network->UE. Figure 8.3.2.1.1 from 3GPP TS 24.501 [10]

Based on the QoS rules, UE determines mapping between UL User Plane traffic and QoS Flows. The Precedence value defines which QoS Rules will be applied for that particular Service Data Flow (SDF). UE marks the UL Protocol Data Unit (PDU) with the QoS Flow Identifier (QFI) and transmits the UL PDUs using the corresponding access specific resource for the QoS Flow based on the mapping provided by (R)AN. Service Data Adaption Protocol (SDAP) Layer has as main function the mapping of IP Flows to QoS Flows (based on the Packet Filters Set), including the QFI value in the protocol header and further mapping it to DRBs. When UL IP traffic is received and RAN fulfills Maximum Data Burst Volume (MDBV), Guaranteed Flow Bit Rate (GFBR), Aggregate Maximum Bit Rate (AMBR) requirements by properly configuring mapping restriction parameters of Logical Channels at the UE. As an instance, it may configure the Prioritized Bit Rate (PBR) of one Logical Channel (LC) based on Maximum Data Burst Volume (MDBV) required for some QoS Flows.

(R)AN transmits the PDUs over N3 tunnel towards UPF. When passing an UL packet from (R)AN to CN, the (R)AN includes the QFI value in the encapsulation header of the UL PDU. Furthermore, it performs transport level packet marking in UL on a per QoS Flow basis and the transport marking value is defined based on the 5QI, the Priority Level and the ARP priority level of the associated QoS Flow. UPF verifies whether QFIs in the UL PDUs are aligned with the QoS Rules provided to the UE or implicitly derived by the UE (in the case of Reflective QoS). UPF and UE perform Session-AMBR enforcement and the UPF performs counting of packets for charging (see section 5.7.1.5 from [1]).

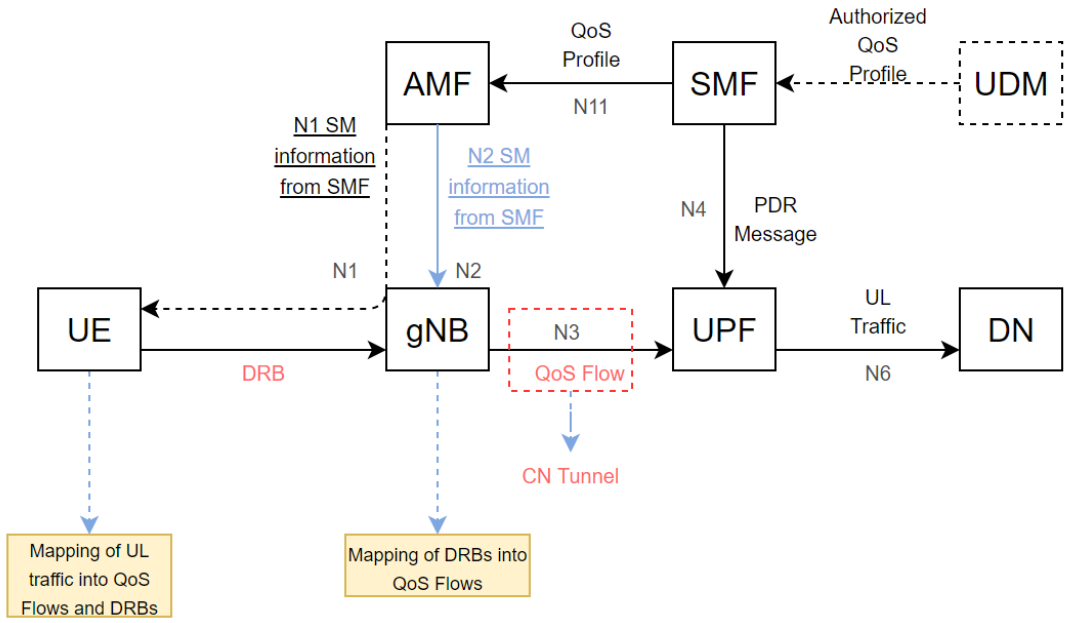

Figure 5.5: Uplink QoS Mapping - Signaling Flow

5.3.1.1 Reflective QoS Mapping

Reflective QoS is controlled on per packet basis by using the Reflective QoS Indication (RQI) in the encapsulation header on N3 (and N9) reference point with the QoS Flow Indicator (QFI) and together with a Reflective QoS Timer (RQ Timer) value. RQ timer may be either signalled to the UE upon PDU Session Establishment (or upon PDU Session Modification as described in clause 5.17.2.2.2) or set to a default value. When the 5GC determines that Reflective QoS has to be used for a specific IP Traffic belonging to a QoS Flow, the SMF shall provide the Reflective QoS Attribute (RQA) within the QoS Flow's QoS profile to the NGRAN on N2 reference point. Furthermore, the SMF shall include an indication to use Reflective QoS in the corresponding Service Data Flow (SDF) information provided to the UPF via N4 interface. When the UPF receives this indication for an SDF, the UPF shall set the Reflective QoS Indicator (RQI) in the encapsulation header on the N3 reference point for every DL packet corresponding to this SDF. When an RQI is received by (R)AN in a DL packet on N3 reference point, the (R)AN shall indicate to the UE the QFI and the RQI of that DL packet (see section 5.7.5 [1]).

Upon reception of a DL packet with Reflective QoS Indicator (RQI):

- if a UE derived QoS rule with a Packet Filter corresponding to the DL packet does not already exist, it shall create a new UE derived QoS rule with a Packet Filter corresponding to the DL packet. Furthermore, it shall start the timer set to the RQ timer value.
- if the DL packet exists, the UE shall restart the timer associated to this UE derived QoS rule

5.3.2 Mapping procedure in Downlink

Downlink QoS mapping is performed at User Plane Function (UPF) for classifying the received Data traffic from the Data Network (DN) into QoS Flows. In order to classify the traffic into QoS flows and mark the corresponding packets, User Plane Function

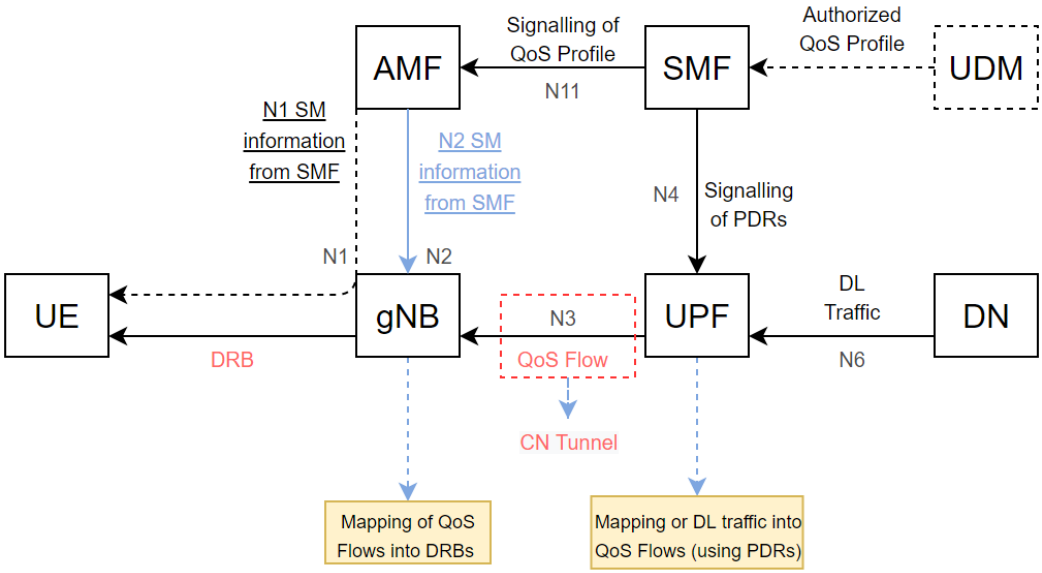

Figure 5.6: Downlink QoS Mapping - Signaling Flow

(UPF) uses Packet Filters and the QoS Flow Identifier (QFI) value signaled from Session Management Function (SMF). Several packet filters may correspond to the same QoS Flow Indicator (QFI) for one PDU Session. Hence, several data flows can be mapped to the same QoS Flow and marked with QFI value at the packet header. Furthermore, based on Forwarding rules signaled by Session Management Function (SMF), User Plane Function (UPF) will perform routing and forwarding actions. As aforementioned in previous sections, along with Packet Detection Rules, Session Management Function (SMF) signals also rules for monitoring the QoS enforcement and traffic usage. The latter results are further reported back to Session Management Function (SMF) which may take actions such as deallocating the resources of a Non-GBR Flow when those are needed by a GBR Flow. User Plane Function (UPF) performs Session-AMBR enforcement and performs counting of packets for charging. Therefore, Session Management Function (SMF) ensures that QoS requirements are met for GBR flows in particular. Whereas for Non-GBR flows, Session Management Function (SMF) ensures that AMBR assigned is not exceeded.

User Plane Function (UPF) transmits the packets of the PDU Session in a single tunnel between 5GC and (R)AN, including the QFI in the encapsulation header. In addition, User Plane Function (UPF) may include an indication for Reflective QoS activation in the encapsulation header (if RQA enabled for that QoS Flow). Furthermore, it performs transport level packet marking in DL on a per QoS Flow basis by applying the packet marking value signaled by SMF. When QoS Flows are received by NG-RAN, the second step shall be performed by mapping the QoS Flows to access-specific resources. The mapping into Data Radio Bearers (DRBs) is done based on the QoS Flow Indicator (QFI), QoS Profile, N3 tunnel associated with the DL packet and the explicit rule is defined by gNB Radio Resource Management (RRM) (refer to section 5.7.1.6 from [1]).

5.4 Impact on Radio Protocol Layers

QoS Flow is used on Radio Protocol Stack for differentiating the traffic from the same UE. In order to manage the QoS traffic in 5G, another user plane protocol has been

introduced , Service Data Adaption Protocol.

One PDU Session may contain more than one Data Radio Bearers (DRBs) and QoS Flows. In [1] clause 5.7.1.6 is stated that User Plane Function (UPF) transmits the PDUs of the PDU Session in a single tunnel between 5GC and (R)AN. Furthermore, the User Plane Function (UPF) includes the QFI in the encapsulation header. Therefore, all QoS Flows of the same PDU Session shall have the same QoS Flow Indicator (QFI) value.

As aforementioned in the beginning on the chapter, in 5G will be possible to map more than 1 QoS Flows into one Data Radio Bearer (DRB).

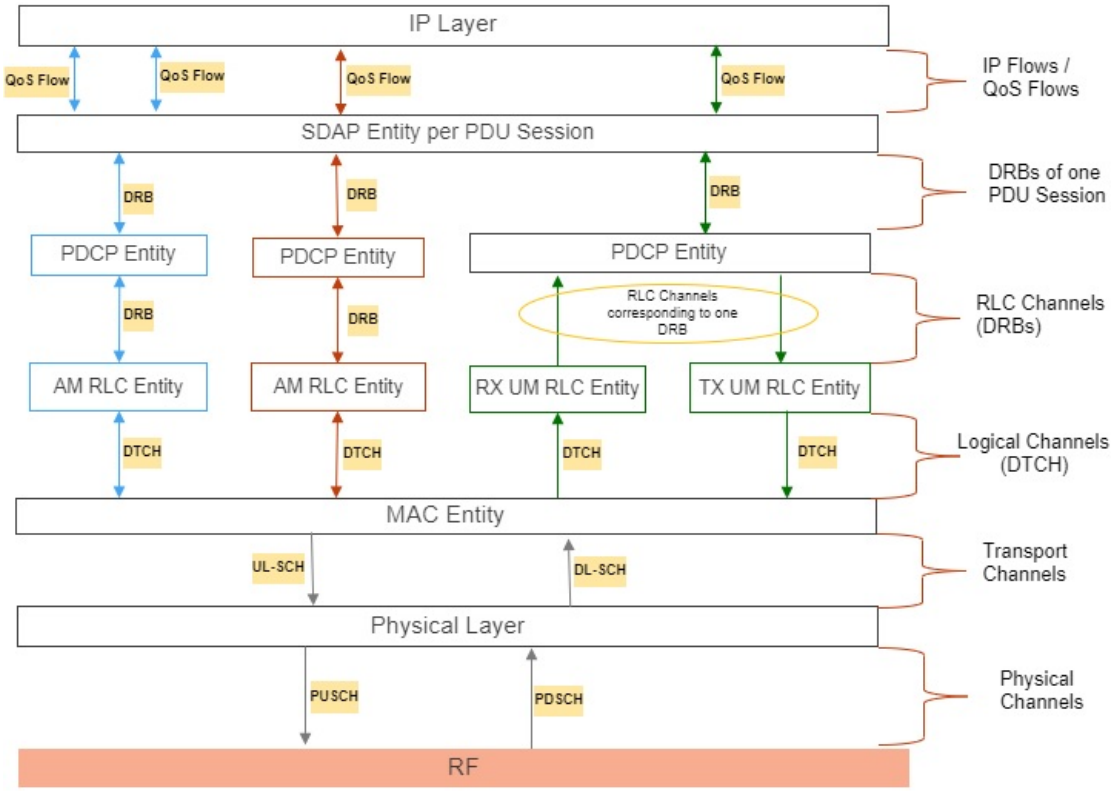

Figure 5.7: User Plane Stack for one PDU Session.

To conclude, in the Figure 5.4 is represented how one PDU Session is handled by each Radio Protocol Sublayer. One Service Data Adaption Protocol (SDAP) Entity handles all the data flows corresponding to one PDU Session. Service Data Adaption Protocol (SDAP) Sublayer is responsible for mapping the QoS flows into Data Radio Bearers (DRBs). One PDU Session can have more than one Data Radio Bearer (DRB) and each Data Radio Bearer (DRB) is associated with one Packet Data Convergence Protocol (PDCP) Entity and more than one RLC Entities (based on number of Data Radio Bearers (DRBs) established and TM). In the Figure 5.7 the PDU Session contains three Data Radio Bearers (DRBs) and four QoS Flows where two QoS Flows (see the QoS flows marked in blue) are mapped into one Data Radio Bearer (DRB) since on 5G it is possible. Whereas the other are QoS Flows are mapped each to one Data Radio Bearer (DRB). Furthermore, The DRBs marked in blue and in red, are associated with one AM Radio Link Control (RLC) Entity. On the other hand, the DRB marked in green is

associated with two UM Radio Link Control (RLC) Entities, the Receiving Radio Link Control (RLC) Entity and the Transmitting Radio Link Control (RLC) Entity. It is important to be mentioned that all QoS Flows within the same PDU Session shall have the same QoS Flow Indicator (QFI) value.

5.5 Summary

Enabling differentiated treatment of traffic flows, ensures the session connectivity of high priority applications. The differentiation is done in flow level, separating the traffic flows coming from the same UE. Qos Model in 5G will make possible to distinguish the types of traffic handled by one UE. Service Data Adaption Protocol (SDAP) sublayer maps the IP traffic into QoS Flows and Radio Access Network Resource based on configurations from Radio Resource Control (RRC) layer.

Chapter 6

Enhancements in 5G MAC Layer

6.1 Introduction

Network Slicing and Quality of Service (QoS) have impacted the architecture and functionality of Media Access Channel protocol at UE side and also at gNB side. Since MAC layer is in charge of managing and scheduling the transmission grants UEs, scheduling algorithm shall be aware of slices and QoS requirements. The physical resources shall be shared between the slices and further assigned between UEs within one slice. Furthermore, serving delay-sensitive applications requires solutions that deal with signaling overhead for UL grants. Therefore, 3GPP has set the requirements for scheduling the packets with low latency requirements [15]. However, the scheduling algorithm to be used at gNB is left to implementation. Furthermore, 3GPP has defined the framework for MAC sublayer at UE only. Several research groups and organizations such as: EUROCOM [36], Universitat Politecnica de Catalunya & Fondazione Bruno Kesslerare [35], are working on this topic to find the best approach to serve different Network Slices with QoS requirements.

6.2 MAC Architecture

MAC Architecture is based on MAC Entities. MAC sublayes at UE may be configured with one MAC Entity or two MAC Entities when the Secondary Cell Group (SCG). RRC Layer is in control of MAC sublayer and configures the MAC Entities. 3GPP has proposed in the one possible implementation solution for MAC Sublayer (Figure 6.1).
The MAC sublayer operates on the transport channels and logical channels The MAC sublayer uses the transport channels to send/receive data from Layer 1 and are categorized as Downlink Channels or Uplink Channels (see table 4.5.2-1 [15]). Whereas on Logical channels it provides data transfer services to/from RLC Sublayer. Logical channels are categorized as Control Channels and Traffic Channels (see table 4.5.3-1 [15].
MAC Entity has several functionalities (see section 4.4 from [15]):

- Mapping between logical channels and transport channels. 3GPP has set the rules of mapping the traffic from logical channels to transport channel and the opposite and can be found on section 4.5.4 from [15].

- Scheduling information reporting

- Multiplexing of MAC SDUs from one or different logical channels onto transport blocks (TB) to be delivered to the physical layer on transport channels

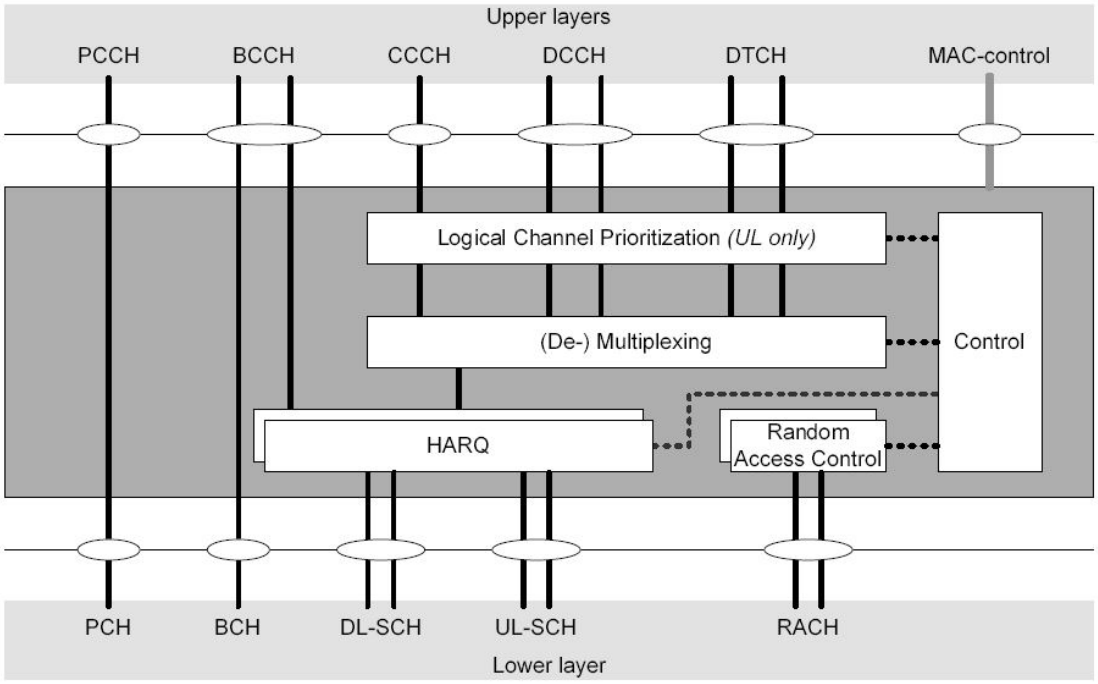

Figure 6.1: MAC Architecture. Figure 4.2.2-1 from [15]

- Demultiplexing of MAC SDUs to one or different logical channels from transport blocks (TB) delivered from the physical layer on transport channels;
- Error correction through HARQ
- Logical Channel prioritisation.

MAC Entity shall handle several procedures such as: Random Access Procedure, DL-Scheduling data transfer, UL -Scheduling data transfer, HARQ operation, Discontinuous Reception (DRX), MAC CE handling etc. (see section 5 of [15]). In the following sections is described in more details Scheduling procedure and Logical Channel Prioritization.
MAC Protocol handles each functions and procedure through the use of Control Elements (CEs) which are type of messages exchanged by MAC layer. Each CE has a specific format type, length and functionality defined by 3GPP (see section 6.1.3 from [15]).
The list of defined Control Elements are presented in table: Table 6.2.1-1 (LCID Values for DL-SCH) and Table 6.2.1-2 (LCID Values for UL-SCH) of [15].

6.3 Scheduling Procedure

Scheduling is the process of allocating resources for transmitting data. As in LTE (actually in all cellular communication), NR scheduling is dictated by Network and UE is just following what network tells. Overall scheduling mechanism in NR is pretty much similar to LTE scheduling, but NR has finer granularity than LTE especially in terms of time domain scheduling at physical layer.
The scheduling algorithm to be applied by RAN has not been specified. It is the crucial part of Radio Access Network dealing with radio resource allocation in real time, serving different users and services with diverse requirements. Hence, it is left to implementation depending on type of applications that will be served in the network and other network

requirements. Nonetheless, a lot of researchers are actively working on this topic and diverse solutions have been proposed.

In order to perform resource allocation every Transmission Time Interval (TTI), MAC Scheduler requires as input the following parameters:

- *Scheduling Requests* from UE to gNB
- *Buffer Status Report (BSR)* from UE in order to assign the UL Grants
- *Power headroom reports (PHR)* are needed to provide support for power-aware packet scheduling.
- *Channel measurements* needed to decide the coding rate and modulation to be used.
- *QoS Requirements* per each flow

The scheduler at gNB side uses the above mentioned measurements for assigning the physical resources to the UEs and to set the best Modulation & Coding Scheme (MSC).

At high level view, NR Scheduling is not much different from LTE scheduling.However, new approaches have been proposed to deal with 5G requirements such as: Pre-emptive Scheduling for downlink transmissions or Configured Grant for Uplink transmission.

6.3.1 Uplink Scheduling

Uplink Scheduling is based on requesting transmission grants from the UE side. Furthermore, UE shall inform the gNB about the amount of data to be transmitted through Buffer Status Reports. Before the data transmission occurs, signaling messages are exchanged between UE-gNB.

6.3.1.1 Dynamic Uplink Scheduling

Dynamic scheduling is the mechanism that assigns the Physical Uplink Shared Channel (PUSCH) resources by Downlink Control Information (DCI). When UE has data queued in LCs (Logical Channels) and no UL grant is pre-configured, UE shall signal to gNB a Scheduling Request (SR) and Buffer Status Report (BSR). gNB provides the Uplink Grant on PDCCH and then the UE may transmit the uplink data. This procedure leads to more delay due to signaling overhead. Therefore, it shall not be applied for URLLC transmission. On the Figure 6.2 is shown the signaling flow for assigning an uplink grant.

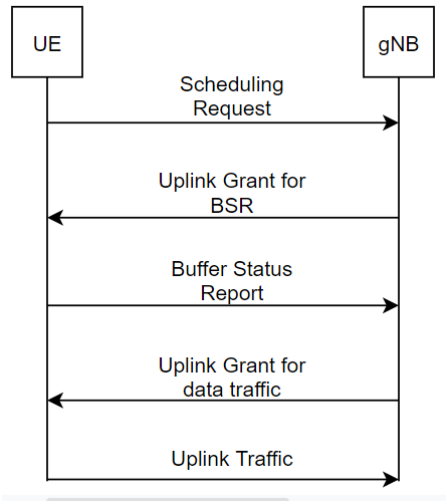

Figure 6.2: Dynamic Uplink Scheduling

6.3.1.2 Semi-Persistent Uplink Scheduling

Semi-Persistent Scheduling (SPS) is configured by RRC per Serving Cell. Multiple configurations can be active simultaneously only on different Serving Cells. SPS enables UL/DL transmissions faster and reduces the signalling overhead for scheduling grant. Furthermore, it has been proposed as a solution for scheduling low latency packets in order not to be buffered in the queue for long time.

Activation and configuration of SPS is controlled by RRC layer.

3GPP has defined 2 types of configured uplink grants:

- **Configured Grant Type 1**: The Uplink Grant is provided by RRC and stored as configured uplink grant at UE side (Figure 6.3). Configured Uplink Grant Type 1 is similar to LTE semi-persistent scheduling (SPS) where UL data transmission is based on RRC reconfiguration without any L1 signaling. RRC provides the grant configuration to UE through higher layer parameter named as Configured-GrantConfig (Refer to section 6.3.2 [5]) including the rrc-ConfiguredUplinkGrant parameter. Potentially SPS scheduling can provide the suitability for deterministic URLLC traffic pattern, because the traffic properties can be well matched by appropriate resource configuration.

- **Configured Grant Type 2**: the uplink grant is provided by Physical Downlink Control Channel(PDCCH) (Figure 6.3). On this grant type, an additional L1 signaling, Downlink Control Indication (DCI) is introduced. RRC only provides the higher layer parameter ConfiguredGrantConfig (Refer to section 6.3.2 [5]) and does not include the parameter rrc-ConfiguredUplinkGrant. The Uplink Grant is activated and deactivated through Downlink Control Indication (DCI) where the latter one can enable fast modification of semi-persistently allocated resources. In this way, it enables the flexibility of UL Grant Free transmission in term of URLLC traffic properties for example packet arrival rate, number of UEs sharing the same resource pool and/or packet size.

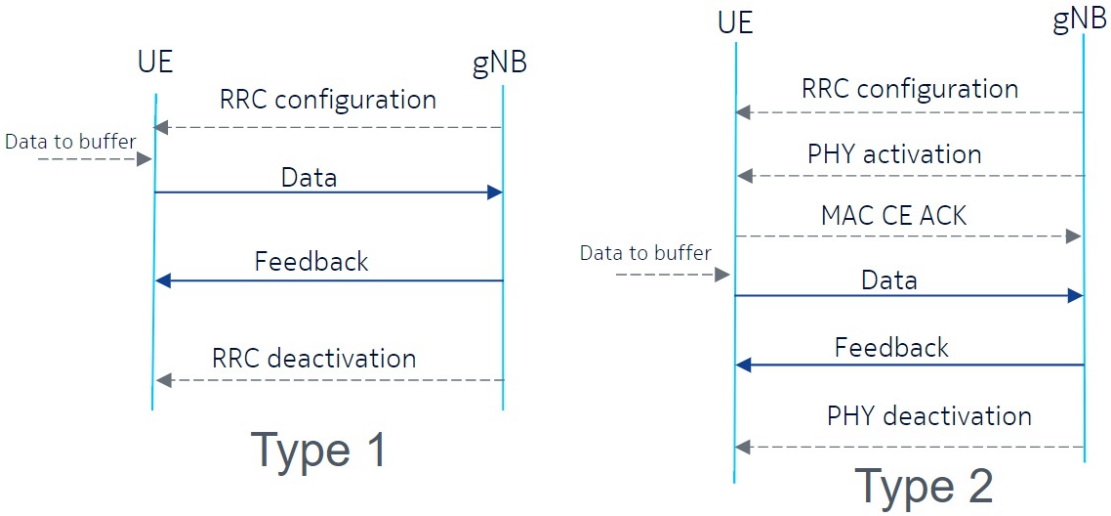

Figure 6.3: Configured Uplink Grant Type 1 & 2

Since requesting for an uplink grant includes a lot of overhead due to signalling between UEgNB, configuring the UL grant reduces the transmission latency of the packet. However, is important mentioning that the resource utilization cannot be maximized since the grants are configured in advance and the amount of data being transmitted is not known (See section 10.3 [11] & section 5.8.2 [15]).

6.3.2 Downlink Scheduling

In the downlink, the gNB can dynamically allocate resources to UEs via the Cell-Radio Network Temporary Identifier (CRNTI) on Physical Downlink Control Channel (PDCCH). The UE always monitors the PDCCH(s) in order to find possible assignments when its downlink reception is enabled (activity governed by DRX when configured). When Carrier Aggregation (CA) is configured, the same C-RNTI applies to all serving cells. The scheduling algorithm is left to implementation and it can use different measurement reports for assigning the resources such as: Signal to Noise Ratio (SINR), Channel Quality Indicator (CQI), Reference Signal Receive Power (RSRP), Block Error Rate (BLER). Some well-known scheduling algorithms are: Round Robin (RR), Weighted Round Robin (WRR), Proportional Fair Scheduling (PFS), Earlist Deadline First (EDF). WRR can be

6.3.2.1 Downlink SemiPersistent Scheduling

For the DL SPS, a DL assignment is provided by PDCCH, and stored or cleared based on L1 signalling indicating SPS activation or deactivation.

The following parameters are configured by RRC:

- *csRNTI* CSRNTI for activation, deactivation, and retransmission;
- *nrofHARQProcesses* the number of configured Hybrid automatic repeat request (HARQ) processes for SPS;
- *periodicity* periodicity of configured downlink assignment for SPS.

6.3.2.2 Dynamic Downlink Scheduling

In the downlink, the gNB can dynamically allocate resources to UEs via the CRNTI on PDCCH(s). A UE always monitors the PDCCH(s) in order to find possible assignments when its downlink reception is enabled (activity governed by DRX when configured).

6.3.2.3 Downlink Preemptive Scheduling

The gNB may preempt an ongoing (Physical Downlink Shared Channel) PDSCH transmission to one UE with a latency-critical transmission to another UE. The gNB can configure UEs to monitor interrupted transmission indications using Interruption Radio Network Temporary Identifier (INTRNTI) on a Physical Downlink Control Channel (PDCCH). If a UE receives the interrupted transmission indication, the UE may assume that no useful information to that UE was carried by the resource elements included in the indication, even if some of those resource elements were already scheduled to this UE (see section 10.2 [11]). The example shown in figure 6.4 represents the downlink shared radio channel and the transmission grants assigned to UE1 on a time-line frame. There are two UEs attached to the gNB simultaneously and UE2 is receiving delay traffic. Therefore, it shall be treated with high priority from the scheduler. When UE1 downlink data is being transmitted, gNB may receive on the buffer DL data for UE2. In order to fulfill delay requirements for UE2, it interrupts the data transmission of UE1 and send the data of UE2. On the other hand, UE1 received the scheduled transmission, where part of it is punctured. In this situation, gNB shall re transmit the portion of data lost to UE1.

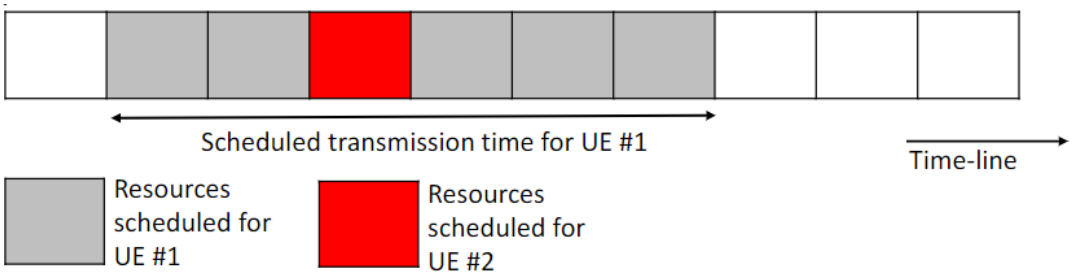

Figure 6.4: Pre-emptive Scheduling in Downlink

6.4 Logical Channel Prioritization

In order to differentiate the application data on UL transmission, Logical Channel Prioritization (LCP) is configured at the UE side. Logical Channel Prioritization (LCP) is applied for UL transmission at UE side and it is controlled by RRC entity. Radio Resource Control (RRC) configures per each Logical Channel (LC) some priority parameters that define the priority of transmission for that specific LC. Furthermore, Radio Resource Control (RRC) controls the mapping procedure of Logical Channel (LC) to be selected for UL TX with the received UL Grant based on some parameters explained in section 5.4.3.1.1 [15]:

- *allowedSCS-List* which sets the allowed Subcarrier Spacing(s) for transmission;
- *maxPUSCH-Duration* which sets the maximum Physical Uplink shared Channel (PUSCH) duration allowed for transmission

- *configuredGrantType1Allowed* which sets whether a configured grant Type 1 can be used for transmission;
- *allowedServingCells* which sets the allowed cell(s) for transmission

From the aforementioned parameters, it is possible to configure specific UL Grants for URLLC transmissions by configuring the grant type 1 and by setting a higher subcarrier spacing (SCS). LCs are arranged into Logical Channel Groups and one MAC entity can have a maximum of 8 LCGs (Logical Channel Groups).
The mechanism for selecting the appropriate Logical Channel (LC) to be served and allocating the resources is described in section 5.4.3.1 [15].

The uplink traffic scheduling at UE side based on Logical Channel Prioritization (LCP) procedure is represented in figure 6.5. Logical Channels with data traffic to be transmitted, are placed with an decreasing priority order. The MAC PDU which length is defined by the uplink grant defined by gNB, shall be filled with data from all selected Logical Channels. Therefore, Prioritized Bit Rate defines the amount of data to be served first per each Logical Channel (LC). On the second round, all the data remained in the buffer for high priority Logical Channel (LC) will be served until the UL grant resources are exhausted. In this way, high priority data is served first for meeting the KPIs and the low priority Logical Channel (LC) are not starving (see section 5.4.3.1.3 from [15]).

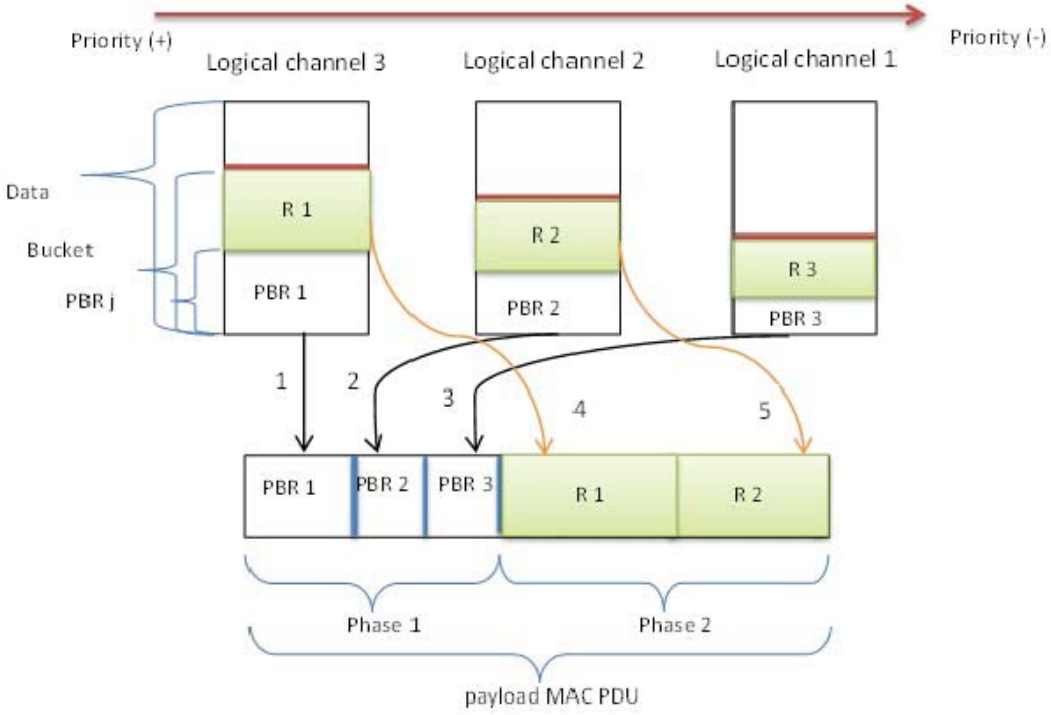

Figure 6.5: Logical Channel Prioritization procedure at UE. Figure from cite[]

6.5 Buffer Status Reporting

Buffer Status Reporting (BSR) is provided from UE to gNB in order to update with amount of data buffered per each LC. The procedure is controlled by RRC which configures some parameters per MAC entity (see section 5.4.5 [15]). Furthermore, 3GPP has

defined also the requirements for triggering Buffer Status Reporting (BSR) on particular scenarios and which type of Buffer Status Reporting (BSR) to be used. The rules are defined for MAC Sublayer at UE side since that protocol shall generate the Buffer Status Reporting (BSR) based on the amount of data stored on the buffers.

6.6 Research on 5G Scheduling Algorithms outside of 3GPP

On the previous sections, it is described the scheduling procedure in Uplink and Downlink direction, mentioning also the solutions proposed by 3GPP e.g. Configured Grant for Uplink and Pre-emptive Scheduling in Downlink. However, the above mentioned techniques have been proposed for URLLC use-case mostly, without restricting the implementation of MAC sublayer at gNB. Since 3GPP has not defined any framework for scheduling algorithm at gNB side, it allows the vendor to experiment and implement different solutions. However, in order to ensure interoperability between vendors, 3GPP has set the requirements for UE implementation. In this way it ensures that all UEs shall be served despite the gNB vendor.

Several research groups such as EUROCOM, Polytechnic University of Cataluna, Fondazione Bruno Kessler, are working on this topic and a 2-Level Scheduling mechanism has been proposed for supporting Network Slicing. Since Network Slicing implies the creation of logical networks on top of the same physical infrastructure, it is needed first to share the physical resources between all active slices. Therefore, a Common Scheduler can be implemented for assigning the resources to each slice. Then, a second level scheduling, Slice Specific Scheduler, shall handle the requests of UEs between each slice and assign the resources to them. Each level of scheduling can run different algorithms based on the service requirements. Polytechnic University of Cataluna and Fondazione Bruno Kessler [35] have suggested as common scheduler to be used Round Robin (RR) algorithm that treats the slices equally by assigning the same portion of the available resources. On the other hand, Weighted Round Robin (WRR) algorithm can also be applied by enforcing the percentage parameter and treating the slices based on the priority. The last solution makes possible to enforce the desired distribution of resources in terms of Physical Resource Blocks (PRBs). Whereas, for second level scheduler or slice specific scheduler, different scheduling algorithms can be applied per each slice based on the service KPIs. Therefore, for eMBB slices the Proportional Fair Scheduling (PFS) may be used to handle each UE in a fair manner and to optimize resource utilization. But also a Round Robin (RR) algorithm can be applied since the UEs shall be treated equally within the same slice slice. Whereas, for mMTC slice it is proposed to apply again PFS scheduling and treating all the connected devices equally. Another proposal would be to apply Semi-Persistent Scheduling or Configured Uplink Scheduling since those devices have periodic transmission and do not require high demands for throughput or latency. On the other hand, URLLC slices have strict requirements for low latency and high reliability transmission and a Round Robin or Proportional Fair Scheduling (PFS) does not fulfill the service KPIs. Eurocom [36] has proposed to apply Earliest Deadline First (EDF) algorithm for URLLC slices which ensure to meet the latency requirements of the traffic. Furthermore, their solution implies running different algorithms for each transmission direction at the slice level. The proposed solution (see section III / B from [36]) serves the traffic based on the slice KPIs and improves network performance.

2-Level Scheduling Algorithm proposed by EUROCOM is shown in figure 6.6. The Common Scheduler it is named as Common Resource Manager (CRM) whereas the Slice Specific Scheduler is named as Slice Specific Resource Manager (SSRM). Their

solution implies that the common scheduling algorithm is connected with the Network Orchestration Layer in order to receive commands for upgrading the Modulation & Coding Scheme (MCS) value and the Physical Resource Blocks (PRBs) to be assigned per every slice. The Orchestration Layer, which is composed of Slice Agents and Slice Orchestrator, shall be connected with every slice, monitor & report if the KPIs are being met. When the service requirements are not being fulfilled, the Orchestration Layer shall instruct the Common Scheduler (CRM) to assign the resources of the affected slice differently (e.g. decrease Modulation & Coding Scheme (MCS) value for ensuring the reliability of the packets).

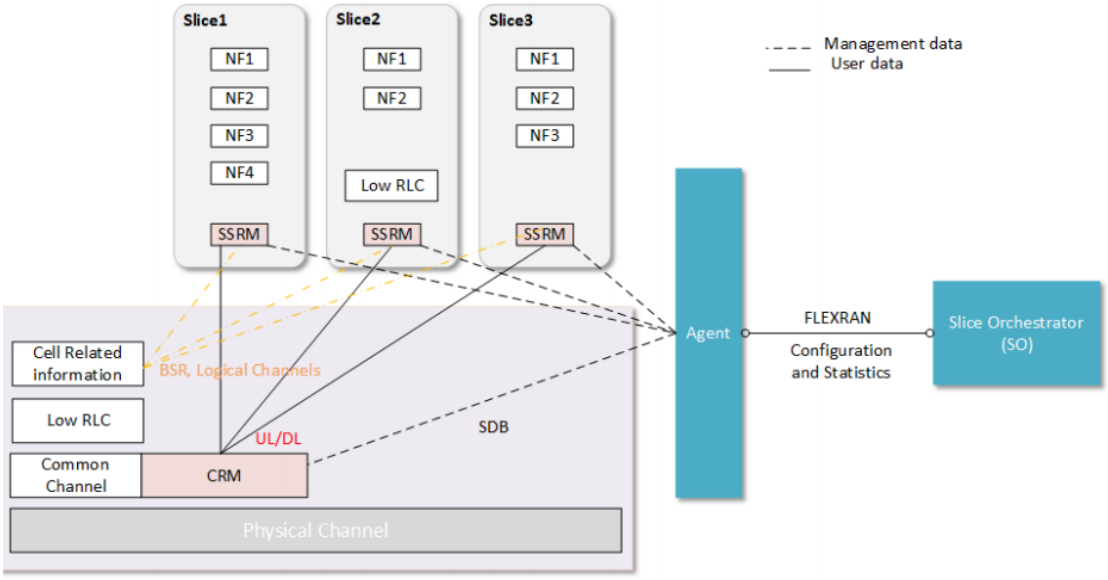

Figure 6.6: 2-Level Scheduling Model by EUROCOM. Figure 2 from [36]

6.7 Summary

This chapter has briefly described the MAC Layer at UE and gNB side and the enhancements for supporting Network Slicing. MAC Layer contains the Scheduling Algorithm which shall be adopted accordingly for serving the Network Slices available in the system. 3GPP has standardized the framework for MAC sublayer at UE side and several procedures are defined such as: Logical Channel Prioritization, Buffer Status Reporting, Uplink Grants. In this way, the vendors working on gNB implementation shall follow those rules for serving UEs of different vendors. Furthermore, the scheduling algorithm which is the most crucial part of the network, is left to vendor's implementation. 3GPP has proposed some techniques to be used for dealing with delay sensitive traffic, e.g. Pre-emptive Scheduling in Downlink or Configured Uplink Grant. Since this is a crucial part for supporting Network Slicing, several research groups are involved and a 2-Level Scheduling mechanism has been proposed. The first level scheduler or Common Scheduler shall assign the physical resources to each active slice and then another slice-dedicated scheduler will handle the UEs. The scheduling algorithms can be selected in accordance with the service requirements of the slice etc.

Chapter 7

Simulation Environment

7.1 PAL Setup Introduction (Confidential)

7.2 System Parameters (Confidential)

7.3 Ideal Simulation Scenarios

This section describes emulation scenarios which are meaningful from researches point of view. The main goal is to evaluate the performance in terms of achieved throughput, packet delay, packet loss and CPU usage of the components when Network Slicing and/or QoS is enabled.
From the context of PRIMO-5G project, the following use-cases are derived: URLLC and eMBB. Therefore, one possible scenario (Figure 7.1) would be to enable one slice of each type: URLLC and eMBB with/without particular QoS Flows to evaluate the network performance when UL/DL traffic is being generated.

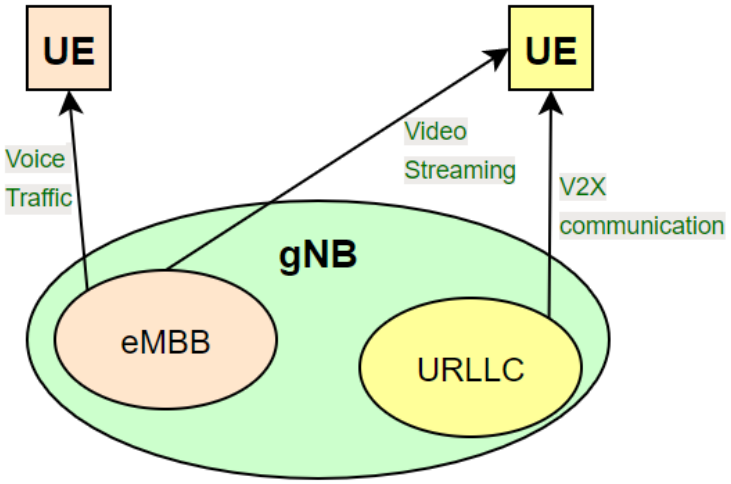

Figure 7.1: Scenario - 2 slices & QoS Flows Enabled

Furthermore, enabling the QoS Flows with 5QI values listed in the Table 7.1 can ensure that the KPIs are met for the following application types: videostreaming, voice traffic and Vehicle to Everything communication (V2X). It is important to perform the same emulation scenario first without QoS and then with the QoS flows in order to evaluate the impact that those features would have on network performance (delay, throughput, packet loss).

5QI	**QoS Flow Type**	**Application**
1	GBR	Conversational voice (Rescue team)
75	GBR	V2X messages
7	Non-GBR	Live streaming
GBR: Guaranteed Bit Rate, Non-GBR: Non Guaranteed Bit Rate		

Table 7.1: QoS Flows - Table 5.7.4-1 from [1]

Another scenario would be enabling two slices (URLLC + eMBB) with shared resource type and evaluate if the KPIs for URLLC traffic would be fulfilled.

For emulating voice traffic, UDP traffic could be generated with small packet size. In contrast, for video streaming application UDP traffic would be used again but with long packet size. Instead, for control messages, TCP traffic could be generated since re-transmissions imply high reliability.

Through the sub scenarios shown below (Table 7.2, it would have been possible to evaluate the impact of Network Slicing and/or QoS on network performance and on achieving the Key Performance Indicators (KPIs) of each slice. Furthermore, by sharing the radio resources between an URLLC and an eMBB slice, would have been possible to evaluate the network performance under normal and congested state. The target would be to fully serve the UE that would receive control messages (i.e. URLLC use-case) and maintain stable the voice traffic connection (maximum packet loss 1%).

Scenario	**Nr. of Slices**	**QoS**	**Slice Type**	**Slice Resource Types**
1	1	Default	eMBB	Isolated
2	1	Enabled Table 7.1	eMBB	Isolated
3	2	Enabled Table 7.1	eMBB	Isolated
4	2	Enabled Table 7.1	eMBB+URLLC	Isolated
5	2	Default	eMBB+URLLC	Isolated
6	2	Enabled Table 7.1	eMBB+URLLC	Shared
7	3	Enabled Table 7.1	eMBB+URLLC	Isolated

Table 7.2: Summary of Ideal Simulation Scenarios [1]

7.4 Selected Scenarios

Due to system capabilities described in the previous sections 7.2, the scenarios described on the previous section 7.3 have been adopted to the following ones:

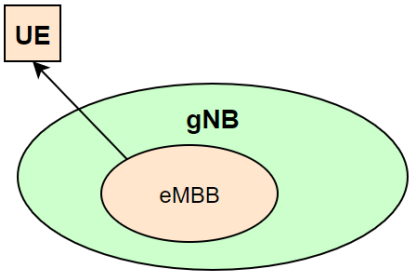

Figure 7.2: Sub-Scenarios with 1 eMBB Slice

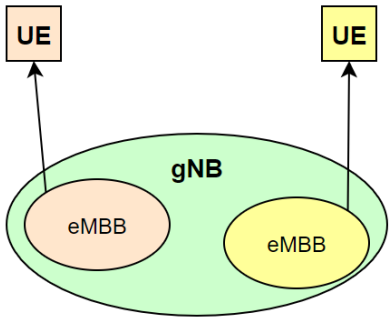

Figure 7.3: Sub-Scenarios with 2 eMBB Slices

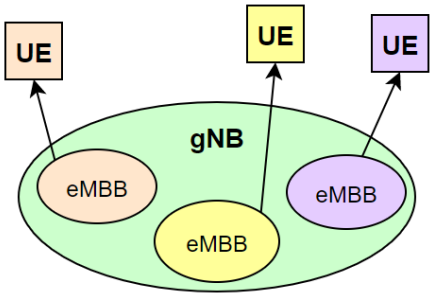

Figure 7.4: Sub-Scenarios with 3 eMBB Slices

Only eMBB slices have been enabled and the system performance is tested with QoS Flows enabled and disabled. The main purpose of the practical analysis was to validate the functionality of Network Slicing and QoS features on the 5G Emulation System. Then, the network performance in terms of achieved throughput, packet loss and CPU Usage when Network Slicing and/or QoS is enabled, is evaluated.

7.4.1 Network Slicing - Scenario

The first scenario to be tested implies the configuration of Network Slices eMBB type only. This scenario itself contains 3 sub-scenarios which are represented in the figures 7.2, 7.3 and 7.4. The summarized scenarios with Network Slicing enabled can be found in table 7.3 and the tests are Downlink-egocentric. The QoS feature will be disabled. 22 Physical Resource Blocks (PRBs) are assigned to each slice, where 11 PRBs are used for UL and 11 PRBs are used for DL transmission.
First, one eMBB slice type is enabled and one UE is attached with the network slice.

Downlink UDP traffic is generated from the Video Server towards the UE and the bandwidth is set to 10 Mbps. The same test is repeated but with two eMBB slice enabled and 1 UE attached to each slice and then for three eMBB slices, as it is summarized on the Table 7.3. The results in terms of received throughput, Jitter and Packet Loss are collected and evaluated for each sub scenario. Furthermore, the CPU Usage of the gNB and UESIM components is also evaluated. It is expected that Network Slicing feature would require more system resources mostly at gNB components since Distributed Unit (DU) shall handle the scheduler and resource sharing. In order to better evaluate the impact of Network Slicing in the system, the same test is performed but without Network Slicing enabled.

The configuration of Network Slicing feature is explained in the following section 7.5.

The results and the evaluations are described in the following section 7.8.

Sub-scenario	**Nr. of Slices**	**QoS**	**Slice Type**	**Slice Resource Types**	**Traffic Bandwidth**
1.1	1	Disabled	eMBB	Isolated	10 Mbps
1.2	2	Disabled	eMBB	Isolated	10 Mbps
1.3	3	Disabled	eMBB	Isolated	10 Mbps
1.4	NO	Disabled	NO	NO	10 Mbps

Table 7.3: Default Scenarios with Network Slicing

7.4.2 Quality of Service Scenario

The second scenario implies the activation of Quality of Service feature. In this scenario only 1 Network Slicing eMBB type is configured. The QoS Flow, Guaranteed Bit Rate type, is enabled and the Guaranteed and the Maximum Bit Rate are set to 10 Mbps. The above parameters refer to the physical layer, therefore the limitation of the bit rate is done at the Physical Layer and not at Application Layer.
The tested sub-scenarios with Network Slicing and QoS feature enabled are summarized in table 7.4.

Sub-scenario	**Nr. of Slices**	**QoS**	**Slice Type**	**UDP Bandwidth**	**Packet Length**
2.1	1	Enabled	eMBB	8 Mbps	1400 B
2.1	1	Enabled	eMBB	10 Mbps	1400 B
2.1	1	Enabled	eMBB	10 Mbps	100 B
2.7	1	Enabled	eMBB	15 Mbps	1400 B

Table 7.4: Scenarios with Quality of Service

From the sub scenarios described above, will be verified if the network accomplished to provide the guaranteed bit rate.
The results and outcomes are discussed in the following section 7.8.

7.4.3 Various Packet Length Scenario

Packet length impact on system performance will be evaluated by performing several test with/without QoS feature. The packet length parameter can be considered for generating control messages on URLLC use case. On sub-scenarios 3.1-3.4 the Network Slice eMBB type is enabled. Whereas, the last sub-scenarios shown on the Table 7.5 are

performed without Network Slicing enabled. The goal of these tests is to highlight the impact that packet length has on the CPU Usage at gNB and UESIM components and the application throughput . The configuration of Quality of Service feature is explained in the following section 7.6.

Sub-scenario	Nr. of Slices	QoS	Slice Type	UDP Bandwidth	Packet Length
3.1	1	Enabled	eMBB	10 Mbps	1400 B
3.2	1	Enabled	eMBB	10 Mbps	1000 B
3.3	1	Enabled	eMBB	10 Mbps	800 B
3.4	1	Enabled	eMBB	10 Mbps	500 B
3.5	1	Enabled	NO	800 Mbps	1400 B
3.6	1	Enabled	NO	800 Mbps	1000 B
3.7	1	Enabled	NO	800 Mbps	800 B
3.8	1	Enabled	NO	800 Mbps	500 B

Table 7.5: Scenarios with various packet length

7.4.4 Scenarios with different MCS index

Another parameter that can be tuned is Modulation & Coding Scheme index which is directly related with the Modulation scheme and Coding Rate. This parameter has a strong impact on the achieved throughput. The system supports only the values taken from Table 1 & 2 (see Table 5.1.3.1-1 & Table 5.1.3.1-2 from [16]). The selected Modulation & Coding Scheme (MCS) indexes to be tested are the lised in the following table 7.6:

Sub-scenario	Nr. of slices	MCS Index	Modulation Order (Qm)	Target code Rate (R)	Spectral efficiency
4.1	1	7	2	526	1.0273
4.2	1	10	4	340	1.3281
4.3	1	17	6	438	2.7305
4.4	1	28	6	948	5.5547

Table 7.6: MCS index tested. Table 5.1.3.1-1 [16]

On this scenario only one eMBB slice is enabled and the QoS flows are disabled.
The results obtained from the above mentioned scenario are shown on the following section 7.8.

7.5 Network Slicing Configuration (Confidential)

7.6 QoS Configuration (Confidential)

7.7 Traffic Generation

From the Video Server it is generated downlink UDP traffic towards the active UEs. Iperf is a network testing tool, built on a client/server model, can be used to measure

maximum UDP and TCP throughput between the client and server stations. Iperf command is used for generating application traffic and has several parameters that can be tuned:

- -u : UDP traffic
- -s: Packet size
- -b: Bandwidth for UDP traffic

Iperf generates application traffic the measurements are done in the transport layer. Therefore, the packet size set with the above mentioned command, includes also the header from transport protocol.

Iperf reports the following results:

- Jitter (latency variation)
- Datagram loss
- Bandwidth
- Reported Interval (set to 1 second)

By tuning the bandwidth of UDP traffic, iPerf creates a constant bit rate UDP stream. The server detects UDP datagram loss by ID numbers in the datagrams. To measure packet loss instead of datagram loss, the datagrams should be small enough to fit into one packet. That can be achieved by using the s parameter. Jitter is the smoothed mean of differences between consecutive transit times. The jitter is the latency variation and is calculated as the differences between consecutive transit times. Therefore, it does not depend on the latency itself.

The results reported from iperf at Transport Layer and also the results in terms of Throughput reported from the UESIM Terminal (measurements gathered at the Physical Layer) are collected at UESIM component.

Instead, for measuring the packet delay, PING command is used. Ping uses ICMP protocol to generate ECHO messages and based on the timestamp on ECHO REPLY packet, it calculates the Round Trip Time (RTT). RTT is reported in milliseconds (ms) and determines the total time required for the packet to be received from the UESIM, processed and sent back to the sender. PING command allows to tune the packet size through the parameter s. The RTT of the packet received is reported every 1 second . Furthermore, at the end of the test it reports how many packets have been lost. PING runs on top of UDP protocol.

7.7.1 Packet Analysis on RAN Stack

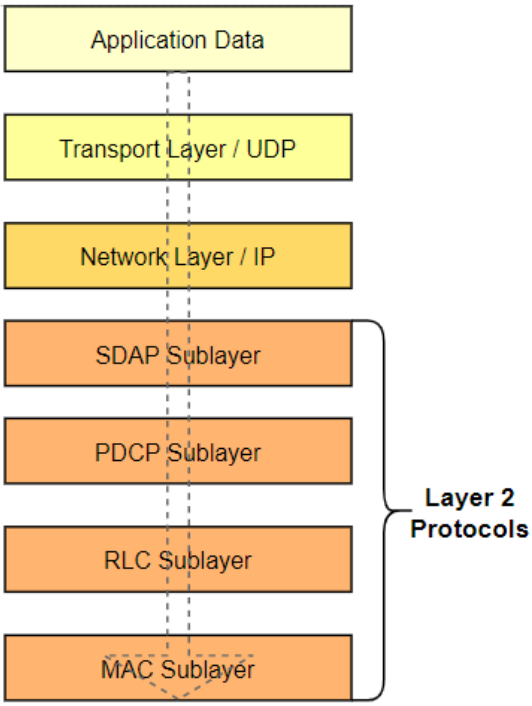

Figure 7.5: Packet Analysis on RAN Protocol Stack

Packet length by default has been set to 1400 B which includes the application traffic and UDP packet header of 8 B. Since the packet is further encapsulated on the IP layer and Radio Protocol Layers (Figure 7.5), the packet length transmitted on the abstracted radio channel will be more than 1400 B. In order to avoid packet fragmentation at Network Layer, the IP Layer Protocol Data Unit (PDU) shall be smaller than 1500 B, which is the value of Maximum Transport Unit (MTU) parameter. MTU value in the system is configured by default to 1500 B.

In the Figure 7.6 is represented the data flow and how the transport Block at MAC Layer is created. One Transport Block may contain more than one MAC SDU, depending on their length and on transport block size. In the figure are shown four IP flows mapped in two Radio Bearers. The IP packet of each flow is processed by the radio protocol layer and extra header is added before sending it to the physical layer.

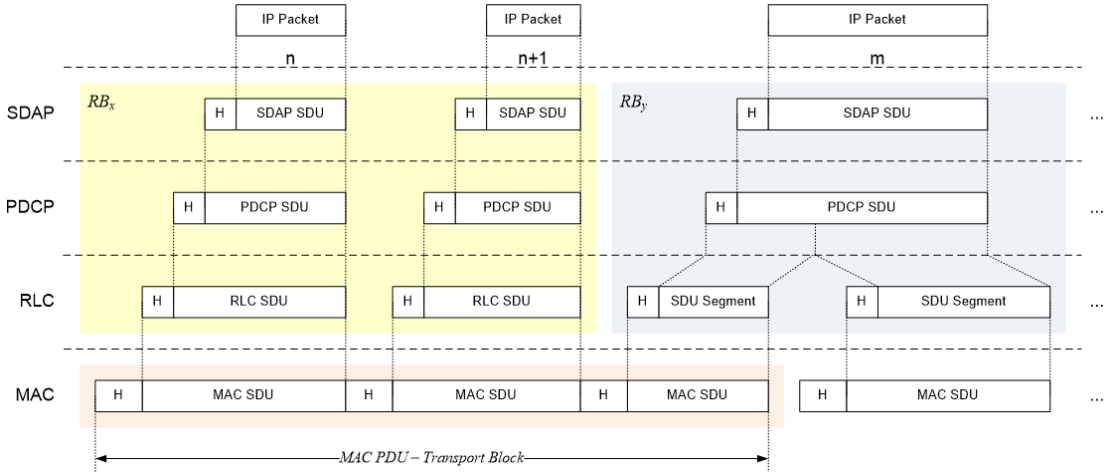

Figure 7.6: L2 Data Flow Example. Figure 6.6-1 from 3GPP TS38.300 [11]

7.8 Simulation Results Partially Confidential

The simulation framework for analyzing the performance of the Network Slicing and QoS under different configurations was defined in section 7.4. The present chapter will first focus on presenting results for each of the simulation sets defined. Furthermore, the results obtained from different tests will be compared to better evaluate the impact on system performance parameters.

7.8.1 Network Slicing - Default Use-Case

In the following Table 7.7 are shown the results obtained from tests described in 7.3. Throughput, jitter and packet loss results are gathered from IPERF application. Each slice is a logical network using 11 downlink Physical Resource Block but running on the same system components. From the performed tests, can be proven that the slices are working as independent network and the performance in terms of slice throughput, jitter or packet loss, is not affected by other slices running at the same time. The UEs are always receiving 10 Mbps, the same bandwidth that was initially generated by the Video Server. Furthermore, the jitter which represents the delay variance between consecutive packets, has not been changed which means that the traffic is being treated and handled equally by the system.

Sub scenario	**UE Nr**	**Throughput_Iperf (Mbps)**	**Jitter (ms)**			**Packet Loss (%)**
			MIN	MAX	AVG	
1.1	UE1	10	0.067	0.103	0.079	0
1.2	UE1	10	0.063	0.091	0.075	0
	UE2	10	0.063	0.091	0.075	0
1.3	UE2	10	0.073	0.114	0.092	0
	UE2	10	0.062	0.124	0.088	0
	UE3	10	0.086	0.138	0.103	0

Table 7.7: Default UseCase Throughput Results

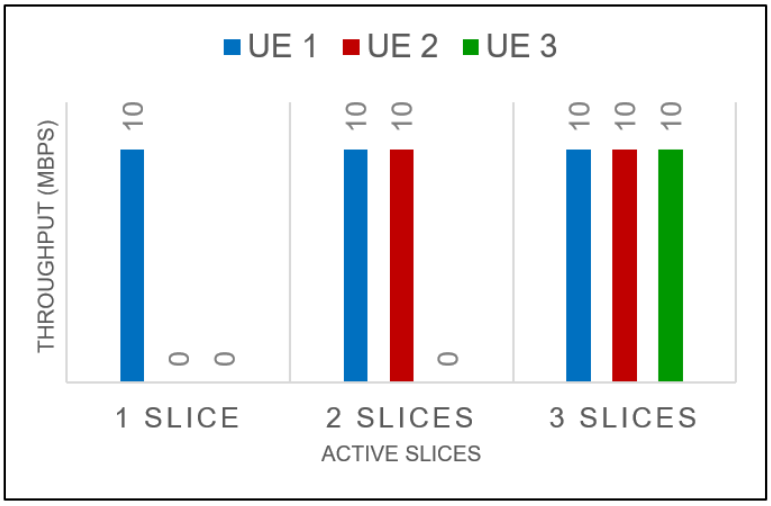

Figure 7.7: Network Slice Throughput - Default Scenarios 7.3

Network Slicing feature requires more processing tasks to be handled at CU and DU. gNB shall enable the slices and share the radio resources between them. Furthermore, DU component shall handle also scheduling procedure for each slice separately which implies more tasks and system resources.

The received throughput of UE's is not affected when there are more slices enabled highlighting the traffic isolation between the slices.

Author's Note: CPU Usage Results are confidential

7.8.2 Quality of Service Scenario

Quality of Service is enabled in order to ensure a set of QoS parameters per each QoS Flow. In the selected scenarios described in the previous section 7.4, it is mentioned that the maximum allowed bit rate and guaranteed bit rate of the QoS Flow is set to 10 Mbps. The latter parameters are limited in the physical layer, meaning that total amount of traffic allowed per one UE is controlled in the Physical Layer. Hence, in order to ensure that UE receives all the data sent from the Video Server, the bandwidth of UDP traffic which represents only the application data, shall be less than 10 Mbps. The generated traffic is processed and the packet headers are added right before transmitting it to the Physical Layer.

In the first test shown in the table 7.8, 8 Mbps application traffic through Iperf is generated. The impact of packet headers is observed by measuring this traffic in Physical Layer where we receive 8.18 Mbps. Since the amount of traffic at the Physical Layer is less than the Maximum Allowed Bit Rate (10 Mbps), the User Equipment has received the data with 0 % packet loss. Whereas, when the amount of traffic generated by Video Server is equal to 10 Mbps (refer to second subscenario in table 7.8, the UE receives less application data because of dropped traffic at the buffers since the total amount of traffic sent in the Physical Layer was more than the Maximum Allowed Bit Rate. Furthermore, when the packet length is smaller (sub scenario 2.3), the received data is less than sub scenario 2.2 which highlights the impact of packet headers in the application throughput.

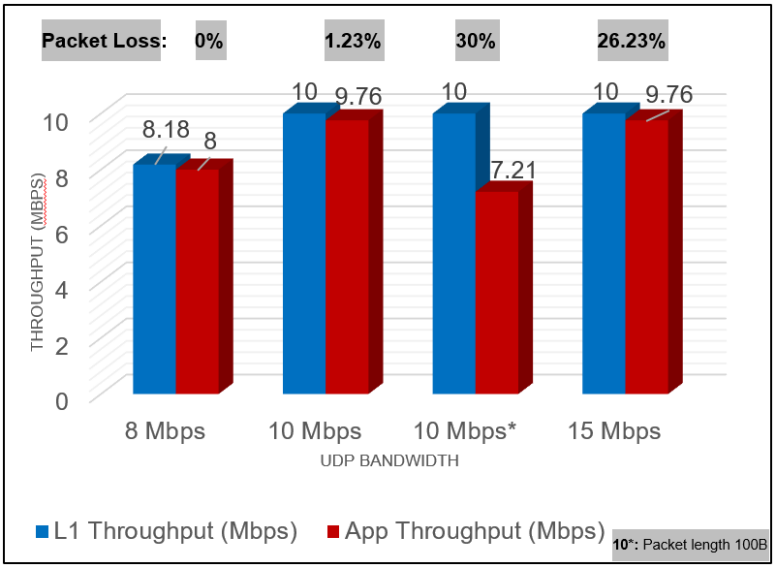

Figure 7.8: QoS Impact on Throughput

Sub scenario	Packet Length	UDP Bandwidth (Mbps)	L1 Throughput (Mbps)	Throughput_Iperf (Mbps)	Packet Loss (%)
2.1	1400	8	8.18	8	0
2.2	1400	10	10	9.76	1.23
2.3	100	10	10	7.21	30
2.4	1400	15	10	9.76	26.23

Table 7.8: QoS Use-Case - Throughput Results

Despite the fact that UE did not receive all the application traffic initially generated, the Bit Rate in the physical layer has been limited and guaranteed for the QoS flow based on the QoS Profile configuration.

7.8.3 Various Packet Length Scenario

The results described on this section are gathered based on the scenarios described on the previous section 7.4.3. In the section 7.7.1 is analyzed the packet format including the headers which have been introduced by the protocol layers. In the following table 7.9 are shown the results related to the impact of headers (i.e. different packet length) on the L1 throughput. As the packet length becomes shorter (from 1400 Bytes to 500 Bytes), the Bit Rate measured at L1 increases significantly. Furthermore, the packet headers have more impact when the amount of traffic generated per unit of time is bigger, as it can be noted when 800 Mbps is being generated. The results shall be taken into consideration for different use-cases that might be tested in the future and to ensure that the system has enough resources to handle the traffic. As an instance, small packets can be used for emulating URLLC traffic (command messages) or voice traffic. Depending on the number of active traffic flows in the network and total amount of application traffic to be generated, the packet length impact has to be considered so the network will not be congested.

Sub scenario	Packet Length	L1 Throughput (Mbps)	Throughput_Iperf (Mbps)	Packet Loss (%)
3.1	1400	10.22	10	0
3.2	1000	10.31	10	0
3.3	800	10.39	10	0
3.4	500	10.62	10	0
3.5	1400	817.7	800	0
3.6	1000	824.8	800	0
3.7	800	831	800	0
3.8	500	849.6	800	0

Table 7.9: Packet Length impact on Throughput

Author's Note: CPU Usage Results are confidential

Small packets have a strong impact also on the CPU usage of UE and gNB in particular. Along with forwarding the data, gNB shall perform traffic management (traffic shaping, traffic scheduling), resource scheduling between the UEs etc. As the packets tend to become smaller, gNB shall process much more packets in a unit of time in order to forward the traffic towards UE. Whereas at the received side (i.e. UE), the packets need to be processed, reassembled and forwarded to higher layers.
On the other hand, small packet length has also a strong impact on increasing the Physical Layer throughput significantly as the amount of generated traffic increases. As it is represented in the figure 7.9, for the same amount of application traffic generated, 10 Mbps, the throughput in Physical Layer tends to increase as the packet length is smaller.

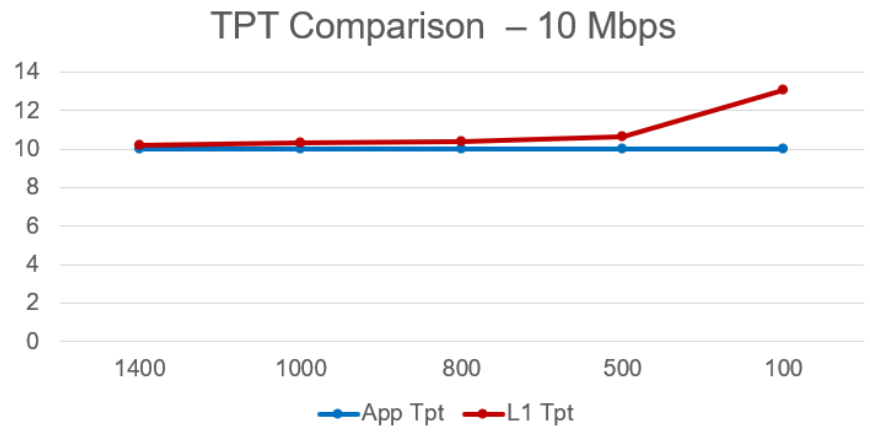

Figure 7.9: Iperf vs L1 Throughput Comparison - Packet Length Sub scenario (10 Mbps)

7.8.4 Different MCS Index Scenario

The results gathered after performing the same test for four different Modulation & Coding Scheme (MCS) are shown in the following table tab:TBS and Maximum estimated TPT. From the gathered results and as illustrated in the figure **??**, the CPU Usage of UESIM and gNB component is constant and it is not affected by MCS index itself. However, since bigger MCS index allows achieving higher throughput, it can be said that the CPU Usage is indirectly impacted by the configured MCS value.

MCS Index	Mod. Scheme	Spectral Efficiency	TBS (bits)	estimated TPT (Mbps)	MAX TPT Achieved (Mbps)
7	QPSK	1.0273	3254	26	37.82
10	16 QAM	1.3281	4207	33	48
17	64 QAM	2.5664	8130	65	92
28	64 QAM	5.5547	17597	140	204

Table 7.10: TBS and Maximum estimated TPT

Author's Note: CPU Usage Results are confidential

In the table 7.10 is shown the maximum throughput achieved throughput by one Slice when different MCS indexes are configured. As it is expected, bigger values of MCS index allow the system to achieve higher throughput. This comes due to the fact that higher MCS indexes apply higher modulation schemes such as 64 QAM.

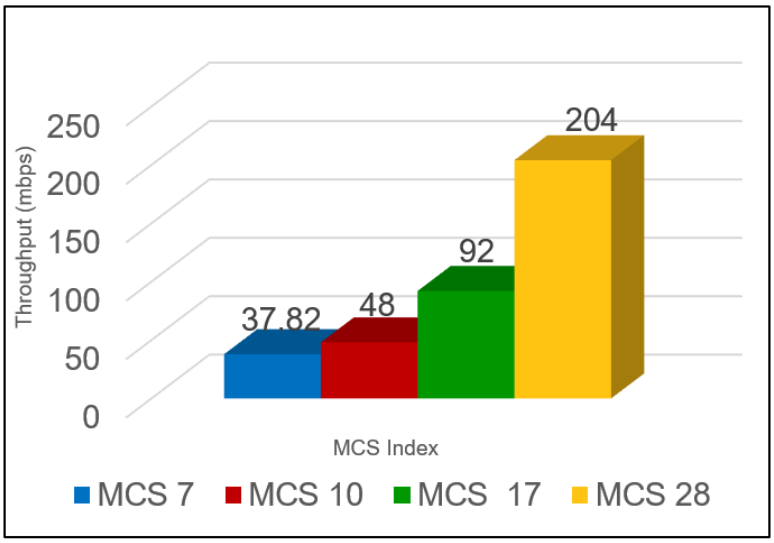

Figure 7.10: MCS Impact on Maximum Achieved Throughput

Chapter 8

Conclusions

The 3rd Generation Partnership Project (3GPP) is an international organization which develops protocols for mobile telecommunications since 2G. The organization itself is a union of members across the entire eco-system and consists of chipset, handset, infrastructure companies as well as network operators and regulation bodies, that keep unifying standardization efforts towards higher-quality communications. 5G New Radio (NR) is the first cellular standard covering millimeter-wave (mmWave) frequencies. Furthermore, the 5G network is thought as a flexible and programmable solution for network operators making possible the configuration of logical networks knows as Network Slicing. Since International Telecommunication Union (ITU) has defined three different use cases for 5G network eMBB, URLLC, mMTC with strict requirements, it is difficult and inefficient to run those in one network. Therefore Network Slicing has been introduced as a solution for serving all applications in on different logical networks but on the same physical network. Network Slicing can be enabled through the virtualization of network functions (VNF) and Software Defined Network (SDN). In this way each network slice is configured to fully serve one particular application, e.g. eMBB or URLLC, and fulfill the KPIs. On the other hand, Quality of Service can be applied to ensure the requirements are met for the above mentioned use cases and related applications/verticals.

This thesis analyzed network slicing guidelines also considering functional split options defined by 3GPP as well as ORAN aspects. Furthermore, QoS Framework in 5G Systems has been analyzed. Another important aspect investigated during this thesis is the impact that Network Slicing and QoS have on Radio Access Network (RAN) Protocol, with the specific focus on MAC Layer. This theoretical work has been applied to a physical network and emulations for different scenarios have been carried.

3GPP specification regarding 5G system architecture, Core Network and Radio Access Network part, is analyzed in the first part. The Next Generation - Radio Access Network (NG-RAN) and 5G Core Network Architecture have been defined by 3GPP in Release 15. Release 15 corresponds to the standards enabling the first phase of 5G deployment. The present interpretation offers the reader a clear understanding collected from multiple specification documents. This research assembles all necessary chunks to provide an overview of 5G network and also thorough references to find further implementation-oriented details. 5G Core Network reference architecture has been set in Release 15, along with the control plane functionalities and the reference interfaces between the components. Service-Based Architecture has been introduced as a new communication method between control plane functionalities where each network functions can expose its services on the service-based interface and the other authorised functions can access it. Each network component can be implemented as a physical network functions or as

virtual functions running on a general-purpose server.

RAN architecture is defined in 3GPP based on Split Option-2 or, so called, PDCP-RLC split. Therefore, the interfaces for connecting the CU and DU component have been set for this option. However, several functional split option are proposed by 3GPP but without well-defined interfaces. The defined architecture is a logical architecture which sets how the protocol layer will be implemented between the components. The real deployment is left to implementation and the network operators can chose to run the CU & DU on different Data Centers or the same one based on service requirements. This provides more freedom to the network operator.

In the second part, Network Slicing analyzed the guidelines given by 3GPP Rel-15 regarding requirements for implementing it in 5G Systems. The network slice is a composition of access network resources, core network resources and transport network. Therefore, in order to enable one slice, configurations are required at RAN and Core Network as well. Furthermore, managing and assigning the radio resources requires a centralized orchestration component. 3GPP has set only the requirements of the Orchestration & Management entities but has not defined a framework. This is being defined by O-RAN Alliance, which is another organization driven by network operators researching on topic not covered by 3GPP. O-RAN is working to defined the framework for implementing O&M in the network. Therefore, 3GPP in Release 16 has defined only requirements, network slicing creation, communication messages between network components and in Release 17 they are discussing the implications in RAN.

In the third part, Quality of Service in 5G systems is analyzed based on the specifications defined by 3GPP. Quality of Service allows the differentiation of application traffic for one User Equipment. Therefore, the traffic flows shall be treated and handled with different priorities in the network assuring the service KPIs. Three types of QoS Flows are defined: Guaranteed Bit Rate (GBR), non-Guaranteed Bit Rate (non-GBR) and Delay-critical Guaranteed Bit Rate (GBR). GBR and Delay-critical GBR requires guaranteed bit rate resources. Each QoS Flow type has associated also a set of QoS Parameters to be provided in the physical layer such as delay or packet loss. The research done for this technology, provides an overview of Quality of Service in 5G regarding the QoS mapping procedures and refers to specifications for more details on implementation.

Researched Scheduling algorithms in 5G have been described by referring to several research papers regarding the optimal scheduling to be implemented at gNB. 3GPP organization defines only the requirements for the User Equipment whereas the architecture and procedures of gNB MAC Layer are left to vendor implementation. Which means that the scheduling algorithm itself is not discussed by 3GPP and the vendors have to come with their own solutions. Logical Channel Prioritization, Buffer Status Report procedures have been specified for MAC Layer at User Equipment. Regarding scheduling in 5G it shall be slightly updated since logical networks will operate on top of the same physical infrastructure. Therefore, 2-Level Scheduling Algorithms have been proposed by several research organizations as a solution to manage the radio resources between logical slices and within each slice. In order to better serve delay sensitive traffic, 3GPP specified means for semi-persistent scheduling on Uplink where periodic slots are assigned to the UE for data transmission. Whereas in Downlink it has dito: means specified for pre-emptive scheduling by interrupting the on-going data transmission and serve the delay sensitive data. However, the scheduling algorithm implemented at the gNB is up to vendors and the organizations are proposing few requirements to improve

the service.

In the next part, the theoretical aspects from the previous parts have been applied to a physical setup. However, this setup w/ productive code is not yet full-featured as 3GPP specification would allow but focuses on validating Network Slicing and Quality of Service in the system. For that purpose, emulation scenarios have been derived from the supported features of the system. In order to enable Network Slicing and QoS features, all the network components starting from SMF, AMF at Core Network, CU, DU at gNB and UESIM have been configured appropriately. From the individual results obtained for each simulation set tested, the Network Slicing allows to enable isolated and independent logical networks that fully serve Users. Furthermore, Quality of Service ensures the guaranteed bit rate on the physical layer and treats serves in a fair manned the QoS Flows on the same type.

Network Slicing in 5G will be one of the key components that will enable flexibility and act as a unified network framework to provision cost effective, reliable, service-guaranteed and secure network services to various industries enabling network operators to sell tailored slices of network functionality to different types of end users where they can quickly configure and operate smarter networks to serve dynamically bandwidth or latency sensitive applications and services.

References

[1] *3GPP TS 23.501 V16.1.0*; System Architecture for the 5G System; Stage 2. (June 2019)

[2] *3GPP TS 23.502 V16.3.0*; Procedures for the 5G System (5GS); Stage 2. (December 2019)

[3] *3GPP TS 28.530 V15.1.0*; Concepts, use cases and requirements. (December 2018)

[4] *3GPP TR 28.801 V15.1.0*; Study on management and orchestration of network slicing for next generation network. (January 2018)

[5] *3GPP TS 38.331 version V15.9.0*; NR; Radio Resource Control (RRC) protocol specification. (March 2020)

[6] *3GPP TR 38.801 version 14.0.0*; Study on new radio access technology: Radio access architecture and interfaces (March 2017)

[7] *3GPP TS 38.806 version 15.0.0*; Study of separation of NR Control Plane (CP) and User Plane (UP) for split option 2 (December 2017)

[8] *3GPP TR 38.816 V15.0.0*; Study on CU-DU lower layer split for NR (December 2017)

[9] *3GPP TS 38.401 V16.0.0*; NG-RAN; Architecture description (December 2019)

[10] *3GPP TS 24.501 V16.3.0*; Non-Access-Stratum (NAS) protocol for 5G System (5GS);(December 2019)

[11] *3GPP TS 38.300 version 16.0.0*; NR; NR and NG-RAN Overall Description;Stage 2 (December 2019)

[12] *3GPP TS 37.324 version 15.1.0*; Service Data Adaptation Protocol (SDAP) specification (September 2018)

[13] *3GPP TS 38.323 version 15.6.0*; NR; Packet Data Convergence Protocol (PDCP) specification (June 2019)

[14] *3GPP TS 38.322 version 15.5.0*; NR; Radio Link Control (RLC) protocol specification (April 2019)

[15] *3GPP TS 38.321 version 15.8.0*; NR; Medium Access Control (MAC) protocol specification (December 2019)

[16] *3GPP TS 38.214 version 16.0.0*; NR; Physical layer procedures for data (December 2019)

[17] *3GPP TS 38.211 version 16.0.0*; NR; Physical channels and modulation. (December 2019)

[18] *3GPP TS 38.212 version 16.0.0*; NR; Multiplexing and channel coding. (December 2019)

[19] *3GPP TS 38.104 version 15.5.0*; NR; Base Station (BS) radio transmission and reception (May 2019)

[20] *3GPP TS 28.531 version 16.4.0*; Management and orchestration; Provisioning; Release 16 (December 2019)

[21] *3GPP TSG-RAN meeting n.86, RP-193169*; Study on enhancement of RAN Slicing; Release 17 (December 2019)

[22] *ETSI GS NFV-MAN 001 V1.1.1*; Network Functions Virtualisation (NFV); Management and Orchestration (December 2014)

[23] *ETSI TS 128 533 V15.1.0*; 5G, Management and orchestration; Architecture framework (April 2019)

[24] *NGMN Alliance*; Overview on 5G RAN Functional Decomposition. (February 2018)

[25] *NGMN Alliance*; 5G End-to-End Architecture Framework v2.0. (February 2018)

[26] *O-RAN-WG1.OAM Architecture -v01.00*; O-RAN Operations and Maintenance Architecture (July 2019)

[27] *O-RAN-WG6.CAD-V01.00.00*; Cloud Architecture and Deployment Scenarios for O-RAN Virtualized RAN (October 2019)

[28] *Sridhar Bhaskaran, Huawei Technologies*; 3GPP 5G Control Plane Service-Based Architecture.

[29] *Gabriel Brown, Huawei Technologies*; Service-Oriented 5G Core Networks (Heavy Reading)

[30] *Huawei Technologies CO., LTD.*; 5G Network Architecture; A High-Level Perspective.

[31] *Cinzia Sartori; Nokia Bell Labs*; Network Slicing in Public and Private 3GPP 5G Networks. (March 2019)

[32] *Sankaran Balasubramaniam; Nokia Bell Labs*; End to End Network Slicing in 5G System, 3GPP Standards Perspective.

[33] *Wanqing Guan, Luhan Wang, Zhaoming Lu, Beijing Laboratory of Advanced Information Networks*; A Service-oriented Deployment Policy of End-to-End Network Slicing Based on Complex Network Theory. IEEE Access, April 2018

[34] *Menglan Jiang, Massimo Condoluci, Toktam Mahmoodi; King's College London, London, UK*; A Network slicing management & prioritization in 5G mobile systems. *European Wireless, pp. 197-202* (2016)

[35] *K.Koutlia, R.Ferrús, E.Coronado, R.Riggio, F.Casadevall, A.Umbert, J.Pérez-Romero*; Design and Experimental Validation of a Software-Defined Radio Access Network Testbed with Slicing Support. *Hindawi Wireless Communications and Mobile Computing* (2019)

[36] *EURECOM Sophia Antipolis*; Providing low latency guarantees for slicing-ready 5G systems via two-level MAC scheduling (April 2018)

[37] *University of Versailles - PRISM Laboratory, Vedecom Institute*; A Cross-Layer QoS Solution for Resource Optimization in LTE Networks (June 2016)

[38] *PriMO-5G Deliverable 1.1*; PRIMO-5G USE CASE SCENARIOS (February 2019)

Acknowledgements

On this moment of submission of my thesis, I would like to thank all those good people, whom I have come across in this path of my journey, whose lives have inspired me and from whom I've learnt to live the life.

I am grateful to Professor Roberto Garello, for making the connection with National Instruments GmbH Dresden. I thank him sincerely, for his consideration and support towards this opportunity.

I express my sincere thanks to my Academic Advisor, Dr. Carla F. Chiasserini, for the valuable advice, excellent guidance and encouragement. Her insight into the subject has always made me realize and understand the subject in a broader perspective. I extend my special thanks to her for accepting me as her student and supporting me through this journey.

I express my gratitude to my advisor at NID, Dr. Achim Nahler, for his help and continuous support. His immense patience, encouragement, discussions and positive attitude have always kept me motivated throughout the course of the study as well as life.

I am thankful to Dr. Michael Löhning for his valuable suggestions and continuous support. I also acknowledge the invaluable help provided by Dr. Walter Nitzold in completing my thesis. Many thanks to Martin Anderseck his initial guidance in configuring the emulation environment and through the debugging sessions. I wish to thank all the colleagues of NID for their the hospitality and support they have given.

Many thanks to all my friends who supported me during this time, Sara, Joana, Inva, Mustafa. I wish to remember the help provided by them by encouraging me to stay positive despite my tough times. I wish to thank also Ervin for his encouragement, understanding and the positive attitude. A special thanks to all other friends and people I met on this long journey. I cannot forget the contribution of my sister, Pamela who has been a constant reservoir of cheers at all those times when I needed it the most.

Finally, my deep and sincere gratitude to my family for their continuous and unparalleled love, help and support. My parents and my Uncles are the ones who have made this thesis a reality by having faith in me and for for encouraging me to strive and achieve higher goals in life.