Implementation of 5G Private Networks in Railway Communications

Pedro Martins de Melo Rodrigues Mateus

Thesis to obtain the Master of Science Degree in

Telecommunications and Informatics Engineering

Supervisor: Prof. Luís Manuel de Jesus Sousa Correia

Examination Committee

Chairperson: Prof. Fernando Manuel Valente Ramos

Supervisor: Prof. Luís Manuel de Jesus Sousa Correia

Members of Committee: Prof. António José Castelo Branco Rodrigues

: Eng. Fernando Manuel Lopes Santana

May 2024

I declare that this document is an original work of my own authorship and that it fulfils
all the requirements of the Code of Conduct and Good Practices of the
Universidade de Lisboa.

Acknowledgements

This master thesis is an important milestone in my life, and it would not have been possible without several people that, in a way or another, contributed to the making of this work.

First, I shall express my deep appreciation to professor Luís Correia who supervised this thesis. Not only he helped me with his advice, knowledge, and explanations, he also served as an example of someone who is dedicated, disciplined and thoughtful. Most importantly, he contributed with his time, the most important asset.

I would like to thank Eng. Fernando Santana, Eng. Tiago Costa and Eng. Nuno Frigolet from Thales for their valuable partnership and the support. The meetings and discussions were fruitful and provided key insights for the development of this thesis.

To all GROW members and the dear “family of telecommunications” for the time together, conversations, love, and fun moments. For being good listeners when things were not great.

To my parents. Not only for bringing me to life but giving me everything to have a great one. To my brother and sister for being there and giving unconditional support.

To my good friends that shared the arduous and demanding path with me since the beginning, side by side, in this great school, Instituto Superior Técnico.

And finally, to the good friends from the earliest stages of life, the innocent times, the childhood, and adolescence. I met these at my home village, Ponte de Lima, the place I will always return to. A special appreciation to my friends from the centre, from Correlhã and from Freixo! Thank you for your friendship.

To all the people that are not included in these groups but are certainly included in my memories for their contribution to my life experiences which I will take with me for the future.

Thank you!

Abstract

The purpose of this thesis is to study and identify the network architectures and private network configurations that best meet the requirements of railway applications, such as latency and capacity. Some of the outputs of the developed model are the provided throughput, the required throughput by all users in the system, as well as delays across the network that compose the E2E Total Latency. The possibility of using a MEC node is also considered for latency reduction as well as different radio configurations. Moreover, private network deployment options considered were isolated, shared and slice. The studied scenario is the Subway of Lisbon, taking in account metrics like the number of passengers, distances between stations, metro lines and number of carriages. The obtained results show that the shared option between the railway and the commercial operators is the only solution that satisfies the requirements of all studied services, of which some are railway specific, and others are associated with passengers. With this option, there is RAN sharing with the railway critical applications hosted in the railway core and the passenger related applications are hosted in the CSP's Core. For passenger services, one obtained a required capacity of 1.355 Gbps with a margin of 134 Mbps to the provided, and a required capacity of 30 Mbps with a margin of 16 Mbps to the provided for railway services. Regarding latency, one obtained a total node latency in the 2 ms to 20 ms range for the studied services.

Keywords

5G, Private Networks, Railway Communications, MEC, Subway

Resumo

O objetivo desta tese é estudar e identificar as arquiteturas de redes e configurações de redes privadas que melhor respondem aos requisitos das aplicações ferroviárias, como latência e capacidade. Alguns dos resultados do modelo desenvolvido incluem o ritmo de transmissão fornecido, a capacidade exigida por todos os utilizadores do sistema, assim como os atrasos na rede que compõem a Latência Total. A possibilidade de utilização de um nó MEC para redução de latência é considerada, bem como diferentes configurações rádio. As opções de implementação de redes privadas consideradas foram isolada, partilhada e em fatia. O cenário é o Metro de Lisboa e são tidas em conta algumas métricas como o número de passageiros, distâncias entre estações, linhas de metro e número de carruagens. Os resultados obtidos mostram que a opção partilhada entre os operadores ferroviário e comercial é a única solução que satisfaz as exigências de todos os serviços estudados, alguns dos quais são específicos do sistema ferroviário e outros associados aos passageiros. Com esta opção, há partilha de RAN com as aplicações ferroviárias críticas hospedadas no núcleo do operador ferroviário enquanto as aplicações relacionadas com os passageiros são hospedadas no núcleo do operador comercial. Para os serviços de passageiros obteve-se uma capacidade requerida de 1,355 Gbps com uma margem de 134 Mbps em relação ao prestado, e uma capacidade requerida de 30 Mbps com uma margem de 16 Mbps ao prestado para os serviços ferroviários. Relativamente à latência, obteve-se uma latência total de nó no intervalo de 2 a 20 ms para os serviços estudados.

Palavras-chave

5G, Redes Privadas, Comunicações Ferroviárias, MEC, Metropolitano

Table of Contents

List of Figures

List of Tables

List of Acronyms

3GPP	3rd Generation Partnership Project
4G	4th Generation of Mobile Communication Systems
5G	5th Generation of Mobile Communication Systems
5GC	5G Core
AF	Application Function
AI	Artificial Intelligence
AL	Air Link
AMF	Access and Mobility Management Function
API	Application Programming Interface
AR	Augmented Reality
ATO	Automatic Train Operation
ATP	Automatic Train Protection
AUSF	Authentication Server Function
BBRS	Broad Band Radio System
BS	Base Station
CBRS	Citizens Broadband Radio Service
CCTV	Closed-Circuit Television
CN	Core Network
COTS	Commercial off-the-shelf
CP	Cyclic Prefix
CQI	Channel Quality Indicator
C-RAN	Cloud – Radio Access Network
CSC	Communication Service Customer
CSP	Communication Service Provider
CU	Central Unit
DL	Downlink
DN	Data Network

DU	Distributed Unit
E2E	End-to-End
eCPRI	enhanced Common Public Radio Interface
EDC	External Data Centre
EMBB	Enhanced Mobile Broadband
EPC	Evolved Packet Core
ERTMS	European Rail Traffic Management System
ETCS	European Train Control System
ETSI	European Telecommunications Standards Institute
EU	European Union
FDD	Frequency Division Duplexing
FR1	Frequency Range 1
FR2	Frequency Range 2
FRMCS	Future Railway Mobile Communication System
gNB	generation Node B
GSA	Global Mobile Suppliers Association
GSM	Global System for Mobile Communications
GSM-R	Global System for Mobile Communications-Railway
IMT	International Mobile Telecommunications
IoT	Internet of Things
ISD	Inter Site Distance
ITU	International Telecommunication Union
ITU-R	International Telecommunication Union - Radio
ITU-T	International Telecommunication Union - Telecommunications
KPI	Key Performance Indicator
LTE	Long Term Evolution
LTE-R	Long Term Evolution-Railway
MAC	Medium Access Control
MCPTT	Mission-critical push-to-talk
MEC	Multi-Access Edge Computing

MIMO	Multiple Input Multiple Output
mMTC	Massive Machine Type Communications
NEF	Network Exposure Function
NF	Network Function
NFV	Network Function Virtualisation
NFVI	Network Function Virtualisation Infrastructure
Ng-eNB	next generation – evolved Node B
NG-RAN	Next Generation – Radio Access Network
NOP	Network Operator
NR	New Radio
NRF	Network Repository Function
NSA	Non-Standalone
NSaaS	Network Slice as a Service
NSC	Network Slice Customer
NSI	Network Slice Instance
NSP	Network Service Provider
NSSAAF	Network slice-specific authentication and authorisation
NSSF	Network Slice Selection Function
O&M	Operations and Management
OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access
OSS	Operation Support System
PAPR	Peak-to-Average Power Ratio
PCF	Policy Control Function
PCFR	Policy and Charging Rules Function
PDCP	Packet Data Convergence Protocol
PDU	Packet Data Unit
PIS	Passenger Information System
PLMN	Public Land Mobile Network
PNF	Physical Node Function

QoS	Quality of Service
RAN	Radio Access Network
RAT	Radio Access Technology
RB	Resource Block
RLC	Radio Link Control
RRC	Radio Resource Control
RRM	Radio Resource Management
RRU	Remote Radio Unit
RU	Radio Unit
SA	Standalone
SBA	Service Based Architecture
SC-FDMA	Single-Carrier Frequency Division Multiple Access
SCP	Service communication proxy
SDN	Software Defined Network
SF	Scaling Factor
SISO	Single Input Single Output
SLA	Service Level Agreement
SLS	Service Level Specification
SMF	Session Management Function
SMS	Short Message Service
TDD	Time Division Duplexing
TETRA	Terrestrial Trunked Radio
TL	Transport Link
TN	Transport Network
TTI	Transmission Time Interval
UDM	Unified Data Management
UDR	Unified Data Repository
UE	User Equipment
UIC	International Union of Railways
UL	Uplink

UNIFE	European Rail Supply Industry Association
UPF	User Plane Function
URLLC	Ultra-Reliable Low Latency Communication
VNF	Virtualisation Network Function
VPN	Virtual Private Network
VR	Virtual Reality

List of Symbols

δ_{AL_Rx}	Air Link Delay in the receiver side
δ_{AL_Tx}	Air Link Delay in the transmitter side
δ_{CORE_Proc}	Core Processing Delay
δ_{CORE_Rx}	Latency of the core of the network in the receiver side
δ_{CORE_Trans}	Core Transmission Delay
δ_{CORE_Tx}	Latency of the core of the network in the transmitter side
$\delta_{Critical\_threshold}$	Latency Critical Threshold
δ_{CU_Proc}	CU Processing Delay
δ_{CU_Queu}	CU Queuing Delay
δ_{CU_Rx}	Latency of the CU node in the receiver side
δ_{CU_Trans}	CU Transmission Delay
δ_{CU_Tx}	Latency of the CU node in the transmitter side
δ_{DU_Proc}	DU Processing Delay
δ_{DU_Queu}	DU Queuing Delay
δ_{DU_Rx}	Latency of the DU node in the receiver side
δ_{DU_Trans}	DU Transmission Delay
δ_{DU_Tx}	Latency of the DU node in the transmitter side
δ_{E2E}	E2E Latency
δ_{EDC}	Latency of the External Data Centre
δ_{EDC_Proc}	EDC Processing Delay
δ_{EDC_Trans}	EDC Transmission Delay
δ_{MEC}	Latency of the MEC node
δ_{MEC_Proc}	MEC Processing Delay
δ_{MEC_Trans}	MEC Transmission Delay
$\delta_{Max\_Total\_Propagation}$	Maximum propagation delay to satisfy the service requirements
δ_{Prop}	Propagation Delay

δ_{Queu}	Queuing Delay
δ_{RU_Proc}	RU Processing Delay
δ_{RU_Queu}	RU Queuing Delay
δ_{RU_Rx}	Latency of the RU node in the receiver side
δ_{RU_Trans}	RU Transmission Delay
δ_{RU_Tx}	Latency of the RU node in the transmitter side
$\delta_{Service}$	Maximum latency requirement for the service
δ_{Tot_Node}	Total node latency which is the sum of all delay contributions expect propagation
δ_{Trans}	Transmission Delay
δ_{UE_Proc}	UE Processing Delay
δ_{UE_Rx}	Latency of the UE in the receiver side
δ_{UE_Trans}	UE Transmission Delay
δ_{UE_Tx}	Latency of the UE in the transmitter side
$\mathit{\Delta f}$	Subcarrier Spacing
$\Delta f_{NR\mu}$	Subcarrier Bandwidth
Δf_{ref}	Default subcarrier bandwidth (15 kHz)
η_r	Ratio of Capacity
ρ_{lat}	Latency adaptation parameter
ρ_{func}	Ratio of functionalities assigned to the MEC node
ρ_{DU}	Ratio of functionalities assigned to the DU
ρ_{CU}	Ratio of functionalities assigned to the CU
ρ_{RU}	Ratio of functionalities assigned to the RU
ρ_{UE}	UE Processing delay ratio
μ	Numerology
μ_s	Subcarrier Load
A	Number of Antenna Ports
A_f	Factor that accounts for the losses in the system
B	System Bandwidth

B_c	Control and Signalling Bandwidth
$BW(j)$	Bandwidth for the j-th carrier
D	Packet size in bytes
$D_{serv,p}$	Packet size in bytes from a specific service with specific priority p
F_{DL}	Fraction of the slot that belongs to DL
F_{DL}	Fraction of the slot that belongs to UL
J	Number of aggregated component carriers
M	Margin of Latency, value between 0 and 1
M_c	Modulation Order for Control Signals
M_{info}	Information in the MAC header
M_{mod}	Order of Modulation
M_{users}	Number of users connected to the simulated node
$N_{PRB}^{BW(j),\mu}$	Maximum Resource Block allocation in bandwidth $BW(j)$ with numerology μ
N_{SC}	Number of Subcarriers
N_{layers}	Number of Layers in the system
N_{layers},c	Layers for control and signalling
N_{symb}	Number of Symbols
Nu	Number of users connected to the RU
$Nu\_passengers$	Number of terminals using passenger services connected to the RU
Nu_rail	Number of terminals using railway services connected to the RU
$O^{(j)}$	Overhead for the control channels for the j-th carrier
PRB	Physical Resource Blocks Available
$Q_m^{(j)}$	Maximum modulation order for the j-th carrier
R	Maximum throughput offered by the RU
R_{FH_6}	FH throughput for Splitting option 6
$R_{FH\_7.1}$	FH throughput for Splitting option 7.1
$R_{FH\_7.2}$	FH throughput for Splitting option 7.2
$R_{FH\_7.3}$	FH throughput for Splitting option 7.3

R_{FH_8}	FH throughput for Splitting option 8
R_{MH_2}	MH throughput for Splitting option 2
$R_{TDD/DL}$	Maximum throughput offered by the RU in the DL with TDD mode
$R_{TDD/UL}$	Maximum throughput offered by the RU in the UL with TDD mode
R_c	Control and Signalling Rate
$R_{codemax}$	Code rate value
R_{max}	Maximum throughput of the link
R_p	System Peak Rate
$R_{required}$	Throughput required by the users in the system
$R_{required\_passengers}$	Throughput required by the terminals in the system using passenger services
$R_{required\_railway}$	Throughput required by the terminals in the system using railway services
R_s	Data Rate of the service
$R_{s\_passengers}$	Data Rate of the simulated passenger service
R_{s_rail}	Data Rate of the simulated railway service
S	Sampling Rate in samples per second
T_s^{μ}	Average OFDM symbol duration for a numerology μ
d	Distance between source and destination of the signal
d_{BH_Rx}	Backhaul Link distance in the receiver side
d_{BH_Tx}	Backhaul Link distance in the transmitter side
d_{E2E_max}	Maximum E2E Distance to satisfy the service requirements
d_{FH_Rx}	Fronthaul Link distance in the receiver side
d_{FH_Tx}	Fronthaul Link distance in the transmitter side
d_{MH_Rx}	Midhaul Link distance in the receiver side
d_{MH_Tx}	Midhaul Link distance in the transmitter side
d_{TL_Rx}	Transport Link distance in the receiver side
d_{TL_Tx}	Transport Link distance in the transmitter side
d_{max}	Maximum Distance between UE and EDC/MEC to satisfy the service

	requirements
$d_{ue-CSPCore}$	Maximum distance between a UE in the subway system and the core of the CSP operator
$d_{ue-RailCore}$	Maximum distance between a UE in the subway system and the core of the railway operator
$f^{(j)}$	Scaling factor for the j-th carrier
j	j-th carrier
p	Service's Priority
v	Velocity of the signal in the link
$v_{layers}^{(j)}$	Number of layers of the transmitter for the j-th carrier

Chapter 1

Introduction

This chapter gives a description of the main subjects of the thesis. It starts with the mobile communication systems evolution towards 5G, Network Slicing and Private Networks and finally Railway Communications. It ends with the objectives and contents of this work.

1.1 Overview and Motivation

Over the years, from 1970 until now, several mobile communication systems were developed and commercialised around the world, starting from 1G to the current 5G, as a result of the transverse technological evolution and of the constant change of requirements in terms of user and applications.

These applications have become more demanding with respect to the bandwidth they require, such as 4K videos with ultra-high definition, the strict latency that is a key requisite in some cases (e.g., gaming, remote surgery and critical systems) and the massive increase in connected devices, which brings another challenge, with the Internet of Things (IoT) and everything that is connected, including our home appliances, traffic lights, sensors in the water/electrical grids, garbage bins, drones and many more.

5G is the most recent technology that helps solving many of these issues with advanced functionality and innovative techniques. 5G is already commercially available to the common user in several countries, but this technology has the highest potential with enterprises. Enterprises want to reduce costs, make processes more efficient and increase profits, and digitalisation is the pivotal factor. 5G will revolutionise and help making this happen along with other technologies, like Artificial Intelligence (AI), Cloudification, Robotics and Virtual Reality.

There are some enablers for 5G as well as new concepts and new requirements, as illustrated in Figure 1.1.

Figure 1.1 – 5G enablers, new concepts, and requirements (extracted from [BGCB17]).

As one can see in Figure 1.1, there are a couple of enablers to 5G, such as the advances in radio interface with the Multiple Input Multiple Output (MIMO) capabilities, the introduction of a new paradigm of Software Defined Networks (SDN) and Network Function Virtualisation (NFV), which is about virtualising resources, avoiding specific hardware, and making use of central controllers that automatically manage the whole network. The Cloud revolutionised the way companies host their applications and networks are built. Network components, like the Radio Access Network (RAN) and the Core, are hosted in the Cloud, which optimises resource management.

There are also new concepts that are fundamental in 5G and that will have the highest impact on users, such as the new architecture that is service-based and cloud-native, once more, making the system more efficient, and one that is closely related to this thesis, which is Network Slicing and Private

Networks. This capability will allow enterprises to have a dedicated and personalised network solution according to a Service Level Agreement (SLA) in a quite fast, secure, efficient, and flexible manner. A slice is a segment of the network with resources allocated to an organisation that is isolated from the rest of the network. To have a complete end-to-end network slice, one must have slicing capabilities in all components of the system, which include RAN, transport, and core.

There are a lot of sectors that are investing in Private Networks already and there is a huge potential for new use-cases to take advantage of this technology. Figure 1.2 depicts the different sectors deploying private mobile networks in trials and commercially as of 2022, according to Global Mobile Suppliers Association (GSA).

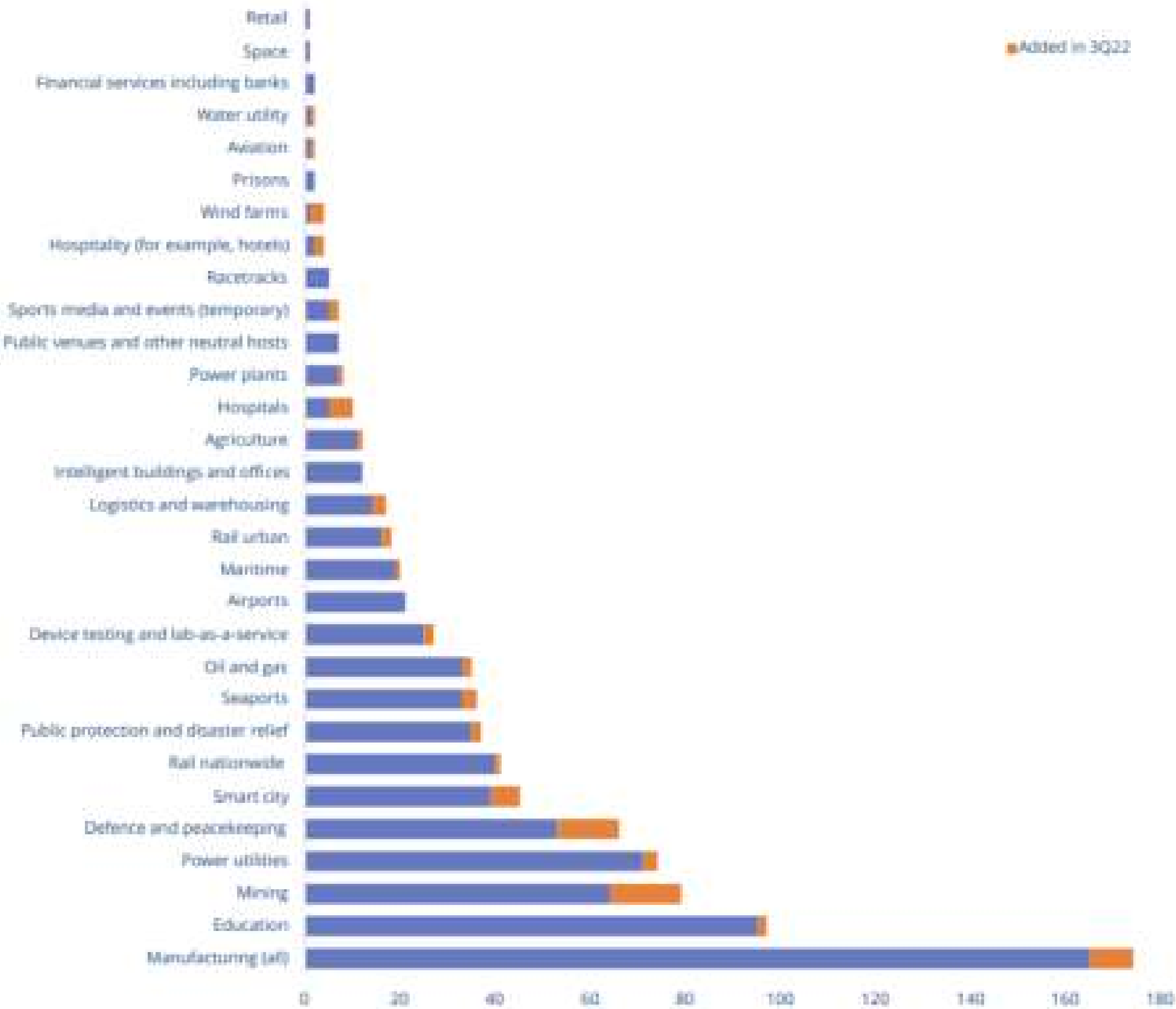


Figure 1.2 – Different sectors deploying private networks as of 2022 (extracted from [GSAP22]).

One can see in Figure 1.2 that the sector that has deployed the most private networks is Manufacturing with more than 150 deployed networks. There are many more sectors that are deploying private networks, like Education, Mining, Power utilities, Defence and Peacekeeping, Smart Cities, Ports and Rails as well, which is the focus of this thesis.

The deployment of private networks has different possible implementation scenarios with different spectrum options for enterprise. Enterprises that have a strict latency requirement normally have the core installed on premises with a Multi-access Edge Computing (MEC) node that runs applications and processing tasks, therefore reducing network congestion, latency and improving performance. Railway systems are a typical case of implementation of private networks, given the need for an isolated communication system for the several critical applications that it requires. In fact, one can see in Figure 1.2 that there are more than 50 private railway networks deployed according to GSA, by merging both the Rail Nationwide and Rail Urban sectors [GSAP22].

Railway Communications are key to the functioning of the whole rail system and that is why there is the necessity of having a robust, efficient, and secure communication system that serves all rail applications. Initially, all countries developed their own railway communication systems, however, this brought border issues and competition problems, and so there was a need to come up with a homogenous system across Europe to promote interoperability, competition and optimise the investments [GSMR23].

The main implemented standard system was GSM-R, which is based on GSM and was developed by the International Union of Railways (UIC), rail stakeholders and ETSI. GSM-R allows for voice and data communication between rail staff, as well as train signalling together with European Train Control System (ETCS). Figure 1.3 illustrates a simplified GSM-R scheme.

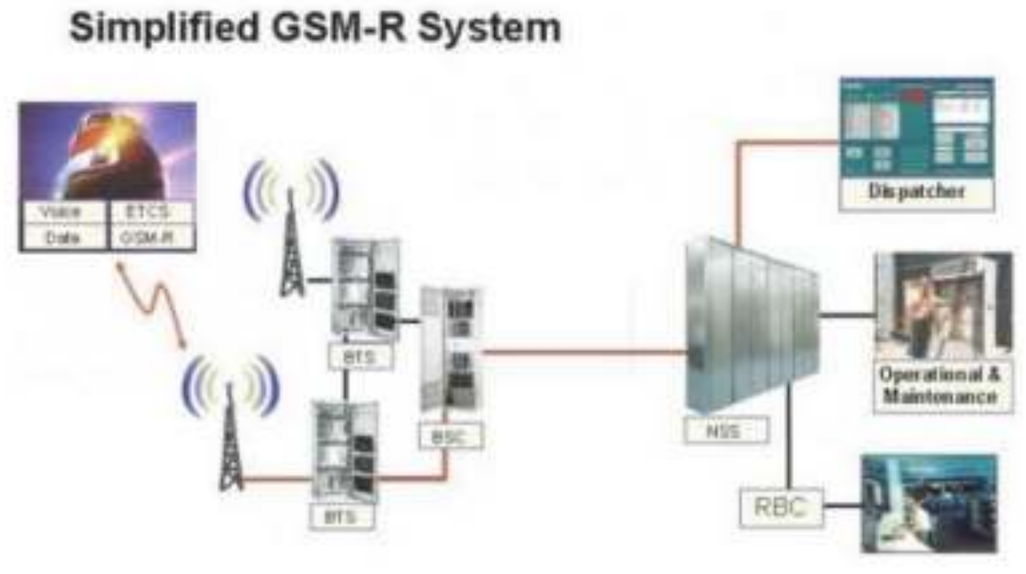


Figure 1.3 – Simplified GSM-R system (extracted from [GSMR23]).

GSM-R is an old system, based on the second generation of mobile communication systems. GSM-R has some limitations, and it is not up to the future railway needs with trains becoming more and more faster, and applications becoming more demanding. Moreover, there is the expectation of being obsolete by 2030 with the end of support from suppliers.

For that reason, efforts are being made to develop a new system that replaces GSM-R, by UIC and other organisations, pertinently called Future Rail Mobile Communications System (FRMCS), which is based on 5G and will support new functionalities, applications, services, and requirements.

1.2 Objectives and Content

The objective of this thesis is to study the implementation of 5G Private Networks in railways for the different applications and scenarios that are involved, in collaboration with Thales.

This thesis is composed of five chapters. The first chapter introduces this work, the motivation and content of the work. The second chapter provides the fundamental aspects of the main topics of this thesis to better comprehend the following chapters, such as 5G foundations, it's network architecture and radio interface, C-RAN architecture and the splitting options, the concept of network slicing and needs for private networks, an overview of the services and applications, existing technologies in railway communications and finally state of the art.

The third chapter presents the model to be implemented, the mathematical expressions that compose the model, the inputs, and outputs as well as intermediary processes, the service requirements, latency and capacity models and their flowcharts and in the end the model assessment.

Then, in the fourth chapter, one describes the results of the work, namely, the capacity and latency results for different private network configurations. Moreover, the scenario analysed, Metro of Lisbon, is characterised with the main features and qualities. The accepted private network options are described in the end of this chapter.

The last chapter, the fifth chapter, gives the conclusions for this work as well as future work to be considered.

Chapter 2

Fundamental Concepts

This chapter gives a detailed description of all fundamental concepts for the development of this thesis and theoretical foundations to every topic related to the thesis including 5G, Network Slicing and Private Networks, Services and Applications, Railway Communications and State of the Art.

2.1 5G aspects

2.1.1 Network Architecture

The deployment of 5G NR networks is being done in 2 different phases: the **Non-Standalone** and the **Standalone** modes, as illustrated in Figure 2.1.

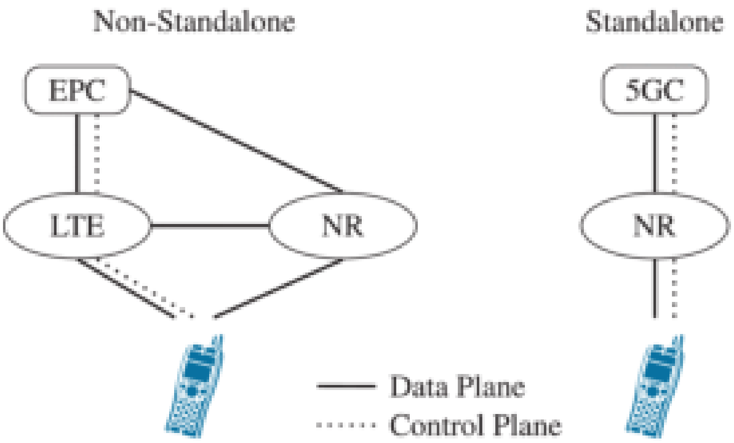


Figure 2.1 - Non-Standalone (NSA) and Standalone (SA) 5G network (extracted from [RSIF21]).

As illustrated in Figure 2.1, there are 2 main deployment phases:

- **Non-Standalone** (NSA) - this is a preliminary version of 5G deployment, where the legacy 4G LTE infrastructure is needed to allow operation. As seen on the left-hand side of Figure 2.1, the network connects the User Equipment (UE) via a control-plane running over the 4G LTE network, often referred to as the 4G anchor. The data-plane is provided by the 5G NR network and 4G as well, to allow dual connectivity. Both control and data planes are forwarded to Evolved Packet Core (EPC).
- **Standalone** (SA) - this is the mode of deployment that makes the 5G network independent of the 4G one. As seen on the right-hand side of Figure 2.1, both control and data planes are forwarded to 5G Core (5GC).

The 5G Core Network Architecture standards and specifications were developed by 3rd Generation Partnership Project (3GPP), the organisation that develops international standards for mobile communications. The 5GC follows a **Service-Based System Architecture**, which is composed of **Network Functions** (NFs) and interfaces between them. According to [3GPP19], an NF is a “defined processing function in a network, which has defined functional behaviour and 3GPP defined interfaces”. It can be implemented either as a network element on a dedicated hardware or a virtualised instance in the cloud.

Figure 2.2 shows the non-roaming 5G service-based architecture (SBA) for the core network. One can see the NFs (shown as green rounded boxes) and their service interfaces (shown as round white circles) along with their relations. According to the Release 15 of 3GPP [3GPP19] and [WiSt21], the several NFs and their functions are:

- Authentication server function (AUSF): Performs authentication between UE and the network.

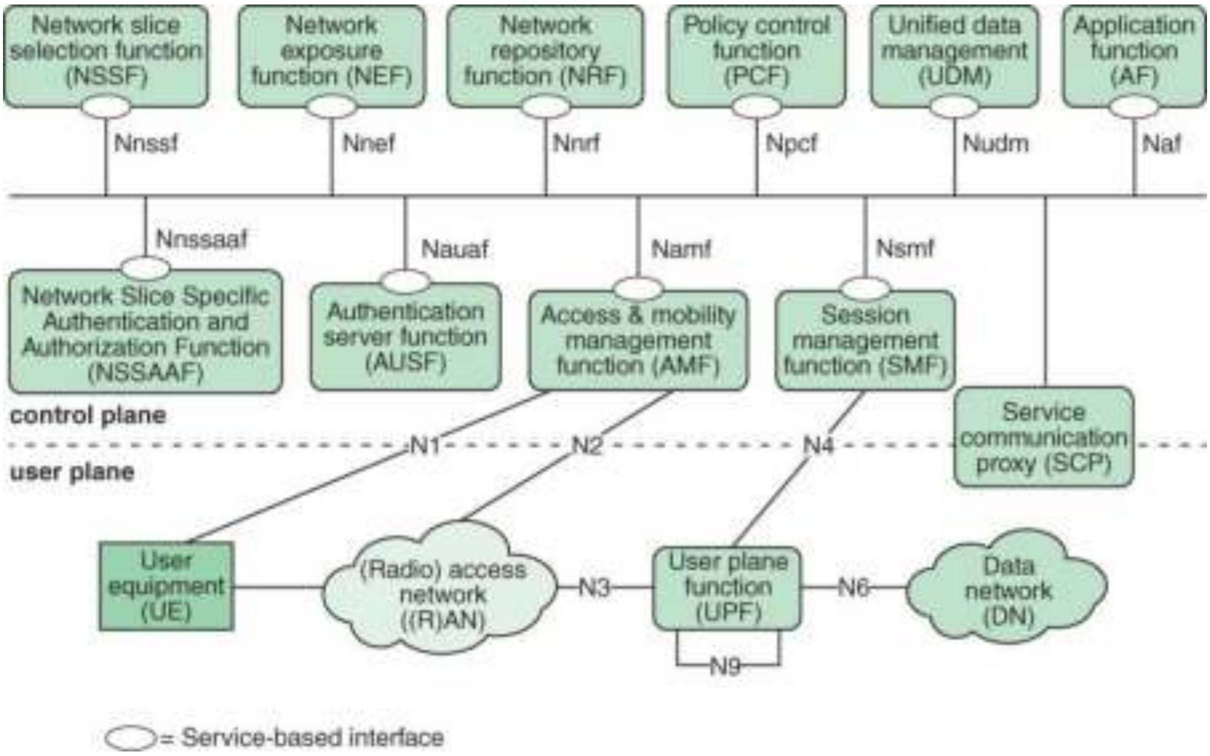


Figure 2.2 - Non-Roaming 5G System Architecture (extracted from [3GPP19]).

- Access and mobility management function (AMF): Responsible for Registration management, Connection management, Reachability management, Mobility management and Access Authentication and Authorisation.
- Network exposure function (NEF): Provides an interface for outside applications to communicate with the 5G network to obtain network-related information.
- Network repository function (NRF): A database that stores information about each NF and allows them to discover services from other NFs.
- Network slice selection function (NSSF): Selects the set of Network Slice instances serving the UE. Determines the Slice ID and the AMF set to be used to serve the UE.
- Network slice-specific authentication and authorisation (NSSAAF): Performs authentication and authorisation specific to a slice.
- Policy control function (PCF): Supports unified policy framework to govern network behaviour and provides policy rules to Control Plane functions to enforce them.
- Session management function (SMF): Deals with Session Management, UE IP Address allocation, management, and Roaming functionality.
- Unified data management (UDM): Manages the authorisation and subscription management of the UEs. Stores subscription data from the same UE.
- User plane function (UPF): Handles the user plane path of Packet Data Unit (PDU) sessions including packet routing and forwarding, packet inspection and Quality of Service (QoS).
- Application function (AF): interacts with the 3GPP Core Network to provide services, for example to access NEF.

- User equipment (UE): Allows a user access to network services. An example is a mobile phone. For 3GPP specifications, the interface between the UE and the network is the radio interface.
- Radio Access Network (RAN): Provides access to a 5G core network. This includes the 5G RAN and other wireless and wired access networks.
- Data network (DN): Allows UE to be logically connected by a session. It may be the Internet, a corporate intranet, or an internal services function within the mobile network operator's core (including content distribution networks).
- Service communication proxy (SCP): Allows NFs and NFSs to communicate directly or indirectly. The SCP enables multiple NFs to communicate with each other and with user plane entities in a highly distributed multi-access edge compute cloud environment. This provides routing control, resiliency, and observability to the core network.

There is also the new RAN architecture in 5G. Figure 2.3 presents the overall RAN architecture.

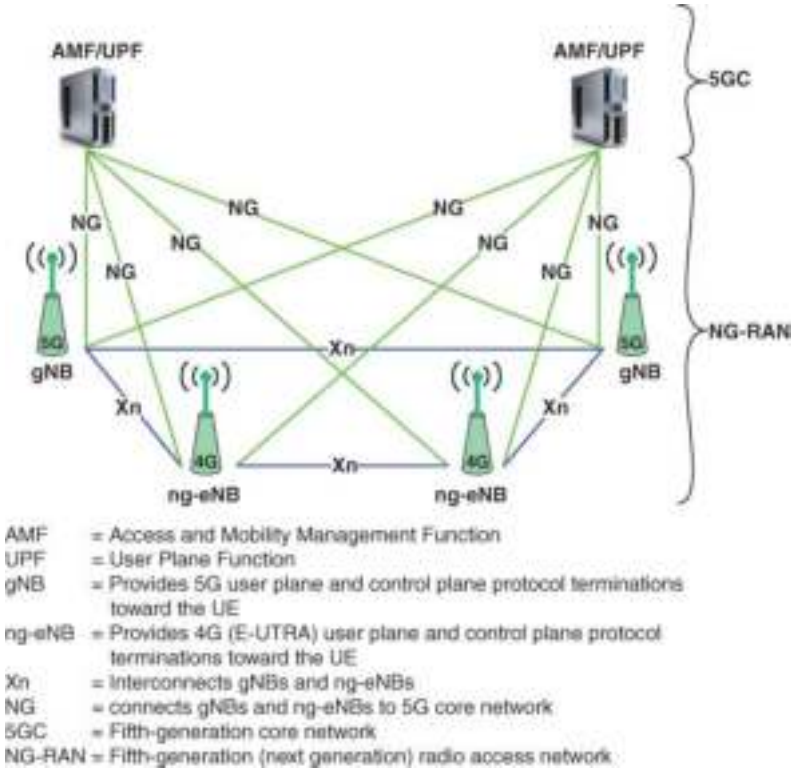


Figure 2.3 – RAN System Architecture (extracted from [WiSt21]).

By analysing Figure 2.3, the overall system is divided into the Next Generation – Radio Access Network (NG-RAN) and the 5GC. The NG-RAN is composed of two types of Base Stations (BSs):

- next generation – evolved Node B (ng-eNB): Provides 4G user and control plane to the UE and connects to the 5GC via the NG interface.
- generation Node B (gNB): Provides 5G user and control plane to the UE and connects to the 5GC via the NG interface as well.

It can also be seen that the Xn interface connects both gNBs and ng-eNBs with each other. The NG interface connects both gNBs and ng-eNBs to the 5GC, specifically, to the AMF node with the interface NG-C and to the UPF with the interface NG-U.

2.1.2 Radio Interface

This section focuses on NR radio interface aspects, such as: frequency bands used in NR, duplexing modes, multiple access technologies, modulation schemes and radio frame structure.

According to 3GPP Release 15 [3GPP18], there are two range of frequencies defined for NR, namely:

- **FR1:** from 0.45 GHz to 6.0 GHz - known as the sub-6 GHz.
- **FR2:** from 24.25 GHz to 52.6 GHz - known as millimetre waves.

Each band can have one of two duplexing modes: **Time Division Duplexing** (TDD) in which both Uplink (UL) and Downlink (DL) use the same frequency band but in different time slots, and **Frequency Division Duplexing** (FDD) where both UL and DL are transmitted at the same time but in different frequency ranges within the band. In FDD, one needs guard bands between UL and DL due to possible interference, whereas in TDD one needs precise synchronisation at both the receiver and transmitter. While FDD is widely used, it requires more spectrum than TDD, contains spectrum only for interference mitigation with guard bands and has a higher hardware cost. TDD, on the other hand, is becoming a favourable option for 5G, due to typical asymmetric data rates that are most common in the current days and spectrum efficiency enabling dynamic bandwidth allocation.

There are several bands across the different frequency ranges with different bandwidths. There is the **low band** (below 2 GHz) which main goal is to provide better wide outdoors and indoor coverage. Then, the **mid-band** (between 2 GHz to 8 GHz) that is considered the ideal band for 5G, providing both coverage and capacity. Finally, one has the **high-band** (above 24 GHz) that allows for very high-capacity connections and specific applications. However, this band is limited, since the signals at these frequencies have a very low range and poor propagation characteristics. This band has not seen wide commercial deployment yet, unlike the others as of today.

In Portugal, the most relevant bands are the n28 of **0.7 GHz** with a bandwidth per operator of 10 MHz and the band n78 of **3.6 GHz** with 100 MHz bandwidth per operator. Regarding private networks spectrum, there are some options, namely: dedicated licence, shared or unlicenced. In several countries, there is already dedicated spectrum for private networks (e.g., in France, USA, UK, Denmark and others). Regarding NR multiple access, there is a different method for each link:

- DL: **Orthogonal Frequency Division Multiple Access (OFDMA)** - multiple closely spaced orthogonal subcarrier signals, each with 15 kHz (or multiples of it) are transmitted.
- UL: **Single Carrier Orthogonal Frequency Division Multiple Access (SC-OFDMA)** – all subcarriers are jointly transmitted.

Regarding the modulation scheme, NR uses subcarrier adaptive modulation depending on the radio network conditions. It can go up to 256-QAM in both DL and UL. NR has a special feature that differentiates it from the previous generation LTE, which is the subcarrier not having a fixed bandwidth. This feature allows to support different kinds of services and flexibility, which is key. The subcarrier bandwidth is given by:

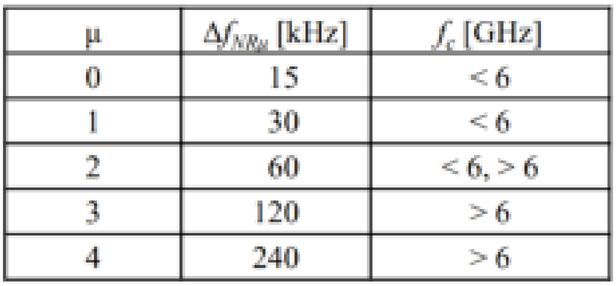
$$\Delta f_{NR\mu[\mathrm{kHz}]} = 2^{\mu} \times \Delta f_{\mathrm{ref}\,[\mathrm{kHz}]} \qquad (2.1)$$

where:

- μ is the numerology.
- Δf_{ref} is the reference frequency of 15 kHz, which is the default subcarrier bandwidth.

Table 2.1 illustrates the possible configurations for the numerology, respective subcarrier's bandwidth, and frequency of operation. For a central frequency over 6 GHz, one uses numerologies of 2 or bigger with a wider subcarrier bandwidth.

Table 2.1 – Numerology configurations for subcarrier bandwidth (extracted from [Corr20]).



μ	$\Delta f_{NR\mu}$ [kHz]	f_c [GHz]
0	15	< 6
1	30	< 6
2	60	$< 6, > 6$
3	120	> 6
4	240	> 6

In NR, each user is given a set of Resource Blocks (RBs) where each RB has 12 subcarriers independently of the numerology. Concerning the NR frame structure, one frame has 10 subframes with 1 ms length each. Each subframe has a number of slots that is 2^{μ} and each slot has 14 OFDM symbols.

2.2 Cloud-RAN Architecture

Cloud Radio Access Networks (C-RAN) is a centralised cloud-based architecture and virtualisation of hardware allowing for the sharing of processing amongst several sites and scalability to add or remove services as required [LaCC19]. 5G RAN is composed of three logical units, each one with different functions: Radio Unit (RU), which consists of the Lower Physical Layer; Distributed Unit (DU), which consists of the Radio Link Control (RLC) and Medium Access Control (MAC) layers; a Centralised Unit (CU), which consists of the upper layers of the 3GPP protocol stack, such as Packet Data Convergence Protocol (PDCP) and Radio Resource Control (RRC). The link between DU and CU is called the Middlehaul, while the link between the RU and DU is the Fronthaul.

Figure 2.4 illustrates the 5G transport network with the different units and network links. One can see the group of Remote Radio Units (RRUs) or RUs connected to the DUs through the fronthaul network with enhanced Common Public Radio Interface (eCPRI), which allows to split the baseband functions and put some of them in the RU. For the fronthaul network, a star or ring topology may be used. Then, one has the link between the DU and the CU through the F1 interface. The midhaul network typically uses a ring topology. The CU can be replaced or complemented with a MEC node. However, the MEC node can be deployed in other configurations, for instance, between the RU and the DU.

Finally, there is the backhaul that bridges the CU to the Core Network through the NG interface, normally by a ring or a mesh topology. For fronthaul and midhaul networks, there is the assumption that a DU only belongs to one CU and that one RU only belongs to a DU, making it a point-to-point service. Normally, several RUs are connected to one DU that aggregates the traffic, and several DUs are connected to one CU that aggregates the traffic as well, which is the approach taken in this thesis.

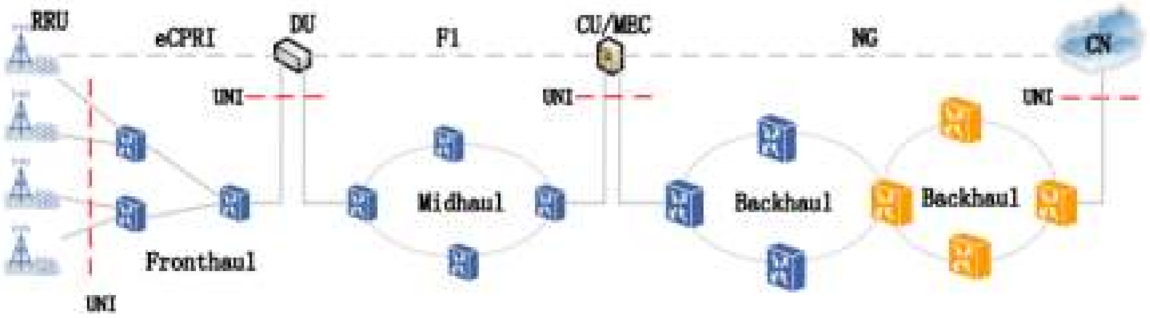


Figure 2.4 – 5G Transport Network (extracted from [ITUT18]).

There are functional splits that determine how the different BS functions are located, whether closer to the user or more centralised. 3GPP has proposed eight functional split options, including several sub-options (7.1,7.2,7.3), ranging from the physical layer and data link layer to the network layer which can be split up into the Distributed Unit (DU), Centralised Unit (CU) and the Radio Unit (RU). Each option offers different trade-offs between centralisation benefits and fronthaul network requirements with impact on capacity and latency parameters. The split of functions allows a flexible design and implementation of the network with different capacity and latency requirements and can bring advantages in terms of energy efficiency.

Figure 2.5 depicts the several split options across all functions of the protocol stack, the red lines showing the different options. The functions to the left of the red arrow are hosted in either the CU or DU, and the functions to the right reside on the RU.

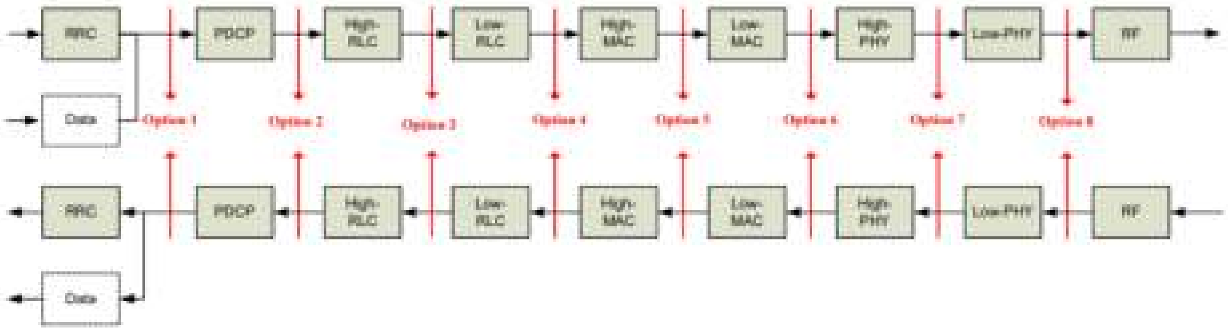


Figure 2.5 – Representation of the split options (extracted from [ITUT18]).

The centralisation of the baseband functionality allows to virtualise many of the network functions, with software hosted on COTS servers. Option 8 fully centralises the functionality of the 5G RAN stack with simple and cheap RUs with low processing functions (only RF), lower power consumption and allowing network upgrades to be made at the CU requiring fewer site visits. Moreover, this simplified RU would be Radio Access Technology (RAT) agnostic, further reducing the masts footprint; however, it demands highly from the fronthaul network with very high capacity and latency requirements. On the other hand, Option 1 places all baseband processing in the RU making it very complex, with high energy consumption. All remaining options ranging from 2 to 7 vary in the level of functions that reside in the RU as opposed to being hosted in the DU/CU [ITUT18]. Option 7.2 is the one that got more interest in industry and the one that is more common in operators' implementation, due to this good balance between RU complexity, fronthaul bandwidth and inter-cell cooperation.

2.3 Network Slicing and Private Networks

Network Slicing is the key factor to provide a flexible network in an efficient way by having different logical networks (slices), which are isolated, over a physical infrastructure. These slices can be given to a set of users with different service requirements. End to End (E2E) Network Slicing involves all parts of the network architecture spanning from the device, access network, core network, transport network and network management system [GSMA21].

Each network slice consists of dedicated or shared Network Functions (NFs) and resources that are virtualised. These resources can be computation, storage, bandwidth in transport network and radio resources, which can be dynamically given to each client. Network automation and orchestration will have a big role in managing these slices and these resources in a flexible and efficient way. Figure 2.6 depicts the general network slicing architecture.

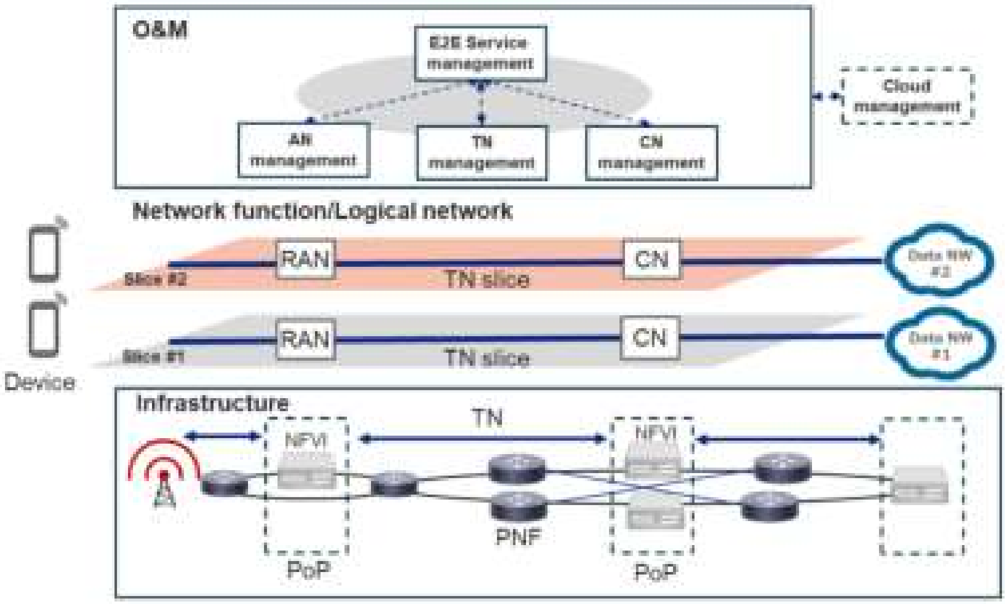


Figure 2.6 – Network slicing architecture (extracted from [GSMA21]).

The network slicing architecture follows a modular design with separation of concerns. Figure 2.6 shows 3 main stratums (going from the bottom to the upper side of the figure):

- **Infrastructure –** Includes all hardware and software. One can have physical nodes that represent a Physical Node Function (PNF) and/or a distributed cloud environment, namely a Network Function Virtualisation Infrastructure (NFVI). One has cell sites that compose the RAN and fibre connections and routing equipment that compose the Transport Network (TN).
- **Network and Application function –** This layer includes the collection of slices that are logical partitions of the network. Each slice is the interlocking of a RAN sub slice, TN slice and Core Network (CN) slice.
- **Operations and Management (O&M) –** The layer with the intelligence to manage all resources and all slices. It is the Operations and Support System (OSS) that allows the deployment and operation of slices.

Each slice performance characteristics is defined in the Service Level Agreement (SLA) between the service provider and the customer. The consumer requirements to the operator are called Service Level Specification (SLS), and may include throughput, latency, and reliability. One key aspect for the deployment of each network slice is the isolation among them. Poor isolation can lead to poor performance or security/privacy concerns. According to [GSMA21] isolation “can be defined as the ability of a Network Slice Provider (NSP) to ensure that congestion, attacks, and lifecycle-related events (e.g., scaling in/out) on one Network Slice Instance (NSI) does not negatively impact other existing NSIs.”

With respect to dynamic slicing, there is also a phased approach to slicing in each of the network segments, namely CN, RAN and TN. CN slicing will be the first natural step, as 5GC design already comes with a cloud-native, microservices architecture with support to network slicing. Then, the next step is RAN with a slicing-aware Radio Resource Management (RRM) policy and finally TN slicing to achieve E2E slicing. TN slicing will be a challenge due to the existing heterogenous technologies (IP/MPLS, optical, microwave) and different resources and topologies. Software Defined Network (SDN) will be crucial helping to cope with this challenge.

Figure 2.7 shows the phased network slicing approach for each network segment.

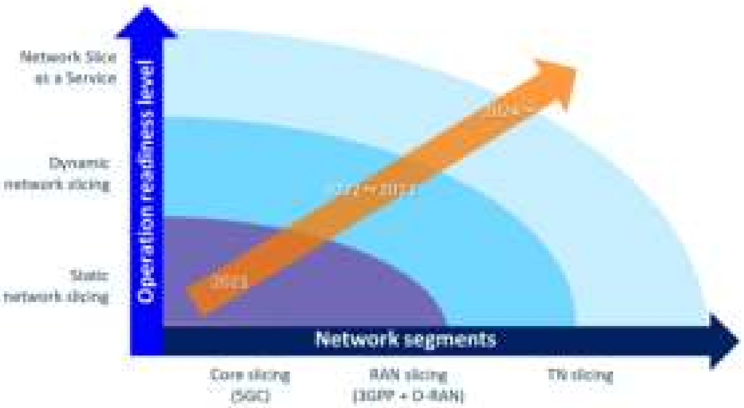


Figure 2.7 – Phased network slicing approach (extracted from [GSMA21]).

One can see the phased approach to slicing throughout the coming years and the network segments to be technology ready for slicing, namely CN, RAN and TN. The goal is to have **Network Slice as a Service** (NSaaS) from 2024 onwards with E2E slicing, which means operators will be able to offer customised network slices in a fast, automated, and efficient way to customers, and give them access to management capabilities with Application Programming Interfaces (APIs).

Network slicing can be used to deploy a **Private Network.** According to [3GPP21a], Private Networks or Non-Public Networks “are intended for the sole use of a private entity such as an enterprise, and may be deployed in a variety of configurations, utilising both virtual and physical elements. Specifically, they may be deployed as completely standalone networks, they may be hosted by a Public Land Mobile Network (PLMN), or they may be offered as a slice of a PLMN.” This allows enterprises to have their own private network, customised to their needs and requirements. Private 5G Networks can offer high data rates, coverage, low latency, and the ability to connect more devices. There are already some implementations on several areas, like Manufacturing, Logistics and Warehousing, Education, Transports and several more.

There are the following modes of deployment of 5G Private Networks:

- **Isolated or Standalone Network –** The entire network is deployed and operated at the customer's premises and is completely isolated from the public network. The enterprise has full control over the network and allows for security and predictable latency. However, it requires a big economical investment and high technical knowledge. All nodes composing the 5G Network Architecture are located at the enterprise premises, including 5GC, UPF and MEC, Figure 2.8.

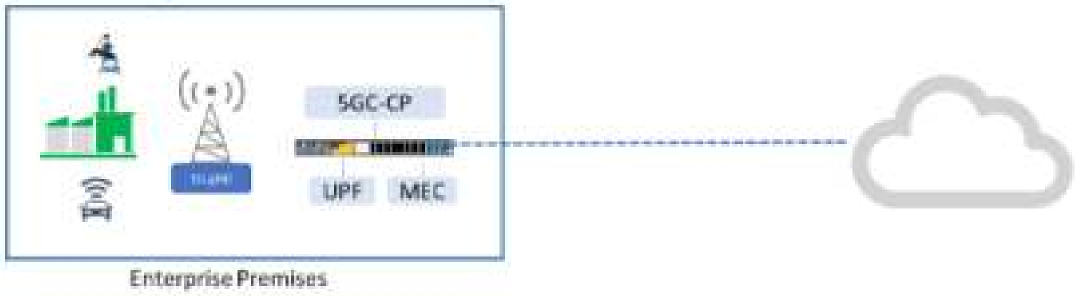


Figure 2.8 – Isolated Private Network (extracted from [WiCh22]).

- **Shared Network –** The network infrastructure is shared with the mobile operator. In this case, depending on the business requirements, the number of components that is managed by the customer and the operator varies, e.g., one can have RAN sharing or core sharing. In a case where **latency** is an important requirement (e.g., smart factory), both UPF and Multi-Access Edge Computing (MEC) should stay in premises, as one can observe at the top image from Figure 2.9, otherwise, some of these components can be at the operator's premises.

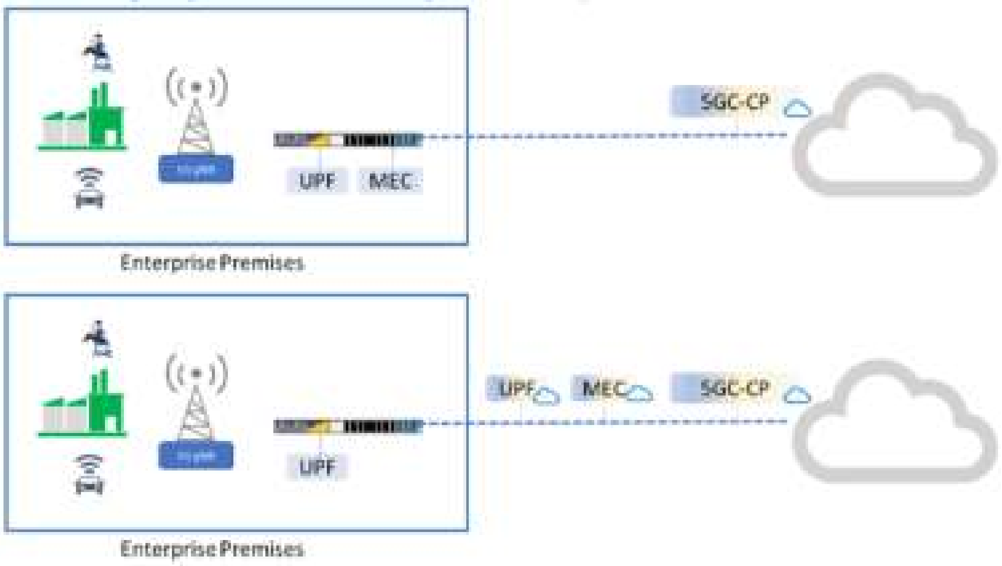


Figure 2.9 – Shared Private Network (extracted from [WiCh22]).

- **Network Slice –** A network slice is allocated to the customer with specific characteristics that are defined in the SLA. This approach has the lowest infrastructure investment cost, but there is low control over the network, higher latency and dependency on the operator respecting QoS. This option is more suitable for wide area deployments, such as smart cities IoT connections, autonomous vehicles or even railways, where centralisation of UPF or MEC is not as important. If anything, one needs several edge nodes that are spread over a large area to still guarantee low latency or high capacity.

Figure 2.10 – Private Network Slice (extracted from [WiCh22]).

With respect to **spectrum**, a 5G private network can use different types of spectrum:

- Licenced – An enterprise can purchase/lease spectrum from the national regulator or from a mobile operator.
- Shared – Verticals may use spectrum that is shared with others, which requires a management system to avoid interference. For instance, in the US verticals can have access to private shared spectrum for their 5G Private Networks with Citizens Broadband Radio Service (CBRS) band.
- Unlicenced – Free to use spectrum with no licence required, having the disadvantage of potential interference from anyone who is using this band.

2.4 Services and Applications

According to ITU-R [ITUR15], there are 3 main service categories of 5G:

- **Enhanced Mobile Broadband (EMBB) –** addresses the access to multi-media content, services and data, and is a consequence of the increase in data traffic and applications that require high bandwidth and high data rates, such as Augmented Reality (AR), Virtual Reality (VR), very high-resolution videos or online gaming.

- **Ultra-reliable and Low-Latency Communications (URLLC) –** dedicated for mission critical applications that require very low latency and high reliability (e.g., autonomous vehicles, industry automation, remote medical surgery, or smart grids automation).

- **Massive Machine Type Communications (mMTC) –** characterised by a very large number of connected devices with specific connectivity requirements, typically transmitting low volumes of data, are low cost and have a long battery life, e.g., IoT devices in a factory or a smart city.

Figure 2.11 illustrates the 3 main services and some examples of usage scenarios, with Smart cities as an example for mMTC, self-driving cars as a scenario for URLLC and high data rates for EMBB.

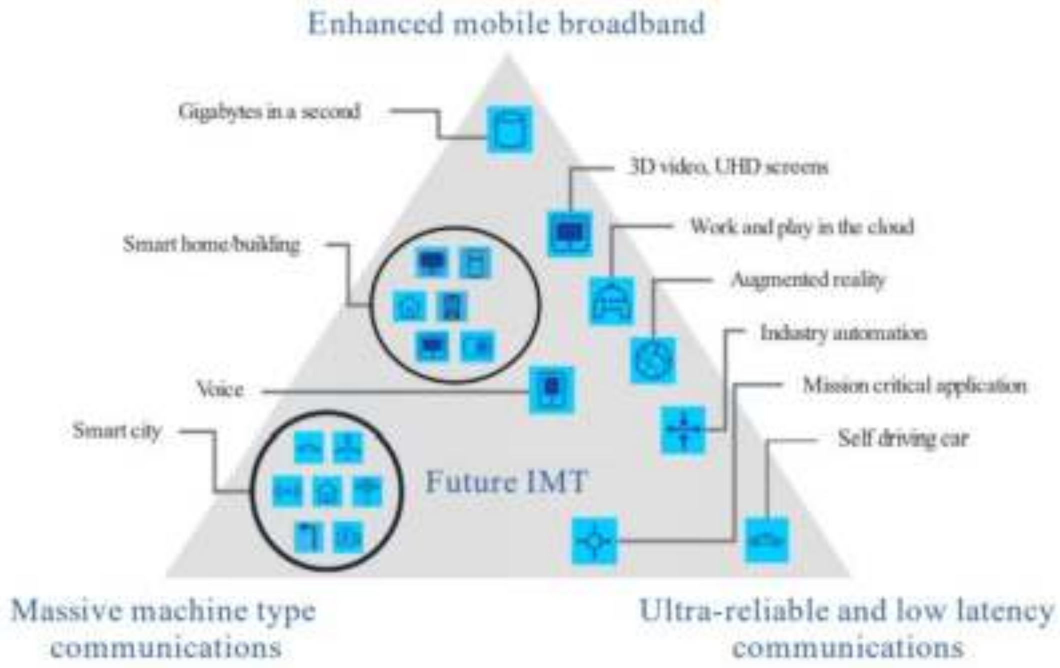


Figure 2.11 – Pillars of 5G and usage scenarios (extracted from [ITUR15]).

Each service category (EMBB, URLCC, mMTC) has its own technical requirements, which are described by important parameters or Key Performance Indicators (KPIs) [ITUR15]:

- **Peak Data Rate –** Maximum theoretical data rate per user/device (in Gbps).
- **User Experienced Data Rate –** Real data rate experienced per user/device across the coverage area (in Gbps or Mbps).
- **Latency –** Time that is takes from data to be transmitted from source to destination (in ms).
- **Mobility –** Maximum speed at which a defined QoS and seamless handover can be achieved (in km/h).
- **Connection Density –** Total number of connected devices per area (in km^2).
- **Energy Efficiency –** On the network side, it refers to the quantity of information bits transmitted/received to/from users, per unit of energy consumption of the RAN (in bit/Joule). On the device side it is the same, but for the energy consumption of the communication module.
- **Spectrum Efficiency –** Average data throughput per unit of spectrum and per cell (bit/s/Hz).
- **Area Traffic Capacity –** Total traffic throughput served per geographical area (in Mbit/s/m^2).

Figure 2.12 depicts the most important metrics for each one of the service categories. This can be seen with an indicative scaling that range from “high”, “medium” and “low”. The relevance of each metric depends on the usage scenario and on the type of service. In the EMBB scenario, experienced data rate, peak data rate, spectrum efficiency, area traffic capacity and network energy efficiency all have high importance. Whereas with mMTC, connection density has the most importance and the network energy efficiency with medium importance, since one has a lot of devices connected that must have a long operational lifetime. Finally, with URLLC, one clearly gives the highest importance to latency. Mobility is important as well with applications, such as autonomous driving that need low latency and high mobility.

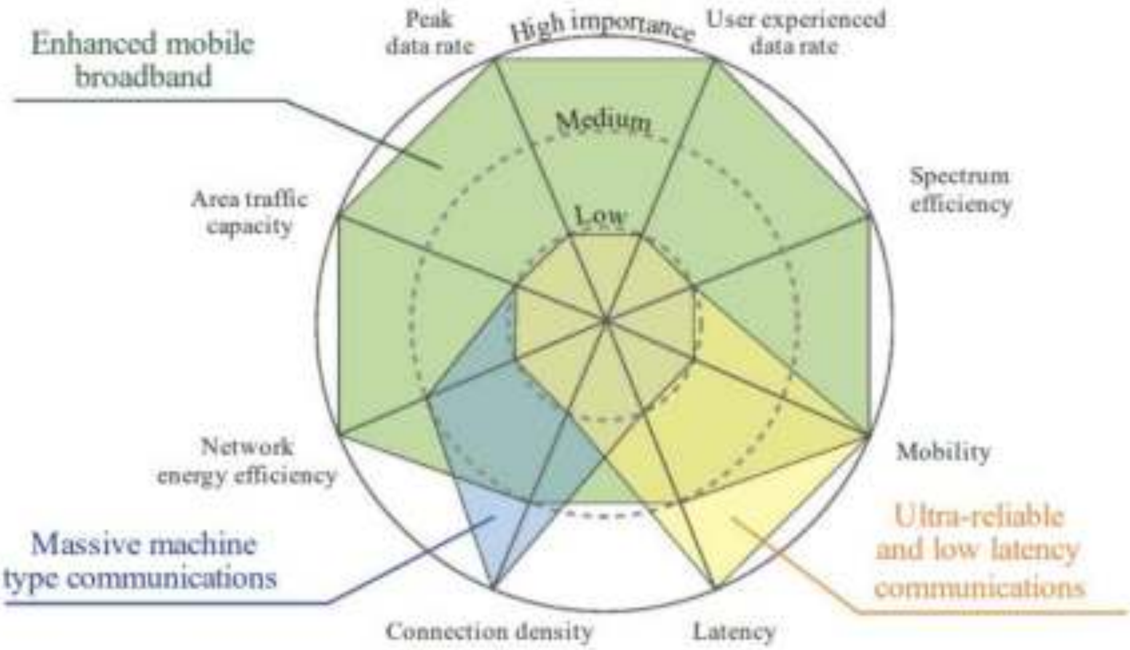


Figure 2.12 – Different 5G service metrics in different usage scenarios (extracted from [ITUR15]).

Figure 2.13 illustrates the target values of IMT-2020 (5G) of each metric compared to IMT-advanced (4G). The user experienced data rate is expected to reach 100 Mbps in 5G, ten times higher than 4G. This is a big increase that allows for EMBB applications. Moreover, the connection density is expected to be 1 million devices per square kilometre, which will enable mMTC scenarios. The latency is expected to get to 1 ms. Such latencies will be key in URLCC applications and use cases.

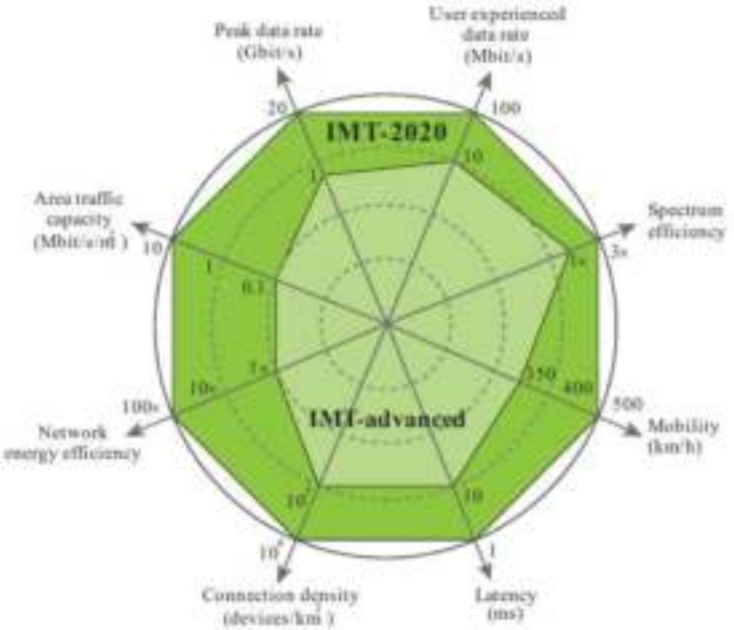


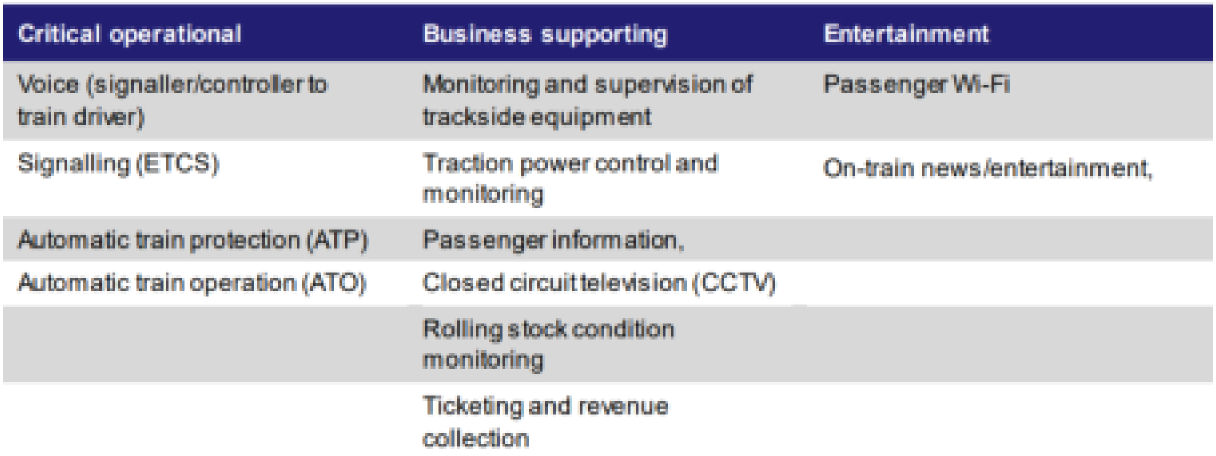
Figure 2.13 – Comparison of IMT-advanced (4G) and IMT-2020 (5G) (extracted from [ITUR15]).

Finally, IMT-2020 is also expected to have a high mobility that can go up to 500 km/h with acceptable QoS. This is particularly important in railways communications, which is the focus of this project.

2.5 Railways Communications

There are different communications systems in railways with different applications and characteristics. Railway communications can be categorised as, Figure 2.14: **critical operational**, such as railway signalling or voice communications, **business supporting**, such as Closed-Circuit television (CCTV), or **entertainment**, which can be passenger Wi-Fi [ERAS14]**.** To support these different railway communication services, one needs a special communication system that can provide the necessary requirements and specifications, which include **GSM-R**, **LTE-R** and **BBRS.**

GSM-R (Global System for Mobile Communications-Railways) is a system based on GSM with some enhancements specific for railway operation. GSM-R offers short message service (SMS), voice and data services. It provides a communication system between the train drivers and the control centres with features like group communications, emergency calls and priority levels.



Critical operational	Business supporting	Entertainment
Voice (signaller/controller to train driver)	Monitoring and supervision of trackside equipment	Passenger Wi-Fi
Signalling (ETCS)	Traction power control and monitoring	On-train news/entertainment,
Automatic train protection (ATP)	Passenger information,	
Automatic train operation (ATO)	Closed circuit television (CCTV)	
	Rolling stock condition monitoring	
	Ticketing and revenue collection	

Figure 2.14 – Railway communications services (extracted from [ERAS14]).

In GSM-R, there are 4 MHz of bandwidth for both UL and DL. The bands are between 876-880 MHz for UL and 921- 925 MHz for DL. In Figure 2.15, one can see the bands for GSM-R, for public GSM and for an extended bandwidth for railway ER-GSM that can be granted by the regulator when the bandwidth is not sufficient.

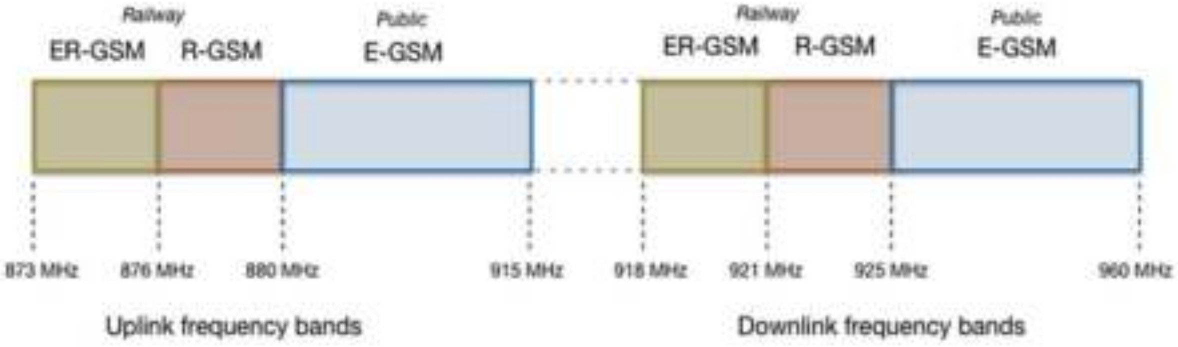
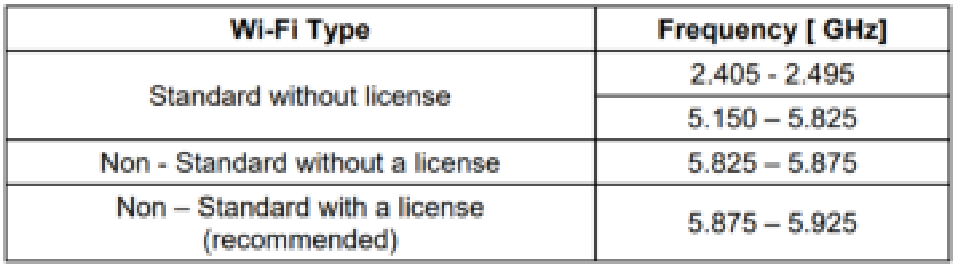


Figure 2.15 – GSM-R, ER-GSM and Public GSM bands (extracted from [Wolf18]).

LTE-R is the upgrade from GSM-R, based on LTE. When compared to GSM-R, it offers higher data capacity, lower latencies, and higher security. LTE-R was not widely deployed in Europe, and one might watch Next Generation railway communications based on 5G overtake this communication system.

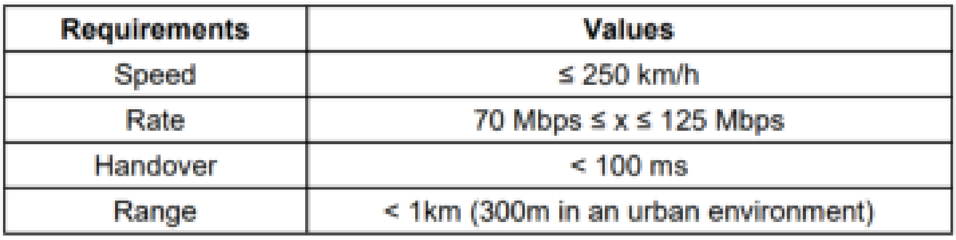
BBRS is a solution developed by Thales, which is based on Wi-Fi. It requires BSs spread over the railway track and access points on the trains. It offers bidirectional communication and typically operates in the 2.4 GHz or 5 GHz, with channels of 20 MHz or 40 MHz depending on the system. Table 2.2 illustrates the different working frequencies for BBRS. It is recommended to use the licenced frequencies due to interference reasons, security and as the non-standard spectrum has 10 to 100 times more signalling power than the standard one. Table 2.3 shows the performance requirements provided by this solution. It is important to note that the range of this system is less than 300 m in urban scenarios, which implies a dense distribution of BSs, due to the high frequencies that are used and their propagation characteristics.

Table 2.2 – Working frequencies for BBRS (extracted from [Corr19]).



Wi-Fi Type	Frequency [GHz]
Standard without license	2.405 - 2.495
	5.150 – 5.825
Non - Standard without a license	5.825 – 5.875
Non – Standard with a license (recommended)	5.875 – 5.925

Table 2.3 – BBRS system performance (extracted from [Corr19]).



Requirements	Values
Speed	≤ 250 km/h
Rate	70 Mbps ≤ x ≤ 125 Mbps
Handover	< 100 ms
Range	< 1km (300m in an urban environment)

In Europe, there was a need to standardise a railway communication and control system to increase competitiveness and remove barriers to international journeys. A unique European train control system was created, the **European Railway Traffic Management System (ERTMS)**, which was developed by the European Rail Supply Industry Association (UNIFE), International Union of Railways (UIC), the European Union and GSM-R industry. ERTMS has two components [ERTM22]:

- **European Train Control System (ETCS)** – an automatic train protection system (ATP) to replace the national existing systems, consisting of using standardised equipment, such trackside or train on-board equipment, to calculate maximum safe speed of the train, automatic control and signalling for the drivers. There are 3 levels of implementation with less or more track equipment necessary, and continuous or intermittent exchange of information between track and trains.
- **GSM-R** – the radio system, previously explained, to provide data and voice communication. It consists of BSs masts close to the railway with around 7 km to 15 km between them.

In Figure 2.16, one can see the ERTMS/ETCS level 2 that comprises the train on-board equipment, such as the GSM-R antenna, the driver's interface, the European Vital Computer and Balise Readers. Moreover, there is trackside equipment, like the Eurobalises, GSM-R mast BS that exchange information with the train and other components, like the Radio Block Centre that is the brain of the operation.

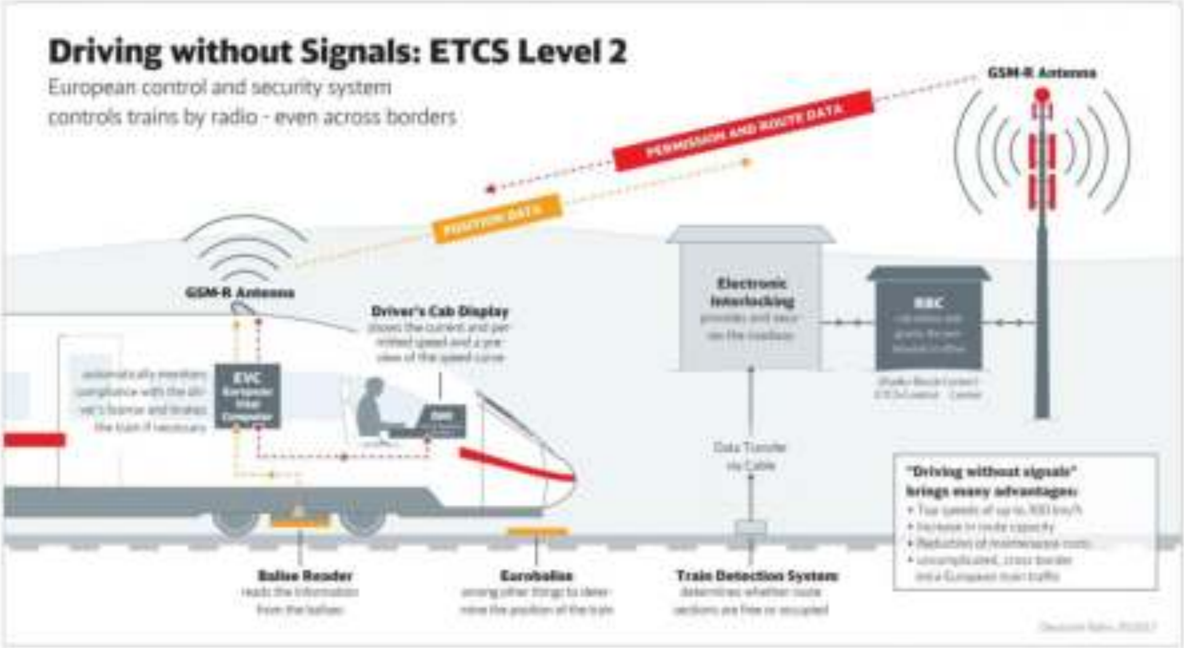


Figure 2.16 – ERMTS/ETCS level 2 (extracted from [DBRa20]).

2.6 State of the Art

GSM-R is predicted to become obsolete in 2030 with end of equipment support by manufacturers. so there is a need to come up with a new communication system for railways. Future Railway Mobile Communication System (FRMCS), which is based on 5G, is expected to be the successor, and was initiated by UIC by developing the first User Requirements Specifications. FRMCS is a key enabler to fully digitalise the railway industry, support an increasing level of automatic train operations (ATO), new services and applications, while aiming for cost effectiveness, future proof concepts and seamless migration of GSM-R. Figure 2.17 shows the general architecture of FRMCS with interoperability to GSM-R from 3GPP.

FRMCS will have dedicated 5G bands for railways, but, for now, the spectrum available is not enough for modern railway services requirements, such as critical video communications. Currently, the bands in which FRMCS can operate are the GSM-R ones ([874.4, 880.0] MHz and [919.4, 925.0] MHz) and the TDD band [1900, 1910] MHz, which was approved by the Electronic Communications Committee [Eric22a]. The 1900 band is more suitable for FRMCS, since it can fulfil critical application requirements with low to medium throughput ([0.1, 20] Mbps). Moreover, the European Commission has adopted the Implementation Decision on the use of the maximum of 10 MHz for Intelligent Transport Systems, namely, urban rail and metros [EuCo20].

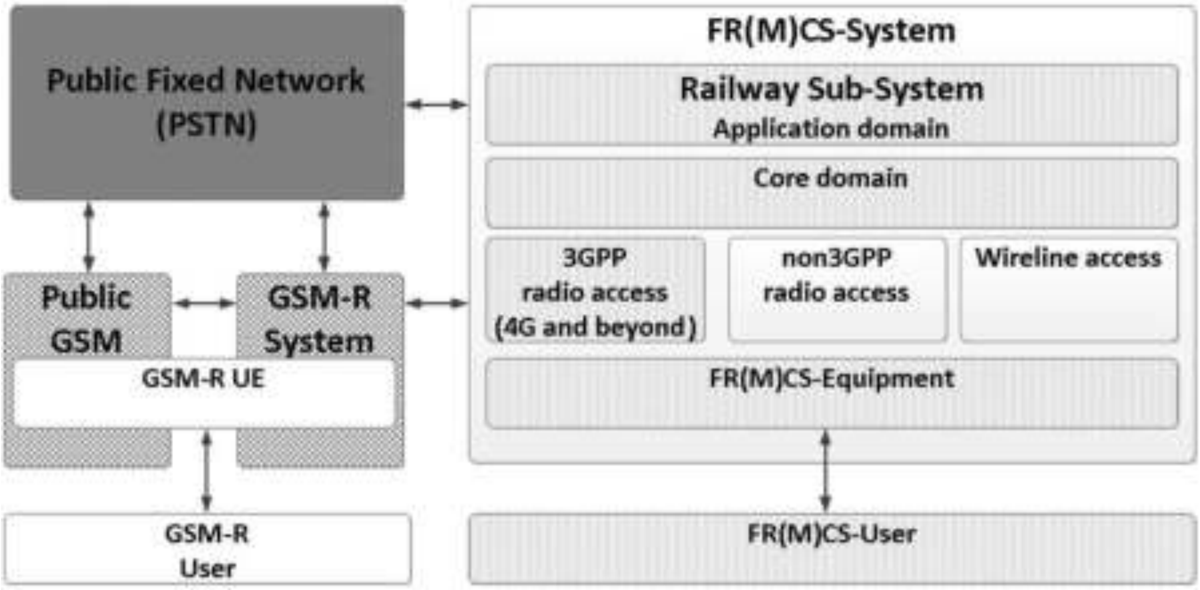


Figure 2.17 – 3GPP FRMCS Architecture with legacy GSM-R (extracted from [3GPP21b]).

One option to mitigate the lack of spectrum and satisfy high throughput demanding applications (more than 30 Mbps on Uplink) is to create hybrid networks with 5G slicing capabilities so Communication Service Providers (CSPs) can supplement the capacity required [Eric22b]. One would have RAN sharing between the railway operator and the CSP providing both coverage and capacity. Figure 2.18 illustrates a hybrid architecture where one has both the railway operator network and the CSP network with RAN slicing capabilities. Critical applications, such as voice communication (between train drivers, maintenance staff and central office dispatchers) and train signalling would be hosted by the railway operator network whereas applications such as CCTV, video streaming and passenger entertainment, would be hosted by the CSP network. Moreover, the latter would serve as a fallback network in case any issue happens.

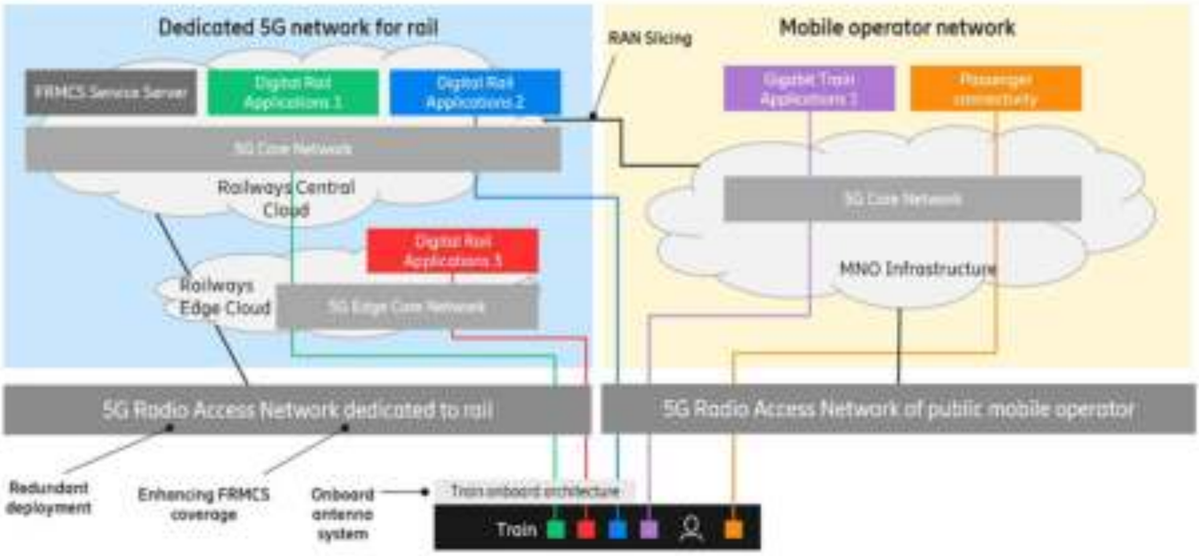


Figure 2.18 – Hybrid Architecture for 5G Railway deployment (extracted from [Eric22c]).

Figure 2.19 depicts the RAN sharing model between a CSP and the railway operator with different spectrum possibilities, such as the 1900 MHz railway dedicated band for critical applications, the CSP 4G/5G spectrum for business applications, like internet access to passengers, and finally unlicenced spectrum, which can be useful giving extra capacity in crowded areas, e.g., in train stations.

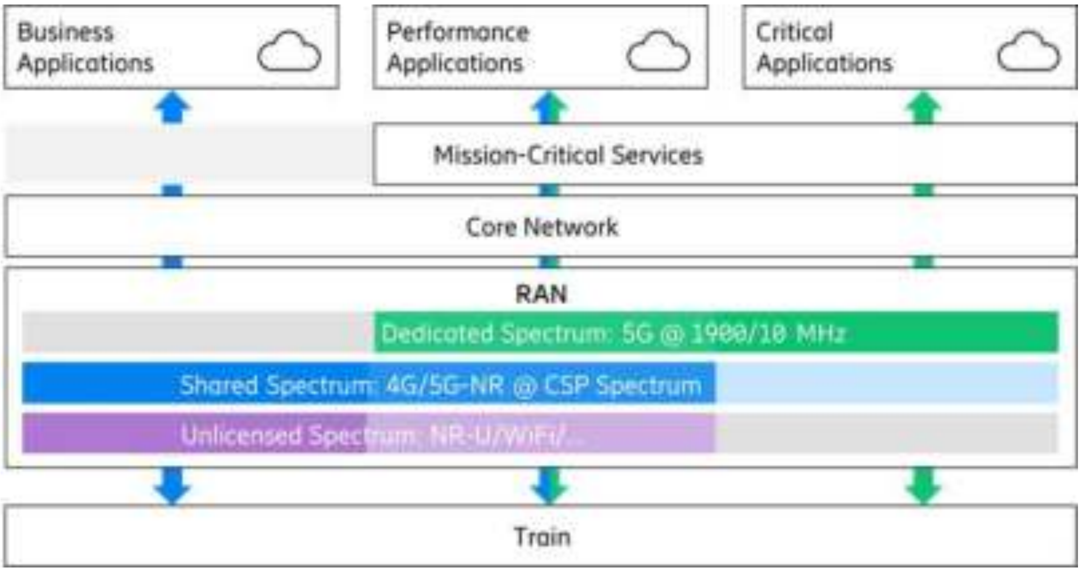


Figure 2.19 – RAN sharing for different railway applications (extracted from [Eric22d]).

5G Rail is a project funded by the European Union, which goal is to validate the first set of FRMCS specifications by developing and doing pilot tests. There are two labs in which they test and prototype FRMCS along with GSM-R. This system interoperability between GSM-R and FRMCS is key to the successful system migration allowing the re-use of existing infrastructure and flexibility in terms of network deployment. The first lab is situated at Nokia premises in Hungary and focus on voice applications, ETCS, CCTV and video. Figure 2.20 shows the lab architecture.

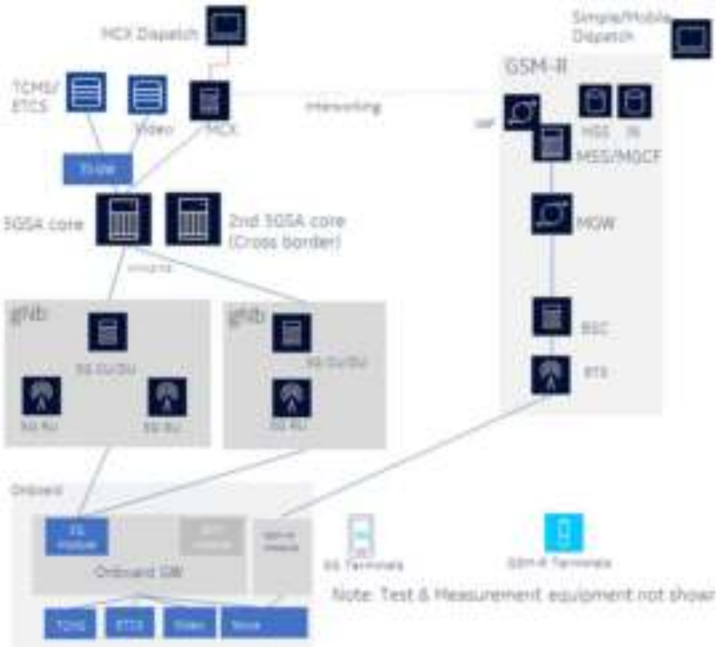


Figure 2.20 – 5G RAIL - Hungary Test lab for FRMCS (extracted from [5Gra22]).

The second lab is situated in France, driven by Kontron Transport, which provides the 5G SA infrastructure, as well as Thales, which provides an access via VPN to its Passenger Information System (PIS), and Alstom that brings ETCS/ATO expertise. Figure 2.21 illustrates the lab components.

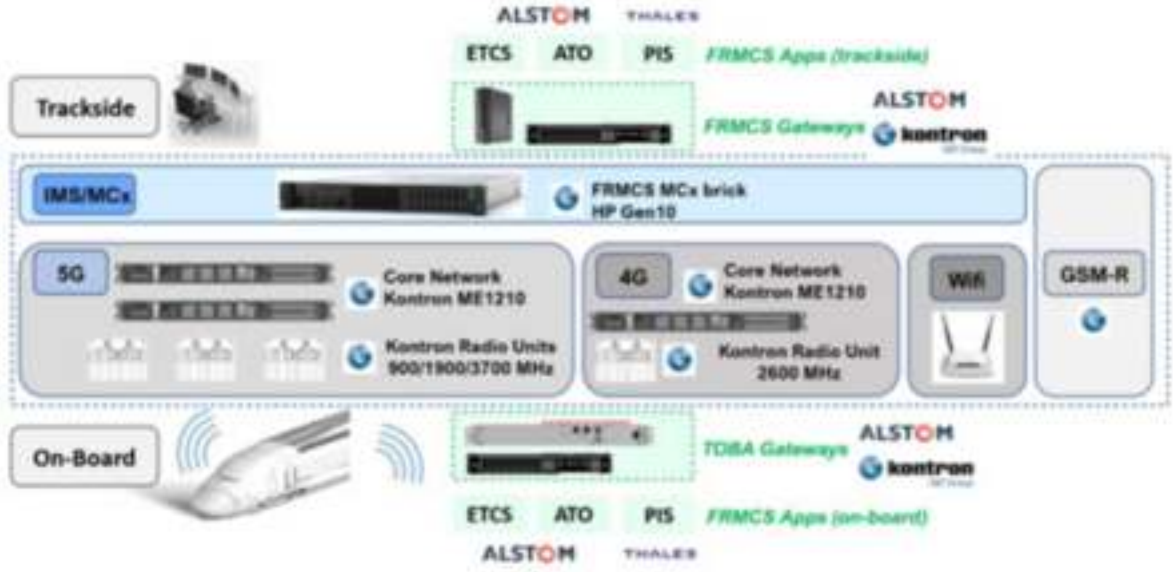
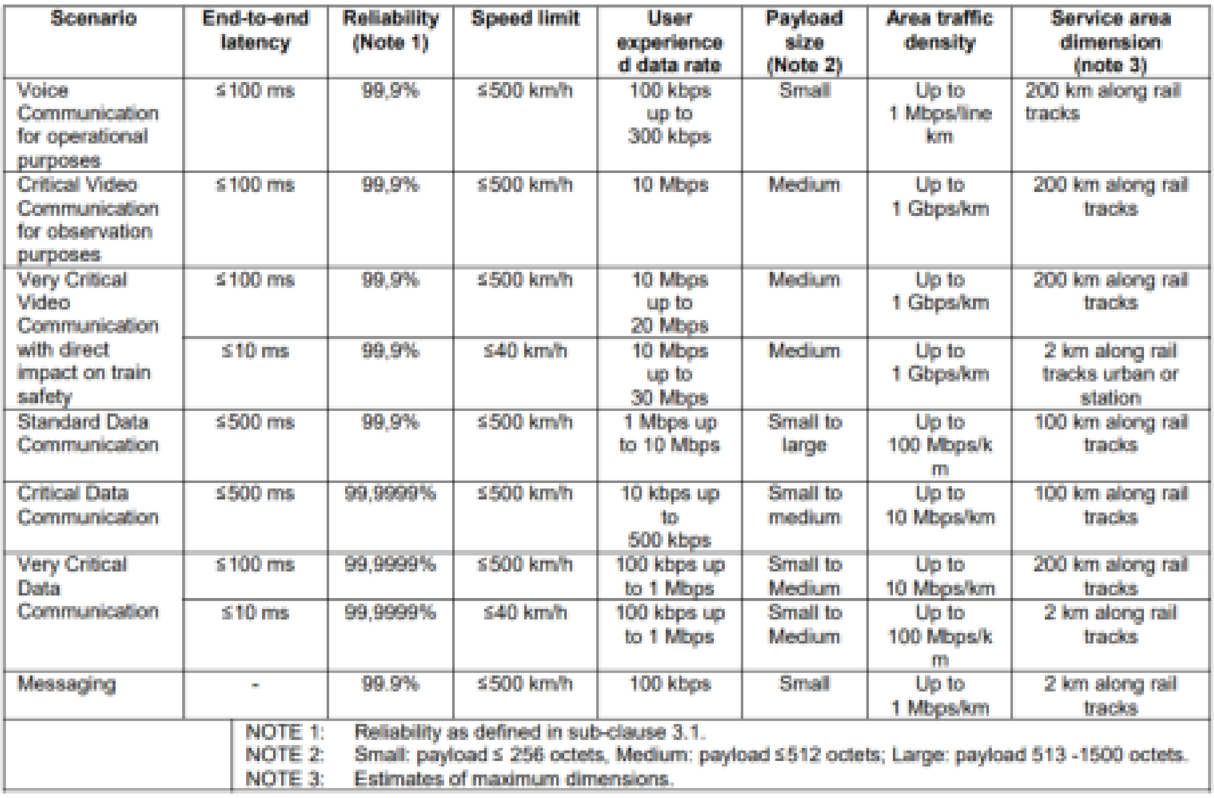


Figure 2.21 – 5G RAIL - France Test lab for FRMCS (extracted from [5Gra22]).

3GPP has specifications on the performance requirements for several applications in railway scenarios for mainlines. These can be seen in Table 2.4, which include services like voice, data, video, messaging, and metrics such as latency, data rate and train speed limit.

Table 2.4 – Performance Requirements for railway scenarios (extracted from [3GPP20]).



Scenario	End-to-end latency	Reliability (Note 1)	Speed limit	User experienced data rate	Payload size (Note 2)	Area traffic density	Service area dimension (note 3)
Voice Communication for operational purposes	≤100 ms	99,9%	≤500 km/h	100 kbps up to 300 kbps	Small	Up to 1 Mbps/line km	200 km along rail tracks
Critical Video Communication for observation purposes	≤100 ms	99,9%	≤500 km/h	10 Mbps	Medium	Up to 1 Gbps/km	200 km along rail tracks
Very Critical Video Communication with direct impact on train safety	≤100 ms	99,9%	≤500 km/h	10 Mbps up to 20 Mbps	Medium	Up to 1 Gbps/km	200 km along rail tracks
	≤10 ms	99,9%	≤40 km/h	10 Mbps up to 30 Mbps	Medium	Up to 1 Gbps/km	2 km along rail tracks urban or station
Standard Data Communication	≤500 ms	99,9%	≤500 km/h	1 Mbps up to 10 Mbps	Small to large	Up to 100 Mbps/km	100 km along rail tracks
Critical Data Communication	≤500 ms	99,9999%	≤500 km/h	10 kbps up to 500 kbps	Small to medium	Up to 10 Mbps/km	100 km along rail tracks
Very Critical Data Communication	≤100 ms	99,9999%	≤500 km/h	100 kbps up to 1 Mbps	Small to Medium	Up to 10 Mbps/km	200 km along rail tracks
	≤10 ms	99,9999%	≤40 km/h	100 kbps up to 1 Mbps	Small to Medium	Up to 100 Mbps/km	2 km along rail tracks
Messaging	-	99.9%	≤500 km/h	100 kbps	Small	Up to 1 Mbps/km	2 km along rail tracks

NOTE 1: Reliability as defined in sub-clause 3.1.
NOTE 2: Small: payload ≤ 256 octets, Medium: payload ≤512 octets; Large: payload 513 -1500 octets.
NOTE 3: Estimates of maximum dimensions.

Chapter 3

Model and Simulator Description

This chapter gives an overview of the developed model with the theoretical background, its inputs and outputs, its implementation as well as a characterisation on the analysed services.

3.1 Model Overview and Service Requirements

This thesis involves the implementation of a model that studies the different possible private network architectures for different railway scenarios. This model has a list of different inputs, processing functions and finally the outputs that provide the results. Then, one analyses if the results are optimal and if one architecture is possible according to the existing constraints.

It is noteworthy to mention that the developed model and sub-models are based on [Carv22] but adapted and modified to meet the needs of this thesis. The components that were extracted, adapted, modified, and created are explicitly described in Section 3.4.

Figure 3.1 depicts a scheme with the model overview, the inputs required, the intermediate steps and outputs of the program.

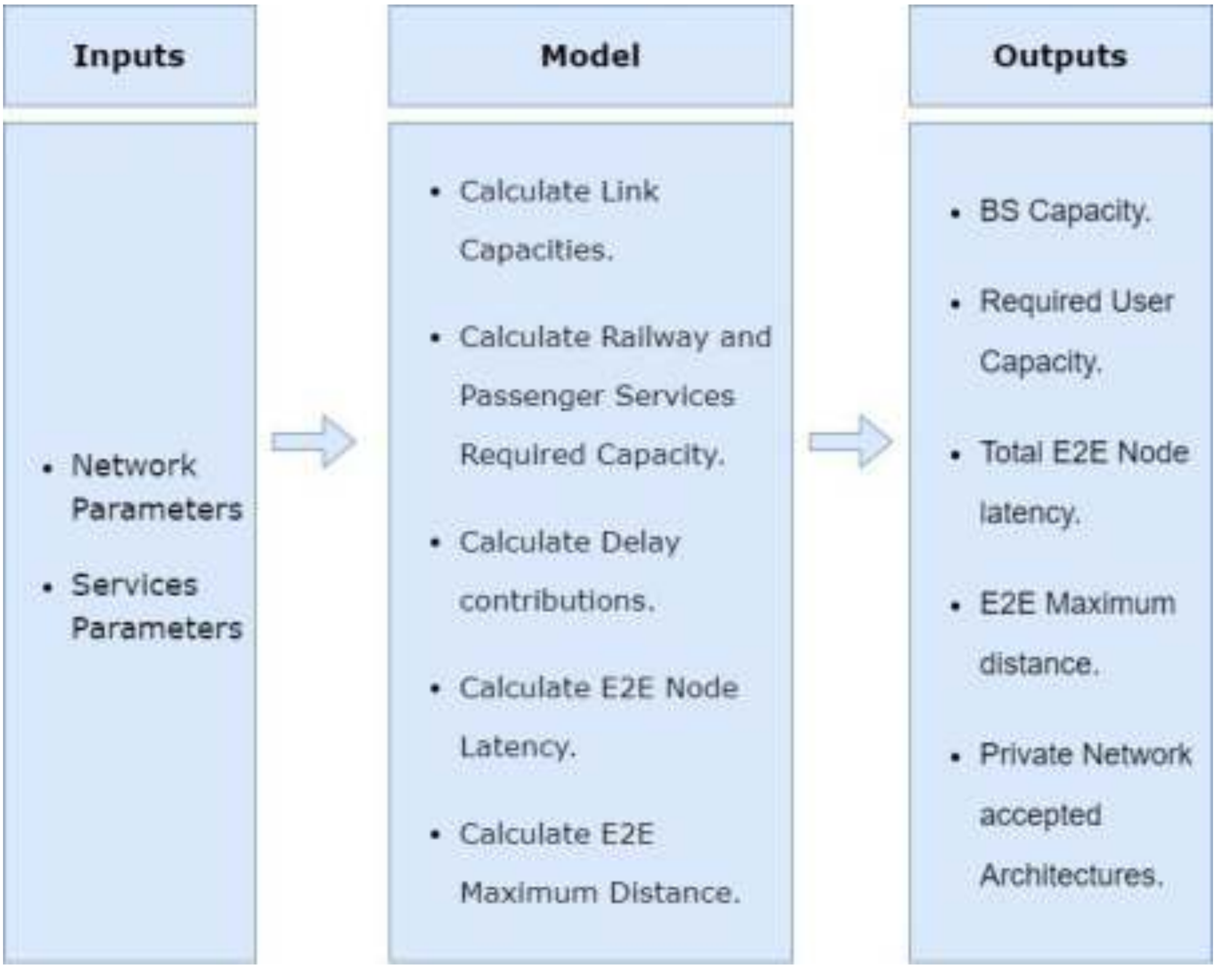


Figure 3.1 – Model Overview.

For the inputs of this model, one has two categories: Network and Services Parameters. The Network parameters and specifications consist of information related to the network itself, the hardware installed and characteristics, the different type of links, i.e.:

- Network Architecture – The nodes arrangement (RU, DU, CU) and collocation.
- MEC Option – The existence of a MEC node in the network and its location.

- FH links – Used splitting option for Fronthaul links.
- MH links – Used splitting option for Midhaul links.
- BH links – BH links capacity.
- Transport links – TL links capacity.
- Radio Parameters – Number of MIMO layers, numerology, total bandwidth, Channel Quality Indicator (CQI), Scaling Factor (SF), Frame Structure (FS).
- Number of Users – Number of users connected to each type of node.
- RU/DU/CU Service Mix – Percentage of Users that are using a specific service connected to a specific node.
- User Distance to the RU – Distance between the UE and the RU to calculate the propagation latency in the air.
- Service to be simulated – Service that is simulated from the list of services that can be of two categories: railway specific or passenger specific.
- Maximum user distance to the Core of the railway's operator and CSP.

As for Services parameters, it includes:

- Maximum E2E Latency – Maximum latency supported by the service.
- Data Rate – Data rate of the service.
- Priority – A smaller value corresponds to a service with higher priority.
- Packet Size – Packet Size of the Service.
- Latency Adaptation Parameter – Assignment of processing capabilities of the node to a service. More processing capabilities are assigned to services with a lower value resulting in a lower node latency.
- Link Type – If service is of the type of DL, UL or 50/50.

All these parameters are stored as variables and used to create the output parameters. The intermediate steps and functions use equations that are described in the following sub-chapter. These processes are seen in the middle box of Figure 3.1 with the calculations of the required expressions.

For each set of input parameters that describe the network and scenario, there are output parameters that are analysed and determine whether the specific configuration is possible. The output parameters for the model are:

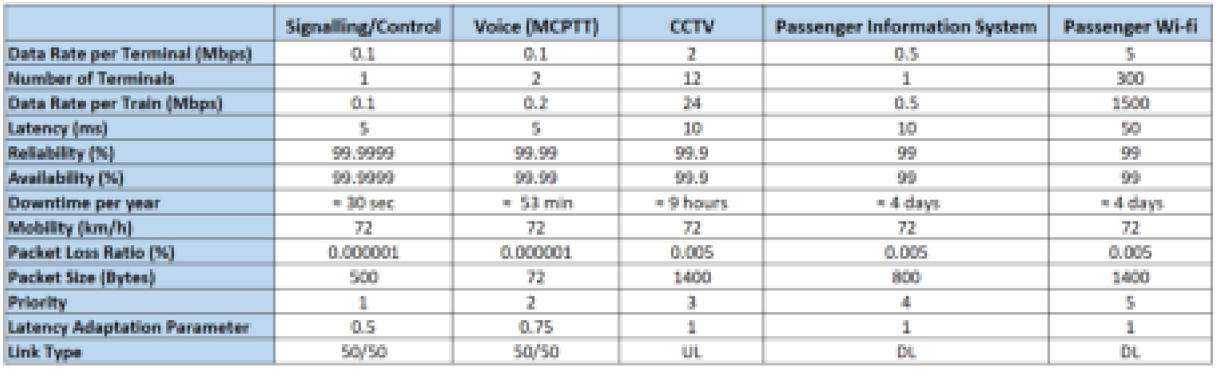
- RU Capacity – The capacity that the RU provides in both DL and UL.
- RU Required Capacity – The capacity that is required for the RU considering the number of users that are connected and the services that these are using, both in DL and UL.
- Link Throughputs – Throughputs of the different links in the network.
- Total Node Latency – Latency of the network with all contributions, except the propagation delay between the nodes, which cannot exceed the maximum latency for the service.
- E2E Maximum Distance – The maximum E2E physical distance between the UE and the node which hosts the application, to guarantee a specific service latency. This parameter helps to know where the nodes can be installed and helps excluding possible implementation solutions.
- Accepted Private Network – The types of private networks that satisfy all the requirements.

The set of services that are analysed fit into two categories:

- **Passenger Services** – All services and applications that are used by or dedicated to railway passengers, e.g., Passenger Information System (PIS) or Passenger Connectivity (Wi-Fi).
- **Railway Specific Services** – All services and applications dedicated to the railway systems and operations which include the railway signalling systems, the Voice service Mission-critical push-to-talk (MCPTT) and the CCTV.

These different types of services have different requirements and specifications, which need to be taken in account when analysing a possible architecture for the system. Moreover, some services have a higher priority than others, due to criticality and more strict rules. The requirements for the different services are illustrated in Table 3.1. The data is suited for railway-subway scenarios.

Table 3.1 – Service Requirements and Specifications.



	Signalling/Control	Voice (MCPTT)	CCTV	Passenger Information System	Passenger Wi-fi
Data Rate per Terminal (Mbps)	0.1	0.1	2	0.5	5
Number of Terminals	1	2	12	1	300
Data Rate per Train (Mbps)	0.1	0.2	24	0.5	1500
Latency (ms)	5	5	10	10	50
Reliability (%)	99.9999	99.99	99.9	99	99
Availability (%)	99.9999	99.99	99.9	99	99
Downtime per year	≈ 30 sec	≈ 53 min	≈ 9 hours	≈ 4 days	≈ 4 days
Mobility (km/h)	72	72	72	72	72
Packet Loss Ratio (%)	0.000001	0.000001	0.005	0.005	0.005
Packet Size (Bytes)	500	72	1400	800	1400
Priority	1	2	3	4	5
Latency Adaptation Parameter	0.5	0.75	1	1	1
Link Type	50/50	50/50	UL	DL	DL

One can see in Table 3.1 the 5 main services that are analysed, listed from the one with higher priority in the left to the one with lower priority on the right. For each service, one has: data rate per terminal, projected number of terminals, data rate per train (being the product of the previous two) which is the data rate required per train, latency, reliability, availability, downtime per year, mobility parameter, packet loss ratio, packet size, priority, latency adaptation parameter and link type.

3.2 Latency Model

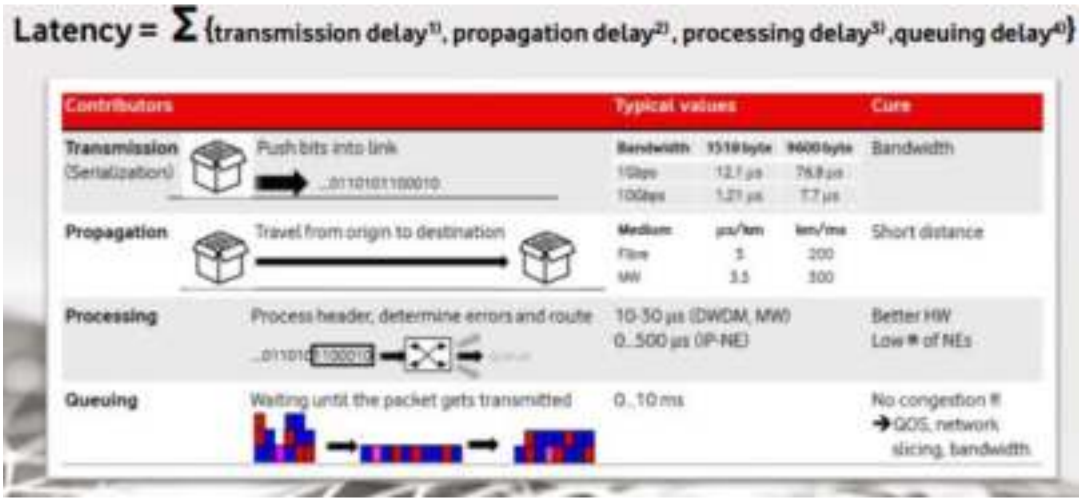
3.2.1 Latency Contributions

The network latency can be split into several delays that accumulate across the network, [Carv22]:

- Transmission Delay – Time that it takes to transmit bits in the links. Increasing the available bandwidth reduces this delay.
- Propagation Delay – Time that it takes a packet to travel between two points. One needs to reduce the physical distance between points to reduce this delay. The type of medium used to transmit information also influences this value, e.g., one has a lower delay using microwaves as opposed to optical fibres.
- Processing Delay – Time that it takes for a node to process the packet in a specific node. This processing depends on the number and complexity of functionalities of the node and the processing capacity. One can reduce this delay by improving processing algorithms or using better hardware.

- Queuing Delay – Time required to wait between the packet arriving to the node and being ready to be processed in the node. This delay depends on the throughput in the output of the node, the traffic aggregated in the node and the priority level associated to the service. This type of delay can be reduced by using priority-base queuing, QoS and increasing bandwidth.

Figure 3.2 shows the several type of delays that when sum up together result in the total network latency.



$$\text{Latency} = \sum \{\text{transmission delay}^{1)}, \text{propagation delay}^{2)}, \text{processing delay}^{3)}, \text{queuing delay}^{4)}\}$$

Contributors		Typical values	Cure
Transmission (Serialization)	Push bits into link ...0110101100010	**Bandwidth** / **1518 byte** / **9600 byte** 10Gbps / 12.1 µs / 76.8 µs 100Gbps / 1.21 µs / 7.7 µs	Bandwidth
Propagation	Travel from origin to destination	**Medium** / **µs/km** / **km/ms** Fibre / 5 / 200 MW / 3.3 / 300	Short distance
Processing	Process header, determine errors and route ...011010 100010	10-50 µs (DWDM, MW) 0...500 µs (IP-NE)	Better HW Low # of NEs
Queuing	Waiting until the packet gets transmitted	0...10 ms	No congestion # ➔ QOS, network slicing, bandwidth

Figure 3.2 – Different types of delays that contribute to the latency (extracted from [VODA18]).

In order to calculate the latency in the 5G network, one must calculate the individual delays of all nodes, namely, UE, RU, DU, CU, 5G Core, External Data Centre (EDC) and MEC nodes, and the links between them, according to the specific deployment of nodes. If the MEC node does not exist, one has the following contributions [Domi19]:

- Processing in the UE.
- Transmission latency bound to the links (Air Link, Fronthaul, Midhaul, Backhaul and Transport Link).
- Propagation latency in the links (AL, FH, MH, BH, TL).
- Queuing latency in the nodes (RU, DU, CU, Network Core).
- Processing latency in the nodes (RU, DU, CU, Network Core and External Data Centre).

When studying the possible architectures, one must take into consideration the use of MEC nodes as well. These are relevant specially in applications with strict latency requirements or even in private networks by hosting specific applications at the edge of the network. For instance, a MEC node could be deployed to prevent data from being transmitted to the Core that could be located hundreds of kilometres distant and eliminating the core transmission and processing delay. There are 4 possible MEC node deployment options, [Carv22]:

- MEC node between the RU and the DU (Option RU/DU) – optimal solution to reduce latency but difficult to implement in practice. This option reduces radio latency and Fronthaul throughput and can achieve very low E2E latency (below 1 ms).
- MEC node between the DU and CU (Option DU/CU) – Viable solution to reduce the propagation latency between the DU and CU, which comprises the midhaul, typically with long lengths.

- MEC node between the CU and the Network Core (Option CU/Core) – This option saves the processing delay that takes part in the core of the network and the propagation delay as well in the Backhaul. It is not optimal, since there are 3 nodes between the UE and the MEC node.
- MEC node between the Network Core and the External Network– Reduces the latency between the network core and any external network that can be located far away. There is still the delay associated with the rest of the nodes, therefore making it not the best solution in strict latency application requirements.

In the case of nodes that are collocated, the MEC node could still be deployed to reduce latency. One neglects the distance between the nodes if these are collocated. The MEC node can be deployed between any pair of nodes, even if they are co-located.

The best scenario to take advantage of the MEC node in reducing the E2E latency is deploying it as the DU, thus, the Option RU/DU. It is worthwhile to say that is impossible to replace logically the RU with the MEC node due to the complexity, high cost and scalability requirements. Hence, with the presence of a MEC node the main differences on the latency are the existence of the MEC processing delay and the latency accumulated between the UE and MEC that varies with its location. With a MEC node, there is no need to send data forward in the network reducing all the delays associated with the nodes that are deployed after it.

Figure 3.3 illustrates an overview of the network and the respective latency contributions that compose the E2E latency, which is the sum of the transmitter latency in the upper part and the receiver latency in the bottom part. One sees the nodes and their position in the network including the MEC node, which can have different deployment options.

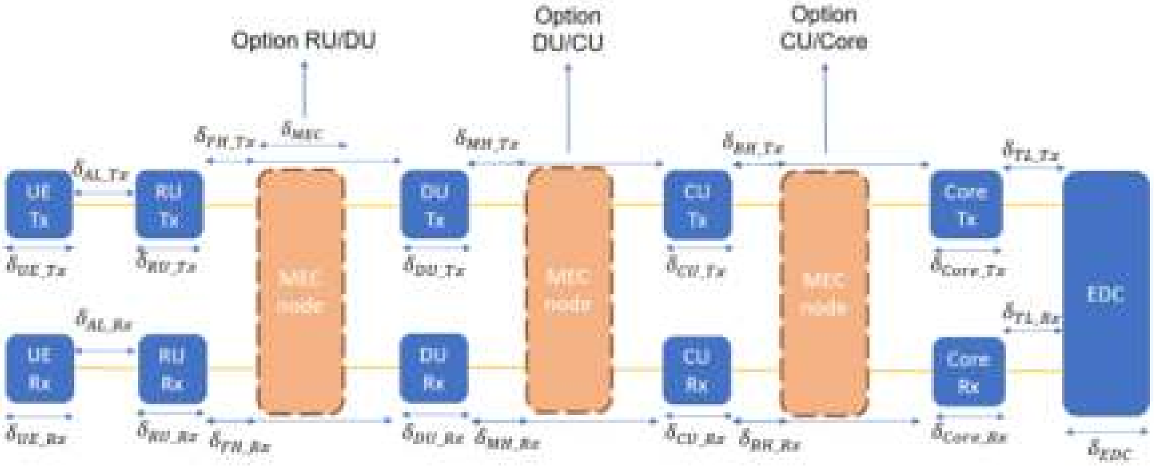


Figure 3.3 – Latency components in the 5G Architecture (extracted from [Carv22]).

All delays that contribute to the latency have expressions that help developing the model for latency calculation, which is based on [Carv22]:

$$\delta_{UE\_Tx} = \delta_{UE\_Proc} + \delta_{UE\_Trans} \tag{3.1}$$

$$\delta_{UE\_Rx} = \delta_{UE\_Proc} \tag{3.2}$$

$$\delta_{RU\_Tx} = \delta_{RU\_Rx} = \delta_{RU\_Proc} + \delta_{RU\_Queu} + \delta_{RU\_Trans} \tag{3.3}$$

$$\delta_{DU\_Tx} = \delta_{DU\_Rx} = \delta_{DU\_Proc} + \delta_{DU\_Queu} + \delta_{DU\_Trans} \tag{3.4}$$

$$\delta_{CU\_Tx} = \delta_{CU\_Rx} = \delta_{CU\_Proc} + \delta_{CU\_Queu} + \delta_{CU\_Trans} \tag{3.5}$$

$$\delta_{CORE\_Tx} = \delta_{CORE\_Rx} = \delta_{CORE\_Proc} + \delta_{CORE\_Trans} \tag{3.6}$$

$$\delta_{EDC} = \delta_{EDC\_Proc} + \delta_{EDC\_Trans} \tag{3.7}$$

$$\delta_{MEC} = \delta_{MEC\_Proc} + \delta_{MEC\_Trans} \tag{3.8}$$

where:

- δ_{UE_Tx}, δ_{UE_Rx} - Accumulated latency in the UE in transmission and reception.
- δ_{UE_Proc}, δ_{UE_Trans} - Processing and Transmission Delay in the UE.
- δ_{RU_Tx}, δ_{RU_Rx} - Accumulated latency in the RU in transmission and reception.
- δ_{RU_Proc}, δ_{RU_Queu}, δ_{RU_Trans} - Processing, Queuing and Transmission Delay in the RU.
- δ_{DU_Tx}, δ_{DU_Rx} - Accumulated latency in the DU in transmission and reception.
- δ_{DU_Proc}, δ_{DU_Queu}, δ_{DU_Trans} - Processing, Queuing and Transmission Delay in the DU.
- δ_{CU_Tx}, δ_{CU_Rx} - Accumulated latency in the CU in transmission and reception.
- δ_{CU_Proc}, δ_{CU_Queu}, δ_{CU_Trans} - Processing, Queuing and Transmission Delay in the CU.
- δ_{CORE_Tx}, δ_{CORE_Rx} - Accumulated latency in the Core in transmission and reception.
- δ_{CORE_Proc}, δ_{CORE_Trans} - Processing and Transmission Delay in the Core.
- δ_{EDC} - Accumulated latency in the External Data Centre.
- δ_{EDC_Proc}, δ_{EDC_Trans} - Processing and Transmission Delay in the External Data Centre.
- δ_{MEC} - Accumulated latency in the MEC node.
- δ_{MEC_Proc}, δ_{MEC_Trans} - Processing and Transmission Delay in the MEC node.

First, the transmission latency in the links, which depends on the amount of data to be transmitted and the data rate of the link, [Carv22]:

$$\delta_{Trans\,[\mathrm{ms}]} = \frac{8\,D_{\,[\mathrm{Bytes}]}}{R_{\,[\mathrm{Gbits/s}]}}\,10^{-6} \tag{3.9}$$

where:

- D – Packet size in bytes.
- R – Data Rate of the link.

Concerning propagation latency, it depends on the type of used medium, which affects the velocity of the signal in the link and the distance between the origin and the destination of the information. One considers having optical fibre as the medium used for the FH, MH, BH and transport links, and only the air link for the link between the UE and the RU. The general expression is the following:

$$\delta_{Prop\,[\mathrm{ms}]} = \frac{d_{[\mathrm{m}]}}{v_{[\mathrm{km/s}]}} \tag{3.10}$$

where:

- d – Distance between source and destination of the signal.
- v – Velocity of the signal in the link.

Regarding processing latency, it consists of the time a node takes to process the packet, for instance, processing the header, check errors and routing the packet. For the calculation of the processing delay in the UE, one needs to know that different number of OFDM symbols are possible depending on the type of slot, and different numerologies are possible as well, with different subcarrier spacings. Scaling up the base subcarrier spacing $\Delta f = 15\ \mathrm{kHz}$ by 2^u (μ = 1, 2, · · ·), the Transmission Time Interval (TTI) duration is scaled down by 2^u, enabling faster transmission and lower processing time [MGEl18].

Table 3.2 shows the processing delay ratios for the UE that depend on the numerology used. This ratio allows for the calculation of the value of the processing delay of the UE.

Table 3.2 - UE processing delay ratios ρ_{UE} (adapted from [MGEl18]).

Subcarrier Spacing (kHz)	15	30	60	120
ρ_{UE}	$\frac{2}{14}$	$\frac{2}{14}$	$\frac{3}{14}$	$\frac{4}{14}$

The processing delay in the UE is given by [Carv22]:

$$\delta_{UE\_Proc\,[\mathrm{ms}]} = \delta_{UE\_Trans}\,\rho_{UE} \tag{3.11}$$

The processing latency in the nodes depends on the chosen splitting option, since the splitting option dictates how many functionalities the node has. For instance, with the option 8 split, RU has 1 of the total 9 functionalities, therefore, one considers the processing latency of the RU with this option to be one ninth of the total processing delay. The same thinking applies for the other splitting options. Moreover, one assumes that all the functionalities of the protocol stack have the same complexity and weight for the calculation of the processing delays.

The expression for the processing delay in the RU is [Carv22]:

$$\delta_{RU\_Proc\,[\mathrm{ms}]} = \delta_{UE\_Trans}\,\rho_{RU}\,\rho_{lat} \tag{3.12}$$

where:

- ρ_{RU} – the ratio of functionalities assigned to the radio unit.
- ρ_{lat} – a parameter that is used to adapt the processing resources to the latency.

Then, the processing delay for the DU is [Carv22]:

$$\delta_{DU\_Proc\,[\mathrm{ms}]} = \delta_{UE\_Trans}\,\rho_{DU}\,\rho_{lat} \tag{3.13}$$

where:

- ρ_{DU} – the ratio of functionalities assigned to the distributed unit.

The processing delay for the CU is given by [Carv22]:

$$\delta_{CU\_Proc\,[\mathrm{ms}]} = \delta_{UE\_Trans}\ \rho_{CU}\ \rho_{lat} \tag{3.14}$$

where:

- ρ_{CU} – the ratio of functionalities assigned to the central unit.

Table 3.3 shows the several processing latency ratios for the RU, DU, and CU.

Table 3.3 – Processing Latency ratios for RU, DU, and CU (extracted from [Carv22]).

	Fronthaul Splitting Option				
	8	**7.3**	**7.2**	**7.1**	**6**
$\boldsymbol{\rho_{RU}}$	1	$\frac{25}{11}$	$\frac{19}{11}$	$\frac{15}{11}$	3
$\boldsymbol{\rho_{DU}}$	6	$\frac{52}{11}$	$\frac{58}{11}$	$\frac{62}{11}$	4
$\boldsymbol{\rho_{CU}}$	2	2	2	2	2

As for the processing latency in the core, the expression is the following [Carv22]:

$$\delta_{CORE\_Proc\,[\mathrm{ms}]} = \frac{4}{2385}\ D_{[\mathrm{bytes}]} + \frac{469}{477} \tag{3.15}$$

Concerning the processing delay in the MEC node and the External Data Centre, it is related to the hardware capabilities of the node, namely, the frequency of the CPU. If the frequency decreases the processing latency increases. The expressions for the delays are [Carv22]:

$$\delta_{MEC\_Proc\,[\mathrm{ms}]} = 4 \cdot 10^{-5} \cdot D_{[\mathrm{bytes}]} \cdot \rho_{func} \tag{3.16}$$

$$\delta_{EDC\_Proc\,[\mathrm{ms}]} = 1.33 \cdot 10^{-5} \cdot D_{[\mathrm{bytes}]} \tag{3.17}$$

where:

- ρ_{func} – number of functionalities that the node needs to execute

At last, the queuing delay is the time it takes when packets wait in a node before being transmitted or processed and account for the latency. There is the concept of priorities where more important services are served first, and this must be considered for the delay of different services. The size of the packets and the throughput of the link after the node affect this delay as well. For this model, one does not assume queuing delay for the core node since it is adapted to congestion.

Therefore, the expression for the queuing delay in a certain node for packets that belong to a specific service with a specific priority [Carv22] is:

$$\delta_{Queu}\,[\text{ms}] \;=\; 10^3 \sum_{p=1}^{M_{users}} \frac{8\, D_{serv,p}[\text{Bytes}]}{R_{max}[\text{bps}]} \tag{3.18}$$

where:

- $D_{serv,p}$ – Packet size in bytes from a specific service with specific priority p.
- R_{max} – Maximum throughput of the link after the node.
- M_{users} – Number of users connected to the node using a service with the same or higher priority than the user under study.

3.2.2 Network Latency

To calculate the E2E Latency one divides the algorithm into segments, namely, the node latency contributions and then the latency propagation contributions depending on the number of nodes and architecture. Since there are no real distances between the nodes in the architecture, one dimensions the propagation latencies and corresponding distances between nodes to guarantee that the minimum latency service requirements are met. Therefore, one assumes initially that the propagation latency is equal to 0 ms in the expressions and obtain the E2E latency, which represents the total node latency. Then, the margin between the E2E maximum required latency for a service and the obtained E2E total node latency is the maximum possible value for the accumulated propagation latency in the network for a specific service. With this value one dimensions the lengths of the links of the architecture. This method is represented in the following expression:

$$\delta_{Max\_Total\_Propagation}[\text{ms}] \;=\; \delta_{service}\,[\text{ms}] \;-\; \delta_{Tot\_Node}[\text{ms}] \tag{3.19}$$

where:

- $\delta_{Max\_Total\_Propagation}$ – Maximum propagation delay to satisfy the latency service's requirement.
- $\delta_{service}$ – Maximum latency requirement for this service.
- δ_{Tot_Node} – Total node latency which is the sum of all delay contributions expect propagation.

The total node latency in the case of no MEC node scenario in the network is described by:

$$\delta_{Tot\_Node}\,[\text{ms}] \;=\; \delta_{UE\_Tx} + \delta_{AL\_Tx} + \delta_{RU\_Tx} + \delta_{DU\_Tx} + \delta_{CU\_Tx} + \delta_{Core\_Tx} + \delta_{EDC} + \delta_{Core\_Rx} + \delta_{CU\_Rx} + \delta_{DU\_Rx} + \delta_{RU\_Rx} + \delta_{AL\_Rx} + \delta_{UE\_Rx} \tag{3.20}$$

For the first scenario with the MEC node located between the CU and the Core (Option CU-Core) the total node latency is:

$$\delta_{Tot\_Node}\,[\text{ms}] \;=\; \delta_{UE\_Tx} + \delta_{AL\_Tx} + \delta_{RU\_Tx} + \delta_{DU\_Tx} + \delta_{CU\_Tx} + \delta_{MEC} + \delta_{CU\_Rx} + \delta_{DU\_Rx} + \delta_{RU\_Rx} + \delta_{AL\_Rx} + \delta_{UE\_Rx} \tag{3.21}$$

The second scenario consists of having the MEC node installed between the DU and the CU (Option DU-CU). The total node latency is described by:

$$\delta_{Tot\_Node}\,[\text{ms}] \;=\; \delta_{UE\_Tx} + \delta_{AL\_Tx} + \delta_{RU\_Tx} + \delta_{DU\_Tx} + \delta_{MEC} + \delta_{DU\_Rx} + \delta_{RU\_Rx} + \delta_{AL\_Rx} + \delta_{UE\_Rx} \tag{3.22}$$

Finally, the third possible scenario consists of installing the MEC node between the RU and the DU

(Option RU-DU) which results in the following total node latency:

$$\delta_{Tot\_Node}\,{\scriptstyle [\mathrm{ms}]} = \delta_{UE\_Tx} + \delta_{AL\_Tx} + \delta_{RU\_Tx} + \delta_{MEC} + \delta_{RU\_Rx} + \delta_{AL\_Rx} + \delta_{UE\_Rx} \tag{3.23}$$

The use of the MEC node influences both the total node latency in the network and the total propagation latency, since it can replace the links and the nodes after it, while the chosen architecture (collocations of nodes for example) influences only the propagation delay by eliminating the distance between certain nodes. If two nodes are collocated the distance between them is 0.

The E2E latency is given by:

$$\delta_{E2E}\,{\scriptstyle [\mathrm{ms}]} = \delta_{Tot\_Node}\,{\scriptstyle [\mathrm{ms}]} + \delta_{Prop}\,{\scriptstyle [\mathrm{ms}]} \tag{3.24}$$

In order to obtain the maximum E2E distance that satisfies a specific service latency requirement with a certain architecture, one first calculates the maximum propagation delay:

$$\delta_{Max\_Total\_Propagation}{\scriptstyle [\mathrm{ms}]} = \frac{d\_max\,{\scriptstyle [\mathrm{m}]}}{v\,{\scriptstyle [\mathrm{km/s}]}} \tag{3.25}$$

One solves (3.25) in order to d_max and replaces $\delta_{Max\_Total\_Propagation}$ with (3.19) resulting in the following expression for the maximum E2E distance:

$$d_{E2E\_max}\,{\scriptstyle [\mathrm{m}]} = (\delta_{service}\,{\scriptstyle [\mathrm{ms}]} - \delta_{Tot\_Node}\,{\scriptstyle [\mathrm{ms}]})\, v\,{\scriptstyle [\mathrm{km/s}]} \tag{3.26}$$

Since optical fibres are not typically installed in a straight line, the factor 1.67 is added to compensate, [Carv22]. The expression with these changes is:

$$d_{E2E\_max}\,{\scriptstyle [\mathrm{km}]} = (\delta_{service}\,{\scriptstyle [\mathrm{ms}]} - \delta_{Tot\_Node}\,{\scriptstyle [\mathrm{ms}]})\, \frac{v\,{\scriptstyle [\mathrm{km/s}]}}{2 \times 1.67} 10^{-3} \tag{3.27}$$

The E2E distance is composed of several segments of the network and is given by:

$$d_{E2E\_max}\,{\scriptstyle [\mathrm{km}]} = d_{FH\_Tx} + d_{MH\_Tx} + d_{BH\_Tx} + d_{TL\_Tx} + d_{TL\_Rx} + d_{BH\_Rx} + d_{MH\_Rx} + d_{FH\_Rx} \tag{3.28}$$

where:

- $d_{FH\_Tx}\,/d_{FH\_Rx}$ - Fronthaul link distances.
- $d_{MH\_Tx}\,/d_{MH\_Rx}$ - Middlehaul link distances.
- $d_{BH\_Tx}\,/d_{BH\_Rx}$ - Backhaul link distances.
- $d_{TL\_Tx}\,/d_{TL\_Rx}$ - Transport link distances.

One can then know the maximum distance between the UE and the EDC or MEC by dividing the E2E Distance by 2:

$$d_{max}\,{\scriptstyle [\mathrm{km}]} = \frac{d_{E2E\_max}\,{\scriptstyle [\mathrm{km}]}}{2} \tag{3.29}$$

where:

- d_{max} - Maximum Distance between UE and EDC/MEC to satisfy service requirements.

One also calculates the latency critical threshold, for which the total latency of the service should not be higher for safety reasons. This threshold is given by:

$$\delta_{critical\_threshold}[\mathrm{ms}] = M \,.\, \delta_{service}\,[\mathrm{ms}] \tag{3.30}$$

where:

- $\delta_{critical\_threshold}$ – Latency critical threshold.
- M – Margin of Latency, value between 0 and 1. Lower values represent stricter policies.

Afterwards, one can identify the accepted private networks deployments that satisfy the maximum distance requirement. For this, one needs the maximum distance between a UE in the subway system and the core of the railway operator $d_{ue-RailCore}$ and the maximum distance between the UE and core of the CSP $d_{ue-CSPCore}$. One considers the EDC node to be collocated with the core since this is common for the studied services and that $d_{ue-RailCore} < d_{ue-CSPCore}$, since the core of the railway operator is normally located in the subway system premises and that the core of the CSP can be located far away from the subway system area. If $d_{max} < d_{ue-RailCore}$, a MEC node is needed such that the distance between the UE and the MEC is lower than d_{max}. If $d_{ue-RailCore} < d_{max} < d_{ue-CSPCore}$, then one would choose the isolated private network or shared private network, both using the core of the railway operator. Lastly, if $d_{max} > d_{ue-CSPCore}$, one could choose all types of private networks (Isolated, Shared or Network Slice) using just one or both available cores.

3.3 Link Throughputs and Capacities

According to 3GPP TS 38.306, the theoretical maximum throughput, or data rate that NR offers for a given number of aggregated carriers is computed as follows [3GPP22]:

$$R_{[Mbits/s]} = 10^{-6} \cdot \sum_{j=1}^{J}\left(v_{layers}^{(j)} \cdot Q_m^{(j)} \cdot f^{(j)} \cdot R_{codemax} \cdot \frac{N_{PRB}^{BW\,(j),\mu} \cdot 12}{T_s^{\mu}} \cdot \left(1 - O^{(j)}\right)\right) \tag{3.31}$$

where:

- J is the number of aggregated component carriers and j is the j-th carrier.
- $v_{layers}^{(j)}$ – number of layers of the transmitter that depend on the MIMO system (maximum of 8 for the DL and 4 for the UL).
- $Q_m^{(j)}$ – Maximum modulation order (2 for QPSK, 4 for 16-QAM, 6 for 64-QAM, 8 for 256-QAM).
- $f^{(j)}$ – Scaling factor which is related with the MIMO layers and modulation order (can take values 1, 0.8, 0.75 and 0.4).
- $R_{codemax}$ – Constant defined by 3GPP which is related with the modulation order (the maximum takes the value 0.926). The possible values can be seen in Annex B.
- $N_{PRB}^{BW\,(j),\mu}$ – Maximum Resource Block allocation in bandwidth $BW\,(j)$ with numerology μ. The possible values are represented in Table 3.4.
- T_s^{μ} – Average OFDM symbol duration for a numerology μ, i.e., $T_s^{\mu} = \frac{10^{-3}}{14 \cdot 2^{\mu}}$.
- $O^{(j)}$- Overhead for the control channels (0.14 for DL and 0.08 for UL).

The values for the number of RBs, $N_{PRB}^{BW\,(j),\mu}$ are represented in Table 3.4. Instead of the numerology, the Subcarrier Spacing is shown.

Table 3.4 – Number of RBs for a certain bandwidth and numerology (extracted from [Carv22]).

Number of RBs														
SCS [kHz]	**Bandwidth [MHz]**													
	5	**10**	**15**	**20**	**25**	**30**	**40**	**50**	**60**	**80**	**90**	**100**	**200**	**400**
15	25	52	79	106	133	160	216	270	-	-	-	-	-	-
30	11	24	38	51	65	78	106	133	162	217	245	273	-	-
60	-	11	18	24	31	38	51	66	79	107	121	132	264	-
120	-	-	-	-	-	-	-	32	-	-	-	66	132	264

The general expression (3.31) assumes the total bandwidth for a single transmission, either DL or UL. For a TDD band, one must consider the slot structure and divide the total bandwidth by the UL and DL transmissions. Therefore, the expression is adapted to TDD mode with the data rate of UL and DL depending on the format used for the slot. The expressions for the data rate of DL and UL are:

$$R_{TDD/DL\,[Mbits/s]} = R_{[Mbits/s]}\ F_{DL}\ A_f \tag{3.32}$$

$$R_{TDD/UL\,[Mbits/s]} = R_{[Mbits/s]}\ F_{UL}\ A_f \tag{3.33}$$

where:

- $F_{DL}/\,F_{UL}$ – The fraction of the slot that belongs to the DL/UL (frame structure).
- A_f – Factor that accounts for the losses in the system. This factor is important to simulate real scenarios in which the link conditions are not optimal.

In order to compare if the data rate provided is enough for the scenarios and different type of services one needs to calculate the required data rate for the services and users. The calculation consists of summing the data rates required of the total users depending on which service they are using (this formula applies for both required capacity in the UL and DL direction):

$$R_{required\,[Mbits/s]} = \sum_{1}^{Nu} R_{s\,[Mbits/s]} \tag{3.34}$$

where:

- Nu – Number of users connected to the RU.
- R_s – Data Rate of the service for that user.

It is important to note that the data rate requirements of the different services can be UL, DL or both depending on the specific service. If the required capacity is higher than the provided, the considered architecture is discarded as well as the considered private network. This total required capacity comprises the passenger services and the railway services required capacity:

$$R_{required\_railway\,[Mbits/s]} = \sum_{1}^{Nu\_rail} R_{s\_rail\,[Mbits/s]} \tag{3.35}$$

$$R_{required\_passengers\,[Mbits/s]} = \sum_{1}^{Nu\_passengers} R_{s\_passengers\,[Mbits/s]} \tag{3.36}$$

where:

- Nu_rail – Number of terminals using railway services connected to the RU.
- $Nu\_passengers$ – Number of terminals using passenger services connected to the RU.
- R_{s_rail} – Data Rate of the railway service for that terminal.
- $R_{s\_passengers}$ – Data Rate of the passenger service for that terminal.

Moreover, one calculates the ratio of capacity, with the following equation:

$$\eta_r = \frac{R_{required\,[Mbits/s]}}{R_{provided\,[Mbits/s]}} \tag{3.37}$$

where:

- η_r – Margin of capacity. If lower than 1, provided capacity is enough and if greater than 1, is not enough relative to the required by the users.
- $R_{required\,[Mbits/s]}$ – Throughput required by the total of users.
- $R_{provided\,[Mbits/s]}$ – Throughput provided by the BS.

Having calculated the data rate provided by the BS to the UE in DL and UL and the required throughput, one needs to calculate the throughputs of the rest of the nodes and links of the network, namely, the FH, MH, BH and TL. To calculate the throughputs of the FH and MH, the splitting options must be considered as it corresponds to more or less processing being done in certain nodes and more or less throughput required in the links consequently. It is important to note that option 7.2 is the first to be analysed, since it is the one that is more commonly used in NR systems. The other splitting options to be considered are the 8, 7.2, 7.3, 6 for the FH and 2 for the MH. Annex C shows the link throughputs and the used values for the following described expressions according to the splitting option considered.

Firstly, the Option 8 (RF/PHY) is the one that is used in 4G as the traditional split between the radio unit and the baseband unit (BBU). This option has the highest and constant data rate on the FH as the processing is mostly centralised and scales with the number of antennas. The throughput of the RU in the FH in both UL and DL is expressed by [DOCO16]:

$$R_{FH\_8\,[Mbits/s]} = S_{[megasymbols/s]} \cdot B_{[bits/symbol]} \cdot A \cdot 5 \tag{3.38}$$

where:

- S – Sampling Rate in samples per second.
- B – Bitwidth.
- A – Number of Antenna Ports.

Option 7.1 (low PHY), compared to Option 8, brings a drop of the fronthaul bitrate however it is still constant, and it scales with the number of antennas as well. It should be used when there is high fibre capacity between the RU and DU/CU. The bit rate in the DL and UL is given by [DOCO16]:

$$R_{FH\_7.1\,[Mbits/s]} = 2000 \cdot N_{SC} \cdot N_{symb} \cdot A \cdot B_{[bits/symbol]} + M_{info\,[Mbits/s]} \tag{3.39}$$

where:

- N_{SC} – Number of Subcarriers.
- N_{symb} – Number of Symbols.
- M_{info} – Information in the MAC header.

Option 7.2 (Low PHY/High PHY) is the main option to analyse, has a more complex DU than Option 7.1 but less shared processing in the CU. The bit rate in the FH in the DL and UL is described by the following expression [DOCO16]:

$$R_{FH\_7.2\,[Mbits/s]} = 2000 \cdot N_{SC} \cdot N_{symb} \cdot N_{layers} \cdot B_{[bits/symbol]} + M_{info\,[Mbits/s]} \tag{3.40}$$

where:

- N_{layers} – Number of Layers in the system.

In Option 7.3 (High PHY) the scrambling and modulation functions are in the DU, which reduces the fronthaul throughput as the bits are assigned to symbols. The data rate that is generated in the fronthaul for UL and DL is given by [Carv22]:

$$R_{FH\_7.3\,[Mbits/s]} = \left(2000 \cdot N_{SC} \cdot N_{symb} \cdot N_{layers} \cdot B_{[bits/symbol]}\right) \cdot \mu_s + M_{info\,[Mbits/s]} \tag{3.41}$$

where:

- μ_s – Subcarrier Load.

Option 6 (MAC-PHY) has a significant bandwidth reduction compared to option 7 however the increased delay may affect some processes of the BS like HARQ timing and scheduling. This split separates the link layer from the physical layer. According to [LaCC19] the resulting CU would only include the link layer and the network layer functions, which represents 20% of the baseband total processing. The throughput in the UL and DL that is generated in the fronthaul is given by the expression [Carv22]:

$$R_{FH\_6\,[Mbits/s]} = \left(R_{p\,[Mbps]} + R_{c\,[Mbps]}\right)\left(\frac{B_{[MHz]}}{B_{c\,[MHz]}}\right)\left(\frac{N_{layers}}{N_{layers},c}\right)\left(\frac{\log_2 M_{mod}}{\log_2 M_c}\right) \tag{3.42}$$

where:

- R_p – System Peak Rate.
- R_c – Control and Signalling Rate.
- B – System Bandwidth.
- B_c – Control and Signalling Bandwidth.
- N_{layers}, c – Layers for control and signalling.
- M_{mod} – Order of Modulation.
- M_c – Modulation Order for Control Signals.

As for the Midhaul, one considers Option 2 (RLC-PDCP). In option 2 one has a separation of the User plane (UP) and Control Plane (CP) and is suited for scenarios where a low fronthaul bitrate is necessary which could be a wireless link. The Radio Link Control (RLC) layer and the Packet Data Convergence Protocol (PDCP) layer are centralised in the CU while the other functions are executed in the DU. The throughput that is generated in the MH link in the UL and DL is given by [Carv22]:

$$R_{MH\_2\ [Mbits/s]} = \left(R_{p\ [Mbps]}\right)\left(\frac{B_{[MHz]}}{B_{c\ [MHz]}}\right)\left(\frac{N_{layers}}{N_{layers,c}}\right)\left(\frac{\log_2 M}{\log_2 M_c}\right) + R_{c\ [Mbps]} \tag{3.43}$$

Finally, for the Backhaul and the Transport Network one assumes fixed values that are typical in different solutions and are based on [ITUT18]. For the BH link (between the CU and the Core) one has typical throughput values of 10 Gbps to 25 Gbps and for the Transport Link between the Core and the External Data Centres one has throughputs of 100 Gbps or more.

3.4 Model Implementation

The model was implemented in MATLAB version R2022a, where there is a main sub-program where all variables are stored, and the output parameters are produced. There are other sub-programs as well that are functions that calculate intermediary values that are needed to achieve the output parameters.

In the following model flowcharts (capacity, latency and general) there are processes represented by boxes with different colours. The boxes with light orange colour correspond to processes that were based and extracted from [Carv22], as mentioned in the model description. The boxes with light purple colour correspond to processes that were adapted from [Carv22] and all the rest is work from this thesis.

The model flowchart for the capacity part is shown in Figure 3.4.

The goal of the capacity part is to check if a certain type of private network can satisfy the capacity requirements of all the services. Thus, it starts by calculating the provided capacity by the RU in the DL and UL. To do this, one needs to calculate the number of resource blocks that depend on the numerology and the bandwidth. Then, one determines $R_{codemax}$ and the modulation order via the channel quality indicator input. The overhead factor is chosen according to the user service. After calculating the provided throughput by the RU, one calculates the throughput for the DL and the UL using the frame structure parameter.

Then, there is the calculation of the total required capacity according to the RU service mix and the number of terminals assumed in the scenario. One calculates the required capacity in the DL and UL that depends on the service link type. Then, the calculation of the required capacity for the different categories of services, the passenger and railway in both UL and DL. Finally, one checks if the provided capacity is enough compared to the required one for all types of services. If it is, the margin is calculated and the private network type is accepted and if not, one produces the results and discards this type of private network.

The latency model aims to verify which types of architectures and types of private networks satisfy the requirements of the services in terms of latency. The model flowchart for the latency part is shown in Figure 3.5.

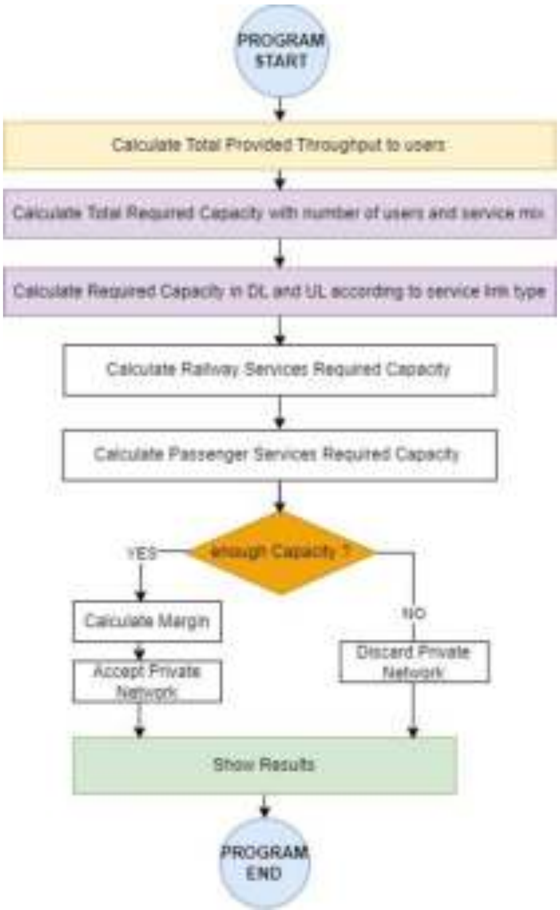


Figure 3.4 – Flowchart of Capacity Model.

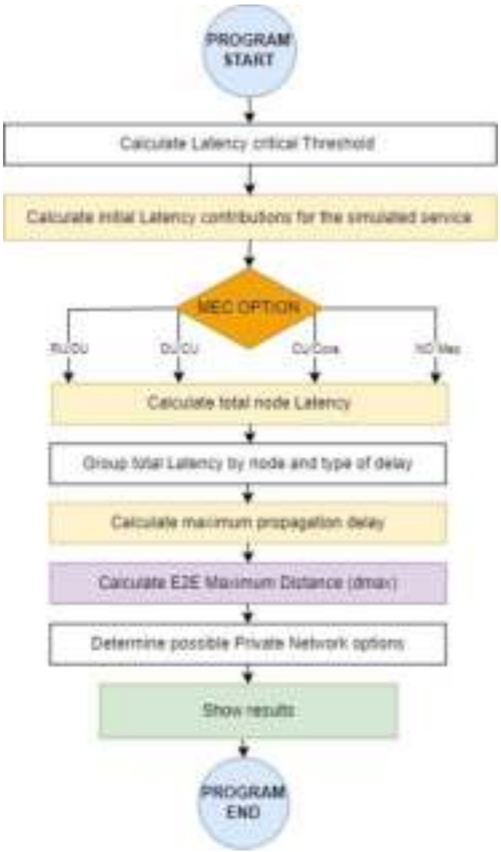


Figure 3.5 – Flowchart of Latency Model.

The first step is to calculate the latency critical threshold of the service. Then, one computes the initial latency contributions for the service that is being simulated, namely, the UE delays and the air link delays which are common to every architecture. The second step is conditioned by the MEC deployment option in order to calculate the total node latency, which is composed by all delay contributions from the nodes considered in the architecture.

One computes the processing, queueing and transmission delay of each node that applies, for the transmission and reception or both depending on the service itself. The next process is to group the total node latency by node and type of delay to better understand and analyse the several latency contributions. After, the calculation of the maximum propagation delay is done by subtracting the total node latency obtained, from the service latency requirement. With it, one can calculate the E2E Maximum Distance and in finally determine the accepted private networks for each service.

The general model flowchart is shown in Figure 3.6.

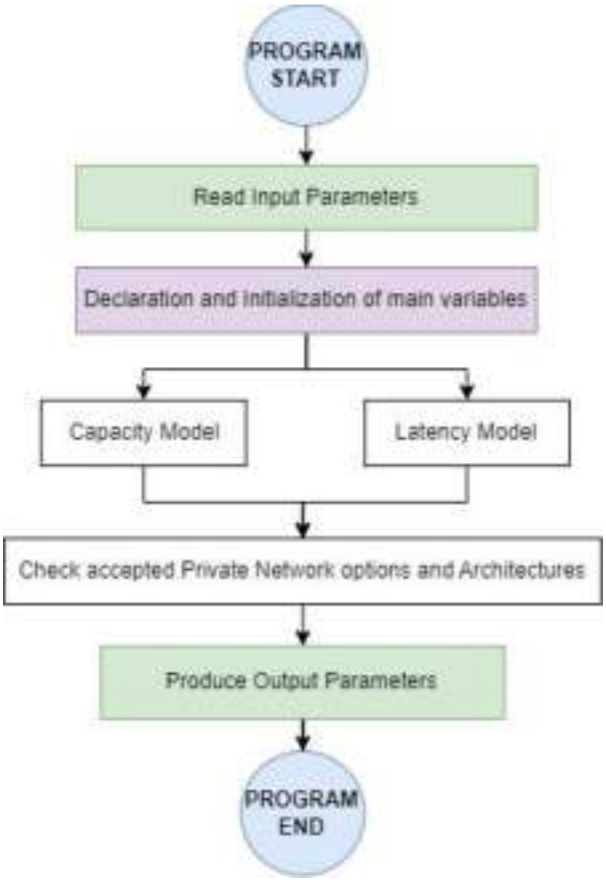


Figure 3.6 – Flowchart of General Model.

The general model encompasses both the previous sub-models of capacity and latency and joins them to produce the final output results. It starts by reading all the input parameters necessary that were mentioned in Section 3.1, parses the data from the Excel and does the declaration and initialisation of the main variables. The main program then, splits into both sub-programs of capacity and latency sections and in the end, after both sub-programs ended, does the final determination of the accepted private network options and architectures that satisfy all the requirements of the services. It then produces the final output results.

3.5 Model Assessment

One crucial part before applying the model, is to assess and validate it. This part focuses on validating the model by executing some tests and comparing it with the theoretical values and expectations. This process covers several steps from the checking of the input parameters, the intermediate steps and code blocks and finally the output results.

The first block of tests includes the initial phases of the simulator, namely the processing of the input files and variables instantiation. Moreover, it comprises the calculation of the provided and required throughput which corresponds to the capacity sub-model.

The sequence of tests that were performed for the throughput section are illustrated in Table 3.5.

Table 3.5 – Tests performed to validate the input parameters and capacity section model.

Test ID	Description
1	Validate the reading of the input files, the correct storage in created variables with the correct value and type of data.
2	Validate the number of resource blocks available for different scenarios.
3	Validate the correct definition of the modulation order to be used according to the Channel Quality Indicator (CQI) Index.
4	Validate the correct calculation of the provided throughput.
4.1	Verify if the throughput increases with the increase of MIMO layers.
4.2	Validate the correct calculation of the UL and DL provided throughput with the frame structure parameter.
5	Validate the computing of the required throughput for passenger and railway services. Validate the correct calculation of the capacity ratio.
6	Verify if private network option is correct if enough capacity is provided or not.

In Test 4.1, one checks if the increase of the MIMO layers result in an increase of the provided throughput. To check this, one varies the number of MIMO layers that can take up values of 2,4 and 8 and calculates the throughput. The rest of the parameters considered for the test are illustrated in Table 3.6.

Table 3.6 – Input Parameters for testing the provided throughput calculation.

Modulation	Scaling Factor	Rmax	Bandwidth [MHz]	Numerology	Overhead
256QAM	1	0.9260	100	1	0.14

The last phase of tests and validations regard the latency sub-model that consists of calculation of the E2E total node Latency, the different delays and contributions, the maximum E2E Distance to satisfy a service requirement and the accepted private network type. The several tests for this section can be seen in Table 3.7.

Table 3.7 – Tests performed to validate the latency section model.

Test ID	Description
7	Verify the correct calculation of the Latency critical Threshold.
8	Validate the calculation of the E2E Latency and the maximum E2E Distance.
8.1	Validate the latency contributions common to every scenario.
8.2	Verify if the Air Link Delay increases with the distance.
8.3	Validate the calculation of the transmission, processing, and queuing delays.
8.4	Verify the total node latency with and without the MEC node.
8.5	Check the correct calculation of the maximum propagation delay and maximum E2E distance for each scenario.
9	Check if choice of private network type is correct according to the conditions.

In this last phase, the delays that are common to all scenarios of deployment were tested, namely, the UE Transmission time and Processing time in both Transmitter and Receiver. The Air Link Delay was tested as well for different distances between the UE and the RU, and it increased in a linear shape as expected being approximately 1μs with 300 metres and 3μs with 900 metres.

The queuing delay and processing delay in the different nodes were validated as well, the first increasing with the number of users in the system and the second increasing when the node had more functionalities according to the splitting option and a higher packet size. The use of MEC node in the different possible options was tested as well, resulting in lower latencies when placing the MEC node closer to the UE replacing this way all the nodes after it.

The E2E total node latency was checked if resulted in the sum of all contributions expect the propagation delay with and without the existence of the MEC node. One verified the correct calculation of the maximum propagation delay and the maximum E2E Distance according to the formulas and obtained acceptable results when comparing with the work of [Carv22]. Finally, the determination of the types of private networks for different services was tested according to the conditions of the distances of the UE to the core of the CSP and railway operator.

Chapter 4

Results Analysis

This chapter provides the description of the studied scenario as well as the analysis of the results obtained from the developed model.

4.1 Scenario Description

The scenario under study is the subway system of Lisbon, “Metropolitano de Lisboa”. In this metro system, there are 4 lines, a total of 56 subway stations with almost 45 km of extension and 333 total carriages. Leaky cables can be used throughout the tunnels for 5G coverage and technical rooms installed at every station with node installation.

Figure 4.1 shows the overall diagram of the Lisbon metro with all the stations and the four lines.

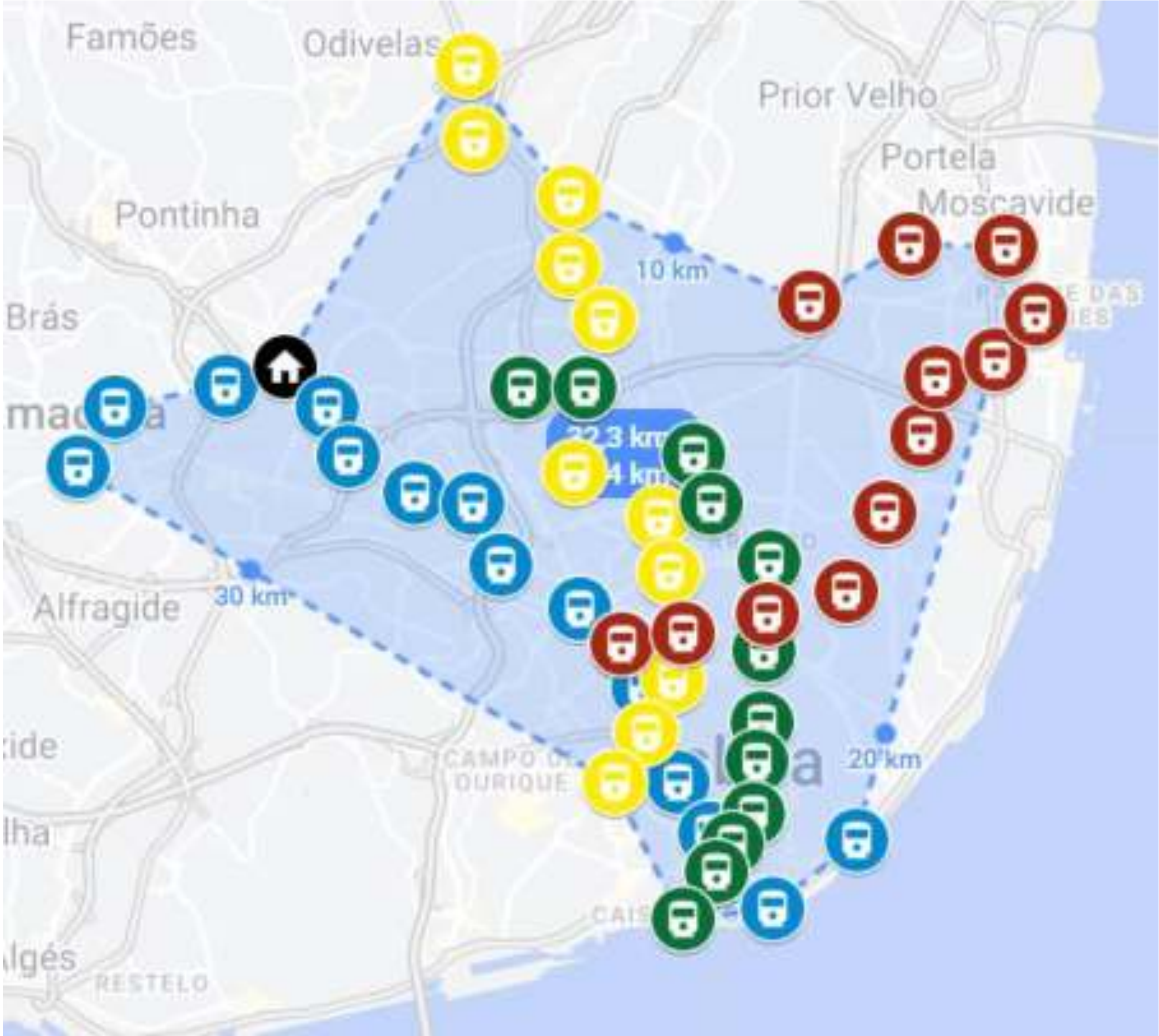


Figure 4.1 – Diagram of Lisbon Metro.

One can see all 56 stations, the red, green, yellow, and blue lines represented by the coloured stations. This diagram also shows the coverage area of the metro system, the light blue area, which accounts for approximately 50 km^2. The total area of the city of Lisbon is around 100 km^2. The dark station with the house symbol represents the location of the core of the metro network, in Pontinha. Besides, one should note that this subway system is mainly underground with some sections at the surface.

The distance between stations is an important parameter for this work, namely, in the calculation of the delay between the UE and the RU that is located at the nearest metro station. Therefore, resorting to

Google Maps one estimated the distances between every station and in the end, calculated the average distance, which is the value taken in account for simulations. The distances between every station and the method that was used can be seen in Annex A. Table 4.1 shows the final calculated average distance between stations, and their maximum and minimum.

Table 4.1 – Minimum, Average and Maximum Distance between Lisbon metro stations.

Minimum [m]	**Average [m]**	**Maximum [m]**
300	780	1400

One can see that the average distance equals 780 m, the maximum distance of 1 400 m, which corresponds to the connection of Campo Grande to Alvalade, and the minimum distance of 300 m associated with the connection between Martim Moniz and Rossio. It is worth noting that there is an error associated with the determination of the distances using Google Maps, which in this case, equals to 4 km of subway line extension: this error comes from summing all distances determined with Google Maps and comparing to the subway line extension, which is public information from Metropolitano de Lisboa (which is 44.5 km), but it is negligible, having low impact on the results.

Table 4.2 depicts some characteristics of the Lisbon metro that are useful for this work.

Table 4.2 – Characteristics of Metropolitano de Lisboa.

Number of lines	4
Number of stations	56
Number of carriages	333
Number of carriages per train	6
Maximum Capacity per carriage	165
Average number of people per train	300
CCTV cameras per carriage	2
CCTV cameras per train	12
Critical voice terminals per train	2
Geographical area of Metro [km^2]	50

One assumed a subway train with 6 carriages, a maximum speed of 72 km/h, 2 CCTV cameras per carriage, 1 central server that receives the Passenger Information Data and replicates it into screens on each carriage, 2 critical voice terminals for redundance per train, and an average of 50 passengers per carriage. A carriage (depending on the model) has an average capacity of 38 seated and 127 passengers standing, which totals a maximum number of people of 165 per carriage.

As for services, as described in the previous chapter, they are characterised by their requirements, such as maximum E2E Latency and Data Rate, and are classified as railway specific, such as signalling, Mission Critical Voice Services and CCTV, and passenger specific like PIS and Passenger Wi-Fi.

The reference radio parameters used in simulations are described in Table 4.3, namely, MIMO layers, numerology, Channel Quality Indicator, Scaling factor, Frame Structure ratio, and Average Factor.

Table 4.3 – Radio Reference Values in Subway scenario.

MIMO layers	Numerology	Average Factor	DL Frame Structure	Scaling Factor	CQI	Modulation
4	1	1	0.85	1	12	256QAM

Regarding the available bandwidth, it depends on the type of private network that is being considered. The accessible spectrum changes for the several options. Table 4.4 shows the available bandwidth according to the studied private network.

Table 4.4 – Available Bandwidth per private network.

Private Network	Available Spectrum [MHz]
Slice	100
Standalone / Isolated	10
Shared	100 + 10

The number of users connected to an RU in a certain point of the subway or at a certain moment depends on several factors, such as number of passengers that are in a certain carriage or a certain metro, time of the day, day of the week or if there is a specific event happening in the area flooding the train with passengers. One considered the average number of passengers per train in the simulation.

The percentage of users that are using a specific service among the total services available and the service mix, has an influence on the total data rate required by the RU as a higher percentage of users using services that require more data rate, require a higher total data rate and the same for the reverse where a lower percentage of users using services with high throughput requirements result in a lower required total throughput.

For the scenario under analysis, the mix in Table 4.5 was defined for the 5 total services, signalling systems, voice applications, CCTV, Passenger Information System and Passenger Connectivity.

Table 4.5 – Service Mix in Subway scenario.

Signalling [%]	Voice MCPTT [%]	CCTV [%]	PIS [%]	Passenger Wi-Fi [%]
1	2	5	1	91

This percentages were determined according to the expected number of terminals that are using a specific application in relation to all the other applications.

Certain services can send data in a unique way or in both ways, which is important to determine in which link way (UL, DL, or both) services data flows. Therefore, as mentioned in the previous chapter, each service was given a parameter 0, 1 or 2 to know which link service uses: 0 for DL, 1 for UL and 2 if the service uses both ways. Table 4.6 shows the link types of each service.

Table 4.6 – Link Types for each service.

Signalling	Voice MCPTT	CCTV	PIS	Passenger Wi-Fi
2 (UL/DL)	2 (UL/DL)	1 (UL)	0 (DL)	0 (DL)

The distance between UE and RU considered for simulation, accounting for the propagation over the air, is 390 m, which is half the average distance between stations. The reasoning for this option can be seen in Figure 4.2.

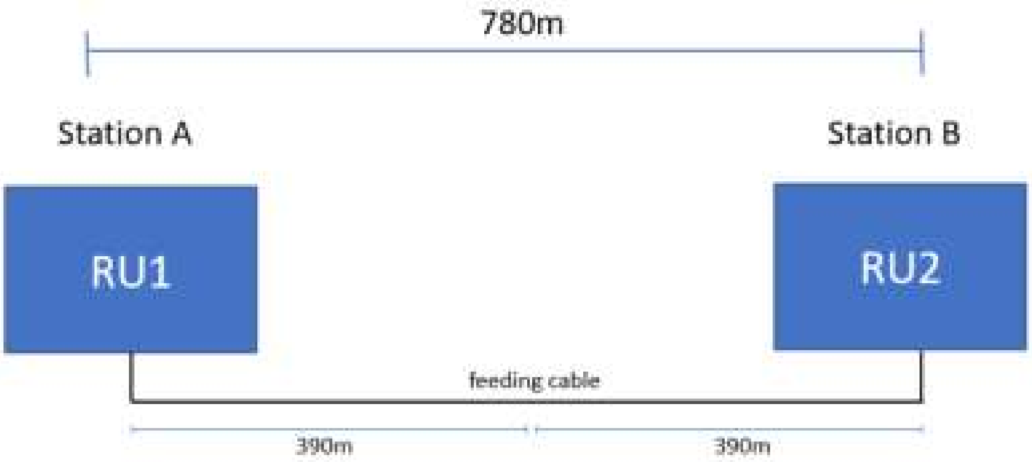


Figure 4.2 – Scheme with distances to RU in the subway line in Lisbon Metro system.

The distance considered for simulations, 390 m, is the maximum distance for a UE in relation to the closest RU and closest subway station. In the middle of the subway line, there is a “break” of the feeding cable, with a new one starting belonging to the next station. This is the handover break point for UEs.

Concerning the total number of users in the subway system in a certain moment and the number of users per node, one assumed a value of 300 users per RU (according to Table 3.1) and a total of 56 RUs distributed by the 56 total Lisbon subway stations, which results in a total average number of passengers of 16 800 users at a certain moment. One also assumed a type of hierarchy where each DU can support 4 RUs and each CU can support 7 DUs, therefore resulting in 1 200 passengers per DU and 8 400 passengers per CU with a total of 56 RUs, 14 DUs and 2 CUs. This is one of many possible architectures for the distribution of the nodes, which then has impact on the queuing latency.

The model considers the base architecture with independent RU, DU, and CU nodes for a certain service E2E connection. However, if one considered the co-location of certain nodes, the distance between them would be 0 and then the resulting maximum distance would allow for a more flexible implementation and location of the nodes.

The goal of the next sections is to present output results according to input parameters. These inputs are also varied in order to simulate different practical scenarios of usage and discover the impact on the end results. Input parameters were varied as follows:

- Number of Users – The total number of users and consequently number of users per node was varied. Increases of passengers in the trains and the possibility of connecting 2 trains to the same RU is seen with this variation.
- Splitting Option – The splitting option was varied to observe the impact on the total node latency and network link throughputs.
- CQI – The CQI has variations due to the channel conditions. This value was varied to simulate poor radio channel conditions in the subway.
- MIMO Layers – The number of MIMO layers was varied to obtain better throughputs.
- Frame Structure – The percentage of bandwidth allocated to UL and DL was varied according to different scenarios.
- Link Type – The link type of the services was varied which as well changes the required capacity of a specific service in a specific link.

4.2 Throughput Analysis

4.2.1 Private Network – Network Slice Analysis

One is assuming, in this first instance, the full use of the operator's spectrum for all services in a slice type of private network, where all system components are owned by the operator with an SLA defined between the mobile network operator and the railway operator. Therefore, the bandwidth used for this part of the simulation is 100 MHz.

Figure 4.3 illustrates the throughput results in the UL and the DL.

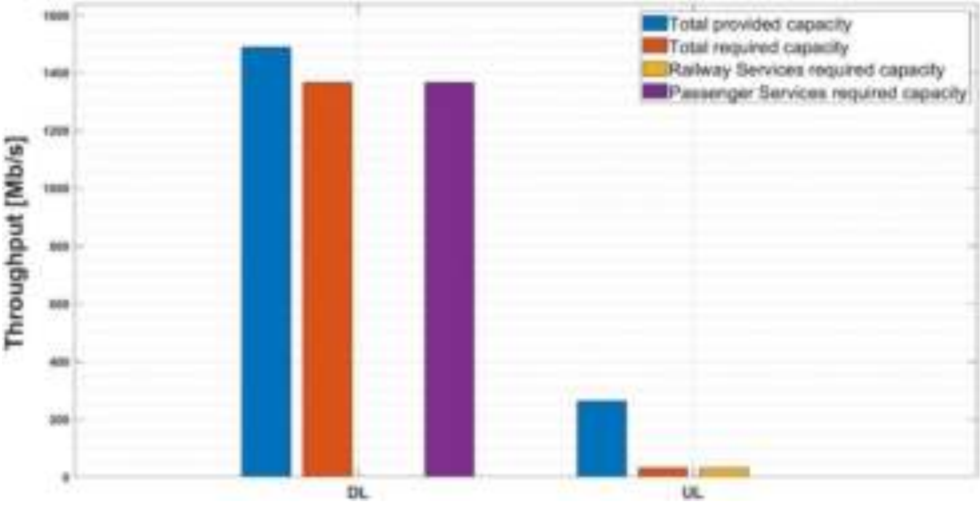


Figure 4.3 – Required and Provided Capacity with network slice private network.

One obtained a total throughput of 1.752 Gbps provided by the RU in both the UL and DL. This results in a total available throughput of 1.489 Gbps in DL and 263 Mbps in UL, according to the frame structure.

In DL, there is a total required capacity of 1.367 Gbps for all users that are using the services and applications, in which 1 366 Gbps corresponds to Passenger Services and 0.45 Mbps to railway specific services. In this link, the railway services require a very low capacity compared with the passenger services. It is, therefore, possible to provide the required throughput for these scenarios in the DL, with a margin of 122 Mbps and a capacity ratio of 0.92, a fine ratio if some of the input factors fluctuate, such as the number of passengers.

In UL, one has a total required capacity of 30 Mbps, which can be provided by the available 263 Mbps. All throughput is required by railway services and applications, which does not exactly correspond to reality, since one assumes passenger Wi-Fi to be a DL main service and there may be some data sent in UL by this service. However, one has a high margin of 232 Mbps and a capacity ratio of 0.12 for possible variations of the required UL throughput.

Assuming a scenario where, instead of 300 total passengers per train, one has 400 (an increase of 100 passengers), requirements can no longer be satisfied, and the provided capacity is not enough to guarantee the required capacity in DL, Figure 4.4. This increase would mean an average of 17 extra passengers per carriage, which can very well happen in busy peak hours. The required throughput would be 1.823 Gbps in DL, an increase of 33% over the original required throughput, which in the UL the provided capacity is still enough in relation to the required one of 40.6 Mbps.

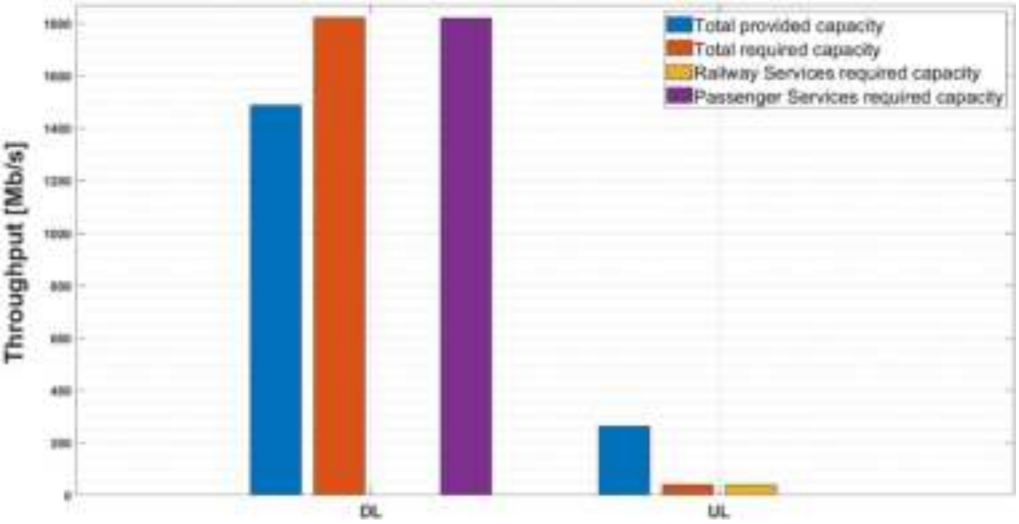


Figure 4.4 – Required and Provided Capacity with an increase of 100 passengers.

The maximum increase of passengers that would still satisfy the requirements of services would be 27 passengers per train and approximately 4 new passengers per carriage in this scenario, making a total maximum of 327 passengers to still satisfy requirements. For this reason, one cannot satisfy the required capacity if 2 trains were connected to the same RU, doubling the number of users to 600. This situation would happen when 2 metro trains would be circulating in different directions but side by side in the line, and therefore connected to the same RU. In optimal conditions, however, the provided capacity would be enough for the scenario with 2 trains connected to the same RU in both UL and DL, Figure 4.5.

In this scenario, one would have the maximum MIMO layers of 8 and the best possible CQI with a value of 15 and consequently the highest order of modulation. In other words, one would have the best radio conditions. The total provided capacity would be 3.974 Gbps in DL and 701 Mbps in UL making a total of 4.675 Gbps available at the RU. The required capacity, in this case, with 2 carriages, would be 2.734 Gbps in DL and 61 Mbps in UL.

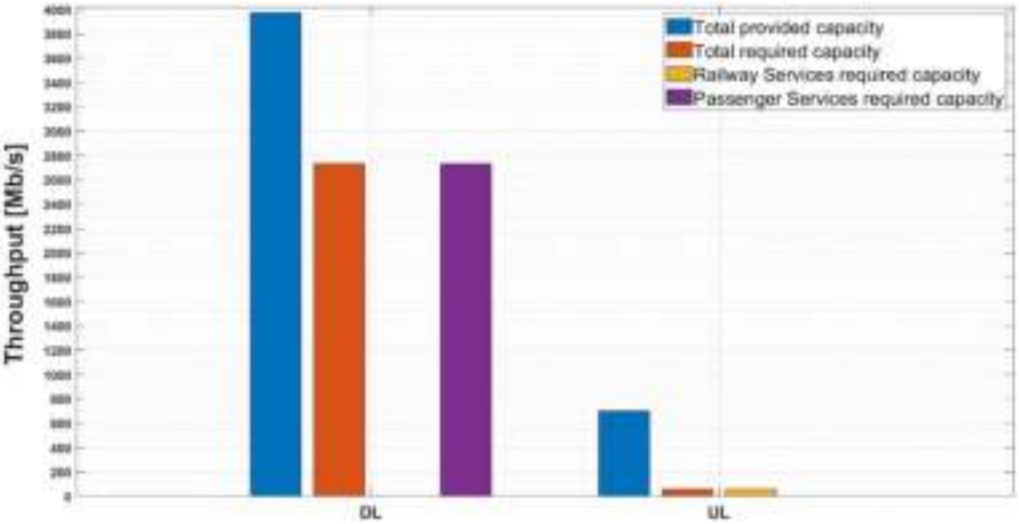


Figure 4.5 – Required and Provided Capacity for two metros with optimal conditions.

It is important to note although, that, increasing the MIMO layers to 8 in the context of subway infrastructure with the leaky cables would be highly unfeasible, due to the increased number of cables to install, which are separated between each other in the upper wall of the subway tunnels, but still, a possibility to consider in theory. One could also analyse capacity results if Passenger Wi-Fi service would be considered a service that sends data equally in DL and UL, which, would change the parameter of the service Link Type to 2, Figure 4.6.

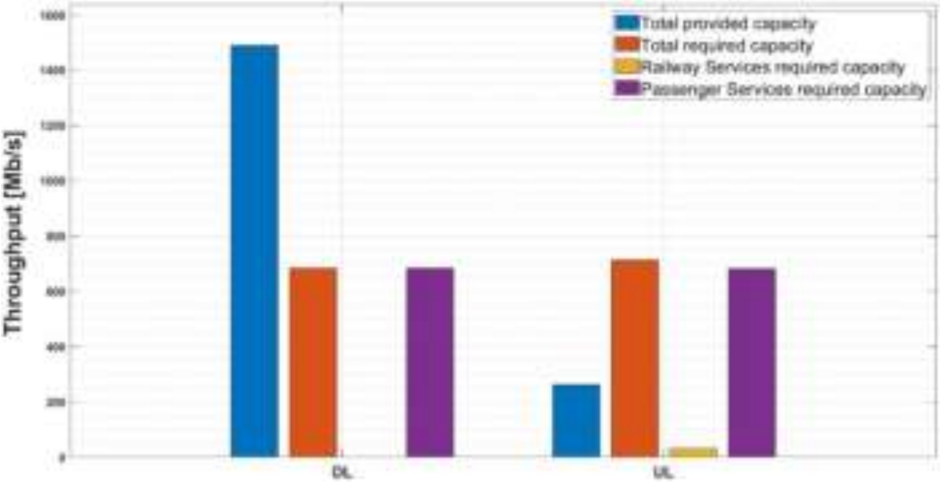


Figure 4.6 – Required and Provided Capacity if Wi-Fi service is 50/50 UL/DL.

One can observe that the network would not satisfy requirements in this example, as the provided capacity in UL is not enough compared to the required one, since the service of passenger Wi-Fi now requires the same capacity in both DL and UL. As for DL, the required capacity, which was 1.367 Gbps in the original scenario is now reduced to half, 684 Mbps. The capacity ratio is 0.46 with this change and the margin between the required and the provided capacity is now 804.67 Mbps, much higher than before allowing for higher variations in the DL. The railway services required capacity remains the same for both UL and DL as expected.

On the other hand, in UL, the provided capacity remains the same, with a value of 263 Mbps whereas the required capacity is now 713 Mbps, an increase of 682 Mbps compared to the original required 30 Mbps. The railway services required capacity remains the same again, as expected, and the passenger Wi-Fi required capacity, which was 0 in the original scenario, is now 682 Mbps. Consequently, in this link, it is not possible to satisfy requirements with the original network characteristics. Even, if one assumed the best radio conditions, with 8 MIMO layers and a CQI value of 15, it would not be possible to satisfy requirements in UL with a provided capacity of 701 Mbps, lower than the required capacity of 713 Mbps.

One could, although, change the network parameter of the Frame Structure to allow for a more convenient distribution of the total capacity in DL and UL. The original value is 0.85 meaning that 85% of the total capacity is attributed to DL and 15% to UL. These values are typical in operator implementations, since there is usually much more data flowing in DL than in UL. One could, for example, change this value to 0.5 resulting in an equally distributed capacity for UL and DL, Figure 4.7.

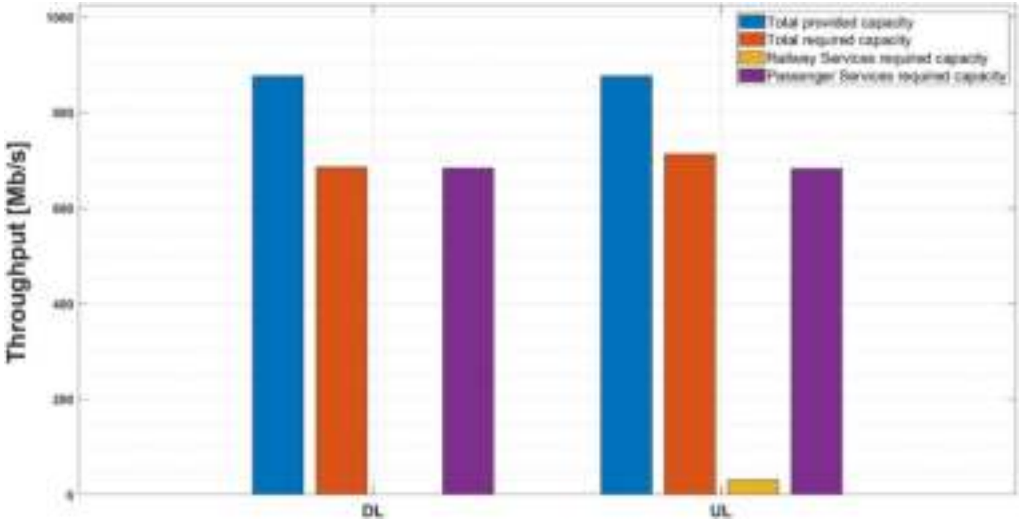


Figure 4.7 – Required and Provided Capacity with a Frame Structure value of 0.5.

With this change of the FS value, the network now satisfies the requirements both in DL and UL for the case that the passenger Wi-Fi service is given a Link Type value of 2. The provided capacity in DL is 876 Mbps compared to the original 1.489 Gbps with a margin of 192 Mbps to the required capacity of 684 Mbps and a capacity ratio of 0.78. In UL, the provided capacity is the same as in DL, 876 Mbps due to the FS value of 0.5, with a margin of 163 Mbps to the required capacity of 713 Mbps and a ratio value of 0.81. Both margins in DL and UL are high enough for variations in data usage. Railway services and passenger services required capacity remain the same. One could find the optimal network design for a given scenario of data usage and network characteristics by maximising the margins and reducing the capacity ratio in both DL and UL allowing for higher variations of data usage in both links.

Therefore, as for the initial scenario, this type of private network can satisfy requirements and provide enough capacity for the services that terminals are using, both railway ones and passenger specific.

4.2.2 Private Network – Isolated/Standalone Network Analysis

This sub-section analyses capacity results when one considers that the network is fully owned by the railway operator in an isolated or standalone type of private network. In this case, one has a limited bandwidth of 10 MHz, either in the 5.9 GHz band with the Intelligent Transport Systems bandwidth harmonised by the European commission or 10 MHz in the 1.9 GHz band made available by the Electronic Communications Committee, as seen in Chapter 2.

Figure 4.8 illustrates the throughput results provided by the RU in both UL and DL for this scenario.

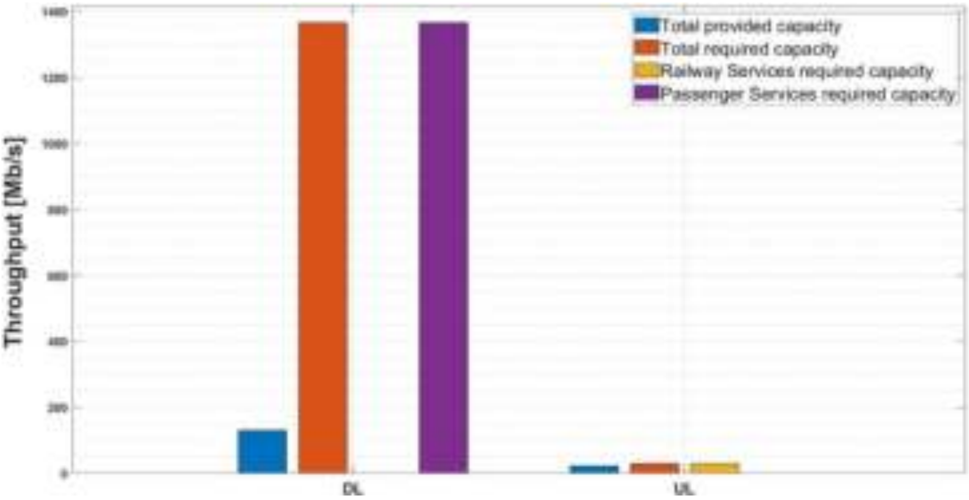


Figure 4.8 – Required and Provided Capacity with standalone private network.

There is a total available throughput of 154 Mbps provided by the RU in UL and DL, which is expectedly low due to the low bandwidth and resource blocks available compared to the previous scenario. There is a decrease of 91% in the total capacity with this type of private network compared to the slice type of private network. In DL, there is a total provided capacity of 131 Mbps and in UL there it is 23 Mbps.

Regarding the required capacity by the terminals that are using the services, since one has not changed the parameters of services and number of users, it remains the same as the previous scenario. It is obvious that this network cannot satisfy the capacity requirements in both DL and UL by a large margin in the DL of 1.236 Gbps with a capacity ratio of 10.4 and a slim margin of 7 Mbps in UL with a ratio of 1.32. One can see, that with this service mix, the service concerning passenger Wi-Fi is the one that consumes and requires more data.

Even, when changing network parameters to optimal ones, considering 8 MIMO layers, a CQI value of 15, the network still cannot provide enough capacity for the services in DL, Figure 4.9. These changes result in an increase of 62.5% of the total provided capacity compared to the case without optimal radio conditions. In DL, there is 349 Mbps provided by the RU, which is much higher compared to the 131 Mbps without optimal conditions, but still significantly lower than the required 1.367 Gbps. Whereas for UL, one has 62 Mbps, higher than the 23 Mbps without optimal radio conditions, which for this link is enough for the required capacity of 30 Mbps corresponding to railway specific services.

Ultimately, this type of private network cannot provide enough capacity for all the services, the railway and passenger ones what the terminals are using.

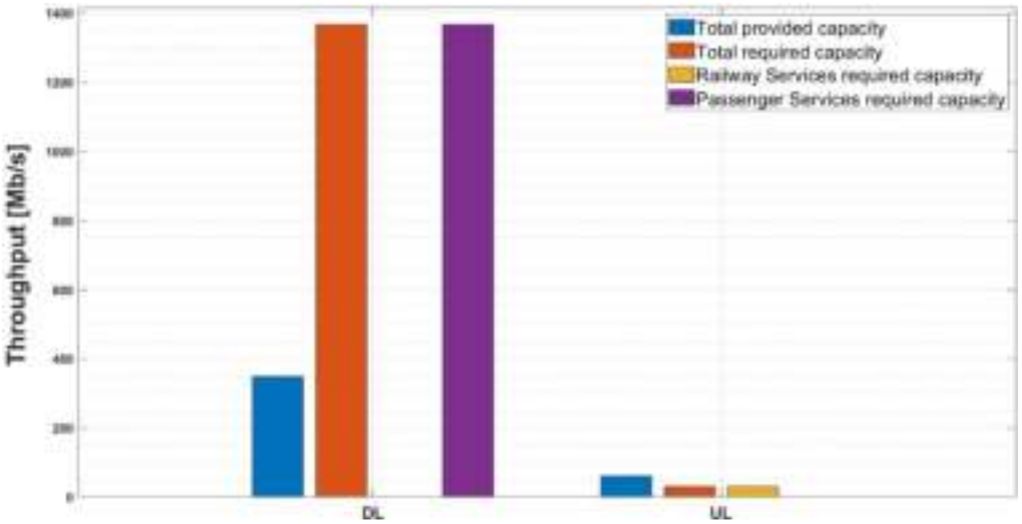


Figure 4.9 – Required and Provided Capacity with standalone network with optimal radio conditions.

4.2.3 Private Network – Shared Network Analysis

This sub-section analyses network capacity results, when considering a shared network type of private network, which means that both the metro operator and the CSP share network resources.

This design follows the solution by Ericsson presented in Chapter 2, where one has RAN sharing between both stakeholders in a hybrid type of network, where railway services are hosted by the metro operator and passenger services are hosted by the CSP in order to provide extra capacity and isolation between the two categories of services. For the purpose of this capacity analysis, one considers that the physical RAN infrastructure is common to both and that one has the 100 MHz of the operator and the 10 MHz belonging to the metro operator combined of available spectrum.

Figure 4.10 illustrates the throughput results provided by the RU in both UL and DL and the required capacity for passenger services, namely, passenger connectivity Wi-Fi and Passenger Information System. These services would use the 100 MHz available from the CSP.

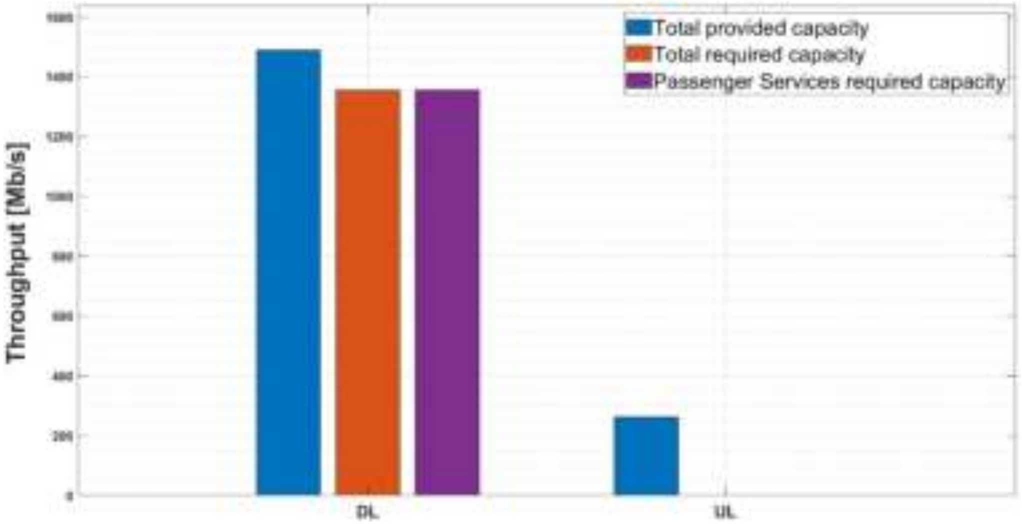


Figure 4.10 – Required and Provided Capacity with shared private network for passenger services.

The provided capacity is the same as in the first type of private network, where one uses the 100 MHz from the operator. On the other hand, the required capacity changes as one does not count railway services in this scenario. For DL, one has a required capacity of 1.355 Gbps with a margin of 134 Mbps, a ratio of 0.91 and a margin of 0 for UL with no required capacity, since one assumes both PIS and Passenger Wi-Fi to be DL predominant services. Nevertheless, the capacity available at UL would be used for the data that would be transmitted in UL in both services in case there was data flowing in that way. One could, even, increase the value of the number of slots in DL with the Frame Structure parameter to increase the margin in DL between the required and provided capacity.

Thus, the part of the network hosted by the CSP can satisfy requirements for passenger services with a reasonable margin. Secondly, one must know if the railway network can satisfy the requirements of the railway services, such as the signalling systems, the critical voice terminals, and the CCTV cameras.

Figure 4.11 shows the capacity results in DL and UL for the railway services that are hosted by the railway operator network with 10 MHz of spectrum. The provided capacity is the same compared with the second type of private network, the isolated one, as one has only the railway operator spectrum. However, as for the required capacity of the network, there is only railway services. In DL, the provided capacity is enough to guarantee the requirements of services, namely, the signalling systems and voice terminals, which use very low bandwidths compared to the provided by the BS. The required throughput for this link is of 0.456 Mbps with a significant margin of 131 Mbps relative to the offered one and a ratio of 0.003, which demonstrates the significant disparity between values. This margin allows for future updated signalling systems and voice terminals to use much more bandwidth.

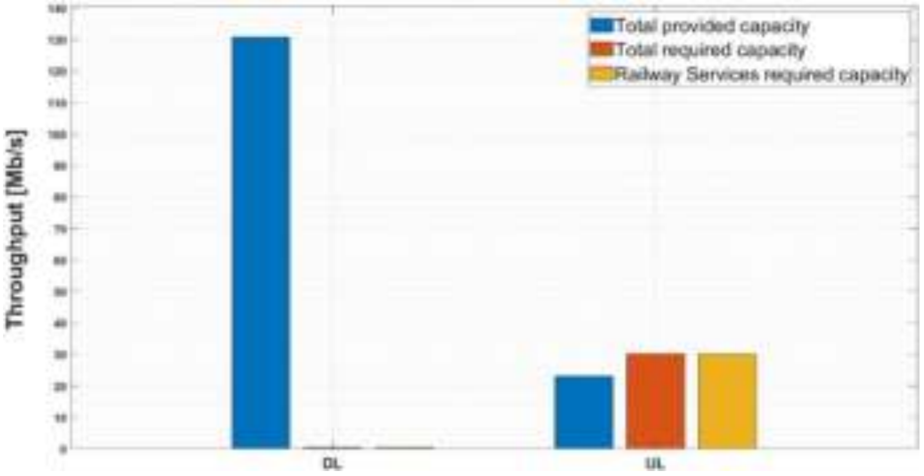


Figure 4.11 – Required and Provided Capacity with shared private network for railway services.

In UL, the provided capacity of 23 Mbps is not enough for the required one of 30 Mbps, which account mainly for CCTV cameras inside the carriages. The capacity ratio is 1.3 for this case. Nonetheless, since there is a large amount of bandwidth in DL that is not used, one can change the frame structure slot parameter from 0.85 to 0.7 and this way increase the provided capacity in UL to meet requirements. This result is shown in Figure 4.12. The maximum value of this parameter to still satisfy the service data rates is 0.8, with the lowest margin between the provided and required capacity in UL and a capacity ratio of approximately 1.

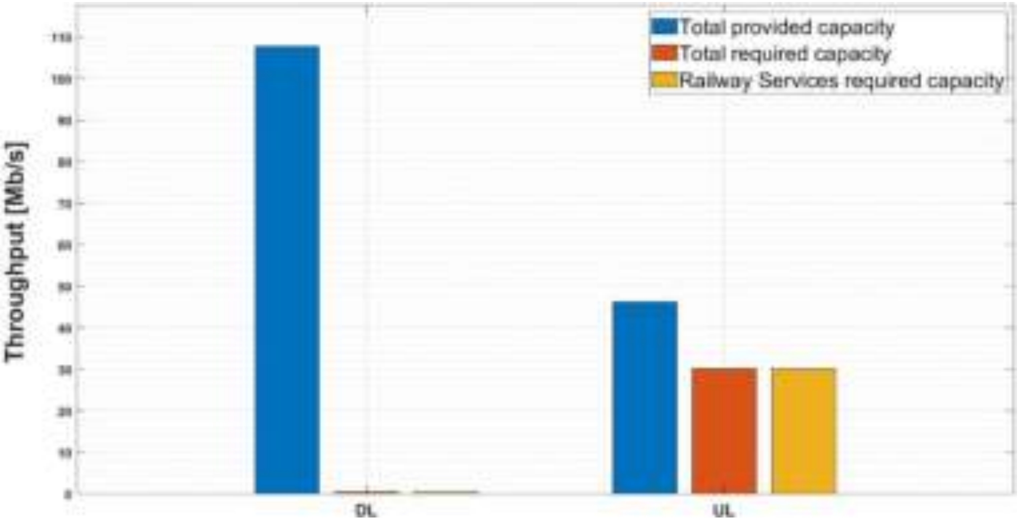


Figure 4.12 – Required and Provided Capacity with shared private network for railway services with frame structure parameter value of 0.7.

With this change, both DL and UL provided capacity is enough to satisfy all services. In UL, the provided throughput is now 46 Mbps with a margin of 16 Mbps relative to the required 30 Mbps. The capacity ratio becomes lower than 1 accordingly, with a value of 0.65.

Finally, one concludes that this type of private network, from the capacity point of view, can also satisfy all service and user requirements. It provides more capacity than all the other options and segregates railway services from passenger services, this way providing a desired isolation between different categories of services and extra layer of security.

4.3 Latency Analysis

This section analyses the delays, the total latency, the propagation maximum delay for every service. One examines the architectures and private networks that satisfy the services maximum supported E2E latency. One assumes, for the latency analysis, that network nodes and physical installation is the same for every type of private network. In a typical subway, there is the coexistence between the physical network infrastructure of the CSP and the railway's operator infrastructure. The only exception could be the position of the Core or MEC node, which, for the railway operator, would be hosted somewhere in the subway premises and for the network operator, could be located outside the subway premises and further away. One also assumes the margin for the critical threshold to be 90%. The criteria of this latency analysis for the choice of the type of accepted private network is the maximum E2E distance, which dictates how far away one could locate the core of the network or the MEC node.

4.3.1 Signalling and Control Service

The first service is the signalling and control system, which is the most critical and the one with the highest priority, requiring 5 ms of maximum E2E delay and 100 kbps of data rate.

Figure 4.13 depicts the total node latency results for each MEC deployment option: the maximum E2E Latency of 5 ms is represented by the red bold line, which must not be exceeded to satisfy the requirement; the blue dotted line stands for the Critical Threshold of the service, with 90% (margin of latency is 0.9) of the maximum E2E Total Node Latency, is considered a safety value that the service should not exceed, taking the value of 4.5 ms for this service.

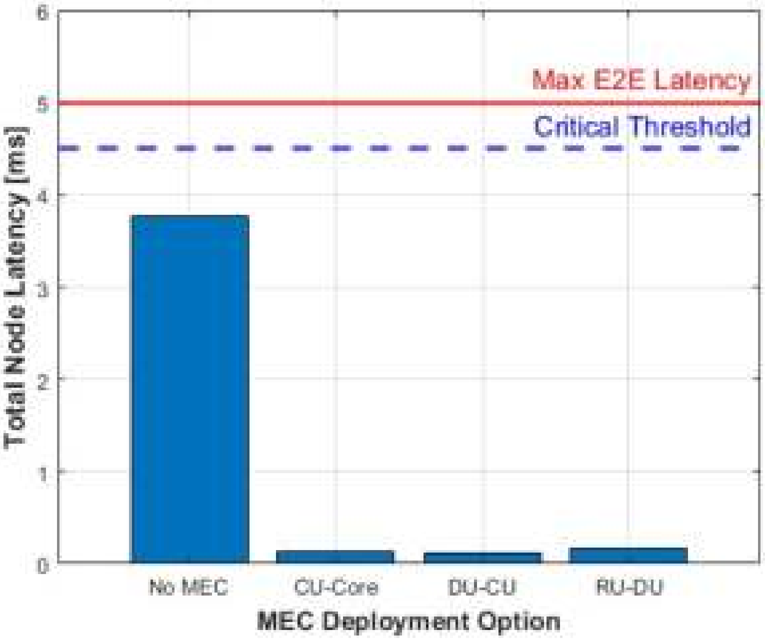


Figure 4.13 – Total Node Latency for Signalling and Control Service.

All MEC deployment options satisfy the service latency requirement, and all are below the critical threshold as well. One obtained a total node Latency of 3.77 ms for the no MEC deployment option with an acceptable margin of 0.73 ms to the Critical threshold. One should choose this deployment option then, which eliminates the need for the MEC node that brings extra cost and extra configurations. Nevertheless, for the MEC node between the CU and the Core option, the total node latency drops significantly to 0.14 ms, a 96.3% reduction in the latency compared to the scenario without MEC node. This reduction comes from the delay in the core, more specifically, the processing delay in this node, which takes values in the order of 3 ms.

For the DU-CU option, the total node latency is 0.12 ms and for the RU-DU option one obtained a total node latency value of 0.17 ms. There is a slight increase in the latency for this last option compared with the two other options using MEC node for a specific reason, which is the combined required processing located in this new MEC node. By placing the MEC node closer to the UE and replacing it with increasingly more nodes, one eliminates the node delays of the replaced nodes and the extra transmission delays, however, the MEC node processing delay increases due to the combined processing functions of the replaced nodes that now belong to the MEC node. For this scenario, this deployment option was the breaking point for which adding the MEC node increases the total node latency, which is the opposite goal of the purpose of the MEC node.

As mentioned before, the splitting option considered for the simulation was the 7.2, however, one analysed the impact of the splitting option choice on the total node latency for this service in particular, with the MEC deployment option RU-DU. The results are shown in Table 4.7.

Table 4.7 – Impact of the splitting option on the total node latency.

Splitting Option	8	7.3	7.2	7.1	6
Total Node Latency [ms]	0.182	0.168	0.174	0.179	0.157

The change of the splitting option changes the processing functions in the different nodes and the links throughputs. For this specific scenario, the splitting option that resulted in the lowest total node latency is option 6 and the highest option 8; however, one cannot conclude the same as different network configurations lead to different results. Moreover, and seen in Table 4.7, the influence of the splitting option decision on the latency is extremely low, in the 25 μs range for this scenario.

One can also analyse the contribution of each type of delay to the total node latency, namely, the queuing, processing, and transmission delays to better understand in which circumstances each of the delays contributes more or less and in which order. To do this, one used the original splitting option 7.2 and the first MEC deployment option, the option CU-Core. One used this MEC deployment option and not the no MEC option, as in the latter, the processing delay in the Core contributes almost all the total node latency, around 98%, and the results would not be conclusive on the other types of delays. Thus, one uses the first MEC option without the Core, to better understand the impact of the delay contributions on the rest of the network. Figure 4.14 shows the total node latency in both Transmission and Reception grouped by type of delay. One assumes a transmitter being the same as the receiver (e.g., a signalling system located in the train that sends/receives information to/from the signalling centre system). The total node latency includes the information that the UE sends to the MEC node which has the signalling application and the response that is sent to the same UE. Thus, the total node latency is the sum of the transmission and reception latency.

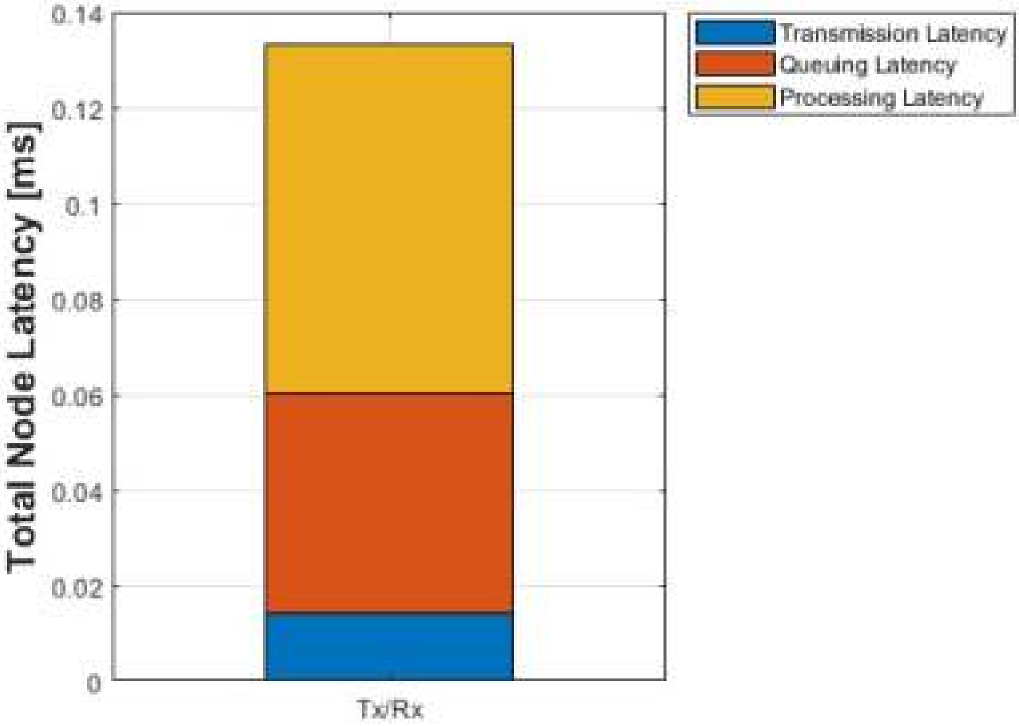


Figure 4.14 – Total Node Latency grouped by type of delay for Signalling and Control Service.

The total node latency is 0.136 ms for this scenario, as mentioned above. Starting with the transmission latency, it takes a value of 0.014 ms and this type of delay contributes the least to the total, around 10% of the total. This happens as the links have high very high capacities and the packet size of this service is relatively low.

Then, there is the queuing latency, which is approximately 0.046 ms in both the Transmitter and Receiver. It corresponds to roughly 34% of the total node latency. The queuing latency for this service is extremely low as well as the service has the most priority and overtake the others when arriving at the nodes to be processed.

Finally, the processing latency is approximately 0.073 ms, which corresponds to 54% and is the delay type that contributes more to the total node latency. The processing latency, once again, considers the processing functions that are done in each node which depend on the splitting option and the latency adaptation parameter that is relative to the service. There is a remaining 0.003 ms (2%) to make up for the total node latency that corresponds to the air link propagation delay that is not included in neither of the types of delay shown in the figure, as it is a single isolated value and does not fit any of the delay types represented.

The same type of analysis can be done considering the type of node instead of the type of delay. Thus, Figure 4.15 illustrates the total node latency grouped by type of node.

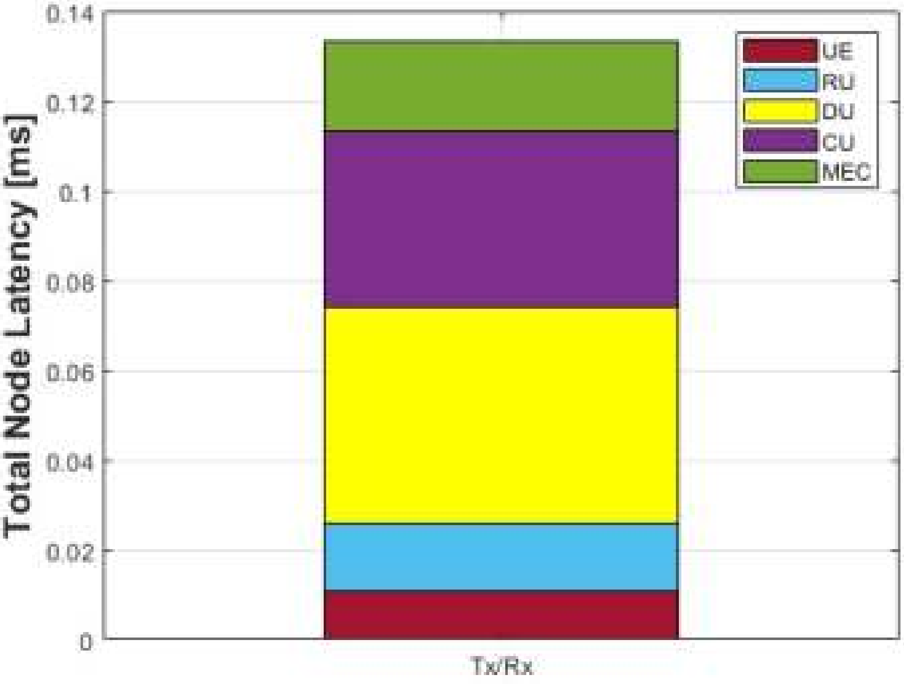


Figure 4.15 – Total Node Latency grouped by type of node for Signalling and Control Service.

Figure 4.15 shows the contributions to the delay from the several nodes in the network, namely, the UE, RU, DU, CU and MEC. The MEC node in this deployment option replaces the Core and the External Data Centre. The UE has approximately a share on total node latency of 0.011 ms (8%), the RU has 0.015 ms (10%), the DU has 0.048 ms (36%), the CU has 0.039 ms (29%) and finally the MEC node takes a value of 0.020 ms (15%).

The DU and the CU contribute the most to the total node latency, 65% combined, as the processing is heavier in these nodes and the queuing latency increases in the nodes that aggregate more traffic, namely, the CU. Once again, the air link propagation delay is the remainder of the total node latency and since is not considered a node delay, was not included in this figure.

4.3.2 Mission Critical Push-to-Talk Voice Service

The MCPTT Voice Service is mainly used by train drivers to communicate to the staff, other subway drivers or the central operations station and it is critical in cases of emergency. It has the second highest priority due to its importance. This service has a requirement of 5 ms of maximum E2E delay and 100 kbps of service data rate.

Figure 4.16 shows the total node latency results for each MEC deployment option. All MEC deployment options are valid to satisfy the service maximum latency requirement of 5 ms and none of them exceed the critical threshold. The option with no MEC results in a total node latency of 2.28 ms and a margin of 2.22 ms in relation to the critical threshold, and thus one would choose this option for this service.

The next option, the CU-Core option, implies a total node latency of 0.078 ms, a 96.6% reduction to the first option due to the elimination of the core node and the external data centre. The DU-CU option has 0.044 ms of total node latency and finally the last option RU-DU in which the MEC node replaces a total of 4 nodes, the DU, CU, Core and EDC, takes a value of 0.031 ms for the total node latency.

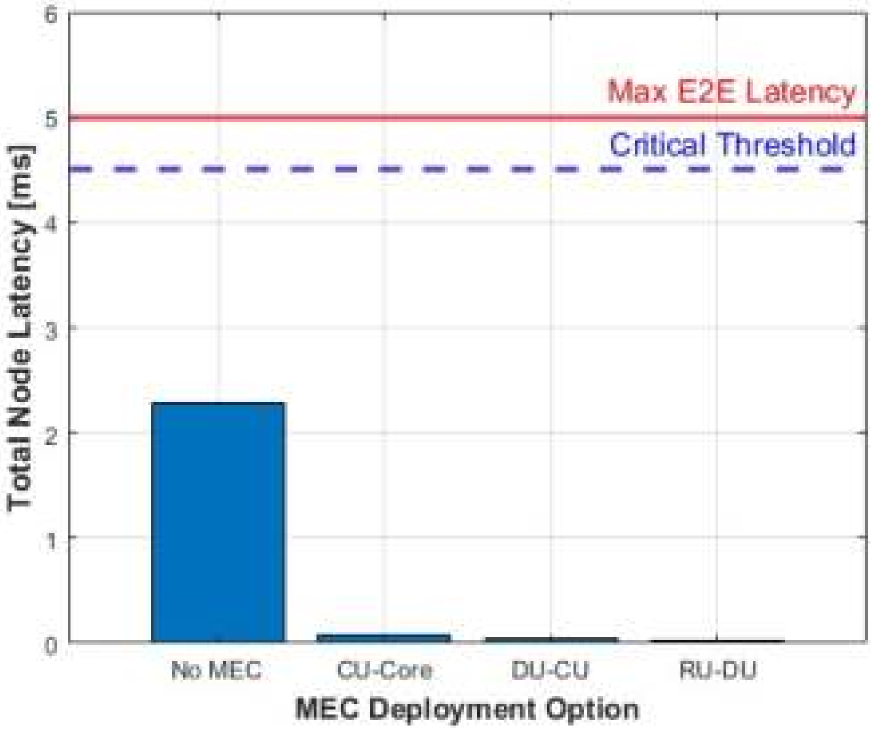


Figure 4.16 – Total Node Latency for MCPTT Voice Service.

Figure 4.17 illustrates the contribution of each type of delay to the total node latency for this service. The splitting option 7.2 was chosen as well as the second MEC deployment option CU-Core, same as with the previous service. In this service, one assumes a UE that sends the voice information to the

MEC node that hosts the MCPTT Voice application, which then sends data to another UE in the subway ecosystem. The total node latency comprises the time that it takes the data to travel from one UE to another. Both UEs are in the subway system, and one assumes the same radio characteristics and network configurations for both.

Once again, the total node latency for this MEC option is 0.078 ms. The transmission latency is 0.002 ms (3%), the lowest type of delay due to the service's short packet size (72 Bytes) and the high link capacities. Next, as one can observe in Figure 4.17 the queuing latency is the highest type of delay for this service. It is 0.059 ms with a share of approximately 76% of the total node latency. Despite being an extremely low value, it is the highest of the three, because it is not the service with the highest priority. For this reason, the queuing delay in the nodes increase as packets with higher priority are served before this service. At last, the processing latency is approximately 0.014 ms (18%). The processing latency has in account the packet size as well as the latency adaptation parameter, which is 0.75 for this service. The remaining 0.003 ms (3%) are associated with the air link propagation delay and do not show up in the figure as previously explained.

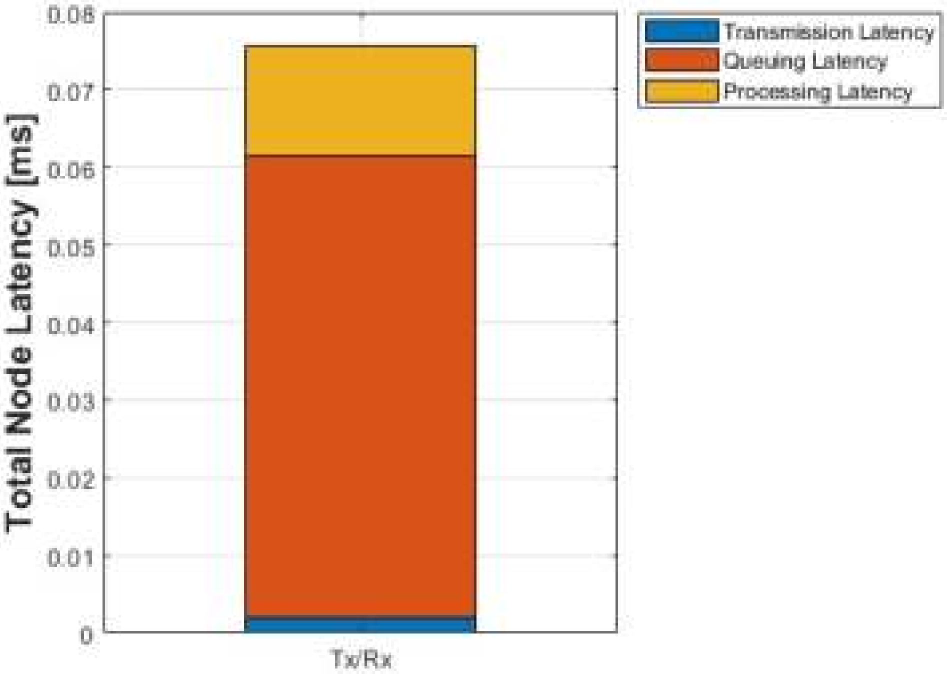


Figure 4.17 – Total Node Latency grouped by type of delay for MCPTT Voice Service.

Next, one can see the impact on the total latency of each specific node. Figure 4.18 shows the total node latency grouped by type of node, and all contributions in terms of delay from nodes in the network. Firstly, the UE has a total node latency of 0.002 ms (2%) as the packet size is extremely low and there is no queuing latency considered for the UE. The RU contributes with 0.005 ms (7%) to the total node latency, the DU has 0.029 ms of delay (37%), the CU with 0.037 ms (47%) and lastly, the MEC node has a delay of 0.003 ms (4%). The air link propagation delay has the remainder of the total node latency, approximately 3%. The queuing latency increased in this service, even more in the CU, which is the last aggregator node before the MEC node where the application is hosted. However, in the UE and the MEC node latency is extremely low as the service has the lowest packet size of all services.

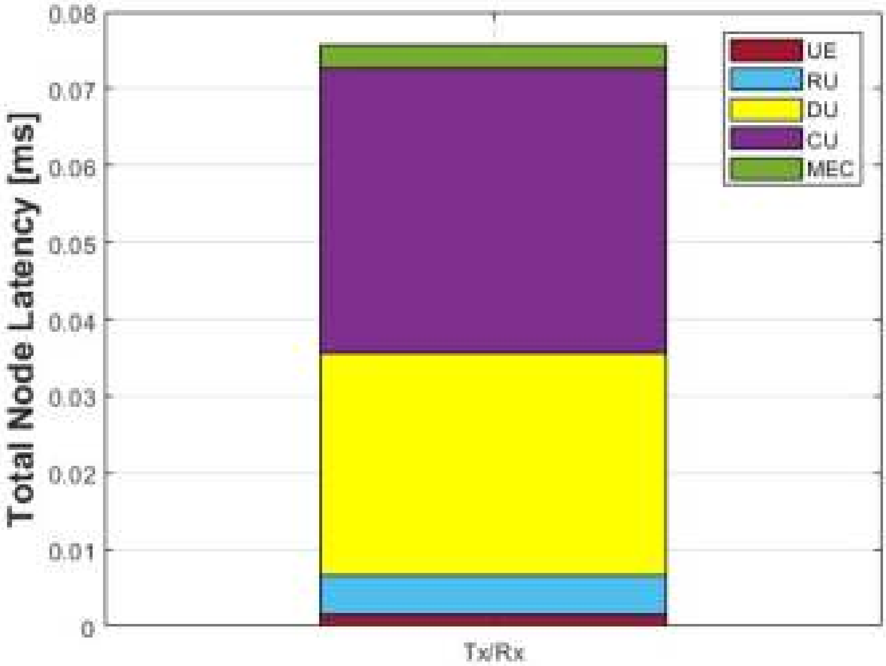


Figure 4.18 – Total Node Latency grouped by type of node for MCPTT Voice Service.

4.3.3 CCTV Service

This service consists of cameras that are spread across subway stations and inside trains for monitoring and activity recording, used for security and surveillance, crime prevention, public safety, and evidence in legal cases. This service has a requirement of 10 ms of maximum E2E delay and 2 Mbps of service data rate.

Figure 4.19 illustrates the total node latency for the CCTV service with each MEC deployment option. The critical threshold is 9 ms, 90% of the total maximum latency requirement of 10 ms. This service, unlike the others previously seen, involves a terminal or camera that only sends information to the node where the CCTV application is hosted (can be in the EDC or MEC node). This is the type of service where the response is not necessary and therefore the total latency for this service only takes in account the transmission from the UE to the application (UL) and eliminates the time that the response would take to go back to the same UE or even another UE.

All MEC options satisfy the service requirements, once again, and none exceed nor get close to the critical threshold. The No MEC option results in a total node latency of 3.95 ms which is far below the 9.5 ms threshold with a margin of 5.55 ms. This would be again the deployment option of choice as one would not need to implement the MEC node in the network that brings extra cost and complexity.

With the next MEC option, CU-Core, results a total node latency of 0.65 ms, which is an 84% decrease in latency to not using the MEC node. For the DU-CU option, one obtained a total node latency of 0.46 ms and lastly, for the RU-DU option, a total node latency of 0.51 ms, a slight increase due to the same reason as in the signalling and train control service.

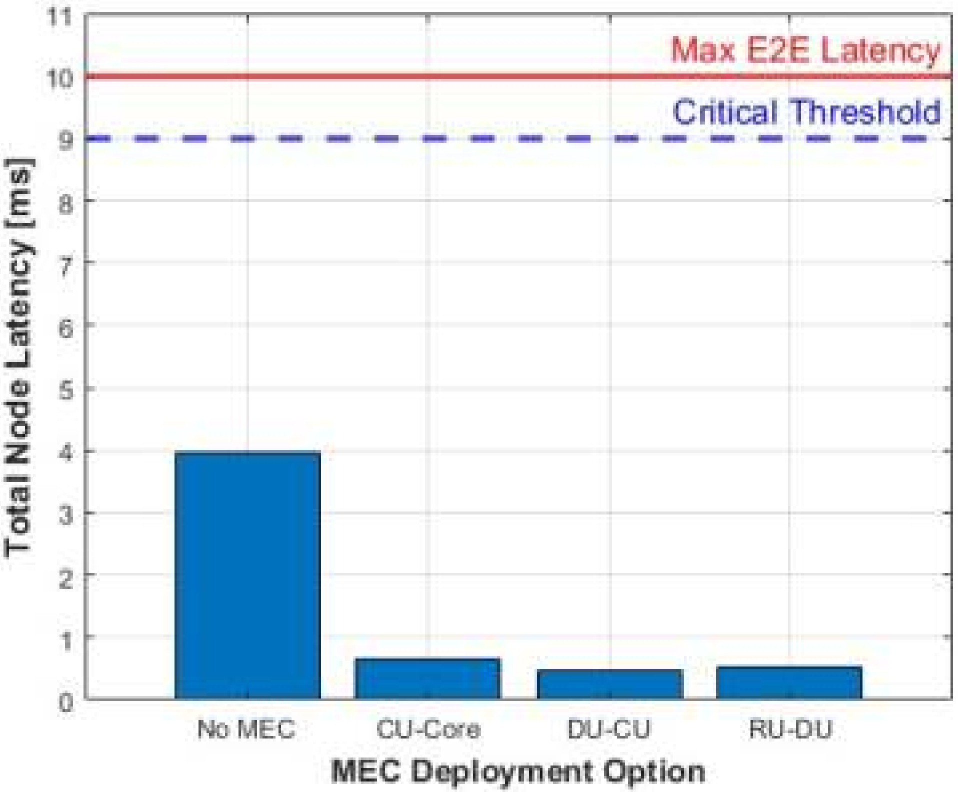


Figure 4.19 – Total Node Latency for CCTV Service.

One considered the exact same scenario, but in the case of poor radio channel conditions. The goal is to see the impact of the radio conditions on the total node latency. The CQI, which is the parameter that indicates the quality of the radio channel, was set from 12 to 5. One could imagine a poor area of the subway system with poor coverage or some type of extra attenuation. The conditions of the radio channel affect the UE transmission delay, which consequently affect the processing delays in the nodes.

The results show an increase of approximately 0.5 ms on the total node latency for every MEC deployment option, a 12% increase. Thus, even a slight change on the input, specifically, on the radio characteristics is not enough, for this service, to exceed the maximum latency requirement nor the critical margin.

Next, one can analyse the impact of each type of delay on the total node latency. The MEC option CU-Core was used as well as the splitting option 7.2. Figure 4.20 shows these results, and as seen previously, for this service one only considers the one-way latency to the CCTV app, therefore with only a transmitter. The total node latency, with this MEC option, is 0.650 ms, as mentioned above. The transmission latency is 0.028 ms (4%), the queuing latency is 0.350 ms (53%) and the processing latency is approximately 0.270 ms (42%). The remainder is the air link propagation delay with 0.001 ms (1%). The transmission delay remains the lowest type of delay, mainly due to the very high link capacities of the network. The queuing latency increases, compared to the previous service, as the priority of this service is lower. The number of users was kept constant. At last, the processing latency increases as well with almost half of the latency share in this service, because of the packet size increasing (1 400 Bytes) and the latency adaptation parameter taking the maximum value of 1.

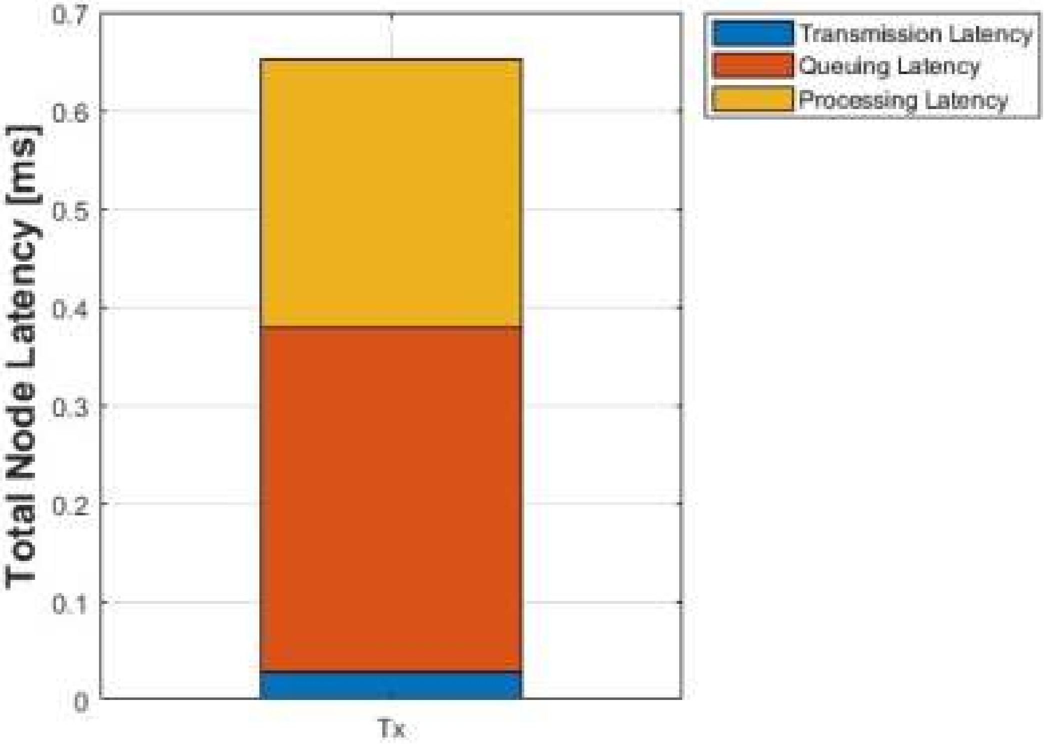


Figure 4.20 – Total Node Latency grouped by type of delay for CCTV Service.

After this, the analysis on the effect of the type of node in the total node latency is done, following the same rules as the previous services. Figure 4.21 shows the total node latency grouped by type of node.

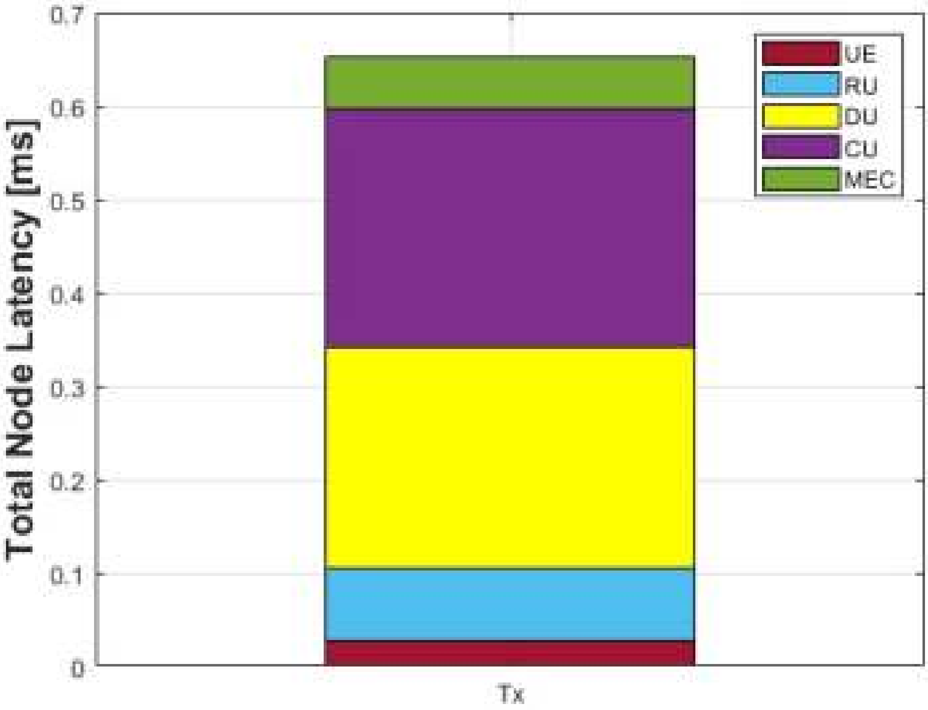


Figure 4.21 – Total Node Latency grouped by type of node for CCTV Service.

By analysing the figure, the UE has approximately a total node delay of 0.027 ms (4%), the RU has 0.078 ms (12%), the DU has 0.240 ms (36%), the CU has 0.250 ms (38%) and the MEC contributes to the total node latency with 0.056 ms (9%). The remainder corresponds to the air link propagation with 0.001 ms (1%). The MEC node contribution increases due to the increase of the packet size compared to previous services. However, the DU and CU are still the nodes that account for more than 70% of the total node latency as a consequence of the processing and queuing delays associated with these nodes.

4.3.4 PIS Service

The PIS service, which stands for Passenger Information System, is the fourth service to be tested. It is composed of all the systems in the metro system that provide information to passengers via audio, text or video transmitted through digital displays inside the carriages or via speakers. It can provide information on the stops, scheduling, arrival and departure times, announcements on delays, disruptions or emergencies, advertising, etc. This service has a requirement of 10 ms of maximum E2E delay and 0.5 Mbps of service data rate.

Figure 4.22 shows the total node latency for this service for all the MEC deployment options. One considers this service to be mainly DL as these systems only receive the information to be given to the passengers and therefore one only considers the one-way latency from PIS servers to all equipment in the carriages that broadcast the information.

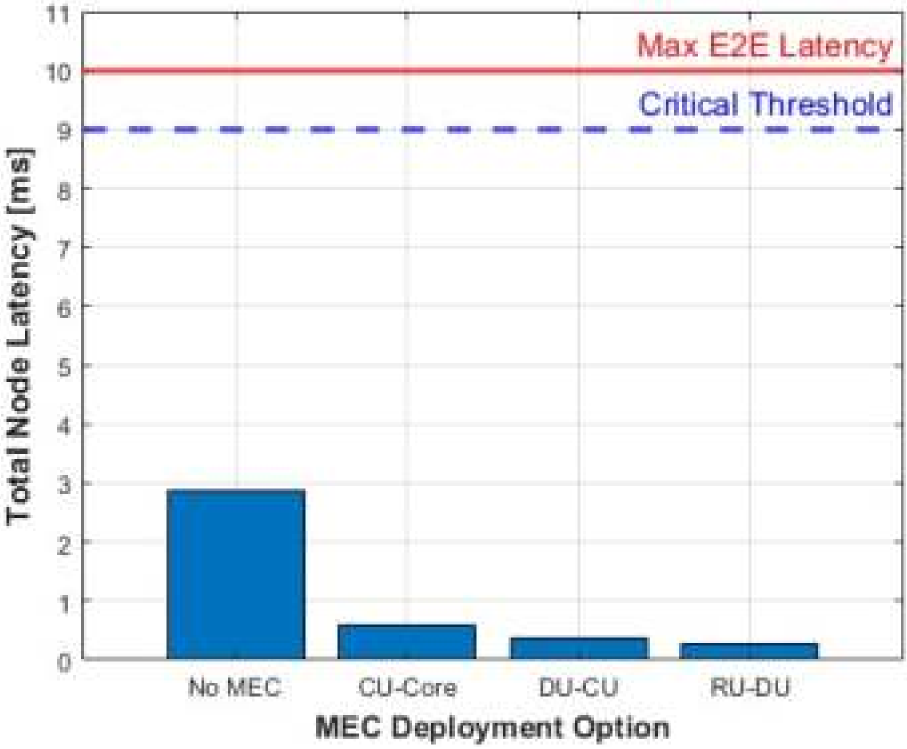


Figure 4.22 – Total Node Latency for PIS Service.

All the MEC deployment options are possible again to satisfy the latency requirement. With no MEC node, the total node latency is 2.880 ms, which results in a margin of 6.120 ms to the critical threshold. There is no need, one more time, to install this node for this service. Furthermore, the option CU-Core implies a total node latency of 0.580 ms (80% reduction to not having MEC). With the option DU-CU, one has a total node latency of 0.350 ms and finally, a latency of 0.280 ms with the hardest option to implement in practical terms, the RU-DU option.

Next, one analyses the grouped delays by type – Transmission, Queueing and Processing for this service. Figure 4.23 represents the total node latency grouped by type with the MEC option CU-Core and splitting option 7.2. One only considered the one-way latency in the receiver equipment, as explained before.

The total node latency for this scenario with this MEC deployment option is 0.580 ms, which can be seen in Figure 4.22. The transmission latency is 0.016 ms (3%), the queuing latency is 0.400 ms (67%) and the processing latency is 0.170 ms (29%). The remainder is the air link propagation delay with 0.001 ms (1%).

Comparing with the previous service, the CCTV that has a one-way latency as well, one deduces that both processing and transmission latency percentages decreased as this service has a lower packet size (800 Bytes) with the same latency adaptation parameter. However, the queuing latency percentage increased as a result of the service having a lower priority and therefore an increased delay in the node's queue. Overall, the total node latency decreased in relation to the CCTV service.

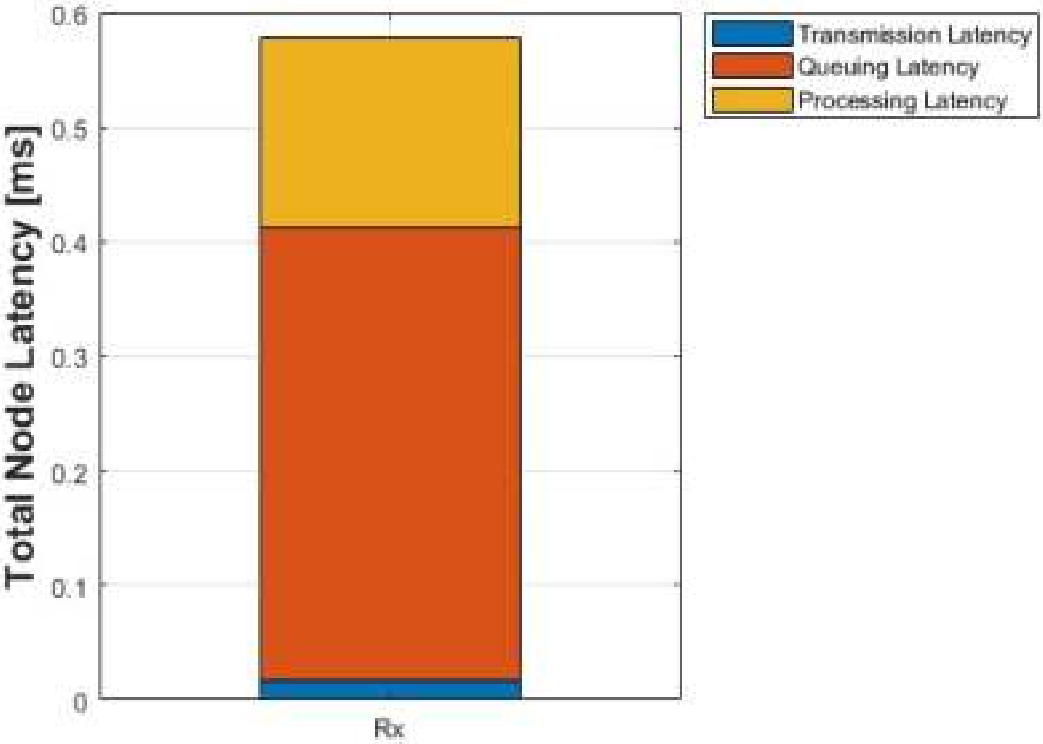


Figure 4.23 – Total Node Latency grouped by type of delay for PIS Service.

Next, one inspected the total node latency grouped by node in the same scenario. Figure 4.24 shows the total node latency grouped by the type of node in the network.

One can see that the UE has approximately a total node latency of 0.017 ms (3%), the RU takes a value of 0.032 ms (5%), the DU has 0.240 ms (42%), the CU has 0.260 ms (44%) and lastly, there is 0.032 ms (5%) of delay in the MEC node. The remainder corresponds to the air link propagation with 0.001 ms (1%).

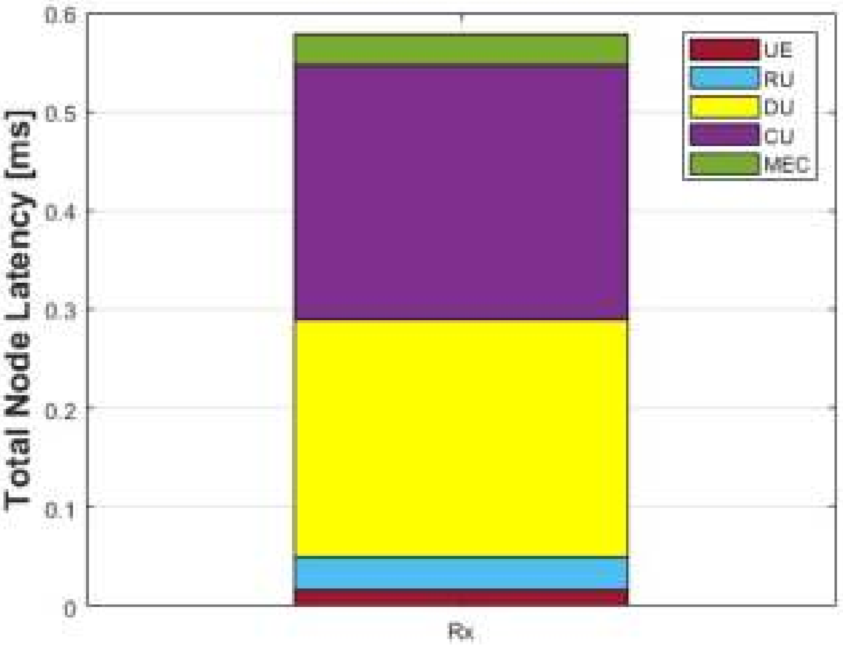


Figure 4.24 – Total Node Latency grouped by type of node for PIS Service.

The contributions of each node to the total node latency are similar to the previous service of CCTV, although, the delays in the UE, RU and MEC node account for 13% of the total as opposed to the 25% in the CCTV, because of the reduction in the packet size and therefore decreased delay in processing and transmitting the packets in these nodes. Both DU and CU contribute the most to the total latency which is 86% for this service and 74% for the CCTV.

4.3.5 Passenger Connectivity (Wi-Fi) Service

The last service to analyse, regarding latency, is the passenger connectivity or Wi-Fi. This is the service with the least priority and with the least strict latency requirement. The goal of this service is to provide an Internet connection to passengers that are using the subway. This service has a requirement of 50 ms of maximum E2E delay and 5 Mbps of service data rate.

Figure 4.25 shows the total node latency with all the MEC node deployment options for this service.

The maximum latency and the critical threshold are respectively, 50 ms and 45 ms for this service. All MEC options satisfy the service latency requirements by a considerable margin.

It is important to note that the implementation of the MEC node for this service is not practical and unlikely to be implemented, as it would be impossible for the MEC node to host all possible internet applications that the UE is using or communicating. It could only have local applications for some Internet services. Therefore, the option with no MEC node is the best and most suitable and it satisfies the requirements as well.

This option, no MEC, results in a total node latency of 19.6 ms with a fine margin of approximately 25 ms to the critical threshold. The other options are not considered for the reason explained before.

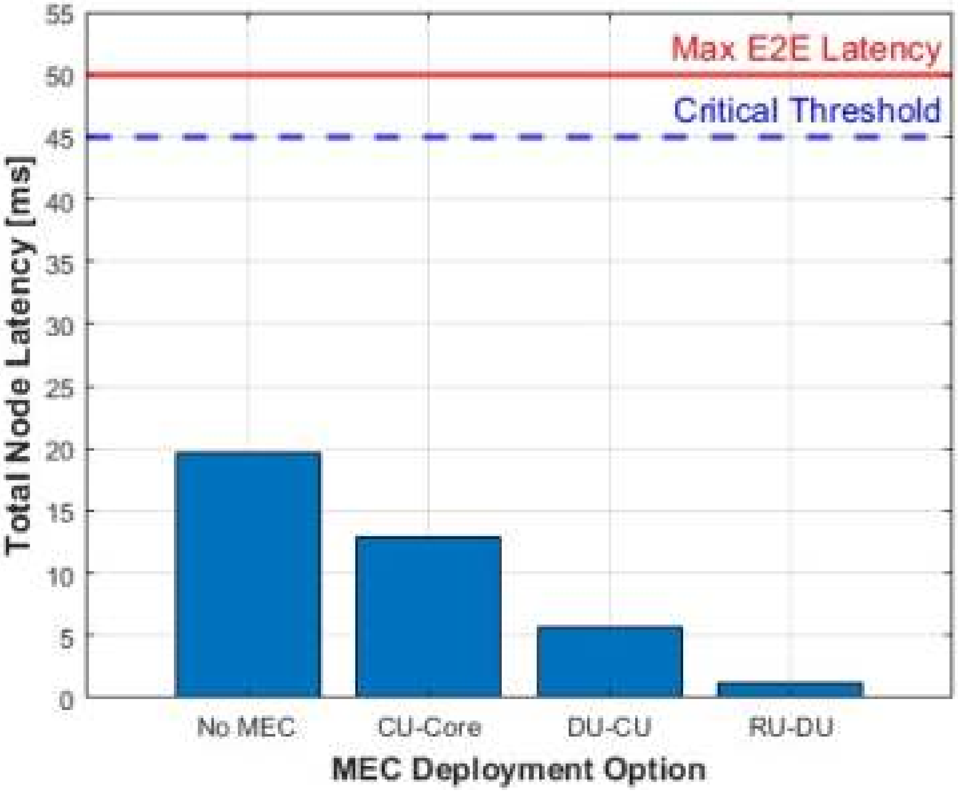


Figure 4.25 – Total Node Latency for Passenger Wi-Fi Service.

Figure 4.26 illustrates the total node latency grouped by type of delay with the no MEC option as the others are not considered. The standard splitting option 7.2 was used.

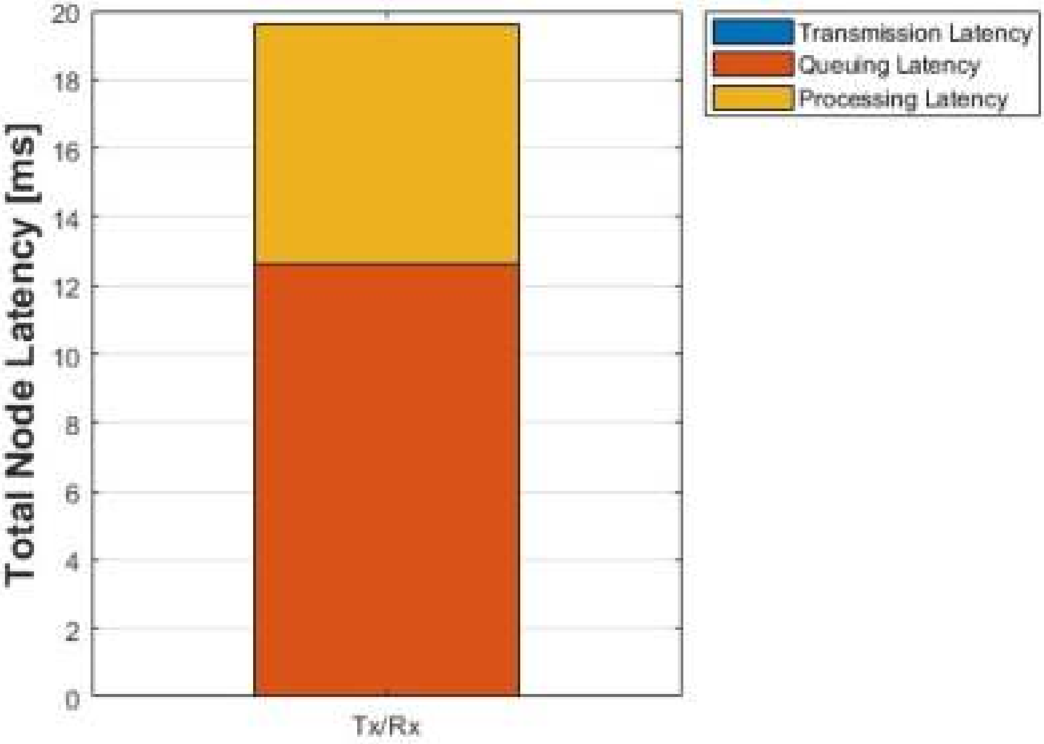


Figure 4.26 – Total Node Latency grouped by type of delay for Passenger Wi-Fi Service.

From the total node latency of 19.6 ms, the transmission latency accounts for 0.040 ms (0.2%), an extremely low percentage compared with the others due to the very high link capacities of the network links. The processing latency of the nodes is approximately 7 ms (35%) and the queuing latency is 12.6 ms (64%). The considerable increase in the queuing and processing delay are caused by the low priority of the service, the packet size for this service (1 400 Bytes) and the maximum latency adaptation parameter of 1.

Figure 4.27 shows the total node latency grouped by node in the network.

In this scenario, one has the Core node and the External Data Centre (EDC) node since no MEC node option is considered. By analysing the figure, the total node latency of 19.6 ms is composed of: the UE delay with 0.030 ms (0.15%), the RU with 0.790 ms (4%), the DU has 4.730 ms (24%), the CU with 7.370 ms (38%), the Core with 6.660 ms (34%) and finally the EDC with 0.019 ms (0.1%). The remainder is the air link propagation delay with 0.003 ms (0.01%).

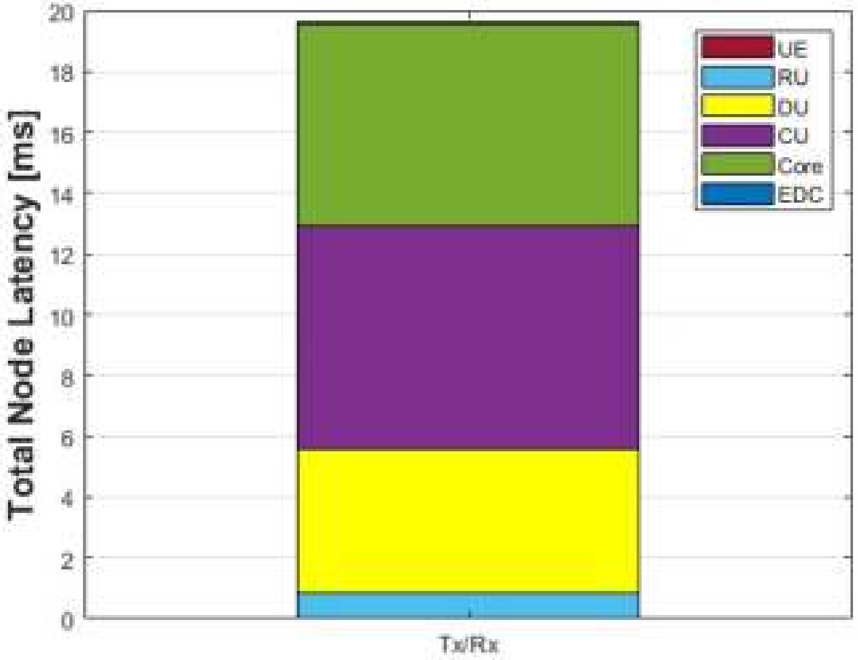


Figure 4.27 – Total Node Latency grouped by type of node for Passenger Wi-Fi Service.

The UE, air link propagation and the EDC delays are negligible as in total account for 0.26%. One supposed for the EDC no queuing latency and a fast data processing since these are typically large servers with expensive hardware and software typically owned by hyperscalers. The DU and the CU account for 62% of the total with higher processing and queuing delays. Finally, the core node has a higher processing delay as it is a complex node where multiple operations must be done.

In addition, one evaluated the impact of the number of passengers in the whole subway system on the final total node latency. The number of users affects the queuing delay, and a higher number of users result in a higher queuing delay. One considered the scenario of having an increase of 150 passengers per train, which results in a total increase of 8 400 passengers in the whole metro system, which means a total of 25 200 passengers against the original total of 16 800 total passengers. With these changes, one obtained a total node latency of 25.9 ms with the no MEC deployment option, which means an increase of 25% of the total node latency. One concludes that, even with this increase on the number of passengers, the network can still satisfy the latency requirements of the service with a large margin.

4.3.6 Maximum E2E Distance and Accepted Private Networks

After having simulated all services and computed the total node latencies for all of them, one can calculate the maximum propagation delay and consequently the maximum E2E distance. This is the maximum distance from the UE to the end node where the service application is hosted, which is the EDC node or the MEC node if used in the architecture. This is the maximum distance that satisfies the latency requirement and so one needs to analyse the possible location of the EDC or MEC node for each service and check the type of private network that is possible for each service.

Since none of the services need the implementation of the MEC node, as analysed in the previous sub-chapters, one considers the EDC or the core of the network (considering them to be collocated) to be the end node.

As seen in the capacity analysis, the types of private networks are network slice, isolated network, or shared network. In the network slice, one considers that all infrastructure is owned by the CSP, and therefore the core of the network is the core of the CSP located somewhere in the Lisbon region (not disclosed). In the isolated network, all the infrastructure is owned by the railway operator, and therefore the core of the network is the core of the subway operator, which in this case (Metropolitano de Lisboa) is in Pontinha, next to the subway station. Finally, the shared network uses both cores and RAN sharing.

Table 4.8 shows the latencies, maximum E2E distance and accepted private network for each service.

Table 4.8 – Latencies, maximum E2E distance and accepted private network for each service.

	Signalling	**MCPTT Voice**	**CCTV**	**PIS**	**Passenger Wi-Fi**
Total node latency [ms]	3.77	2.28	3.95	2.88	19.6
Maximum latency [ms]	5	5	10	10	50
Maximum propagation delay [ms]	1.23	2.72	6.05	7.12	30.4
Maximum Distance [km]	33	73	326	384	819
Accepted Private Networks	• Isolated • Shared	• Isolated • Shared • Slice	• Isolated • Shared • Slice	• Isolated • Shared • Slice	• Isolated • Shared • Slice

The E2E Maximum Distance corresponds to the variable d_{max} for each service. Moreover, the maximum distance from a UE in the subway ecosystem to the core of the rail operator located in Pontinha $d_{ue-RailCore}$ is approximately 15 km. Assuming that at least one CSP has a core node in the Lisbon area, that the city of Lisbon has an area of 100 km^2 and that the total area covered by the subway system is approximately 50 km^2, the maximum distance between a UE in the metro system and the core of any CSP, $d_{ue-CSPCore}$, with a margin of error due to several factors, is considered to be 35 km.

The latency requirement for all services is satisfied and no MEC node is necessary in any service. For the Signalling, MCPTT Voice and Passenger Wi-Fi services, the maximum distance was divided by 2, since these are two-way latency services, whereas the CCTV and PIS services are one-way latency services in UL and DL, respectively. The maximum propagation delay increases from the service of the left to the service of the right as the latency requirement becomes less strict and with less priority. Consequently, the maximum E2E distance increases as well, having a value of 33 km for the signalling service and 819 km for the Wi-Fi one.

All services, expect for the signalling one, allow for the deployment of any type of private network – isolated, shared and network slice, since the condition $d_{max} > d_{ue-CSPCore}$ is verified for these services. The signalling service is the only one where this condition does not apply, and one would have to choose between the isolated or shared private networks option.

4.4 Network Configurations

This section covers the network configurations that are accepted in each of the analysis that was made in the previous chapters, namely, the capacity and latency analysis. These were quantitative analysis which studied the possible architectures in terms of capacity and latency taking in account the services and their requirements.

Table 4.9 shows the private networks accepted for each of the analysis and globally.

Table 4.9 – Accepted Private Networks for Capacity, Latency and Global analysis.

	Capacity	Latency	**Global**
Standalone / Isolated	✗	✓	✗
Shared	✓	✓	✓
Slice	✓	✗	✗

The three types of private networks that were studied are standalone/isolated, shared between the CSP and the rail operator and finally a network slice, virtual network portion with a guaranteed SLA with the

CSP.

By observation of Table 4.9, the global analysis only accepts the private network if it is accepted in both capacity and latency analysis. It is an intersection between the capacity and the latency columns, therefore, only if it is accepted in both columns is consequently accepted in the end, seen in the global column. The global column refers to the final solution and private networks that are accepted having taken in account all the previous requirements.

In terms of capacity, all options except the standalone/isolated satisfy the requirements and therefore are possible to implement. Whereas in the latency evaluation all options are valid except the network slice. Finally, combining both analyses, the only private network that is possible to implement as it satisfies all the requirements is the **shared option**.

The different categories of services are considered for each analysis as well namely, the critical services that include the signalling and control service, voice MCPTT, CCTV and non-critical that include the PIS and the passenger connectivity. For the shared private network option, these services are separated and isolated as the critical ones are hosted in the railway core and the non-critical are hosted in the core of the CSP. This option is a possible future alternative to the technology used by Thales, called BBRS, which is based on Wi-Fi as seen in Chapter 2.

With the shared private network, besides satisfying the requirements quantitatively, it also brings advantages qualitatively such as : lower capital investment as costs can be shared and equipment from both railway operator and CSP can be re-used, like cables, radios, cell sites, etc.; a desired isolation between railway specific and passenger-oriented applications that brings an extra layer of security, resilience and data privacy; and redundancy as if one of the networks goes down, the other serves as a backup ensuring that railway operations continue even in the event of a disaster.

Chapter 5

Conclusions

This chapter gives the summarisation and the final conclusions of this thesis.

The main goal of this thesis was to analyse and study which types of 5G Private Networks could satisfy a set of requirements for a list of services that are found in railway communication systems. A model was developed that takes several input parameters into account, such as, network and user parameters, and then evaluates latency and capacity that ultimately determine which network deployment options and network configurations are acceptable for the scenario that is being studied.

Chapter 1 presents a brief introduction to the subject, an overview of mobile communications systems and how it links with Private Networks and railway communications. The enablers and new concepts around 5G are mentioned in this chapter as well as how 5G Private Networks can improve railway systems in terms of performance, efficiency, and security. Finally, this chapter ends with the main motivations for the development of this work and the objectives of this thesis and its content.

Chapter 2 provides the fundamental concepts of the main topics related to this work. It includes technical explanations and general architectures about 5G, its network architecture, both Core and RAN, a section for the radio interface that describes frequency bands, duplexing modes, multiple access methods, modulation, the concept of numerology, and how it can affect radio latency. Moreover, a brief specification about C-RAN, Network Slicing and Private Networks is done. 5G Services and Applications are detailed in this chapter as well as the state-of-art for this work with the most relevant information found in papers, articles, and company white papers.

Chapter 3 concerns the model that was developed for this thesis. The first section presents the model overview, which comprises its inputs like network and services parameters, intermediary calculations on capacities and latencies, and finally the outputs of the model. This section also covers the characteristics of the services to be studied, of which some are related to railway passengers and others specific to railway's operations. The second section describes the latency model, and the expressions to calculate the several delays in the network, such as, transmission, queuing, processing, and propagation. The different possible network architectures with the different nodes are explained in this section and the possibility to incorporate a MEC node is detailed. There are several deployment options for the MEC node, and each option has an impact on latency as the MEC node replaces one or more nodes. The splitting option is also considered in this section as it impacts the throughputs on network links and processing functions in each node. Moreover, this section shows the expressions to calculate the maximum total propagation delay in the network to satisfy service requirements, the total node latency, the maximum E2E distance between the UE that transmits and the one that receives or between the UE that transmits and the External Data Centre if the latency is only considered to be one-way. The critical threshold is computed for the service in question and in the end, one presents the reasoning to identify the accepted private network deployments that satisfy the maximum distance requirement.

The third section of this chapter relates to link throughputs and capacities. The throughput provided by the BS is computed according to radio parameters and available bandwidth, as well as the capacity required by users in DL and UL for both railway and passenger related services. The capacity ratio is determined, which tells if the cell is saturated or not. Besides, the throughputs for the rest of the network are computed based on the chosen splitting option.

The implementation of the model is found in the fourth section of this chapter, where different flowcharts are illustrated, showing how the general model and sub-models function in order to achieve the intended results. The last section depicts the model assessment, which includes the list of tests and validations made to assess the program and its functions. The results were compared with expected results and results found in other works.

Chapter 4 presents the main results for this thesis. It starts by describing the scenario that is used in the analysis, which is the Lisbon Subway System. The main characteristics of the system are provided, such as the number of stations, geographical area, location of the network core, distance between stations, number of carriages per train, average number of users per carriage, CCTV cameras per carriage, and critical voice terminals per train. One then presents the reference input values for the scenario, like radio configurations, number of users, service mix, and link types for each service.

Regarding throughput results, the analysis is done for the three types of private networks: network slice, isolated/standalone and shared. The amount of available spectrum is the big differentiator among these solutions, besides the differences in architectures. For the network slice, one obtained a total throughput of 1.752 Gbps provided by the RU in both UL and DL, which translates in 1.489 Gbps in DL and 263 Mbps in UL, according to the frame structure. The required capacity is 1.367 Gbps in DL and 30 Mbps in UL. This type of private network satisfies the capacity constraints and provides enough capacity for the services with a ratio of 0.92 in DL and 0.12 in UL.

As for the isolated/standalone private network, there is a total of 154 Mbps available, from which 131 Mbps are in DL and 23 Mbps in UL. There is a decrease of 91% in the total capacity with this type of private network compared to the slice type of private network. It cannot provide enough capacity in either UL or DL. Even considering the optimal network parameters with the best radio channel conditions, the provided capacity is still not enough.

Finally, the shared private network involved combining the spectrum from both the railway operator and the CSP, and hosting the railway dedicated services in the railway core while the passenger specific services are hosted in the core of the CSP. Regarding the passenger services, the required capacity is 1.355 Gbps in DL with a ratio of 0.91 and 0 Mbps in UL. Having the same provided throughput as in the network slice with the spectrum of the CSP, these services have their capacity guaranteed. Whereas for railway services the required capacity is 0.456 Mbps in DL and 30 Mbps in UL with the same provided capacity as in the isolated private network. There is enough capacity in DL and insufficient in UL. By changing the frame structure parameter to 0.7 and allocating more bandwidth to UL, the problem is solved and this private network can thus satisfy all the requirements concerning capacity.

The latency analysis and results are presented in the third section of Chapter 4. This section analyses the total node latency, maximum propagation delay, maximum E2E distance for each service bearing in mind each service's maximum latency requirement.

For the railway signalling and control application, the most critical service, the total node latency is 3.77 ms without using the MEC node, with a margin of 0.73 ms in relation to the critical threshold. The maximum latency for this service is 5 ms. When using the MEC node, the total node latency drops to below 1 ms with any deployment option, a 95% reduction in the latency. The impact of the splitting option in the total node latency is in the 25 μs range for this service and scenario. The maximum E2E distance between the UE and the Core is 33 km.

For the MCPTT Voice service, one obtained a total node latency of 2.28 ms without any MEC node and a maximum E2E Distance of 73 km. The maximum latency for this service is 5 ms as well. The third service is the CCTV cameras with a latency requirement of 10 ms. The latency for this service only takes in account the transmission from the UE to the CCTV application (UL) and eliminates the time that the response would take to go back to the same UE or even another UE. The total node latency is 3.95 ms with a maximum propagation delay of 6.05 ms and consequently a maximum E2E Distance of 326 km. By worsening the radio channel conditions by changing the CQI to a value of 5, one obtains a 12% increase on the total node latency for every MEC deployment option. As for the fourth service, the PIS, the total node latency is 2.88 ms and a maximum E2E distance of 384 km. The last service of passenger Wi-Fi with a maximum delay of 50 ms results in a total node latency of 19.6 ms with a margin of 25 ms to the critical threshold and a maximum E2E distance of 819 km. One observes as well that by increasing the number of passengers per train by 150, which results in a total increase of 8 400 passengers in the whole metro system, one obtains a total node latency of 25.9 ms, a 25% increase on the total. Still, the service requirement is satisfied.

Accordingly, there is no need to install a MEC node in the network as every service requirement is satisfied without one. In addition, one observes that the queuing delay ranges from 0.04 ms to 12.6 ms and that depends mainly on the service’s priority. The processing delay is in the 0.014 ms to 7 ms range and varies with the packet size while the transmission latency ranges from 0.01 ms to 0.04 ms taking very low values due to the very high-capacity links and high throughputs. Furthermore, the nodes that contribute the most to latency are the DU and CU, as these aggregate more traffic and have more complex processing functions. In terms of latency, all private network types are accepted except for the slice network, as it does not allow satisfying the requirement of the signalling service. This service has a maximum distance to the core of 33 km, which is lower than the distance to the CSP core of 35 km.

In conclusion, several architectures were studied and analysed, but the only solution that satisfies all requirements quantitatively and qualitatively is the shared private network between the CSP and the railway operator. It provides the capacity required by users and at the same time provides the technology so every service can be satisfied in terms of latency.

In future work, new applications that are even more demanding and stricter should be considered like VR/AR applications in the railway system, driverless trains and IoT related applications with train sensors. Additionally, new available and harmonised spectrum for private networks and railway communications could be analysed, as well as the implications on more complex scenarios with a heavy user load and co-existence of cellular 5G with Wi-Fi to provide better performance and extra capacity.

Annex A

Distances between Metro Stations

This Annex presents the distances between Lisbon metro stations and the method used.

Table A.1 – Distances between metro stations for Blue Line.

Stations	Distance [m]
Reboleira – Amadora Este	800
Amadora Este – Alfornelos	1 300
Alfornelos – Pontinha	700
Pontinha – Carnide	500
Carnide – Colégio Militar / Luz	700
Colégio Militar / Luz – Alto dos Moinhos	900
Alto dos Moinhos – Laranjeiras	700
Laranjeiras – Jardim Zoológico	800
Jardim Zoológico – Praça de Espanha	1 000
Praça de Espanha – S.Sebastião	500
S.Sebastião - Parque	700
Parque – Marquês de Pombal	500
Marquês de Pombal – Avenida	700
Avenida – Restauradores	700
Restauradores – Baixa-Chiado	500
Baixa-Chiado – Terreiro do Paço	600
Terreiro do Paço – Santa Apolónia	1 200

Table A.2 – Distances between metro stations for Yellow Line.

Stations	Distance [m]
Odivelas – Senhor Roubado	800
Senhor Roubado – Ameixoeira	1 300
Ameixoeira – Lumiar	800
Lumiar – Quinta das Conchas	700
Quinta das Conchas – Campo Grande	800

Campo Grande – Cidade Universitária	1 000
Cidade Universitária – Entre Campos	1 100
Entre Campos – Campo Pequeno	600
Campo Pequeno – Saldanha	700
Saldanha – Picoas	500
Picoas – Marquês de Pombal	700
Marquês de Pombal – Rato	700

Table A.3 – Distances between metro stations for Green Line.

Stations	**Distance [m]**
Telheiras – Campo Grande	700
Campo Grande – Alvalade	1 400
Alvalade – Roma	500
Roma – Areeiro	1 000
Areeiro – Alameda	600
Alameda – Arroios	400
Arroios – Anjos	800
Anjos – Intendente	500
Intendente – Martim Moniz	700
Martim Moniz – Rossio	300
Rossio – Baixa-Chiado	400
Baixa-Chiado – Cais do Sodré	700

Table A.4 – Distances between metro stations for Red Line.

Stations	**Distance [m]**
S.Sebastião – Saldanha	800
Saldanha – Alameda	1 000

Alameda – Olaias	900
Olaias – Bela Vista	1 000
Bela Vista – Chelas	1 000
Chelas – Olivais Olivais – Cabo Ruivo	600 700
Cabo Ruivo – Oriente	700
Oriente – Moscavide	900
Moscavide – Encarnação	1 100
Encarnação – Aeroporto	1 300

Table A.5 – Average, Maximum and Minimum distance between Lisbon metro stations.

Average [m]	**Maximum [m]**	**Minimum [m]**
780	1400	300

Medir distância

Clique no mapa a adicionar ao seu percurso

Distância total: 1,30 km (4 267,91 pés)

Figure A.1 – Method used to discover distances between Lisbon metro stations using Google Maps Distance Calculator (in the example one can see distance between Aeroporto and Encarnacão).

Pontinha

Odivelas

Tempo estimado 00:46

Distância 15.2 km

Origem: Pontinha

Destino: Odivelas

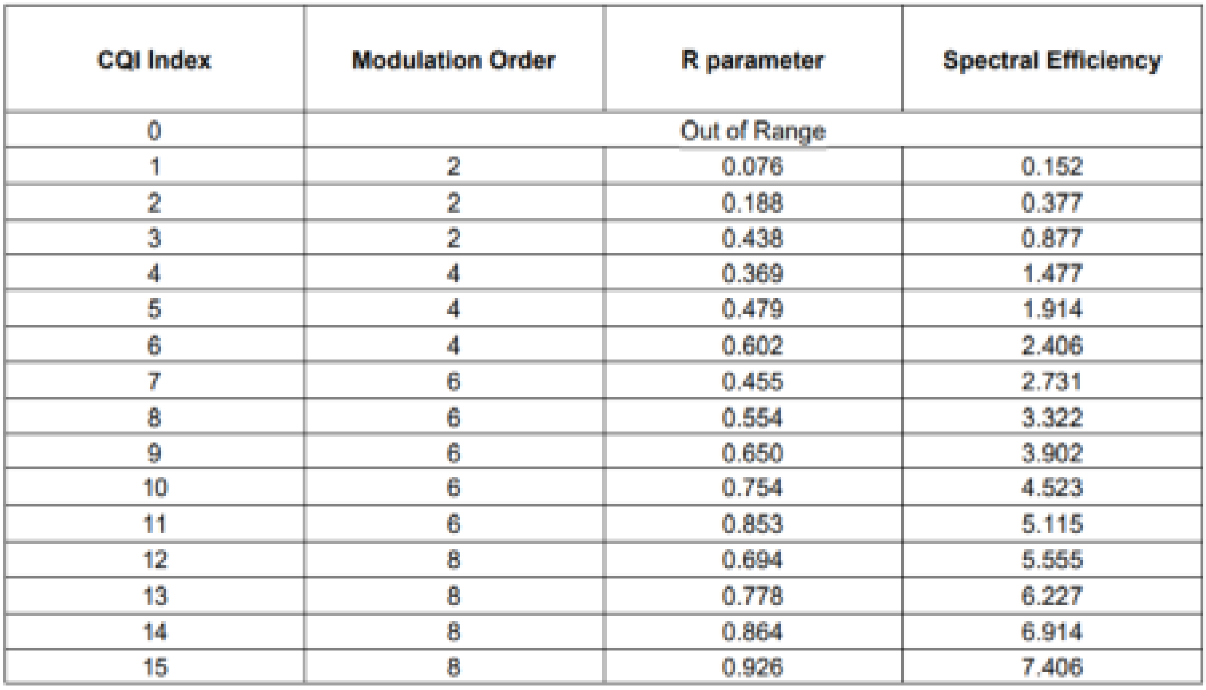
Figure A.2 – Maximum distance from a UE to the core of the network located in Pontinha.

Annex B

Code Rate, Modulation and CQI

This Annex presents the list of options for the CQI indexes, modulation orders and R parameter.

Table B.1 – CQI indexes, Modulation orders and R parameter list (extracted from [Carv22]).



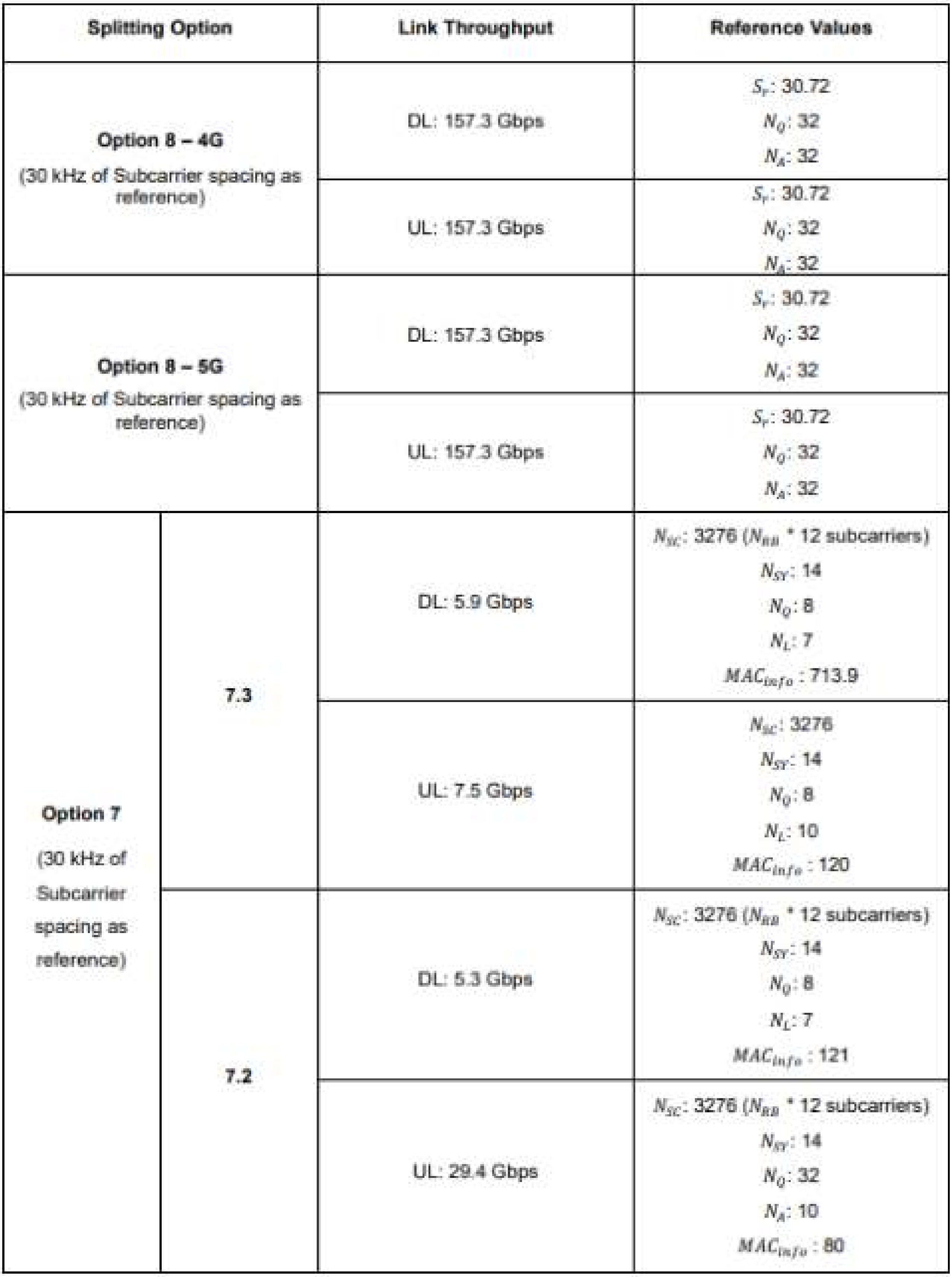
CQI Index	Modulation Order	R parameter	Spectral Efficiency
0	Out of Range		
1	2	0.076	0.152
2	2	0.188	0.377
3	2	0.438	0.877
4	4	0.369	1.477
5	4	0.479	1.914
6	4	0.602	2.406
7	6	0.455	2.731
8	6	0.554	3.322
9	6	0.650	3.902
10	6	0.754	4.523
11	6	0.853	5.115
12	8	0.694	5.555
13	8	0.778	6.227
14	8	0.864	6.914
15	8	0.926	7.406

Annex C

5G Link Throughputs

This Annex presents the link throughputs and parameters for different splitting options.

Table C.1 – 5G Link Throughputs according to different splitting options (extracted from [Carv22]).



Splitting Option		Link Throughput	Reference Values
Option 8 – 4G (30 kHz of Subcarrier spacing as reference)		DL: 157.3 Gbps	S_r: 30.72 N_Q: 32 N_A: 32
		UL: 157.3 Gbps	S_r: 30.72 N_Q: 32 N_A: 32
Option 8 – 5G (30 kHz of Subcarrier spacing as reference)		DL: 157.3 Gbps	S_r: 30.72 N_Q: 32 N_A: 32
		UL: 157.3 Gbps	S_r: 30.72 N_Q: 32 N_A: 32
Option 7 (30 kHz of Subcarrier spacing as reference)	**7.3**	DL: 5.9 Gbps	N_{SC}: 3276 (N_{RB} * 12 subcarriers) N_{SY}: 14 N_Q: 8 N_L: 7 MAC_{info} : 713.9
		UL: 7.5 Gbps	N_{SC}: 3276 N_{SY}: 14 N_Q: 8 N_L: 10 MAC_{info} : 120
	7.2	DL: 5.3 Gbps	N_{SC}: 3276 (N_{RB} * 12 subcarriers) N_{SY}: 14 N_Q: 8 N_L: 7 MAC_{info} : 121
		UL: 29.4 Gbps	N_{SC}: 3276 (N_{RB} * 12 subcarriers) N_{SY}: 14 N_Q: 32 N_A: 10 MAC_{info} : 80

	7.1	DL: Same as option 7.3	Same as option 7.3
		UL: Same as option 7.2	Same as option 7.2
Option 6 (250 Mbps of throughput as reference)		DL: 6.8 Gbps	R_p: 250 R_c: 5 B: 100 B_c: 20 N_L: 8 $N_{L,C}$: 2 M: 256 M_c: 64
		UL: 8.4 Gbps	R_p: 96 R_c: 44 B: 100 B_c: 20 N_L: 8 $N_{L,C}$: 1 M: 64 M_c: 16
Option 2 (250 Mbps of throughput as reference)		DL: 6.7 Gbps	R_p: 250 B: 100 B_c: 20 N_L: 8 $N_{L,C}$: 2 M: 256 M_c: 64
		UL: 5.0 Gbps	R_p:83 B:100 B_c: 20 N_L: 8 $N_{L,C}$: 1 M: 64 M_c: 16

References

[3GPP18] 3GPP, *User Equipment (UE) radio transmission and reception*, Internal Report TS 138 101-1, Release 15, V15.3.0, Oct. 2018. Available: https://www.etsi.org/deliver/etsi_ts/138100_138199/13810101/15.03.00_60/ts_13810101v150300p.pdf

[3GPP19] 3GPP, *System Architecture for the 5G System, Internal Report TS 23.501*, Stage 2 (Release 15), V15.3.0, Sep. 2019. Available: https://www.etsi.org/deliver/etsi_ts/123500_123599/123501/15.03.00_60/ts_123501v150300p.pdf/

[3GPP20] 3GPP, *Mobile communication system for railways*, Release 16, Nov. 2020. Available: https://www.etsi.org/deliver/etsi_ts/122200_122299/122289/16.01.00_60/ts_122289v160100p.pdf.

[3GPP21a] 3GPP, *Service requirements for the 5G system*, TS 22.261, Release 16, V16.14.0, April. 2021, Available: https://www.etsi.org/deliver/etsi_ts/122200_122299/122261/16.14.00_60/ts_122261v161400p.pdf

[3GPP21b] 3GPP, *Study on Future Railway Mobile Communication System*, Release 17, Mar. 2021. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3162

[3GPP22] 3GPP, *User Equipment (UE) radio access capabilities*, Release 17, May 2022. Available: https://www.etsi.org/deliver/etsi_ts/138300_138399/138306/17.00.00_60/ts_138306v170000p.pdf

[5Gra22] 5G Rail – Newsletter Number 1, June 2022, Available: https://5grail.eu/wp-content/uploads/2022/06/5GRAILNewsletter-No1_Final.pdf

[BGCB17] D.Bega, M.Gramaglia, C.Cano, A.Banchs, X.Perez, *Toward the network of the future: From enabling technologies to 5G concepts,* Research Article, June 2017, Available :https://onlinelibrary.wiley.com/doi/10.1002/ett.3205

[Carv22] A.Carvalho, *Analysis of Strategies for Minimising End-to-End Latency in 4G and 5G Networks,* M.Sc. Thesis, Instituto Superior Técnico, Lisbon, Portugal,2021, Available: https://grow.tecnico.ulisboa.pt/wp-content/uploads/2022/03/Thesis_AfonsoC_vfinal.pdf

[Corr19] M.Correia, *Analysis of the Influence of Pantographs in Railway Communications,* Instituto Superior Técnico, Lisboa, Portugal, 2019, Available: https://grow.tecnico.ulisboa.pt/wp-content/uploads/2020/01/Disserta%C3%A7%C3%A3oMagaliCorreiap%C3%B3s_defesa.pdf

[Corr20] L.Correia, *Mobile Communication Systems*, Lecture Notes, Instituto Superior Técnico, University of Lisbon, Lisbon, Portugal, Feb. 2020.

[DBRa20] DB Rail Academy, *ERTMS was an initiative of the European Union to promote easier border crossings in the railway sector*, Oct. 2020, Available: https://railacademy.deutschebahn.com/en/news/ertms-was-initiative-european-union-promote-easier-border-crossings-railway-sector

[DOCO16] NTT DOCOMO, *R3-162102: CU-DU Split: Refinement for Annex A*, 3GPP, Sophia Antipolis, France, Oct. 2016.

[Domi19] S.Domingues, *Analysis of the Performance of Multi-Access Edge Computing Network Slicing in 5G*, M.SC thesis, Instituto Superior Técnico, Lisbon, Portugal, Nov.2019. Available: https://grow.tecnico.ulisboa.pt/wpcontent/uploads/2020/07/Thesis_SergioD_vPublic.pdf.

[ERAS14] Final report for the European Railway Agency, *Survey on operational communications (study for the evolution of the railway communications system),* Feb. 2014, Available: https://www.era.europa.eu/studies?page=1

[Eric22a] Ericsson - 5G-powered FRMCS, Mar. 2022. Available: https://www.ericsson.com/4953a2/assets/local/reports-papers/white-papers/5g-powered-frmcs.pdf

[Eric22b] Ericsson - Keeping railways on track with FRMCS, Jan. 2022, Available: https://www.ericsson.com/en/blog/3/2022/1/keeping-railways-on-track-with-frmcs,

[Eric22c] Ericsson - How railways benefit from 5G, Oct. 2022, Available: https://www.ericsson.com/en/blog/2022/10/ericsson-booth-tour-at-innotrans-2022

[Eric22d] Ericsson - Railway communications on the express track: Wide area networks for digital rail, Mar. 2022. Available: https://www.ericsson.com/en/blog/2022/3/railway-communications-on-the-express-track

[ERTM22] ERTMS – ERTMS in brief, https://www.ertms.net/about-ertms/ertms-in-brief/, Dec. 2022

[EuCo20] European Commission, Harmonisation of the 5.9 GHz spectrum band for real-time information exchange will improve road and urban rail transport safety, Oct. 2020. Available:https://digital-strategy.ec.europa.eu/en/news/harmonisation-59-ghz-spectrum-band-real-time-information-exchange-will-improve-road-and-urban-rail

[GSAP22] GSA, *Private-Mobile-Networks December-2022 Summary Report*, Internal Report, Dec. 2022, Available: https://gsacom.com/paper/private-mobile-networks-december-2022-summary-report/

[GSMA21] GSM Association, *E2E Network Slicing Architecture*, 3 June 2021, Available: https://www.gsma.com/newsroom/wp-content/uploads/NG.127-v1.0-2.pdf

[GSMR23] UIC - GSM-R, https://uic.org/rail-system/gsm-r/, Jan. 2023.

[ITUR15] Recommendation ITU-R M.2083-0, *IMT Vision – Framework and overall objectives of* *the future development of IMT for 2020 and beyond*, Sep. 2015, Available: http://www.itu.int/rec/R-REC-M.2083-0-201509-I/en

[ITUT18] ITU-T, *Transport network support of IMT-2020/5G*, Feb. 2018, Available: https://www.itu.int/dms_pub/itu-t/opb/tut/T-TUT-HOME-2018-PDF-E.pdf

[LaCC19] L.Larsen, A.Checko and H.Christiansen, “A Survey of the Functional Splits Proposed for 5G Mobile Crosshaul Networks”, *IEEE Communications Surveys & Tutorials*, vol. 21, no. 1, pp. 146-172, Firstquarter 2019, doi: 10.1109/COMST.2018.2868805.

[MGEl18] D.Maaz, A.Galindo and S.Elayoubi, *URLLC User Plane Latency Performance in New Radio*, 2018 25th International Conference on Telecommunications (ICT), Saint-Malo, France, 2018, pp. 225-229, doi: 10.1109/ICT.2018.8464912.

[RSIF21] J.Rischke, P.Sossalla, S.Itting, F.Fitzek and M.Reisslein, “5G Campus Networks: A First Measurement Study”, *IEEE Access*, vol. 9, pp. 121786-121803, 2021, doi: 10.1109/ACCESS.2021.3108423.

[VODA18] Vodafone, *Overview and Predictive Analysis for Latency Optimized Telecommunication Networks*, Hochschule Rhein Main Russelsheim University, Russelsheim, Germany, Oct.2018. Available: https://www.hsrm.de/fileadmin/persons/khofmann/Gastvortraege/Vortragsfolien/20181026-BurkLemberg-vodafone-5G.pdf

[WiCh22] Ufispace, *Private 5G - Benefits, Architectures, and Requirements*, 30 June 2022, Will Chang, Available: https://www.ufispace.com/company/blog/what-is-private-5g-benefits-architectures-requirements

[WiSt21] W.Stallings, *5G Wireless: A Comprehensive Introduction,* Addison-Wesley Professional, Massachusetts, United States, 2021.

[Wolf18] A. Wolf, *D4.1 Technological Forces and Railway Performance Conditions, Technical Viability Analysis, Mistral – Communication Systems for Next-Generation Railways*, Apr. 2018. Available: https://projects.shift2rail.org/download.aspx?id=a2d32c6b-b31e-4de8-b484-ccfaf32d7f8c