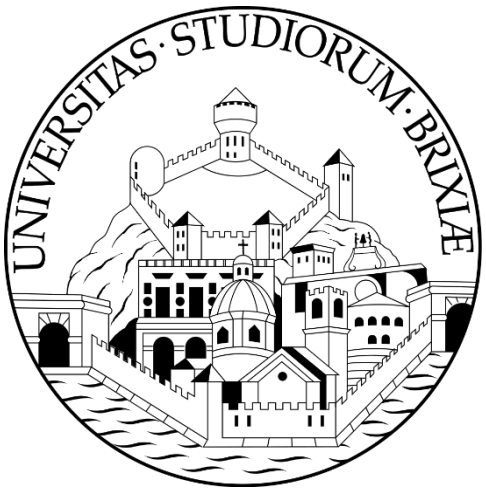
Università degli Studi di Brescia

Dipartimento di Ingegneria dell'Informazione

Corso di Laurea Magistrale in Tecnologie delle Comunicazioni e Multimedia



The Configuration, Setup, and Performance Evaluation of 5G Networks with Open-source Software

Supervisor: Prof. Francesco Gringoli

Student: M. Bashir Qanaa

Matricola: 737942

Academic Year: 2022/2023

Through the pages of this final project, I conclude an immersive journey I started 2 years ago when I first arrived in Italy to achieve my Master's degree. A journey that was full of ups and downs, laughter and tears, and most importantly a journey where I learned more about myself and shaped the person I'm today.

However, none of this would've been possible without the care that God has surrounded me with, opened many doors, guided me in all of my steps, being the light in the dark, the friend in the loneliest of days, and the strength in the hardest of days.

To my parents who were always there for me despite the distance, you're the source of support and inspiration and without you, I would've never reached this day nor the man I'm today.

To my Grandmother in her grave, I wish you were still alive to see me walking through the road of knowledge and education that today because of you is possible.

To my dear sisters, thank you for being there for me, offering me cheerfulness and advice along the way, thank you for all the precious moments we had together.

To the University of Brescia which facilitated my studies for the past two years providing all sources of comfort and support, I say thank you.

And finally, many thanks and greetings to all the professors I met during my studies and in particular Professor Maria Antonietta Vincenti managing my mobility period in Germany to eventually be an exceptional experience.
And most importantly, Professor Francesco Gringoli who was a significant source of support and guidance through not just the final project but also through any obstacles that a foreign student can face abroad, treated me like a son and a friend.

Abstract

Over the years the need for a solid communication system that provides both high data rates and mobility has grown dramatically in correspondence with the rising demand for data exchange and the increase in users number. Several generations have been released and each one was a step ahead of the previous one with new features and concepts.

4G and LTE (Long Term Evolution) main focus was a higher throughput compared to 3G. The key technologies that have made 4G possible are MIMO (Multiple Input Multiple Output) and OFDM (Orthogonal Frequency Division Multiplexing). Furthermore, This generation supported interactive multimedia, voice, and video. it's more scalable and supports Ad hoc and multi-hop networks. LTE was also capable of operating in two-dimensional resource scheduling (in time and frequency), allowing support to multiple users in a time slot; in contrast, 3G technology performs one-dimensional scheduling, which limits service to one user for each time slot.

The aim of this thesis work is to keep track of the evolution of these cellular networks and study which adjustments were made. We will take on a deep dive into the 5G network analyzing its architecture, building blocks, usage scenarios, and the technologies implemented, and eventually report some lab experiments conducted with the aim of building a fully functioning SA-5G network by using Off-The-Shelf Hardware and Open-source software running and processed by a powerful computer. The main goal of such deployment is to study and analyze the new authentication mechanisms, more specifically the SUCI concealment, and validate its benefit when compared to the legacy solution by LTE. Furthermore, we will do tests related to connection stability between our network deployed using Open-Source implementations, and report the achieved data rates with different devices and any compatibility issues.

Contents

List of Figures

List of Tables

List of Acronyms

1G	First Generation	**GUAMI**	Globally Unique AMF ID
2G	Second Generation	**GBR**	Guaranteed flow bit rate
HSPA+	Evolved High Speed Packet Access	**SBA**	Service-Based Architecture
4G	Fourth Generation	**REST**	Representational state transfer
LTE	Long-term Evolution	**GTP**	GPRS Tunnelling Protocol
SMS	Short Message Messaging	**FDM**	Frequency-division multiplexing
Mbps	Megabit per second	**FDD**	Frequency Division Duplexing
Kbps	Kilobit per second	**TDD**	Time Division Duplexing
GPS	Global Positioning System	**CP-OFDM**	Cyclic Prefix OFDM
MIMO	Multiple Input Multiple Output	**DFT-SC-OFDM**	Single Carrier OFDM
OFDM	Orthogonal Frequency-division multiplexing	**PAPR**	Low peak-to-average-power ratio
Ad Hoc	For a specific solution	**QPSK**	Quadrature Phase Shift Keying
5G	Fifth Generation	**SNR**	Signal to Noise Ratio
SA	Standalone	**SDAP**	Service Data Adaptation Protocol
IOT	Internet of Things	**PDCP**	Packet Data Convergence Protocol
GSM	Global System for Mobile	**RLC**	Radio Link Control

NSA	Non Standalone	**MAC**	Medium Access Control
SDN	Software Defined Radio	**PHY**	Physical Layer
NFV	Network Function Virtualization	**FAPI+**	Functional Application Platform Interface
5G-PPP	The 5G Infrastructure Public Private Partnership	**DU**	Decentralized unit
SC	Service Customer	**CU**	Centralized unit
SP	Service Provider	**RAN**	Radio Access Network
NOP	Network Operator	**RT-RIC**	Near real-time Radio Intelligent Controller
VISP	virtualized infrastructure providers	**NAS**	Non-access stratum
DCSP	Data Centre Service Provider	**TAC**	Tracking Area Code
API	Application Programming Interface	**NSSF**	Network Slicing Selection Function
NR	New Radio	**TAI**	Tracking Area Identifier
SGW	Serving Gateway	**CHF**	Charging Function
PGW	Packet Network Data Gateway	**SDF**	Service Data Flow
CUPS	Control plane and User plane Separation	**PFD**	Packet Flow Description
UE	User Equipment	**LI**	Lawful Intercept
GTPU	General Packet Radio Service Tunneling Protocol	**MEC**	Multi-Edge Computing
PDU	Packet Data Unit session	**AKA**	Authentication and Key Agreement

UPF	User-Plane Function	**NSSF**	Network Slicing Selection Function
LAN	Local Area Network	**NRF**	Network Repository Function
RRC	Radio Resource Control	**SCP**	Service Communication Proxy
PCO	Protocol Configuration Option	**NEF**	Network Exposure Function
DNN	Data Network Name	**UDR**	Unified Data Repository
AMF	Access and Management Function	**SEPP**	Security Edge Protection Proxy
SMF	Session Management Function	**DRB**	Data Resource Block
UDM	Unified Data Management	**QFI**	Quality of Service Identifier
PCF	Policy and Charging Function	**PLMN ID**	Public Land Mobile Network
gNB	gNodeB	**TEID**	Tunnel Endpoint Identifier
SUPI	Subscription Permanent Identifier	**PUCCH**	Physical Uplink Control Channels
NAI	Network Access Identifier	**ARQ**	Automatic Repeat-request
IMSI	International Mobile Subscriber Identity	**CSI**	Channel-state information
MSIN	Mobile Subscription Identification Number	**TRP**	Multiple Transmission Reception Point
MCC	Mobile Country Code	**URLLC**	Ultra-reliable, low-latency communications
MNC	Mobile Network Code	**PSS**	Primary Synchronization Signal
GUTI	Globally Unique Temporary Identifier	**PBCH**	Physical Broadcast Channel

SSS	Secondary Synchronization Signal	**DAPS**	Dual Active Protocol Stack
SSB	Synchronization Signal Block	**GNSS**	Global Navigation Satellite System
IAB	Integrated Access and Backhaul	**PDCP**	Packet-Data Convergence Protocol
PRS	Positioning Reference Signal	**MCG**	Master Cell Group
SDAP	Service Data Application Protocol	**BCCH**	Broadcast Control Channel
SCG	Secondary Cell Group	**CCCH**	Common Control Channel
PCCH	Paging Control Channel	**DTCH**	Dedicated Traffic Channel
DCCH	Dedicated Control Channel	**TTI**	Transmission time interval
TF	Transport Format	**BCH**	Broadcast Channel
PCH	Paging Channel	**MIB**	Master Information Block
DL-SCH	Downlink Shared Channel	**UL-SCH**	Uplink Shared Channel
UCI	Uplink Control Information	**DCI**	Downlink Control Information
PUSCH	Physical Uplink Shared Channel	**PDSCH**	Physical Downlink Shared Channel
PDCCH	Physical Downlink Control Channel	**PUCCH**	Physical Uplink Control Channels
PRACH	Physical Random-Access Channel	**T_U**	Useful Symbol Time
T_{cp}	Cyclic Prefix	**DC subcarrier**	Direct Current Subcarrier
EPS-AKA	Evolved Packet System based Authentication and Key Agreement	**SEAF**	Security Anchor Function

AUSF	the Authentication Server Function	**SIDF**	Subscription Identifier De-concealing Function
5G-AKA	Fifth Generation Authentication and Key Agreement	**SUCI**	Subscription Concealed Identifier
TIMSI	Temporary Mobile Subscriber Identity	**KAUSF**	Key of AUSF
KSEAF	Key of SEAF	**KAMF**	Key of AMF
RAM	Random Access Memory	**EPC**	Evolved Packet Core
SIM	Subscriber Identity Module	**USRP**	Universal Software Radio Peripheral
Ki	Secret Authentication Key	**MSISDN**	Mobile Station Integrated Services Digital Network
OP	Operator Code	**OP_c**	Derived Operator Code
NGAP	Next-Generation Application Protocol	**RF**	Radio Frequency
OAI	Open Air Interface	**QAM**	Quadrature Amplitude Modulation
SCTP	Stream Control Transmission Protocol	**UESIM**	User Equipment SIM
AES-KEY	Advanced Encryption Standard	**Web UI**	Web User Interface
KHz	Kilo Hertz	**MHz**	Mega Hertz

Introduction:

The arrival of 4G offered higher data rates and high-quality streaming capabilities that were nearly impossible in the previous generation while still supporting voice call functions on the go. However, with the increasing number of users, streaming demands, and media consumption by super-advanced mobile devices the 4G network has started to reach its technical limits in terms of how fast it can transfer data across the spectrum not to mention the evolution of IOT (Internet of Things) and the data-driven industries that required more and more devices to be connected and work reliably and securely at the same time which added additional pressure on cellular networks to keep up and provide such services.

5G networks have been targeted to meet the requirements of a highly mobile and fully connected society so unlike its predecessors, it takes the expectations to a totally different level with its new Radio Interface and new Core promising very low latency communications that allows several applications to provide services almost immediately, reduced congestion, and faster speeds by utilizing higher frequency ranges. This can be seen in time-critical situations like automotive and crash detection, transportation management, robotics automation, and remote surgery. This means that 5G is not only for consumers mainly but can be used for many different applications and services.

The progress was monitored and standardized by the 3rd Generation Partnership Project (3GPP) which was a collaboration between several telecommunication associations around the globe with the aim of defining the rules and standards for mobile networks based on evolved GSM (Global system for mobile communication) core and supporting the radio technologies they implement, the security capabilities, and the services.

Currently, the 3GPP website indicates several versions of releases starting from 15 which are related to enhanced broadband for SA and NSA networks, while releases 16 and 17 focus on improving existing features in the areas of MIMO and beamforming and introducing special use cases for low latency applications. Release 18 evolves 5G in the areas of Artificial intelligence and Machine learning [2].

In the following pages, we will take a closer look at the architecture and foundation that 5G operates on. Furthermore, discussing the Radio interface and technologies used and how everything connects together. Followed by that we concentrate on the experiments performed in the LAB where we report the results we obtained from connecting several UEs to the SA-5G network which we deployed using different radio interfaces along a full 5G core implemented through Open-Source Software and broadcasting the signal of the network through radio antennas connected to a USRP. We will try to report the obtained throughputs, and connection stability while testing different setups and later dissect the traffic exchanged through the communication channel to understand the initial access protocols and validate the authentication mechanisms used, either the legacy one or the new version relying on SUCI concealment.

Chapter 1:
5G Overview

1.1 The Overall Architecture of 5G:

The fundamental pillars of a 5G network are the ability to perform end-to-end (E2E) network slicing, service-based architecture, Software-Defined Networking (SDN), and Network Functions Virtualization (NFV) and because of virtualization and standardized interfaces, we're able to achieve rapid deployment and guaranty more compatibility between service providers, hardware manufactures and consumers' needs. 5G-PPP Phase I/II collaborative research projects have defined various possible customer-provider relationships between verticals, operators, and other stakeholders so the main entities we have are the **Service Customer (SC)** that uses the services offered by a **Service Provider (SP)** that can offer several functionalities depending on the request so it offers traditional telecommunication services, Digital Services like broadband or IOT, and it can offer Network Slice as a Service (NSaaS) that allocates a slice of the network with all its functionalities and configurations. **The Network Operator (NOP)** is responsible for orchestrating resources, potentially from multiple **virtualized infrastructure providers (VISP)** which are entities Providing virtualized infrastructure services, designs, and builds. Finally, we have **Data Centre Service Provider (DCSP)** which differs from VISP by offering raw traditional resources in centralized locations whereas VISP aggregates several technology domains and offers them through a unified API.

The adoption of network slicing, i.e., executing multiple logical mobile network instances on a shared infrastructure requires a continuous reconciliation of customer-centric service level agreements (SLAs) with infrastructure-level network performance capabilities[4]. e.g., Service Customers used to request the creation of telecommunication services from Service Providers. Previously, operators performed these mappings in a manual way on a limited number of services or slice types. However, With the increasing number of these customer requests, an E2E framework for Service Creation and Service Operations needs to implement a high level of automation for the management of network slices.

1.2 An Overview of 5G Non-Standalone:

The 3GPP standards provided the operators with a path that implemented a gradual transition to a full 5G network, NSA 5G is the initial step towards that ultimate solution and it helped make the move as smooth as possible for both the manufacturers and consumers. The idea of NSA is that it connects the NR (5G Radio) to the legacy 4G core network. However, it still relies on the 4G eNB for all control plane signaling with the control plane core network. Thus, not being able to operate by itself without the help of a master 4G node which's why it's called non-standalone.

In legacy 4G networks the serving gateway (SGW) and the packet data network gateway (PGW) were responsible for both signaling and data exchange, unlike 5G where a separation was introduced between the user plane and control plane (CUPS). So, to make the transition more manageable the same idea of CUPS was implemented in the 4G core so that after this the SGW and PGW would have their own control and user plane. So finally in such an

arrangement, the User equipment (UE) that's trying to connect to the 5G NSA network would rely on the master eNB for signaling and procedures like the session mobility management, UE IP address allocation, charging policy, and so on are all handled by the 4G core. Only the network access is 5G in this case so the device would communicate to the 5G NR, and NR would send the data traffic to the SGW via the S1-U tunnel and is encapsulated in general packet radio service tunneling protocol (GTPU).

1.3 The main concepts and entities in 5G SA:

In this section, we will define some of the most important concepts and headlines in the world of 5G networks and then we will discuss in detail all the main entities in the network and finally describe the flow of a PDU session more clearly.

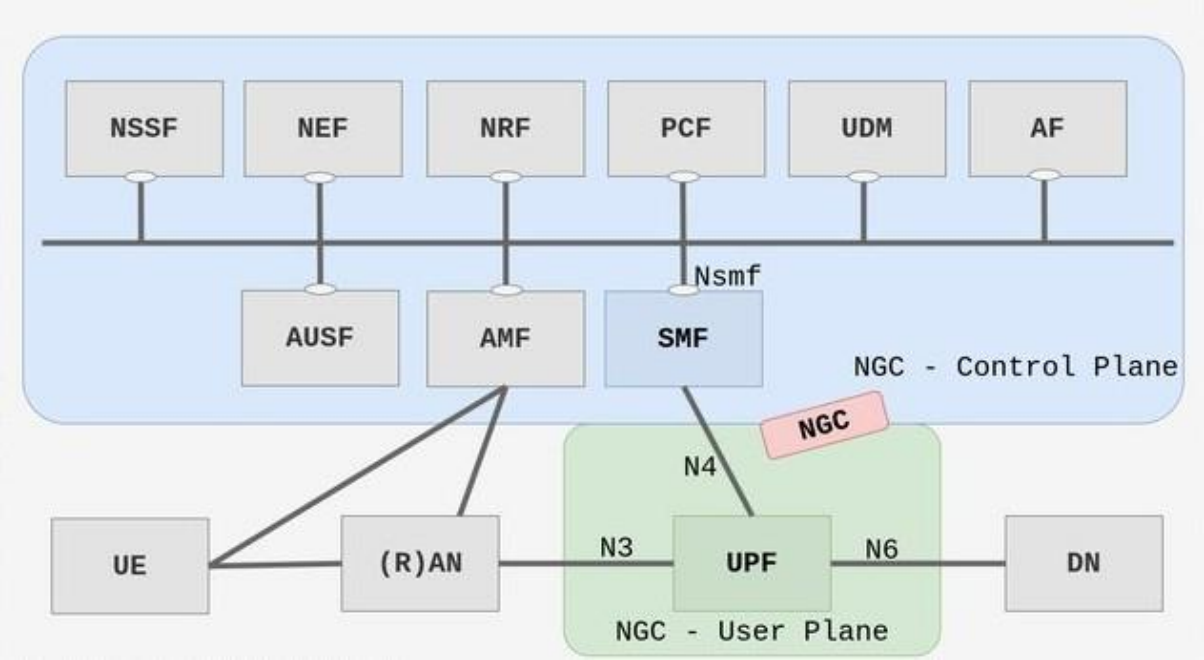


Figure 1-1: A simplified 5G Architecture with the main elements and components[3].

1.3.1 Packet Data Unit session:

The purpose of a PDU session is to carry data between the UE and the UPF (user plane function). This connection is mainly managed, initialized, and terminated by all the control plane entities in the core whereas only the UE, gNB, and the UPF are considered to be data plane entities. The PDU session can have 3 types depending on the usage scenario, the first one would be IP type which's used for the purpose of moving traffic from the UE to the network, the second one is ethernet and it's used in cases where the UE is part of a LAN and connected to the UPF so it can provide the UE with a layer 2 connectivity. The third type is unstructured and in such cases, the network doesn't know the session formats and it only operates as a tunnel or pipe for packet transfer that's mostly from IOT devices.

After a successful establishment of the session, the UE can start making calls and browsing the web. First, The UE will start RRC (radio resource control) connection request to the gNB with a request for the creation of a PDU session. The UE indicates its desirable network slice, the data network name, and a PDU session ID that's self-generated. It also sends its session management capabilities and the PCO (protocol configuration options). In case the UE didn't

specify any values for the DNN or the network slice default values are used. Furthermore, the following entities are involved:

- AMF: The access and mobility management function.
- SMF: The session management function.
- UDM: Unified data management.
- PCF: Policy and charging function.

The request is processed by the AMF and sent to the SMF, after that the SMF interacts with UDM for subscription details about the user, PCF for policy and charging details, and UPF for the n4 TEID (tunnel endpoint identifier), after that the SMF responds to AMF with success, which is forwarded to the UE by the GNB.

1.3.2 Subscription Permanent Identifier (SUPI):

SUPI is a unique identification code assigned to all sim cards used in the 5G network, it can be in a format similar to the traditional international mobile subscriber identity (IMSI) or network access identifier (NAI) in case of non-sim devices. it consists of 15 or 16 decimal digits consisting of MCC, MNC, and MSIN explained as follows:

MCC (mobile country code) and MNC (mobile network code) are 3 digits codes used to identify a mobile network across the globe. Whereas the MSIN (Mobile Subscription Identification Number) is a 10 digits identifier used to distinguish a UE user by the service provider.

1.3.3 5G Globally Unique Temporary Identifier 5G-GUTI:

The SUPI is never sent in the clear via the radio access network because if it got intercepted by intruders or attackers a DOS attack is likely to happen. So the UE is assigned a temporary GUTI to distinguish it over the radio network. The AMF is responsible for the allocation of the 5G-GUTI during the network registration phase, it consists of two sections, the AMF ID and the temporary identifier for the UE that's being changed frequently.

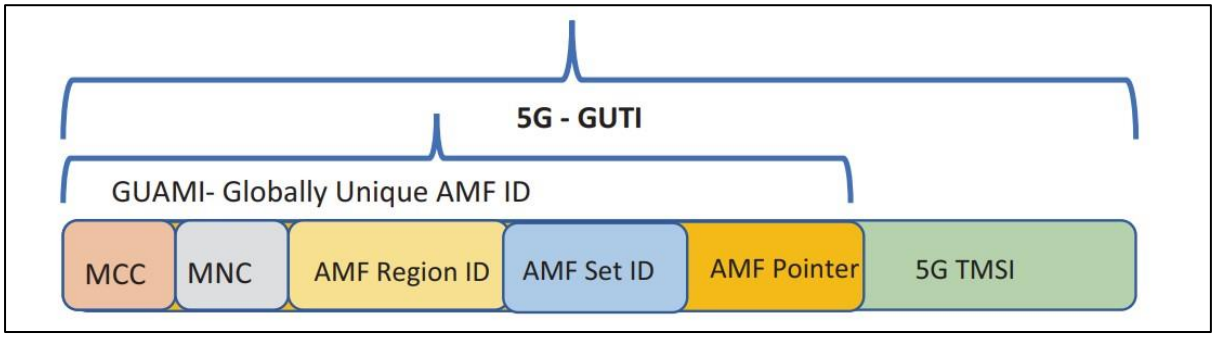


Figure 1-2: 5G GUTI Structure, source [5].

Where the AMF ID (GUMAI) is formed by the combination of:

- AMF Region ID that indicates the specific region of the AMF.
- AMF Set ID: indicates a set in that region.
- AMF Pointer: indicates an individual AMF in that set.

In addition to the MCC and MNC as explained before.

1.3.4 The Quality of Service in 5G:

The PDU session that the UE establishes can consist of many Quality Of Service Flows that can be distinguished by a unique identifier (QFI), unlike the model in LTE where the process of adding an additional flow required a new bearer to be created. In 5G the UE can add or remove different flows from the PDU session through a PDU modification request to indicate different needs for priority and latency. When the UE first initializes the session a non-guaranteed bit rate (GBR) QoS flow is set up and later on additional flows for voice or video can be added.

1.3.5 Service-Based Architecture SBA:

Unlike LTE, 5G is built with the concept of a Service-Based Architecture that's mostly used to implement modularity in Software Applications and it's newly introduced to the telecom world through 5G. So in SBA, the 5G core functionality is achieved by several functions providing services to other authorized functions in the network via what's called service interfaces like the N7 interface between the SMF and the PCF. Therefore, The Representational State Transfer (REST) architectural design model is used to support the communication between these functions a model that's derived from the client server-based architecture, unlike LTE's traditional telecom protocols like diameter and GTP.

1.3.6 New Radio (NR):

NR is one of the most significant changes introduced in 5G, antennas are critical elements in the communication process and previous generations of mobile networks usually used only one antenna both at the transmitting end and receiving one. The principle of MIMO (Multiple input and multiple outputs) was introduced with the introduction of 3G and 4G and it achieved higher speeds and signal-to-noise ratio as antennas at the receiving end can combine data arriving from multiple paths and thus improving the capacity of radio transmissions. At the transmitter, the data is divided into individual streams before transmission which results in a more reliable connection. MIMO was also used in congestion with OFDM (Orthogonal Frequency Division Multiplexing) a form of modulation where data is distributed over several narrow band sub-carrier frequencies with spacing between them which makes the connection robust and improves the spectral efficiency.

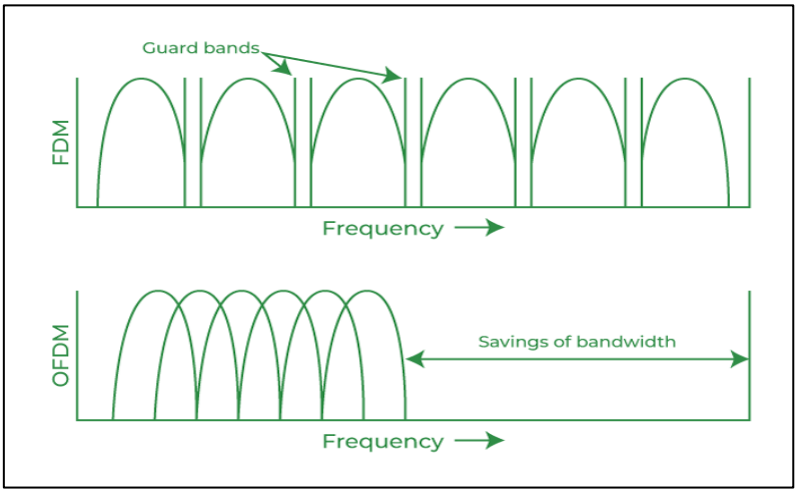


Figure 1-3: OFDM vs FDM, Source: https://www.geeksforgeeks.org.

In 5G, Massive MIMO is also used and results in more optimization, 4G used to rely on frequency division duplex (FDD) to implement MIMO while 5G relies on an additional method of Time Division Duplex (TDD) and a huge number of antenna arrays. Each one has its own strengths for example, FDD MIMO is more suitable for improving the coverage and speeds but in scenarios where network capacity is concerned, TDD MIMO can perform better. TDD lets the UE and gNB communicate over the same channel in the uplink and downlink using the same frequency but over different time slots so it's simpler and faster to implement but requires more base stations as we use shorter wavelengths and thus this increases the costs of hardware. On the other hand, FDD builds communication over two frequencies in the up and downlinks and has higher data rates for the same distance and a lower number of base stations but suffers more from spectrum wastage. So in 5G, Massive MIMO implementation results in higher bandwidth and more users in the same physical area.

The type of OFDM version used in 5G is similar to the one used in LTE in the downlink signal and it is CP-OFDM (Cyclic prefix OFDM) where the end of the OFDM frame is attached to the beginning of it and a cyclic prefix length is chosen so that it can accommodate interference caused by delays. In 5G's NR uplink signal, CP-OFDM has also been considered alongside DFT-SC-OFDM (Single Carrier OFDM) the one used in LTE considering the fact that OFDM signals have a high peak-to-average power ratio (PAPR), requiring a linear power amplifier with overall low efficiency but with SC-OFDM the transmission power is higher due to the lower PAPR. On the other hand, CP-OFDM performs achieves better compatibility with multi-antenna technologies, high spectral efficiency, and low implementation complexity. Moreover, CP-OFDM is well-localized in the time domain, which is important for latency-critical applications and TDD deployments. It is also more robust to oscillator phase noise and Doppler than other multicarrier waveforms. Robustness to phase noise is crucial for operation at high carrier frequencies (e.g., mmWave band).[6]and finally, apart from transmission schemes, 5G has a higher level of QAM(***Quadrature amplitude modulation***) up to 256 compared to 128 in LTE. The higher the QAM the throughput with the increase of the constellation points but this comes with a penalty for our waveform being more susceptible to noise and interference so these higher levels of modulation are only used when there's high SNR(Signal-noise ratio), QPSK is also used as the lowest order modulation in cases where high levels of interference are present or the link quality dropped significantly so it can provide a more robust connection.

NR operates in two frequency bands, the sub 6 GHz band with channel sizes from 5 MHz up to 100 MHz and the millimeter wave range band ranging from 24-100 GHz with channels from 50 MHz to 400 MHz Furthermore, NR's most significant role is the management of the radio resources ranging from radio admission and radio bearer control to the scheduling of resources for UEs in downlinks and uplinks and it's also responsible for IP and Ethernet header compression, encryption and integrity of information. During the attachment phase with the UE, the gNB handles the process of assigning an AMF from a list already configured, the selection process should imply a balanced distribution of load and the gNB must be able to detect failures in any AMF so that a new one can be assigned. After that comes the functionality of routing the data plane data towards the UPF (User plane function) and the control plane signals to the AMF and SMF.

In some situations where the UE is in idle mode and can't receive any messages, gNB delivers paging messages to communicate with the device as soon as it gets back online mostly in a different cell than the one it was lastly in, paging is initiated by the SMF that sends a request to the AMF communicating forward with gNB that eventually handles the scheduling and paging procedures. And finally, the task of analyzing the measurement reports that contain data about the signal strength received from the UEs is performed by the gNB to ensure proper action for handovers and mobility.

As mentioned before, the NR architecture introduces a different approach to the eNB in the sense that it performs CUPS. Therefore, we can find two main units in NR as follows, Control unit (CU) and Distributed unit (DU), and each is split into CP(Control plain) and UP(User plain) sections. This split in 5G assigns the management tasks of the high levels of the communication protocol stack to the CU. Therefore, it controls SDAP, PDCP, and RRC. Whereas, the lower levels are managed by the DU like RLC, MAC, and the Physical layer. The CU can be connected to several DUs in the same gNB but only one CU can be found. The two sections are connected through the F1 interface while the CU terminates with the AMF through the NG interface and both have different versions like NG-U or F1-C depending on the entity they connect to you meaning a connection is made between the DU and the control plane of the CU in the latter case.

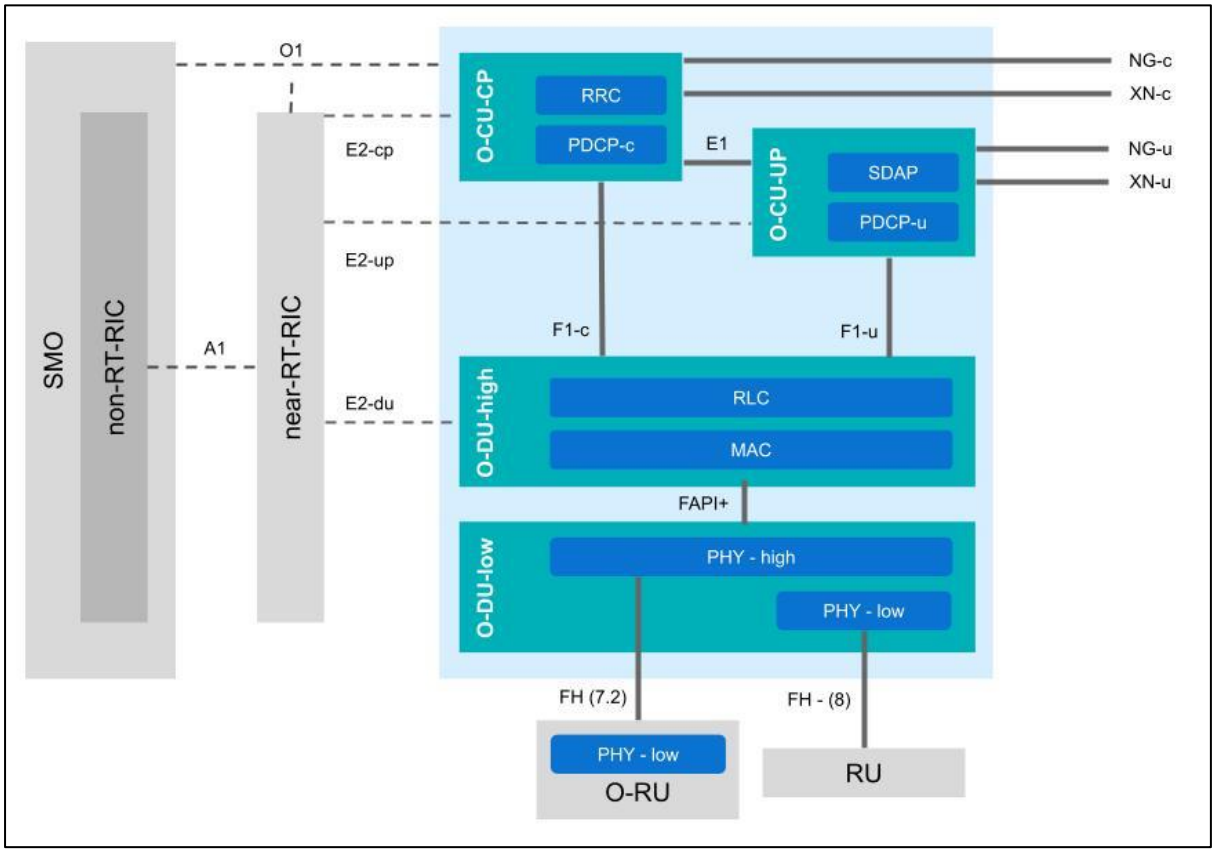


Figure 1-4: gNB high-level architecture and splits, source: https://www.srslte.com.

So as we can see in Figure 1-4 the DU is divided into high-level dealing with RLC and MAC while the low level is concerned with the PHY management, and both levels are connected through the FAPI+ interface. We can also observe the E1 interface connecting the two sections of the CU and the F1 interface either F1-U or F1-D both terminating at the high-level DU. Finally, the E2 interface is the one that starts from a specific section of the gNB and terminates at the near-RT RIC(near real-time radio intelligent controller) that's responsible for optimizing and controlling RAN functions. NR is only the entrance gate of this big network and the first step that a packet takes in its journey so now we will start looking at the several other stations along the way.

1.3.7 The Access and Mobility Function AMF:

In 4G the functions of mobility and session management were jointly managed by a single entity (MME) while in 5G another separation has been executed so we have the AMF and the SMF instead respectively handling different tasks. The AMF is the endpoint for N1 NAS signals and it's responsible for their security and encryption, Moreover, it grants permissions and authorizations to access the network and check subscription rights and policies during mobility between different radio cells. One very important function is the handover procedure carried out when switching from 5G to 4G so that continuity in the connection is guaranteed and all settings are preserved, this's achieved by an interface (N62) connecting the AMF with the MME in the LTE core supporting the GTPv2 protocol.

As mentioned before the AMF is an integral part of the paging procedure performed to deliver calls or messages to the user when it's in idle mode. This's done by when the SMF notifies the AMF to deliver data after it has been notified by the UPF so that the AMF starts scheduling and paging and executing retransmissions through the gNB.

The tasks extend to cover several selection parameters that the AMF is responsible for, most importantly it selects the SMF based on different criteria like TAC for a specific region, slice, and DNN. It's also used to query requests to the NSSF (Network slice selection function) for choosing the appropriate network slice. Depending on the geographical area or the density of subscribers, the network is mostly divided into several AMF sections where each section can have many AMF sets and each set accommodating several AMFs, so identifiers are used to distinguish between them, these identifiers along other parameters are all registered in the AMF and are as follows:

- The serving network ID.
- The MCC and MNC.
- Region ID.
- The set ID.
- The AMF pointer that's used to identify the AMF within a set.
- The slice ID.
- TAC: The tracking area code, used to identify every tracking area "cell".
- TAI: The tracking area identifier used to identify a tracking area globally. so TAI consists of (MMC, MNC, and TAC).

The AMF is linked with the Authentication Server Function (AUSF) that processes the authentication of users through the implementation of the EAP authentication server and it stores all the necessary keys retrieved by the AMF for integrity and security.

1.3.8 The Session Management Function SMF:

The SMF's main responsibility is the establishment, modification, and termination of PDU sessions for the UE with the allocation of a specific IP address for each particular session when a PDU session request arrives through the AMF, the SMF chooses a UPF for that session and in a similar way to the AMF selection procedure, the choice can be based on the DNN, slice or location. It also setups the connection between the UE and the data network and handles the management of that connectivity through the UPF, so it can communicate with the PCF to query the policy rules that govern the network and traffic steering and passes them to the UPF for enforcement.

Furthermore, charging interfaces are supported by the SMF either online or offline charging and it can also collect the data consumption reports calculated by the UPF and forwards them to the CHF.

Similarly to the AMF and to the gNB, the SMF contributes to the paging procedure as it forwards the data received from the UPF to the AMF eventually performing and managing scheduling procedures for the UE that's in idle mode. The SMF communicates with several entities and elementary configurations need to be executed, these can include configuring the IP table pool that it's used to allocate IP addresses to the UEs, and the setup of NRF(Network respiratory function) endpoint is essential for the procedure of discovering other network functions on the network and planning a backup solution in case of failures. Moreover, the SBI(Service based infrastructure) endpoints have also to be configured so the SMF can communicate with the AMF, UPF, etc. Finally charging configurations like online charging or offline charging can be designed in the SMF.

1.3.9 The User Plane Function UPF:

One of the most significant nodes in the 5G core network is the UPF, the bridge that connects the network infrastructure to the external data network, It's considered the gate and point of connection for all the PDU sessions providing mobility within and between RATs, and sending end marker or more to the gNB. It performs frame routing and forwarding, including the procedure of Uplink Classification which's the directing data to a specific DN depending on the traffic filters, and it can be a branching point for the support of multi-homing PDU sessions. Moreover, the UPF is responsible for downlink packet buffering and downlink data notification triggering.

The SDF (Service data flow) and the PFD (Packet flow description) received from the SMF give the UPF the ability to detect applications and perform deep packet inspection. Furthermore, it's responsible for the data traffic billing and charging and the lawful intercept (LI) collector interface. It handles the enforcement of the policy rules and manages the QoS related to the user plane.

The new Service Based Architecture (SBA) and CUPS in 5G provide the flexibility to deliver user plane functionality at the edge as well as the network core. So, the UPF can be co-located

with local and central data centers at both locations. This enables multi-access edge computing (MEC), which delivers resources at the edge of the network to offer low-latency, ultra-reliable, and mass-volume 5G applications as we will see later.

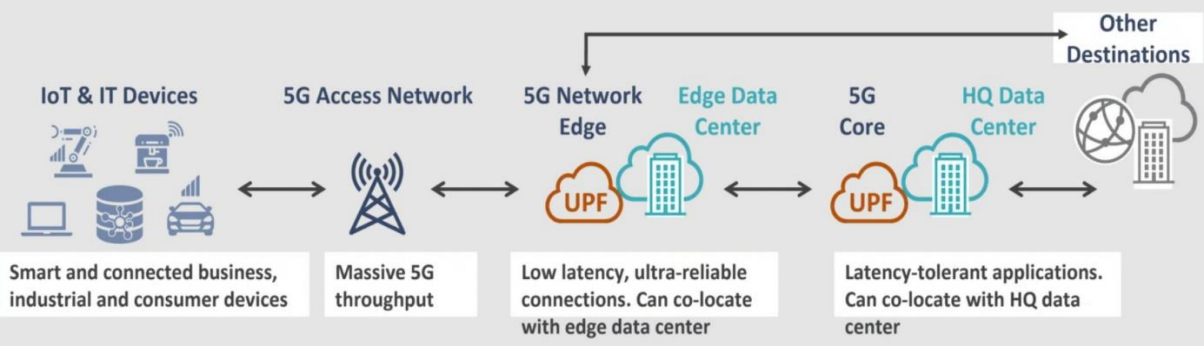


Figure 1-5: The importance of the CUPS and the UPF for low latency [7]

1.3.10 The Policy Control Function PCF:

The PCF is responsible for defining the framework that manages the network behavior, it delivers a set of rules and conditions for the control plane entities in the network for enforcement. It also takes policy decisions based on the subscription information specific to each user that can be queried from the UDR (Unified Data Repository) and then these rules can be forwarded to the SMF as an update and later on passed to the UPF for enforcement.

The PCF plays a critical role in the process of providing VoNR (Voice over NR) because it communicates with the IMS to get the rules required for creating a Voice PDU Session.

1.3.11 The Charging Function CHF:

The task of the CHF is quite simple and is significantly different from the setup found in LTE, the CHF in 5G supports and handles all types of charging and billing on the network of all its kinds either online or offline charging, we recall the difference between them being that in offline mode the charging reports are created for each user for a particular session and by the end of it the report is closed. Meanwhile, in online mode, the user session is affected in real-time and might be terminated if he runs out of balance.

The CHF is the entity responsible for the production of the charging records, these records are calculated relying on the traffic and data reports delivered by the SMF and passed to it by the UPF that did the calculations of data consumption, the reports can also include a location or network slice preference so that different billing conditions are applied, a quota is reported back to the SMF in return. So all in all the CHF is a service-based interface as we will discuss later.

1.3.12 The Unified Data Management and The Unified Data Repository Functions UDM and UDR:

The UDM is a management entity handling the user's identification, authorization, and registration. The identification process is done through the knowledge of the SUPI of each user while the authorization can be granted based on subscription rights and it also generates the AKA authentication credentials. On the other hand, the UDR is the actual place where the UDM does store all the data, it stores and delivers policy data for the PCF, subscription data

about the users, the SUPI, and application data and parameters related to the session continuity by keeping the SMF/DNN assignments.

There're two implementations that can be considered here for the UDM, either stateful or stateless. The stateful approach keeps data locally, while a stateless version stores data externally in the UDR which keeps subscriber data separate from the functions it supports. This way, the database is separate, which improves stability and flexibility. The problem is that multiple entities can't update the same points of data at once, which can cause delays in the network.

1.3.13 The Network Slice Selection Function NSSF:

Network slicing is a new concept introduced in 5G and it's one of its most significant elements, the idea of slicing states that the operator can create several virtual instances of the 5G network on the same infrastructure where each slice can have its own fixed resources to deliver for a type of traffic or application. This can guarantee a stable performance for specific use cases for example the network can offer IoT its own slice and enhanced broadband can have a separate slice.

The NSSF is the entity used for the slice selection process so that it matches the application purposes of the UE and it also chooses the serving AMF respectively. It also handles the selection of the configured and allowed NSSAI (Network Slice Selection Assistance Information). More details about slicing will be discussed in the 5G SA chapter.

1.3.14 The Network Repository Function NRF:

As we mentioned before, the 5G core implements an SBA, and in such cases when a consumer needs to contact a producer it needs to know its contact details or more accurately be able to discover that entity so that it can start requesting/providing services. The NRF is responsible for answering these types of requests as it provides the consuming entities or the SCP (Service communication proxy) with information about the available NFs on the network and the support of SCPs via SCPs instances. Furthermore, it handles the maintenance, monitoring, and updating of the NFs lists and sends notifications about them.

We can distinguish 4 main cases:

- Case A: the consumer already has the contact details of the producer.
- Case B: the consumer doesn't know the contact details hence it asks the NRF to provide it.
- Case C: the consumer first contacts the NRF for the details but it's answered with a set of available NFs instead of only one so next it communicates with the SCP that depending on specific logic chooses the most suitable producer based on several factors like load balancing.
- Case D: in this case, the consumer contacts the SCP directly which in its turn communicates with the NRF and selects an appropriate producer.

1.3.15 The Network Exposure Function NEF and Application Function AF:

With the NEF the capabilities and services offered by the NFs can be safely and securely exposed to applications and functions inside or outside the operator's network like third

parties and application functions. The NEF receives information from the NFs about the services exposed to other functions and then the NEF stores this data as “structured data” in the UDR. The information can be now accessed by the NEF and re-exposed and used later by authorized entities securely such as enterprises for monitoring purposes or the enforcement of application policies. Some examples can be the location of UE, reachability, roaming status, loss of connectivity, setting charging rates, and requesting a specific QoS for a session, etc.

This leads us to the AMF which’s nothing but the representation of these entities that try to access the core and have an influence on the network and its traffic, QoS, and policies. These functions can access and interact with the other functions in the network directly in case they’re recognized by the operator or they can use the NEF otherwise.

1.3.16 Security Edge Protection Proxy SEPP:

The SEEP provides E2E security, confidentiality, and integrity protection through an interconnection channel between different 5G networks. According to the 3GPP security standards as defined in TS 33.501 [8] the SEEP implements a security negotiation interface (N-32-c) and an encrypted application interface (N32-f). Furthermore, the topologies of the two networks are completely abstracted from each other.

1.4 The PDU Session establishment:

Now that we’ve seen the main components in the core and their tasks, we can describe the process of connecting a UE to the DN through the establishment of a PDU session.

The communication tunnel that links the gNB to the UPF is called the N3 tunnel, whereas the communication link between the UE and the gNB is the Data Radio Bearer (DRB). Each PDU session has its own N3 tunnel and one or several QoS flows inside it that can be identified by the QFI. However, In case we have 3 QoS flows in the tunnel there will be 3 DRBs between the gNB and the UE.

When the UE wants to connect to the DN, it has to establish a PDU session in the network by sending a PDU Session Establishment Request to the AMF which contains the DNN to which the user wants to connect, the preferred network slice ID or the one he was registered in before, the PDU session ID that’s generated by the UE as we’ve seen before. If the UE doesn’t include the slice ID or the DNN, the AMF can use the default values configured in it. Based on the PDU session request type “one of three” as we mentioned before, the AMF can decide if the request is for a new session or associated with an existing one. The AMF proceeds with the selection of the SMF that will serve this session by contacting the NRF, the selection process can be based on the load balance, allowed network slice, PLMN ID, and TAI. Now if the PDU request type is initial and the AMF does not have an association with the SMF for that PDU session ID, it sends a session context creation request to the SMF including the session ID, serving network ID, network slice ID, and DNN.

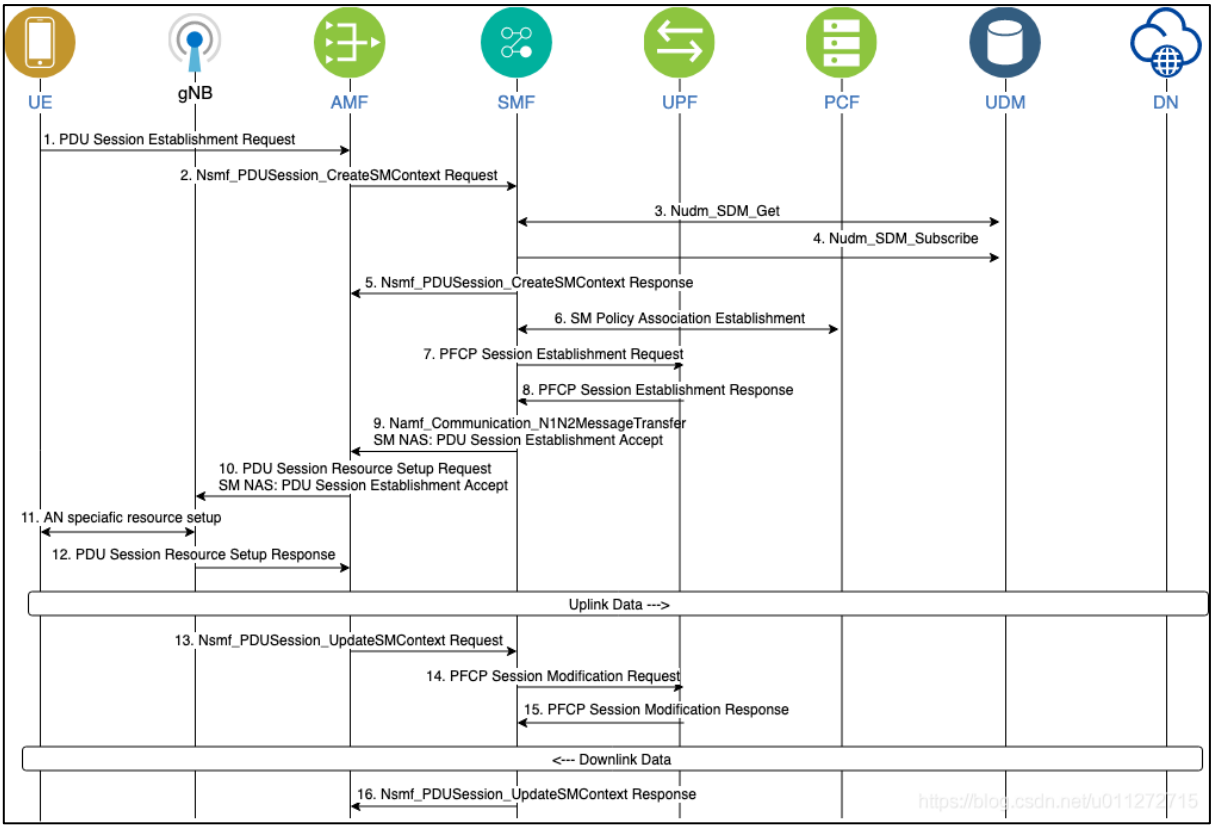


Figure 1-6: PDU session establishment, source:https://blog.csdn.net/u011272715/article/details/104251748.

The SMF now can communicate with the UDM and after the successful registration and fetching of the user's subscription data and the creation of the session management context, the SMF sends the context identifier to the AMF. The SMF can proceed with the selection of a PCF so it can receive the default policy and charging rules for the session. Furthermore, the SMF allocates the IP address for the UE and selects the UPF that will bridge the UE to the DN. The SMF sends the QoS parameters that need to be enforced by the UPF along the QFIs to distinguish between different flows for the services. It also sends the UE IP address, traffic steering rules and the policy rules to be enforced.

The UPF creates the tunnel endpoint ID (TEID) and confirms back to the SMF which in its turn sends the N1 and N2 messages to the AMF and the gNB respectively, the N1 message contains the PDU session details and session creation confirmation, the N2 message includes the N3 tunnel details along with the endpoint of the UPF and QoS flow info for the gNB. The AMF forwards the message to the gNB and then to the UE to inform it about the IP address and the QoS rules. Now the UE can start sending data in the uplink direction after the creation of the DRB with gNB that enforces the previous QoS conditions and with the known endpoint of the N3 tunnel of the UPF towards the DN.

Finally, the gNB assigns a tunnel endpoint ID for its side of the connection and delivers it to the AMF and SMF which will, in turn, send a modification message to the UPF over the N4 interface to tell the UPF about the gNB endpoint and lastly through the newly created N3 tunnel with both ends known the UPF can now send downlink data towards the UE and this way the PDU session establishment is concluded and the UE is connected to the DN.

Chapter 2:
5G NR Overview

From this chapter forward we will focus only on the NR technology and as we've seen earlier it has introduced many new approaches and advancements to the legacy eNB promising low latency communications, and enhanced mobile broadband implementing an enormous number of antennas and new beamforming technologies and of course, utilizing higher frequencies in order to take advantage of the wider transmission bandwidths by gaining more frequency spectrum. And finally modifying the radio operations to an ultra-lean design also allows for higher data rates and lower latency. So in this chapter, we will mention the features and main enhancements to NR briefly before expanding more in the following chapters to frame structures and signal types.

2.1 The Operation in higher frequencies:

In 5G an approach of using both low and high frequencies is vital. 5G NR operates in two frequency ranges, FR1 operation is done in the Sub 6 GHz and FR2 operating in between 6-100 GHz according to 3GPP. The utilization of frequencies in high orders and working in the mm-Wave region provides more spectrum availability thus enabling the usage of very wide bandwidths. This can dramatically offer high traffic capacity and extreme data rates. However, operating in such a way has the disadvantage of the signal's attenuation with these higher frequencies leading to poor coverage scenarios, especially in the case of non-line-of-sight and outdoor to indoor propagation. While the effect of this problem can be reduced partly with the usage of MIMO and beam-centric techniques, an unignorable drawback in coverage remains.

For such reasons, 5G doesn't give up entirely on the low-frequency bands and operates in both so it offers a high-frequency layer that can accommodate a huge number of users and it reduces the pressure on the low-frequency bandwidth-limited layer.

2.2 NR Ultra-lean design:

In legacy LTE and before, a heavy concern was present regarding the number of transmissions that the cells make regardless of the number of data traffic they carry. These "always on" signals had a minor issue with performance in LTE as they only occupied a very small portion of the transmissions. However, in very dense networks designed for high data rates, the traffic load per node can be low. Thus making the issue we discussed very prominent.

These signals affect the network signal's maximum achievable power and cause interference to other nearby nodes which leads overall to a reduction in the data rates. In NR the ultra-lean design approach focuses on reducing these signals like base station detection, broadcast of system information, and always-on reference signals for channel estimation. So for example in NR the reference signals are expected to be present only when there's data transmission, unlike LTE where they're always being transmitted.

2.3 Low latency:

The concept of low latency in NR has been improved drastically through two approaches. Firstly locating the reference signals and the downlink controlling signals at the beginning of the transmission and most importantly not using time-domain interleaving in OFDM symbols. This way, the decoding is hugely reduced as the device can start processing and analyzing the data almost immediately without prior buffering. Therefore requirements on the network and the device has been reduced significantly in the processing time aspect.

One more case is the modifications to protocols like MAC and RLC concerning the headers structure so that processing is possible without knowing the data length which's the opposite in LTE where the knowledge of the data amount is required therefore making low latency more challenging. This can be very important in the uplink link if the device has only a few OFDM symbols until the transmission time.

In the case of downlink data transmission, the device tries to decode the signal and then reports back to the base station. In the case of errors in the received data, the network performs a retransmission procedure to the device. However, this approach of sending the whole transport block again is mostly insufficient as the OFDM-interfered symbols between two devices might only be so few so it would make more sense to resend the interfered code block groups instead of the whole data block. And lastly, the scheduling In NR can be configured in a way that a device has already been allocated radio resources so that it can use them periodically either in the uplink or the downlink and it can communicate with the network immediately without having to wait for the scheduling request-cycle thus enabling low latency.

2.4 Control channels:

With respect to the physical layer control channels that are used to carry out scheduling data to devices in the downlink and the reports in the uplink. A major difference to LTE is the flexibility in the way they're transmitted so they can be in one or more control resource sets, unlike LTE where the full bandwidth of the carrier is used. Instead in NR, it can only occupy a portion of the carrier bandwidth so this way it can handle devices with different bandwidth capabilities. Another difference is the implementation of beamforming for control signals which was the reason for designing a new reference signal structure where each control channel has its own reference signal.

On the other hand, Physical Uplink Control Channels (PUCCH) are used to transmit Hybrid automatic repeat-request (Hybrid-ARQ) acknowledgments which are the ones the users send back to the network after receiving downlink data. It's also used for channel-state feedback for multi-antenna operation, and scheduling request for the uplink data that awaits transmission. The PUCCHs can have different formats, the format depends highly on the duration of the PUCCH transmission and the length of the data. When can use the short version of PUCCH and include it in the last two or three symbols of a slot, therefore, delivering a very fast response of the hybrid-ARQ acknowledgments considering that the delay from the end of the transmission to the reception of the acknowledgment of the device is in the order of an OFMD symbol which shows how far the improvements to latency in 5G has been

improved in many ways. On the other hand, longer PUCCHs are used in cases where insufficient coverage is an issue, and in the case of data blocks being small with respect to the data transmission, the usage of hybrid- ARQ is not performed but other codes can be selected.

2.5 Beam-Centric Design and Multi-Antenna:

NR in 5G has introduced a huge number of antenna elements, these antennas can be a boost in the high-frequency case to extend coverage and use the concept of beamforming while in the low frequency-case, it's primarily used to enable massive MIMO as we discussed before and reduce interference.

In the downlink signals, CSI reports (Channel-state information) can be included so that feedback can be obtained regarding the operation of massive multi-antenna schemes so signals and channels in NR including the ones used for control or synchronization have been modified and designed to support beamforming.

NR is more flexible and it's offering the possibility of performing analog beamforming, this technique results in a very high gain and a narrow transmission that can reach the coverage area, the idea is used in high-frequencies to shape the beam after a digital-to-analog conversion and it results in a beam that can only be formed in one direction at a given time instant and then beam-sweeping is applied so that the beam is repeated in several OFDM symbols but different beams. However, with a large number of antennas and narrow beams, the tracking of different beams can be challenging and fail often so beam-recovery procedures are introduced. Moreover, beam-management procedures help the receiver select a beam either for data or control reception. In the cases of low-frequency bands, the possibility to separate users in the spectrum is enhanced by using high-resolution channel-state information feedback using a linear combination of DFT vectors, or uplink-sounding reference signals targeting the utilization of channel reciprocity.

Additionally, in NR twelve orthogonal demodulation reference signals are allocated for multi-user MIMO transmission purposes and NR devices can at maximum receive up to 8 layers of MIMO in the downlink and four layers in the uplink. Moreover, with high frequencies and when using high constellation order the phase noise power is higher and it can degrade the performance of the demodulation so a phase tracking reference signal is introduced.

Finally, the support for transmissions to a single device from several transmission nodes (multi-TRP) is also added, including the necessary control signaling enhancements. Multi-TRP can provide additional robustness toward the blocking of signals towards the device from the base station, something which is particularly important for URLLC scenarios.

2.6 Initial Access:

The UE goes through an initial procedure of finding a network cell, receiving the network parameters, getting registered on the network, and finally be able to establish a PDU session. These steps are almost the same between NR and LTE except for a few changes that come from the dependency of NR on the lean-design principle and the beam-centric design.

the Primary Synchronization Signal (PSS) and Secondary Synchronization Signal (SSS) are a pair of downlink signals used to handle synchronization for the device and find the network. In

addition to PSS and SSS a third element is transmitted is the Physical Broadcast Channel (PBCH) which contains a set system's minimum info. The PSS, SSS, and PBCH are jointly referred to as a Synchronization Signal Block (SSB).

In LTE, the SSB block is located at the center of the carrier and is transmitted once every 5 ms so a device will surely find at least one SSB block if it exists at the same frequency otherwise it needs to search for all the possible carrier frequencies in 100 kHz carrier raster. However, in 5G the ultra-lean principle emphasizes the importance of network energy performance so the SSB block is being sent only once in every 20 ms instead of 5 ms in LTE. This in return means that the device will spend more time searching. This sparsity in the frequency domain for the SSB block means that it will mostly not be located at the center of the NR carrier so Instead of searching for an SS block at each position of the carrier raster, a device only needs to search for an SS block on the sparse synchronization raster which provides enough time for the process for the initial cell search and the improvement of the network energy performance.

2.7 Mobility Enhancements:

The aspect of device mobility is critical and very significant in any cellular network and 5G NR design took into consideration providing low latency and robust handovers between the serving cells to avoid loss of the service or interruption times.

Working at high frequencies means that an intense usage of beamforming is performed and due to the sweeping used the interruption time can be very long compared to the one in the low-frequency spectrum. So a new mechanism had to be introduced called Dual Active Protocol Stack (DAPS). DAPS is an approach to secure the next connection before eliminating the current one. The measurements and reports received from the device are used to identify a desirable handover based on the reception power of the current serving cell and the neighboring ones. Such a technique doesn't work well in situations where the signal from the serving base station drops suddenly and the handover command can't be delivered but this has also been solved by informing the device in advance about the possible cells that it can switch to and under what conditions. So in this way, the device can perform a handover by itself and conclude it even in cases of loss of the connection link with the serving base station.

2.8 Device Power:

Modifications to the power consumption aspect have also been considered for NR in 5G.

Which that in cases where discontinuous reception and bandwidth adaptation were the main focus. In a very high-level description, a sufficient example would be using resources only when in demand so when the device is in a state of data exchange with the network the NR can activate and adapt the link to be more reliable and support low latency communications with flexible MIMO schemes supporting a large number of layers. However, in situations where the device is not transferring or receiving data these features can be reduced in terms of latency demands and MIMO levels which results in less power consumption in the UE.

2.9 Fewer delays to the carrier aggregation activation:

A rapid setup and activation of the additional carriers are quite important to take benefit of the carrier aggregation that allows for higher data rates. So if the activation is not fast enough so that the carriers are used before the end of the transaction no real advantage has been accomplished this way and it's unrealistic from a power consumption aspect to keep all the carriers active which by the way would be of huge benefit to the low latency issue. Therefore, NR introduced a new mechanism to support delivering early measurements of the serving and nearby cells in addition to reductions in the signaling overhead and the latency of activating other cells. Previously in LTE without such early reports, the network needed to operate on less efficient single-layer transmission until the necessary channel-state information is available. However, now in such cases, the network can select a suitable MIMO scheme more quickly. Early reports have also been supported by keeping the context of the device and security configurations which allows the RCC connection to be resumed even after several slots of inactivity, unlike LTE where an extensive singling for setting up the security protocols is concluded. So in total, we end up in a situation where measurements can be delivered as early as during the resume procedure.

2.10 Integrated Access and Backhaul IAB:

In addition to fiber backhaul, NR has extended to wireless IAB which enables the use of NR to cover locations from a central node to distributed cells where the IAB node is configured in the network as a donor node that appears to the UEs to be normal but it creates cells of its own and additional IAB nodes can connect to the network via these cells which introduces a concept of multi-hop wireless backhauling. It's anticipated that the mm spectrum would be the most relevant as higher frequency spectrums are usually unpaired with the operation being in TDD.

2.11 Positioning in NR:

GNSS (The Global Navigation Satellite System) has been used for years for the purposes of localization and positioning. However, it requires satellite visibility in order to provide accurate measurements. Moreover, services like logistics require precise coordinates to function correctly. Therefore, additional positioning services provided by the cellular networks are required. NR relies on the so-called location servers, these servers are responsible for the distribution of information related to positioning like device capabilities, assistance data, measurements, position estimates, etc. so it can support the other entities involved in the process of positioning.

There're two techniques used that can be used either in conjunction with each other or separately depending on the accuracy required. The first method is Down-Link Positioning, it utilizes a new reference signal called positioning reference signal (PRS) that's different from the LTE version in the allocated bandwidth thus offering better estimation during the calculation process that starts with the device receiving several PRSs from multiple base stations and it measures and sends a report about the difference in the time of arrival between them then these reports are fed to the location server that will be able to estimate the device's location.

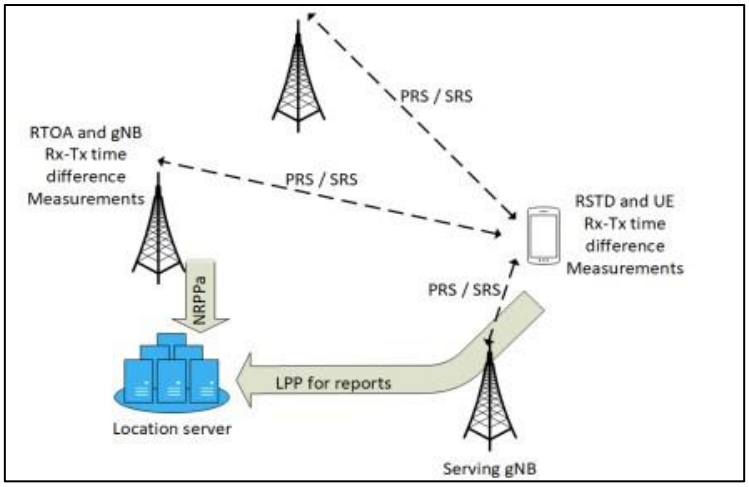


Figure 2-1: Downlink-Positioning in NR [9].

On the other hand, Up-Link based positioning relies on other reference signals that are sent by the device to multiple base stations, these signals are the sounding reference signals (SRSs) which allow the base station to calculate several parameters like the measured power and the angle-of-arrival in case of receiver beamforming usage, and the difference between downlink transmission time and uplink SRS reception time. All of these measures are delivered to the location server afterward.

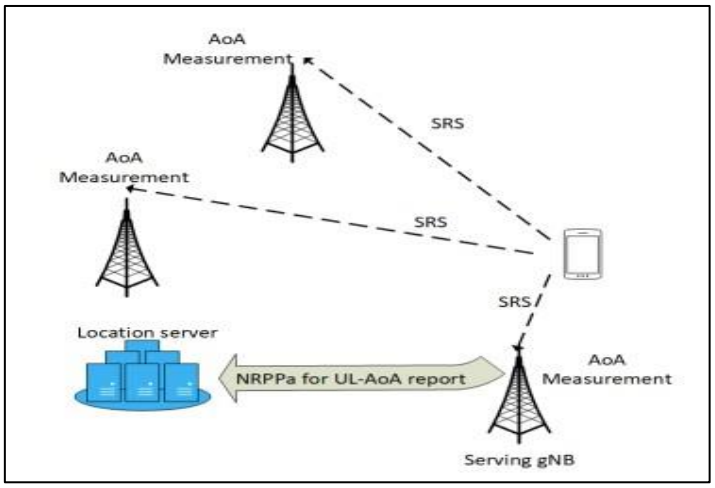


Figure 2-2: Uplink-Positioning in NR [9].

Chapter 3:
The Radio Access Architecture

After we've seen some of the main features and enhancements in NR in the previous chapter, we will focus a bit more deeply on the radio interface architecture in this one. We will be discussing the protocol stack exploring its different types of protocols being split into two parts represented in User-Plane and Control Plane and introduce the changes from the previous version of cellular networks.

The Radio Protocol Stack:

The communication protocols in 5G are very similar to LTE with the exception that now we can define the stack in the CUPS approach so some protocols will be serving only in the user plane and some only in the control plane while the others are common between the two. To illustrate more details as we can see in *Figure 3-1* the UE stack shares with the control plane the PDCP, RLC, MAC, and PHY protocols while NAS and RRC are used in the control plane of the UE and SDAP in the user plane. In the gNB SDAP is used in the user plane and again RRC in the control plane.

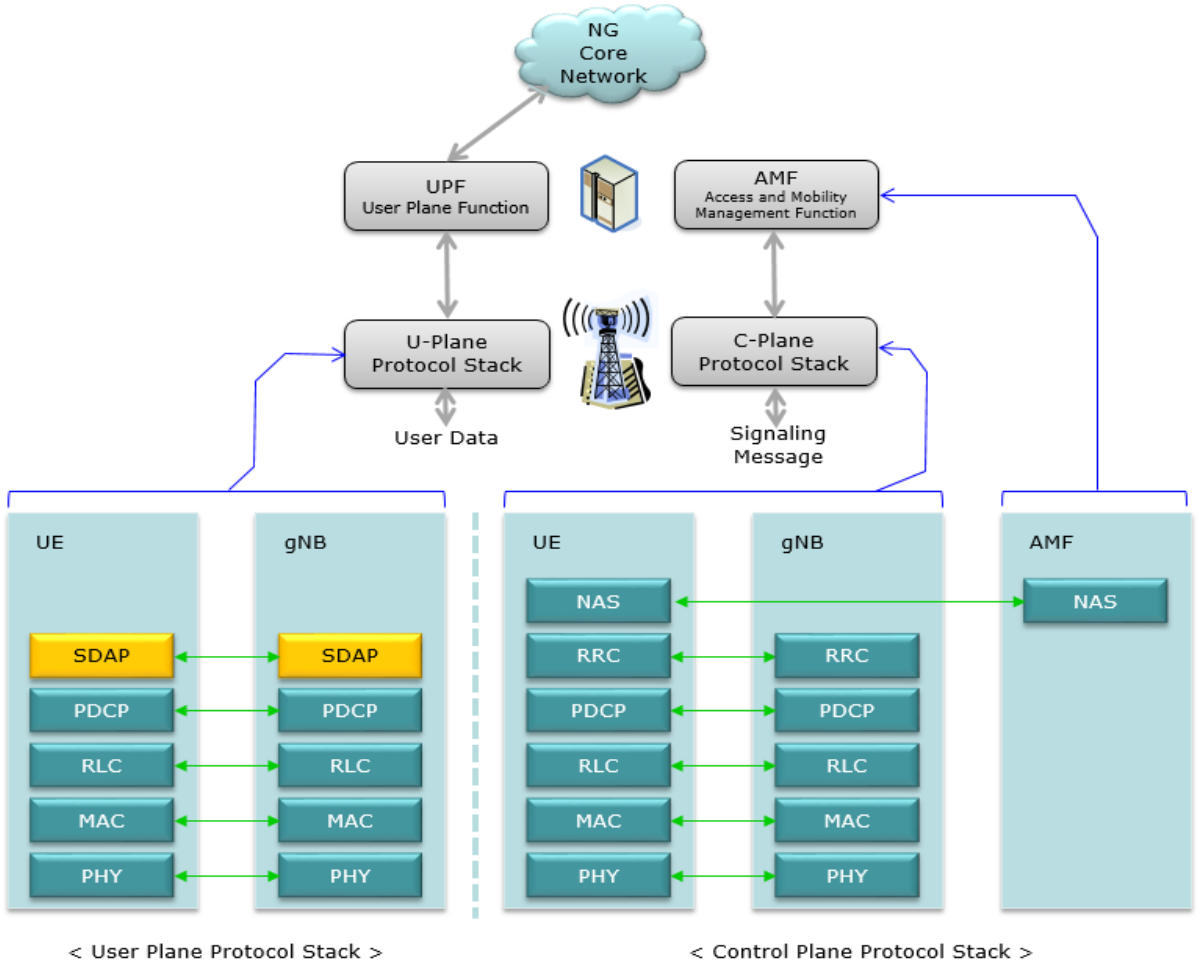


Figure 3-1: User-Plane and Control-Plane in 5G's Protocol Stack [10]

3.1 User-Plane Protocols:

3.1.1 Service Data Application Protocol (SDAP):

this protocol has been introduced with NR and it's not relevant in LTE, it's responsible for mapping the QoS bearers to the radio bearers according to their quality of service requirements and transferring the PDUs to the upper layers and it's also responsible for marking packets with QFI either in the uplink or the downlink connections.

3.1.2 Packet-Data Convergence Protocol (PDCP):

PDCP handles the process of ciphering and deciphering to protect against eavesdropping attacks and, for the control plane part it ensures integrity protection to ensure that control messages originate from the correct source. The PDCP protocol also performs IP header compression to reduce the number of bits to transmit over the radio interface. The compression approach is based on robust header compression and a set of standardized header-compression algorithms can also be used for several other mobile-communication technologies.

In cases where high reliability is essential, PDCP can perform duplication on the packets and transmit them to multiple cells which increases the possibility of at least delivering one correct packet and on the receiver end it can perform selection diversity which is the removal of duplicate frames. Moreover, duplication removal can be helpful in handover scenarios where some PDUs might be received in duplicate either over the connection from the old gNB or the new one. In addition to that the protocol forwards the undelivered packets to the new gNB and retransmits any undelivered uplink data from the device as the hybrid-ARQ buffers are flushed upon handover. Lastly, PDCP is responsible for separating the data between two nodes on the radio network like in the case of dual connectivity where we have the main master node and the secondary one as discussed before therefore the data bearer is split into two sections and PDCP handles the distribution between the two nodes. This's illustrated in Figure 3-2 and it shows the three types of bearers that can be seen in the radio nodes. So we can either have an MCG bearer when radio protocols are only located at the master node, an SCG bearer when the radio protocols are located only at the second node, or a Split bearer.

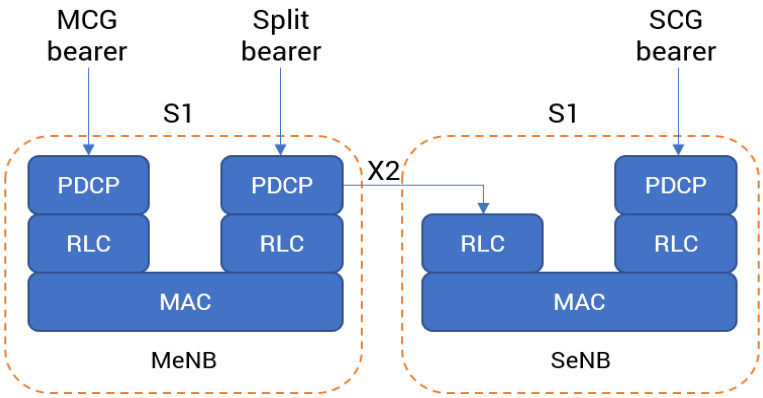


Figure 3-2: Dual connectivity with split bearer [11]

3.1.3 The Radio Link Control (RLC) Protocol:

The RLC protocol performs segmentation procedures for the RLC SDUs from the PDCP to suitable PDUs. Moreover, it deals with the retransmission of the wrongly received PDUs or the detection and removal of duplicate ones. RLC can be configured in three modes depending on the usage needs. The first mode is the transparent mode which doesn't add any extra headers or services while the second mode is the unacknowledged mode which supports segmentation and error detection and finally, the acknowledged mode which gives the ability to retransmit in case of errors.

In the segmentation task, the scheduler decides on a suitable transport block size is chosen and previously in LTE, the RLC PDU wasn't allowed to assemble until the scheduling decision is taken. This results in an additional delay until the uplink transmission and the low latency requirement of NR can't be met this way. So instead in 5G, concatenation has been removed from RLC so that the RLC PDUs can be assembled in advance and the device upon receiving the scheduling info can send a suitable number of PDUs to the MAC "depending on the allocated block size" layer and to completely fill up the entire transport block the last PDU may contain a segment of an SDU.

Errors detection and the process of retransmission is also one of the RLC responsibilities to ensure the delivery of error-free packets to the higher levels, this is accomplished by a monitoring mechanism between the transmitter and receiver that keep track of the identifiers in the headers of the PDUs "different from the PDU identifier" and if any discontinuity occurs or seen in the feedback reports towards the source, a retransmission procedure is initialized.

3.1.4 The Medium Access Control (MAC) Protocol:

MAC is the layer that interfaces with the RLC layer from the top and the PHY layer from below It's responsible for mapping the information between logical and transport channels.

It's also responsible for hybrid ARQ retransmissions and the procedures for scheduling in addition to multiplexing data over several carriers when the technique of carrier aggregation is implemented.

3.1.4.1 Logical and Transport channels:

Logical channels are the bridge that connects the MAC layer to the RLC and they can have different classifications depending on the type of data they carry. We can observe two kinds of logical channels, the ones that carry user data and are known as *Traffic channels* or the ones that deliver configuration settings and control parameters known as *control channels.*

In NR we have the following *control channels*:

- The Broadcast Control Channel (BCCH): generally, devices need to know the configuration of the network's settings and how to interact with the system so such information is provided by the gNB to all devices in a particular cell using the BCCH control channel.
- The Paging Control Channel (PCCH): it's sent to multiple cells in order to perform paging in case the network can't distinguish the location of the UE.

- The Common Control Channel (CCCH): this channel delivers control signals from the base station to the UEs and it's used during the initial phase when there's no RRC connection yet.
- The Dedicated Control Channel (DCCH): it's used for the individual configuration of devices and for sending control signals to and from the user.
- The Dedicated Traffic Channel (DTCH): same as DCCH but it handles sending data to and from the device.

On the other hand, when we talk about *transport channels* that connect the MAC layer to the PHY layer we mention that they're defined by the method and configurations they use to deliver data over the antenna. The data in transport channels are organized in transport blocks where every TTI (Transmission time interval) can transmit one transport block. Each transport block can have a TF (transport format) that basically defines the configuration used to transmit, this can include the coding schemes, the modulation rate, and block size.

In NR we notice the following types of transport channels:

- The Broadcast Channel (BCH): mainly, this channel is used to broadcast the MIB (Master Information Block) that the devices need to look for before they connect to the network as it provides information about the network parameters.
- The Paging Channel (PCH): used to send the paging messages from the PCCH logical channel.
- The Downlink Shared Channel (DL-SCH): the main downlink data transmission in NR is handled by the DL-SCH channel and it supports all NR features like dynamic scheduling, hybrid ARQ, and spatial multiplexing.
- The Uplink Shared Channel (UL-SCH): contrary to DL-SCH, this channel handles the transportation of data in the uplink direction.

One of the main tasks of the MAC as we mentioned before is the mapping of the logical channels to the transport ones. Each logical channel has its own RLC entity and supports priority, several logical channels can be multiplexed together to the same transport channel, and a corresponding demultiplexing procedure is performed at the receiver where each PDU is forwarded back to its RLC entity. With low latency in mind, the implementation of the MAC header in NR has been modified compared to the previous technique used in LTE where the MAC header that corresponds to a MAC PDU is located at the beginning which means that we weren't able to identify the set of SDUs to include in the PDU structure before receiving the scheduling decisions. On the other hand, in NR the MAC header for each SDU is placed at the beginning of it which allows for faster preprocessing of the PDU.

An additional feature that the MAC layer can offer is the ability to include MAC control elements so that they can be included in the transport channels, these elements are mostly used for in-band control signaling and this can offer a faster way to deliver control signals without the need to limited in respect to the payloads conditions or the physical layer control signaling. There're several MAC control elements and some of them the ones that are related to scheduling, random access, activation/deactivation of previously configured component carriers, or PDCP duplication detection. Lastly, the MAC layer is a fundamental stage of the

carrier aggregation process as it's the entity responsible for distributing the data from several flows to the component carriers where logical channels are multiplexed to transport channels that will be delivered per component carrier that has its own Hybrid-ARQ entity.

3.1.4.2 Scheduling:

The shared-channel principle in NR implies the dynamic sharing of time-frequency resources between the users and is mainly handled by the scheduler that's part of the MAC layer and is responsible for the allocation of resource blocks in the frequency domain and OFDM symbols in the time domain. The scheduler sends scheduling information once per slot to a set of devices based on the dynamic scheduling principle and scheduling decisions are not necessarily conducted at the end or start of a slot which's another useful thing for low latency.

Uplink and Downlink scheduling decisions are taken separately from each other and the scheduler in the downlink connection dynamically chooses which devices to transmit to and the set of resource blocks that will be delivered. Whereas, TF is handled by the gNB along the logical channel multiplexing. In contrast, the uplink scheduler selects which devices are allowed to transmit on their UL-SCH and which resources are allocated, the device in this case handles the logical channel multiplexing and the scheduler only chooses the TF.

Dynamic scheduling is vendor specific and it's not standardized in 3GPP but it can be aided by the *channel-state-information (CSI)* that we mentioned before, a report that the device sends to the gNB in order to provide the channel quality whereas sounding reference signals are used in the uplink. Reducing the overhead caused by the control signaling of the dynamic scheduler can be achieved by using predefined and configured schemes where in the downlink a similar approach to semi-persistent scheduling in LTE is used by sending a scheduling pattern in advance to the device that later on will start receiving based on that pattern. Whereas in the uplink 2 types can be seen, one in which RRC configures all the parameters and activates the transmission according to that scheme and the second where the activation is done through L1/L2 signal.

3.1.4.3 Hybrid ARQ:

This retransmission technique is part of the MAC layer and ensures the robustness of the communication link by making it error-free. This technique can't be applied to all types of traffic but is only supported for the DL-SCH and the UL-SCH and hence broadcast messages that are delivered to many users are not supported. The main working principle of ARQ is started after the reception of the transport block, if the decoding result wasn't successful the receiver can indicate that in the acknowledgment bit sent back to the transmitter that must know to which ARQ process the received acknowledgment is linked which can be solved by is using the timing of the acknowledgment relative to the downlink data transmission for association with a certain hybrid-ARQ process.

Unlike LTE, NR technology in 5G utilizes an asynchronous hybrid-ARQ to support dynamic TDD where there is no fixed uplink or downlink resource allocation. It also offers better flexibility in terms of prioritization between data flows and devices. Each process has its own identifier that's used which's used to address which process will be transmitted and where the retransmissions are in the order of the original data sequence. In NR, up to 16 hybrid-ARQ

processes are introduced and when used in parallel, from the device point of view some bits might be delivered out of order to the original one so for example if packet number 5 is decoded successfully and received before packet number 4 because it needed to be retransmitted and this's mostly acceptable in all applications otherwise forcing in sequence delivery can be done through the PDCP protocol. On the other hand, the RLC protocol doesn't provide in-sequence delivery to offer less latency. As discussed earlier, the RLC layer is also capable of performing retransmissions However, the reason for having two mechanisms is due to the reason that hybrid ARQ provides fast retransmissions but due to errors in the feedback the residual error rate is typically too high while RLC ensures (almost) error-free data delivery but slower retransmissions. Hence, the combination of hybrid ARQ and RLC provides a combination of small round-trip time and reliable data delivery.

3.1.5 The Physical Layer (PHY):

Finally, tasks like coding, modulation, the processing of the hybrid-ARQ and multi antennas, and the allocation of the time-frequency resources are all handled by the physical layer. And as mentioned before it handles the mapping of the transport channels to the physical ones. Each physical channel is mapped to a transport channel and they correspond to the frequency-time resources to be allocated. Moreover, there're the L1/L2 control channels that don't map into a transport channel and have two types, the Downlink Control Information (DCI) and the Uplink Control Information (UCI).

The DCI channel is used to send information to the UE related to the reception and decoding of the downlink data whereas the UCI provides the gNB with information from the device about the Hybrid-ARQ feedback and metrics parameters to the scheduler. So in the physical layer of NR, we can observe the following channels:

- The Physical Downlink Shared Channel (PDSCH): this's the main physical channel that handles unicast transmissions and other purposes like sending paging information and providing response messages related to random access.
- The Physical Uplink Shared Channel (PUSCH): the uplink version of PDSCH and it carries signaling and user data.
- The Physical Broadcast Channel (PBCH): a broadcast message to deliver essential settings to the devices so that they can access the network.
- The Physical Downlink Control Channel (PDCCH): handles control signaling in the downlink related mainly to scheduling decisions for the reception of the PDSCH and scheduling resources allocation that enable the PUSCH.
- The Physical Uplink Control Channel (PUCCH): this channel allows the UE to report the Hybrid-ARQ bit to identify if the reception of a transport block has been successful or not in addition to the capability of sending measurements reports back to the base station to help the process of dynamic scheduling and sending a request for obtaining resources for the transmission of uplink data.
- The Physical Random-Access Channel (PRACH): used for random-access.

3.2 Control-Plane Protocols:

On the control plane side, we have the NAS functionality that basically operates between the AMF in the 5G core network and the UE. It handles many responsibilities like assigning IP addresses during session management, authentication, registration, mobility functions, and security. And another significant protocol in the control plane is the RRC (Radio Resource Control) which operates between the gNB and the UE and deals with the control-plane signals for the network access procedures. its responsibilities are as follows:

- Providing the UE with connection information to communicate with a network cell through a broadcast technique.
- Configuring the connection parameters to establish the connection between the UE and gNB.
- Sending paging requests to the device and delivery of system information updates.
- Mobility functions such as cell reselection.
- Measurements reports configuration and delivery.
- Handling and managing the different specifications of devices' capabilities through the report they send during their initial connection phase.

RRC messages are delivered to the device through *signaling radio bearers (SRBs)* and each one is mapped to the common control channel (CCCH) at the initial phase of communication and then to the dedicated control channel (DCCH).

The device in LTE networks was able to be in two RCC states during its connectivity period. The first one is *RCC_IDLE* in which the device and core network are not connected and there's no RCC context nor the device allocation to a particular cell. Therefore, a downlink data transmission is not available but the device might wake up periodically to receive paging messages if any from the network. On the other hand, the uplink is also shut down and might only be initialized in case of random access. And finally, mobility is managed by the device only. The second state is *RCC_CONNECTED* in which the RCC context is preserved in both the device and the core network so the cell that the device is connected to and the identifier used for signaling purposes is known.

Moreover, the mobility in this case is handled by the gNB. However, in NR there's a third state called *RCC_INACTIVE* that was introduced and it offers significant features as it allows the device to stay in sleeping mode and preserve power like the idle state but it keeps the RCC context saved in both the device and the core to reinitiate the connection much faster because in LTE there was a problem in the latency and signaling load caused by the frequent switching from the idle state to the active one especially in cases where repetitive reception of data packets occurs.

Chapter 4:
The Transmission Architecture, Initial Access, and Security.

In the fourth chapter, we take a look at the transmission structure of the used transmission schemes in NR to try to understand how they're composed and what benefits they serve. Later we proceed with describing the initial random access procedure to hook the device to the network and allocate transmission resources to it to eventually end the chapter with an overview of the authentication algorithm that's being used in 5G and how different is it from the previous version in LTE with a few words about the strategies used to immigrate from LTE to 5G.

4.1 Transmission structure:

As we've already mentioned in Chapter 1, NR keeps using OFDM as its main waveform due to its robustness in the medium and ease of assembling its structure for different signals and channels either in the time or frequency domain. The difference to LTE is mainly in the uplink/downlink arrangement as NR uses not only DFT-SC-OFDM in the uplink but also CP-OFDM that's used in the downlink to achieve less phase noise at high frequencies and is more compatible with multi-antenna technologies as it offers the possibility of using not just one layer but up to four.

Now to discuss things even further, we have to mention the importance of the sub-carrier spacing and the cyclic prefix lengths in OFDM due to their huge impact on the transmission link. Choosing a large sub-carrier spacing normally results in reducing the phase noise and allows the implementation of larger bandwidths. However, from a cyclic prefix point of view, the overhead in the channel increases when we increase the spacing. In LTE where the usage scenario was outdoor deployments with a carrier frequency of around 3 GHz, the subcarrier spacing of 15 KHz with a cyclic prefix of around 4.6 µs was sufficient. Whereas in NR, several operation modes exist ranging in cell sizes and frequency bands from sub-1-GHz to mm-waves. Thus, relying on a single numerology wouldn't be efficient because, in frequencies close to 1 GHz, cells are usually very large so choosing a large cyclic prefix length is required to make up for the delay spread and the carrier spacing would be in the order of LTE scenarios between 15-30 KHz. However, NR also operates in the mm-wave range where cell sizes are smaller and the phase noise is much higher which requires a bigger spacing interval. Thus, a shorter cyclic prefix along beamforming is preferred in such deployments.

Therefore, NR has defined 15 KHz as the default value to offer compatibility with LTE with dynamic change available in the range of 15-240 KHz proportional to the change in the cyclic prefix values, and in *Table 1*, we can see that the Useful Symbol Time T_u depends on the spacing and the total length of the OFDM time is the sum of the useful symbol time and the cyclic prefix length. Although in LTE there were two cyclic prefixes, either normal or extended that were mainly used in situations where a very large delay was present, in NR only one mode "normal' is available.

Subcarrier Spacing (KHz)	Useful Symbol time T_u (µs)	Cyclic Prefix T_{CP} (µs)
15	66.7	4.7
30	33.3	2.3
60	16.7	1.2
120	8.33	0.59
240	4.17	0.29

Table 1: Supported transmission numerologies in NR.

4.1.1 Time Domain Structure:

NR transmissions in the time domain are structured in frames with a length of 10 ms, these frames contain 10 subframes of length 1 ms that are further divided into 14 OFDM symbols. And this arrangement is similar to the one found in LTE with a subcarrier spacing of 15 KHz which means that the cyclic prefix for the first and eighth symbols are slightly larger than the rest. Scaling the baseline structure by powers of two allows us to derive the higher subcarriers' spacings in NR. So one slot will split into two OFDM slots of the next order so the 14 OFDM symbols will have a length of 0.5 ms in the case of 30 kHz subcarrier spacing and so on, but the frame length will always be 1 ms with 2^{μ} slots representing the dynamic scheduling unit and no matter which numerology we're using this approach of scaling by a factor of 2 helps maintain the symbol boundaries and simplifies using different numerologies on the same carrier.

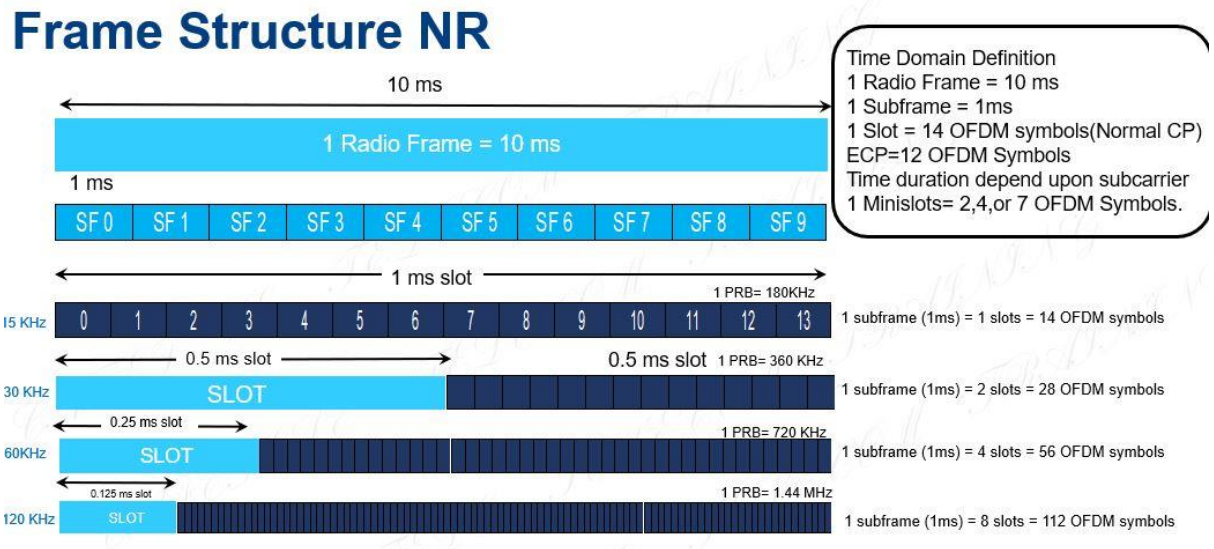


Figure 4-1: Frames, subframes, and slots in the time domain of NR, source: https://cafetele.com/5g-nr-frame-structure.

So from the structure seen in *Figure 4-1,* we conclude that a slot has a fixed length, and increasing the subcarrier spacing will result in a shorter duration. While this can be considered a good approach to achieve lower latency we have to pay attention to the decrease that happens to the cyclic prefix length as a result which might not be suitable in all scenarios and a method to keep the cyclic prefix around the values defined for the 15 KHz while opting for the shorter slot duration is required. Therefore in the case of 60 KHz spacing an extended cyclic prefix has been defined and even though it's not the most practical way to provide lower latency due to the increased overhead. Therefore, it's only used in cases where a large

delay is expected to happen. However, a more efficient way to achieve low latency has been introduced in NR and it's the ability to use only a part of a particular slot for a transmission. So in case of transmission that requires very low latency, the data can start at any OFDM symbol without the need to wait for a slot edge and many features can be achieved this way like the ability to direct an already going transmission process to a different device or even in the case of analog beamforming where we can only use one beam at a time to transmit and the need for the multiplexing of devices in the time domain is needed and considering the fact that in the mm-wave range, we have very large bandwidths so a few OFDM symbols are more than enough to send the whole payload with the excessive use of a complete slot. Lastly, it allows the device to transmit on the channel immediately after it confirmed its availability instead of waiting for the beginning of a new slot in case of operation in the unlicensed spectrum where a listen-before-you-talk approach is used so it can guarantee no other device is going to take the channel during that period.

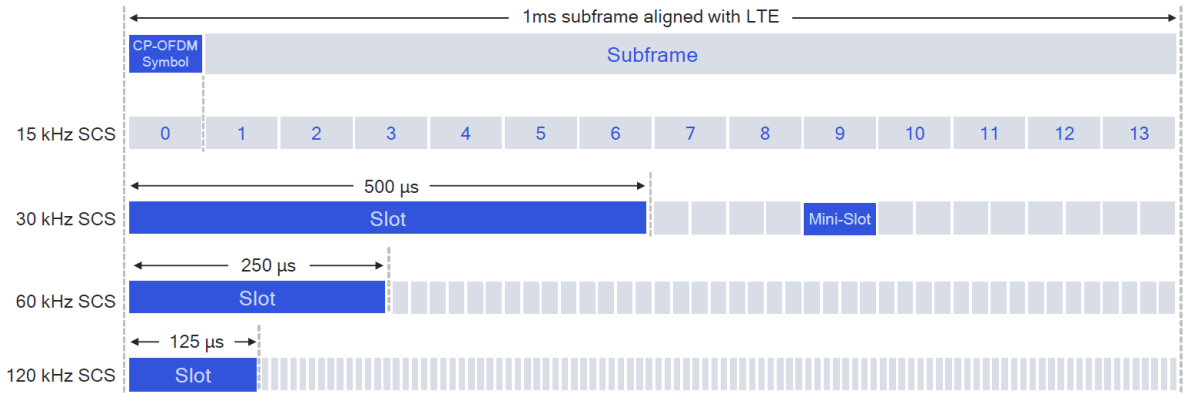


Figure 4-2: Mini slot concept, source: https://devopedia.org/5g-nr-transmission-time-interval.

4.1.2 Frequency Domain Structure:

Previously in LTE, it was a requirement that all devices must support the maximum transmission bandwidth of 20 MHz. whereas in NR, several bandwidths are available up to 400 Mhz while this was a huge enhancement it also came with some drawbacks as not all devices in 5G can support these huge channel sizes and it would be so expensive to force such thing so the problem in NR is that a 5G device might be able to detect the whole part of the carrier so as a result we end in a situation where the carrier is not perfectly aligned with the carrier at the center frequency so in NR the need to use the DC subcarrier was mandatory with the acceptance of some performance degradation due to interference.

Moreover, a frequency resource element in NR is composed of one subcarrier during one OFDM symbol and 12 consecutive symbols resemble a resource block. These resource blocks are processed in many numerologies, unlike LTE where only one unified numerology was used. And they're designed so that the frequency range is the same for two resource blocks with a spacing of Δf with respect to one resource block with a spacing of 2Δf. The resource grids are what's used to describe the boundaries between the blocks exploiting different numerologies, the device needs to know where these blocks are located in the carrier and this was an easy process considering the unified numerology. However, in NR the need for reference points is required to help assist this task. So the actual physical resource blocks that's used to send the transmission are located

in a location in reference to that point. So for example we can say that the physical resource block 0 corresponding to the spacing of Δf is x resource blocks away from the reference point. Finally, the first allocated resource block for a subcarrier spacing say of 2Δf is positioned further away from the carrier edge than for subcarrier spacing Δf to allow a larger portion of the spectrum to be used for spacing of a lower level.

4.1.3 Bandwidth Parts:

Bandwidth parts are introduced in NR to handle the issue of devices not capable to process the full transmission bandwidth. Furthermore, the large bandwidths available in NR require taking the limits of the UEs on the network as the increase of bandwidth will lead to an increase in the power consumption in the device and in situations like LTE where the full carrier bandwidth was used to transmit control signals so a new approach in 5G was needed to solve such issues. A bandwidth part is built by defining a particular numerology and a particular set size of consecutive resource blocks in that numerology.

Only one bandwidth part is active in one carrier and a device can be configured for up to 4 parts for both the downlink and the uplink for a serving cell.

4.2 UE'S Initial Access in NR:

The initial access procedure in NR helps the It helps

UE gets its initial uplink resources and performs synchronization with the gNB.

The procedure has two types, either contention-based or contention-free. The first approach is applied when the UE isn't yet synchronized or has lost its synchronization with gNB while the contention-free one is used when the gNB has already synchronized successfully to a gNB previously and both ways include the process of sending a random access preamble from the UE to the gNB on time or frequency resources that were indicated to the UE on the control channels.

The base station transmits synchronizations signals (PSS and SSS) and broadcast channels (PBCH) on a frequent basis and then the device does beam measurements and picks the best beam during the synchronization process and consequently decodes the MIB/SIB on that beam which carries the system and network parameters. the UE now uses that beam to transmit the RACH preamble to try random access and the gNB replies with a "RA Response" message. After that the UE sends an "RRC connection request" and the base station replies with "RCC connection setup" followed by an SS block and the CSI-RS which is used to generate the beam/CSI report. This reports gets requested by the gNB and the UE sends it back. Finally, after the procedure is completed a dedicated connection is created between the gNB and the UE with a certain connection ID.

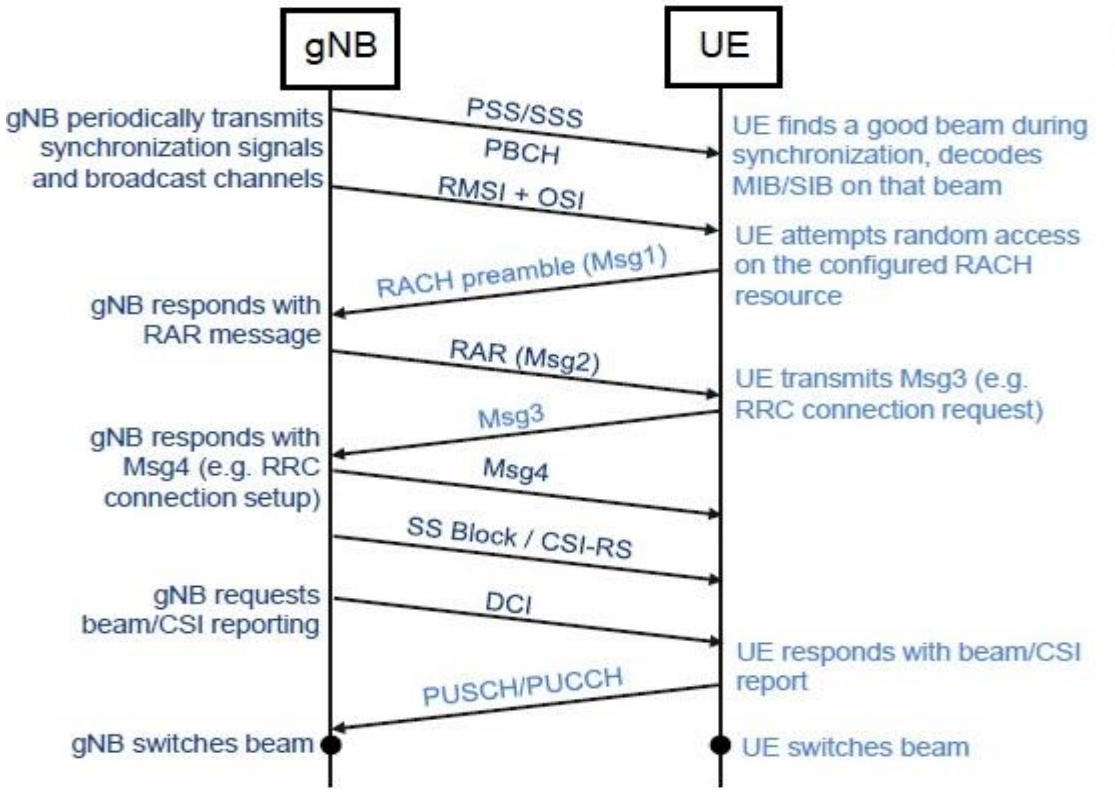


Figure 4-3: Initial access in 5G [12].

4.3 Security and Authentication in NR:

The Authentication of users on cellular networks is a very significant matter as it ensures the safe interaction between the network and different UEs, it also guarantees no attempts to fake messages or pose fake base stations. Therefore, enhancement to this aspect of the network is crucial.

Previously in LTE, the authentication algorithm used was 4G EPS-AKA, while it offered key derivation and secure messages transmissions it had a major flow represented in transmitting the user's identity "IMSI" in the clear during the first access, such act allows for a potential "Man in the middle" attacks where the attacker can change the IMSI or use it. and although a GUTI approach is also used in LTE and it can be sent instead of the IMSI this has been proved to have some serious drawbacks as the temporary identifier is not changed as frequently as it should [13] and the way that the GUTI is being allocated in is predictable [14]. Moreover, when the network sends an "Identity Request Message" to the UE, it will simply send its permanent identifier in the clear with an "Identity Response Message". And lastly, the authentication decision is made only by the serving network and the home network is only responsible for providing the Authentication vectors Avs. Therefore, these issues needed to be taken into consideration when designing an authentication approach for 5G.

As discussed in Chapter 1 that 5G's core was designed to implement a service-based architecture, so it's built on the concept of several entities with functions they provide to

other authorized ones in the core and regarding the security and authentication matter, the main functions involved are as follows:

- The Security Anchor Function (SEAF): this function is considered to be like a middleman between the UE and the home network, it can reject an authentication request but approval can't be made and it's done only by the home side.
- The Authentication Server Function (AUSF): it's the main entity responsible for the final decision about accepting the UE and it's located in the home network. It relies on the UDP to compute the keys and authentication data.
- Unified data management (UDM): it calculates authentication data and selects an authentication method based on the UE identity.
- The Subscription Identifier De-concealing Function (SIDF): the responsibility of this function is to decrypt the encrypted "IMSI" using the network's private key that's associated with the public key used by the UE.

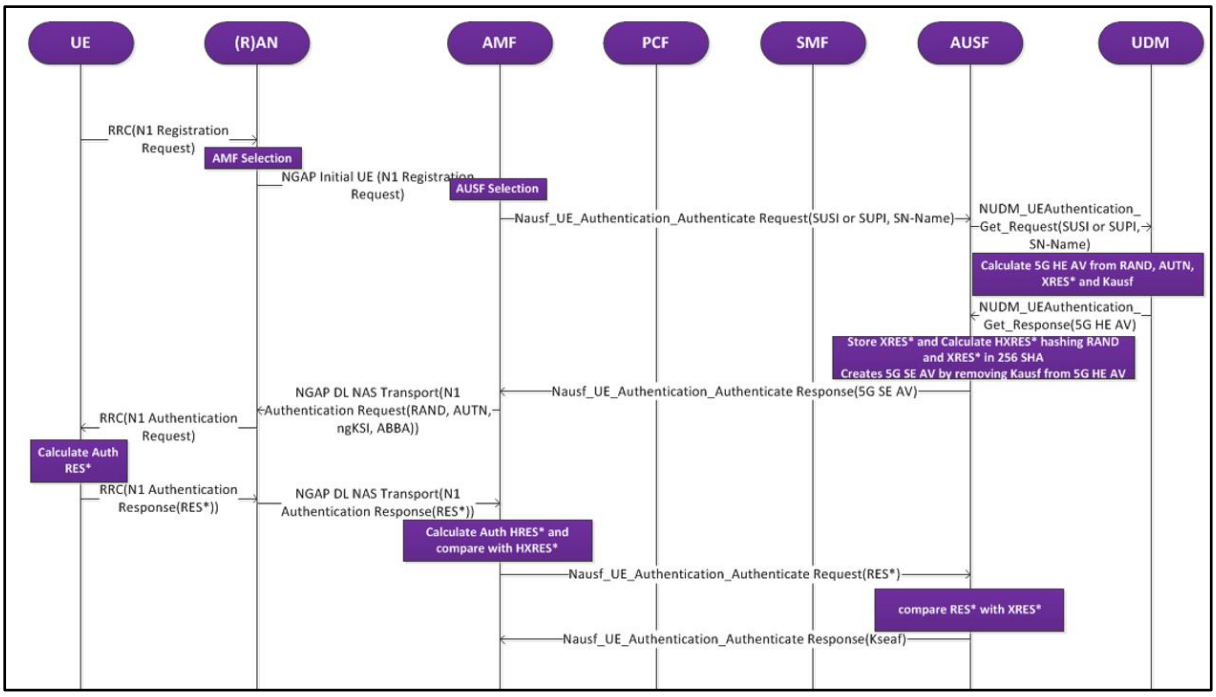


Figure 4-4: 5G Authentication procedure and entities [15].

4.3.1 The 5G-AKA Algorithm:

The authentication procedure starts when the UE transmits a signaling message to the network requesting the registration, it sends its identity either in a 5G-GUTI format or in a SUCI (Subscription Concealed Identifier) format which's basically the IMSI but encrypted using the public key of the network that was distributed to UEs. The UE usually sends a SUCI when it still hasn't been allocated any TIMSI by the network yet.

The identifier is delivered through the gNB to the AMF to the SEAF which will start the authentication procedure by first contacting the AUSF which starts by verifying if the serving network that's requesting the service is trusted. Upon saucerful authentication, the request moves forward to the UDM which replies with an "Authentication Response" which basically

holds an authentication vector that includes an authentication Token (AUTH Token), XRES Token, and the Key KAUSF along the SUPI if applicable. If the request that was delivered to the UDM included the SUCI identifier, the UDM triggers the SIDF to encrypt it so that the UDM can choose the authentication method that's configured to the UE.

The AUSF stores the KAUSF and can compute a hash of the received XRES and sends it along the AUTH Token to the SEAF in an authentication response message. Then the SEAF stores the HXRES and sends an authentication request to the UE which contains the AUTH Token which the UE can validate by using the key of the network upon successful validation the UE proceeds with the computation of a RES token that can be validated by the SEAF and forwarded to the AUSF for taking the final decision. Finally, the AUSF computes the SEAF KEY (KSEAF) and sends it to the SEAF along with the SUPI if applicable. The SEAF proceeds further by deriving the AMF KEY (KAMF) and sending it to the AMF which in its turn derives the keys required for the integrity and confidentiality connections between the AMF and the UE and it also derives the KgNB and sends it to the base station which's used by it to protect communications with the UE. The UE has the long-term key that can be used to derive the previously mentioned Keys and proceed with its communication with the network.

4.4 Dual Connectivity DC and 5G NSA:

Implementing 5G by mobile operators requires installing a whole new architecture to benefit from the 5G new capabilities represented in high throughput speeds and low latency communication. However, the migration process requires a lot of time and steps. Therefore, the NSA option offered a smooth transition for operators by using the already existing LTE infrastructure and modifying it a bit to offer users 5G capabilities. As mentioned before, the NSA deploys two access points, the legacy eNB handling all control signals and the new gNB delivering data to the UEs at high speeds. This's a concept known as Dual Connectivity. This approach alongside CUPS forms the main building blocks of 5G NSA.

DC can have many flavors and options in the way it can be implemented. Three pillars define which type of DC we're using, the core network (LTE or 5G), the master node, and the secondary node. In NE-DC for example, the master and secondary nodes are both gNBs and the core that they connect to is the new 5G core. Whereas in Multiple Radio Access Technology Dual Connectivity, the nodes are different from each other with variations in the core that they connect to.

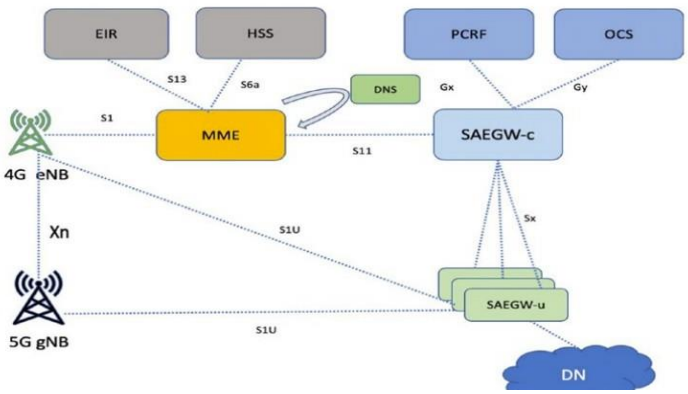


Figure 4-5: DU and 5G NSA implementation [16].

Chapter 5:
Deploying a SA 5G Network using Open-Source Software and COTS Hardware.

In the previous chapters, we got familiar with the main concepts, entities, and features in 5G, and while we've not mentioned everything we did focus on what would be beneficial for us to understand what has been implemented during the LAB.

However, in this chapter, we will discuss and explain the steps and procedures taken to deploy a SA 5G network by simply relying on easily accessible software and purchasable hardware (even though it's not always cheap). The architecture of the project from a simple perspective is built by using Open5GS implementing the 5G Core and entities and using the srsRAN Project that deployed the gNB of this network supported by an SDR and a powerful PC. Although this approach does have some limitations, it helps understand the service-based architecture of 5G and explore the different entities so we will be focusing on testing this stack and validating its reliability in a small radius to avoid any frequency regulation issues. This can include observing the different UEs' behaviors when communicating with the network, their throughput, and setting up their parameters to attach to the core successfully. Moreover, considering that we will be deploying this network using Radio Antennas there will be some considerations to the SDRs used and their operation limitations and capabilities. Finally, we will be analyzing some traffic movement between the gNB and the core to understand how the packets are being transported and observe the authentication process involved.

5.1 Overview of srsRAN Project and Open5GS:

The first part of the stack used in the LAB is srsRAN, it's simply an open-source project implementing the NR gNB with almost all of its features. So it does include the functional split we discussed earlier shown in Figure 1-4 with all the necessary protocols and interfaces making the gNB operate in a way that has a Centralized Unit (CU), the Decentralized Unit, and the Radio Unit. These splits are further separated into the User-Plane and the Control-Plane part for the CU and the DU.

This project has been tested by srsRAN with different devices working based on different frequency bands and transmission modes (TDD or FDD) and in *Table 2* we can see some of them.

It's clear that different devices are working based on different configurations and they can report different behaviors. So from the table, we can notice that some devices are not working in TDD mode for example and some tend to operate in a specific frequency band only with certain channel sizes. Not to mention the uncertainty regarding the SDR type or the possibility of needing to activate a hidden option to activate NR only option. All of these issues can be related to the different specs and capabilities of the devices depending on what bands they support and so on.

Make	Model	CPU	FDD	TDD	Configuration	SDR	Notes
OnePlus	Nord 5G	Snapdragon 765G	n3 - 5, 10, 20 MHz	n78 - 20, 40, 80, 100 MHz	IMSI 00101 gNB PLMN 90170	B210 3, Leo Bodnar GPSDO N310 4, External Clock	Seems to not like Test PLMN Has been tested in the shielding box and on a desk
Xiaomi	11 Lite 5G NE	Snapdragon 778G 5G	n3 - 10, 20 MHz	n78 - 20 MHz	gNB PLMN: 00101	B210, Leo Bodnar GPSDO	Test SIM: Sysmocom SJA2 5 Activate hidden SA option by dialing #*#726633#*#
OnePlus	8T	Snapdragon 865 5G	NA	n78 - 20 MHz	IMSI 00101 gNB PLMN 90170	B210, Leo Bodnar GPSDO	Requires use of the "roaming hack" Requires GPSDO to keep a stable connection.
Telit	FN990A28	Snapdragon X62	n7 - 5, 10, 20 MHz n71 - 5, 10, 20 MHz	n78 - 20, 30 MHz	IMSI 00101, 99970 gNB PLMN: 00101, 99970	USRP B210	Test SIM: Sysmocom SJS1, SJA2 (must disable service 124) For Symocom Breakout Board requires SIMIN to be "Active High" MS operation mode: normal or type_approval (both work)

Table 2: Tested UEs with srsRAN Project [1].

The latest version of the srsRAN project currently offers support for both TDD/FDD along all frequency bands, it also allows for choosing between two subcarrier spacings, either 15 or 30 KHz, All physical channels, all RRC procedures excluding mobility functions, and finally all the MAC procedures excluding the power control functions. New features are promised for the upcoming versions with support for paging, multiple PDU sessions, and so on. However, currently, it still doesn't support operating in band 79 (both 15 and 30 kHz sub-carrier-spacing) and Band 34, 38, 39 (15 kHz sub-carrier-spacing).

On the other hand, the second most important element in the setup is Open5GS which deploys and implements the 5G SA core or even the 5G NSA/4G core by defining software components and entities using C programming language and it supports IPv6, Hanover procedures and AES encryption. However, with limitations in aspects like Roaming, Emergency calls support, and no possibility of interworking with EPC.

5.2 The Installation and Setup Procedures:

The setup we considered for deploying the SA Network consisted of installing both srsRAN and Open5GS on a PC with the condition of having enough processing power and running Linux as an operating system. So the desktop we used runs Ubuntu 22.04.2 LTS with 18 cores Core-i9 processor and 16 GB of RAM. Such high demands in the power aspect are required to run and process the gNB along the SDR and avoid dropping in the connection or suffering from underflows that might happen due to the sampling rate.

Furthermore, the choice of SDRs used in the experiments was limited to the available hardware so one option was the USRP B210 which's a Dual Channel Transceiver capable of operating in a frequency range of (70 MHz-6 GHz) with a channel width of 56 MHz able to work in TDD and FDD[18]. and another option was the USRP N300 operating in a frequency range of (10 MHz-6 GHz) with up to 100 MHz of instantaneous bandwidth also operating in both transmission modes [19].

Regarding the UE, the options for our experiments were testing the Google Pixel 7, Google Pixel 6, and the Xiaomi 11 5G all using a reprogrammable sysmocom SIM card.

This SIM card offers high flexibility as it allows the reprogramming of all parameters like the IMSI (The international mobile subscriber identity), MSISDN (Mobile Station Integrated Services Digital Network), and Ki (The Secret Authentication Key), and OP_C (Derived Operator Code) or OP (Operator Code) [20].

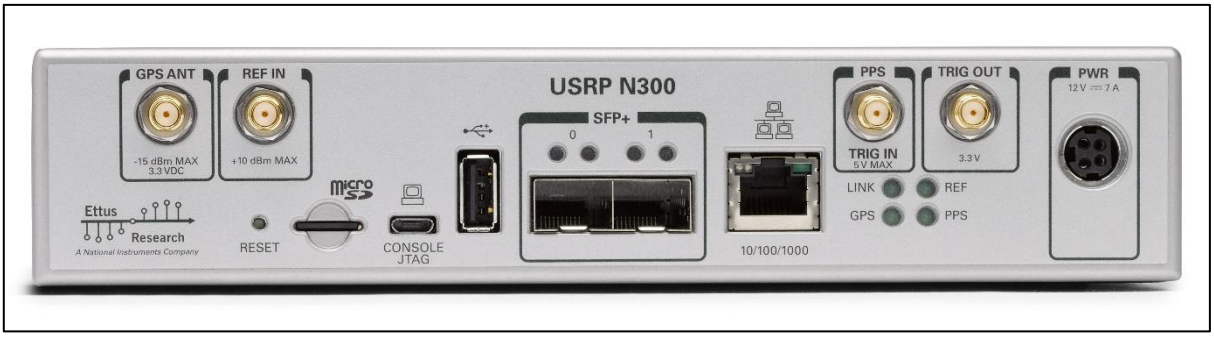


Figure 5-1: USRP N300 interface [19]

5.2.1 The Installation:

The installation process is pretty straightforward for either Open5GS or even srsRAN Project, it involves installing the dependencies and required libraries like MongoDB which we had an issue with compatibility in it regarding the Ubuntu version so we ended up downgrading to a lower release of MongoDB. In Figure 5-2 , we can take a simple look at the code used to import and install Open5GS for example.

```
$ sudo apt update
$ sudo apt install gnupg
$ curl -fsSL https://pgp.mongodb.com/server-6.0.asc | sudo gpg -o /usr/share/keyrings/mongodb-server-6.0.gpg --dearmor
```

Figure 5-2: Open5GS Installation.

In srsRAN, apart from installing the dependencies, we have to install the RF drivers so that we can make the system communicate with the SDRs which was in our case version 13.5.

And now after we clone and build the srsRAN Project repository and install the gNB we can run a Linux script that helps tune the system to deliver the most possible performance to handle the demanding computational load required by 5G. and as an additional tool we can install the WebUI of Open5GS which offer a friendly user interface to help add subscribers data to the core.

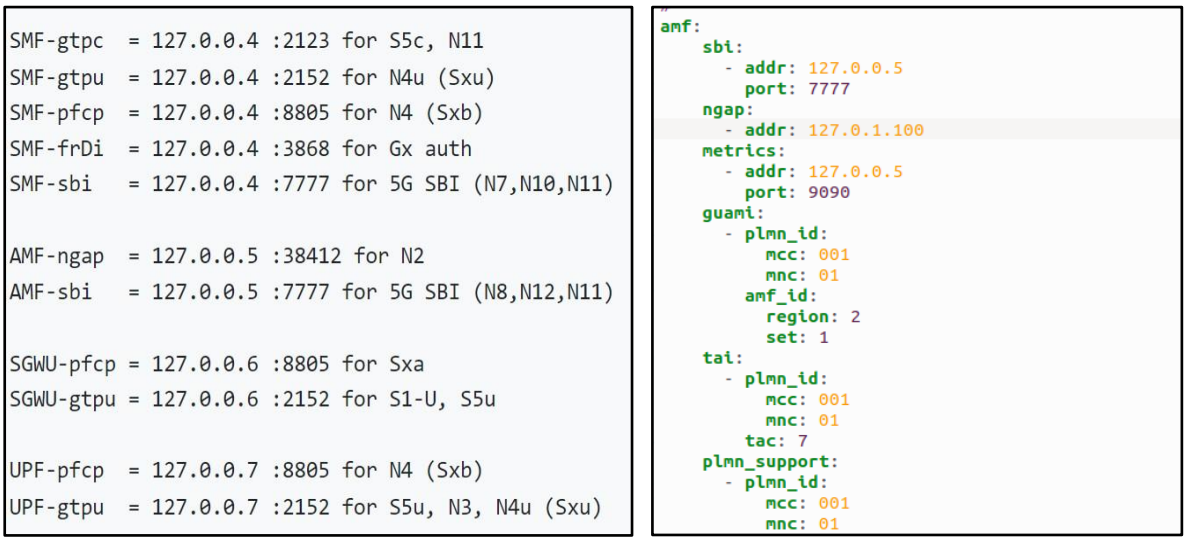
5.2.2 Configuring the parameters and entities of the network:

As we've already seen, 5G core is a service-based architecture which means that in order to achieve the functionality of the network, all entities must be able to communicate with the provider of the services that they need. So the first thing to do after we finished installing the stack is to configure the addresses of the network blocks so that they can see each other.

Starting from the 5G core, we need to specify the NGAP bind address of the AMF and the GTPU bind address of the UPF so that they connect and exchange data, this step is only required if we've installed the components on different machines. Otherwise, the core functions can communicate with one another and communicate over the loopback address space (127.0.0.X).

And another thing to edit is the PLMN and TAC values in both the gNB configuration file and the AMF so that the UE using the SIM we programmed can see the network and register itself in it. The values we used are the international test values for 5G which's 001/01 which basically indicates the values of the MCC (Mobile Country Code) and MNC (Mobile Network Code).

On the other hand, we have the TAC value which represents the Tracking Area Identifier which's a unique value to identify a cell in the core. The value of TAC can be optional to choose but it has to be compatible between the gNB and the AMF because otherwise it would mean we're setting up the gNB in a different area of the AMF which eventually means they can't communicate together.



```
SMF-gtpc  = 127.0.0.4 :2123 for S5c, N11
SMF-gtpu  = 127.0.0.4 :2152 for N4u (Sxu)
SMF-pfcp  = 127.0.0.4 :8805 for N4 (Sxb)
SMF-frDi  = 127.0.0.4 :3868 for Gx auth
SMF-sbi   = 127.0.0.4 :7777 for 5G SBI (N7,N10,N11)

AMF-ngap  = 127.0.0.5 :38412 for N2
AMF-sbi   = 127.0.0.5 :7777 for 5G SBI (N8,N12,N11)

SGWU-pfcp = 127.0.0.6 :8805 for Sxa
SGWU-gtpu = 127.0.0.6 :2152 for S1-U, S5u

UPF-pfcp  = 127.0.0.7 :8805 for N4 (Sxb)
UPF-gtpu  = 127.0.0.7 :2152 for S5u, N3, N4u (Sxu)
```

```
amf:
  sbi:
    - addr: 127.0.0.5
      port: 7777
  ngap:
    - addr: 127.0.1.100
  metrics:
    - addr: 127.0.0.5
      port: 9090
  guami:
    - plmn_id:
        mcc: 001
        mnc: 01
      amf_id:
        region: 2
        set: 1
  tai:
    - plmn_id:
        mcc: 001
        mnc: 01
      tac: 7
  plmn_support:
    - plmn_id:
        mcc: 001
        mnc: 01
```

Figure 5-3: The Default Core Functions address over the same machine and the modified AMF's file.

As we can see in Figure 5-3, the different interfaces of each function have a different binding address. we can also see the configuration file of the AMF specifying the address of the NGAP protocol interface and the values of the MCC and MNC.

```
amf:
  addr: 127.0.1.100                                  # The address or hostname of the AMF.
  bind_addr: 127.0.0.1                               # A local IP that the gNB binds to for traffic from the AMF.

rf_driver:
  device_driver: uhd                # The RF driver name.
  device_args: type=n3xx            # Optionally pass arguments to the selected RF driver.
  clock: external                   # Specify the clock source used by the RF.
  sync: external                    # Specify the sync source used by the RF.
  srate: 30.72                      # RF sample rate might need to be adjusted according to selected bandwidth.
  tx_gain: 61                       # Transmit gain of the RF might need to adjusted to the given situation.
  rx_gain: 35                       # Receive gain of the RF might need to adjusted to the given situation.

cell_cfg:
  dl_arfcn: 368640                  # ARFCN of the downlink carrier (center frequency).
  band: 3                           # The NR band.
  channel_bandwidth_MHz: 20         # Bandwith in MHz. Number of PRBs will be automatically derived.
  common_scs: 15                    # Subcarrier spacing in kHz used for data.
  plmn: "00101"                     # PLMN broadcasted by the gNB.
  tac: 7                            # Tracking area code (needs to match the core configuration).
```

Figure 5-4: gNB Configuration file.

And of course, the procedure has to be done in the UPF configuration file in case we decided not to rely on the default IP addresses. Furthermore, in the srsRAN project folder, we can find the gNB configuration file and also modify it to connect to the IP address of the AMF, we can also change other parameters related to the SDR, so for example we can specify the sampling

rate, frequency band, channel size, antenna gain and so on. We also must be careful to change the TAC value in this file to match the one specified in the core settings, in our case “7”. Now, finally, In order to bridge the UPF to the internet we had to enable IP forwarding and add NAT rules to the IP Tables.

5.2.3 Registering the Subscriber Information:

Now after we managed to configure the main entities in the core and the gNB, it’s time to add UEs to the network. However, we need to program the SIM card so that it has the compatible MCC/MNC which in turn requires modifying the IMSI.

The process of adding users to the network can be done as explained before through the WebUI which will save the credentials and user’s parameters in addition to the APN and so on in the UDR as seen in Figure 5-5.

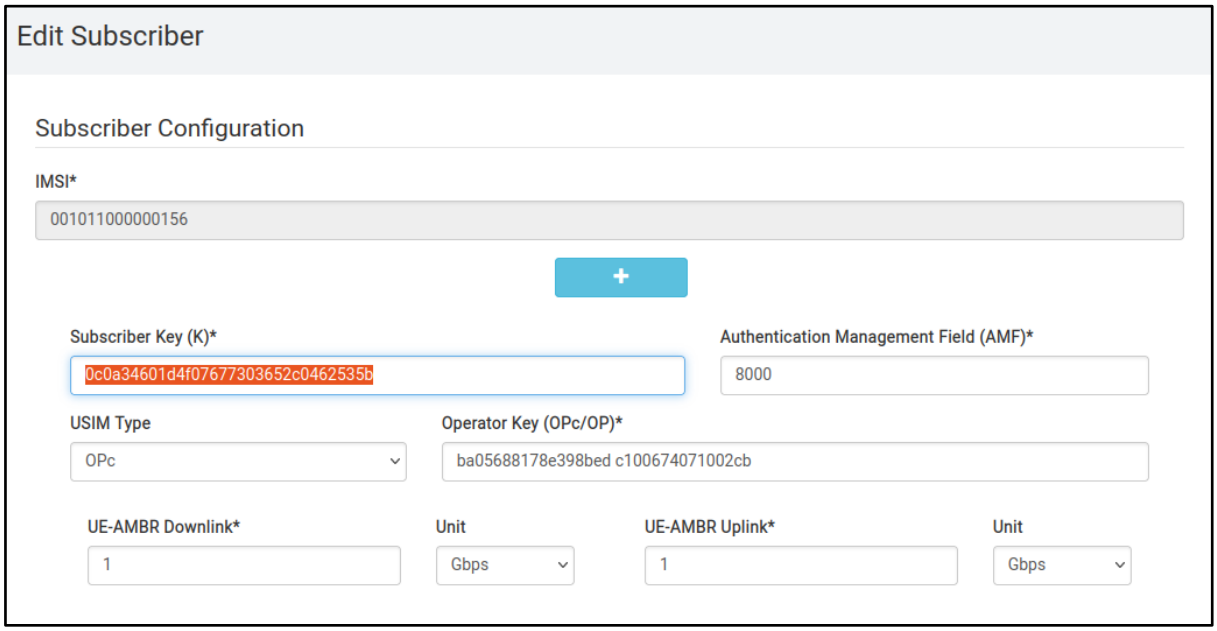


Figure 5-5: WebView Portal of Open5GS.

On the other hand, the device that we use has to support 5G SA at first, then we can add an Access Point name compatible with the one previously configured with the Open5GS WebUI to the phone’s list and preferably force the device to use the “NR only” option.

5.3 Testing the Network:

5.3.1 The tested configurations and their outcome:

After the successful configuration of the network and its entities, we can proceed with attempting to connect the UEs to the core. The gNB folder in the srsRAN Project provides several configuration files to use and they can be different in terms of operation, so we can have files for running the gNB in either TDD or FDD mode, different frequency bands, and channel sizes etc.

The first attempt we tried was to use the Xiaomi 11 lite 5G with USRP B210 operating in TDD mode. The phone was simply incapable of discovering the network either by manual or automatic search modes and changing the frequency band along with forcing “NR Only” didn’t give back any positive results. So as a second attempt, we switched to using the USRP N300

in FDD mode in the frequency band N3, this setup allowed the device to find and see the network successfully and it was able to complete the registration phase and start to exchange data with the UPF only to crash and lose connectivity after that in a few seconds. So while the Xiamoi was able to finally see the network, it couldn't hold a stable connection for a long time and data browsing wasn't even an option due to the connection drop issue. It's safe to say that this phone was tested before by connecting it to srsRAN LTE and I didn't suffer from any problems reporting high throughput speeds of approximately close to 50 Mbits/s using USRP B210 in TDD mode.

Due to the limited results we obtained, we had to proceed with testing another phone especially since it was already reported in Table 5-1 that there could be different behaviors and feedbacks between different setups. The second experiment involved using the Pixel 7 which also didn't report any successful feedback in discovering the network in the first place when considered in USRP B210 in TDD mode. on the other hand, the phone did manage to capture the signal of the network by performing a manual search but any attempts to connect or register the phone were unsuccessful whereas the attempts made with Pixel 6 couldn't even discover the network in any case. As a final attempt to make srsRAN work, we added an external clock source to our setup which means an added cost to an already expensive setup. The external clock should offer more precise synchronization between the Tx and Rx which in turn would lead to a more robust communication.

By first running the tests using USRP B210 in TDD mode, we noticed no difference. However, surprisingly, the Pixel 6 was able to find and successfully connect to the network when we switched to USRP N300 in FDD and more specifically in band N3 and 15KHz of subcarrier spacing along 20MHz channel width. This configuration gave a positive outcome and allowed the Pixel 6 to exchange data with the internet. Whereas the rest of the devices were still unlucky in managing to at least find the network. Although we did find a working configuration and setup with the Pixel 6. Regarding the performance issues, the Pixel 6 was able to connect, register, and exchange data successfully with the network. However, we noticed that the phone loses coverage and drops the connection entirely after a few minutes "which's much longer than Xiaomi" and it wasn't possible to reestablish the connection unless we switch off and on the gNB and put the phone to airplane mode and back again which means the although this setup was working, it wasn't so stable after all. However, Another issue we noticed during the experiment was the throughput data rates. The same mentioned devices did achieve very acceptable speeds over srsRAN LTE using the same hardware including the PC and the SDRs. However, the Pixel 6 on srsRAN Project didn't achieve the throughput we were hoping to see from a 5G network connectivity, the speeds did reach 8 Mbits/s at most and in some conditions fluctuated between 1-6 Mbps. Considering an NR connectivity, these results were unsatisfactory. And even any attempts to change some parameters like sampling rate, and channel size didn't enhance the communication. In fact, choosing a 5MHz channel did lower the throughput to almost none as seen in Figure 5-6, and going for the higher sampling rate exhausted the system and cause overflows in the transmissions.

```
Cell pci=1, bw=5 MHz, dl_arfcn=368640 (n3), dl_freq=1843.2 MHz, dl_ssb_arfcn=368670, ul_freq=1748.2 MHz

==== gNodeB started ===
Type <t> to view trace
t

            -------------DL----------------|------------------UL--------------------
 pci rnti  cqi  mcs  brate   ok  nok  (%) | pusch  mcs  brate   ok  nok  (%)    bsr
   1 4601   11    0      0    0    0   0% |   n/a    0      0    0    0   0%    0.0
   1 4602    5    6   4.1k    7    0   0% |   9.0   11    23k    4    2  33%    0.0
   1 4602    5    0      0    0    0   0% |   n/a    0      0    0    0   0%    0.0
   1 4602    5    0      0    0    0   0% |   9.6   14   4.2k    1    0   0%    0.0
   1 4602    5    0      0    0    0   0% |   n/a    0      0    0    0   0%    0.0
   1 4602    5    0      0    0    0   0% |   n/a    0      0    0    0   0%    0.0
```

Figure 5-6: data exchange with 5MHz channel width in srsRAN.

Alternatively, we wanted to see if changing the 5G gNB with OAI's (Open Air Interface) solution which's also an Open-source software implementation for the radio interface of NR. We did only consider the radio part of OAI and didn't install their 5G core solution which in theory shall be more compatible. As a result, we did test OAI radio with Open5GS again relying on the same phones with the same setup we were implementing and operating the gNB and the Open5GS on different computers and configuring the IP addresses as explained before.

OAI does offer several configuration options different in frequency bands, and channel sizes but also with more advanced features like MIMO capabilities which for sure require more computational power and more antennas. We did try the configuration files associated with the SDR we had at the LAB as reported before and we were lucky to find a working setup that was in the frequency band of 78 with a subcarrier spacing of 30 KHz compatible with the N300 SDR and operating in TDD mode. The results obtained from this experiment were astonishing.

Firstly, we were able to connect two phones successfully to the network, both the Xiaomi and the Pixel 6. Moreover, the connection was very stable during idle mode and during data exchange unless we moved a bit far from the antenna range which we kept relatively small for regulation issues. And finally, the throughput was drastically higher than the one obtained in the srsRAN solution as we were able to measure very high speeds on both phones reaching approximately 130 Mbps in the downlink and limited to 10 Mbps in the uplink. We used Network Signal Guru for the analysis of the throughput and the cell configuration. Therefore, we registered the following figures:

Figure 5-7:Cell configuration (Left) and Analysis about the resource blocks, QAM (Right).

As indicated in Figure 5-7 the operation mode was TDD with N78 and a numerology of 30 KHz. Whereas the reported bandwidth appears to be 40 Mhz because as we mentioned in the previous chapters NR supports a dynamic approach to the channel size depending on the transmission medium condition. We can also see on the right side of the figure that allocated resources for the UE in the uplink and the downlink with analysis about the QAM level used in each situation. So we can see in the downlink direction a higher order is used varying mostly between 64 QAM at the beginning of the reception and 256 QAM for most of the following period, whereas the uplink situation does report high order but it sticks mostly to 16 QAM.

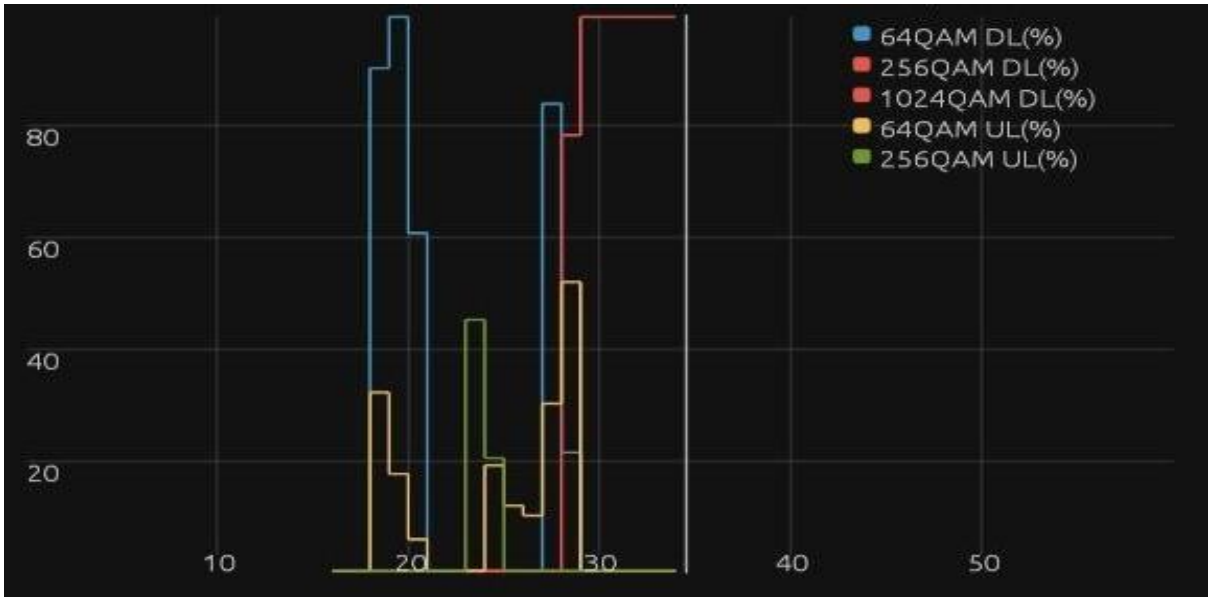


Figure 5-8: QAM analysis.

In Figure 5-9 we can again see the modulation scheme used in addition to the measurement speeds in each direction.

Figure 5-9: Downlink and Uplink analysis.

Moreover, we can report in Figure 5-10 the average throughput obtained from each phone using Speedtest for Android.

Figure 5-10: Throughput of different devices.

5.4 Analyzing the Network's Traffic:

The aim of this step is to dissect and analyze the traffic exchanged across the communication channels in the network we deployed, more specifically the packets and signals related to the initial connectivity and registration. And finally, we can also take a look at the security aspect that's being used.

5.4.1 Initial Connection Setup:

The exchange traffic that we focused on was between the most significant entities in the core, in Figure 5-11 which was obtained by analyzing the traffic via using Wireshark we can see an overview of the exchanged messages with the AMF distinguished with an IP address of (127.0.1.100) and the gNB with an IP address of (127.0.0.1).

```
127.0.0.1       127.0.1.100     SCTP     114 INIT
127.0.1.100     127.0.0.1       SCTP     338 INIT_ACK
127.0.0.1       127.0.1.100     SCTP     310 COOKIE_ECHO
127.0.1.100     127.0.0.1       SCTP      50 COOKIE_ACK
127.0.0.1       127.0.1.100     NGAP     118 NGSetupRequest
127.0.1.100     127.0.0.1       SCTP      62 SACK (Ack=0, Arwnd=106441)
127.0.1.100     127.0.0.1       NGAP     118 NGSetupResponse
127.0.0.1       127.0.1.100     SCTP      62 SACK (Ack=0, Arwnd=31249946)
```

Figure 5-11: traffic exchange between the AMF and the gNB.

We can notice that at first, the gNB starts the process of establishing the SCTP connection with the AMF over the N2 interface, the initialization starts by sending an INIT SCTP packet that contains INIT Chunk that includes the IP addresses that are used by the gNB in addition to the number of inbound streams that can be received from the other endpoint and the outbound streams that can be sent at maximum. The AMF replies to that packet with an acknowledgment message containing an INIT_ACK chunk that also includes the used IP addresses and streams.

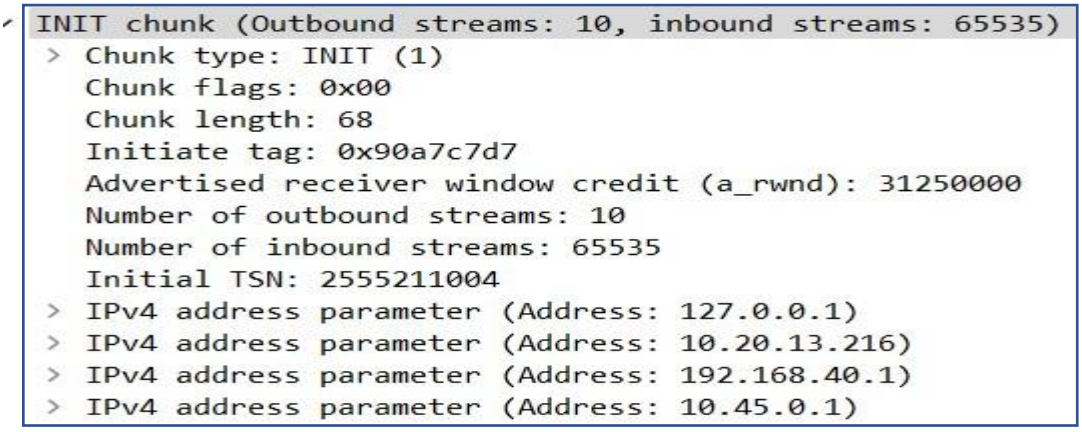


Figure 5-12: INIT chuck related to the gNB.

Then the establishment is done by sending the COOKIE ECHO/COOKIE ACK messages. From this point on, the endpoints can exchange SCTP frames that contain DATA chuck in the sending endpoint and the receiver acknowledges with a frame containing a SACK chunk.

Now the NGSetup_Request and NGSetup_Response are used between the two entities to establish the NGAP connection that's used to handle configuration updates, UE context transfer, and PDU session management. It's also used later to convey downlink and uplink NAS messages as a payload. The Setup request sent by the gNB contains the PLMN identity specifying the MNC/MCC, the gNB ID, and the supported TAC list as seen in Figure 5-13.

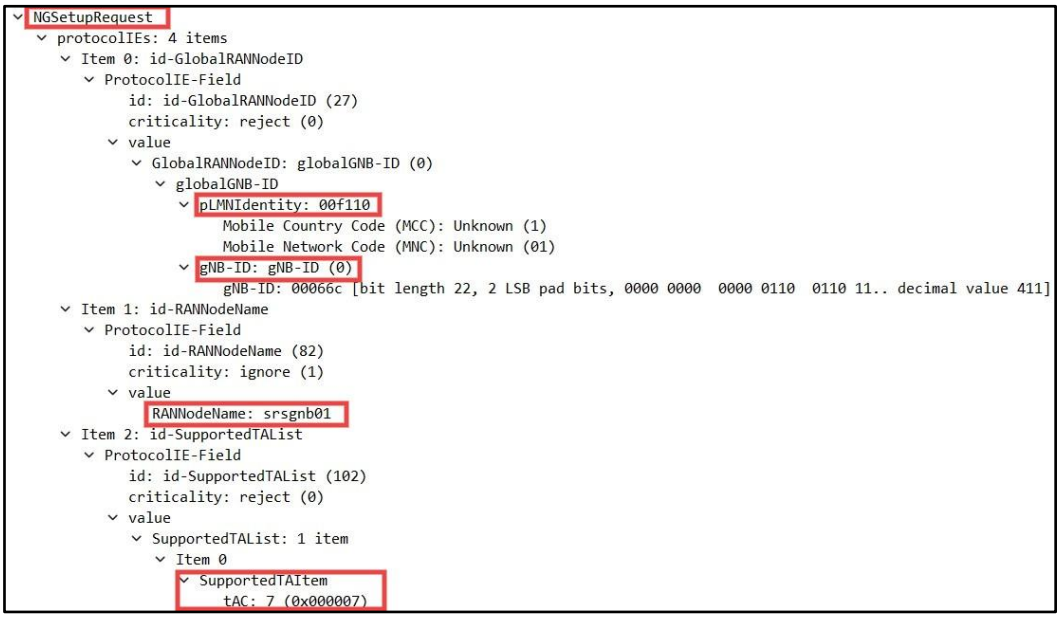


Figure 5-13: NGAP's Setup Request.

On the other hand, the NGAP_Response indicates a successful communication and includes the AMF's name and ID along the serving ID which indicates the AMF region, set as mentioned and discussed in the AMF description in Chapter 1. It also includes the AMF capacity and the supported PLMN list.

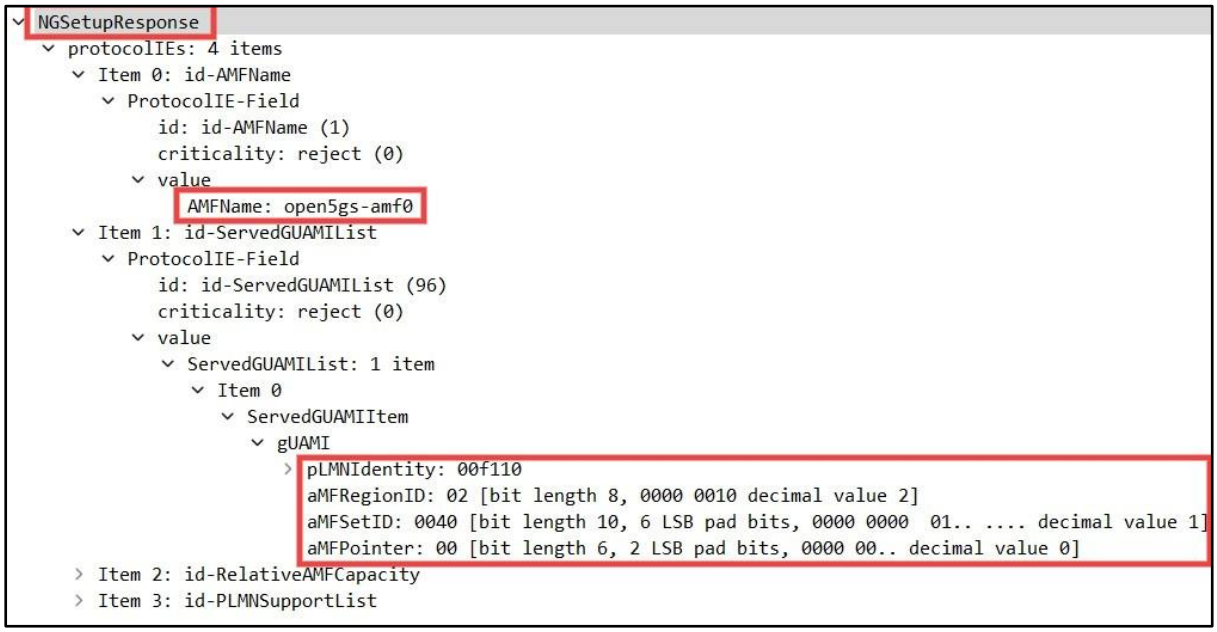


Figure 5-14: NGAP's Setup Response.

And from this point on, the UE can send a registration request to the AMF.

This request includes many parameters including the NAS-PDU ID and the supported security algorithms by the UE. It also includes the MNC/MCC as usual but most importantly this time the UE has to declare its identity to the network and this includes sending the IMSI to the AMF which can be done in several ways as we will discuss later.

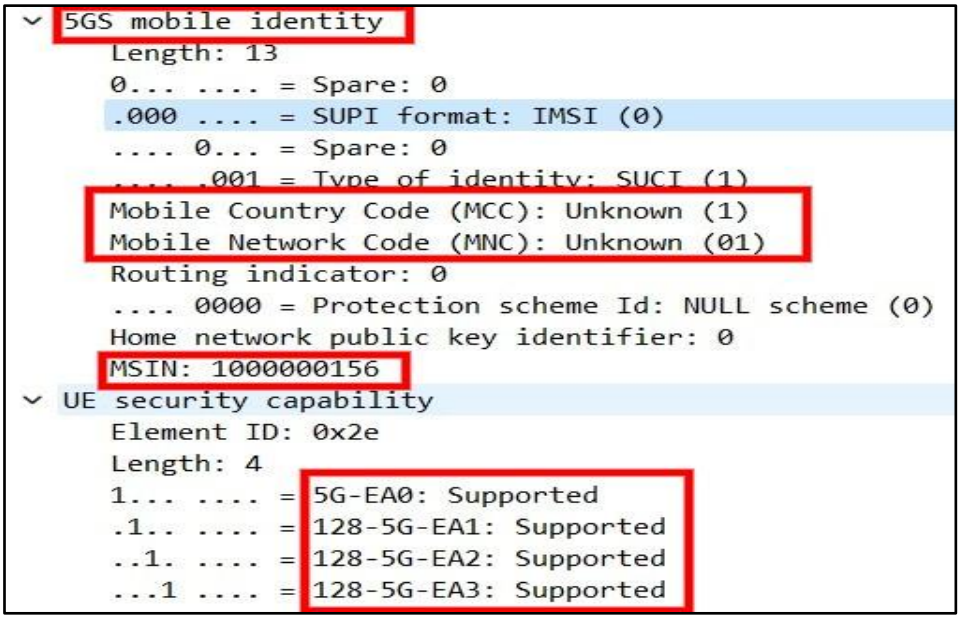


Figure 5-15: Registration request parameters.

The next message will be sent from the AMF towards the UE and it includes an authentication challenge that the UE has to verify its authenticity using the pre-shared secret key Ki and if the value generated by the UE matches the one sent by the network then it considers it to be legitimate and then sends an authentication response to the AMF with another challenge so the AMF by its turn can now verify the UE. After that, a security mode command message can be sent by the AMF to derive the security keys to eventually have a secure and protected communication channel.

127.0.0.1	127.0.1.100	NGAP/N…	134 InitialUEMessage, Registration request
127.0.1.100	127.0.0.1	NGAP/N…	146 SACK (Ack=1, Arwnd=106496) , DownlinkNASTransport, Authentication request
127.0.0.1	127.0.1.100	NGAP/N…	142 SACK (Ack=1, Arwnd=31250000) , UplinkNASTransport, Authentication response
127.0.1.100	127.0.0.1	NGAP/N…	126 SACK (Ack=2, Arwnd=106496) , DownlinkNASTransport, Security mode command

Figure 5-17: Registration request flow.

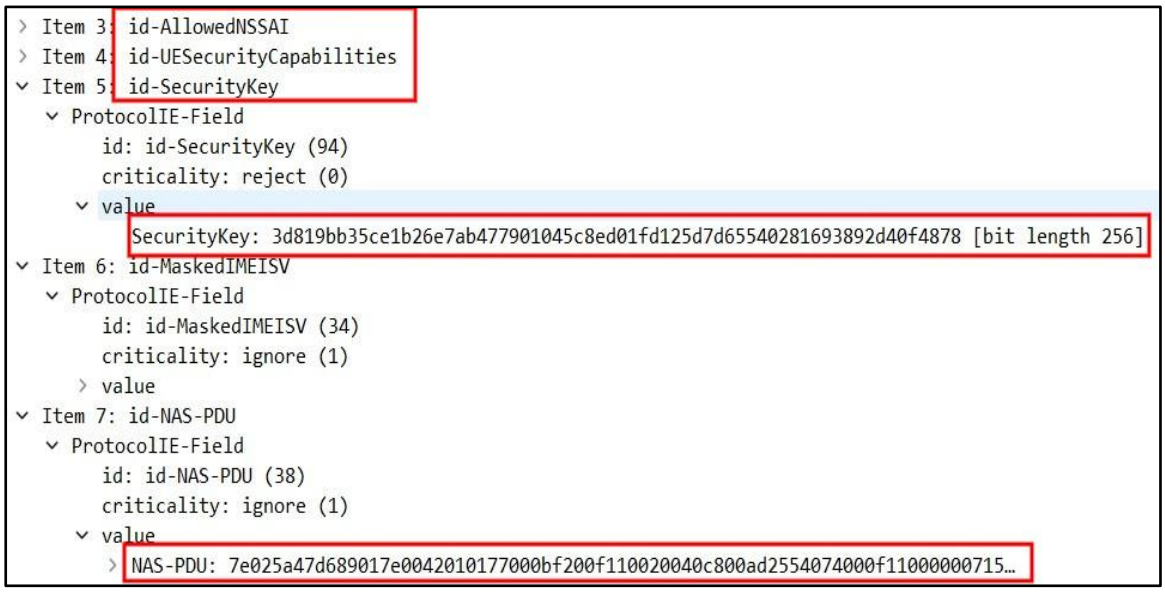


Figure 5-16: Initial Context's parameters.

Lastly, the AMF sends the UE an Initial Context _Setup_Request message that includes a key used to derive the subkeys used for protecting the RRC channel so that subsequent communication between the gNB and the UE is protected. and it also includes slice information, allowed mobility, and supported security functions along the PDU session context.

5.4.2 Security features in NR:

5.4.2.1 5G-AKA:

Like we have mentioned earlier a new authentication framework was introduced in NR, and it consists of the SEAF (Security Anchor Function) that's the middleman between the UE and the network, and The AUSF (Authentication Server Function) that's responsible for making decisions related to the UE authentication and finally the SIDF (The Subscription Identifier De-concealing Function) which deciphers the encrypted IMSI or the so-called SUCI (Subscription Concealed Identifier).

When the SEAF receives an authentication request from the UE, it has to include some form of identity which as we mentioned is the IMSI in the clear, whereas here the UE must either

send a SUCI or TMSI (Temporary Mobile Subscriber Identity) if it has already been assigned one. So if the SUCI is provided towards the AUSF, it will invoke the SIDF to obtain the IMSI which will be used to select the appropriate authentication for the user and after that, the process of validating the network and the UE starts.

5.4.2.2 5G-AKA and Open5GS:

What we were interested in during these experiments is to see whether such procedures are taken during the initial access traffic in our deployed network and verify if Open5GS supports such an authentication technique. In fact, as seen in Figure 5-14, the type 5G identity appears to be SUCI but in reality, we don't see any ciphered values but the real IMSI of our UE. This's happening because some modifications need to be performed in order to activate the SUCI concealment. Firstly, we need to uncomment the security schemes in the UDM configuration file that describe the combination of the public/private keys that can be used where each scheme is indicated by an index that the UE uses to tell the home network which private key to use to decipher the SUCI. On the other hand, the SUCI security features are already active in the SIM but we need to program it in a way that specifies these combinations of keys, more specifically the Public keys of the network that can be used.

After executing the previous steps we tied running the experiment again to see if the concealment will be used this or if the IMSI will be sent in the clear like before and as expected

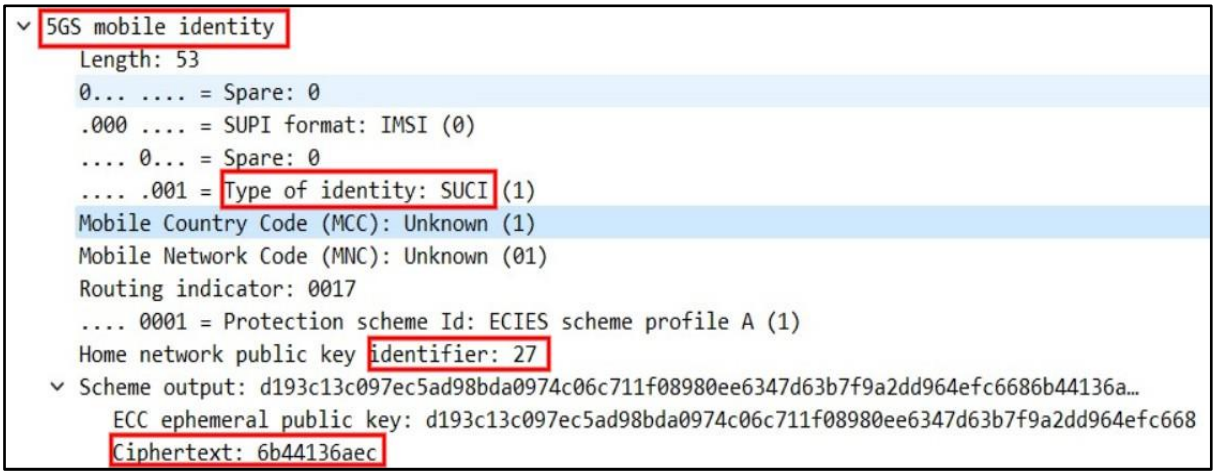


Figure 5-18: SUCI concealment.

we were able to verify that indeed this step is proven to support the new authentication method of NR as seen in Figure 5-17, now the 5G identity field includes a cipher text value and the UE is indicating the protection scheme index value used by its side. Moreover, when we attempt to reconnect the phone again without resetting the network state we can also verify that the UE will have a TMSI value for the upcoming registration requests instead of the SUCI, so it's only used during the first registration request and after that a frequently changed TMSI shall be more than secure for later interactions when the network can already identify the UE.

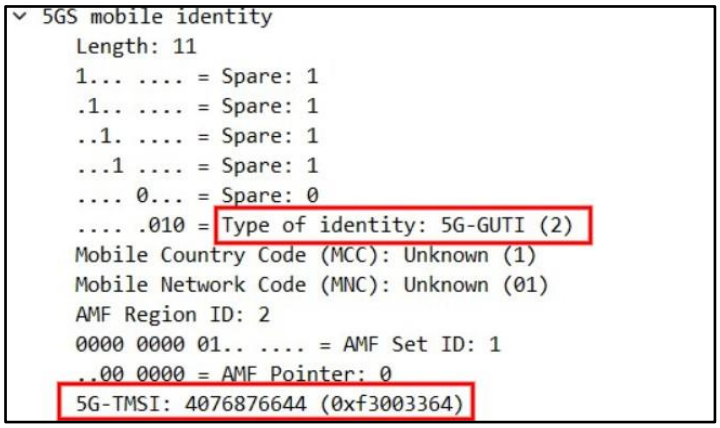


Figure 5-19: TMSI.

After running several iterations of tests, we also noticed that the value of the SUCI is changing with each registration request and it's never the same, such behavior is mainly related to the ciphering techniques used to conceal the IMSI or the so called SUPI (Subscription Permanent Identifier). While 5G relies on Symmetric-Key algorithms in most of its security parts, it does use Asymmetric-Key encryption when generating the SUCI. The scheme that's being used is ECIES (Elliptic Curve Integrated Encryption Scheme) which can operate in two profiles that define different parameters for example the Diffie-Hellman primitive. Furthermore, the main element contributing to the variation in the SUCI cipher text is related to the encryption steps executed at the UE side, we mentioned before that the Public key of the home network is stored in the UESIM whereas the Private one is in the UDM function, these are the main keys of the network and never change but the UE generates a new pair of Public/Private Keys more specifically ephemeral ones that do change with each iteration and registration request so they're never used again. To understand this a bit more, we can explain how the Curve25519 works, we define a curve and choose a point on it and then we choose a random large number N (this's the private key). Using point addition we add our point to itself N times and finally, the X-coordinate is the Public-key.

The fact that the key is 256 bits long makes it almost impossible for someone to guess the same large number N we choose (2^256=10^77) and the chance of repeating the same key pair would be minimum. Now the Private ephemeral key and the network public key are used to generate the Shared-Key using a Key Agreement function whereas the Public ephemeral key is sent to the network and also used along with The Shared-Key to produce the AES-Key which's our encryption key and a MAC Key is generated for authentication purposes.

On the other hand, the UDM can generate the same Shared Key by using the network's private key and the ephemeral Public key sent to it and then the process of generating the AES key would be exactly the same. In fact, we can indeed verify the ephemeral key derivation by looking at the packets shown in Figure 5-18 where we can see that the UE sends the Public ephemeral key along the SUCI.

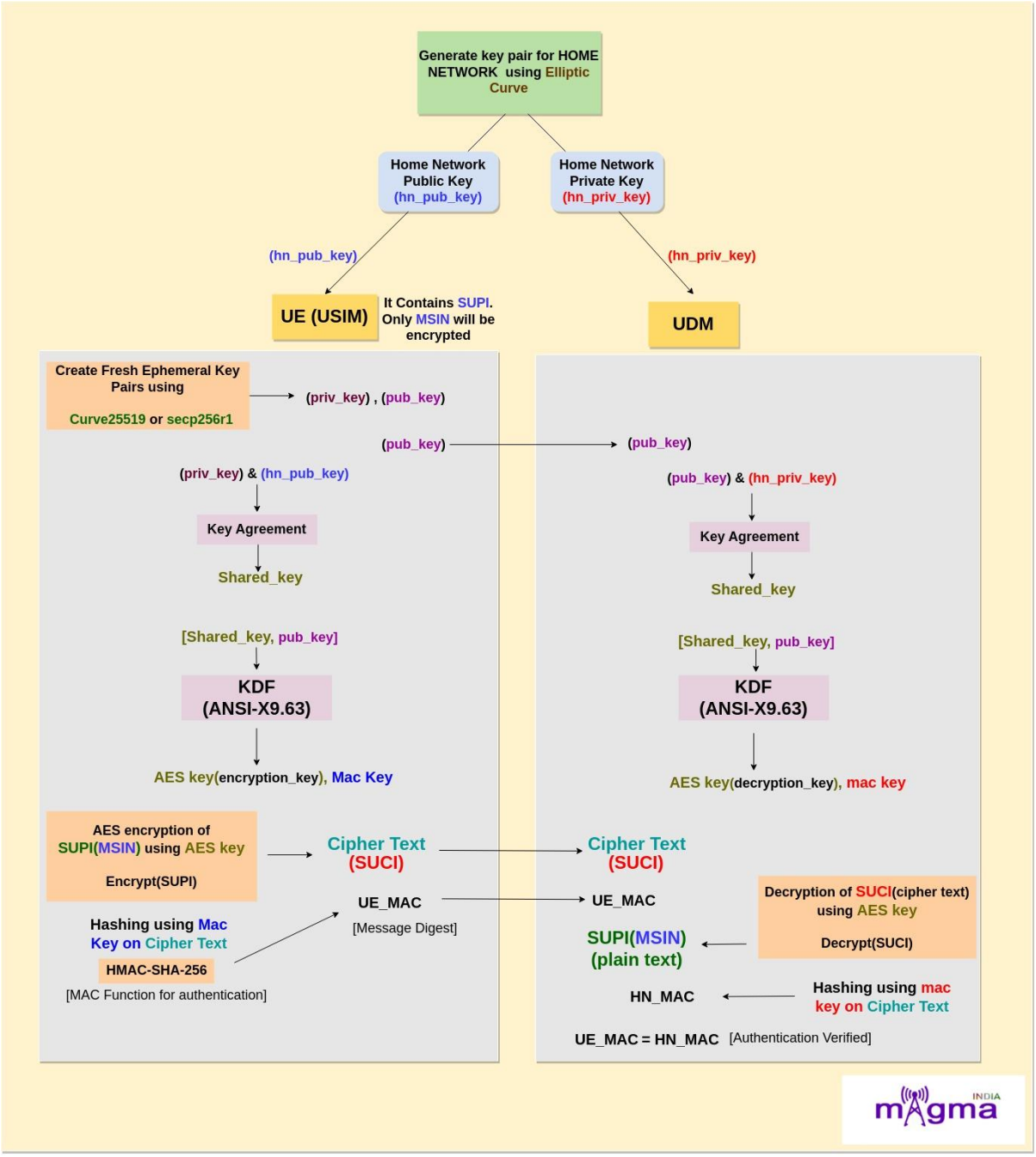


Figure 5-20: MSIN encryption steps [21].

5.5 Conclusions

Finally, after doing these experiments and testing the software stack provided by OPEN5GS, srsRAN and OIA we were able to validate the performance and capabilities of such deployment techniques. The power of these experiments is mainly represented in the simplicity of the application and deployments to allow many users to test, understand and analyze 5G networks closely especially since almost everything is implemented in software that's also Open-Source which gives the opportunity to many people to contribute to these project and enhance their features for the upcoming versions so for example as we mentioned earlier new release can include more in-depth characteristics like Power consumption,4x4 MIMO for srsRAN, Handover and so on. On the other hand, deploying such systems can be frustrating to minor issues represented in compatibility issues, bugs, or instability so it's definitely not ready to be used professionally but rather as an initial step that might lead to using 5G systems indoors depending on the frequency restriction agreement. Furthermore, despite its simplicity as a solution, it still requires powerful hardware and very expensive SDRs to obtain somehow stable performance or even use the full potential of MIMO in the case of OAI.

However, at the end of the day being able to deploy such a system successfully would really be beneficiary to testing environments to study the behavior of the network, and analyze traffic more easily like our case where we did focus on the security aspect and so on.

M. Bashir Qanaa

Bibliography

[1] Bhatt, Ashutosh. "Difference between 2G and 3G Technology." *Engineers Garage*, https://www.engineersgarage.com/difference-between-2g-and-3g-technology/.

[2] *3GPP – The Mobile Broadband Standard*. (n.d.). 3GPP. https://www.3gpp.org/

[3] Amaral, T. B. D., Rosa, R. V., Moura, D. S., & Rothenberg, C. E. (2021). *An In-Kernel Solution Based on XDP for 5G UPF: Design, Prototype and Performance Evaluation*. https://doi.org/10.23919/cnsm52442.2021.9615553.

[4] 5G PPP Architecture Working Group. 5G Enhanced Overall System Architecture." 5G PPP View on 5G Architecture, *2019*, pp 20.

[5] Shetty, R.S. (2021) "5G Globally Unique Temporary Identifier," in *5G Mobile Core Network: Design, deployment, automation, and testing strategies*. New York, NY: Apress Media, p. 13.

[6] Zaidi, A., & Baldemair, R. (2017). *In the race to 5G, CP-OFDM triumphs!* ericsson.com. https://www.ericsson.com/en/blog/2017/5/in-the-race-to-5g-cp-ofdm-triumphs.

[7] Awave_Admin. (2022). 5G User Plane Function. *Enea*. https://www.enea.com/solutions/dpi-traffic-intelligence/networking/5g-user-plane-function.

[8] 3GPP TS 33.501, "Security architecture and procedures for 5G system".

[9] *Overview of Positioning in 5G New Radio*. (2019, August 1). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8877160.

[10] 5G | ShareTechnote. (n.d.). https://www.sharetechnote.com/html/5G/5G_RadioProtocolStackArchitecture.html.

[11] Dryjanski, M., PhD. (2020). Dual Connectivity – Practical Aspects. Grandmetric. https://www.grandmetric.com/dual-connectivity-practical-aspects/

[12] 5G NR Initial Access Procedure | 5G NR Random Access Procedure. (n.d.). https://www.rfwireless-world.com/5G/5G-NR-Initial-Access-Procedure.html.

[13] Zhenhua Li, Weiwei Wang, Christo Wilson, Jian Chen, Chen Qian, Taeho Jung, Lan Zhang, Kebin Liu, Xiangyang Li, and Yunhao Liu, "FBS-Radar: Uncovering Fake Base Stations at Scale in the Wild," Proceedings of the Internet Society Symposium on Network and Distributed System Security (NDSS) (February 2017).

[14] Byeongdo Hong, Sangwook Bae, and Yongdae Kim, "GUTI Reallocation Demystified: Cellular Location Tracking with Changing Temporary Identifier," Proceedings of the Internet Society Symposium on Network and Distributed System Security (NDSS) (February 2018).

[15] Professional, P. S. P. M. a. V. S. (2020, January 20). 5G Security (5G AKA Authentication) – 5G Resource Center Blogs. https://www.5gblogs.com/5g-security-5g-aka-authentication.

[16] Shetty, R.S. (2021) "5G NSA Design and Deployment Strategy" in *5G Mobile Core Network: Design, deployment, automation, and testing strategies*. New York, NY: Apress Media, p.110.

[17] COTS UEs — srsRAN Project documentation. (n.d.). https://docs.srsran.com/projects/project/en/latest/knowledge_base/source/cots_ues/source/index.html#tested-cots-ues.

[18] Ettus Research, a National Instruments Brand. (n.d.). USRP B210 USB Software Defined Radio (SDR) - Ettus Research. Ettus Research. https://www.ettus.com/all-products/ub210-kit.

[19] Ettus Research, a National Instruments Brand. (n.d.-b). USRP N300. Ettus Research. https://www.ettus.com/all-products/usrp-n300.

[20] sysmoISIM-SJA2 SIM + USIM + ISIM Card (10-pack) with ADM keys | sysmoISIM-SJA2-10p-adm. (n.d.). Sysmocom Webshop. https://shop.sysmocom.de/sysmoISIM-SJA2-SIM-USIM-ISIM-Card-10-pack-with-ADM-keys/sysmoISIM-SJA2-10p-adm.

[21] Koranga, A. (2022, November 5). ECIES Implementation in 5G Core: SUPI to SUCI conversioN. https://blog.magmaindia.org/2022/11/05/ecies-implementation-in-5g-core-supi-to-suci-conversion%EF%BF%BC/.