Universidade de Aveiro
2022

Pedro Valente Mateus

Desagregação do Core Celular para Serviços no Edge da Cidade

Disaggregated Mobile Core for Edge City Services

Universidade de Aveiro
2022

Pedro Valente Mateus

Desagregação do Core Celular para Serviços no Edge da Cidade

Disaggregated Mobile Core for Edge City Services

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Engenharia de Computadores e Telemática, realizada sob a orientação científica do Doutor Pedro Filipe Vieira Rito, Investigador Auxiliar do Instituto de Telecomunicações de Aveiro, do Doutor Duarte Miguel Garcia Raposo, Investigador Auxiliar do Instituto de Telecomunicações de Aveiro, e colaboração da Doutora Susana Isabel Barreto de Miranda Sargento, Professora catedrática do Departamento de Eletrónica, Telecomunicações e Informática da Universidade de Aveiro.

Dedico este trabalho aos meus avós por me terem acompanhado em todas as fases da minha vida, os meus segundos pais.

o júri / the jury

presidente / president	Professor Doutor António José Ribeiro Neves Professor Auxiliar, Universidade de Aveiro
vogais / examiners committee	Professor Doutor Filipe Marques da Silva Cabral Pinto Professor Adjunto Convidado, Instituto Politécnico de Viseu - Escola Superior de Tecnologia de Viseu
	Doutor Pedro Filipe Vieira Rito Investigador Doutorado, Instituto de Telecomunicações da Universidade de Aveiro

agradecimentos / acknowledgements

Não enumerarei nomes nestes agradecimentos pois muitos passaram pela minha vida em diferentes fases tendo cada um a sua influência, maior ou menor. Apenas deixarei o meu obrigado a todos os que estiveram presentes nestes cinco anos. O meu obrigado a quem me ajudou nas fases escuras e me permitiu ser eu mesmo. O meu obrigado a quem esteve nas melhores fases e não permitiu que piorassem. Um obrigado à minha familia por todo o apoio e paciência ao longo da minha vida. Deixo também um agradecimento ao NAP do Instituto de Telecomunicações por todo o apoio na aprendizagem de um tópico que me era até agora desconhecido. Por fim, agradeço ao Instituto de Telecomunicações pelas condições de financiamento e acolhimento. Este trabalho é financiado pelo projeto S-0009-AV-20, fiQare, fiQare - Future Intelligent Quality Assurance for Release Enhancement, com o apoio da Ubiwhere. Este trabalho é também financiado pelo Fundo Europeu de Desenvolvimento Regional (FEDER), através do Programa Operacional Regional do Centro (CENTRO 2020) do Portugal 2020 e por Fundos Públicos Nacionais através da FCT I.P. (OE) [Projeto SNOB-5G com o nº 045929 (CENTRO-01-0247-FEDER-045929)].

Palavras Chave 5G, Rede do Core, Desagregação, Edge, Serviços.

Resumo

As redes de comunicações móveis evoluíram significativamente ao longo dos anos. A evolução atual (5G) enfrenta um aumento do número de equipamentos e de aplicações com alto consumo de dados, crescendo assim a necessidade de uma densificação da rede. Para resolver este problema, e otimizar a infra-estrutura de uma rede de comunicações 5G, deverá levar-se a cargo o desenvolvimento de uma rede unificada e partilhada, para que nesta possam ser implementadas redes de múltiplos operadores. Com uma rede planeada e partilhada pelos operadores, começam a ser pensadas diferentes formas de tirar o máximo partido da arquitectura. Considerando vários serviços, como Vehicle-to-Everything (V2X) e Internet-of-Things (IoT). É aqui que surgem tópicos como Multi-Acess Edge Computing (MEC), que podem fornecer formas de maximizar a exploração de uma infra-estrutura de cidade de forma a reduzir latências e remover esforço computacional dos servidores na cloud para o edge. Esta dissertação centra-se no desenvolvimento de uma arquitectura neutra para redes 5G com Multi-acess Edge Computing (MEC), com o propósito de ser implementada numa cidade inteligente. No caso concreto desta dissertação, a arquitetura será preparada para ser implementada na cidade de Aveiro, tirando partido dos múltiplos edge nodes e serviços que já possui. A desagregação da Rede do Core (CN) é um ponto fulcral desta dissertação, criando uma forma de ter a gateway mais próxima do edge, permitindo assim que tecnologias MEC melhorem a experiência do utilizador e a performance dos serviços existentes. Foi concebida e implementada uma testbed real para simular com precisão o ambiente da cidade e comprovar a arquitetura concebida para os serviços mencionados. Os cenários de teste envolveram a comparação de um serviço instanciado na cloud e no edge, cenários de mobilidade com handover para verificar a relocalização do user plane e, por último, a garantia de Qualidade de Serviço (QoS) para serviços de maior prioridade num cenário de ocupação de toda a largura de banda. Os resultados obtidos demonstram melhorias tanto no tráfego comum do utilizador como no tráfego orientado a serviços; além disso, este trabalho também descreve e testa as capacidades do core utilizado, bem como as suas limitações, proporcionando uma base para trabalhos futuros.

Keywords 5G, Core Network, Disaggregation, Edge, Services.

Abstract Mobile communication networks evolved staggeringly throughout the years. Current evolution (5G) needs a denser network, due to new vertical-based data-hungry applications and the increase of UEs, leading to additional costs in cellular deployment. To solve this issue and optimize the deployment of 5G, one unified network could be devised and shared by multiple operators to deploy their own networks. With a city network planned, ways to take full advantage of a neutral hosting architecture begin to be executed, which serve various services, like Vehicle-to-Everything (V2X) and Internet-of-Things (IoT). This is where topics such as Multi-Access Edge Computing (MEC) arise, by maximizing the exploitation of the city infrastructure to reduce latency and remove the computational effort from cloud servers to the edge. This thesis focuses on developing a neutral architecture for 5G networks, with Multi-acess Edge Computing (MEC), which can be deployed in a city. In the concrete case of this thesis, this architecture will be deployed in the city of Aveiro, taking advantage of its multiple edge nodes. The disaggregation of the Core Network (CN) is a focal point of this thesis, creating a way to have a gateway closer to the edge, thus enabling MEC technologies to improve end-user and service experience. A disaggregated architecture was proposed, and extended to the city infrastructure, to accurately simulate the city environment and to attest to the city services. The test scenarios involved the comparison of a service instantiated in the cloud and in the edge, mobility scenarios with a handover through the same network to attest user plane relocation, different user plane selection, and lastly, different flow level priority assessment, to higher priority services in a full bandwidth occupation scenario. The obtained results show that the deployment of the user plane in the edge brings significant improvements, both in common user traffic and service-oriented traffic; moreover, this outlines the capabilities of this core solution as well as its limitations providing a foundation for future works.

Contents

List of Figures

List of Tables

Glossary

1G first generation

2G second generation

3G third generation

3GPP 3G Partnership Project

4G fourth generation

5G fifth generation

5GC 5G Core

5GPPP 5G Public-Private Partnership

5QI 5G QoS Identifier

AF Application Function

AMF Access and Mobility Management Function

AN Access Network

APN Access Point Name

ARP Allocation and Retention Priority

AS Access Stratum

ATCLL Aveiro Tech City Living Lab

AUSF Authentication Server Function

BBU Baseband Unit

BSF Binding Support Function

CDN Content Delivery Network

CN Core Network

CP Control Plane

C-RAN Cloud RAN

CU Control Unit

CUPS Control and User Plane Separation

C-V2X Cellular Vehicle to Everything

DN Data Network

DNN Data Network Name

DNS Domain Name Server

DU Distributed Unit

EARFCN E-UTRA Absolute Radio Frequency Channel Number

eMBB enhanced Mobile BroadBand

eNB eNodeB

en-gNB enhanced gNodeB

EPC Evolved Packet Core

EPS Evolved Packet System

E-UTRAN Evolved Universal Terrestrial Radio Access Network

FCCN Fundação para a Computação Científica Nacional

FDD Frequency Division Duplex

GBR Guaranteed Bit Rate

gNB gNodeB

gNB-CU gNodeB Control Unit

gNB-DU gNodeB Distributed Unit

gNB-RU gNodeB Radio Unit

GPRS General Packet Radio Service

GSM Global System for Mobile communication

GTP-C GPRS Tunnelling Protocol Control Plane

GTP-U GPRS Tunnelling Protocol User Plane

HSS Home Subscriber Server

IMS IP Multimedia Subsystem

IoT Internet-of-Things

IPFS InterPlanetary Fyle System

LPWAN Low Power Wide Area Network

LTE Long Term Evolution

MAC Medium Access Control

MEC Multi-access Edge Computing

MME Mobile Management Entity

MMS Multimedia Messaging Service

mMTC massive Machine Type Communications

MOCN Multi Operator Core Network

MORAN Multi Operator Radio Access Network

NAS Non-Access Stratum

NB-IoT NarrowBand-IoT

NF Network Function

NG Next Generation

NGAP Next Generation Application Protocol

Non-GBR Non Guaranteed Bit Rate

NG-eNB Next Generation eNodeB

NG-RAN Next Generation Radio Access Network

NR New Radio

NRF Network Repository Function

NS Network Slicing

NSA Non Standalone

NSSF Network Slice Selection Function

OAI OpenAirInterface

OBU Onboard Unit

Open-RAN Open Radio Access Network

PCF Policy Charging Function

PCRF Policy Charging Rules Function

PDCP Packet Data Convergence Protocol

PDN Packet Data Network

PDU Protocol Data Unit

PFCP Packet Forwarding Control Protocol

P-GW Packet data network gateway

PGW-C Packet data network GateWay - Control plane

PGW-U Packet data network GateWay - User plane

PLMN Public Land Mobile Network

QCI QoS Class Identifier

QoS Quality of Service

RAN Radio Access Network

RIC RAN Intelligent Controller

RLC Radio Link Control

RRC Radio Resource Control

RRH Remote Radio Head

RRM Radio Resource Management

RSRP Reference Signal Received Power

RTT round trip time

RU Radio Unit

S1AP S1 Application Protocol

SA Standalone

SBA Service Based Architecture

SBI Service Based Interface

SCTP Stream Control Transmission Protocol

SD Slice Differentiator

SDN Software-Defined Network

SG Smart Grid

S-GW Serving GateWay

SGWC Serving Gateway Control Plane

SGWU Serving Gateway User Plane

SMF Session Management Function

SMO Service Management and Orchestration

SMS Short Message Service

S-NSSAI Single Network Slice Selection Assistance Information

SSL Security Sockets Layer

SST Slice/Service Type

TAC Tracking Area Code

TAI Tracking Area Information

TCP Transmission Control Protocol

TDD Time Division Duplex

TEID Tunnel Endpoint Identifier

UDM Unified Data Management

UDP User Datagram Protocol

UDR Unified Data Repository

UE User Equipment

UMTS Universal Mobile Telecommunication System

UP User Plane

UPF User Plane Function

URLLC Ultra Reliable Low Latency Communication

V2X Vehicle-to-Everything

VNF Virtual Network Function

VoIP Voice over IP

VoLTE Voice over LTE

vRAN virtualized RAN

CHAPTER 1

Introduction

1.1 Motivation

In the current days, most of the mobile traffic is generated indoors, whether through crowded areas such as concerts, hospitals, sports, universities, stations, shopping malls, or enterprises. According to Cisco [1] this composes 80% of all traffic and is rising at a pace of 20% every year. On the other hand, the introduction of the Industry 4.0 campaign has completely revolutionized industries and factories towards cyber-physical systems and Internet-of-Things (IoT) systems [2]. One popular solution to these problems has been network densification, but it is still difficult to provide uniform cellular coverage for different operators. This originates from the possible attempt of different operators to use the same cell sites, overwhelming a location with their infrastructure. Additionally, these circumstances might signify convoluted and ineffective arrangements, brought on by many operators' infrastructure deployments on the same locations. Therefore, creating a single common network architecture, that all wireless carriers could use, would be a more enticing approach. One of the popular solutions to these issues is an independent access provider operating as a host, for wireless connectivity who is not affiliated with any one of the individual cellular providers [3]. In line with an article from Nokia[1], the company that owns more cell towers in America is not any mobile network operator, but an independent company which then rents the network infrastructure. This Neutral Host approach allows operators to invest more into their business, in areas such as accelerating rollouts of new services, improving network coverage and spread the costs of the infrastructure along the years, as they do not need to make the investment in the infrastructure. Under 5G Public-Private Partnership (5GPPP), multiple projects arose to enable small cell multi-tenancy network operators[2], which can be a target not only to a complete city scenario, but also for an in-building cellular connection. Nonetheless, for a network to be able to host other operators, it must first be prepared to provide a stable infrastructure for the area it is covering.

[1] https://www.nokia.com/networks/insights/neutral-hosts-path-to-5G-profitability/
[2] https://5g-ppp.eu/5g-city/

For the infrastructure to be city-prepared, it must always take into account the services it will provide. Cloud computing is one way to tackle this issue, providing rich computing resources in remote servers, however at a cost of non-negligible latency. The 5G of the mobile network implies capabilities for Ultra Reliable Low Latency Communication (URLLC) at both Radio Access Network (RAN) and CN, but a fully Standalone (SA) deployment of a 5G network is still not a possibility for all. Therefore, a need for a technology that can reduce end-to-end latency is in place, and along with it Multi-access Edge Computing (MEC), which will be key for the provision of services closer to the users. MEC is proposed to provide computing, storage, and communication functionalities at the network edge and therefore alleviate the burdens on cloud servers and provide close-to-the-expected latency values [4], as well as bringing application-oriented capabilities into the CN. It provides itself as a universal access technology to any application that has locality requirements, such as those used in sports arenas or retail malls, or anywhere low latency is needed, such as in 5G Vehicle-to-Everything (V2X), or autonomous car applications [5]. Fourth generation (4G) is still expected to be very much presented in the years to come, therefore common work will be made to make MEC technologies work in 4G architectures, always having in mind that it can be reused to support 5G as well.

This dissertation will focus on deploying a cellular network capable of being scaled to a city-wide deployment, having in mind its needs and the integration of other operators. This thesis will provide the core disaggregation for the neutral hosting and edge-based services, and an effort will be made to have an end-to-end network prepared to take full advantage of its edge nodes and enable MEC scenarios in the resulting approach.

1.2 Objectives and Contributions

The main goal of this thesis is to define and implement a city ready end-to-end disaggregated mobile network compliant with 3G Partnership Project (3GPP) and Open Radio Access Network (Open-RAN) standards to enable edge services. The resulting network will provide an easy to understand architecture that provides Quality of Service (QoS) to designated users, and reliable latency values when compared to cloud solutions. This thesis has the following objectives:

- Characterize the different needs of a city's infrastructure with respect to the new services to be deployed;
- Evaluate the designated tools to deploy an end-to-end mobile network;
- Deploy the aforementioned end-to-end network with an architecture ready to enable MEC scenarios, pertinent to the city services;
- Test the network in different scenarios to confirm the solution within the planned use cases.

This dissertation accomplished the deployment of an end-to-end mobile network with the core disaggregated in a MEC environment. It studied how the used CN and RAN tools can provide Neutral Host functionalities, and it developed the architecture of the network thinking on a set of city services. It managed to successfully disaggregate the User Plane (UP), thus enabling MEC scenarios in which this works statement's was proved, with a set of tests to the infrastructure, a traffic generator and a real use case with IPFS.

1.3 Thesis Outline

This document is divided into six chapters, beginning with chapter 1, Introduction, which this section concludes. Following, chapter 2, State-of-the-Art, presents background theoretical concepts required to understand the topics in discussion throughout the work (e.g., 5G Access and Core Networks, coexistence of 4G and 5G technologies, and Edge Services), and some related work. Chapter 3, A City-Scale Architecture, explains the need of a common city infrastructure, ready for multiple operators and distinct services, and the current status in research and development of works in this matter. Chapter 4, Real Deployment, demonstrates the tools used for the deployment of the access and core network, and how they act together to implement the proposed approach. Chapter 5, Experimentation and Results, demonstrates the functionality of the proposed approach in scenarios designed to illustrate the city environment, and discusses the obtained results. Chapter 6, Conclusion and Future Work, summarizes the discussion of the obtained results in direct answer to the objectives stated in Chapter 1, and includes possible research directions for the future.

CHAPTER 2

State-of-the-Art

Mobile communication networks evolved tremendously throughout the years, from the first two generations, only supporting voice and text services, to the transition to broadband access. At the time of writing, the 4G is still the most dominant technology, with data rates measured at megabits per second. However, the demand keeps growing, meaning that there is a need to make advancements to new technologies, that is, the 5G of mobile networks.

This new generation expects to transition from a single device to an agglomeration of edge devices and services. This also comes with the need of a much lower latency, an increase in speed and bandwidth, denser networks that allow more devices in the same physical area as before, reduced power consumption, and increased security. The whole mobile communications ecosystem revamped with this new generation, from the RAN to the CN. The RAN saw its giant antennas converted into small cells, and a possibility to split its components into various parts, allowing each component to form a specific and independent objective. Permitting each vendor to work on a specific area instead of the whole solution, from end to end, the concept of Open-RAN was created, in which each component has its interfaces standardized letting that different vendors' products work together. In the CN, the architecture changed to a Service Based Architecture (SBA) implementing network principles and a cloud-native design approach. Each Network Function (NF), formed by a combination of software code called microservices, offers one or more services to the other NFs. As all these Network Functions (NFs) are composed of software, they provide a flexible deployment that eases the changes to the edge scenarios previously mentioned. Moreover, a new concept of Network Slicing (NS) is presented, in which each Network Function is solely deployed to serve a specific service. Each slice is dedicated to a service according to its requirements, whether it be low latency, high bandwidth, or mobility [6], [7].

This chapter introduces the evolution of cellular networks in section 2.1, all from the first generation to the current fifth generation, explaining what advancements each generation brought and new features, as well as current architectures and their purposes. Section 2.2 explains concepts of 4G and 5G Radio Access Network (RAN), the evolution that these areas saw from one generation to the other and why they were needed. It will also introduce the concept of Open-RAN, why it is needed, and its architecture. The next section (2.3) presents the CN to demonstrate the new modules of the architecture and their new capabilities and functionalities. Since this thesis focuses in both Long Term Evolution (LTE) and Next Generation (NG) (5G) technologies, it should be reviewed how these two can coexist, this is reviewed in section 2.4. Finally, section 2.5 gives an insight into how Multi-access Edge Computing (MEC) technologies can bring an improvement to various services and how the CN is deployed to serve this type of approach. Section 2.8 summarizes the chapter.

2.1 Cellular networks

The need and demand for mobile and Internet are increasing every day. The mobile industry has initiated the creation of technology and the evolution of the mobile industry since the early 1970s [8]. This had the intent to provide mobile connectivity to everyone. Since then, cellular networks have noticeably evolved, going through various generations to represent continuous achievements in mobile technology, as seen in Figure 2.1. This section will introduce the various generations of cellular networks and explain their key features.

Beginning with the first generation (1G), in the 1970s, as afore mentioned, commercial cellular networks were introduced, having fully implemented standards established ten years later. The radio signals used in this generation were analog, which means that the voice calls were modulated to a higher frequency instead of encoding them into digital signals. The problem with these analog signals is that they degrade over time and space. This led to reliability issues, signal interference, and a lack of security against eavesdroppers. Digital representation of analog stored as signals allows for more significant amounts of data to be carried out more effectively.

Following, second generation (2G) came, and its successor 2.5G, which improved some aspects of data transfer and the Internet. The second generation, introduced in the late 1990s, used digital signaling and Global System for Mobile communication (GSM) technology

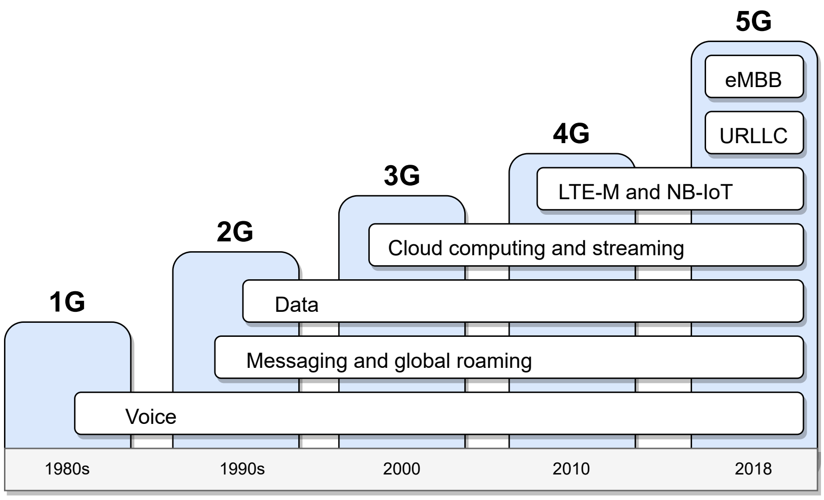


Figure 2.1: Cellular network evolution throughout time

to introduce digital voice and data to be sent across the network. This ultimately allowed end-users to roam for the first time, and improved security and phone calls. This new technology also introduced new concepts such as Short Message Service (SMS), Multimedia Messaging Service (MMS), internal roaming, conference calls, call hold, and others. Finally, 2.5G introduced the packet network, providing high-speed data transfer and Internet. More precisely, these new standards included General Packet Radio Service (GPRS), in 2.5G, and EDGE (enhanced data rates in GSM), in 2.75G. The latter one supporting flexible data transmission rates and providing continuous connection, receiving and sending e-mail messages. Another improvement on the service provider side was that the providers were capable to charge for the amount of data sent, rather than their connection time.

Third generation (3G), known as Universal Mobile Telecommunication System (UMTS) was introduced in 2010, set out to ease greater voice and data capacity, and support a more comprehensive range of different applications, and better data transmission at a lower cost. Besides higher speeds and broader bands, this generation brought fixed wireless Internet access, video calls, chatting, conferencing, and others. It also improved the security significantly within the 3G, as well as including network access, domain security, and application security [9], [10].

2.1.1 The evolution of 4G

4G is the first generation to use LTE technology, created by the 3GPP, to attempt to deliver higher download speeds as well as better latency, instant messaging services, better voice quality, better download speeds, and others. Besides, it is also the first IP-based mobile network, handling voice as just another service through Voice over LTE (VoLTE) and the beginning of the attempt to offer QoS technologies. One of the significant benefits of the IP-based network stated before is the ability to seamlessly handover for both voice and data to the technologies from the previous different generations' infrastructure.

At a high level, the network is constituted by a Core Network (CN), Evolved Packet Core (EPC), and an Access Network, Evolved Universal Terrestrial Radio Access Network (E-UTRAN). The Access Network is constituted by a single node, the eNodeB (eNB) that connects to the UEs. The CN is made up by many logical nodes, responsible for the control of the UE and establishment of the bearers, since it only provides a bearer path of a certain QoS. An external node called IP Multimedia Subsystem (IMS) is used to control multimedia applications, such as Voice over IP (VoIP). The CN nodes are:

- **MME** - the Mobility Management Entity is the main control and signaling processing node, responsible for initiating paging and authentication of the UE. The protocols that run between the UE and CN are known as Non-Access Stratum (NAS) protocols. It retains information on the location of each user selecting the appropriate gateway during the initial registration process, at the tracking area level. It is vital in handovers between LTE and lower generation networks. Signaling load can be increased by grouping multiple MMEs, which is done to support more end-users;

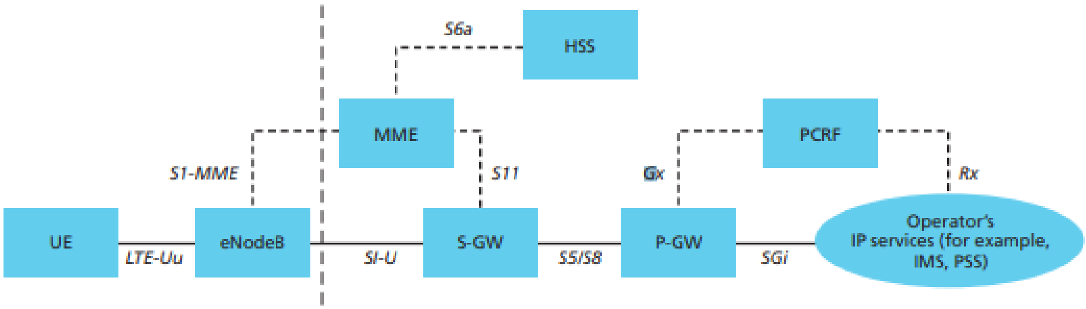


Figure 2.2: Evolved Packet System Network architecture [11]

- **HSS** - the Home Subscriber Server holds the user's subscription data, such as QoS profiles, access restrictions for roaming, and the Packet Data Networks (PDNs) accessible to each user. This is done according to a specific Access Point Name (APN) or a PDN address (indicating subscribed IP address). Additionally, it dynamically holds information on which MME the user is attached;
- **PCRF** - the Policy Control and Charging Function is responsible for decision-making, policy control, and control of the flow-based charging functionalities in the Packet data network gateway (P-GW). It ensures that a user's subscription profile has its requirements met;
- **S-GW** - Serving Gateway serves as the local mobility anchor for the UE and the data bearers as the user moves from eNodeB to eNodeB. It deals with the UEs, not only in mobility scenarios, but also when it goes idle, temporarily buffering downlink data. In contrast, the MME initiates paging of the UE to reestablish the bearers. It carries out some administrative functions in the visited network, like gathering information for charging and lawful interception. Besides this, it serves as the mobility anchor for internetworking with other 3GPP technologies like UMTS and GPRS;
- **P-GW** - The Packet Gateway or (PDN Gateway) is in charge of allocating IP addresses to the UEs, QoS enforcement and flow based charging according to the rules given by the Policy Charging Rules Function (PCRF), filtering the downlink user IP packets into bearers with distinct QoS rules.

The access network of LTE, E-UTRAN, is only constituted by eNBs. E-UTRANs architecture is usually called flat because there is no centralized controller in typical user traffic. The eNBs are connected to the EPC via `S1` interface, more precisely to the MME by the `S1-MME` interface and to the Serving GateWay (S-GW) by the `S1-U` interface, as depicted in the figure 2.2. Nevertheless, the eNBs also connect to each other via the `X2` interface. The protocols that run between UEs and eNBs are known as Access Stratum (AS) protocols [11]. The E-UTRAN user plane protocol stack consists of Packet Data Convergence Protocol (PDCP), Radio Link Control (RLC) and Medium Access Control (MAC) sublayers that terminate in the eNB. In mobility scenarios and in the absence of a controller node, the data buffer is performed by the eNB. Data protection is of the responsibility of the PDCP layer and the other two layers start afresh in a new cell.In the control plane the layers refered before perform the same functions, except there is no header compression function. The responsibility of establishment of radio headers and configuration of the lower layers falls upon the Radio Resource Control (RRC), which is done through signaling procedures between the eNB and the UE.

One of the advantages to previous generations of mobile networks is that the radio controller function is integrated into the eNB, allowing close interactivity in all protocol layers of the RAN, reducing latency and improving efficiency. Furthermore, having the controller in this distributed way, eliminates the need for intensive processing controllers or high availability, which diminishes costs and decreases the importance of a single node failure. Even though not having a centralized controller does bring a lot of advantages, it also brings some consequences, a few of them in mobility scenarios, in which the UE has to have all its related information transferred by the network from eNB to eNB. This forces the eNB to be prepared to avoid data loss during handovers.

According to the 3GPP standardization, UMTS and LTE present two different air interfaces [12]. The first one is Frequency Division Duplex (FDD) and the second Time Division Duplex (TDD). Network providers normally select only FDD in uplink and downlink. However, some providers chose to optimize resources by changing or adding TDD into their network, due to limited radio frequency resources. When both of these techniques are used together, it is called a coexistence network. Even though the two can coexist, the coexisting of the systems is complex, as interference occurs when a base station or UE tries to transmit or receive data at the same frequency or time. There are various coexistence scenarios, such as FDD-FDD, TDD-TDD and FDD-TDD. Each has its advantages, disadvantages, and optimization techniques to mitigate the interference between both types of networks, which are not specified in this document.

2.1.2 5G network

The world evolved, and with it, the necessities of services and end-users applications. New services with diverse and challenging performance requirements across all industries appeared, forcing a new generation of mobile communications. To face this, the 3GPP standards development organization created a new 5G system architecture, which includes a 5G New Radio (NR) access, and a new 5G Core (5GC) network, as depicted in figure 2.3. Studies suggest that services will be the main purpose of 5G architectures, separating each service type by technical requirements and economic models. By doing this separation, each service is developed solely to be optimized to its requirements. Some services' requirements and characteristics [13] can be seen below:

- Network operations;
- Critical communications applications;
- Enhanced Mobile BroadBand (EMBB);
- V2X;
- Massive Machine Type Communications (MMTC) or IoT.

These categories can be condensed and spread among three main 5G service verticals. These being eMBB, mMTC and URLLC [14]. The eMBB is the first phase of the 5G ecosystem and can be seen as a development of the already existing 4G services, its main requirements are: high data rates and low latency communications with the goal of achieving better QoS. The main services that it serves are mobile broadband, high definition videos, and virtual/augmented reality.

The mMTC primarily considers services and applications related to high density IoT devices, all with the ultimate goal to achieve connectivity solutions for smart cities. It is designed to be latency-tolerant, efficient for small amounts of data whether sent or received, and to work in low bandwidths. The main requirement is that mMTC can support high connection density, with in the order of one million devices per square kilometer, which is ten times what is currently possible with the 4G technology.

URLLC mainly takes into account latency-sensitive applications with very high dependability, like automated driving, industrial control, and tactile Internet. It is therefore focused on developing the digital industry. URLLC must provide low latency, minimum packet loss and data transmission times of 1ms. To achieve this, the development of high-performance gadgets is required, which will provide even bigger difficulties for electronic hardware designers and software programmers. For particular applications, software developers may need to use new protocols, coding approaches, and provisioning strategies to guarantee the highest level of security and dependability.

Current architectures are made to run on proprietary software meaning they are not exactly flexible. Also, such software can be very restricted to what the vendor develop, making their integration with third-party applications challenging. Another issue is that there can not be efficient resource utilization as these are not dynamic, and the type of service is usually not known. With all this in mind, the 3GPP defined that an Service Based Architecture, where the functionalities of the control plane and standard data repositories will be delivered as independent, renewable, and self-contained NFs, which will then connect to the desired service employing a Service Based Interface (SBI) with a Rest API through HTTP/2 and JSON. In this type of network architecture, services are deployed in a distributed, cloud-native environment that can provide failover and scaling mechanisms, such as Kubernetes, allowing the core to run on non-proprietary infrastructure. In such scenario, every NF can be deployed as Virtual Network Function (VNF), meaning that as software, it can run on virtual machines or containers efficiently. This VNF, in conjunction with a Software-Defined Network (SDN) allows for a customized service on a programmable platform implementation, permitting the software-based network control and management plane to decouple from the hardware-based forwarding plane. As opposed to 4G, all the elements from the 5G SBA are cloud-native, meaning that they are not nodes anymore, but functions, representing that they can interact

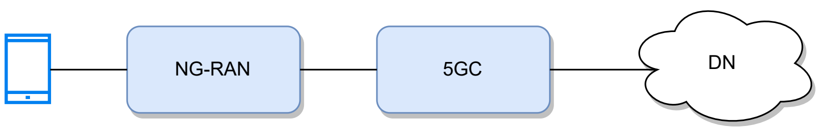


Figure 2.3: End-to-end 5G network

with each other. More details on the core architecture will be mentioned in subsection 2.3.1. This Network Function is grouped following a certain logic: being called a slice, each one has a specific purpose, following different QoS requirements and latency features. These slices will then be addressed to each one of the services mentioned above [15].

To support this new way of thinking in the core architecture the RAN had to be rethought. With this in mind, 3GPP created the Next Generation Radio Access Network (NG-RAN) that defines new principles, made to be universal:

- the data transport networks and the signaling should be logically separated;
- the 5GC NFs and the NG-RAN are completely distinguished from the transport functions;
- the mobility of the RRC protocol is entirely controlled by the NG-RAN;
- NG-RAN interfaces are controlled by a logical element, and multiple logical nodes can be implemented by a physical network element.

In NG-RAN there are two types of nodes that can be linked to a 5GC, being the gNodeB (gNB), and the Next Generation eNodeB (NG-eNB). The gNB uses New Radio protocols of Control Plane (CP) and UP to serve NR devices while the NG-eNB makes use of LTE CP and UP protocols to serve LTE devices. The connection between two gNBs is made through the `Xn` interface, that is resourceful as it implements Radio Resource Management (RRM) functions and by using packet forwarding can achieve a lossless mobility among neighbour cells. The gNB is constituted by two functional entities, the gNodeB Control Unit (gNB-CU) and the gNodeB Distributed Unit (gNB-DU) which are linked through the `F1` interface, the RAN topic will be revisited in the following section 2.2.

This section laid out a resume of cellular network evolution, going more in depth for 4G and 5G. In the following sections a bigger insight will be given to the AN and CN, presenting what 5G brought as innovation as well as its coexistence with the previous generation.

2.2 Access Network

The mobile communications market has begun to make the change from a well-established fourth generation to a new generation, and with it, new functionalities are being established to meet the requirements referenced in 2.1.2. With the need to move closer to the end-user to meet such requirements, the traditional RAN components have been divided and virtualized. This section will explore the evolution of RAN and why this evolution was needed, with a description of its modules in the two more recent generations of mobile communications. Moreover, there is an approach to the Open-RAN's architecture and general concept. Even though this work focuses more on the CN, the implemented testbed constitutes a Open-RAN-compliant RAN that is also 5G architecture-ready, therefore it is important to address the progress and work made in these technologies.

2.2.1 RAN evolution, from D-RAN to NG-RAN

With 4G the Radio Access Network evolved from having all radio and baseband processing functions at the same location to a distributed architecture, where both are separated. The

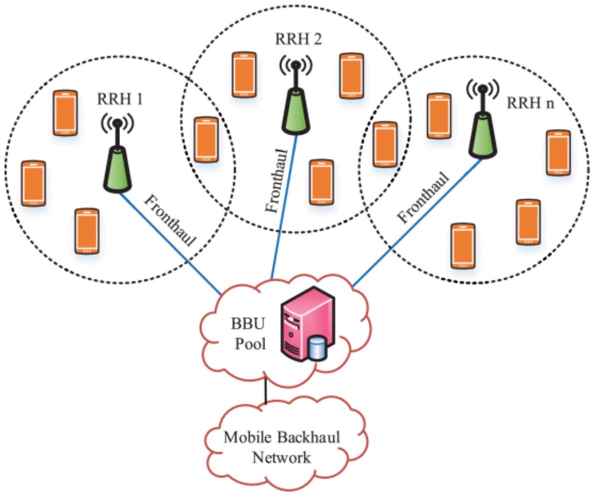


Figure 2.4: Cloud RAN architecture [16]

radio functions passed to the Remote Radio Head (RRH), and the signal processing functions are represented in the Baseband Unit (BBU). These can be placed in different locations either for convenience or efficiency, always having in mind that the each RRH has its own BBU. Even though this architecture worked well for 4G architectures, it did not meet the requirements for the new generation of mobile networks. In part, the solution is unscalable, not efficient enough nor can it deliver 5G's biggest promises, such as of high bandwidth, low latency or cost efficient services, [16]. Therefore, a new type of architecture was conceived, with the idea of virtualizing and separating some of the known components, being Cloud RAN (C-RAN) or centralized RAN. Much like its ancestors, C-RAN divided the RRH and the BBU, but to correspond to the new requirements, it created the concept of a centralized cluster of BBUs that can support various RRHs. This cluster is then connected to the core network via a backhaul link.

When clustering the BBUs into a pool, as stated before, there can be a centralization to a site like cloud or data center which brings various benefits not only in the economic spectrum, as seen in the figure 2.4, since resources can be allocated dynamically as per service demand, but also in internal management due to the flexibility of the deployment, being able to handle various wireless standards. C-RAN can be categorized into two types of architectures: partially centralized C-RAN, and fully centralized C-RAN [16]:

- In a partially centralized C-RAN, Layer 1 related functions are moved from the BBU to the RRH. When compared to the fully centralized architecture, it requires much less transmission between between the BBU and RRH, but leads to low flexibility in the network and is less convenient in multi-cell collaborative signal processing, even if the processing is moved to the RRH;
- In a fully centralized C-RAN, all functions from Layer 1, 2 and 3 are present in the BBU, leaving all the radio functions responsibilities to the RRH. This type of deployment is the one that best meets the C-RAN advantages, but at the price of requiring the availability of large bandwidth and timing of transmission signals between the BBU and RRH.

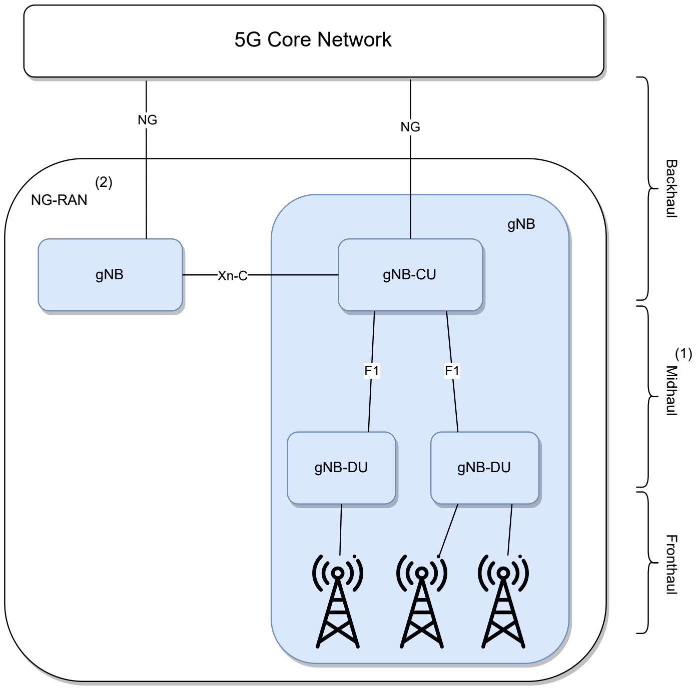


Figure 2.5: NG-RAN architecture

As introduced in section 2.1.2, 5G came with a new concept of RAN, the Next Generation Radio Access Network (NG-RAN), and with it a new node that connects to the 5GC, the gNodeB (gNB). The structure in which the gNB is based is the C-RAN, referenced in the beginning of this subsection. The gNB brought three new terms to the equation: i) the gNodeB Distributed Unit (gNB-DU), which implements lower-level functions; ii) the gNodeB Control Unit (gNB-CU), which implements higher-level functions; iii) and the gNodeB Radio Unit (gNB-RU), that implements radio functions. The initially suggested architecture implements one gNB-CU with multiple gNB-DUs, as seen in figure 2.5(2); the chosen implementation only limits the number of gNB-DU. The idea is that the 5GC does not need to be aware of the internal structure of the gNB, as it only needs to be visible to other gNBs. This also introduces a new concept of midhaul, as evidenced in figure 2.5(1), to connect gNB-CU and gNB-DU. The functions that belonged to the BBU in previous architectures are now separated between all the new units. The high-layer protocol stack functions are provided by the gNB-CU whilst the gNB-DU provides low-layer protocol stack functions, the lower layer and radio functions are provided by the gNB-RU.

2.2.2 Open RAN

The internal interfaces within the single black box used in the conventional method of providing RAN are sealed off and under the control of a single vendor. As we go towards the Open RAN, the base station functions are separated into three groups with open interfaces: a Control Unit (CU), a Distributed Unit (DU), and a Radio Unit (RU). With Open RAN all three of these entities can be fabricated by different vendors, since all the interfaces are open and known to everyone. These concepts bring along one important element, the RIC,

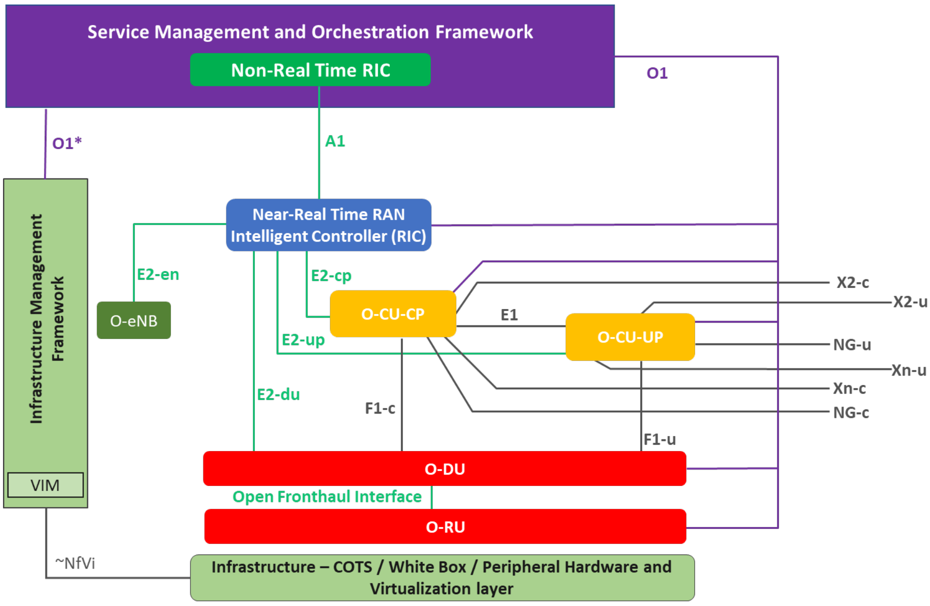


Figure 2.6: Open RAN architecture from [18]

which allows to provide the management functionality, controlling the radio resources and RAN operations [17].

The Open-RAN architecture consists of NFs, a Service Management and Orchestration (SMO) framework to manage said NFs and an O-RAN Cloud to host all the NFs, all this in a high-level view. The NFs are in fact Virtual Network Function (VNF), as they can be virtual machines or containers. The composing elements of the architecture seen in figure 2.6 are:

- **Service Management and Orchestration (SMO)**: The element in charge of managing, automating, and orchestrating RAN components. `O1`, `A1`, and `O2` interfaces are supported;
- **Non-Real Time RAN Intelligent Controller (RIC)**: The functionality inside the SMO in the architecture that offers the A1 interface to the Near-RT RIC is known as Non-RT RIC. By providing policy-based guidance, machine learning model management, and enrichment data to the near-RT RIC function, the goal is to aid and enable intelligent RAN optimization. It performs tasks over a non-real-time, or longer than one second, interval;
- **Near-Real Time RAN Intelligent Controller (RIC)**: With the help of integrated intelligence, this controller delivers further features including QoS management, connection management, and seamless handover control. The Near-RT RIC can be seen of as a strong, safe, and scalable platform that allows for the control and optimization of O-RAN resources and elements in close to real-time. It includes the interpretation and application of SMO policy as well as enrichment data to improve control functionality. A Near-RT Controller performs operations in under one second.

2.3 Core Network

The 5GC core network manages all data, voice, and Internet connections, usually referred to as the mobile exchange and data network. As stated in previous section 2.1.2, it is based on cloud technologies and in an SBA, all having future iterations of mobile communications in mind. Like its predecessors, the 5GC controls subscription information, registers UEs, establishes data sessions, traffic forwarding, and has specific QoS functions. Even though it reutilizes the principles, the 5GC completely reworks some of the functions from the EPC while presenting some new ones to meet the new objectives established for it.

This section explains the mandatory functions from the 5GC core architecture and the purpose of each one of them, as well as the innovations in the CN, mainly in slicing and QoS.

2.3.1 Overview of the architecture

Unlike in the Evolved Packet Core, where the Mobile Management Entity handles the session management and mobility management, these functionalities and procedures are governed by the Session Management Function (SMF) and AMF in 5G [19]. The AMF terminates the connection between the AN and the UE in the Control Plane (CP). A link is created in the Access Network between the User Equipment and the AMF, called NAS, like in 4G. The UPF is the bridge connecting the CP and Data Network (DN), handling the SMF procedures and session management functionalities as actual user data. Nevertheless, the SMF deals with selecting the UPF. The three functions referred deliver the primary functionalities of the 5GC, but there are other functions with other responsibilities which will be referred to later. As stated before, 5G is a Service Based Architecture (SBA), meaning it uses a set of inter-connected NFs to deliver the CP functionalities and the standard data repositories standard data repositories, all this with he capability of each NF having authorization to access each other's services. This type of architecture, present in the 5GC framework, offers a flexible deployment of common applications. The Network Functions offer specific services to other NFs utilizing interfaces of a common framework. The service could be extended to any other service-permitted consumers. Hence, the SBA provides reusability and modularity and enables a virtualized deployment [19]. This uniform interface connections are called Service Based Interface (SBI). The 5GC can be: fixed, mobile or a combination of both; single communication of unicast, multicast and broadcast; and lastly, either wide area access networks or local area accesses. Showing that it will offer the scalability and service ubiquity for a wide variety of use cases.

There will be various deployments of a 5GC (each with different NFs), but some of these are mandatory and need to be used in every architecture, these being:

- **Access and Mobility Management Function (AMF)** - This function uses the `N2` and `N1` interfaces to interact with the radio network and the devices through signaling, it connects to the rest of the core's NFs through the SBIs. It is the main function when it comes to signaling flow calls in all of the network as well as managing connections with the devices, permitting them to register, authenticate and move between different radio cells. One big difference to its predecessor in the EPC architecture is that it does not handle session management or device authentication. All the messages related to session management are forwarded to the SMF, and device authentication is forwarded to the Authentication Server Function (AUSF);

- **Session Management Function (SMF)** - Responsible for managing the UEs sessions, all from their establishment, to eventual modifications and releases. It is also responsible for allocating an IP address for each individual session. Since this function does not communicate directly with an UE, it uses the AMF to send the messages. Much like all the other NFs, it uses the SBI to communicate with other NFs through the `N4` interface. Besides this, this function is responsible for the selection and control of different UPFs, by enforcing traffic for individual sessions and traffic steering. Finally, in conjunction with the Policy Charging Function (PCF), the SMF controls the charging functionality enforcing it in the UPF;
- **User Plane Function (UPF)** - The main purpose of the UPF is to process and forward user data. It connects with external IP networks and acts as a stable IP anchor point for the devices towards external networks, hiding the mobility [20]. This means that even if the UE is moving around in the network the IP packets with the address of a specific UE is always routable from the Internet to a specific UPF. The UPF also analyzes the content of user data packets for policy charging, as mentioned above, or traffic usage reporting. One other remark is that it applies QoS in packets with the destination of external networks, or the radio network. This is useful because if network congestion happens, the transport network can handle each packet with the right priority;
- **Authentication Server Function (AUSF)** - This is a very limited function but an important one nonetheless. It is responsible to authenticate a specific device using the credentials given by the Unified Data Management (UDM), it also provides services for generating cryptographical material for secure updates of roaming information and other parameters in the UE;
- **Unified Data Repository (UDR)** - This is the database for all user subscription data, types of network and user policies. It mainly supplies data as services to other NFs, such as the UDM and the PCF;
- **Unified Data Management (UDM)** - Acts as a face to the UDR, facilitating the management of user subscriptions registered in it. It generates authentication data to

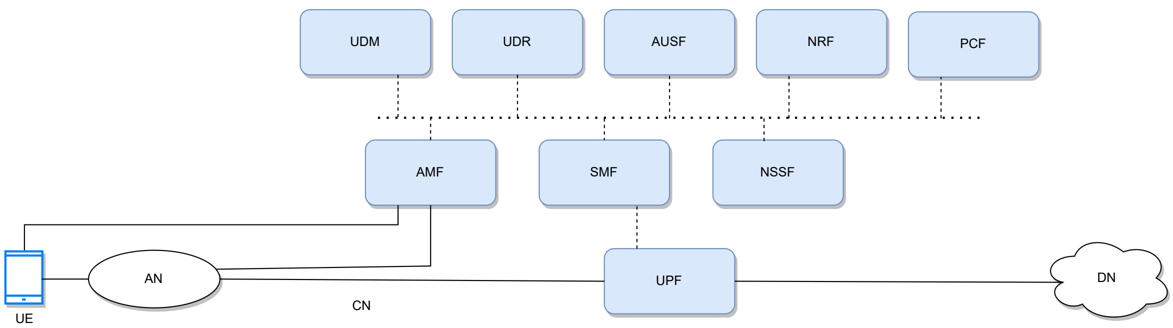


Figure 2.7: 5G core network mandatory components

authenticate attaching devices, as well as authorizing access to specific users depending on their subscription data, such as roaming services. Mostly, it acts according to the requests of the AMF and if there is more than one instance it keeps track of which instance is serving a specific device. This is also applicable to the SMF;

- **Network Repository Function (NRF)** - Is a repository for the profiles of the network functions available in the network. The main purpose is to discover the best suited NFs services to a service requested by the consumer, all without the need to be configured beforehand;
- **Policy Charging Function (PCF)** - According to [20], this is the only one on this list which is not a mandatory component. Similar to the PCRF in the EPC architecture, this function manages users, data sessions and data flows. It presents a new feature as it provides policies directly to the device through the AMF in NAS signaing messages. It can provide two types of rules, one in which it indicates to the UE how application traffic shall be sent over the network, be it data sessions or slices. The other type is used to guide the device in determining which Wi-Fi networks to select, this is applicable in non-3GPP network access.
- **Network Slice Selection Function (NSSF)** - The (set of) network slice instances and the set of AMFs that should serve the UE are chosen by the NSSF. The NSSF, which is aware of all network slices and helps the AMF with cross slice selection, may be assigned to one or a group of network slices.

2.3.2 Slicing and QoS

The previous section presented the reshaping of the CN, that brought new features, such as Network Slicing (NS). NS is considered one of the key features in 5G by the 3GPP. The idea behind the concept of a network slice is that there can be a dynamically created end-to-end tunnel. All the processes should be seamless for the end user as the management is made by the gNB and the CN. The end-user can also access multiple slices over the same gNB according to which service it uses. To better understand what a slice is, it should be taken into account that a slice is defined within a Public Land Mobile Network (PLMN) and includes both the 5GC and NG-RAN networks; A slice is identified by the Single Network Slice Selection Assistance Information (S-NSSAI), which is signaled by the UE to select the slice instance. There are two types of network slicing: hard and soft. While soft slicing network slices may share specific network resources, hard slicing network slices must be entirely separated from one another. According to 3GPP [21] there are two important definitions that should be understood:

- Network Slice: A logical network that provides specific network capabilities and network characteristics.
- Network Slice Instance: A set of Network Function instances and the required resources (e.g. compute, storage and networking resources) which form a deployed Network Slice.

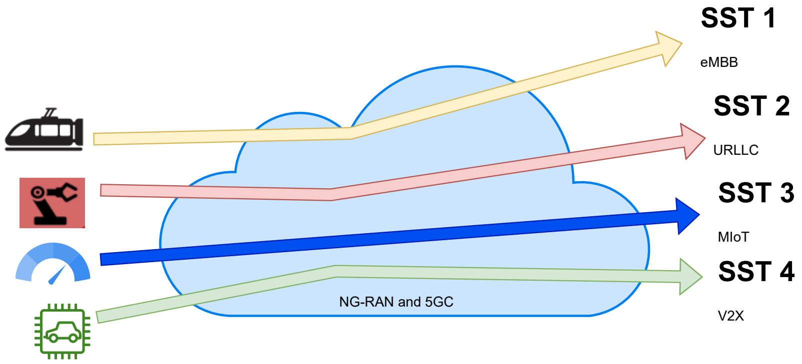


Figure 2.8: 5G Slice/Service Type structure

An S-NSSAI is used by the UE in the AN and consists of a Slice/Service Type (SST) and a Slice Differentiator (SD). The SST refers to the service a slice is supposed to represent, whereas the SD differentiates each slice in the same service type. According to [21] there are four SST standarized values of each major service type addressed in section 2.1.2, which can be seen in figure 2.8.

Slicing is the new concept of 5G, but some already known concepts got reworked and improved, such as QoS. In 4G, there was the definition of QoS Class Identifier (QCI), which identifies the quality of packet communication provided. A bearer was defined with what could be up to nine different QCIs per APN for a UE. As it worked by bearer, much signaling was required by the EPC because each bearer has its tunnels between the eNB and the EPC; not only this, but it also presents very little granularity with only nine levels. With the arrival of 5G, 5G QoS Identifier (5QI) appeared, with two hundred and fifty-four values, presenting a much higher granularity. All these QoS flows of a UE are set up inside a GPRS Tunnelling Protocol User Plane (GTP-U) tunnel, that connects the gNB to the UPF, better explained further on in section 2.5.2. This set up is done with reduced signaling messages in the CN, therefore it is much more efficient. In slicing and QoS flows, the SMF does all the session management. For example, an SMF is chosen based on which slice it represents, which can be done with the S-NSSAI, and by communicating with the PCF it applies the desired QoS rules to a specific user.

2.4 LTE and NG coexistence

As expected, and as 5G is deployed with a 5GC network instead of an EPC network, the initial coverage will be very limited. This has two main reasons, one thing that it takes time to build coverage to all locations, so it can be assumed that this first deployment will be geographically spotty. Secondly, NR technologies are usually deployed on higher frequency bands than the current radio technologies, mostly because this provides superior network capacity. Therefore, and to cover scenarios in which a user wishes to move from a area with 5G coverage to one that does not, a coexistence of the two technologies must exist, meaning connectivity between some of the EPC and 5GC network elements. Ultimately this leads to numerous architectures, in which there is a deployment of solo LTE technologies, solo 5G technologies or a conjugated solution.

This section will address how LTE and 5G can, and should, coexist. It is divided into the core and the RAN. The core subsection will address a core architecture in which both technologies coexist. The RAN subsection will address two types of architectures in NSA and SA and the different types of nodes that represent each of the architectures.

2.4.1 Core

Independent core solutions for LTE and 5G have been approached in 2.1.1 and 2.3.1, respectively. But as previously explained, a full deployment of 5G is not yet possible, in order for this to work there needs to be a coexistence of both technologies. The way this is will work is if a 5G user does not have 5G coverage, it will connect over LTE. In such a scenario, a user will be served by an MME, a S-GW, and will be allocated to a SMF in the 5GC network, this latter one will act as a Packet data network gateway to the Serving GateWay. The subscription data from this specific user will be given to the Home Subscriber Server (HSS) by the UDM, so that it can be provided by the MME. Upon return to a geographic location in which the user has NR coverage, once again, the session context is handed from the MME to the AMF accross the `N26` interface, and user data tunnels from UPF are moved from `S5-U` towards S-GW to `N3` towards the NR radio network [20], shown in figure 2.9. During this transaction, a single stable IP anchor point is maintained by the UPF.

In order for this to work, some requirements need to be met by the CN, besides the need for the interface `N26` to exist, the main ones being:

- Both the SMF and UPF need to support EPC P-GW logic and functionalities;

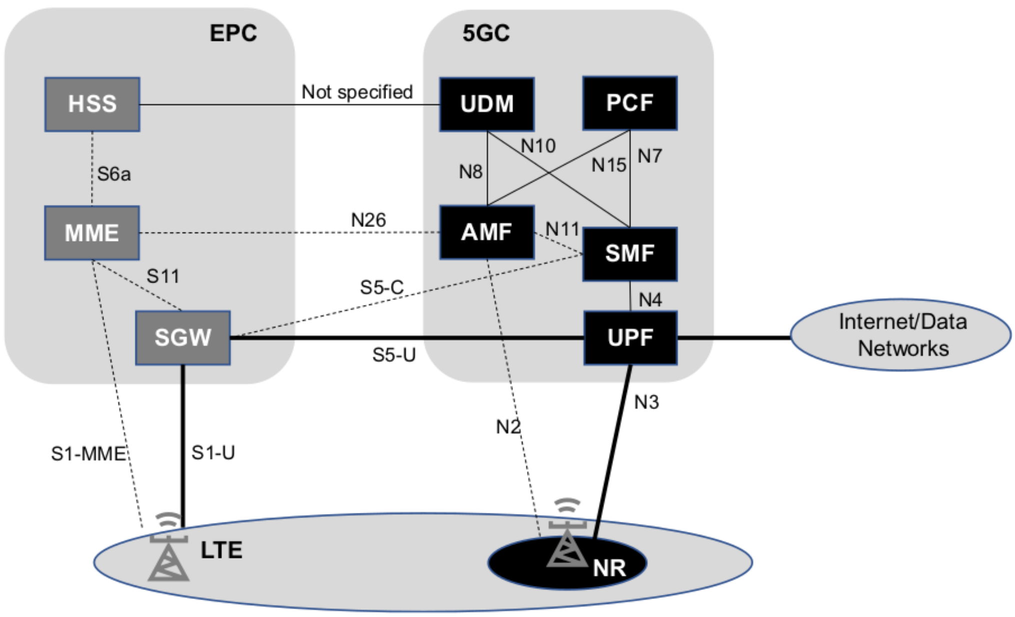


Figure 2.9: 5G Core and EPC interworking [20]

- The PCF needs to support correspondent policy data parameters to work with the Packet data network GateWay - Control plane (PGW-C) functionality of the SMF;
- Existing 4G users must be unaffected by the new 5G architecture.

In the cases that the `N26` interface does not exist the MME and the AMF cannot exchange session information, therefore the information on which anchor to use must be retained in both the HSS and UDM. The result of this is that in some cases there will be an interruption time when there is mobility. When there is an interruption time to do the handover it is called single mode registration, meaning that the UE is only connected to one CN at a time. To avoid this, a variation was made in which the UE is connected to both CN at the same time, this is called dual mode registration. But, this solution implicates a more complex device, and the support for this variation in not required, opposed to single mode.

2.4.2 RAN

As NG-RAN provides both NR and LTE radio access, in this representation, a node can be a gNB or a NG-eNB. The gNB will provide NR, UP and CP services whereas the NG-eNB will provide LTE / E-UTRAN services. NG-RAN brings along various new functionalities to the field of RAN, but one of the more prominent features is the possibility to operate in NSA, and in SA architectures. This means that, if is operating in NSA, the RAN components (which can be both NR and LTE) are connected to an EPC, and on the other hand, if is operating in SA, the gNB is connected to a 5GC Network.

When dealing with an NSA topology the enhanced gNodeB (en-gNB) and the eNB will be connected between them through the `X2` and `X2-U` interfaces, they will also be connected to the MME/S-GW pair through the S1-U and S1 interface correspondingly, as seen in the figure 2.10. This ensures the availability of 5G NR technology without the need to replace the already established core networks, although only the services provided by 4G would be available in 5G NR in terms of higher capacity, lower latency, etc. Before we can have a fully operational 5G SA architecture, the implementation of QoS is relying on a complete deployment of a 4G infrastructure [22]. When operating in SA the NG-eNB and the gNB will be connected between them through the Xn interface and to the AMF/UPF pair through the NG interface.

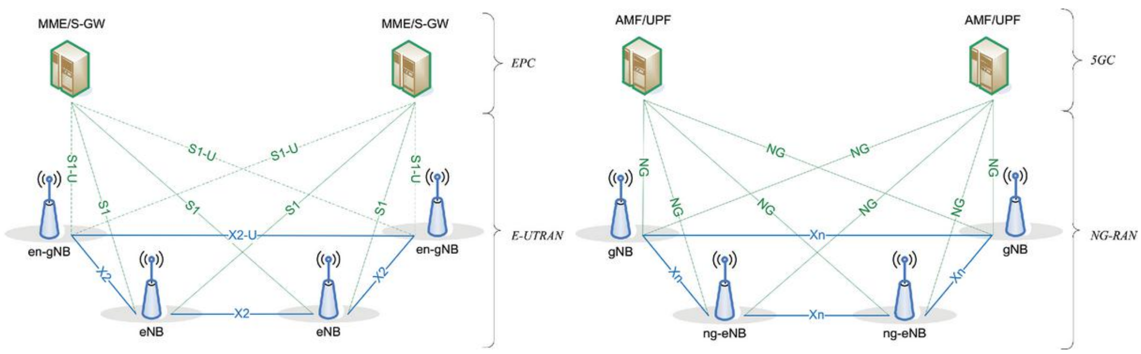


Figure 2.10: Non Standalone and Standalone RAN [22]

2.5 Edge Services

Effectively managing network resources has been a big concern since the advent of smartphones. There was a need to address cloud computing challenges, such as data processing and storage, in addition to the emergence of various and complex mobile apps. The best way to do this is to move the computing part straight to where the end-user is. From that, the concept of edge was brought into picture, along with the concept of MEC. This concept can be used in several scenarios, such as of computation offloading, distributed content delivery, IoT, and many others. Core Networks takes advantage of this by disaggregating the user plane from the control plane, in what is called CUPS. This concept originally came from LTE, but has since then been used by 5G in the selection of the UPF, since the UPF is the user plane of the 5GC.

This section studies MEC, its applications and use cases, as well as the importance of the UPF selection and how this selection is made.

2.5.1 MEC

Multi-access Edge Computing (MEC) intends to integrate technological and telecommunication services by supplying a cloud computing platform at the edge of the radio access network. By providing storage and processing resources at the edge, MEC helps mobile end users to experience less latency and makes better use of the mobile backhaul and core networks. Furthermore, MEC focuses on application-related improvements, taking into account single platforms in the initial phase, such as feedback mechanisms, information, content processing and storage, to name a few [23]. MEC has a series of use cases and applications, some examples are:

- Computation Offloading - Due to the decreased latency and near proximity to the RAN edge, MEC uses less energy. Users thus benefit from quicker execution and better performance which is the main the reason for the usage of computation offloading;
- Distributed Content Delivery and Caching - Distributed caching can improve user experience while lowering backhaul and core network utilization by extending Content Delivery Network (CDN) services to the mobile edge;
- Web Performance Enhancements - The major goal of web performance improvements is to shorten access times and speed up page loads by offering content optimization that takes into account user, network, and device circumstances;
- IoT and Big-Data - MEC can provide computing and storage capabilities near data sources, which can be utilized for processing objectives, such as assuring a quick response to user requests, decreasing IoT data and signaling, or allowing new services.
- Smart City Services - MEC is the perfect platform for video analytics at the edge of the network, such as face recognition or computer vision algorithms, and vehicular communications, all without overtaxing the mobile backhaul and core network while ensuring low latency.

As expected, MEC intends to take advantage of cloud computing and virtualization technologies. Even though these concepts are well established in most aspects of the technology field, it only appeared recently in mobile communications. This is the reason why concepts of VNF, SDN, and NS are relevant as they allow flexibility and multi-tenancy support. These concepts in correlation with the edge will be better explained in this section, with the exception of SDN as it is out of the scope of this thesis, and NS, as it has already been approached in the previous section 2.3.2.

The virtualization of a particular component allows for easy deployment on most platforms. This permits MEC services and components to be placed easily on the edge. One good solution example is Docker[1], which can replace virtual machines ensuring that storage and computation are done with fewer resources. In more detail, applications are built and created using image stacking and extension by containers, which are then saved as images in a repository. The container engine later uses this to run the program on the host. The portability of containers from one virtual machine to another or even to a piece of hardware ensures flexibility on the edge deployment. Virtual Network Functions and Network Slicing are new concepts that also came to boost the deployment of edge services. They allow for a decoupling of the gateway NFs, firewalls and Domain Name Server (DNS) from proprietary hardware. This allows the control and data plane to be deployed in any desired piece of hardware (as long as it meets the minimum requirements) and to be easily separated across platforms, such as virtual machines. For instance, VNF can deploy more resources from one edge platform to another to handle congestion in an ongoing MEC application caused by a flash crowd event.

As shown, MEC is present in a variety of scenarios but its original target always was mobile networks. In a 4G architecture there are many scenarios on how MEC can be deployed, which can be further explored in [5]. One of these architectures is Distributed S/PGW, in which the S-GW and the P-GW are deployed on edge site where as the rest of the nodes are maintained in the operator's core site. Both the S-GW and P-GW can run as VNFs along with the MEC application. The S-GW is chosen by the MME or the control plane S-GW if they're decoupled, according to 3GPP standards and based on Tracking Area Code (TAC), APN or other, all this allows for traffic offloading of specific applications.

In 5G networks, 3GPP permitted the mapping of MEC onto Application Functions (AFs). This function can then use the services and information from other NFs based on configured policies [24]. Besides the AF, the UPF has a key role in the integration of MEC in a 5G network, as from the MEC standpoint it is a distributed and customizable data plane that can steer the traffic towards the MEC application. Having such a customizable entry to such application leads to more complex scenarios, like mobility.

Application mobility is necessary for the MEC system to maintain the application requirements in a mobile environment. Currently, at the time of writing, this entails moving the user-serving program instance, the UPF, to a different location. This enables the architecture to serve critical services in the future of mobile networks like V2X, following an architecture such as the one in the figure 2.11.

[1] `https://www.docker.com`

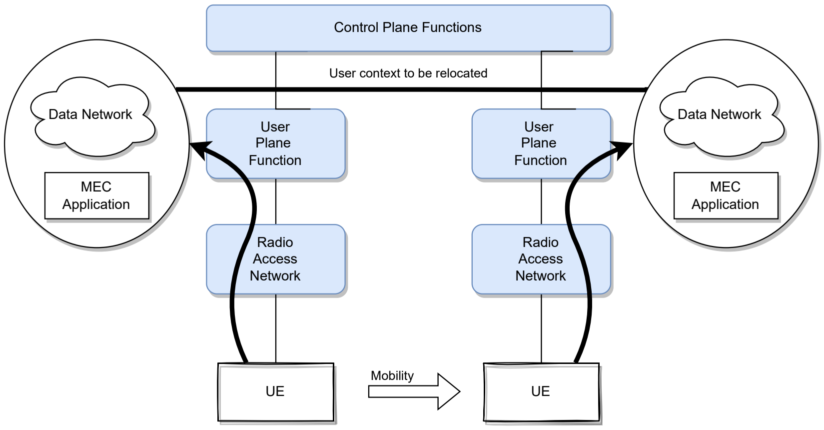


Figure 2.11: MEC application with UPF relocation adapted from [24]

2.5.2 UPF Selection

The UPF is a fundamental component of the 5GC infrastructure, representing the data plane evolution into Control and User Plane Separation (CUPS) strategy. This enables the data forwarding, previously handled by the P-GW, to be decentralized allowing the packet processing and traffic aggregation to be executed closer to the network edge. Some of the main responsibilities of the UPF are:

- A Protocol Data Unit (PDU) session anchor point for mobility scenarios;
- Packet routing and forwarding;
- Per-flow QoS handling;
- The connecting point between the CN and DN, by doing the encapsulation and decapsulation of GPRS Tunnelling Protocol User Plane (GTP-U).

When talking about the last point, of GTP-U encapsulation and decapsulation one needs to be aware that there are two components to the GPRS Tunneling Protocol, or GTP. These being the GPRS Tunnelling Protocol Control Plane (GTP-C) and GPRS Tunnelling Protocol User Plane (GTP-U). The GTP-C is the control protocol for the EPC architecture for DN connections and user plane tunnels, where as the GTP-U in the 5GC only carries out user plane data. The GTP-U tunnels separate traffic into different communication flows. They are identified by a Tunnel Endpoint Identifier (TEID), an IP address and a UDP port. These tunnels are used to connect the RAN to the 5GC, more specifically the UPF, and have bundled inside them all the QoS flows of a certain UE.

As referenced in the previous section 2.5.1 the decouple of the user plane from the control plane enables the usage of edge services. Logically, there will be the need to choose multiple user planes, and since the UPF is the NF that represents the user plane, there is a need to perform a UPF selection. In the 5GC architecture, as mentioned in section 2.3, the SMF is responsible for the selection of the UPF. In the EPC this selection is made by the S-GW. The way this selection is made depends on how the network topology was constructed as it is not standardized, but some key features serve as ground. Before looking at this key features there should be a guarantee that the SMF is actually aware of the existent UPFs, this can be done in a number of different ways:

- Awareness by location: The network topology is configured in a way that the SMF is aware of the UPF's location and in what way the UPFs are connected, ultimately allowing to select UPFs by the location of the UE;
- NRF: A query is made to the NRF to request the existent UPFs, this also delivers basic information on each UPF such as Data Network Name (DNN) and S-NSSAI. This way the SMF does not need to be preconfigured but the information is reduced, not having information such as the UPF's topology;
- N4 protocol: This protocol supports exchange of information and capabilities by the SMF and UPF. This way the SMF learns about all the basic information of the UPF and its load allowing for better decisions.

It was previously mentioned that there were some key features in which the SMF can rely on to choose the UPF, independent of the operator. The operator can then choose which ones are to be used or make sense according to its network topology. Some of them are the following, the full list can be found in [25]:

- UPF's dynamic load;
- UPF/UE location;
- Data Network Name (DNN);
- PDU session type (IPv4, IPv6, IPv4v6);
- S-NSSAI;
- Local operator policies.

The above mentioned principles of UPF selection have in mind a stationary approach, or initial registration of the UE to the CN. But thinking in mobility there may be cases of UPF re-allocation or dynamic selection of UPFs. If mobility is in question then handovers should be taken into account, in the scope of this thesis three handovers are relevant, `S1` and `X2` handovers in LTE and `Xn` handover in NG-RAN. Understanding these handovers, the serving eNB in an LTE handover operation decides whether to start the handover procedure or not. The serving cell eNB also selects the target eNB cell from its neighbor list when deciding whether to do an `X2` or an `S1` handover. The source eNB executes the `X2` handover if the `X2` connection has been established with the target eNB and is currently available; otherwise, it chooses to do the `S1` handover. In NG-RAN it is performed a `Xn` handover which is all and all similar to the `X2` handover in LTE, using the gNB's `Xn` interface. The handover scenarios will be further discussed later in this thesis.

In an `Xn`-based handover with UPF re-allocation both the source and target gNBs are connected to different UPFs. When a UE moves from gNB to gNB the NG-RAN sends a path switch request to the AMF, then the SMF has to select a target UPF and send a session establishment request to it. After acceptance, the SMF exchanges session modification messages to a PDU session anchor to switch the PDU sessions. After a downlink on the old path, the AMF sends a path switch request to the new path [26].

2.6 Emulation and Simulation of 5G networks

Nowadays there are a number of open source projects that allow anyone to easily deploy a functional core and access network. A full list of open source projects can be found in a

GitHub[2] repository. Between these projects there are a few worth mentioning and some were used in this thesis, further on there will also be a small description of these projects:

- free5GC;
- Open5GS;
- OpenAirInterface (OAI);
- srsRAN;
- UERANSIM.

`Free5GC`[3] is an open-source project for 5G mobile core networks with the ultimate goal to implement the 5GC defined in 3GPP Release 15 and beyond. This project migrated the MME, S-GW and P-GW to 5GC maintaining the PCRF and HSS unchanged. In stage 2 the developers stated that they added an handover feature but there was no further specification.

`Open5GS` [27] is a C-language open-source implementation of 5GC and EPC, meaning the core network of NR/LTE network. As some of the main supported features it has multiple encryption algorithms, IPv6 support, multiple PDU session and handover capabilities (`Xn/N2` for 5GC and `S1/X2` for EPC). Besides this it has some known limitations such as no roaming or emergency call and the main one, does not support Interworking EPC (mentioned in 2.4.1).

The `OAI`[4] project supports basic procedures for connection, registration (UE registration, de-registration, and service request) and session management (PDU session establishment, modification and release). This project also provides RAN functionalities. OAI will be further approached in the next section, as it is mentioned in some of the articles referenced in it.

The `srsRAN`[5] is a free and opern source 4G and 5G software radio suite. It features both UE and eNB/gNB applications, and can be used with third-parthy CN solutions to build an end-to-end mobile network. Besides both UE and RAN it also includes a light-weight EPC implementation, with all the modules from the original architecture (unlike `Open5GS` which separates the S-GW and P-GW). As the name indicates its main features are in the RAN components, so they do not explicit which features are implemented in the core solution.

As shown previously, each project has different priorities, thus each implements different functionalities from one another. To make a choice of which tool to use one must know if the solution implements the needed requirements. In order to check some of main functionalities a group of researchers developed a tester [28], to perform some experiments in the three cores referred above. These experiments revolved around some key aspects, such as SMF and UPF selection, registration and authentication, PDU session management and UE context management. Besides, there are other tests done to these projects but they fall of the scope of this thesis. A result of these tests can be seen in table 2.1.

For the development of this thesis and in order to more easily study 5GC architectures and user sessions a RAN simulator may be useful, `UERANSIM`. It is an open source state-of-the-art 5G UE and RAN (gNB) simulator. UE and RAN can be considered as a 5G mobile phone

[2] `https://github.com/calee0219/awesome-5g`
[3] `https://www.free5gc.org/`
[4] `https://openairinterface.org/`
[5] `https://www.srsran.com/`

Table 2.1: Test results resume (adapted from [28])

Features and Procedures	Protocols	Messages	free5GC	Open5GS	OAI
Registration	NAS	Registration Request Registration Accept Registration Complete	✓	✓	✓
Primary authentication key agreement	NAS	Authentication Request Authentication Response	✓	✓	✓
UE context management	NGAP	Initial Context Setup Request Initial Context Setup	✓	✓	✓
PDU session management	NGAP	PDU Session Resource Setup Request PDU Session Resource Setup Response	✓	✓	✓
SMF selection	NAS	Transport	✓	✗	✗
UPF selection	NAS	Transport Session management	✓	✓	✓
Interface management	NGAP	Interface Management	✓	✓	✗

and a base station in basic terms. The project can be used for testing 5G Core Network and studying 5G System[6].

From the overview of each of these solutions, a decision was made to implement Open5GS as the CN provider. This project presented a good documentation that was made not only by the developer, but also by the community who shared its experience and knowledge through personal blogs which are then featured in the official website. On this note, the official site is also very transparent on the features Open5GS[7] provides and hardware tested with the solution. Lastly, from the GitHub repository one can see that the developers are quite active and responsive to issues created. This type of transparency is not as clear in the other core solutions, added to this Open5GS has a strong implementation of CUPS scenarios which is one of the main objectives of this dissertation. Finally, from table 2.1, one can attest that Open5GS checks all the boxes except the SMF selection, whilst Free5GC effectively checks all the boxes.

2.7 Related Work

This section will present different works in the scope of this thesis and in them, three main topics will be approached: open source tools, deployments of end-to-end cellular networks and the involvement of cellular networks in city environments and their services (topics such as Neutral Hosting and MEC). The aim of the first and second topics is to show how a cellular network can be deployed in its entirety, as this involves having a knowledge in the architecture of both 4G and 5G Core Network and RAN. And, with the knowledge from this study, make a choice of which core solution to implement. The choice of presenting open source solutions, also approached in the previous section (2.6), has as main purpose to present solutions that everyone can deploy and work on, in an attempt of learning about 5G as a community.

[6] https://github.com/aligungr/UERANSIM
[7] https://open5gs.org/open5gs/docs/

Firstly, two papers will be studied in the sphere of the first topic, one referring to a wide variety of open source solutions and the second focusing only on one of the CN solutions, OAI. In [29], a multiplicity of open source solutions for the 5G ecosystem are presented and organized in appropriate blocks. This article depicts how each solution fits into the 5G puzzle, and presents an overview of strengths and limitations of each of them. Although it presents a very complete list, it does not give an insight on how each solution really works as the deployment is not made, relying solely on documentation. This is mainly an issue when deploying the project in conjunction with others, as their integration may result in unknown issues, even to the developers. Besides this, it also does not mention the features that do and do not exist in the tools presented. However, this article was relevant to understand the functionalities of other CN solutions as well as new frameworks that could be integrated with a mobile network. Moving on to [30], the state of the development of OpenAirInterface (OAI) is discussed along with a roadmap for the future, its community, public license and its usage in academia and industry. The authors present the possibility of NSA and SA scenarios as well as an explanation of the used interfaces. Real results are shown in a NSA testbed with one eNB and one gNB connected to the OAI EPC. The scalability of the solution is not mentioned, only having a brief mentioning of virtualization with tools such as Docker, in a reference to [31]. This leaves out the possibility to escalate this solution to a city wide environment. From this article it was possible to further understand how this CN solution works, coming from the study made previously (section 2.6), and ensure the decision of using Open5GS.

To approach the second topic, the objective was to find works that mentioned 5G networks from end-to-end, as this will be the work made in this dissertation. With this in mind, another two articles were chosen, the first with a more theoretical approach. The second has the same objective of a full deployment, but presents a more hands-on approach. First, in [7], an overview of the new 5G architecture is shown, together with its minimum requirements and network fundamentals. Furthermore, the authors propose a new spectrum for services of different use cases. This survey is supposed to serve as a guideline for researchers or operators for future 5G network deployments. It provides very good theoretical background of 5G, but it does not provide any practical insight on the matter, not approaching any CN solution nor RAN. In the concepts mentioned in it, there are a few that would be interesting to be mentioned, such as CUPS or MEC. This would be interesting to refer as 5G brought more attention to them mainly in one of its three verticals, URLLC. This article was used as foundation for the 5G concepts approached in this thesis, as it was already mentioned previously in this chapter. Furthermore, the different use cases for the three verticals are also very well depicted by the authors and served as inspiration for the ones defined for this dissertation. In [32], a different approach on 5G is made, focusing more on the history and background of 5G companies and manufacturers, open source software and generic hardware. It also gives a detailed view of how to build and set up a network using specific tools. The deployments are explained in detail, showing that nowadays cellular networks can be deployed in almost any piece of hardware, with the major restrictions being radio functionalities. It would be interesting to have a level ground test to compare all the three deployed solutions, even if not in network performance as it is difficult with simulators, but maybe a more detailed computational performance. One interesting detail that could have been added is, since the authors provide a thorough tutorial in which it details every command that should be used, a script could have been developed and made available through a platform like GitHub to ease the replication of the tutorial.

The third topic of this related work, the involvement of cellular networks in city environments and their services, is grounded in the first two which were presented previously. With an idea of what open source projects exist, and the technical background of the deployment of end-to-end mobile networks, a more thorough analysis of works in their integration with city services can be made. Before starting this analysis, in section 1.1 it was mentioned that multiple projects arose under 5GPPP to enable small cell multi-tenancy network operators, one of these projects is *5G ESSENCE*. In [33], this project is detailed and a enhanced architectural framework is presented. This article mainly refers to RAN sharing by using a controller for it and radio resource abstraction, both managing programmable data planes while having the control plane disaggregated. It concluded that such would reinforce and take advantage of MEC capabilities by having edge-based, virtual environment attached to the small cell. In [34], improvements on the H2020 5GCity[8] project are presented. This project integrates VNF, SDN and MEC on distributed cloud and radio platforms. The project leverages on a neutral host platform to create a flexible slicing allocation, policies definition and scaling of infrastructure resources. MEC is integrated to assure ultra-low latency and high bandwidth, but besides this, it is aimed to provide virtualization of computing at the edge along with virtualized radio functions. The paper was studied with the intend of understanding the 5GCity project and what the technologies to be used in this dissertation are truly capable of, in terms of different virtual services deployments in smart cities.

The usage of MEC frameworks in the edge requires that there is an already solid deployment of VNFs beforehand. This type of systems has huge requirements in regards to infrastructure management and orchestration. For this, [35] describes a solution for open issues identified by ETSI in the integration of MEC with VNFs, having as goal that these VNFs are edge aware solutions for 5G neutral hosts. From this paper one can attest that ETSI is making an effort to standardize architectures for MEC and VNF with the goal of accelerating the adoption of the neutral hosting paradigm. The solutions proposed to the identified issues seemed all very limited to the architecture defined for the project in which the authors are integrated, making it hard to deploy in a different scenario. In [36], a survey was made on the existing work of MEC for 5G, along with a proposition of a framework based on the use of MEC for 5G in a smart city environment. The proposed framework only works with assumptions, even thought it refers cellular networks and more precisely 5G, the authors do not mention how a 5G network will enable MEC scenarios. It proposes an edge sub layer to process data at the edge, by using a service near the base station. This dissertation will follow a similar approach by deploying a service and the UP near the edge. Following the edge deployment

[8]`https://www.5gcity.eu/`

topic, in [37], the authors propose and prototype a CUPS-based edge computing architecture for dynamic mapping and quantified its benefits. Furthermore, they provide experimental results that stand as ground for enabling edge computing scenarios having in mind the year in which it was written. Although merit should be given for deploying a real prototype, in order to simulate edge, the authors simulate the separation through FLARE, which is a deeply programmable node architecture that isolates the VNF. No more explanation is given on how this separation occurs, therefore it is hard to understand how it is made in order to replicate. The authors also give an useful insight of how the modules of the CN selected each UP. It would be interesting to also indicate which other ways the OAI solution allows to make the node selection, as the APN is the only one mentioned. Besides this, the idea that there can be a disaggregation of the CP from UP to enable edge based scenarios is what this dissertation also aims to execute, and this work shows that it is possible even though not on the CN tool that will be used in this work.

This study served to explore open source core solutions other than Open5GS, which is going to be used in this dissertation. The main contestant to Open5GS was OAI. Although it is a very complete project, it appeared to not be as flexible as Open5GS in terms of module replication and code modification. Moreover, Open5GS presents a very active community with a lot of different projects in motion. The research made through these papers also lead to broader goals when approaching city wide services that could be explored with cellular networks. This work shows that ways to enable MEC services, such as core network disaggregation, are the best approach to assure the best results in latency and high bandwidth, which are some of the pillars of what 5G networks will bring. Finally, an understanding of how other projects tackle, at a whole scale, cellular networks as a foundation for the smartness of a city inspired the development of this work. One main conclusion taken from the research made in this section is that there are quite few real deployments of 5G networks in city environments, mainly open source solutions. With 5G having its use cases defined back in 2019, it would be expected that initial testing would have been approached already. Also, the fulfilling of the three main verticals through new technologies, such as MEC, appears to be merely theoretical with nothing more than propositions of frameworks and architectures. The documentation for the deployments that were already made is lacking, giving little to no insight on how one can replicate the work. The CN solutions that are available for everyone were not thoroughly explored, and their full capabilities were not fully tested, not only in the search for flaws, but also for pertinent features that could be taken advantage of.

This dissertation will attempt to make use of the knowledge gained from these works and state-of-the-art study to deploy an end-to-end cellular network, and provide thorough insights of its operation, so that it serves as ground foundation for its own future scalability.

2.8 Summary

In this chapter, background concepts were introduced for the foundation of the full deployment of a mobile network. Firstly, it discussed the general concept of cellular networks together, with a short description of each generation up until the fourth and fifth generations, which are described in more detail. After this, two concepts were approached, constituting an end-to-end mobile network architecture, the Access Network (AN), and the Core Network (CN). To explain the AN, firstly, the complete evolution of the Radio Access Network (RAN) nodes were presented, all from a distributed architecture, from the cloud to where it stands now in Next Generation Radio Access Network (NG-RAN). Next, the new 5GC network was detailed, with the purpose of its mandatory functions and the new features they bring to the table. Features such as slicing which provides fully purposed networks with the intent to isolate specific services, and further on to new 5QI that allow for a much bigger service differentiation, all to fulfill the service requirements established for 5G networks.

The coexistence of the two current main generations was discussed with a grasp of the AN and CN concepts. This coexistence was approached firstly by how they can interact with each other in the CN, how the two cores can coexist and share responsibilities, as well as deal with users changing from 4G to 5G coverage. After that, the RAN was explored, introducing two essential concepts in Non Standalone (NSA) and Standalone (SA) architectures and by which components these architectures are constituted. After this, there was a familiarization with the concepts of edge and Multi-access Edge Computing (MEC) services, intending to understand the needs of a city environment. To pair with this concept, the notion that the gateway of the CN can be brought closer to these services was approached and explained. To finalize the chapter, some simulators tools of AN and CN were displayed and related work to this thesis, along with a discussion of what this work brought to the scientific community, and how they were useful for this dissertation. The state of the art revision touch upon the concepts of Neutral Hosting, and how a smart city can deploy different services, such as V2X. These concepts were used to design a city-prepared end-to-end mobile network architecture which will be approached in the next chapter.

CHAPTER 3

A City-Scale Architecture

The definition of three main service verticals, the enhanced Mobile BroadBand (eMBB), massive Machine Type Communications (mMTC) and Ultra Reliable Low Latency Communication (URLLC), led to a Service Based Architecture (SBA) that opened the doors to smarter cities with dedicated and distributed processing nodes. To fulfill the requirements of the services and to distinguish them evenly, the end-to-end architecture of a hosting mobile network must be prepared for different User Equipment (UE) profiles that connect to it. Section 3.1 explains the importance of having a common infrastructure, that all operators can share in order to save resources that can later be invested into their own business. Furthermore, the needs of a smart city by surveying smart city services; how it should take advantage of service differentiation in the cellular context; and lastly, how spreading out various processing nodes can benefit the existing ecosystem. To fulfill the details approached in this section, a set of requirements need to be met. These requirements are laid out and explained in section 3.2. An overview of the current infrastructure of the city of Aveiro is given in section 3.3, together with an explanation of how a cellular network architecture can improve the already implemented services. To finalize, a chapter summary is made in section 3.4.

3.1 Motivation

For a city to have cellular coverage of a specific operator's network, this operator must guarantee that it has enough equipment to provide such a coverage. With this in mind, operators must study a city map and decide the best points to guarantee coverage all around the city area. Also, in urban scenarios, it should always be considered that the number of buildings and density of users is much higher than in a rural scenario. The more obstacles the signal has to go through, the more loss there will be; this is called outdoor-to-indoor penetration loss. To resolve this issue, indoor facilities started using small cells - which aggregates the RU, CU and DU - to provide wireless network connection [38].

Nevertheless, what if all operators start choosing the same city points to assemble their Radio Units (RUs), even if not the same? What will happen if all operators have the same numbers of RUs across the city? Add to this scenario every big facility deploying its own group of small cells. This would be incredibly costly and undoable in a city infrastructure plan. From this, the concept of Neutral Host arose.

High-interactivity applications with low latency and high throughput requirements are anticipated to be supported by 5G. To provide low latency, the computing applications had to be moved closer to the end user, in a concept denoted by edge computing. This type of scenario allows to operate in a single computing platform, or a collaborative platform of multiple edge nodes that can work together with each other, and with the cloud computing nodes [39]. Since highly interactive applications that are computationally heavy and have high QoS requirements, including low latency and high throughput, require edge computing, the typical cloud computing architecture is insufficient. This inefficiency is primarily due to the distance to the end user, which also ends up costing more energy incurred in communication, by the offload of tasks and data to the cloud. Real-time services should be looked upon to understand further the real need of edge computing, such as the Vehicle-to-Everything (V2X) service which has, as one of its main requirements, the 5G vertical URLLC.

3.2 Requirements

The new generation of mobile networks presented a new type of architecture that came to fulfill new use cases, and reinvented service types with more demanding characteristics. However, before reaching the definition of an architecture for a specific Core Network (CN), it should always be kept in mind that a city's infrastructure must be prepared for multiple operators as much as for different services. The following section will provide ways to respond to this infrastructure issue.

3.2.1 Neutral Hosting

Neutral Hosting revolves around propagating public carrier networks across a privately deployed wireless network. This technology uses LTE or 5G for outdoor connection. This system is often cited to reduce costs and increase flexibility while maintaining network visibility and control. The architectures of LTE and 5G are very different, as explained in chapter 2. LTE follows a monolithic approach with all the components being interconnected and dependent on one another; this centralizes the solution and makes it very difficult to integrate external applications. However, with 5G the Next Generation Core Network architecture was reinvented into a SBA architecture, facilitating the decentralization and integration with external applications, since any authorized Network Function (NF) can access the services provided by any other control plane function [40]. This easy integration with external functions enables the connection of functions from external operators to a shared, neutral host network.

Sharing such a neutral host network across multiple operators can be done in the two crucial areas of mobile networks, with the concepts being called Multi Operator Radio Access Network (MORAN) and Multi Operator Core Network (MOCN). In both approaches, the RAN is shared, however in MOCN only one operator is sharing its CN resources with other operators; on the other hand in MORAN, each operator provides its resources through their isolated CN, only sharing the RAN. A city should provide a planned infrastructure with studied locations for small cell deployment to provide an easily shareable RAN, done through Open-RAN, joint with a CN that has already been set up to provide edge support. With its User Plane (UP) next to the deployed small cell, reducing the backhaul distance, it has an open door for operators to deploy their CNs in a way that the city's end user can have the best Quality of Service (QoS).

3.2.2 City Services

In a smart city, it is shown that the use of edge computing brings many advantages as discussed in section 2.5.1, but it also brings some requirements with it, which are:

- Real-Time Interaction: Ensures low and deterministic latency for the URLLC use cases, from V2X scenarios such as autonomous mobility and accident prevention, to services which require improved QoS. Edge servers deliver different services such as decision-making and data analysis from data gathered from local sensors, crucial to be made in real-time;
- Local processing: Instead of using cloud processing nodes, the usage of edge processing nodes will reduce the amount of traffic between a small cell and the CN, which ensures data governance;
- Higher data rates: Necessary as with the growing amount of edge services comes higher transmission of data to the edge clouds. In this scenario fiber cables between edge nodes can provide higher bit rates, together with improved cellular connection potentiated by mmWave antennas;
- High availability: Offloads cloud services to the edge, making them more easily available as the backhaul distance is reduced.

However, edge computing is not the only way to provide better QoS; the Core Network (CN) can be configured to differentiate itself according to a specific service. Although monolithic, the Evolved Packet Core (EPC) architecture may be developed so that its components can be replicated and dedicated to each service. This principle is perfectly executed by the 5G Core (5GC) through Network Function (NF) alteration in real-time through network slicing and, as previously mentioned, by function replication. The vital city services should be defined in advance to design the CN so that it can have defined functions/components for each service, providing load balancing in the network. Three examples of city services will be presented to further explore the capabilities of a 5G network in a smart city: Vehicle-to-Everything, Low Power Wide Area Network and Smart Grid. A discussion of V2X focusing in the verticals URLLC and eMBB, LPWAN for eMBB giving the example of monitoring a city, and SG for URLLC and mMTC discussing how it can connect all services in an energy perspective. There are many ways in which a complete 5G architecture can improve the development of a smart city. Figure 3.1 depicts a proposition on how a smart city can integrate with a cellular network in a cooperative environment. With these services and their respective needs in mind, this dissertation aims to deliver an architecture based on an already quite developed infrastructure in the city of Aveiro.

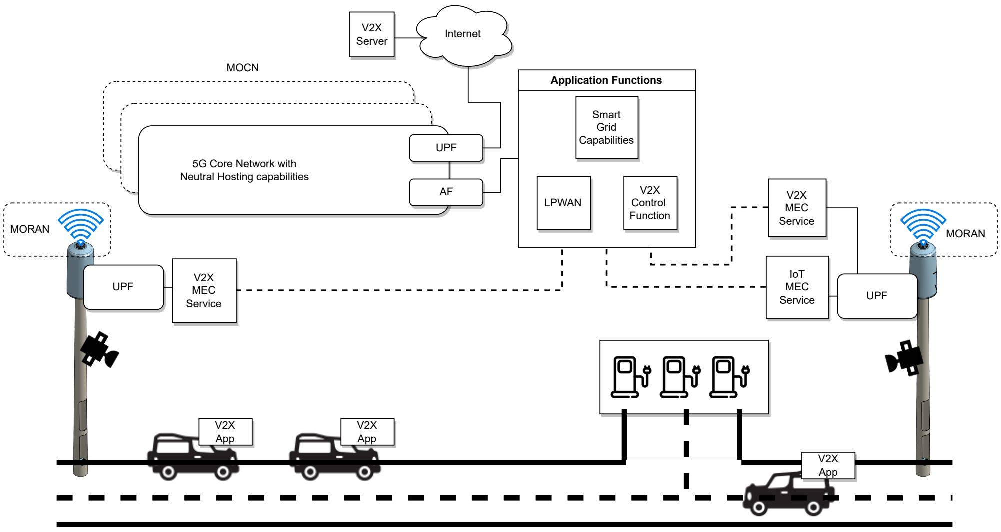


Figure 3.1: City services and cellular network integration proposition

V2X

V2X is one great example of a complete city service that focuses in a variety of aspects, from improving road safety, to saving energy and costs. ITS-G5 was the first technology to appear, in its stack it uses the IEEE 802.11p protocol at 5.9GHz. Cellular Vehicle to Everything (C-V2X) reuses the same band as 802.11p, 5.9 GHz, to implement the PC5 sidelink, while also supporting a cellular connection to increase coverage. In short range coverage, both technologies are quite similar [41], but the possibility to use cellular in C-V2X will allow a broader range of possibilities in its scenarios. Due to this, IEEE 802.11p has been compared to C-V2X. In [42] the authors conclude that it offers some advantages in scenarios such as traffic jam, vehicle speed and others, but C-V2X still needs some time to fulfill the long range ITS requirements. 3GPP proposes an architecture that envisions V2X services included in the EPC and 5GC, which further approves that C-V2X might be the way to advance [43]. V2X has different use cases due to its different communication types, these are usually bidirectional with some examples being vehicle-to-vehicle, vehicle-to-infrastructure, vehicle-to-pedestrian and vehicle-to-network. Some of the 5G C-V2X use cases are, following [44]:

- Infotainment: Making use of the 5G vertical, eMBB, data can be exchanged with the purpose to provide a pleasant vehicular experience offering video, audio or even live TV;
- Telematics: A service that allows vehicles to connect through mobile devices with application services. This allows for vehicle monitoring, automated parking and parking slot discovery;
- Road Safety and Efficiency: This can come in forms of road warnings which must be delivered in high speeds, information from sensor data, cooperative driving and advanced driving service and platooning. This latter one allows for vehicles that are close together to share information on acceleration, braking, heading and speed. All that was mentioned in this point is expected to be delivered in extremely high speeds; therefore, it will make use of the service vertical URLLC.

The availability of edge computing capabilities is critical to support the V2X use cases mentioned before. The lesser processing the vehicle needs to do, the better, reason why moving the processing to the edge is critical. One relevant scenario to consider is the handover, in which if a vehicle moves from one base station to another, and the MEC service moves along with it, then the delay present in this mobility is reduced. Following the vehicular networks example, some use cases require response times of under 120 ms; therefore, they cannot afford to have time losses in data acquisition. Furthermore, in these use cases the processing time of the messages has to be included in that reduced interval. According to [45] two examples of these use cases are: Cross-Traffic Left-Turn Assist, which states that a vehicle is alerted if another is crossing the lanes that it is using to turn left; Emergency braking, in which a vehicle is alerted in case another vehicle is braking.

LPWAN

An increase in Internet-of-Things (IoT) applications popularity leads to their discussion in the context of several private, business, and industrial spheres. The main application areas vary from sensor networks for Smart City infrastructures, smart homes, to energy supply and others. With this in mind, the fifth generation of mobile networks defines requirements for massive IoT in the context of massive Machine Type Communications (mMTC) [46]. Before the arrival of 5G, a very common solution to large numbers of IoT connections was LPWANs, which are expected to replace 2G/3G connections in the future [47]. The 5G system may offer LPWANs with the backbone connectivity and management services, due to its boundary-stretching performance measures, which can greatly lower complexity, cost, and increase the robustness of LPWAN deployment. Besides this, [48] also references a motivation for 5G-LPWAN integration, that LPWAN can provide a trade-off between the limitations of cellular technologies and the requirements of mMTC, which is due to the substantial number of connections, low power consumption, and wide coverage. Two examples of cellular LPWAN technologies are Long Term Evolution for Machines (LTE-M) and NarrowBand-IoT (NB-IoT), with the latter being standardized by 3GPP based on LTE. Even though it shares the same infrastructure as LTE, it can operate in NR and therefore is a future-proof mMTC technology. It is important to note that IoT devices usually offload the data retrieved from sensors to centralized cloud servers, as they have limited computational power and storage proficiency. But since these cloud servers may not meet latency requirements, edge computing is proposed to address these latency issues. The main idea behind this process is to take advantage of the integration of NB-IoT with cellular networks, to provide MEC servers with computing capabilities in one hop distance by radio resource allocation strategies, in order to provide an efficient, low-latency solution to IoT applications [49]. Usually, when mentioning LPWAN network and applications, the main focus is data gathering. One way to deal with data gathering issues is to somehow persist the data in a database and that will solve the problem, but what if the data is needed as soon as possible? In a city scenario, where there is an emergency and the data is needed almost in real time, having a MEC service in the edge will reduce the time of the data reaching the end user from the database. Not only emergency scenarios, but also the industry sector can leverage from this LPWAN technology, by having almost real-time data monitoring.

Smart Grid

The conventional electrical grid is shifting to a modern, internet-based, internet infrastructure, known as SG. It is an electrical system that includes a variety of processes and energy measures, such as smart meters, smart appliances, renewable energy sources, and energy-efficient resources. By utilizing digital communications and technology controls, the SG became a promising area of research with increased fidelity of power-flow control, energy reliability, and energy security [50]. An SG represents an ideal vertical for an extensive 5G deployment, as it covers the verticals of the new generation of cellular networks, according to [51]:

- Needs stricter latency, the greatest availability, and security include supervisory monitoring of mission critical activities (URLLC) such as, cyber monitoring, physical/aerial surveillance, fault localization, isolation/self-healing, and energy re-routing;
- Advanced metering of mMTC applications that provide the widespread and integration of end users' infrastructure while requiring higher standards for capacity and privacy;
- A combination of the aforementioned, such as in smart electric vehicle charging, where 5G technology should be able to combine and handle both low latency and high quantity; vehicle charging is one good example for low latency requirements as there may be hundreds of charging stations in a system and they need to constantly communicate to balance the load, or the system may overload causing failures.

3.3 City Infrastructure

This section describes the current status of the implemented platform in the city of Aveiro, as it was designed with three main goals in mind[1]:

- to develop a fiber-optic network connecting a communication, sensing, and computing infrastructure with radio terminals for short-, medium-, and long-range communication;
- to put in place a sensing platform that can comprehend the quality of the environment and how people behave in a city, and offer innovative approaches to effective traffic management, and people's safety;
- to give third-party partners a platform where they may test their own protocols, methods, prototypes, or look into the data they have acquired in order to develop new services and applications.

There are forty-four edge nodes scattered around the city of Aveiro that support the Aveiro Tech City Living Lab (ATCLL) system, with their locations marked in the map of figure 3.2. The switch, which is housed in the Instituto de Telecomunicações in Aveiro, supports 10 GbE connections and SDN capability, which aggregates the connections between these edge nodes using fiber link technology. These edge nodes are separated into two types of nodes, smart

[1] https://www.aveirotechcity.pt/en/about-us

lamp posts (lamp posts placed near roads) and wall boxes (boxes fixed in existing buildings). This infrastructure covers the majority of the city of Aveiro, where the demographic rate is the highest and therefore will be the main target of a complete 5G network deployment. By taking advantage of the mentioned nodes to provide cellular coverage, one can offer a capable neutral host infrastructure to share with network operators and a guarantee for all ranges of communications as mentioned in the first goal.

Figure 3.2: Fiber and Edge nodes distribution in Aveiro (from [52])

Every edge node has a unique set of components that enable multiprotocol communication, data processing and storage, traffic and environmental sensors, and other functions. Additionally, the entire collection of data can be transmitted via the fiber link towards the cloud data center housed at Instituto de Telecomunicações. They can be static, positioned in the edge nodes (with a fixed position), or mobile, positioned in moving nodes, depending on the sensor location (e.g. cars or buses). All the static nodes possess ITS-G5 and the transition to C-V2X has already begun with some of them already implementing it. Having C-V2X implemented in these edge nodes paves the way for a complete cellular network integration with this service, which then enables all the benefits mentioned in the previous section.

Radars, cameras, and other components have been installed by the infrastructure in a number of edge computing nodes, enabling the collection of data on people's flow and traffic. This data is gathered and processed in real-time at the edge, enabling for visualization in the ATCLL dashboard[2]. With the escalation of the amount of data gathered, there will be a need for greater bandwidth - to support all the sensors - and faster connections, to and from the edge nodes, as depicted when the context of IoT in 5G mMTC was mentioned. This leads to

[2]https://aveiro-living-lab.it.pt

the need of having cellular network enabled MEC services, and IoT application integration, to respond to the addressed requirements. Public buses (which contain Onboard Units (OBUs) of this infra-vehicular network), garbage collector trucks, and local boats (moliceiros) are included in the ATCLL as mobile nodes for mobile sensing and have several environmental sensors installed on them. Finally, with all this, the ever growing need to have more control over all that is happening will arise. Hence why, having an already ready cellular infrastructure to support a SG capable of such control will ease future deployments of new services and objectives. This infrastructure can be followed for a more precise consultation in [53].

The neutral hosting aspect of this groundwork will not only enable multiple operators to take full advantage of a smart city, but empower a forever growing cycle of evolution for both parties. This is because, as providing itself as a service, the neutral hosting infrastructure can bring revenue that leads to self improvement, as stated in the third main goal. The proposed cellular network architecture is prepared to be utilized in the city environment, once every edge node has a Radio Unit, or small cell, in it. This work was developed as if all edge nodes had one, meaning that this dissertation will give a disaggregated CN thought to accommodate an UP in each needed edge node, if need be. This way, the city would have full private cellular coverage, with neutral hosting capabilities for MOCN and MORAN scenarios, with mobility ready exit nodes dedicated for specific services, such as C-V2X and NB-IoT, in specific points of the city. The proposed technological architecture is depicted in figure 3.3: it presents both the approach aimed for the city and a simulator for internal testing.

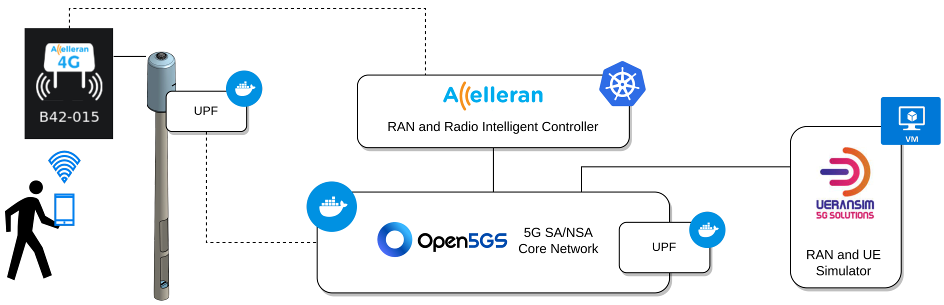


Figure 3.3: Proposed technological architecture

3.4 Summary

With the gathered information from the previous chapter on Core Network (CN) and edge services, this chapter firstly illustrated the need for a city infrastructure to be ready for a cellular deployment that can be shared by all operators. With a purpose set, the requirements to start this chapter were explained following two concepts that serve as foundations for the architecture that will be defined, being Neutral Hosting and the need for service differentiation together with them being deployed in the edge. Lastly, the infrastructure of the city of Aveiro, ATCLL, was presented. Whilst depicting the existing services of this infrastructure, a connection was made to the previously mentioned city-services. The goal of this description was to show the relevance of a well thought definition of requirements, before the implementation of a real CN and testbed. With this in mind, the devised testbed will be presented in the next chapter.

CHAPTER 4

Real Deployment

This chapter introduces how the end-to-end architecture previously discussed was designed and implemented, keeping in mind the needs of a smart city. Section 4.1 presents the used hardware for the RAN and how it can enable the desired scenarios. Section 4.2 gives an insight on the Core Network (CN), explaining which nodes were deployed, what they do and the UE path created inside the CN. It will also be explained the implementation of the 4G/5G NSA, the 5G SA core functions, and the node selection in the UPF. Since the selected core implements both 4G and 5G functions, both CNs were studied as in section 4.3, showing basic flows of EPC functionalities, and doing the same for the 5GC in section 4.4. After this, section 4.5 presents the equipment used for the testbed and how it was implemented to provide valid testing scenarios. Lastly, section 4.6 concludes the chapter.

4.1 Access Network

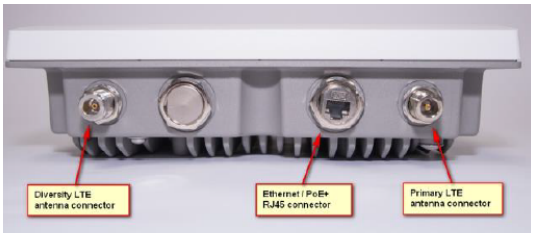
This section will address the used hardware and software to implement the RAN. This includes the small cells and the RAN Intelligent Controller (RIC) that manages connections and radio capabilities. For further insight, it will be explained what can be configured in the small cells and their pre-set configurations, what some of the parameters mean, and how they may affect the scenarios tested as described in the next chapter.

4.1.1 Small Cells

To provide radio capabilities to the architecture, a set of small cells from Accelleran[1] was used in two different locations: 2 units were located in the researching institute where this study was conducted, and another in the city. The pair of small cells located in the institute are in band 42 with Time Division Duplex air interface, and the other works in band

[1] `https://accelleran.com/`

7 with Frequency Division Duplex air interface. These small cells are constituted by two LTE antennas, an Ethernet/Power of Ethernet connector and a GPS antenna connector. The small cell physical ports can be seen in figure 4.1.



(a) Top

(b) Bottom

Figure 4.1: Small cell physical ports

This small cell type is indicated for local and urban areas and enables MEC/5G scenarios. Besides this, it can also be deployed in a disaggregated and virtualized RAN (vRAN). To make full use of the virtualized RAN capabilities, it makes use of its Open-RAN compliance and has a Near Real-Time RIC also from Accelleran, named dRAX.

The eNB configuration files, available through dRAX, present some parameters relevant to the handover procedure explained in the next chapter of this document. This dissertation works with an LTE network, which uses two important concepts when studying eNB interactivity, for instance, as in handover scenarios. These concepts are inter-eNB and intra-eNB: intra-eNB handover has both the source and target cells in the same LTE network, whereas inter-eNB has the source and target cells in different LTE network. This work scope falls upon working in the same network therefore, it follows intra-eNB. The next section will further explain the eNB's parameters and the functionalities of the Near Real-Time RIC.

4.1.2 dRAX

Following open standards like Open-RAN, dRAX supplies tested virtualized software components that enable real-world deployment of multi-vendor, disaggregated Open-RAN. These cloud-native components provide dependable, affordable, and scalable solutions for both 4G and 5G networks. The dRAX components are pre-integrated with a variety of Distributed Unit and Radio Unit solutions from partners and other vendors in the RAN ecosystem, and they cover the essential control and resource management activities of the RAN (Service Orchestration, RIC, CU-CP, and CU-UP). It uses Kubernetes as its orchestration platform in a fully containerized and ready-for-deployment manner. As mentioned the dRAX provides an Open-RAN-compliant completely cloud-native near real-time RAN Intelligent Controller (RIC) that enables the near real-time management and optimization of Open-RAN components and resources. All the components from the ecosystem can be seen in figure 4.2, such as the RIC, a SMO, that such as mentioned in section 2.2.2, provides manageability and orchestration, as well as user interface for monitoring and configuration of the RAN elements.

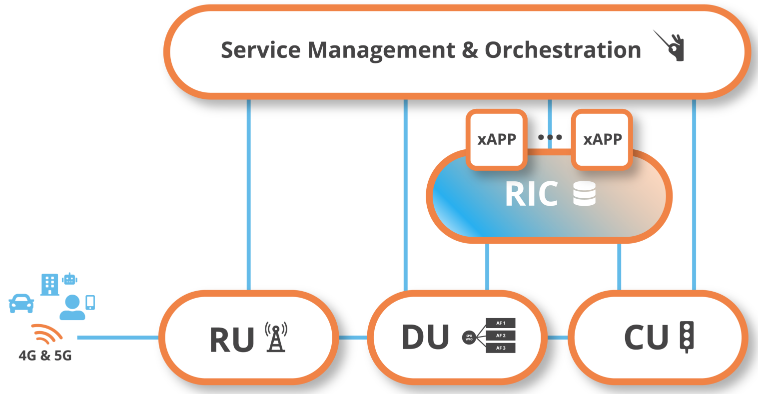


Figure 4.2: dRAX RIC simplified architecture (retrieved from [54])

Finally, the CU and DU are software-based and provide layer 1, 2 and 3 of the RAN together with security features, whereas the RU, seen in the previous section, is the only non-software module.

For the scope of this dissertation, the RIC was not used for its `xApp` integration capabilities but for small cell management and configuration. It allows configuring radio parameters, cell-specific parameters, as well as some other features. The radio configurations are supposed to be set according to the band that the small cell represents; if set wrong, the cell can work erratically. Some examples of these parameters are downlink E-UTRA Absolute Radio Frequency Channel Number (EARFCN), downlink bandwidth, physical cell id, output power, and the cell's band. For cell-specific configurations there are the following parameters to consider:

- Cell ID: Usually obtained by multiplying the physical cell identifier by 256, as it identifies other cells and UEs;
- Tracking Area Code (TAC): Enables MEC, and other scenarios by associating the small cell to a geographical location; the operator, in a logic at choice defines this location;
- Mobile Management Entity (MME) connectivity: allows to reference a MME IP and port to which the small cell will connect to. Multiple MMEs can be defined for load balancing or multi-operator purposes;
- Serving PLMN: Allows to set the PLMN that the cell serves, which must be associated to one of the MME configured previously; multiple PLMNs can be set, but at least one of them must always be defined as primary.

For the features set, there are only three that are set to boolean values of `true` or `false`, hand-in, hand-out and Simple Network Management Protocol (SNMP). The first two serve to define if the cell is to perform handover procedures and the latter for network management; this feature was not tested.

Table 4.1: dRAX small cell configured values

Type	Parameter	Value
cell-selection	`q-rxlev-min-offset`	2
cell-selection	`q-rxlev-min`	-65
cell-selection	`p-max`	23
cell-reselection	`q-hyst`	4
cell-reselection	`thresh-serving-low`	31
intra-frequency-carrier	`q-rxlev-min`	-70
intra-frequency-carrier	`neigh-cell-config`	`no-subframes`
intra-frequency-carrier	`t-reselection-eutra`	0
intra-frequency-carrier	`presence-antenna-port-1`	`TRUE`
intra-frequency-carrier	`cell-reselection-priority`	7

The eNB has a configuration file with various parameters that are relevant for the study of handover scenarios. The values of these parameters appeared to be unalterable, but it was essential to understand what they meant, and what influence the set values have in the behavior of the cell. The parameters [55] that were found pertinent, and their respective information, are the following:

- `q-rxlev-min-offset`: Offset to q-rxlev-min that is only taken into account as a result of a periodic search for a higher priority PLMN while camped normally on a visitor PLMN;
- `q-rxlev-min`: This parameter is the minimum RX (receiver) level required in a cell, and its unit is dBm (measured RSRP). Value is calculated in dBm/0.5;
- `p-Max`: this parameter is used to limit the allowed UE uplink transmission power on the serving cell.
- `q-Hyst`: hysteresis value applied to serving cell for evaluating cell-reselection ranking criteria. It is for cell reselection when RSRP values are used in the evaluation.
- Thresh Serving Low: threshold for serving frequency used in the evaluation of reselection towards lower priority E-UTRAN frequency. The actual value in dBm is obtained by multiplying by 2.
- Cell Reselection Priority: cell reselection priority of the serving frequency. A value of 0 indicates the lowest priority.
- Neigh Cell Config: provides the information related to TDD UL/DL configuration of neighbor cells.
- `t-Reselection` EUTRA: cell reselection timer for intra-frequency E-UTRAN cell reselection. Value in seconds.
- Presence Antenna Port1: indicates whether all the intra-frequency neighboring cells are configured with at least two antenna ports, can be `true` when all the intra-frequency neighboring cells are configured with at least two antenna ports, or `false` when one of the intra-frequency neighboring cells is configured with only one antenna port.
- Cell Reselection Priority: cell reselection priority of the serving frequency. 0 indicates the lowest priority.

There are more parameters for inter-eNB, but since it is not on the scope of the scenario, they will not be referred. The values set for the used small cells can be seen in table 4.1.

4.2 Core Network

The Open5GS [27] solution was the Core Network (CN) selected to deploy the 4G/5G Non Standalone (NSA) and 5G Standalone (SA) network functions. This solution was chosen because, as mentioned in the end of chapter 2, it provides a complete documentation joint with relevant features for this dissertation, and an active community and developers. This is an open-source project comprised of two main planes: the Control Plane (CP) and the User Plane (UP). These planes are physically disaggregated in a Control and User Plane Separation (CUPS) practice. The Open5GS solution was integrated with the previously presented RAN to provide an end-to-end mobile network architecture. This section intends to give insight into Open5GS's components, more specifically into which architectures they are meant to be deployed. Also, the protocols they take advantage of and their main purpose, to finalize it, this section explores how Open5GS defines the path between nodes, and how some nodes can be selected inside the CN.

4.2.1 Open5GS software architecture

As mentioned before, this core solution implements both 5G SA and NSA architectures; these architectures can be deployed together or separately. The 5G NSA architecture components must be configured with the network identification (IP, DNS) of other components they connect to, while the 5G SA architecture makes use of its Service Based Interfaces (SBIs) to be configured with the identification of the Network Repository Function (NRF). This function will then manage the connection to the other functions.

Open5GS makes use of a database made in MongoDB[2] to hold subscriber data. This database is common in both architectures, but is managed by different identities. Even though these architectures can be deployed together, they do not have interworking capabilities. Both architectures with the CUPS scenario are depicted in figure 4.3.

Following figure 4.3 and focusing on the NSA architecture, with its modules presented in green, Open5GS implements standard interfaces defined by ETSI. The components implemented are:

- **Mobile Management Entity (MME)**: Represents the Control Plane (CP) hub for the core. It mainly manages sessions, mobility, paging, and bearers. It links to the HSS, Serving Gateway Control Plane (SGWC) and eNBs/NSA gNBs;
- **Home Subscriber Server (HSS)**: Generates SIM authentication vectors and holds the subscribers' profiles, as it is a manager for the subscriber database;
- **Policy Charging Rules Function (PCRF)**: Enforces subscriber policies and handles charging, retrieves the information from the subscriber database, and uses the SMF to enforce the policies on a UE session;
- **Serving Gateway Control Plane (SGWC)**: Mandated by the MME it is serving, creates sessions and acts as a gateway server to the Control Plane (CP), shares the session management responsibilities with the SMF, and links to the Serving Gateway User Plane (SGWU);

[2] `https://www.mongodb.com`

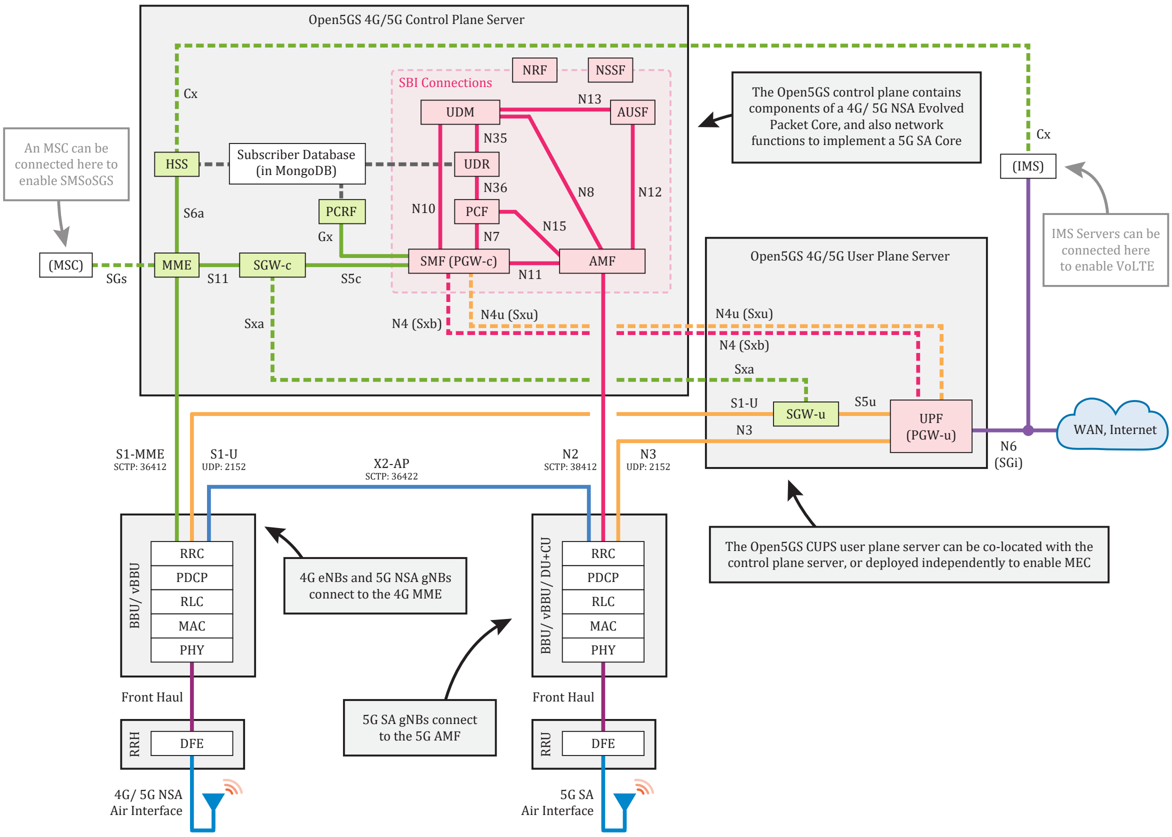


Figure 4.3: Open5GS 4G (in green)/5G (in pink) CUPS architecture from Open5GS website [27]

- **Serving Gateway User Plane (SGWU)**: Managed by the connected SGWC it is serving and links to the eNB/NSA gNB; it holds information on the session created in the SGWC and carries user data packets that come from the RAN to the UPF; can be disaggregated from the CP;
- **Packet data network GateWay - Control plane (PGW-C)**: Demonstrated in figure 4.3 as SMF; it also has a UE session that manages accordingly with the SGWC, that is shared with the UPF to which it is connected; it is also connected to the PCRF and with it enforces policies on the session;
- **Packet data network GateWay - User plane (PGW-U)**: Demonstrated in figure 4.3 as UPF; carries the user data packets to the DN; can be disaggregated from the CP, and if so in conjunction with the SGWU enables Multi-access Edge Computing use cases.

Open5GS also allows third party SMS and IMS servers to connected to the CN. The first can be directly connected to the MME while the latter uses the HSS functionalities to interact with the subscriber database. The SMS functionalities where tested using OsmoMSC[3] and do work, allowing to send SMS to other users registered in the HSS.

The 5G SA architecture of Open5GS, displayed in pink in figure 4.3, also implements standard interfaces that are reference points for the different function connections in the 5GC.

[3] https://osmocom.org/projects/osmomsc/wiki

The implemented functions are:

- **Access and Mobility Management Function (AMF)**: Handles connection and mobility management, which is a subset of what the MME was tasked with; holds connection with the gNB; since it is the main function of the 5GC it is connected to almost all the functions, except the UPF, UDR, NRF and NSSF;
- **Authentication Server Function (AUSF)**: Consults the UDM and authenticates the connecting UE;
- **Unified Data Management (UDM)**: Interface for the UDR; provides subscriber information to the AUSF, SMF and SMF; generates SIM authentication vectors holding subscriber profile together with the AUSF;
- **Unified Data Repository (UDR)**: Retrieves subscribers' profile information from the database and provides it to the UDM;
- **Policy Charging Function (PCF)**: Similar to the PCRF, charges and enforces subscriber policies; it retrieves the information directly from the UDR and applies it in the session managed by the SMF;
- **Session Management Function (SMF)**: Handles all the management of the UE session, which was responsibility of the MME, SGWC and PGW-C; Manages and chooses the existent UPFs;
- **User Plane Function (UPF)**: User Plane (UP) unique function, carries out user data packets between the gNBs, connected to it, and the DN;
- **Network Slice Selection Function (NSSF)**: Responsible for selecting the appropriate network slice for the UE;
- **Network Repository Function (NRF)**: Hold registration of all the existent functions in the 5GC and helps aforementioned functions to discover each other;
- **Binding Support Function (BSF)**: Simple function to support the binding of the functions to their interfaces.

4.2.2 Node Selection

Core Networks can be implemented in a way that allows node selection. This means that some nodes have the ability to choose other nodes in the context of a user session. This selection is according to the parameters of a UE's session, meaning that this selection is only done upon the UE registration to the network. For this selection to take place, the nodes have to be previously connected to each other. Enabled by this node selection, the ability to disaggregate the User Plane in a way that UP does not need to be in the same host as the CP grows stronger. If the CN is able to choose disaggregated nodes according to a session, it can provide a service with its needed nodes at less hops of distance. One example is the MEC use case, in which the pertinent node to be selected would be the UPF. In this case, by choosing an UPF that is deployed conjunction, or in less hops, with the MEC service one can reduce latency times.

Open5GS nodes are implemented, so that node selection is allowed. There are several selection modes available: APN/DNN; TAC/Tracking Area Information (TAI); Cell ID and NR Cell ID; and S-NSSAI. All these are contained in the session of the UE. Besides these modes, there is also a round robin mode to be used if none of the above is specified in the nodes that are available for selection. The idea behind, it is that the selector node keeps a record of the last selected node and, upon the next UE connection, selects the following node in the list; once the whole list is selected, it returns to the beggining, as in a circular buffer. Another use case is, if the available nodes have the same parameter defined, a APN for instance. When such scenario happens, Open5GS makes a round robin selection by all nodes with the same APN. The goal of this algorithm is to apply some load balancing to avoid overload in a certain node; it can be turned off if the core operator desires it. The available nodes for selection and their available selection methods can be consulted in table 4.2.

Table 4.2: Open5GS component selection

Component/Function	Selected by	APN/DNN	TAC	Cell ID	NR Cell ID	S-NSSAI	Round Robin
SGWC (NSA)	MME	✗	✓	✓	✗	✗	✓
SGWU (NSA)	SGWC	✓	✓	✓	✗	✗	✓
PGWU (NSA)	PGWC	✓	✓	✓	✓	✗	✓
SMF (SA)	AMF	✓	✓	✗	✗	✓	✓
UPF (SA)	SMF	✓	✓	✓	✓	✗	✓

All these parameters have already been approached in the scope of this dissertation with the exception of Tracking Area Information, which is the combination of Tracking Area Code with PLMN, usually used when there are multiple networks and/or operators. This is defined in the MME and AMF allowing to distinguish them from other MME/AMF thus creating new networks. Each will then have the corresponding RAN connect to them, serving their specific network.

As attested in table 4.2, SMF is the only component/function that can be chosen according to S-NSSAI. This selection is made differently from the other mentioned nodes: each SMF is described by a set of values that are then called `SMFinfo`. This `SMFinfo` is constituted by S-NSSAI, DNN and TAI, and is sent to the NRF. During the Network Function (NF) discovery process the NRF responds to the AMF with the `SMFinfo` of each SMF and, based on the UE session information, the AMF chooses the SMF. The selection of the SMF is highlighted in this paragraph because in section 2.6, [28] states that this selection is not possible. From this, and returning to table 2.1, one can attest that Open5GS checks all the boxes.

The nodes mentioned in table 4.2 are selected in an OR logic, that is, the node will only be selected according to one of the previous parameters. This selection is then made following an algorithm that gives different priorities to each parameter; since UPF is the function with more selection manners, its selection diagram can be seen in figure 4.4. It should be noted that all the nodes in a selection must be associated with Packet Forwarding Control Protocol (PFCP); this means that the selector node must be associated, via PFCP, with the to-be-selected nodes; otherwise, even if a node corresponds to the session values, it will not be taken into account.

These various forms of selection, and nodes to be selected, grant the proposed architecture with versatile paths for the UE, which then provides load balancing and service differentiation in

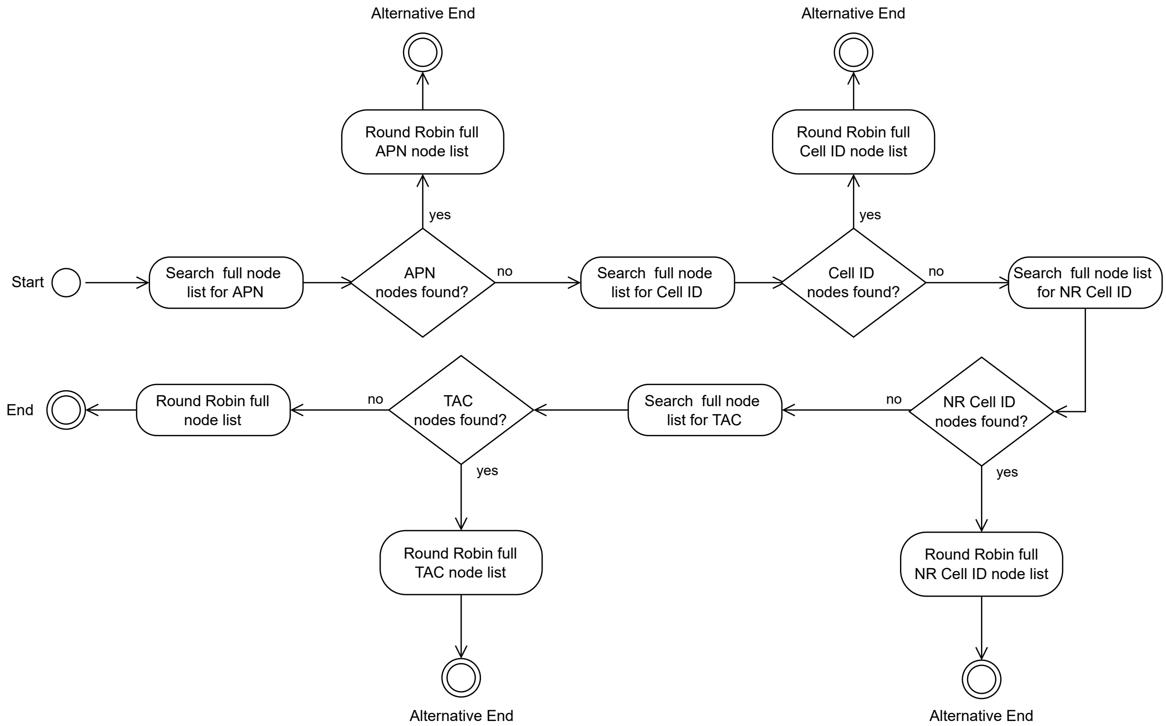


Figure 4.4: Open5GS UPF selection algorithm diagram

both SA and NSA architectures. The ability to differentiate services, whether by geographical area or APN, for instance, is a significant advantage in city scenarios with different needs for different users, in different regions. This is why this node selection was studied in more depth to understand and give the aptness to assemble different CN architectures for different city scenarios, either being mobile or stationary. In these type of scenarios, V2X shows the need for an edge service to accompany the UE, or V2X application, whilst the vehicle is moving. Showing that there must be a selection and reallocation of specific nodes. Furthermore, in more stationary scenarios such as IoT sensors, the selection of a node could provide CN paths directed to the edge database, per example.

4.3 5G NSA Procedure Flows

This section will show how Open5GS's EPC nodes react to some basic operations in mobile networks; this data was retrieved through analysis of packets exchanged between each node.

The packets were captured using a library for network traffic capture named `tcpdump`[4] in each node's container. The represented order in each table is the one the packet appears in the capture. Through the analysis, there are quite a few messages with the content `SACK` that identifies the use of selective ACK, informing that all segments have arrived successfully. In practice, this informs the sender of the last fragment received. The purpose of presenting these flows is to provide an insightful view on the core's functionalities, by executing some functional tests on basic procedures such as eNB association, UE attachment and detachment

[4] `https://www.tcpdump.org`

and other not so simple, such as an handover. The procedures presented were selected as they are needed for every real deployment of an end-to-end cellular network. Moreover, some explanations of the protocols used by these procedures will be given.

4.3.1 eNodeB Association

In order to test this scenario, a reboot was made to a small cell to see its reattachment to the MME; this is why there is an `ABORT` message in the first exchange in figure 4.5. After reboot, the cell initialized and proceeded to establish a connection with the MME, an event in which one can verify two cookie messages appearing: this happens during the initialization of an association and means that the sender of the `COOKIE_ECHO` chunk pretends to report unrecognized parameters. After the association procedure, the cell and the MME keep trading `HEARTBEAT` messages to attest that the address that the original `HEARTBEAT` was sent to is now considered available for normal data transfer. For more information on the Stream Control Transmission Protocol (SCTP) consider visiting [56]. Besides SCTP, this procedure also uses S1 Application Protocol (S1AP), this protocol is used between the eNB and the MME in order to support operations such as E-UTRAN bearer management, transfer of UE context information, NAS signalling transport, paging, and EPC based mobility. When developing a cellular network there are many peaces to be assembled, and they are not always from the same manufacturer. This may lead to error in joining the pieces together, to understand where it failed this flow exemplifies a successful association to understand where it may have failed.

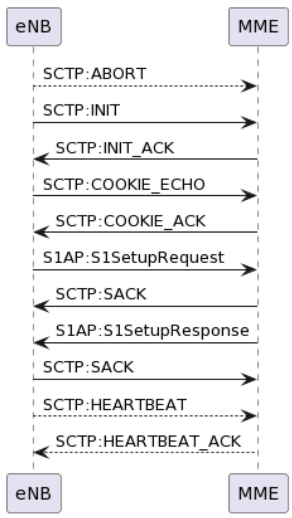


Figure 4.5: eNodeB Association flow in MME

4.3.2 UE attachment

For this test, airplane mode was turned off and on in order to reinstate the cellular interface in the UE. This causes the UE to re-register in the core: firstly the UE communicates with the eNB, then a message is sent to the MME to begin the process in the CN. As seen in figure 4.6, the cell sends an attach request via `S1AP/NAS-EPS` protocols, and then provides the UE's information. The S1AP protocol was explained previously, and NAS was referenced in previous section 2.1.1. Between the core's nodes the used protocol is GTP-C, as it was stated in section 2.5.2. In this specific case GTPv2 was used, which is an updated version of GTP-C. Figure 4.7 illustrates the messages exchanged between the SGWC and other nodes, using GTPv2 and PFCP. The prefix PFCP, or "PFCP session", indicates that messages and procedures are common and used on Sx and N4 reference points and Sx or N4 sessions. It is used to establish associations or sessions between EPC and 5GC nodes according to 3GPP TS 29.244 version 16.5.0 Release 16 [57].

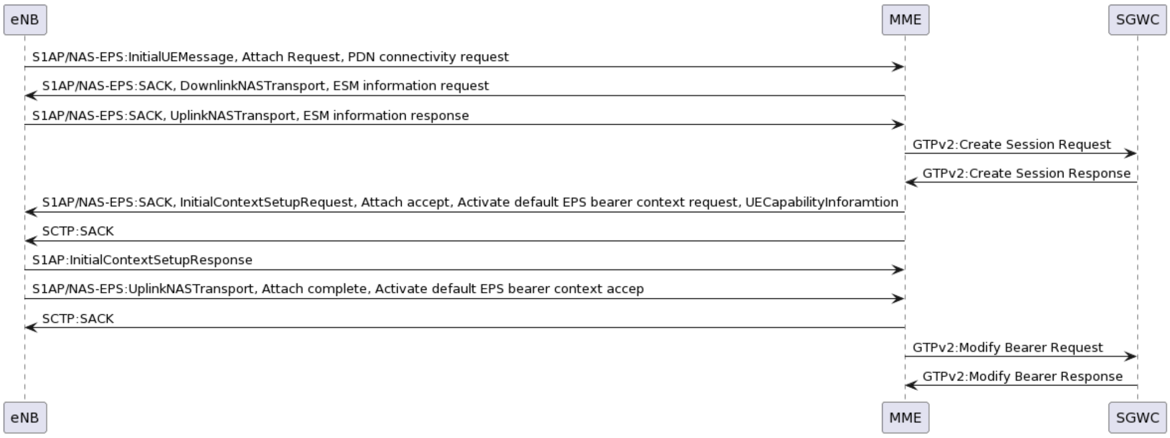


Figure 4.6: UE attachment flow in MME

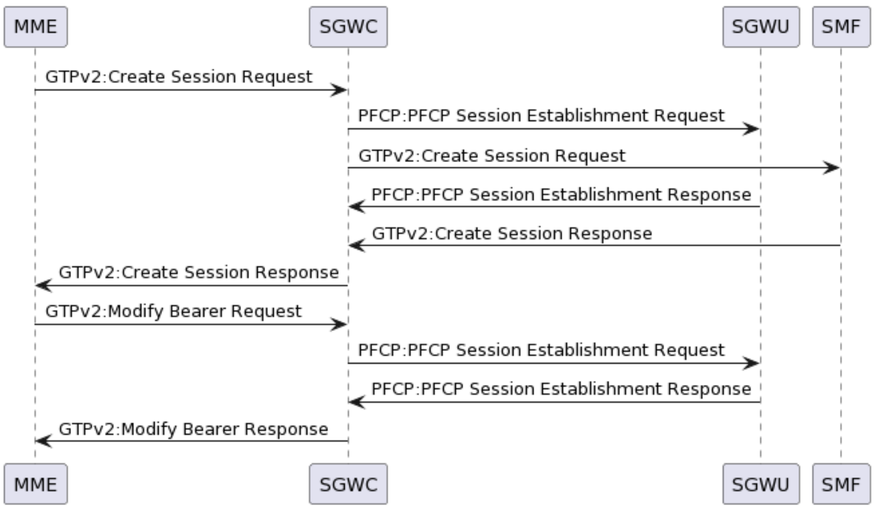


Figure 4.7: UE attachment flow in core network

The session creation request/response, PFCP session establishment/modification and modify bearer request/response are also be seen in the SGWU and the SMF, to both of them these messages are sent from SGWC. PFCP session messages are also sent from the SMF to the UPF.

When a UE attaches to the CN, the MME communicates with the HSS via `Diameter` protocol to check if the subscriber exists in the database.

The PCRF communicates only with the SMF and the database. To the SMF it uses `Diameter` protocol, and Security Sockets Layer (SSL) or Transmission Control Protocol (TCP) to communicate with the database. It uses the SCTP protocol to send `HEARTBEAT` messages.

4.3.3 UE detachment

The results from this test were retrieved from the previous in which airplane mode was turned off and on, causing the UE to be detached. There are multiple reasons for why a UE would initiate a detach procedure, such as the UE being turned off, a SIM card being removed from a UE, or the UE attempting to use a non-Evolved Packet System (EPS) service (like SMS). This procedure is depicted in figure 4.8, which is triggered by the UE, then the eNB sends a message with a detach request to the MME, the MME then sends session deletion requests to the SGWC and upon response releases the UE's context. In figure 4.9 it can be seen that the deletion request sent to the SGWC is propagated throughout the other nodes, that hold information on the UE's session, SGWU and the SMF. After deletion of the PFCP sessions in the CN the SGWC allerts the MME to procede.

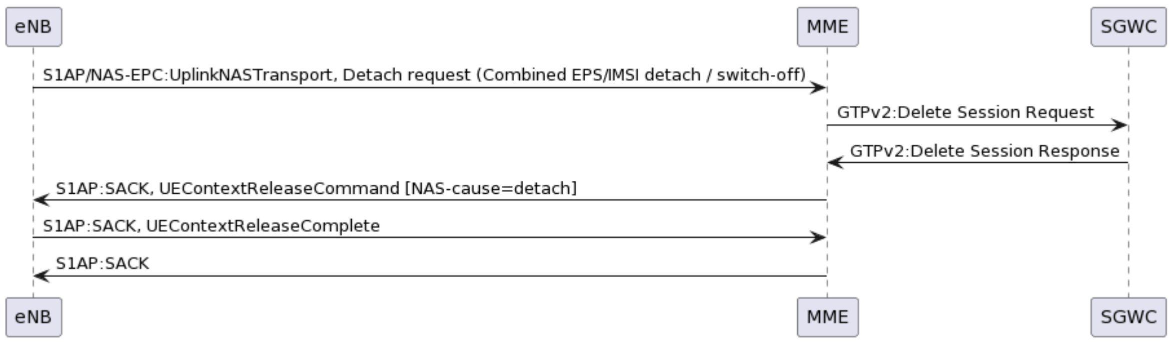


Figure 4.8: UE detachment flow in MME

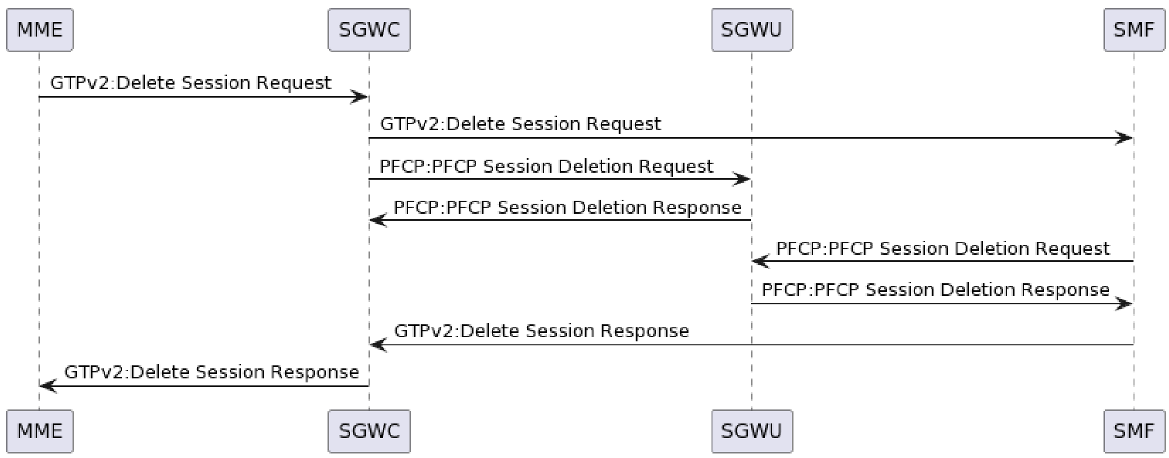


Figure 4.9: UE detachment flow in core network

4.3.4 Unregistered UE

In this procedure, the UE is removed from the database in order to test how Open5GS's MME deals with an unknown UE trying to connect. The core rejects the connection attempt, but it always needs to attest that the UE is not actually registered in the database.

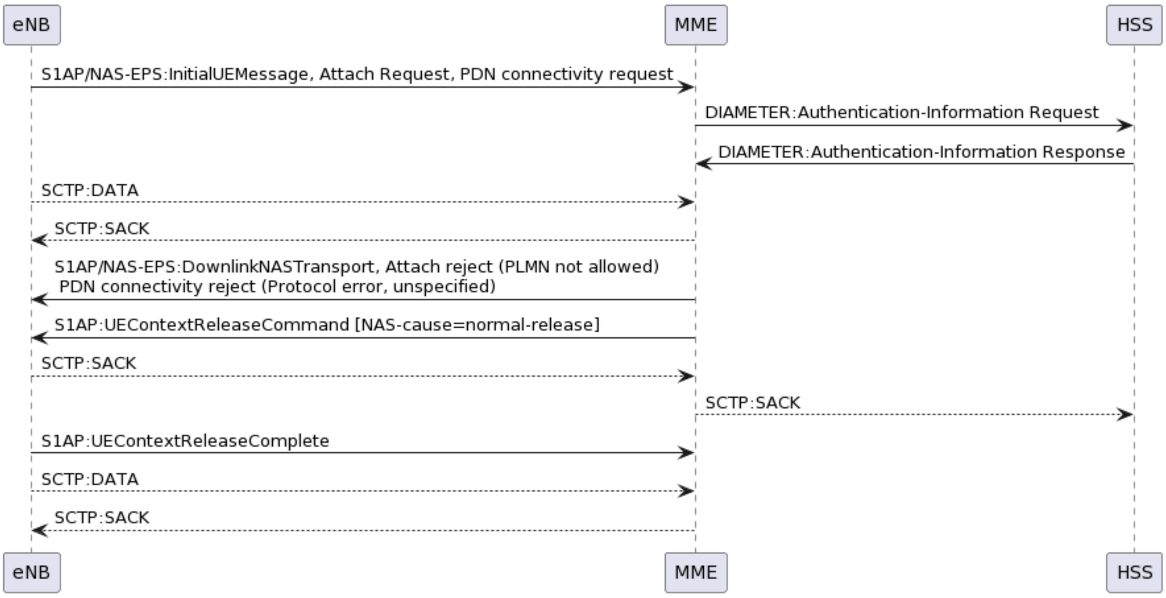


Figure 4.10: Unregistered UE attempt of connection flow

The attempt of registering flow can be seen in figure 4.10. The information regarding the UE is stored in the HSS and this is why the MME request information to it to check if the UE is registered. Upon negative response, the MME reports back to the eNB rejecting the attach and a command through NAS protocol to release the UE. The reason for this release command is that since the UE is not registered, there is no need for it to be camping on the cell. Therefore the eNB releases the UE for it to connect to a cell serving its corresponding operator.

4.3.5 Handover

As referenced in section 3.2.2, the handover is a very important feature for all the services that require mobility, especially V2X. In an handover, it is of major importance that there is no data loss, but besides this fact it would also be relevant that there would a modification of session to a new endpoint. This way the performance of the connection would not degrade with the distance. In the mentioned section it is also approached the change from IEEE 802.11p to C-V2X, but there is also the possibility to perform seamless handover, with the support of SDNs, in vehicular networks - vehicular ad-hoc networks - as referred in [58]. The handover flow can be seen in figure 4.11 and further on in the results section 5.1.2. As the RAN manages most of the procedure of switching from one eNB to the other, the core is only requested to manage the sessions for that specific UE. This is an `X2` handover and was mentioned in section 2.5.2. This is the reason why, in figure 4.11, the first message is already the target-eNB requesting the UE initial message, to which the MME responds with UE's information. With this set in order, a release command is sent to the source-eNB via NAS protocol to release the UE context. After the release is complete, the MME needs to set up the UE context once again and modify the bearers. In order to do this, the PFCP sessions need to be modified, so the SGWC that previously received an order to modify the bearers by the MME sends a message to the SGWU to perform this modification. However, as mentioned before, the handover intended to switch the User Plane, meaning that a new SGWU, represented by the TAC of the target-eNB, should have been chosen. This does not happen, as only the SGWU representing the source-eNB received modification requests sent by the SGWC. Nevertheless, the SGWC also did not send any session modification requests to the SMF as it does in session creation/deletion, shown in the previous sections 4.3.2 and 4.3.3. This meant that the SGWU was not changed; therefore, the User Plane also maintained the same. The way that Open5GS's CN and dRAX's RAN dealt with this situation was by creating a GTP-U tunnel from the target-eNB to the currently used SGWU. All this ultimately results in almost no packet loss during the handover, but the User Plane reallocation does not happen.

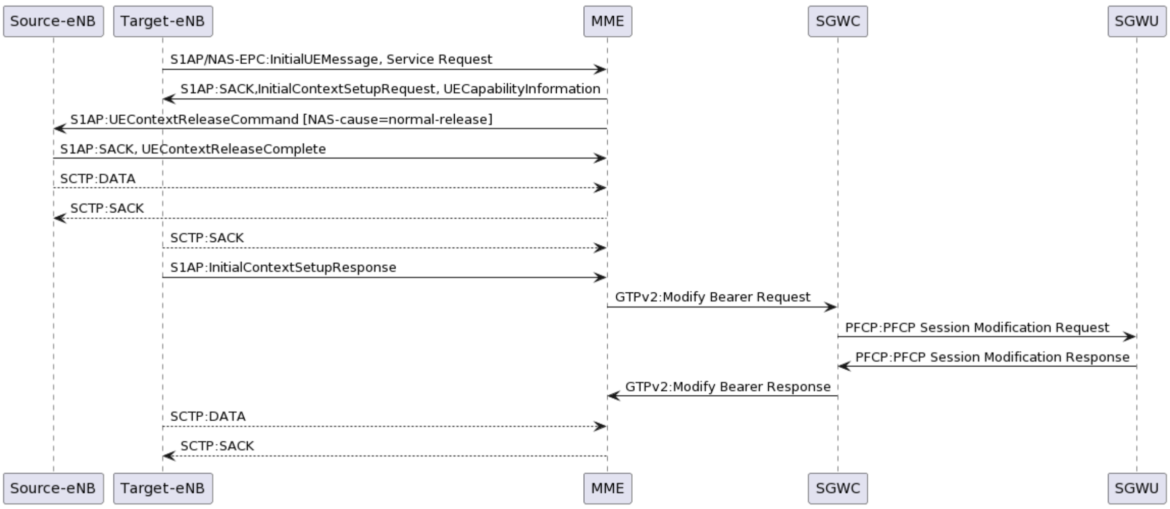


Figure 4.11: Handover flow

4.3.6 Conclusion

In table 4.3 a summary of the used CN components of each procedure is shown. From this table one can attest that the MME is present in all the tested procedures, this is one of the main characteristics of the EPC as it gave the majority of the responsibility of UE management to the MME. One interesting fact in Open5GS is that, as it reuses the SMF from the 5GC the sessions are created in duplicates. Therefore, when a UE is interacted with, the session must be arranged in all the involved components: SGWC, SGWU, SMF and UPF. This is true for all the tested procedures except for the handover in which the SMF session is unaltered, which may be one of the reasons as to why there is no UPF reallocation. The UPF does not appear in the depicted flows as it would be repeated information, as the SMF replicates with the UPF the procedures made between the SGWC and SGWU. The HSS is only used for UE verification on registration procedures, as all the session management is not of its responsibility. Finally, the PCRF is used solely in the UE attachment procedures as the rules defined to it are established upon the session creation.

Procedure	MME	HSS	PCRF	SGWC	SGWU	SMF	UPF
eNodeB Association	✓						
UE attachment	✓	✓	✓	✓	✓	✓	✓
UE detachment	✓			✓	✓	✓	✓
Unregistered UE	✓	✓					
Handover	✓			✓	✓		

Table 4.3: Summary of the used CN components in each procedure

4.4 5G SA Procedure Flows

This section will show how Open5GS's 5GC nodes react to some basic operations in mobile networks; this data was retrieved through analysis of packets exchanged between each node. The same network capture tool was used than before. The represented order in each table is the one in which the packet appears in the capture. UERANSIM, referenced in section 2.6, was used to represent the RAN and an UE. This simulator implements a 5G SA UE and 5G SA RAN, more precisely the gNB-CU. It presents some limitations as it does not implement PDU session modification or any handover, which was pertinent to the scope of this thesis. Nonetheless, it is still a great tool for understanding standard 5GC procedures. The feature set for this simulator can be found in the project's wiki on GitHub[5].

4.4.1 gNodeB Association

A single gNB process was initiated to perform this test through UERANSIM. The configuration files for this gNB were the default ones, with the only parameter altered being the IP of the AMF to which the gNB associates. The association flow of a gNB is similar to the eNB explained in section 4.3.1. Although, instead of using S1AP to send/receive the setup messages, the used protocol is Next Generation Application Protocol (NGAP). This protocol is found in the `N2` reference point between the gNB and the AMF. The main purpose of this protocol is to establish the communication between the RAN and the CN; some examples are configuration updates, UE context transfer, PDU session management, and support for mobility procedures. Besides this, it transports downlink and uplink NAS messages, paging, and UE context releases. After associating the gNB, the AMF keeps sending `HEARTBEAT` messages to confirm that the gNB is still online.

[5] `https://github.com/aligungr/UERANSIM/wiki/Feature-Set`

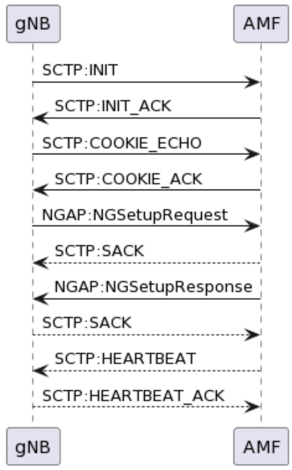


Figure 4.12: gNodeB association flow in AMF

4.4.2 UE attachment

For this test, the gNB previously associated with the CN was kept running, and a single UE was initiated also using UERANSIM. This UE is indicated to connect to the gNB via its configuration file and not through radio features.

The first message is sent by the gNB to the AMF with a registration request to begin the attachment process. After this, the AMF asks the AUSF to authenticate the SIM card trying to register in the network. If the AUSF gives green light to the AMF, the attachment procedure continues, and the AMF reports back to the gNB with a authentication request. The gNB in security mode responds with an authentication information via NGAP/NAS-5GS protocols, and, with this, the AMF can confirm that the UE is registered in the UDM/UDR. If such a verification ends up being true, then the policies according to the information given by the PCF. With the UE authenticated and its information confirmed, the AMF begins setting up the context for the UE by sending a request to the gNB. Upon context setup, the slice information registered for that UE is passed on to the NSSF, which chooses the ideal slice for the UE's service. This slice information is passed to the SMF. The SMF establishes a session with the UPF and the AMF contacts the gNB to create a PDU session. Upon attachment a GTP-U tunnel is created between the UE and the gNB, as mentioned in section 2.5.2. When the explained session creation is finished, the information from the instantiated session can be utilized by routing mechanism in the transport layer, an example of this is Segment Routing MPLS. This mechanism is being studied in another thesis of the researching group [59], making use of this Open5GS CN.

Figure 4.13: UE attachment to 5GC flow

4.4.3 UE re-attachment

When a UE tries reconnecting to the network, it sends a registration request message through the gNB. This UE has already been registered in the network and already had an established session, meaning that the previous session needs to be released first. To do this, the AMF sends a release command to the gNB, and when this release is complete, the AMF follows the same UE attachment procedure explained in the previous subsection. The re-attachment process is explained instead of detachment as the latter could not be performed with UERANSIM.

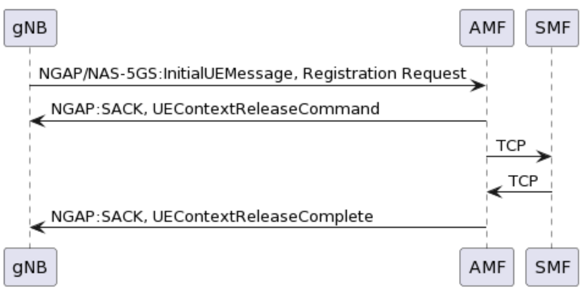


Figure 4.14: UE re-attachment flow in AMF

4.4.4 Unregistered UE

The flow shown in figure 4.15 represents an attempt of registration by a UE that is not registered in the UDR. This verification is initially made by the AUSF, as shown in the attachment and re-attachment procedures. The AUSF verifies that the UE is not registered and immediately reports that to the AMF, which rejects the registration and sends a release command to the gNB.

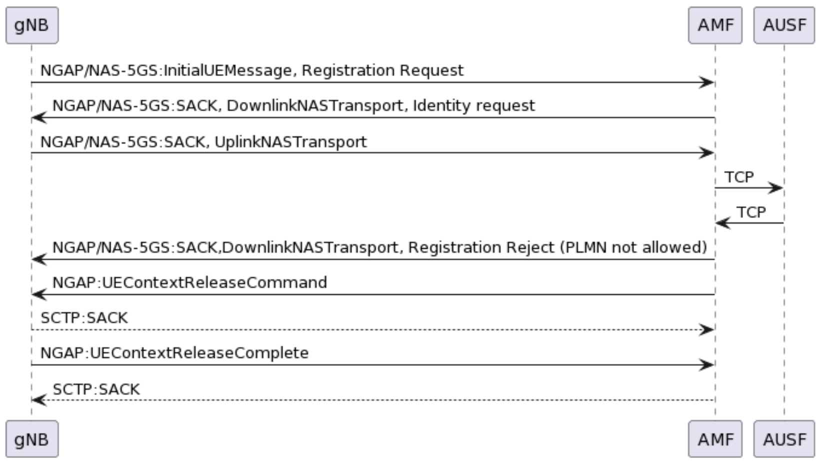


Figure 4.15: Unregistered UE attempt of connection flow

4.4.5 Conclusion

Similarly to the previous conclusion, for the NSA architecture in section 4.3.6, one function is present in all procedures, in this case it is the AMF. But unlike the MME, as the SA architecture possesses more functions, the responsibilities of the AMF are less, some of them being reduced to signaling tasks. It takes advantage of the SBA to be aware and communicate with all the other existent functions. Besides the AMF, only the AUSF, SMF and UPF are present in more than one of the tested procedures. This is due to the fact that the other functions are only used for session establishment, according to the information they are responsible for. It should be noted that these functions are present in procedures other than the ones presented in this dissertation, which are the basic ones. In the 5G architecture the session management is of the responsibility of the SMF, therefore it is the only function in which sessions are created, modified, or deleted, other than the UPF. The UPF manages sessions according to the information sent by the SMF, as their sessions are duplicates. In figure 4.3, there are two functions that are not represented in these flows, being the NRF and UDR. The reason for this is that the NRF only manages the discovery of the functions, not having any responsibilities in UE or gNB management, which was what these procedures tested. The UDR is not present as it is managed by the UDM, therefore it does not communicate with the other functions.

Procedure	AMF	AUSF	UDM	PCF	NSSF	SMF	UPF
gNodeB Association	✓						
UE attachment	✓	✓	✓	✓	✓	✓	✓
UE re-ttachment	✓					✓	
Unregistered UE	✓	✓					

Table 4.4: Review of the used CN components in each procedure

4.5 Cellular Testbed

With a deep knowledge of the CN and RAN at hand, a cellular testbed was tought to build with the available resources and network capabilities, without compromising the use cases of a city scenario defined in chapter 3. For this, a set of virtual machines in different locations was used, together with low resource devices (APUs[6]) and the previously presented real RAN, with LTE small cells. This section will present the deployed testbed, the designed CN, chosen selection manners, and an alteration made to the Open5GS source code to facilitate the desired scenarios.

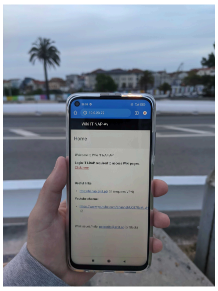
4.5.1 Testbed Overview

The implemented testbed is displayed in figure 4.16; in it, there is a deployment of an SDN along with a SDN controller; this will not be approached as it is not in the scope of this work. Marked with (1) are all the virtual machines, with (2) the used APU, and with (3) the containerized components, which are the UPFs as well as the core components. To serve as UEs, two smartphones with 5G capabilities were used.

IT1 and IT2 represent two different physical locations of Instituto de Telecomunicações; both have data centers in which the virtual machines were deployed. These are geographically close to each other (around 100 meters), and the figure 4.16 intends to demonstrate their separation as it influences ever so slightly the round trip times (RTTs) of the packets. The cloud service virtual machine is located outside the institute's network, hence why it is connected to the Data Network (DN). The switches present in the figure intend to display the proximity of the small cells to the host with the UPF, this will represent the edge in the scenarios approached in the next chapter.

[6] `https://www.pcengines.ch/apu2.htm`

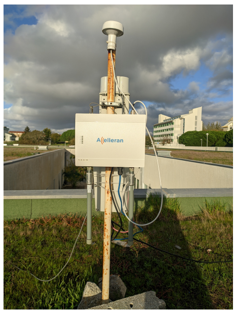
(a) SIM cards



(b) Smartphone 5G

Figure 4.17: UE 5G and Sysmocom SIM cards

(a) Small Cell 1



(b) Small Cell 3

Figure 4.18: Small Cells used for testbed

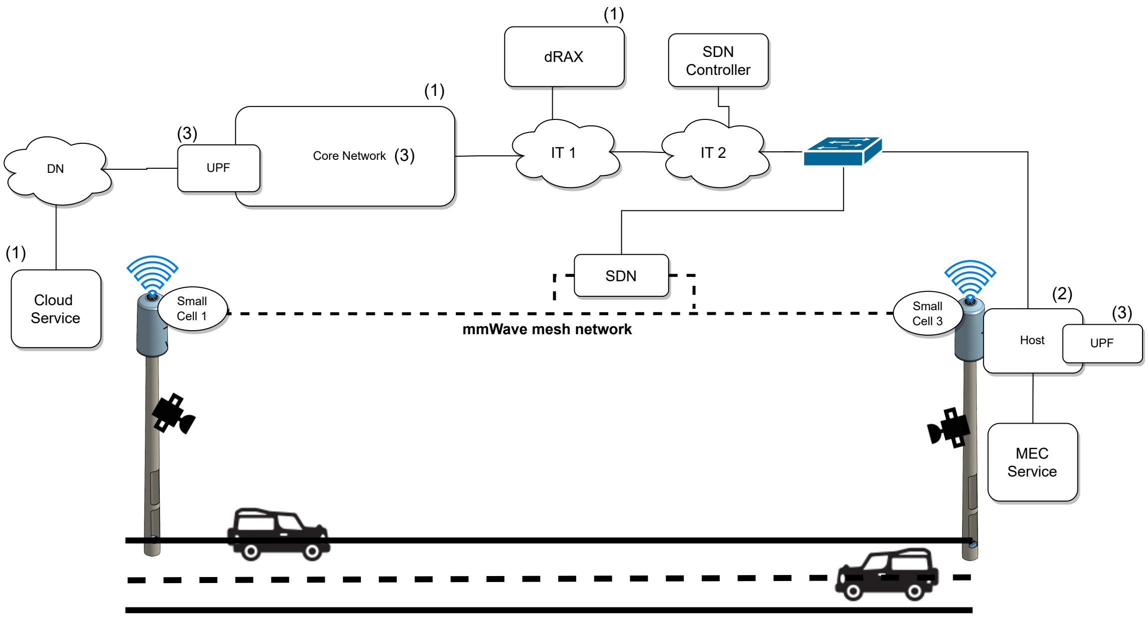


Figure 4.16: Cellular Testbed

The hardware specifications will be shown in the following tables, as well as specifications for the virtual machines, in table 4.6 and table 4.7 respectively, and the used APU in table 4.5.

Table 4.5: PC Engines APU characteristics

Processor	Memory	Storage	Operative Sys.	Linux kernel
AMD G series GX-412TC 1GHz quad Jaguar core with 64bits	4 GB DDR3-1333 DRAM	200 GB	Linux - Debian 11	5.15.1

Table 4.6: Virtual machines characteristics

Processor	Memory	Storage	Operative System	Linux kernel
Common KVM processor	4 GB	35 GB	Linux - Ubuntu 20.04.5 LTS	5.4.0-110-generic

Table 4.7: Cloud Service virtual machines characteristics

Processor	Memory	Storage	Operative System	Linux kernel
Common KVM processor	8 GB	65 GB	Linux - Ubuntu 20.04.4 LTS	5.4.0-125-generic

To serve as User Equipment (UE) two Xiaomi Redmi 9T 5G were used paired with Sysmocom[7] programmed SIM cards for a private network. The smartphone is pictured in figure 4.17b, and the SIM cards in figure 4.17a. The smartphones had the APN defined to `internet`, which was the APN that was configured in the CN so that they could see the private network.

The APU prepared to serve as a host for the UPF and MEC service displayed in figure 4.16, along with the switch to which the small cells are also connected.

Figure 4.19: APU used for Testbed

The two small cells that serve as Radio Unit (RU) for the deployment of this architecture are displayed in figure 4.18, where their connection to mmWave antennas is also depicted. These mmWave antennas are a part of a testbed from previous research projects, being a part of a mmWave mesh network [60], for flexibility in assembling the small cells in different locations.

4.5.2 Open5GS implementation

Section 4.2 showed the versatility of Open5GS, the capability of creating various paths inside the CN allied with the possibility to create replicas of the same components (e.g., as

[7] `https://sysmocom.de`

multiple UPFs) allows for various architectures. According to the needs mentioned in the previous chapter 3, the implemented components are the ones in figure 4.20.

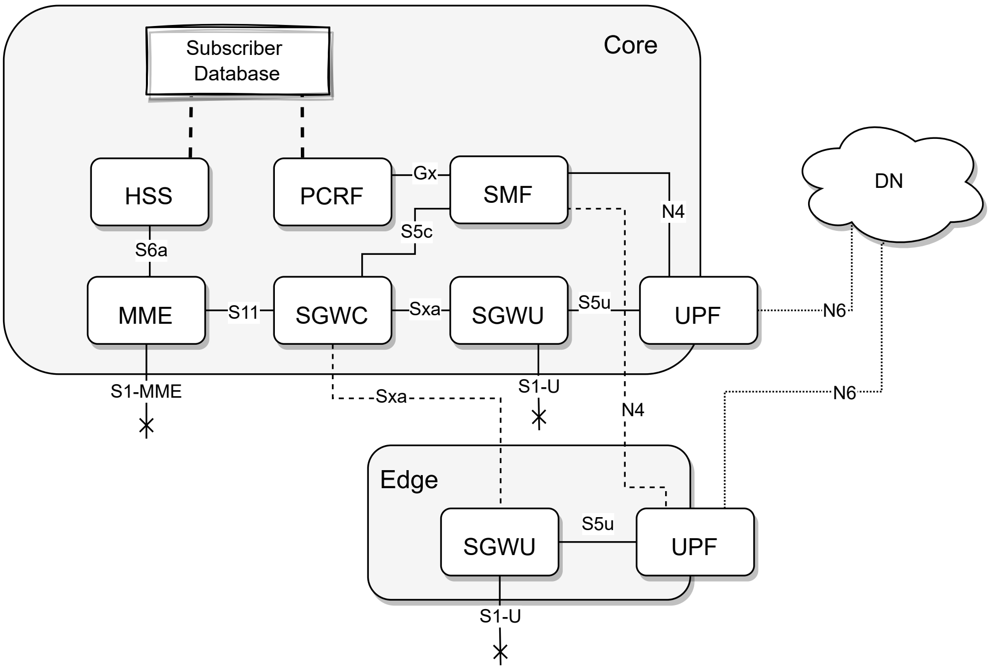


Figure 4.20: Implemented CN components

The figure 4.20 displays the implemented components and interfaces, which are the ones from the 5G NSA architecture due to the available RAN components. The Open5GS CUPS was implemented, disaggregating the second SGWU/UPF pair from the rest of the core. The connections to these modules are made with dashed lines to highlight the disaggregation; the connections represented with a straight line are in the same machine. The interface names can also be seen in the figure mentioned previously, with the MME and SGWU also showing the interfaces that connect with the RAN components; each component's interface serves a different protocol and uses a specific port to communicate with other components; the protocols and ports for each of the components are specified in table 4.8. Besides this, the advertised interfaces are also shown, these belong to components that connect to external equipment to the network.

In order to choose which SGWU/UPF pair should be used, the configuration files of the SGWC and SMF were changed. They were defined to select the core's pair when a UE connects to a small cell that represents TAC 1, and to select the edge's pair when the session's TAC is 17. The number selected to represent the TACs serve to represent not only geographical distance but also service differentiation, with TAC 1 being IoT services and TAC 17 being eMBB types of service and V2X, as the latter ones are desired to be edge services.

In order to instantiate the edge's components in a sustainable and low-resource manner, all the previously explained components are deployed in Docker containers, each of the components having its container with specific network rules. Since every component has its container, it is possible to deploy each component apart from the others; this is what was

Table 4.8: Open5GS 5G NSA Component Inter-Connection Specification

Component-Protocol	Port	Interface	Advertised
MME - S1AP	36412	S1-MME	✓
MME - GTPC	2123	S11	
MME - frDi	3868	S6a	
HSS - frDi	3868	S6a, Cx	
PCRF - frDi	3868	Gx	
SGWC - GTPC	2123	S11	
SGWC - PFCP	8805	Sxa	
SMF - GTPC	2123	S5c, N11	
SMF - GTPU	2152	N4u (Sxu)	
SMF - frDi	3868	Gx auth	
SMF - PFCP	8805	N4 (Sxb)	
SGWU-C - GTPU	2152	S1-U, S5u	✓
SGWU-C - PFCP	8805	Sxa	
SGWU-E - GTPU	2152	S1-U, S5u	✓
SGWU-E - PFCP	8805	Sxa	✓
UPF-C - GTPU	2152	S5u, N3, N4u (Sxu)	
UPF-C - PFCP	8805	N4 (Sxb)	
UPF-E - GTPU	2152	S5u, N3, N4u (Sxu)	
UPF-E - PFCP	8806	N4 (Sxb)	

done to the SGWU/UPF. To make the deployment possible, the UP components had their IPs advertised so that the containers of the CN components could communicate with them. To ease the deployment of the CN and edge CN, the `docker-compose` technology was used to aggregate the desired components, and to test various architectures quickly. This docker deployment was inspired by a project that can be found in GitHub[8].

4.5.3 UPF selection

In section 4.2.2 it was explained how Open5GS performs node selection, including how the SMF performs UPF selection. Whilst studying scenarios for the city environment and its different use cases, there was a conclusion that this approach did not give enough flexibility for node selection. Let us suppose the following scenario. There are two different services, V2X and video, and each service has an APN to identify it, as figure 4.21 depicts; two small cells each with a serving UPF associated with them. Both cells serve V2X UE as this service requires mobility, and the first one also serves a video service node placed closer to it; the first cell is associated with TAC 1 and the second with TAC 2. The UE with APN for the video service, when connecting to cell one, will have the correspondent UPF associated with it as the APNs match, when connecting to cell two and following Open5GS manner of selection. Since no matching APN was found, the UPF will be selected according to the next mode, TAC; since the UE is connected to the cell with TAC 2, it will have its corresponding UPF associated to it, the implemented solution works thus far. But when a V2X UE connects to a cell, since both of them have APN V2X associated to them, no matter which cell the user

[8] `https://github.com/herlesupreeth/docker_open5gs`

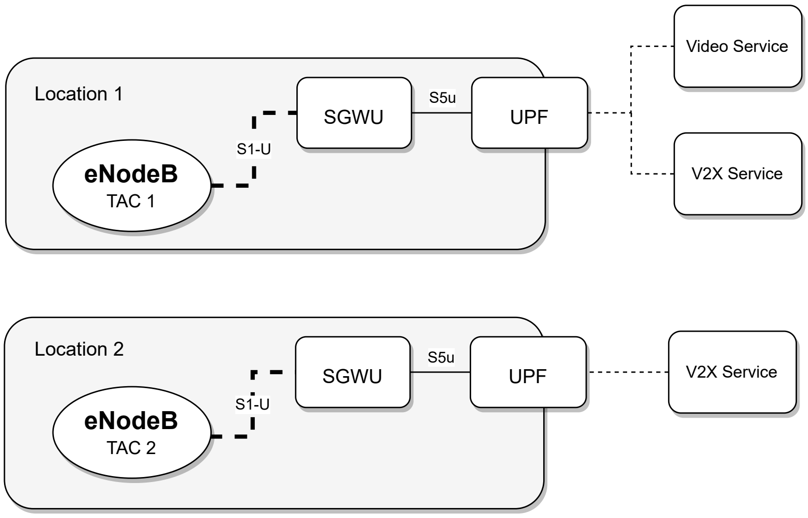


Figure 4.21: Supposed scenario for UPF selection

connects to, Open5GS will always do a round robin selection throughout the APN V2X node list, disregarding the geographical area (defined by TAC) of the cell and corresponding UPF; this means that, when a V2X user connects, its associated UPF will be almost random.

The presented scenario follows the services referred to in chapter 3 in a concrete way. As shown with the previous explanation, the logic explained in section 4.2.2 could not fulfil the supposed scenario. With this in mind, there was the need to add the selection mode in such logic that when a UE connects, all the session parameters must match the parameters defined for a particular node. This way, the V2X service, besides having its APN matched, would also have to have the TAC matched to guarantee it always gets associated with the nearest UPF. For this, the algorithm 1 was implemented in Open5GS's node selectors source code to fulfill the thought about scenarios.

In a real city environment, there will be dozens of different services with different processing nodes at different city points. Making it so that the option of simply removing the APN to solve this scenario is not a real option at all, as without such combination it would not be possible to replicate the city scenario. Therefore, the algorithm change was made taking the solution's scalability into account.

Algorithm 1: PFCP node selection with AND logic

```
Result: A PFCP node is chosen
int i = 0 ;
bool found = false ;
if node has one or more DNN associated then
    /* Check if the UE session DNN is equal to one of the node DNN(s) */
    for list of associated DNNs do
        if node DNN == session DNN then
            found = true ; break;
    /* If no DNN found then this node can be skipped */
    if found is not true then
        return false;
found = false ;
if node has one or more TAC associated then
    /* Check if the UE session TAC is equal to one of the node TAC(s) */
    for list of associated TACs do
        if node TAC == session TAC then
            found = true ;
            break;
    /* If no TAC found then this node can be skipped */
    if found is not true then
        return false;
/* The same logic applied for DNN and TAC is then applied to eCell ID
   and NR Cell ID */
/* If the algorithm reached the end means that all the nodes mode of
   selection are the same as the UE session therefore the node is
   selected */
return true ;
```

4.6 Summary

This chapter presented the tools to deploy the end-to-end mobile network architecture to be integrated in the ATCLL infrastructure, from the Access Network (AN) to the CN, explaining these tools' capabilities and how they can be used to fulfill the needs of the city's services and concepts. In the AN section, it was given an insight into the used hardware and the functionalities of the RAN Intelligent Controller (RIC) software to manage small cells, and then the CN used was introduced.

Open5GS was introduced as the main component of this dissertation, hence the detailed explanation of the architecture and node selection features. Following this, basic procedures are presented inside the Core Network (CN), both in the NSA and SA architectures, by showing flow diagrams of the interactions between the Radio Access Network (RAN) and all the active core components. Then, the RAN and CN tools were shown together in a real deployment of a cellular testbed; the designed implementation was explained with the chosen components. Lastly, an algorithm applied in some components was shown together with the purpose for its existence as an extension to the default behavior of Open5GS.

In the next chapter, this testbed will serve as the platform for understanding how this core solution performs day-to-day activities, and for more specific scenarios related to improving services QoS.

CHAPTER 5

Experimentation and Results

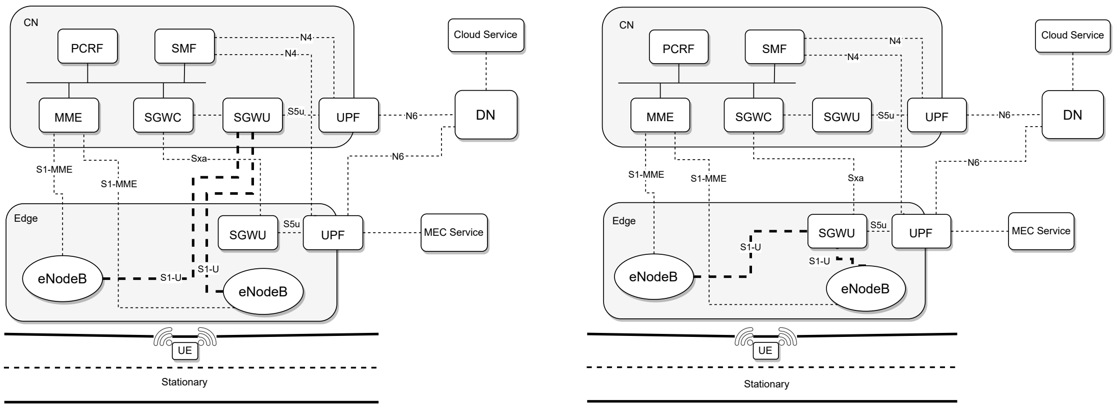
In order to validate the implemented testbed, as explained in the previous chapter, a set of tests was carried out following defined scenarios. To do so, several tools were used to attempt to recreate a real city environment. To understand the scenarios and results, the basic core procedures of how the implemented Core Network (CN) works should be known. The tests were executed in scenarios defined according to the city services highlighted in previous chapters. These tests will be described in section 5.1. Section 5.2 displays the plots obtained from the performed tests measurements, allowing us to conclude over the results. During testing, some limitations were detected in the CN and RAN, as section 5.3 intends to display these limitations and the difficulties they brought. To finalize, section 5.4 gives a summary of this chapter.

5.1 Scenarios

A city has different services with different needs, and for different needs different solutions are required. The Radio Access Network (RAN) and Core Network (CN) architecture was designed to try and respond to these different needs and to try and balance the load of UE that can be connected to it. To validate the chosen architecture, various scenarios were thought to perform tests. This section will present this architecture and will be divided into two scenarios for different services: stationary, such as IoT services; and mobility, such as eMBB mobility type services and V2X.

5.1.1 Stationary

The objective of a stationary scenario is to compare an architecture with the User Plane (UP) jointly with the Control Plane (CP), to an architecture that has the UP in a Control and User Plane Separation (CUPS) approach, by disaggregating the UP from the CP. Both these designs have the Radio Access Network (RAN) associated to the same zone; this can be done by associating the small cell to a Tracking Area Code (TAC) which is then defined in the CN, more specifically the SGWC and SMF, to represent a determined UP. To follow these



(a) UP next to core

(b) UP next to edge

Figure 5.1: Stationary scenarios

architectures, two types of services were projected: Multi-access Edge Computing (MEC) (mentioned in section 2.5.1) service and cloud service. The MEC service was placed in the same host that the CUPS scenario UPF representing edge, whilst the cloud service was placed in a host outside the researching institute's network. These scenarios intend slight to no mobility by the user or device as their main intention is to compare both solutions with equal ground. This can be achieved with the same constant cellular connection. As stated previously, RAN will represent two different zones which will be the main differentiator between the two stationary scenarios.

Figure 5.1a shows both small cells (eNodeB) connected to the CN's SGWU, that had TAC 1 defined for it, causing all the connected UEs to have a session created in the CN SGWU and UPF, making of the latter one the session's associated packet gateway. For this scenario, it is intended that the packets that go through this UPF are destined to the DN, and then to the defined cloud service. Thus simulating an IoT service communicating with a cloud database.

Figure 5.1b shows both small cells connected to the edge's SGWU, that had TAC 17 defined for it, with sessions created in this UP. For this scenario, the service was placed in the same host that the UPF. As mentioned previously, this allows the packets targeted to it not to need to go to the DN, which ultimately intends to reduce the round trip time (RTT), thus improving the connection. With this architecture, it can be shown that a UE that is using a specific service does not need to have longer packet travel times, as long as it has a city architecture prepared to deliver services closer to it.

5.1.2 Mobility

The main objective of the mobility scenario is to represent an actual device moving from one small cell to the other and, with it, changing the UP, thus maintaining the best designated UPF. To do this, the two available small cells represent different zones, each with different TACs (1 and 17) and different UPs, respectively. The idea behind this scenario, shown in figure 5.2 is to test if the CN can reallocate the UP in real time according to the UE's serving cell. For this, a handover procedure is expected to occur. The flow for this procedure was explained in section 4.3.5. Since it has few interactions with the MME, it can be concluded

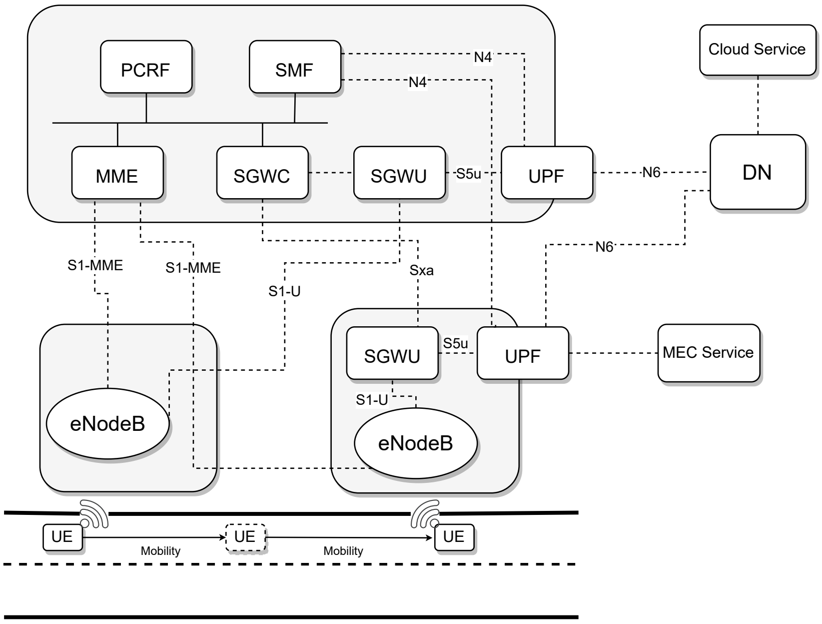


Figure 5.2: Mobility scenarios with UPs in different locations

that it is an `X2` handover, meaning that it is the RAN that manages the handover of the UE with the CN being required to update the current session.

5.2 Analysis of the results

With the metrics collected, this section will present an analysis of the implemented network, mainly focusing on the features that the Open5GS CN brings.

First, network performance tests will be comparing a UP joint with the CP and a CUPS scenario. The second collection of tests will analyze the behavior of the RAN and CN when two UEs are trying to overflow the bandwidth, putting to test the UE priority features provided by Open5GS. Lastly, the mobility scenario will see how the RAN manages the passage of a UE and how this affects the sessions created in the CN.

The `iPerf`[1] tool was used to perform these tests, running an instance of it on the server side and taking advantage of its Python wrapper[2] to run the client. Furthermore, the `pingparsing`[3] tool was also used to test round trip times (RTTs). All the tests were made on smartphones to emulate real scenarios. Termux[4] was used to run Python scripts on the devices. Termux-integrated API[5] was used to measure the strength of the serving cell's signal and other UE metrics. To reach the cloud service, tailscale[6] service of wireguard-like VPN was used. The hops required to reach the cloud service are shown in figure 5.3.

[1] `https://iperf.fr`
[2] `https://pypi.org/project/iperf3/`
[3] `https://pypi.org/project/pingparsing/`
[4] `https://termux.dev/en/`
[5] `https://wiki.termux.com/wiki/Termux:API`
[6] `https://tailscale.com`

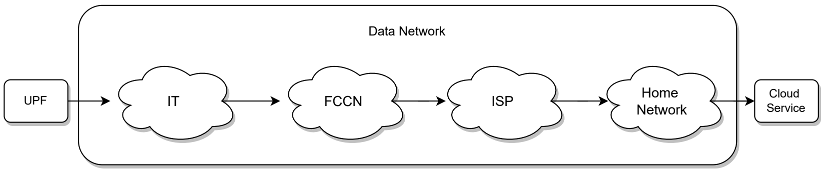
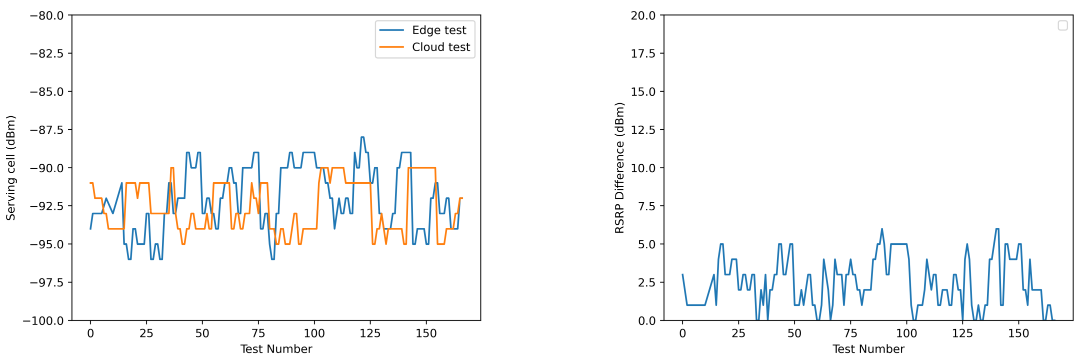


Figure 5.3: Hops description to reach the cloud service

The `iPerf` tool works under a client-server architecture. To get it running, a server must be instantiated, and then a client connects to this server. The measurements presented in this dissertation using the `iPerf` tool are all retrieved through the client's perspective. Even though that both download and upload were tested, the data is always retrieved from where the `iPerf` client was instantiated, which was the smartphone.

5.2.1 Network performance

The tests were performed by setting perfect conditions as best as possible, with a line-of-sight to the small cell, and on the weekend. This is important because the network is inside the university campus, meaning that in order to reach the Internet, it needs to go through all the routing processes made by the Fundação para a Computação Científica Nacional (FCCN) infrastructure, through an ISP and only then does it reach the home network where the cloud service is. Not only that, but during the week, it is highly used by students and researchers all over the country, which can cause unexpected congestion and failures. The small cell RSRP was measured and can be seen in figure 5.4a to attest that different connections would not influence the comparison of both solutions. Table 5.1 shows the average power difference in both tests, of 2.53dbm, which defines the level ground between both tests. If for reasons beyond control it would be detected that the connection was different, the test would be repeated. All the tests shown in this section will be comparing the CP jointly with UP and CUPS scenario, with the first being identified as *cloud* and the latter as *edge*.



(a) Serving cell RSRP

(b) Cell RSRP difference between both tests

Figure 5.4: Serving cell RSRP during testing

Table 5.1: Serving Cell RSRP descriptive statistics in dBm

Test	Mean	Standard Deviation	Min	Quantile 75%	Max
Edge	-91.98	2.13	-88	-90	-96
Cloud	-92.54	1.70	-90	-91	-95
Difference	2.53	1.64	0	4	6

Figure 5.5 and table 5.2 present RTT statistics, the results are from 20 tests and each of the tests transmits and awaits to receive a total of 10 packets. Figure 5.5 displays the average RTT of the 10 packets throughout the 20 tests. It can be seen that when the UP is deployed close to the small cell, the RTT is almost half of the one with the UP joint with the CN, both of them not showing a big variance in values.

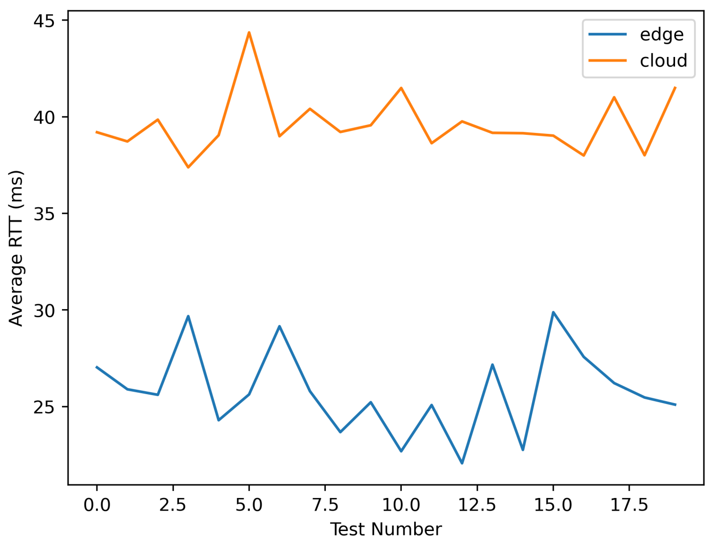


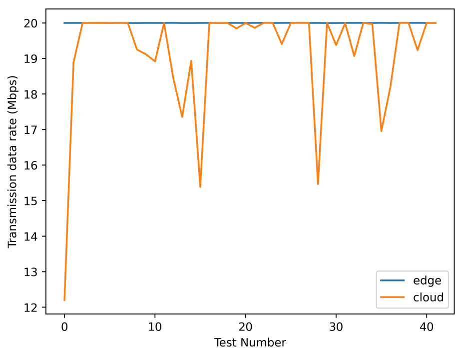
Figure 5.5: Average Round Trip Time

Table 5.2 presents the minimum and maximum values and the standard deviation, the same behavior seen in the previous figure is attested in these metrics. With the edge's metrics having better results than the core's, a similarity in the standard deviation shows that even though it has a higher minimum and maximum RTTs, it still maintains stability. This stability can also come from not having any other UE using the bandwidth of the cell and the constant line of sight in both tests.

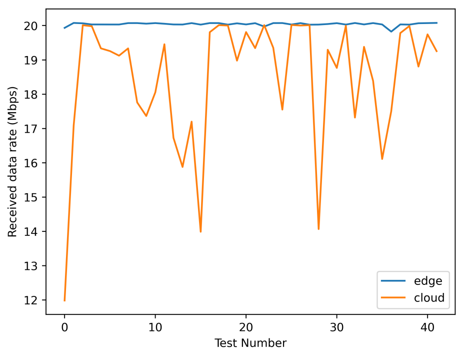
Table 5.2: Round Trip Time Statistics in milliseconds

Test	min	max	mdev
Edge	18.34	35.42	5.36
Cloud	34.84	60.27	7.35

Figures 5.6, 5.7, and 5.8 represent the usage of the `iPerf` script with TCP protocol comparing data upload and download, with transmitted and received data rates (in Mbps) and number of retransmissions as metrics. Each test has an average duration of five seconds, a block size of 1200, and a target data rate of 20 Mbps. In both transmitted and received data rates, the edge scenario fulfilled the target data rate of 20 Mbps in data download (figure 5.6). The same can not be concluded for the cloud tests, that only managed to reach these data rates in a few instances. One reason being is the increased number of hops a packet must go through to reach its destination. The results from the upload come from the same test hence why both figures represent similar results in peaks and descents at the same test numbers; the same goes for the download results.



(a) Transmission data rate (in Mbps)



(b) Received data rate (in Mbps)

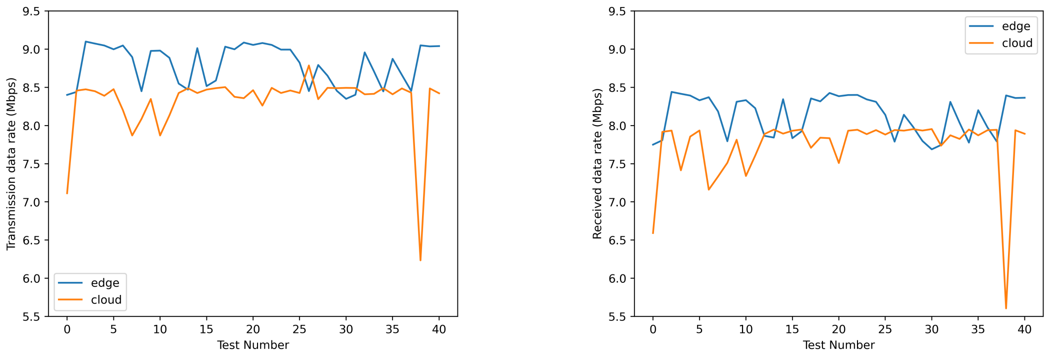
Figure 5.6: TCP Download

Consulting table 5.3, it can be verified that the main differences from the download tests come from the standard deviation and minimum values, in which the edge deployment proved to be very consistent showing almost no value variation, with constant values of 20Mbps; while the cloud deployment showed peaks of low values. If a service that required a constant throughput such as video streaming was being used, the cloud deployment could lead to frame losses in certain instances and drop in video quality.

Table 5.3: TCP Download descriptive statistics in megabits per second

Test	Mean	Standard Deviation	Min	Quantile 75%	Max
Edge TX	20	0	20	20	20
Cloud TX	19.19	1.60	12.20	20	20
Edge RX	20	0.05	19.82	20	20
Cloud RX	18.47	1.87	11.98	19.80	20

When it comes to data upload, the UE never managed to reach the target of 20 Mbps. A reason for this is that the used small cells provide LTE connection, which by itself already limits the upload maximum compared to download. Nevertheless, the edge solution presented constantly better results than the cloud, following the results obtained in the RTT test.



(a) Transmission data rate (in Mbps)

(b) Received data rate (in Mbps)

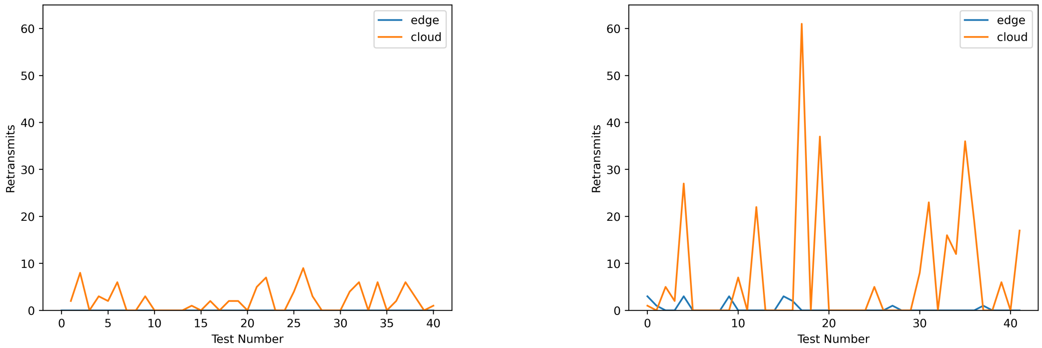
Figure 5.7: TCP Upload

Edge showed slightly better stability, as well as minimum and maximum values in the download test than in the cloud scenario which can be viewed in table 5.4. The deviation values are lower than the download case, which can be explained by the overall lower bit rate of the upload case, leading to lesser inconsistencies.

Table 5.4: TCP Upload descriptive statistics in megabits per second

Test	Mean	Standard Deviation	Min	Quantile 75%	Max
Edge TX	8.80	0.26	8.35	9.04	9.10
Cloud TX	8.31	0.42	6.23	8.49	8.79
Edge RX	8.13	0.26	7.69	8.36	8.44
Cloud RX	7.73	0.44	5.60	7.94	7.95

TCP retransmissions happen when a segment reaches an IP, and there is no acknowledgment for the data before TCP's automatic timer expires, or when duplicate acknowledgements are received which happens in TCP fast retransmission. Even though it usually does not raise many problems and does not cause data loss, it is still used as an indicator of the quality of the connection. The number of retransmissions of the download and upload tests can be seen in figure 5.8. The number of retransmissions follows the pattern of the other two TCP metrics, with the edge deployment having better performance. Further analyzing the edge scenario, in figure 5.8a it can be seen that there were no retransmissions at all, opposed to the download test; as seen previously the upload tests had a lower bit rate than the download tests, this can be an explanation to why there were no retransmissions, as less data is being transmitted there is also less of a chance it failing to do so.



(a) Upload **(b)** Download

Figure 5.8: Number of retransmissions

Table 5.5 shows that the difference seen in the RTTs is not evident across all protocols. The table represents the jitter values for User Datagram Protocol (UDP) in the previous scenarios for data upload and download. Each test has an average duration of 12 seconds and a target data rate of 20 Mbps; the results are similar, with a slight advantage even for the core solution. Although these differences are seen in paper, these numbers are quite derisory, with 2 ms not representing significant differences. The UDP values can be explained by the fact that `iPerf` calculates its metrics on the client side, and since UDP has no congestion control, the client creates a constant bit rate stream.

Table 5.5: UDP Jitter in milliseconds

	Mean	Standard Deviation	Min	Quantile 75%	Max
Upload Edge	1.24	0.15	1.06	1.33	1.62
Upload Core	1.25	0.15	1.02	1.33	1.51
Download Edge	0.80	0.22	0.44	0.86	1.97
Download Core	0.61	0.12	0.45	0.65	1.09

The values presented above are for a block size of 1200 bytes, which the network can handle with almost no jitter. This happens because the jitter is directly correlated with the block size. To prove this correlation, the edge scenario with data upload was repeated with different and increasing block sizes to see how much the jitter would increase, and the results can be seen in table 5.6.

Table 5.6: Jitter statistics with increasing block sizes in milliseconds

Block Size	Mean	Standard Deviation	Min	Quantile 75%	Max
1200	1.24	0.15	1.06	1.33	1.62
1500	1.51	0.18	1.40	1.57	1.72
2000	2.33	0.26	2.14	2.41	2.45
3000	4.29	0.75	3.66	4.35	5.72
10000	15.92	1.40	14.50	16.48	18.25

5.2.2 Priorities

When talking about QoS in mobile networks, it is obligatory to refer to QoS Class Identifier and 5G QoS Identifier, which are a particular identifier that characterizes the quality of packet communication provided by LTE and 5G, respectively. This topic was already slightly approached in section 2.3.2. For more detail, it should be noted that these identifiers are separated into two categories:

- Non Guaranteed Bit Rate (Non-GBR): Dedicated bearer that does not guarantee bit rate, the default bearer is always non-Guaranteed Bit Rate (GBR); these bearers can suffer packet loss under congestion;
- Guaranteed Bit Rate (GBR): Dedicated bearer that guarantees bit rate and a maximum bit rate; these bearers do not suffer packet loss in congestion situations.

Some examples of services that require GBR are: V2X messages, mission critical video and push-to-talk voice on user plane, real-time gaming and electrical distribution. Where as, some examples of Non-GBR services are: low latency eMBB applications for augmented reality, IMS signaling, stream buffering and TCP-based chats. Besides being divided into the two types referred to previously, these identifiers have different priority levels representing packet delay budgets and packet error loss rates. Also, 5QIs define default maximum data bursts and default averaging windows; these values can be consulted in QCI/5QI tables. The main difference between QCIs and 5QIs is that the latter applies to a flow carried at some point in a bearer, while QCI applies to a bearer within which certain types of flow are expected.

In addition to the identifiers, a second QoS parameter contains priority level, pre emption capability, and pre emption vulnerability called Allocation and Retention Priority (ARP). The priority levels go from 1 to 15, where:

- Range of 1 to 8 should be assigned to resources for services that are allowed to receive prioritized treatment inside an operator domain;
- Range of 9 to 15 should be assigned to resources approved by the home network and thus usable when a UE is roaming.

Pre-emption capability determines if a service data flow is eligible to receive resources that were previously allocated to a service data flow with a lower ARP priority level. Pre-emption vulnerability sets the conditions under which a service data flow may lose its resources in favor of a higher ARP priority level service data flow.

Open5GS has a web user interface that allows setting configurations for a specific UE session upon initial registration or after registration updating the rules when a new session is created. Firstly a UE can be assigned to a SST in a range of one to four and then a SD at choice. For session configuration, Open5GS permits the definition of all the parameters explained previously as well as APN and types of internet protocol, such as IPv4, IPv6, or IPv4v6. A specific UE IP address can also be defined; if not, Open5GS will give addresses from a pool defined upon core creation and increment the address for each session created. Finally, it is possible to define the aggregated maximum bit rate for uplink and downlink for the session.

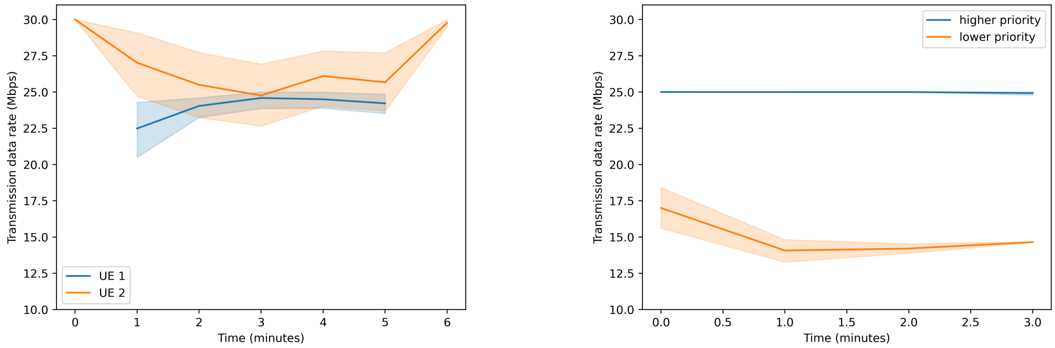
To attest that these configurations work and to confirm that the RAN acts accordingly with the CN, a test was performed with two UEs. Firstly the base scenario, with the same priorities, was tested to see the network behavior, and then different priorities were set to confirm the functionality of Open5GS's configurations. The UE configurations for each test can be seen in table 5.7.

Table 5.7: UEs configurations for priorities test

Same priorities test comparison								
UE	Target data rate (Mbps)	APN	QCI	ARP	Capability	Vulnerability	AMBR Uplink (Gbps)	AMBR Downlink (Gbps)
UE 1	30	`internet`	9	8	Disabled	Disabled	1	1
UE 2	30	`internet`	9	8	Disabled	Disabled	1	1
Different priorities test comparison								
UE	Target data rate (Mbps)	APN	QCI	ARP	Capability	Vulnerability	AMBR Uplink (Gbps)	AMBR Downlink (Gbps)
UE 1	30	`internet`	9	8	Disabled	Disabled	1	1
UE 2	30	`internet`	2	3	Disabled	Disabled	1	1

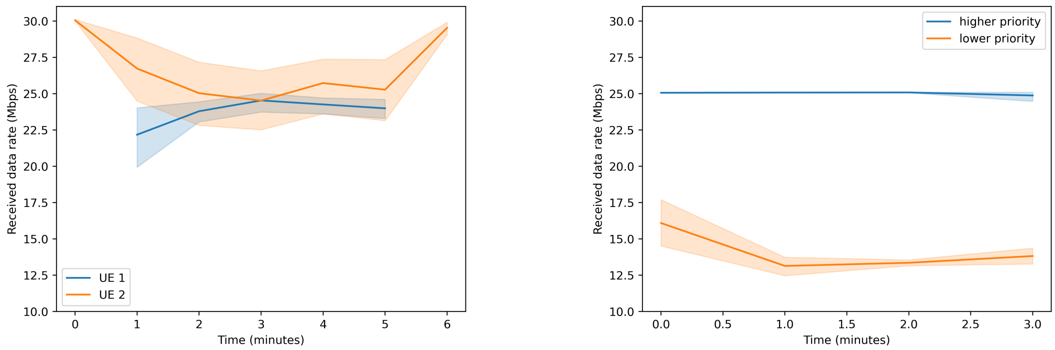
In section 5.2, the target data rate that `iPerf` enforced was 20 Mbps in order to not have packet loss due to network saturation. In this scenario, network saturation was an excellent indicator to see which UE would be given priority. After testing 20 and 25 Mbps, the network managed to deliver. This is why the value of 30 Mbps was chosen to perform these batch of tests. The TCP protocol was chosen as it presented better metrics and more apparent results than the UDP protocol in the previous tests; the test was made with data download.

Figures 5.9 and 5.10 show the sent and received bit rates in the tests with same and different priorities in the UEs. As these plots have multiple `y` values for each `x` an estimated mean was calculated by aggregating them for each `x` value, the shadow that can be seen is the confidence interval of the estimate. When both UEs have the same priorities (figures 5.9a and 5.10a) even if one user has better maximum value of sent and received bit rates, none of the connections is stable, with the network constantly trying to balance the bandwidth to both UE. When one user peaks, the other drops; the better bit rates keep interleaving between both UE in a load-balancing attempt. When applying a higher priority to a UE, the stability of its connection improves (figures 5.9b and 5.10b), with a constant bit rate of 25 Mbps, and the lower priority UE having its bit rate reduced, which also ends up balancing its connection. As mentioned previously, the target data rate was 30 Mbps which was not achieved (a single UE can get to this value, attested in the first seconds of the test in figure 5.9b). However, the network quickly adapts and stays constant in a value it knows it can reach, in this case, the 25 Mbps.



(a) same priorities **(b)** different priorities

Figure 5.9: Download TCP Transmission data rate (in Mbps)



(a) same priorities **(b)** different priorities

Figure 5.10: Download TCP Received data rate (in Mbps)

The two following tables (table 5.8 and 5.9) give a more thorough insight of the values presented in the previous figures. When analyzing the core with same priorities, UE 2 may appear to have better performance, but it should be noted that it also presented more irregular values, with almost double the standard deviation of UE 1. Even though the maximum values of both UEs are distant from one another, the minimum values show the network is clearly attempting to balance the load given to both users.

When analyzing the core with different priorities the conclusions are much easier, with the equipment with higher priority having very equal values throughout all statistics. This is important as in a city environment, there are critical services that require constant throughput values that should not be altered by other equipment's in the network. One good example is the one of video streaming in an emergency situation, in which an hospital can have a clear view of what is happening inside of an ambulance to try and assist the personnel in the vehicle. This vehicle would be an high priority user that would have its bitrate guaranteed.

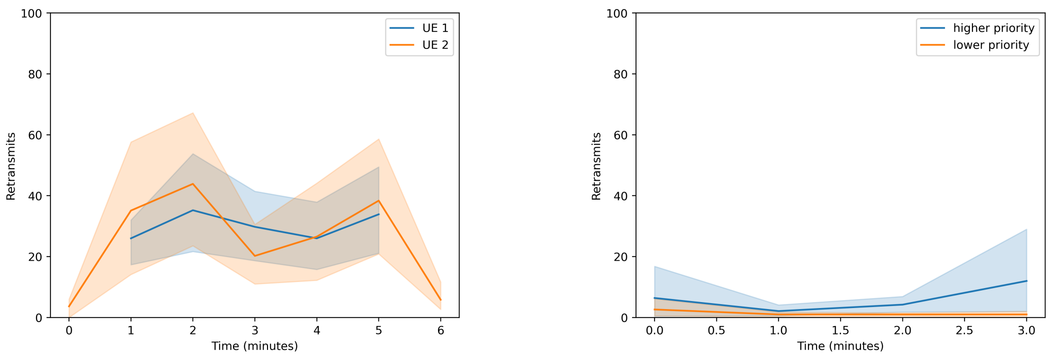
Table 5.8: Download TCP Transmission priorities comparison descriptive statistics in megabits per second

UE	Mean	Standard Deviation	Min	Quantile 75%	Max
1	24.10	1.51	18.51	25	25
2	26.50	3.32	19.10	30	30
higher priority	25	0.04	24.80	25	25
lower priority	15	1.90	11.39	15.35	20.50

Table 5.9: Download TCP Received priorities comparison descriptive statistics in megabits per second

UE	Mean	Standard Deviation	Min	Quantile 75%	Max
1	23.86	1.57	18.12	24.90	25.1
2	26.17	3.37	18.83	29.30	30.11
higher priority	25.10	0.12	25.10	25.10	25.11
lower priority	14.10	1.90	11.25	14.18	19.15

Figure 5.11 shows the number of TCP retransmissions in the two tests executed. The number of retransmissions when both UE have the same priorities follows the pattern of the other metrics with a constant attempt of load balancing by the network between the two UEs with the number of retransmissions jumping from almost 0 to more than 60 (5.11a). Figure 5.11b shows that upon applying more priority to one UE, the maximum number of retransmissions reduces. However, the UE that presents a higher number is the one with higher priority; this is because this user also has a much higher bit rate. Hence, it sends much more packets to the network, causing the retransmission numbers to rise accordingly.



(a) Same Priorities **(b)** Different Priorities

Figure 5.11: Download number of Retransmissions

With these tests, it is shown that it is possible to have a service prioritized to have better throughput showing that in a city environment, it can be assured that its specific requirements can be met.

5.2.3 Mobility

This scenario was tested with a smartphone serving as UE and moving it from one eNB to the other. Each eNB represents one distinct tracking area with its own TAC and a UP distinguished by these TACs. The paths taken are shown in figure 5.12. To better understand how this handover happened, RAN-wise, measurements were taken using Termux API endpoints[7]. The results from these tests can be seen in figures 5.13 and 5.14.

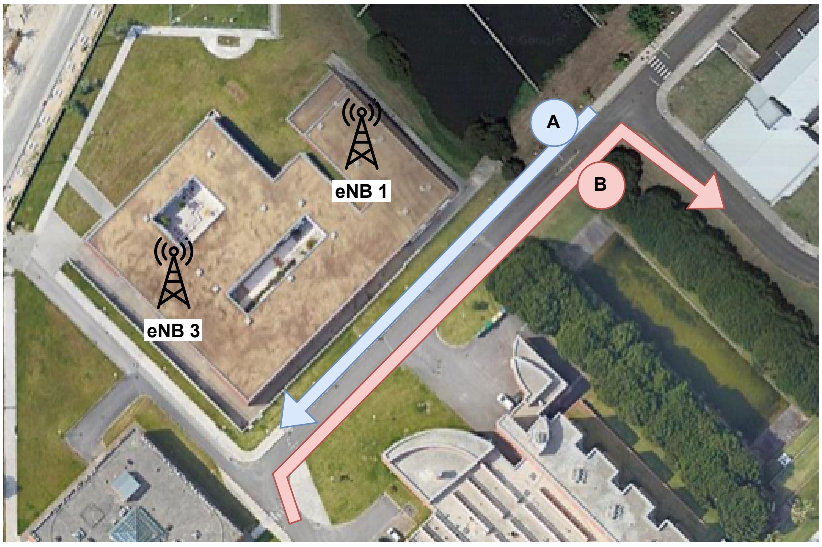


Figure 5.12: Handover paths

In figure 5.12 there are two paths, the blue path (marked with A) starting in eNB 1 and making the handover to eNB 3, and the red path (marked with B) starting in eNB 1 and making the handover to eNB 3. These two eNBs are also quite close, meaning that in some cases, the UE is in range of both. Nevertheless, as it can be quickly seen in both plots, the eNB 3 can reach much better values of RSRP than the eNB 1, even though both of them have the same output power value.

[7]https://wiki.termux.com/wiki/Termux-telephony-cellinfo

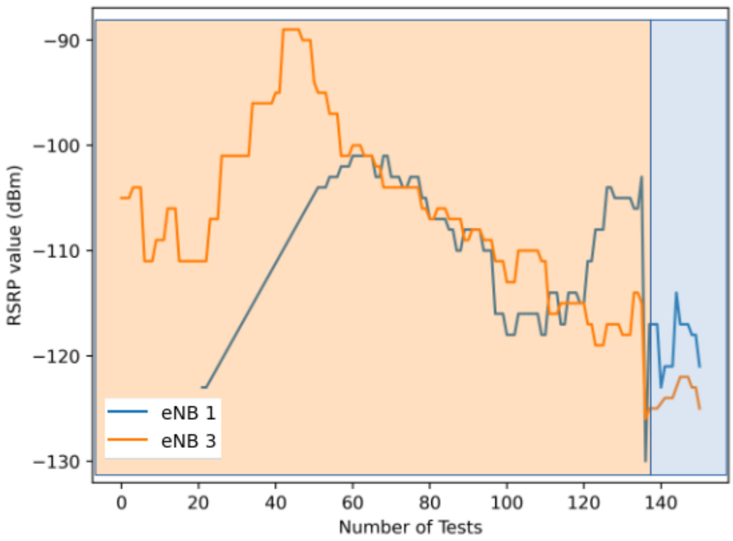


Figure 5.13: Cells RSRP during failed handover procedure

This difference in RSRP causes that, when the UE is moving in the red path, from 3 to 1, the eNB 1 never has a strong enough signal to trigger an handover. As the UE keeps moving further away from the serving cell, it simply loses coverage and only then connects to the eNB 1. In figure 5.13, with the painted areas representing the selected cell, it can be seen that the cell to which the UE is connected. The serving cell only changes when the signal is so poor that a loss of connection happens and upon reconnection it chooses the closest cell. Besides this, even when connected to eNB 1, the signal was already feeble, meaning that the UE continues to search for a different cell to connect to.

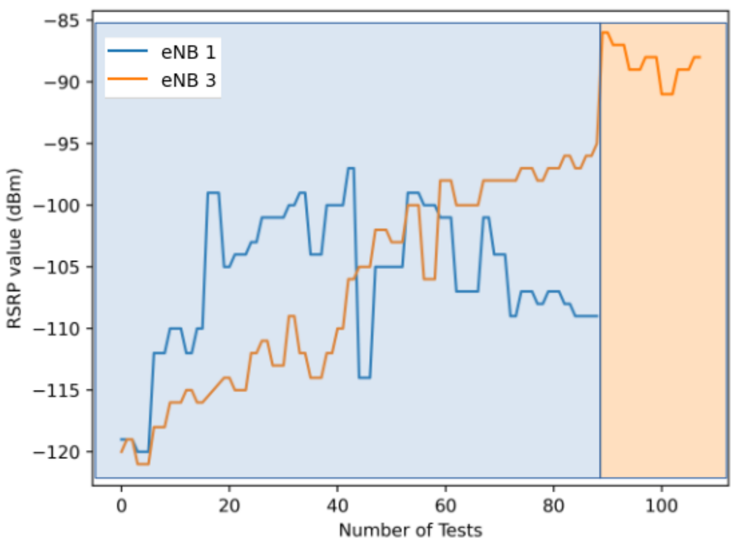


Figure 5.14: Cells RSRP during successful handover procedure

When performing a test in the blue path, from eNB 1 to 3, since eNB 1 has a much weaker signal, eNB 3 can trigger a handover as soon as it reaches the desired threshold set by dRAX. In figure 5.14, the decaying of the serving cell and improvement of the target cell is clear. It should be noted that the handover does not happen because the target-eNB reached a particular value, but because the UE calculated that it needed to search for another cell and found one that matched the re-selection criteria. After the handover, the UE's calculations reported that the serving cell's signals were good. Therefore, it stops searching for a new cell to try to connect to; that is why at the end of figure 5.14, the RSRP values from eNB 1 stop appearing.

5.2.4 IPFS integration

In order to validate the testbed in a stationary scenario with a real life scenario, a IPFS network was integrated with the deployed cellular network. This IPFS network was built by another researcher in a parallel thesis [61]. The testbed shown previously, in section 4.5, was slightly adjusted to accommodate the IPFS nodes, deployed in Nvidia Jetson development boards[8]; this testbed can be seen in figure 5.15.

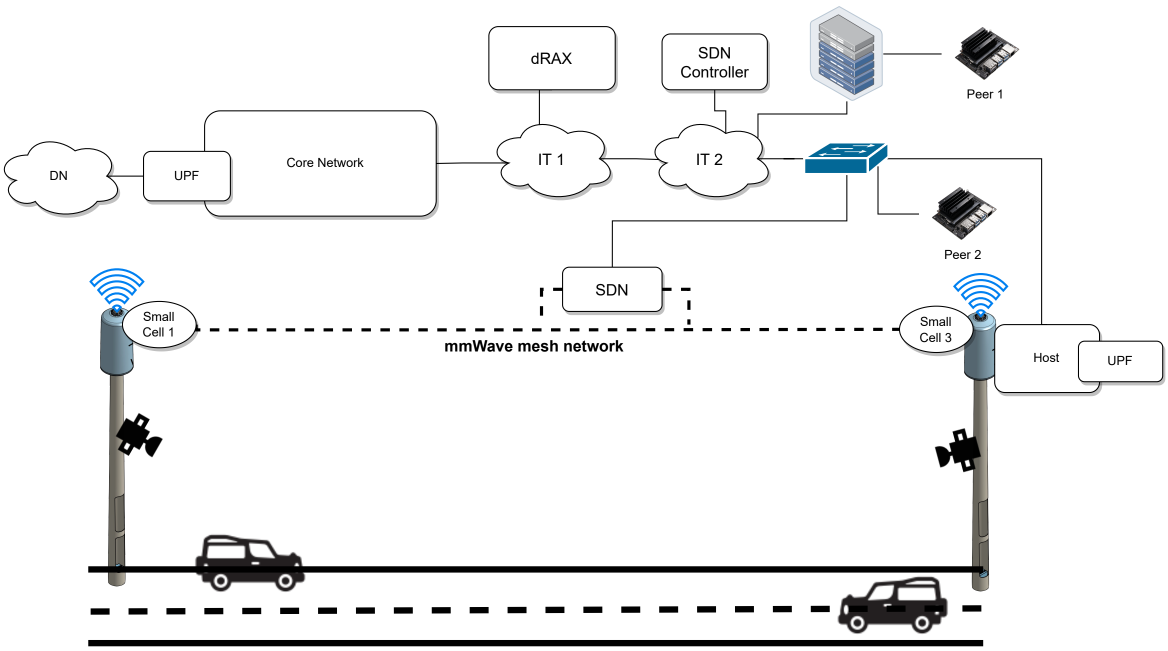
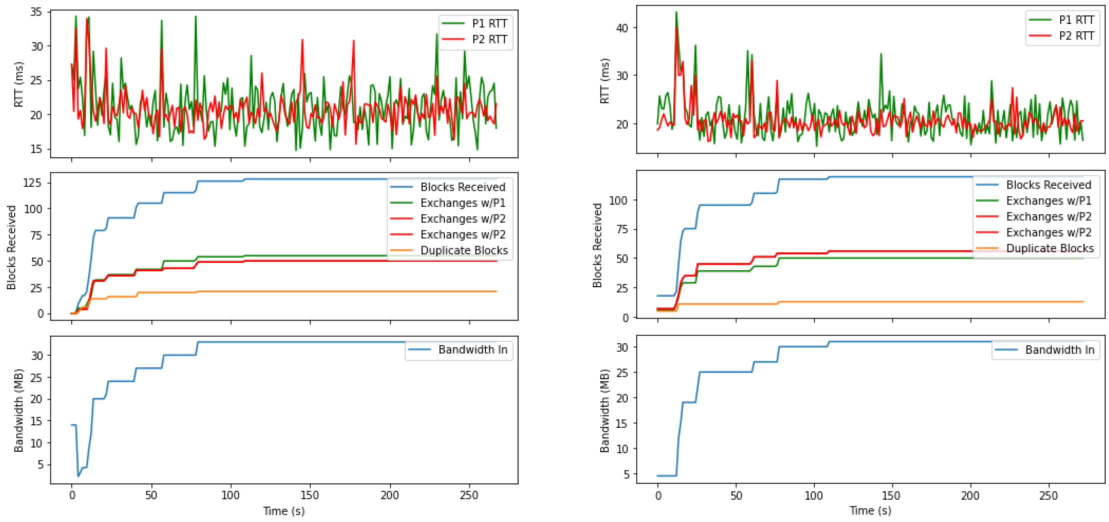


Figure 5.15: Cellular Testbed with IPFS integration

The idea behind this testbed is that the IPFS peers are separated, one closer to the core and the other to the small cell, representing edge; similarly to what was done with the cloud versus edge scenarios shown previously. This way, when a User Equipment is connected to a small cell serving an UPF for the CN peer, one would be chosen as it is the one with lesser latency; the opposite scenario is expected for peer two with the edge UPF. The results of these tests can be seen in figure 5.16.

[8]https://developer.nvidia.com/embedded/jetson-nano-developer-kit



(a) Core Scenario **(b)** Edge Scenario

Figure 5.16: IPFS Test

These tests show, even though not in a clear way, that in the core scenario (figure 5.16a) the peer closer to the CN sent more blocks to the user than the edge peer. This can be seen through the blocks received metric. In the same line of thought, in the edge scenario (figure 5.16b) peer 2 is the one with more exchanges, i.e. blocks received. These results show very little difference: this is due to the proximity of both peers (very few hops) even when comparing their distance, in round trip time (RTT), to both UPFs. This shows how the inclusion of edge services with the selection of the UPF is still unclear, and that very little support exists. Finally, by having these IPFS nodes in the edge, some bandwidth can be saved in the backhaul, as video content is usually heavy in this aspect.

5.3 Limitations and concerns

In subsection 4.3.5 it was mentioned that Open5GS did not have UP relocation in handover scenarios[9]. After some research, an issue on GitHub referenced a similar problem with S-GW relocation: even though the developer states that this feature was implemented, it was unclear if the solution worked or not. The tests indicate that there is a relocation of the GTP-U tunnel to the same SGWU, which is not the desired solution as this ends up not reallocating the UPF and therefore, the traffic maintains in the firstly connect UPF. References on how this relocation should happen are defined by ETSI in [62].

Currently, the UPF is running on an APU which has limited processing power when compared to a virtual machine; this may limit the performance of the UPF. To analyse the performance of the APU, a test was made using UERANSIM and the SA implementation of Open5GS, in it up to fifty UE were used to generate traffic through the `pingparsing` tool mentioned before. The following parameters were set to perform the test: a packet size of 50000 bytes, one second of timeout and 1000 `icmp` packets for each ping.

[9] `https://github.com/open5gs/open5gs/issues/1459`

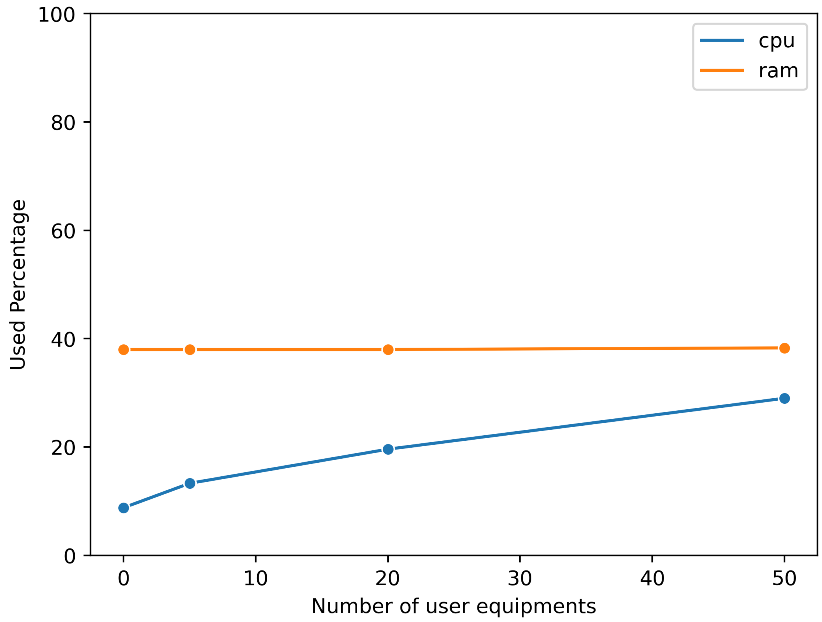


Figure 5.17: Containarized UPF RAM and CPU usage, in APU

The results of the CPU and RAM usage are shown in figure 5.17: they originate from tests with 0, 5, 20 and 50 UEs. This plot shows that, as expected, once the number of UEs starts increasing the CPU usage increases. The APU managed to handle fifty different users without reaching 50% of either CPU and RAM.

5.4 Summary

This chapter described the trial of the testbed, along with each component delivered previously. Two different scenarios were introduced in stationary and mobility, together with architectures of the RAN and CN on which the tests were performed. The stationary tests intend to compare two use cases, edge and cloud. A service was placed in the cloud, and the same service was on edge, which is as close as it can be with the CN gateway, the UPF.

Next, a set of tests was executed to analyze the priority configuration features of Open5GS, in which a service is given more priority in the network. In contrast, the mobility tests intended to attest to the capability of UP relocation in Open5GS. The test results show that services deployed closer to the UPF have better Quality of Service (QoS) than cloud services proving that there is a need to have various points of distribution throughout a city. The priority test results attested to the functionalities of Open5GS, showing that it is possible to have a service with higher priority with a constant bit rate while the network is clogged. The last set of tests, mobility-wise, shows that UP relocation in a handover scenario is not yet possible in Open5GS. Lastly, some limitations of the used infrastructure are laid-out.

CHAPTER 6

Conclusion and Future Work

This chapter presents the global conclusions obtained throughout this dissertation. After, a discussion is made on what else is still to be done in the topic along with what could be continued from this work.

6.1 Conclusion

In the current days, 5G is starting to be commercially advertised, and the idea that we are living in the fifth generation of mobile networks is evident. This generation brought to life several new concepts that were merely theoretical. But as cellular coverage of the fifth generation is being deployed in urban city centers, the rest of the area is still LTE dependent.

The goal of this dissertation was to approach 5G concepts such as Neutral Hosting, URLLC and advanced Multi-access Edge Computing scenarios. To fulfill this, a study of both LTE and 5G end-to-end architectures was done, together with the concept of Control and User Plane Separation allied with MEC technologies. The deployed network always had in mind city scenarios and it was in that sense that it was tested. Upon User Plane disaggregation from the Control Plane, edge and cloud services were deployed to further explain the difference and importance of having a city prepared to provide the services as close to the end user as possible.

Developing an architecture for a city presented its challenges. Although some of them are solvable, there are a few that were out-of-reach. When tackling neutral hosting, one should be aware that 5G's spectrum may not be available to non-operator entities, meaning that one operator has to be the neutral host, therefore arrangements have to be made between the city and all the country's operators. Not all core solutions provide ways to satisfy concepts like neutral hosting or CN disaggregation, choosing to adopt a core solution should be made according to previously defined use cases and goals. The CN is not the only one which should be approached in Neutral Hosting; even though this thesis focuses on the CN, it was also shown that having Open-RAN compliant radio capabilities provides another way to share a city infrastructure.

Moving forward, having a disaggregated CN in which a UP is used to enable MEC scenarios proved to be the right decision for various services, as this solution provides not only better throughput values but also more constant values with little to no variations. The importance that these results have, are always dependent on the service that will take advantage of them. The small variations in throughput values are important in services that require a minimum value of bandwidth that also does not fail, where as the reduced RTTs are important in services that require a ultra reliable low latency. Besides CN re-arrangements, this work has also shown how specific UEs can have different QoS giving them priority over the rest, which can then be transplanted to services.

Lastly, this dissertation showed how handover scenarios are executed with this RAN and CN tools and their limitations, concluding that, even though handovers may be seamless, having UP re-allocation is still not yet a reality. Integration of this cellular network with IPFS, proved the flexibility of mobile networks in various scenarios, and that acting on the base of an infrastructure can improve its overall performance.

In the end, a tested architecture was achieved, although with known and defined limitations, being a good foundation for mobile networks in the infrastructure of cities and other large and crowded areas, as is shown in the integration with the ATCLL project.

6.2 Future Work

Mobile networks are in constant evolution, with the sixth generation (6G) already beginning to be discussed. This dissertation followed fifth generation (5G) principles but worked on a Non Standalone (NSA) architecture, therefore the work done in this thesis can be improved with the following aspects:

- Evaluate the proposed architecture using several services used in the city (V2X, video, IoT), split them into different QoS levels and measure the impact;
- Test MOCN and MORAN scenarios in the proposed architectures with User Equipment (UE) from different networks and study how this impacts the overall network;
- Follow the evolution of the used CN to continuously integrate newly launched features in the implemented architecture; Study other solutions for CNs to attest if they deliver a better implementation of the desired scenarios;
- Extend the solution to a fully 5G Standalone (SA) implementation. This includes the Radio Access Network (RAN) and the Core Network (CN) (which had the basic principles already laid out in this document).

References

[1] U. Cisco, "Cisco annual internet report (2018–2023) white paper," *Cisco: San Jose, CA, USA*, 2020.

[2] R. Bajracharya, R. Shrestha, and H. Jung, "Future is unlicensed: Private 5g unlicensed network for connecting industries of future," *Sensors*, vol. 20, no. 10, p. 2774, 2020.

[3] R. Bajracharya, R. Shrestha, H. Jung, and H. Shin, "Neutral host technology: The future of mobile network operators," *IEEE Access*, vol. 10, pp. 99 221–99 234, 2022. DOI: 10.1109/ACCESS.2022.3207823.

[4] A. Ksentini and P. A. Frangoudis, "Toward slicing-enabled multi-access edge computing in 5g," *IEEE Network*, vol. 34, no. 2, pp. 99–105, 2020.

[5] F. Giust, G. Verin, K. Antevski, *et al.*, "Mec deployments in 4g and evolution towards 5g," 2018. [Online]. Available: www.etsi.org.

[6] O. S. Larry Peterson, *5g mobile networks: A systems approach*, [Online; accessed June 21, 2022]. [Online]. Available: https://github.com/SystemsApproach/5G.

[7] F. Erunkulu, M. Adamu, C. Lebekwe, M. Mosalaosi, and J. Chuma, "5g mobile communication applications: A survey and comparison of use cases," *IEEE Access*, vol. PP, pp. 1–1, Jun. 2021. DOI: 10.1109/ACCESS.2021.3093213.

[8] N. Bhandari, S. Devra, and K. Singh, "Evolution of cellular network: From 1g to 5g," *International journal of engineering and techniques*, vol. 3, no. 5, pp. 98–105, 2017.

[9] M. Keenan. "The evolution of cellular networks." (2020), [Online]. Available: https://www.avnet.com/wps/portal/abacus/resources/article/the-evolution-of-cellular-networks/ (visited on 05/2022).

[10] C. Telecommunications. "The evolution of mobile communications, from 1g to 5g." (2017), [Online]. Available: https://www.carritech.com/news/evolution-mobile-communication-1g-5g/ (visited on 05/2022).

[11] A. Lucent, "The lte network architecture—a comprehensive tutorial," Tech. Rep., 2009. [Online]. Available: https://rintintin.colorado.edu/~gifford/5830-AWL/LTE_Alcatel_White_Paper.pdf.

[12] A. Thongrak, N. A-Mapat, P. Moungnoul, and P. Anunvrapong, "Load sharing technique for coexistence lte-fdd and lte-tdd," in *2019 5th International Conference on Engineering, Applied Sciences and Technology (ICEAST)*, 2019, pp. 1–4. DOI: 10.1109/ICEAST.2019.8802549.

[13] H. R. GABRIEL BROWN PRINCIPAL ANALYST, "Service-based architecture for 5g core networks," Huawei Technologies Co. Ltd., Tech. Rep., Nov. 2017. [Online]. Available: https://www.3g4g.co.uk/5G/5Gtech_6004_2017_11_Service-Based-Architecture-for-5G-Core-Networks_HR_Huawei.pdf.

[14] A. A. Ateya, A. Muthanna, M. Makolkina, and A. Koucheryavy, "Study of 5g services standardization: Specifications and requirements," in *2018 10th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, 2018, pp. 1–6. DOI: 10.1109/ICUMT.2018.8631201.

[15] M. Settembre, "A 5g core network challenge: Combining flexibility and security," in *2021 AEIT International Annual Conference (AEIT)*, 2021, pp. 1–6. DOI: 10.23919/AEIT53387.2021.9627014.

[16] M. A. Habibi, M. Nasimi, B. Han, and H. D. Schotten, "A comprehensive survey of ran architectures toward 5g mobile communication system," *IEEE Access*, vol. 7, pp. 70 371–70 421, 2019. DOI: 10.1109/ACCESS.2019.2919657.

[17] NicoElNino, “The o-ran whitepaper 2021 overview, architecture, and traffic steering use case,” 2016. [Online]. Available: www.rimedolabs.com.

[18] “O-ran minimum viable plan and acceleration towards commercialization,” 2021. [Online]. Available: https://static1.squarespace.com/static/5ad774cce74940d7115044b0/t/60f9b144abdc902712f43475/1626976585796/O-RAN+Minimum+Viable+Plan+and+Acceleration+towards+Commercialization+White+Paper+29+June+2021.pdf.

[19] O. O. Erunkulu, A. M. Zungeru, C. K. Lebekwe, M. Mosalaosi, and J. M. Chuma, “5g mobile communication applications: A survey and comparison of use cases,” *IEEE Access*, vol. 9, pp. 97 251–97 295, 2021. DOI: 10.1109/ACCESS.2021.3093213.

[20] S. Rommer, P. Hedman, M. Olsson, L. Frid, S. Sultana, and C. Mulligan, “Chapter 3 - architecture overview,” in *5G Core Networks*, S. Rommer, P. Hedman, M. Olsson, L. Frid, S. Sultana, and C. Mulligan, Eds., Academic Press, 2020, pp. 15–72, ISBN: 978-0-08-103009-7. DOI: https://doi.org/10.1016/B978-0-08-103009-7.00003-X. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B978008103009700003X.

[21] TSGS, “Ts 123 501 - v16.6.0 - 5g; system architecture for the 5g system (5gs) (3gpp ts 23.501 version 16.6.0 release 16),” 2020. [Online]. Available: https://portal.etsi.org/TB/ETSIDeliverableStatus.aspx.

[22] B. Bertenyi, R. Burbidge, G. Masini, S. Sirotkin, and Y. Gao, “Ng radio access network (ng-ran),” *Journal of ICT Standardization*, vol. 6, pp. 59–76, 1 May 2018, ISSN: 2245-800X. DOI: 10.13052/JICTS2245-800X.614. [Online]. Available: https://riverpublishers.com/journal_read_html_article.php?j=JICTS/6/1/4.

[23] T. Taleb, K. Samdanis, B. Mada, H. Flinck, S. Dutta, and D. Sabella, “On multi-access edge computing: A survey of the emerging 5g network edge cloud architecture and orchestration,” *IEEE Communications Surveys & Tutorials*, vol. 19, no. 3, pp. 1657–1681, 2017. DOI: 10.1109/COMST.2017.2705720.

[24] S. Kekki, W. Featherstone, Y. Fang, *et al.*, “Etsi white paper no. 28 mec in 5g networks,” 2018. [Online]. Available: www.etsi.org.

[25] S. Rommer, P. Hedman, M. Olsson, L. Frid, S. Sultana, and C. Mulligan, “Chapter 6 - session management,” in *5G Core Networks*, S. Rommer, P. Hedman, M. Olsson, L. Frid, S. Sultana, and C. Mulligan, Eds., Academic Press, 2020, pp. 111–136, ISBN: 978-0-08-103009-7. DOI: https://doi.org/10.1016/B978-0-08-103009-7.00006-5. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780081030097000065.

[26] A. Abdulghaffar, A. Mahmoud, M. Abu-Amara, and T. Sheltami, “Modeling and evaluation of software defined networking based 5g core network architecture,” *IEEE Access*, vol. 9, pp. 10 179–10 198, 2021. DOI: 10.1109/ACCESS.2021.3049945.

[27] S. Lee, *Open5gs*, 2020. [Online]. Available: https://open5gs.org/.

[28] L. B. Dominato, H. C. C. de Resende, C. B. Both, J. M. Márquez-Barja, B. O. Silvestre, and K. V. Cardoso, “Tutorial on communication between access networks and the 5g core,” *CoRR*, vol. abs/2112.04257, 2021. arXiv: 2112.04257. [Online]. Available: https://arxiv.org/abs/2112.04257.

[29] L. Bonati, M. Polese, S. D’oro, S. Basagni, and T. Melodia, “Open, programmable, and virtualized 5g networks: State-of-the-art and the road ahead,” 2020. DOI: 10.1016/j.comnet.2020.107516. [Online]. Available: https://doi.org/10.1016/j.comnet.2020.107516.

[30] F. Kaltenberger, A. P. Silva, A. Gosain, L. Wang, and T.-T. Nguyen, “Openairinterface: Democratizing innovation in the 5g era,” *Computer Networks*, vol. 176, p. 107 284, 2020, ISSN: 1389-1286. DOI: https://doi.org/10.1016/j.comnet.2020.107284. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1389128619314410.

[31] I. Trindade, C. Nahum, C. Novaes, D. Cederholm, G. Patra, and A. Klautau, “C-ran virtualization with openairinterface,” *arXiv preprint arXiv:1908.07503*, 2019.

[32] S. Iqbal and J. M. Hamamreh, “A Comprehensive Tutorial on How to Practically Build and Deploy 5g Networks Using Open-Source Software and General-Purpose, Off-the-Shelf Hardware,” *RS Open Journal on Innovative Communication Technologies*, vol. 2, no. 6, Dec. 2021. [Online]. Available: https://rs-ojict.pubpub.org/pub/lqhigiva/release/2.

[33] I. P. Chochliouros, A. Kostopoulos, I. Giannoulakis, *et al.*, "Using small cells for enhancing 5g network facilities," in *2017 IEEE Conference on Network Function Virtualization and Software Defined Networks (NFV-SDN)*, 2017, pp. 264–269. DOI: 10.1109/NFV-SDN.2017.8169871.

[34] M. Paolino, G. Carrozzo, A. Betzler, *et al.*, "Compute and network virtualization at the edge for 5g smart cities neutral host infrastructures," in *2019 IEEE 2nd 5G World Forum (5GWF)*, 2019, pp. 560–565. DOI: 10.1109/5GWF.2019.8911726.

[35] G. Baldoni, P. Cruschelli, M. Paolino, *et al.*, "Edge computing enhancements in an nfv-based ecosystem for 5g neutral hosts," in *2018 IEEE Conference on Network Function Virtualization and Software Defined Networks (NFV-SDN)*, 2018, pp. 1–5. DOI: 10.1109/NFV-SDN.2018.8725644.

[36] TufailAli, NamounAbdallah, AlrehailiAhmed, and AliArshad, "A survey on 5g enabled multi-access edge computing for smart cities: Issues and future prospects," *International Journal of Computer Science and Network Security*, vol. 21, no. 6, pp. 107–118, Jun. 2021.

[37] Z. Sun, P. Du, A. Nakao, L. Zhong, and R. Onishi, "Building dynamic mapping with cups for next generation automotive edge computing," in *2019 IEEE 8th International Conference on Cloud Networking (CloudNet)*, 2019, pp. 1–6. DOI: 10.1109/CloudNet47604.2019.9064135.

[38] V. Jungnickel, K. Manolakis, W. Zirwas, *et al.*, "The role of small cells, coordinated multipoint, and massive mimo in 5g," *IEEE Communications Magazine*, vol. 52, no. 5, pp. 44–51, 2014. DOI: 10.1109/MCOM.2014.6815892.

[39] N. Hassan, K.-L. A. Yau, and C. Wu, "Edge computing in 5g: A review," *IEEE Access*, vol. 7, pp. 127 276–127 289, 2019. DOI: 10.1109/ACCESS.2019.2938534.

[40] ATIS, "Neutral host solutions for 5g multi-operator deployments in managed spaces innovation to lower the cost of capacity and coverage for 5g and existing radio technologies," 2019, [Online; accessed October 10, 2022]. [Online]. Available: https://access.atis.org/apps/group_public/download.php/48403/ATIS-I-0000073.pdf].

[41] M. N. Tahir and M. Katz, "Performance evaluation of ieee 802.11p, lte and 5g in connected vehicles for cooperative awareness," *Engineering Reports*, vol. 4, e12467, 4 Apr. 2022, ISSN: 2577-8196. DOI: 10.1002/ENG2.12467. [Online]. Available: https://onlinelibrary.wiley.com/doi/full/10.1002/eng2.12467%20https://onlinelibrary.wiley.com/doi/abs/10.1002/eng2.12467%20https://onlinelibrary.wiley.com/doi/10.1002/eng2.12467.

[42] M. N. Tahir, S. Fatima, and N. Bashir, "Car-to-car communication using its-g5 & 5g," in *2020 IEEE 5th International Conference on Intelligent Transportation Engineering (ICITE)*, 2020, pp. 331–335. DOI: 10.1109/ICITE50838.2020.9231349.

[43] TSGS, "Ts 123 287 - v16.3.0 - 5g; architecture enhancements for 5g system (5gs) to support vehicle-to-everything (v2x) services (3gpp ts 23.287 version 16.3.0 release 16)," 2020. [Online]. Available: https://portal.etsi.org/TB/ETSIDeliverableStatus.aspx.

[44] S. A. A. Hakeem, A. A. Hady, and H. W. Kim, "5g-v2x: Standardization, architecture, use cases, network-slicing, and edge-computing," *Wireless Networks*, vol. 26, pp. 6015–6041, 8 Nov. 2020, ISSN: 15728196. DOI: 10.1007/S11276-020-02419-8/FIGURES/13. [Online]. Available: https://link.springer.com/article/10.1007/s11276-020-02419-8.

[45] 5. A. Association, "Use case implementation des-cription (ucid) 5gaa automotive association technichal report safety treatment in connected and autonomous driving functions report 2," 2021. [Online]. Available: https://5gaa.org/wp-content/uploads/2021/04/use-case-t21001.pdf.

[46] S. Böcker, C. Arendt, P. Jörke, and C. Wietfeld, "Lpwan in the context of 5g: Capability of lorawan to contribute to mmtc," in *2019 IEEE 5th World Forum on Internet of Things (WF-IoT)*, 2019, pp. 737–742. DOI: 10.1109/WF-IoT.2019.8767333.

[47] I. Analytics, *Lpwan market report 2021–2026*, 2021.

[48] Y. Chen, Y. A. Sambo, O. Onireti, and M. A. Imran, "A survey on lpwan-5g integration: Main challenges and potential solutions," *IEEE Access*, vol. 10, pp. 32 132–32 149, 2022. DOI: 10.1109/ACCESS.2022.3160193.

[49] L. Lei, H. Xu, X. Xiong, K. Zheng, and W. Xiang, "Joint computation offloading and multiuser scheduling using approximate dynamic programming in nb-iot edge computing system," *IEEE Internet of Things Journal*, vol. 6, no. 3, pp. 5345–5362, 2019. DOI: 10.1109/JIOT.2019.2900550.

[50] S. M. Matinkhah and W. Shafik, "Smart grid empowered by 5g technology," in *2019 Smart Grid Conference (SGC)*, 2019, pp. 1–6. DOI: 10.1109/SGC49328.2019.9056590.

[51] H. C. Leligou, T. Zahariadis, L. Sarakis, E. Tsampasis, A. Voulkidis, and T. E. Velivassaki, "Smart grid: A demanding use case for 5g technologies," in *2018 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, 2018, pp. 215–220. DOI: 10.1109/PERCOMW.2018.8480296.

[52] R. E. Lopes, "Software-defined emergency network for a smart city," Ph.D. dissertation, Universidade de Aveiro, 2021.

[53] P. Rito, A. Almeida, A. Figueiredo, *et al.*, *Aveiro tech city living lab: A communication, sensing and computing platform for city environments*, 2022. DOI: 10.48550/ARXIV.2207.12200. [Online]. Available: https://arxiv.org/abs/2207.12200.

[54] Accelleran, *Drax: Cloud-native open ran software*, Accessed in 12/10/2022. [Online]. Available: https://accelleran.com/drax-2/.

[55] "Lte; evolved universal terrestrial radio access (e-utra); user equipment (ue) procedures in idle mode (3gpp ts 36.304 version 8.3.0 release 8)," 2008. [Online]. Available: http://portal.etsi.org/chaircor/ETSI_support.asp.

[56] R. Stewart, "Network working group," Sep. 2007, Accessed in 10/10/2022. [Online]. Available: https://www.rfc-editor.org/rfc/pdfrfc/rfc4960.txt.pdf.

[57] TSGC, "Ts 129 244 - v16.5.0 - lte; 5g; interface between the control plane and the user plane nodes (3gpp ts 29.244 version 16.5.0 release 16)," 2020. [Online]. Available: https://portal.etsi.org/TB/ETSIDeliverableStatus.aspx.

[58] M. Silva, P. Teixeira, C. Gomes, D. Dias, M. Luís, and S. Sargento, "Exploring software defined networks for seamless handovers in vehicular networks," *Vehicular Communications*, vol. 31, p. 100 372, 2021, ISSN: 2214-2096. DOI: https://doi.org/10.1016/j.vehcom.2021.100372. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2214209621000413.

[59] E. Dias, "Towards xurllc strategies with programmable pipelines and network coding," Engenharia de Computadores e Telemática, MSc Thesis, Universidade de Aveiro, 2022.

[60] T. Ferreira, A. Figueiredo, D. Raposo, M. Luís, P. Rito, and S. Sargento, "Improving mmwave backhaul reliability: A machine-learning based approach," *Ad Hoc Networks*, p. 103 050, 2022, ISSN: 1570-8705. DOI: https://doi.org/10.1016/j.adhoc.2022.103050. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1570870522002220.

[61] J. Oliveira, "Edge-based ipfs for content distribution," Engenharia Eletrónica e Telecomunicações, MSc Thesis, Universidade de Aveiro, 2022.

[62] TSGS, "Ts 123 401 - v15.4.0 - lte; general packet radio service (gprs) enhancements for evolved universal terrestrial radio access network (e-utran) access (3gpp ts 23.401 version 15.4.0 release 15)," 2018. [Online]. Available: https://portal.etsi.org/TB/ETSIDeliverableStatus.aspx.