UNIVERSITY OF ALBERTA

Research on 5G core network slicing and its associated security issues

Capstone Project Proposal

Presented by
Peter Awedana Assorow
University of Alberta
Master of Science in Internetworking Edmonton, Canada

Supervisor
Sandeep Kaur

Contents

Table of Figures

List of Tables

Abstract

Network slicing is a key enabling technology of 5G deployment. With this technology, MNOs could flexibly and logically dedicate some of their core network resources to various businesses and enterprises and provide a specific quality of service and policies based on agreed terms. This ability will enable MNOs or CSPs to maximize their revenue and provide more optimized services to meet the varying needs of different market verticals or diverse user experiences.

But the catch is, before network slicing can be implemented fully, several troubling security challenges and vulnerabilities, some of which still need to be well investigated and explored yet need answers. This leaves much to be desired since information and data have to be protected from being compromised. What makes it more worrying is the virtualization of network functions and the shift to a cloud-based and service-based architecture exposing it to the internet, which, as we are aware, has several security implications that are graver than earlier-generation networks which relied on centralized hardware-based functions that provided security endpoints and checkpoints that were relatively easier to monitor. For key players in the network-slicing market to implement and maximize its benefits while gaining the confidence of businesses and consumers, many security and trust questions need to be identified, answered or resolved first.

This project seeks to explore the various security flaws or vulnerabilities of 5G core network slicing in both standalone and non-standalone implementations. To do that, we have to explain how mobile technologies evolved over the years with a focus on the 5G architecture. We then narrow down to the 5G core network and look at Network slicing and its benefits as we seek to understand the implementation of Network slicing while using a case study of how slices will interact with each other in a city like Leduc with various enterprise customers using Telus as an MNO delivering this service. Furthermore, we explore the various security flaws that 5G core network slicing is exposed to. Finally, we look at some meaningful security recommendations and solutions that, if well implemented, could help squash the doubts and insecurities surrounding the implementation of a full-scale 5G core network slicing.

Acknowledgement

I'm incredibly grateful to everyone who helped me complete this project.

Primarily, I would thank God for the strength and wisdom to complete this project successfully. In addition, I give special thanks to my mentor, Sandeep Kaur, under whose guidance I learned a lot about this project. Her suggestions and directions have been valuable in completing this project.

Also, I would thank the Program Director, Prof. Mike MacGregor, Shahnawaz Mir, Sharon Gannon and many others in our faculty for their continuous support and guidance throughout the MINT program.

Furthermore, I thank the University of Alberta for allowing me to study and access various resources and materials that helped me complete this project.

Finally, I thank my parents and friends who have helped me with their prayers, valuable advice and encouragement throughout my studies.

1 CHAPTER 1: A HISTORY OF THE EVOLUTION OF MOBILE TECHNOLOGIES FROM 1G to 5G

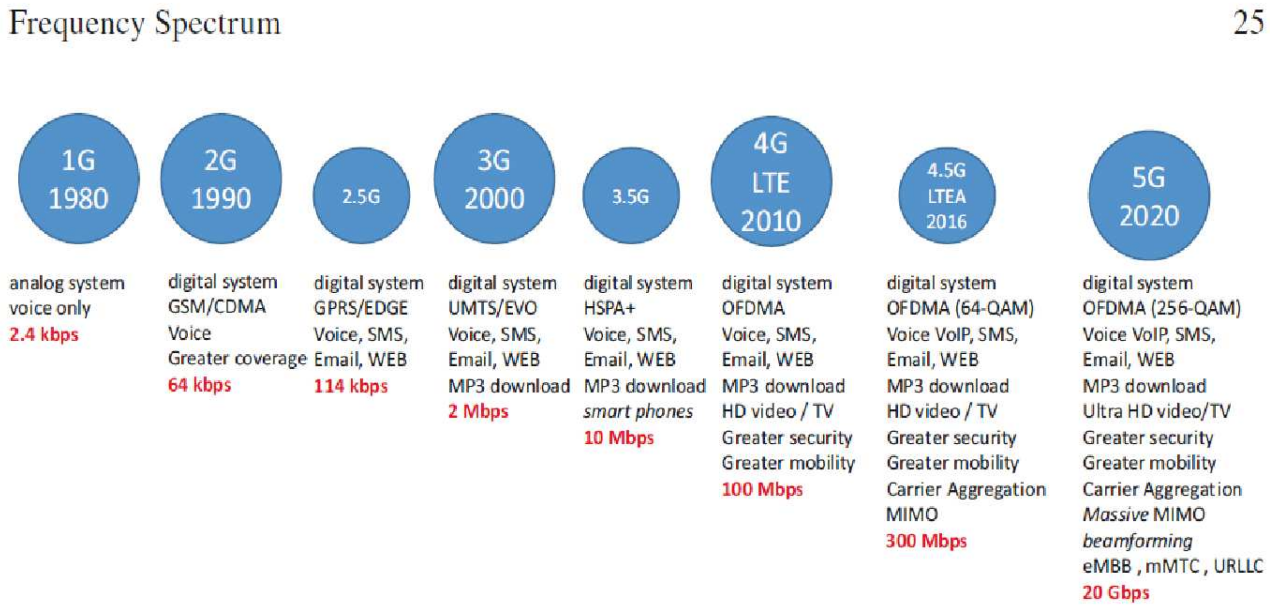


Figure 1-1 Evolution of mobile communication technologies [1, p. 25]

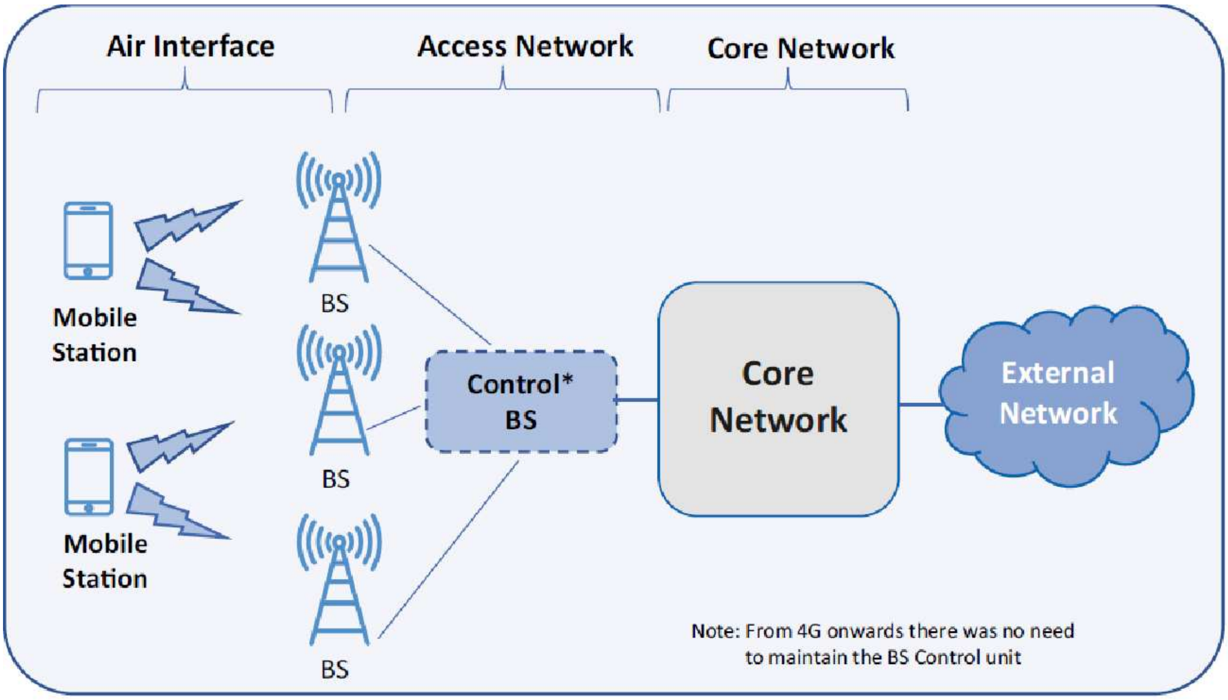


Figure 1-2. An overview of a basic mobile network ([1, p. 84]

The general concept of a network consists of three parts: The Air Interface, Access Network, and Core Network, as shown in Figure 1-2.

We also note that, from the *core,* it is possible to interconnect the system with other mobile phone networks, landline telephony, packet networks and the internet, integrating the cellular Network with the global **telecommunication system.** [1]

1.1 1G MOBILE COMMUNICATION SYSTEM

Mobile radio communications were introduced in the early 20th century. However, car-based telephones were initially tested in 1946 in Saint-Louis.

This system used a single large transmitter on a high-rise building. A single channel was used for sending and receiving, similar to a half-duplex system. To talk, the user pushed a button that enabled transmission and disabled reception. This became known as the push-to-talk system in 1950. To allow users to speak and listen simultaneously, Improved Mobile Telecommunication System (IMTS) was introduced in the 1960s. It used two channels, one for sending and the other for receiving, bringing telecommunication to full duplex mode.

In the 1970s, private companies started developing their systems to evolve existing ones. Those systems were, Analog Mobile Phone System (AMPS) used in America, the Total Access Communication System (TACS) and Nordic Mobile Telephone (NMT) used in parts of Europe, and the Japanese Total Access Communication System (JTACS) used in Japan and Hong Kong. These independently developed communication systems were known as 1st Generation communication. It was first introduced in 1982 by bell labs and is popularly known as Advanced Mobile Phone Systems (AMPS). The key idea was to divide geographical areas into cells, served by a base station, to implement frequency reuse. As a result, AMPS could support 5 to 10 times more users than IMTS could. [2]

The first-generation mobile communication system Used FDMA technology to cater to the limited capacity that it faced. Each user was given one channel out of the bandwidth available, divided into several orthogonal channels. This design resembled a highway divided into lanes, with each car assigned a particular route. This improved capacity only to a certain point. Since this technology was analog, it could only transmit voice signals with limited capacity. Challenges such as poor voice quality, unstable signals, limited coverage and several security flaws also hindered the growth and penetration of 1G. Further compounding the growth of 1G was its need for roaming capability since the communication standards of different countries needed to be more consistent. This pushed the development of second-generation mobile communication technology. [2]

A technology called Frequency Division Multiple Access (FDMA) was adopted and implemented to help increase capacity. FDMA divides the total bandwidth into multiple orthogonal channels. Each user occupies one channel, just as highways are divided into lanes, with each car driving on only one of the assigned lanes. This brought some speech involvement but simultaneously many challenges. Because this mobile communication technology used analog signal transmission, whose capacity was limited; and generally could only transmit voice signals. As a result, there were issues such as low voice quality, signal instability, insufficient coverage, poor security and vulnerability to interference. Furthermore, since the communication standards of different countries were inconsistent, the first-generation mobile communication could not perform “global roaming,” which greatly hindered the development of the first-generation mobile communication technology. Therefore, the second generation of mobile communication technology was on the horizon. [2]

1.1.1 The 1G System architecture

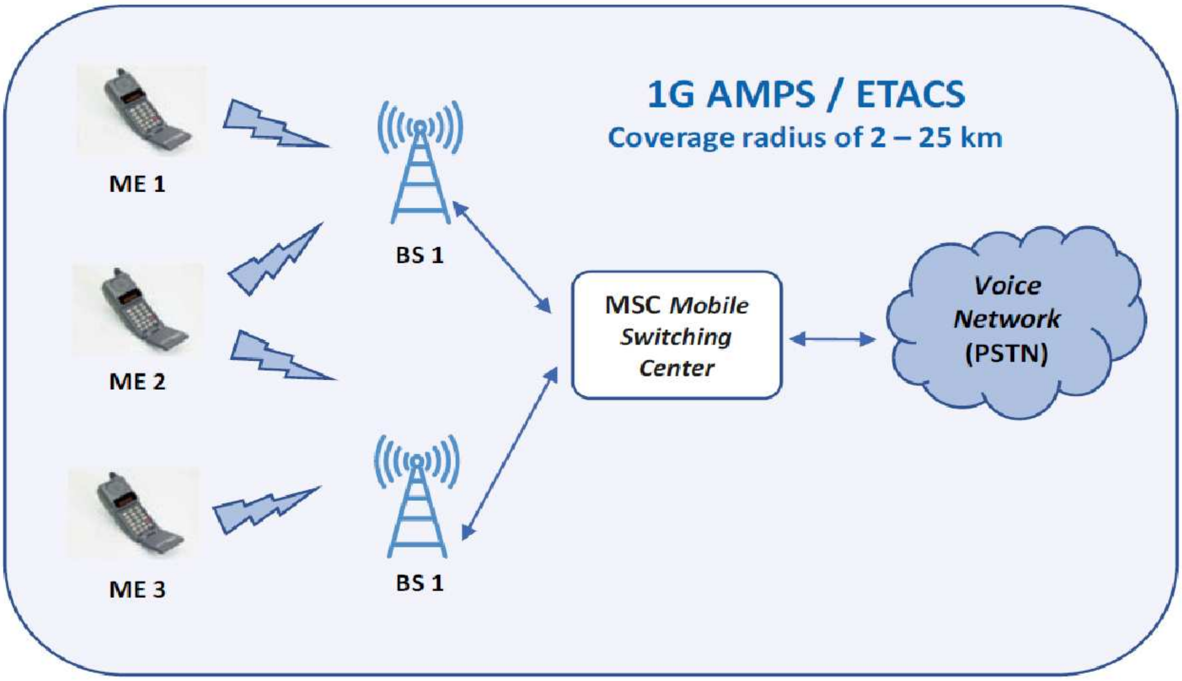


Figure 1-3. Simplified scheme of a 1G mobile network architecture. [1, p. 85]

Here*,* in Figure 1-3, **the Mobile Switching Center (MSC) (Mobile Switching Center**); receives all the traffic or signals from the base station and allows users to interface directly with the Public Switched Telephone Network (PSTN). This architecture was simply because, at the time, 1G could only support voice communication, and thus the core network was fused with the fixed-line Network, which made up the MSC Mobile Switching Center. [1]

1.1.2 Major security Flaws for 1st generation:

- No authentication was designed to identify a user's device uniquely;
- Cloning of phones was possible
- Attackers could eavesdrop on phone conversations
- Weak security on Air Interface
- Full analog mode of communication

1.2 2G MOBILE COMMUNICATION SYSTEM

Various individual telecommunication companies started working together under one umbrella, the European Telecommunications Standards Institute (ETSI) and developed the 2nd Generation system (2G) in 1988. The 2G system was developed to cater to the limitations and flaws of the 1G analog system. It was designed to provide higher-quality signals, higher data rates to support digital services, and greater capacity than what 1G could provide.

The digital modulation system allowed for voice compression through either time (GSM) or code (IS-95 CDMA) multiplexing.

Based on GSM standards, the 2nd Generation System was commercially launched in Finland in 1991. It could deliver data at a rate of 9.6Kbps. [2]

GSM also adopted TDMA (Time Division Multiple Address) to help alleviate issues with capacity and allow for full duplex communication. Here, radio frequencies are divided into time slots for different users to be assigned to each space. Compared with FDMA technology used in 1G, TDMA has the advantages of improved communication quality, good confidentiality and security with a larger system capacity. Still, it has to be precisely timed and synchronized to ensure efficient communication between mobile terminals and base stations, which is technically more complex.

These were the reasons for further developing fully digitalized, second-generation mobile communication systems like code division multiple access, CDMA-based IS-95 in North America.

Unlike GSM, which is deployed globally, CDMA's worldwide deployment is mainly concentrated in the US, South Korea and China. Compared with analog communication, digital communication makes up for the technical shortcomings of the analog communication era to a certain extent.

As time progressed, GSM (Global System for Mobile Communications) became the most popular 2G variant. With GSM came the separation of the terminal and subscriber data storage (SIM, subscriber identity module) for easier changing of the device, with consumers or users being able to keep their mobile telephone numbers. Enhanced services like the short message service (SMS) were also introduced, making communication more efficient. [3]

The 2G, however, needed to provide an internet access service and soon received an upgrade, and became known as 2.5G. The main innovation was the introduction of packet switching (PS – Packet Switched), the same transmission technique adopted by IP networks (TCP/IP – Transmission Control Protocol/Internet Protocol architecture). The 2.5G network was connected to a packet data network (PDN), but the connection to the circuit-switched Network was maintained for the voice channels. The separation of the voice service from the data service allowed for greater efficiency in processing the data service. This innovation received the acronym GPRS (General Packet Radio Service). This was revision 5 (GSM Rev.5), which allowed reaching rates of 115.2 kbps [1]

GSM carriers started developing a General Packet Radio Service (GPRS) service to achieve higher data rates. This system overlaid a packet-switching network on the existing circuit-switched GSM network. GPRS (2.5G) could transmit data at up to 160Kbps in 1995

After GPRS, another phase called Enhanced Data rates for GSM Evolution (EDGE) was introduced in 1997. It introduced an 8 PSK modulation scheme and could deliver data rates of up to 500Kbp/s using the same GPRS infrastructure.

2.75G Generation The 2.5G GPRS also received another innovation that allowed, thanks to a new carrier modulation system (8-PSK = 8 states Phase Shift Keying) and different error correction, increase the spectral efficiency. While GSM/GPRS processed 1 bit/symbol with GMSK modulation, this new 8-PSK modulation was able to process up to 3 bits/symbol, practically tripling the volume of transmitted data and receiving the name EDGE [1]

The internet was becoming popular, and data services were becoming more prevalent. Multimedia services and streaming started growing, and phones started supporting web browsing as the internet boomed in 1998. This forced and pushed the development of the 3rd generation communication systems to accommodate the need for more data and capacity. [2]

1.2.1 The 2G System Architecture

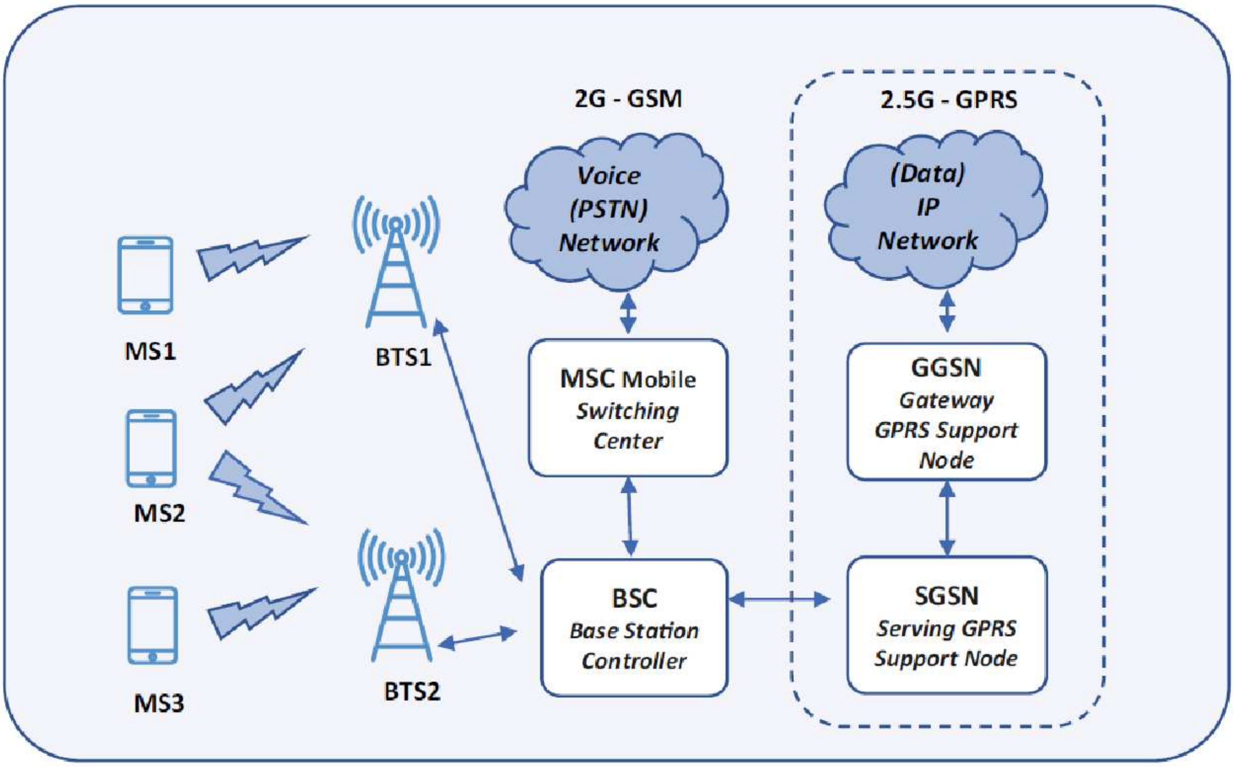

2G - GSM
2.5G - GPRS
Voice (PSTN) Network
(Data) IP Network
MS1
MS2
MS3
BTS1
BTS2
MSC Mobile Switching Center
GGSN Gateway GPRS Support Node
BSC Base Station Controller
SGSN Serving GPRS Support Node

Figure 1-4 the second-generation cellular systems – GSM (2G)/GPRS (2.5G) [1, p. 85]

In 2G and 2.5G, the *BSC* (Base Station Controller) was introduced, aggregating and controlling the traffic or signals from the BTSs and passing the telephone traffic to the *MSC*. The MSC is part of the Circuit Switching (CS), the GSM Core Network. The MSC, a legacy technology of the public switched telephone network, is still the central element of the GSM Core. Circuit Switching (CS); allows 2G to be connected to the PSTN. GSM is still a system focused on voice transmission. In 2.5G generation, the adoption of GPRS architecture allows interconnection with packet network (X.25 and Frame Relay), low-speed data transmission (9.6 to 171 Kbps), and *SMS* (Short Message Service) text message to implement frequency reuse. [1]

1.2.2 Security Flaws of the 2G System:

The most common types of attacks and vulnerability exploits in these types of networks are:

• GSM security flaws - no authentication of the Network is provided to the end user; vulnerabilities in the subscriber identity confidentiality mechanism;

• Impersonation attacks - the attacker tends to impersonate a legitimate user to conduct an attack;

- The attack gains anonymity - the attacker gains information on the user’s habits, calling patterns, etc. which can be used against the end user;

- The attacks against confidentiality - the attacker uses weaknesses in the GSM architecture; and flaws in the protocols between the GSM networks and the end user; significant attacks of the type are brute force attacks, cryptanalysis-based attacks, and non-cryptanalysis attacks;

- Denial of Service (DoS) attacks - the attacker floods the Network to disable the end users from accessing the Network by performing the attack episode using a physical or logical intervention. [4]

Other security vulnerabilities in 2G networks include: (i) obscurity, meaning that none of the security algorithms used by GSM is available to the public, (ii) provision of access security only, (iii) weak and difficult-to-upgrade cryptographic mechanisms, (iv) mobile subscriber visibility missing, and (v) authentication of the user to the Network and not vice versa [5]

1.3 3G MOBILE COMMUNICATION SYSTEM

Apart from the ability to support voice and SMS, 3G was designed to support data transmission; and wireless communication over the internet at a high data rate above 100Kbps mobile communication. 3G mainly has three standards: CDMA2000, WCDMA, and TD-SCDMA. [2] Development of 3G, 3GPP UMTS, and Universal Mobile Telecommunications System succeeded EDGE in 1999. This system uses Wide Band CDMA to carry the radio transmissions, often called WCDMA. The goal of UMTS or 3G wireless system was to provide a minimum data rate of 2MBits/s for stationary or walking users and 384Kbits/s in a moving vehicle. 3GPP designated it as release 99. The upgrades and additional facilities were introduced at successive releases of the 3GPP standard.

3GPP standard;

Release 4: This provided for the efficient use of IP, which was a key enabler for 3G HSDPA 2001

Release 5: This release included the core of HSDPA. It provided reduced delays for downlink packets and a data rate of 14Mbps.

Release 6: This included the core of HSDPA with a reduction in uplink delay…. enhanced the uplink raw data rate of 5. 74Mbps.This release also included MBMS for broadcasting services

Release 7: This release of the 3GPP standard included downlink MIMO operation and support for higher-order modulation of up to 64 QAM. Either MIMO or 64-QAM could be used at a time.

Evolved HSPA provides data rates up to 28Mbps in the downlink and 11Mbps in the uplink [6]

3G allowed more mobile users to connect to the internet and access multimedia, VoIP, and services. Third parties developed more software, which allowed mobile users to use applications for voice and video calls. E.g. WhatsApp and WeChat. More users started using these services over the SMS service since it was more convenient and cheaper. This connected the world and brought people from all walks of life closer to each other. As more users got on board 3G technology and its benefits, there was a pressing need to develop a much faster and more reliable system that could meet the data needs of the evolving world. This pushed the development of 4G. [2] [6]

Formation of governing bodies:

To produce genuinely global standards, a collaboration for both GSM and UMTS was expanded from ETSI to encompass regional Standards Development Organizations such as ARIB and TTC from Japan, TTA from Korea, ATIS from North America and CCSA from China.

The successful creation of such a large and complex system specification required a well-structured organization. This gave birth to 3GPP in 2000, which worked under the observation of ITU-R. ITU-R is one of the sectors of ITU. Its role is to manage the international radio frequency spectrum and to ensure the effective use of the spectrum ITU-R defines technology families and associates specific parts of the range with these families. It also proposed requirements for radio technology.

Three organizations started developing standards to meet standards proposed by ITU-R, i.e. 3GPP, 3GPP2, and IEEE.

The evolution of 3GPP started from GSM to LTE advanced.

The evolution of 3GPP2 started from IS95 to CDMA Revision B

IEEE evolution started from 802.16 FIXED WIMAX to 802.16M

But 3GPP dominated and was widely accepted. Its roadmap was incorporated [6].

1.3.1 3G System Architecture

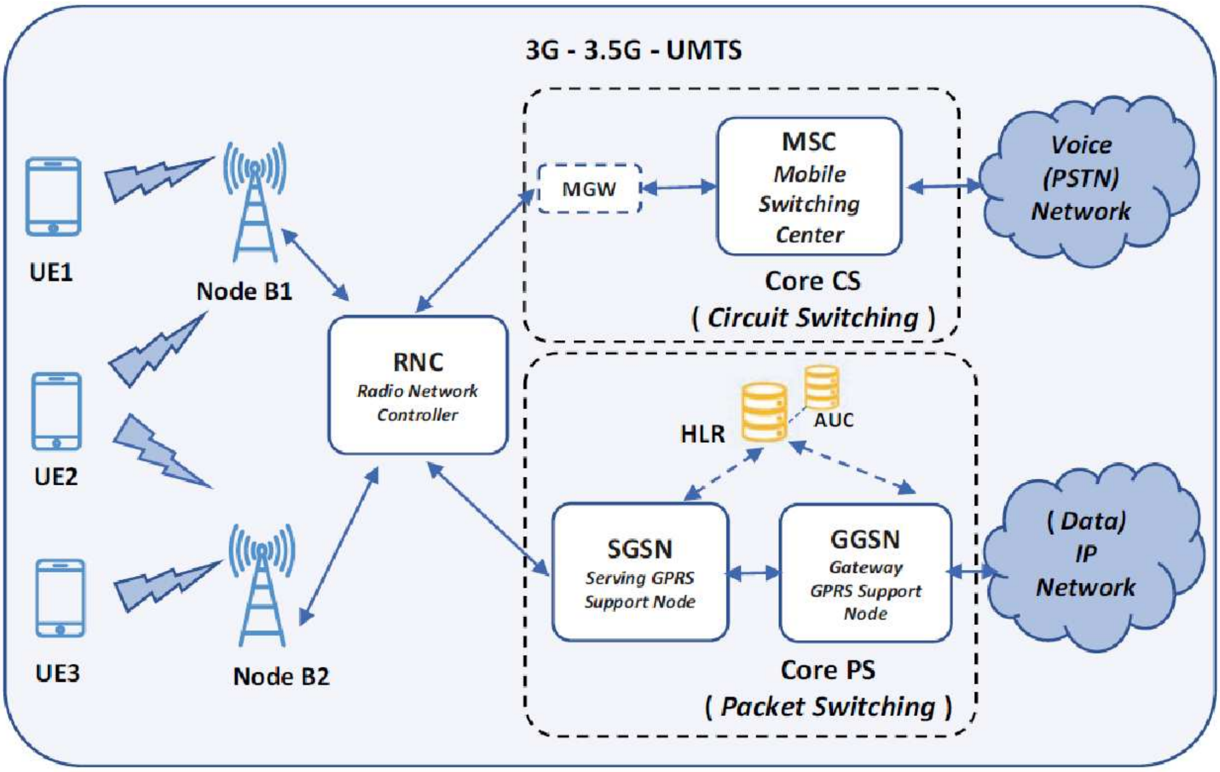


Figure 1-5. Third-generation cellular systems – UMTS (3/3.5G) [1]

In this architecture in Figure 1-5, the BSC is replaced by the RNC (Radio Network Controller), which is more intelligent and can efficiently perform handovers without the help of the MSC and SGSN, which promotes traffic aggregation and controls the NodeBs. SGSN and GGSN; process data traffic, then direct it to PS (Packet Switching), and telephone traffic is sent to CS (Circuit Switching). There, the MSC is used to process the voice calls. The PS interconnects to the packet/internet the CS interconnects 3G to the legacy networks. [1]

The UMTS core network consists of several functional elements, some in the circuit-switched (CS) domain, some in the packet-switched (PS) domain, and some shared by both parts. Active components in the CS domain are primarily based on the GSM network elements and carry data in a circuit-switched manner (i.e., a fixed channel for the duration of the call). These elements are:

Mobile switching center (MSC): Manages circuit-switched calls. The MSC also includes a visitor location register (VLR), which controls mobile stations roaming in the MSC area. When a mobile station (MS) enters a new location area, it starts a registration procedure. An MSC in charge of that area notices this registration and transfers it to a visitor location to register the identity of the location area where the MS is situated. Gateway MSC (GMSC): The interface to the PTSN and GSM networks. [7]

The elements in the PS domain are:

Serving GPRS support node (SGSN): This element was first introduced in 2G GSM systems to support packet-switched functionality. It provides the following functions:

Mobility management (MM): When user equipment (UE) attaches to the PS domain, the SGSN generates MM information based on the mobile's location. [7]

Session management: data sessions are managed by the SGSN, ensuring the quality of service. It also contains the PDP (Packet Data Protocol) contexts, which are logical connections over which the data packets are sent. [7]

Billing: The SGSN monitors the flow of data to generate billing information.

Gateway GPRS support node (GGSN): The GGSN is the central element within the PS domain. It handles interworking between the UMTS packet-switched and external packet-switched networks, thus functioning as a router. When the GGSN receives data addressed to a specific user, it checks whether the user is active and then forwards the data to the SGSN serving the particular UE. [7]

The elements that link the CS and PS domains are:

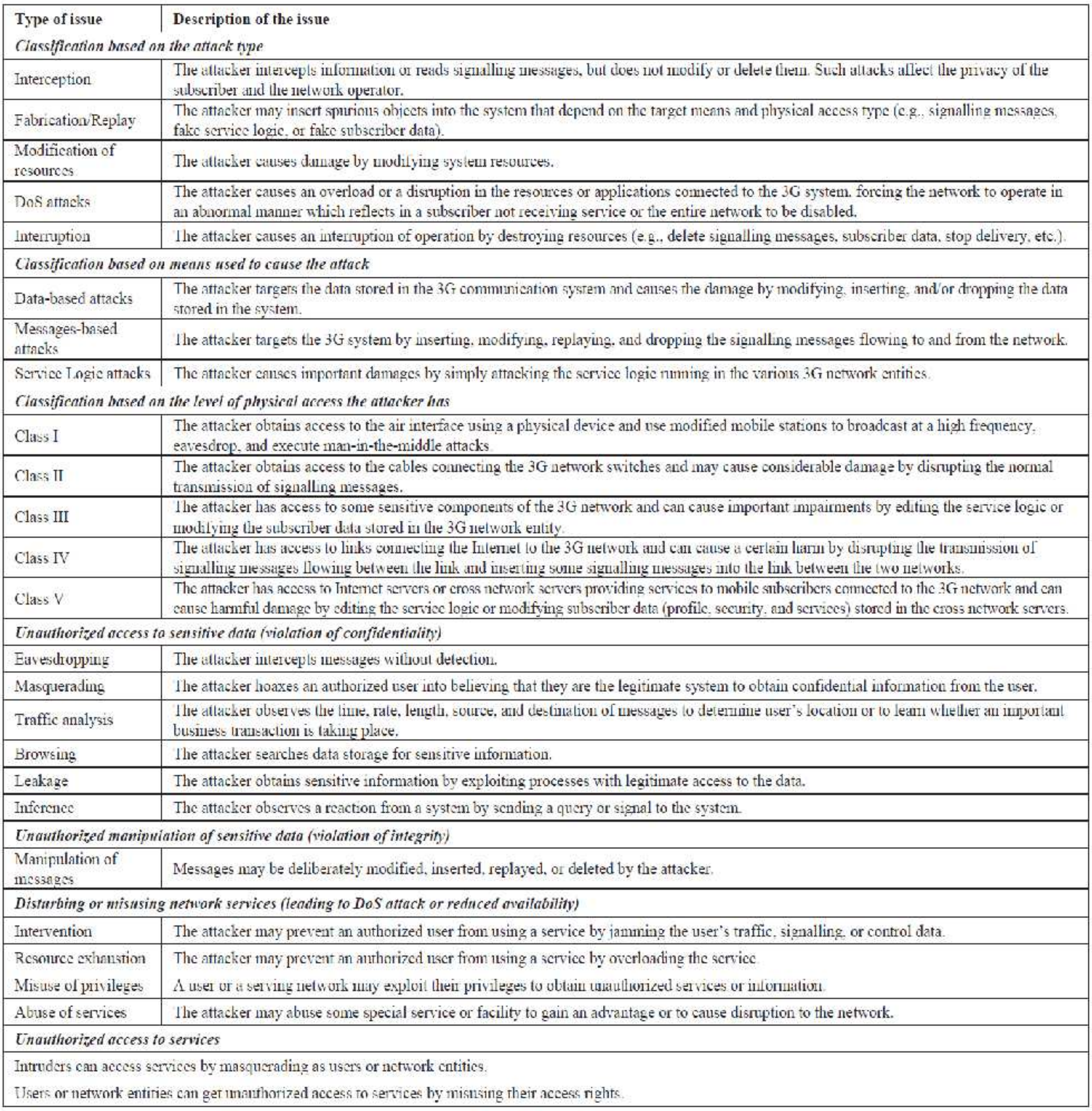
Home location register (HLR): This database contains administrative information about each subscriber; and that individual's last known location. In this way, the UMTS network can route calls to the relevant RNC/Node B. When a user switches on his UE, it registers with the Network, and from this, it is possible to determine which Node B it communicates with so that incoming calls can be routed correctly. Even when the UE is not active (but switched on), it re-registers periodically to ensure that the Network (HLR) is aware of its latest position with the current or last-known location on the Network.

Equipment identity register (EIR): This entity decides whether given user equipment may be allowed on the Network. Each piece of user equipment has a unique identifier known as the International Mobile Equipment Identity (IMEI). This number is installed in the equipment and is checked by the Network during registration.

Authentication center (AUC): The AuC is a secure database with the secret key in the user's UMTS SIM (USIM) card. The USIM card includes a microprocessor that provides enhanced capability compared to a SIM card. [7]

1.3.2 3G Security flaws.

Table 1-1. Various 3G Security flaws and descriptions [5]



Type of issue	Description of the issue
Classification based on the attack type	
Interception	The attacker intercepts information or reads signalling messages, but does not modify or delete them. Such attacks affect the privacy of the subscriber and the network operator.
Fabrication/Replay	The attacker may insert spurious objects into the system that depend on the target means and physical access type (e.g., signalling messages, fake service logic, or fake subscriber data).
Modification of resources	The attacker causes damage by modifying system resources.
DoS attacks	The attacker causes an overload or a disruption in the resources or applications connected to the 3G system, forcing the network to operate in an abnormal manner which reflects in a subscriber not receiving service or the entire network to be disabled.
Interruption	The attacker causes an interruption of operation by destroying resources (e.g., delete signalling messages, subscriber data, stop delivery, etc.).
Classification based on means used to cause the attack	
Data-based attacks	The attacker targets the data stored in the 3G communication system and causes the damage by modifying, inserting, and/or dropping the data stored in the system.
Messages-based attacks	The attacker targets the 3G system by inserting, modifying, replaying, and dropping the signalling messages flowing to and from the network.
Service Logic attacks	The attacker causes important damages by simply attacking the service logic running in the various 3G network entities.
Classification based on the level of physical access the attacker has	
Class I	The attacker obtains access to the air interface using a physical device and use modified mobile stations to broadcast at a high frequency, eavesdrop, and execute man-in-the-middle attacks.
Class II	The attacker obtains access to the cables connecting the 3G network switches and may cause considerable damage by disrupting the normal transmission of signalling messages.
Class III	The attacker has access to some sensitive components of the 3G network and can cause important impairments by editing the service logic or modifying the subscriber data stored in the 3G network entity.
Class IV	The attacker has access to links connecting the Internet to the 3G network and can cause a certain harm by disrupting the transmission of signalling messages flowing between the link and inserting some signalling messages into the link between the two networks.
Class V	The attacker has access to Internet servers or cross network servers providing services to mobile subscribers connected to the 3G network and can cause harmful damage by editing the service logic or modifying subscriber data (profile, security, and services) stored in the cross network servers.
Unauthorized access to sensitive data (violation of confidentiality)	
Eavesdropping	The attacker intercepts messages without detection.
Masquerading	The attacker hoaxes an authorized user into believing that they are the legitimate system to obtain confidential information from the user.
Traffic analysis	The attacker observes the time, rate, length, source, and destination of messages to determine user's location or to learn whether an important business transaction is taking place.
Browsing	The attacker searches data storage for sensitive information.
Leakage	The attacker obtains sensitive information by exploiting processes with legitimate access to the data.
Inference	The attacker observes a reaction from a system by sending a query or signal to the system.
Unauthorized manipulation of sensitive data (violation of integrity)	
Manipulation of messages	Messages may be deliberately modified, inserted, replayed, or deleted by the attacker.
Disturbing or misusing network services (leading to DoS attack or reduced availability)	
Intervention	The attacker may prevent an authorized user from using a service by jamming the user's traffic, signalling, or control data.
Resource exhaustion	The attacker may prevent an authorized user from using a service by overloading the service.
Misuse of privileges	A user or a serving network may exploit their privileges to obtain unauthorized services or information.
Abuse of services	The attacker may abuse some special service or facility to gain an advantage or to cause disruption to the network.
Unauthorized access to services	
Intruders can access services by masquerading as users or network entities.	
Users or network entities can get unauthorized access to services by misusing their access rights.	

1.4 4G LTE MOBILE COMMUNICATION SYSTEM

The initial goal of telecommunication was mobility and global connectivity, but as the technology evolved, the services started expanding. For example, now services are not restricted to voice and SMS only. For this expansion and its successful, efficient execution in LTE, a whole new architecture was adopted for both non-radio part System Architecture Evolution (SAE) and radio part using pure IP architecture; and packet switching.

To fulfill the requirements proposed by ITU-R, a study group was formed, and LTE standardization began in 2004. As a result, many telecom companies collaborated to achieve their shared vision.

LTE (Long Term Revolution) is the long-term evolution of 3GPP based on the UMTS (Universal Mobile Telecommunications System) technology standard developed in 2004. It has a peak download speed of 299.6 Mbit/s and a peak upload speed of 75.4 Mbit/s. [2]

In June 2005, release eight was finally crystallized after refining. Some of the significant features of release eight were reduced delays for both connection establishment and transmission latency; increased user data throughput; increased cell-edge bit-rate; uniformity; reduced cost per bit implying improved spectral efficiency; simplified network architecture; seamless mobility including between different radio access technologies; reasonable power consumption for mobile devices. Advancements in the underlying mobile radio technology fulfilled these requirements.

The three fundamental radio technologies that have shaped the LTE radio interface design were:

Multi-carrier technology, multiple antenna technology and the application of packet switching to the radio interface

Specifications for release eight were completed by December 2007. The first commercial deployment occurred in northern Europe by the end of 2009. 4G took center stage drastically in 2010

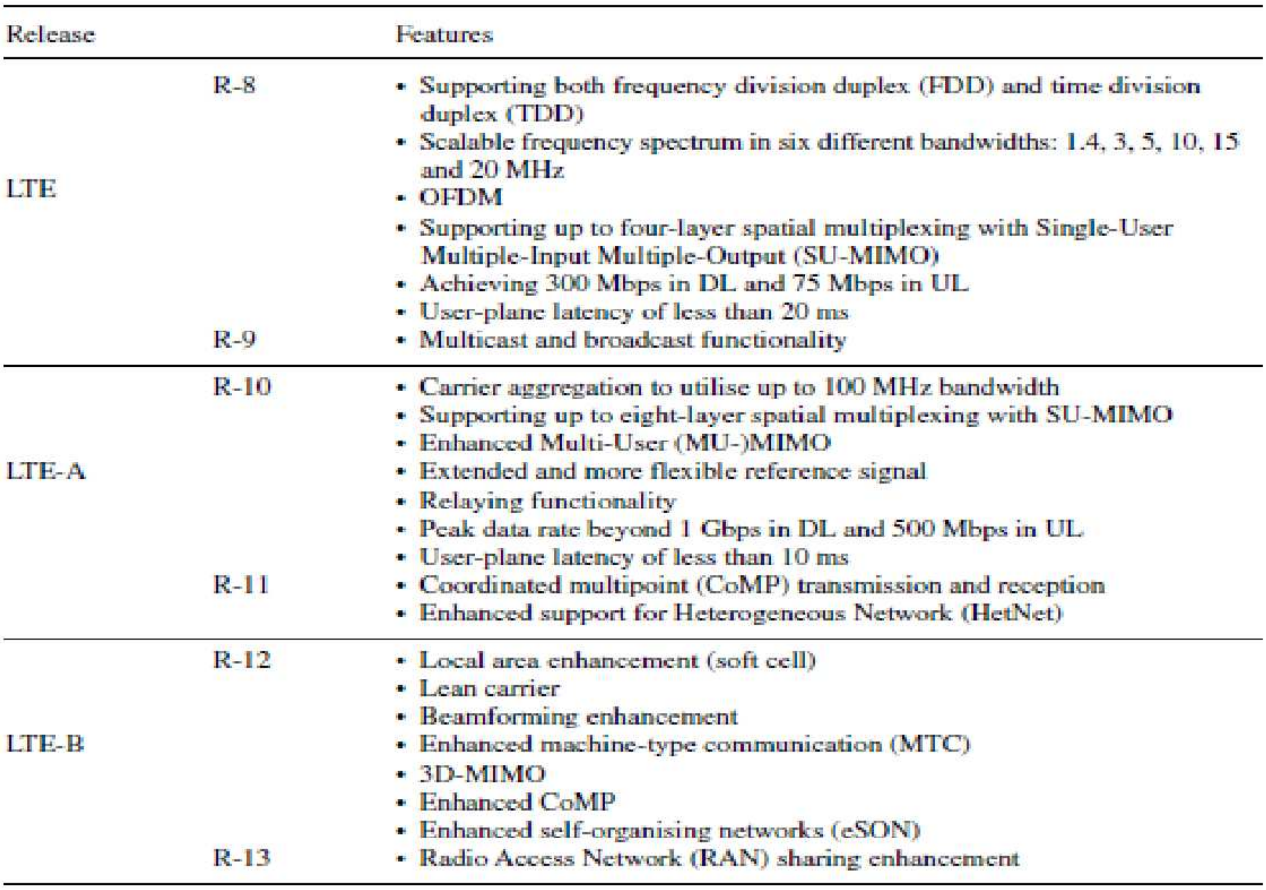
In the subsequent releases, multiple services such as multi-cell HSDPA, HET NET, Coordinate Multipoint, carrier aggregation and Massive MIMO, and many more were targeted for a rich customer experience.

The network speed in the 4G era is almost ten times that of 3G, and high-quality video calls, files and pictures transmission will be unimpeded.

We then moved from the Services to the multi-services approach, from LTE to 5G. [6]

4G gave a new direction to see the service to be provided. It provided broadband speeds, high quality, and high capacity to users while improving security and lowering the cost of voice and data services, multimedia, and Internet access. Applications include enhanced mobile web access, gaming services, IP telephony, high-definition mobile TV, video conferencing, 3D television, and cloud computing. By adding higher speed, which means quality, and a broad "suite" of mobile applications, LTE has added value to mobile telephony. This has enabled operators to get a better return on the huge investments they have already made and have yet to make and to earn higher profits as revenues from voice telephony have been eroded. But value-added services became more important than transport services. As time went by, mobile operators began to enable profits to those using their transport networks without sharing the revenues of the products transported. [1]

Table 1-2. Features of the different LTE releases [8, p. 5]



Release		Features
LTE	R-8	• Supporting both frequency division duplex (FDD) and time division duplex (TDD) • Scalable frequency spectrum in six different bandwidths: 1.4, 3, 5, 10, 15 and 20 MHz • OFDM • Supporting up to four-layer spatial multiplexing with Single-User Multiple-Input Multiple-Output (SU-MIMO) • Achieving 300 Mbps in DL and 75 Mbps in UL • User-plane latency of less than 20 ms
	R-9	• Multicast and broadcast functionality
LTE-A	R-10	• Carrier aggregation to utilise up to 100 MHz bandwidth • Supporting up to eight-layer spatial multiplexing with SU-MIMO • Enhanced Multi-User (MU-)MIMO • Extended and more flexible reference signal • Relaying functionality • Peak data rate beyond 1 Gbps in DL and 500 Mbps in UL • User-plane latency of less than 10 ms
	R-11	• Coordinated multipoint (CoMP) transmission and reception • Enhanced support for Heterogeneous Network (HetNet)
LTE-B	R-12	• Local area enhancement (soft cell) • Lean carrier • Beamforming enhancement • Enhanced machine-type communication (MTC) • 3D-MIMO • Enhanced CoMP • Enhanced self-organising networks (eSON)
	R-13	• Radio Access Network (RAN) sharing enhancement

1.4.1 The 4G System Architecture

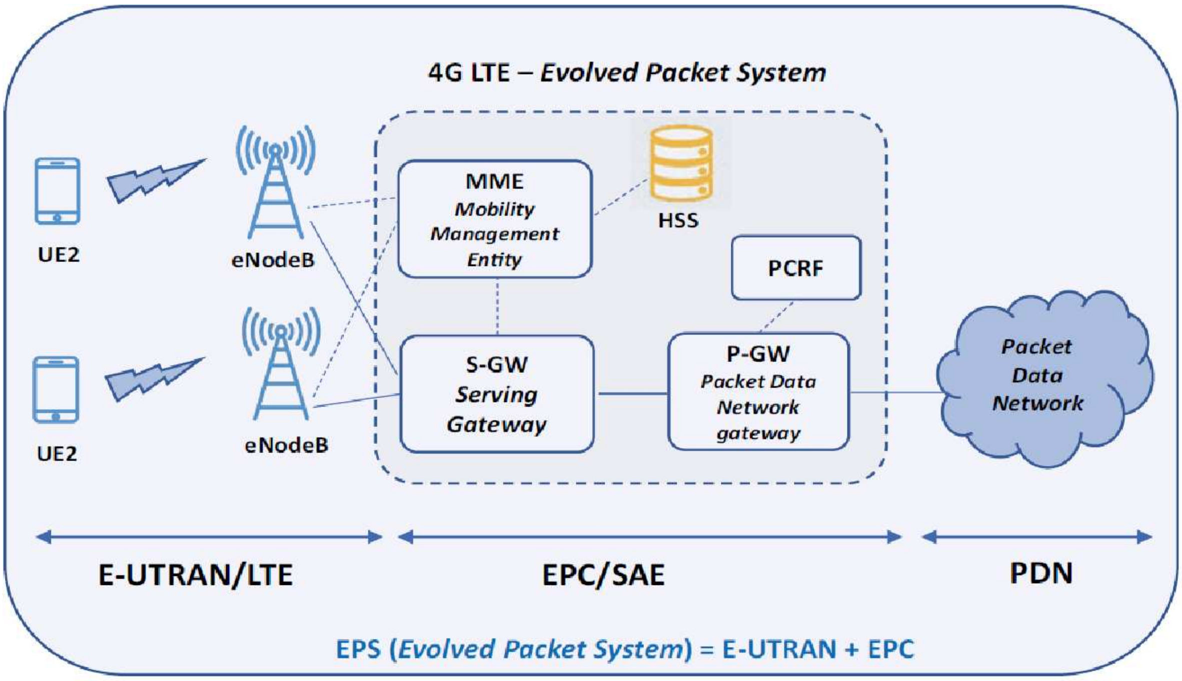


Figure 1-6. 4G LTE access network architecture – evolved packet system. E-UTRAN (evolved universal terrestrial radio access network) + EPC (evolved packet core) CORE [1]

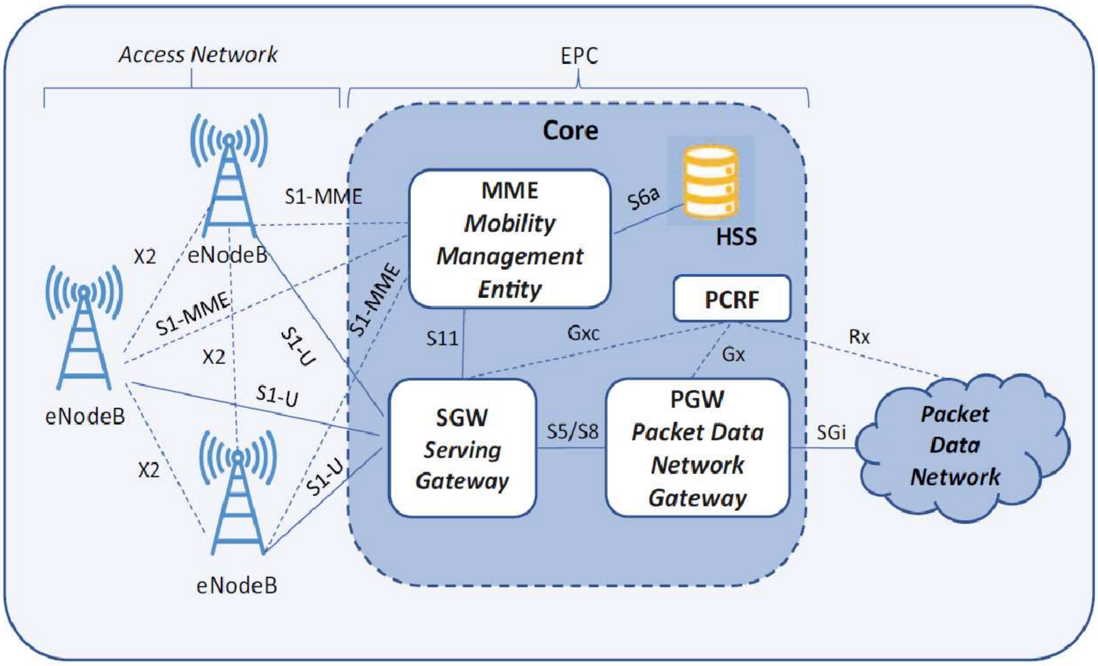


Figure 1-7. E-UTRAN [1]

In E-UTRAN, the base station is evolved NodeB (eNodeB). While NodeB in UMTS is based on CDMA, eNodeB is based on OFDMA. Instead of the existence of an RNC, the eNodeB has control functionality embedded in its operation, thus supporting radio resource control, admission control, and mobility management, which were initially the function of an RNC. [7]

This architecture is open to allow for interconnections between different manufacturers. We see that eNodeBs located in the same area communicate with each other through the interface called X2. The eNodeBs communicate with the Core EPC through the S1 interface, while the Core EPCs communicate through the S10 interface. S1 interfaces can also be established between an eNodeB and other Core EPCs to speed up communication between all network components. [1]

Evolved Packet Core

The operator, or carrier, a network that interconnects all of the base stations of the carrier, is referred to as the evolved packet core (EPC). Traditionally, the core cellular network was circuit-switched, but for 4G, the core is entirely packet-switched. It is based on IP and supports voice connections using voice over IP (VoIP). Figure 1-7 illustrates the essential components of the EPC:

Mobility management entity (MME): The MME deals with control signalling related to mobility and security. The MME is responsible for tracking and paging the UE in idle mode.

Serving gateway (SGW): The SGW deals with user data transmitted and received by UEs in packet form; using IP. The SGW is the interconnect point between the radio side and the EPC. As its name indicates, this gateway serves the UE by routing IP packets. Furthermore, it is the anchor point for intra-LTE mobility (i.e., in the case of handover between eNodeBs). Thus packets can be routed from an eNodeB to an eNodeB in another area via the SGW and routed to external networks such as the internet (via the PGW).

Packet data network gateway (PGW): The PGW is the point of interconnection between the EPC and external IP networks like the internet. The PGW routes packets to and from the external networks. It also performs various functions, such as IP address/IP prefix allocation, policy control, and charging.

Home subscriber server (HSS): the HSS maintains a database that contains user-related and subscriber-related information. It also provides support functions in mobility management, call and session setup, user authentication, and access authorization. [7]

1.4.2 LTE Security Model

Figure 1-8 shows the authentication method of LTE with step‑by‑step details. The authentication process in LTE is initiated by the authentication server when it sends Enhance Authentication Protocol request/Identity message (EAP) to the UE. The UE responds by replying to the EAP‑response/Identity message containing the identity message and Network Access Identifier (NAI). Upon receipt of the EAP‑resEAP response message, the authentication server tried to access the UE's certificate from its record. Next, the authentication server generates the EAP‑Request/Authentication and Key Agreement (AKA) ‑Challenge message using the standard AKA process.

The authentication server sends the EAP‑Request/AKA‑Challenge message encrypted by the UE's public key to the UE. The UE then decrypts the EAP‑Request/AKAChallenge message using its pre-key; and sends the EAP‑Response/ AKA‑Challenge to the authentication server. Next, the authentication server decrypts the information using the server's private key and verifies the EAP‑response/AKAChallenge message using the AKA algorithm. If the message is correct, the EAP server sends the EAP success message to the UE. [9]

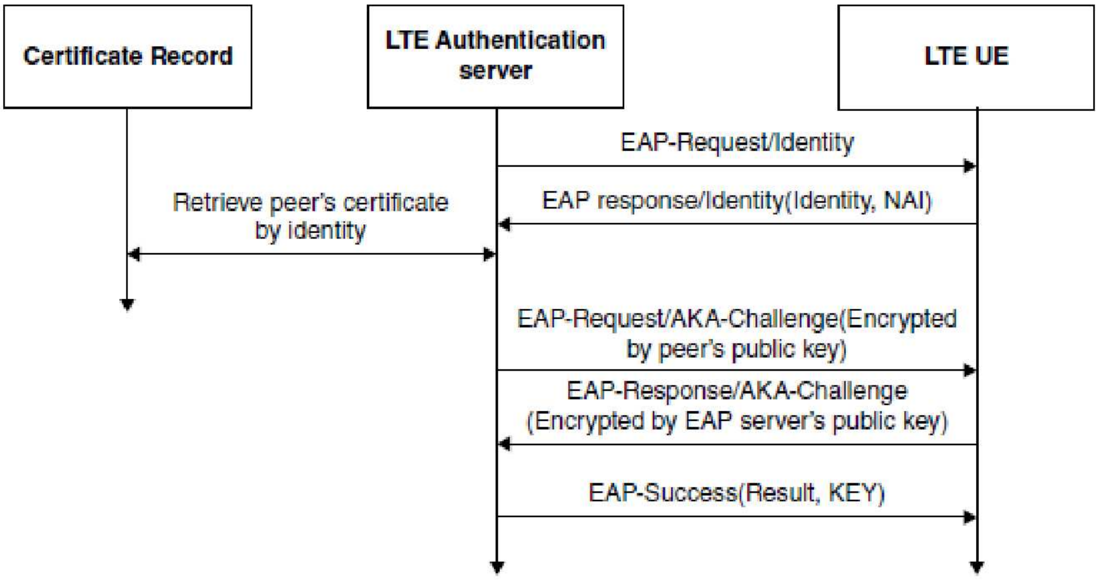
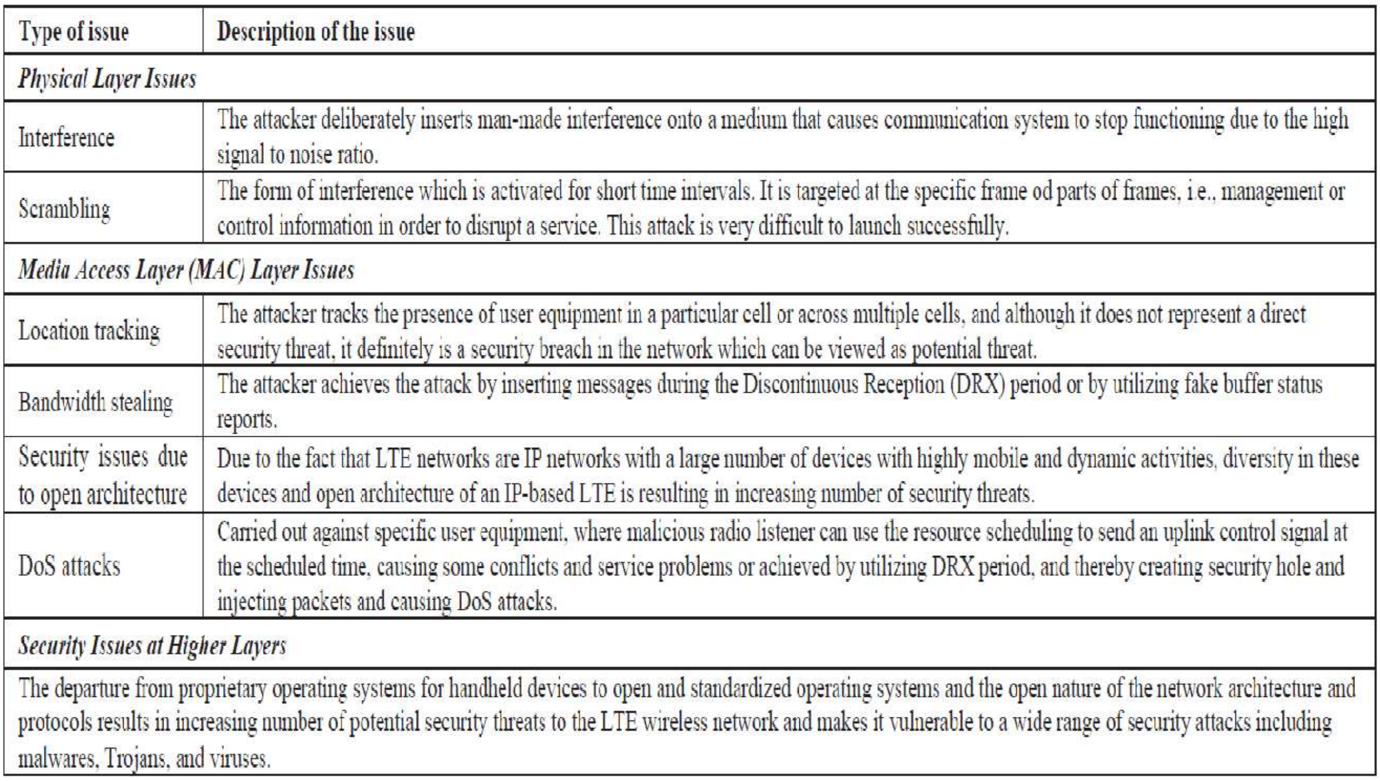

Certificate Record
LTE Authentication server
LTE UE
EAP-Request/Identity
EAP response/Identity(Identity, NAI)
Retrieve peer's certificate by identity
EAP-Request/AKA-Challenge(Encrypted by peer's public key)
EAP-Response/AKA-Challenge (Encrypted by EAP server's public key)
EAP-Success(Result, KEY)

Figure 1-8. Authentication process in LTE [9]

1.4.3 Security attacks in 4G/LTE networks

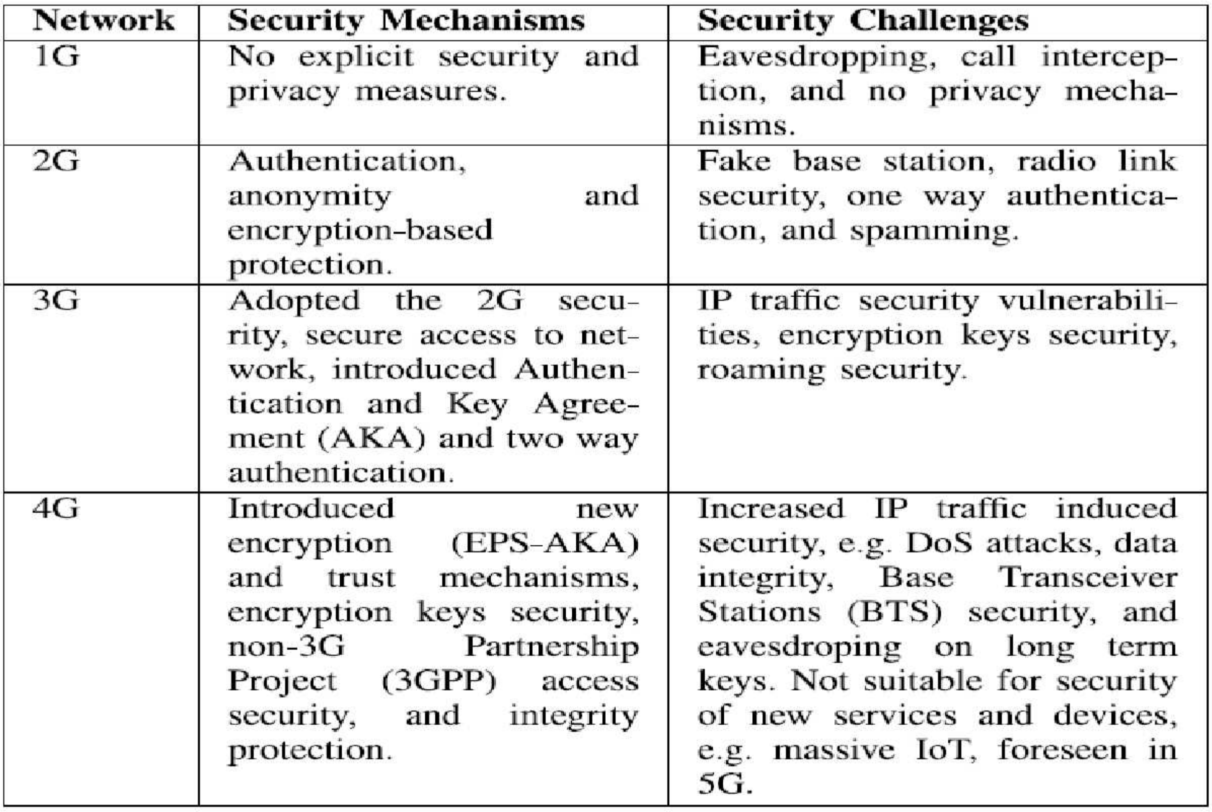
Table 1-3. Common security attacks/ flaws in 4G/LTE networks [5]



Type of issue	Description of the issue
Physical Layer Issues	
Interference	The attacker deliberately inserts man-made interference onto a medium that causes communication system to stop functioning due to the high signal to noise ratio.
Scrambling	The form of interference which is activated for short time intervals. It is targeted at the specific frame od parts of frames, i.e., management or control information in order to disrupt a service. This attack is very difficult to launch successfully.
Media Access Layer (MAC) Layer Issues	
Location tracking	The attacker tracks the presence of user equipment in a particular cell or across multiple cells, and although it does not represent a direct security threat, it definitely is a security breach in the network which can be viewed as potential threat.
Bandwidth stealing	The attacker achieves the attack by inserting messages during the Discontinuous Reception (DRX) period or by utilizing fake buffer status reports.
Security issues due to open architecture	Due to the fact that LTE networks are IP networks with a large number of devices with highly mobile and dynamic activities, diversity in these devices and open architecture of an IP-based LTE is resulting in increasing number of security threats.
DoS attacks	Carried out against specific user equipment, where malicious radio listener can use the resource scheduling to send an uplink control signal at the scheduled time, causing some conflicts and service problems or achieved by utilizing DRX period, and thereby creating security hole and injecting packets and causing DoS attacks.
Security Issues at Higher Layers	
The departure from proprietary operating systems for handheld devices to open and standardized operating systems and the open nature of the network architecture and protocols results in increasing number of potential security threats to the LTE wireless network and makes it vulnerable to a wide range of security attacks including malwares, Trojans, and viruses.	

.

Table 1-4. A summary of Security mechanisms and challenges from 1G to 5G [10]



Network	Security Mechanisms	Security Challenges
1G	No explicit security and privacy measures.	Eavesdropping, call interception, and no privacy mechanisms.
2G	Authentication, anonymity and encryption-based protection.	Fake base station, radio link security, one way authentication, and spamming.
3G	Adopted the 2G security, secure access to network, introduced Authentication and Key Agreement (AKA) and two way authentication.	IP traffic security vulnerabilities, encryption keys security, roaming security.
4G	Introduced new encryption (EPS-AKA) and trust mechanisms, encryption keys security, non-3G Partnership Project (3GPP) access security, and integrity protection.	Increased IP traffic induced security, e.g. DoS attacks, data integrity, Base Transceiver Stations (BTS) security, and eavesdroping on long term keys. Not suitable for security of new services and devices, e.g. massive IoT, foreseen in 5G.

1.5 5G MOBILE COMMUNICATION SYSTEM

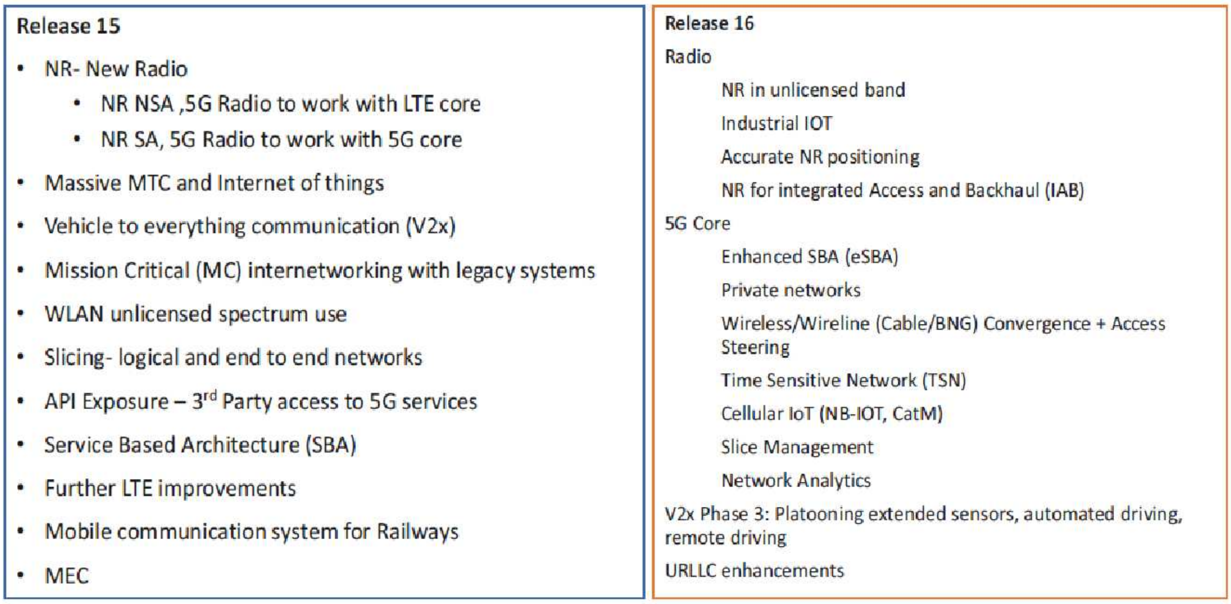
5G represents a considerable evolution. It builds on everything revolutionary about 4G. It adds new possibilities, making it the structural technology that will underpin the new applications needed for Industry 4.0 and the development of the IoT. Given the vast number of new services and applications, 5G had to become highly flexible and efficient, being able to serve from straightforward applications, such as the reading of a temperature sensor that sends data to a platform once or twice a day, to applications that require very high throughput and sensor readings in fractions of a millisecond, such as the control of a mechanical arm. 5G aims to work with extremes, serving the most diverse types of existing technologies and those yet to be created. [1]

5G was designed not just for human-to-human communications but to encompass machine-to-machine and human-to-machine communications.

5G is designed to support various services with different data traffic profiles (e.g., high throughput, low latency and massive connections) and models (e.g., IP data traffic, non-IP data traffic, short data bursts and high throughput data transmissions). In addition, various PDU session types are supported, including IPv4, Ipv6, Ipv4v6, Ethernet and Unstructured.

Some of the new applications will enable our Multi-gigabit wireless mobile broadband Fixed broadband wireless access, Augmented reality (AR) and virtual reality (VR), Autonomous vehicles, Vehicle-to-everything (V2X) communications, The mobile Internet of Things (IoT) and The wireless industrial IoT (IoT)

The 3GPP standards for 5G began with Release 15, which set the ground for new radio (NR) and the basis for non-standalone (NSA) 5G networks that leveraged the existing LTE core networks; the early drop for this was in 2018. It also detailed some enhancements in the LTE core, such as control and user plane separation, to better cater to 5G adoption. Next, release 15 details for the 5G standalone (SA) core networks. Then release 16, released in June 2020, added more features for 5G. Let us understand how we started with 5G, the contents of various releases, the details of Release 16, and what is planned for Release 17 that makes it so exciting. Infrastructure-wise, a significant difference between 4G and 5G is that 4G started the movement to a virtualized network, and 5G pushed it further to a containerized infrastructure. Release 16 introduced more features, mainly focusing on industrial usage. Figure 1-9 illustrates the details of Release 16. Additionally, different versions of Release 16 will continue over the next few quarters. [11]



Release 15

- NR- New Radio
  - NR NSA ,5G Radio to work with LTE core
  - NR SA, 5G Radio to work with 5G core
- Massive MTC and Internet of things
- Vehicle to everything communication (V2x)
- Mission Critical (MC) internetworking with legacy systems
- WLAN unlicensed spectrum use
- Slicing- logical and end to end networks
- API Exposure – 3rd Party access to 5G services
- Service Based Architecture (SBA)
- Further LTE improvements
- Mobile communication system for Railways
- MEC

Release 16

Radio

NR in unlicensed band

Industrial IOT

Accurate NR positioning

NR for integrated Access and Backhaul (IAB)

5G Core

Enhanced SBA (eSBA)

Private networks

Wireless/Wireline (Cable/BNG) Convergence + Access Steering

Time Sensitive Network (TSN)

Cellular IoT (NB-IOT, CatM)

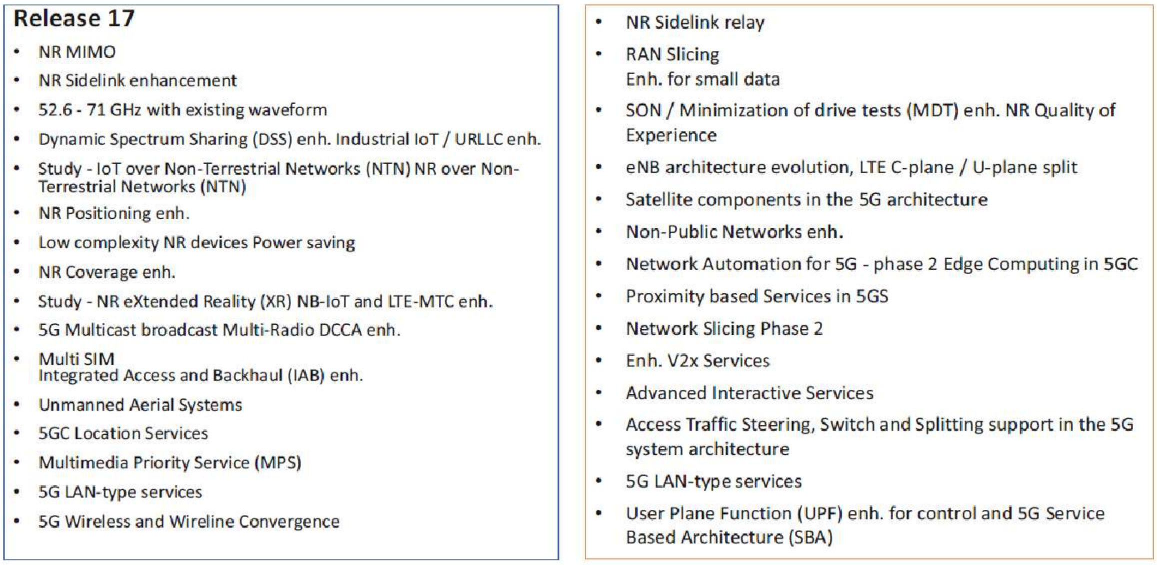
Slice Management

Network Analytics

V2x Phase 3: Platooning extended sensors, automated driving, remote driving

URLLC enhancements

Figure 1-9. Releases 15 and 16 contents [11]



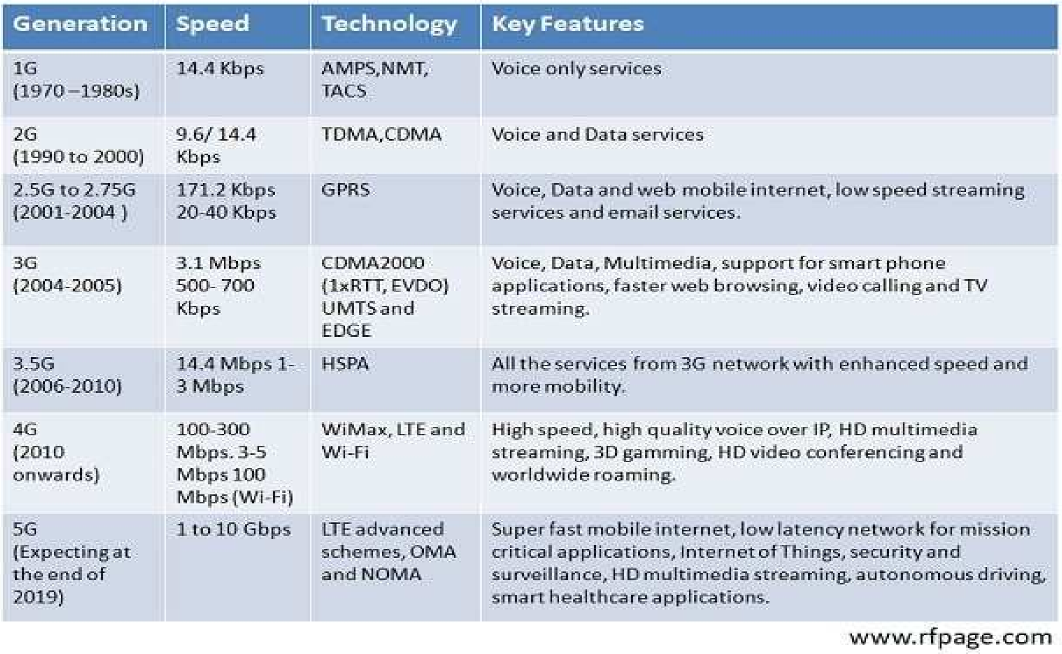
Release 17

- NR MIMO
- NR Sidelink enhancement
- 52.6 - 71 GHz with existing waveform
- Dynamic Spectrum Sharing (DSS) enh. Industrial IoT / URLLC enh.
- Study - IoT over Non-Terrestrial Networks (NTN) NR over Non-Terrestrial Networks (NTN)
- NR Positioning enh.
- Low complexity NR devices Power saving
- NR Coverage enh.
- Study - NR eXtended Reality (XR) NB-IoT and LTE-MTC enh.
- 5G Multicast broadcast Multi-Radio DCCA enh.
- Multi SIM
  Integrated Access and Backhaul (IAB) enh.
- Unmanned Aerial Systems
- 5GC Location Services
- Multimedia Priority Service (MPS)
- 5G LAN-type services
- 5G Wireless and Wireline Convergence

- NR Sidelink relay
- RAN Slicing
  Enh. for small data
- SON / Minimization of drive tests (MDT) enh. NR Quality of Experience
- eNB architecture evolution, LTE C-plane / U-plane split
- Satellite components in the 5G architecture
- Non-Public Networks enh.
- Network Automation for 5G - phase 2 Edge Computing in 5GC
- Proximity based Services in 5GS
- Network Slicing Phase 2
- Enh. V2x Services
- Advanced Interactive Services
- Access Traffic Steering, Switch and Splitting support in the 5G system architecture
- 5G LAN-type services
- User Plane Function (UPF) enh. for control and 5G Service Based Architecture (SBA)

Figure 1-10. Release 17 contents [11]

Release 17, as shown in Figure 1-10, is set to introduce full-scale deployment of 5G and came up with further enhancements for the 5G network. 5G technology will provide speeds that transcend the prior technologies discussed. Thus we can accommodate the needs of both consumers and industries in their various applications. 5G is expected to provide speeds of up to 10GB/s and 1 ms or less latency. 5G will enable service providers to provide more capacity; hence data-intensive applications can be catered to. With 5G, we can accommodate the data needs of humans, machines and smaller devices that require the internet or data to function efficiently. Per its standards, 5G inherently caters to be ultra-reliable and has provisions to have no connection loss, enabling it to be adapted by critical applications in healthcare for applications such as remote surgery. [11]



Generation	Speed	Technology	Key Features
1G (1970 –1980s)	14.4 Kbps	AMPS,NMT, TACS	Voice only services
2G (1990 to 2000)	9.6/ 14.4 Kbps	TDMA,CDMA	Voice and Data services
2.5G to 2.75G (2001-2004)	171.2 Kbps 20-40 Kbps	GPRS	Voice, Data and web mobile internet, low speed streaming services and email services.
3G (2004-2005)	3.1 Mbps 500- 700 Kbps	CDMA2000 (1xRTT, EVDO) UMTS and EDGE	Voice, Data, Multimedia, support for smart phone applications, faster web browsing, video calling and TV streaming.
3.5G (2006-2010)	14.4 Mbps 1-3 Mbps	HSPA	All the services from 3G network with enhanced speed and more mobility.
4G (2010 onwards)	100-300 Mbps. 3-5 Mbps 100 Mbps (Wi-Fi)	WiMax, LTE and Wi-Fi	High speed, high quality voice over IP, HD multimedia streaming, 3D gamming, HD video conferencing and worldwide roaming.
5G (Expecting at the end of 2019)	1 to 10 Gbps	LTE advanced schemes, OMA and NOMA	Super fast mobile internet, low latency network for mission critical applications, Internet of Things, security and surveillance, HD multimedia streaming, autonomous driving, smart healthcare applications.

www.rfpage.com

Figure 1-11. How 5G compares to other mobile networks [12]

1.5.1 Key Goals of 5G

- Very high throughput (1-20 Gbps).
- Ultra low latency (<1ms).
- 1000x bandwidth per unit area.
- Massive connectivity.
- High availability.
- Dense coverage
- Low energy consumption and
- Up to 10 years of battery life for machine-type communication [13]

Below are some of the forecasts set by 5G PP (A Joint initiative between the EU Commission and the European ICT):

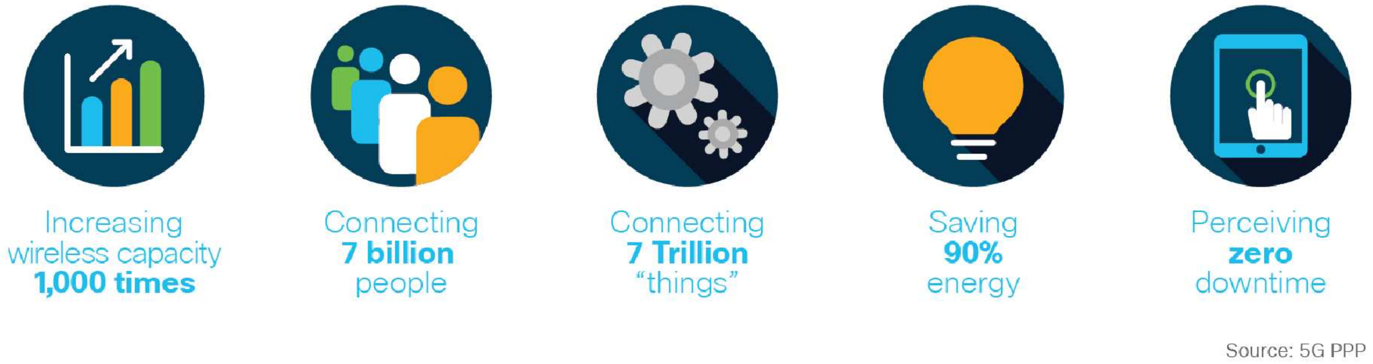


Figure 1-12. Forecasts set by 5G PP (A Joint initiative between EU Commission and European ICT) [13]

1.5.2 5G USE CASES

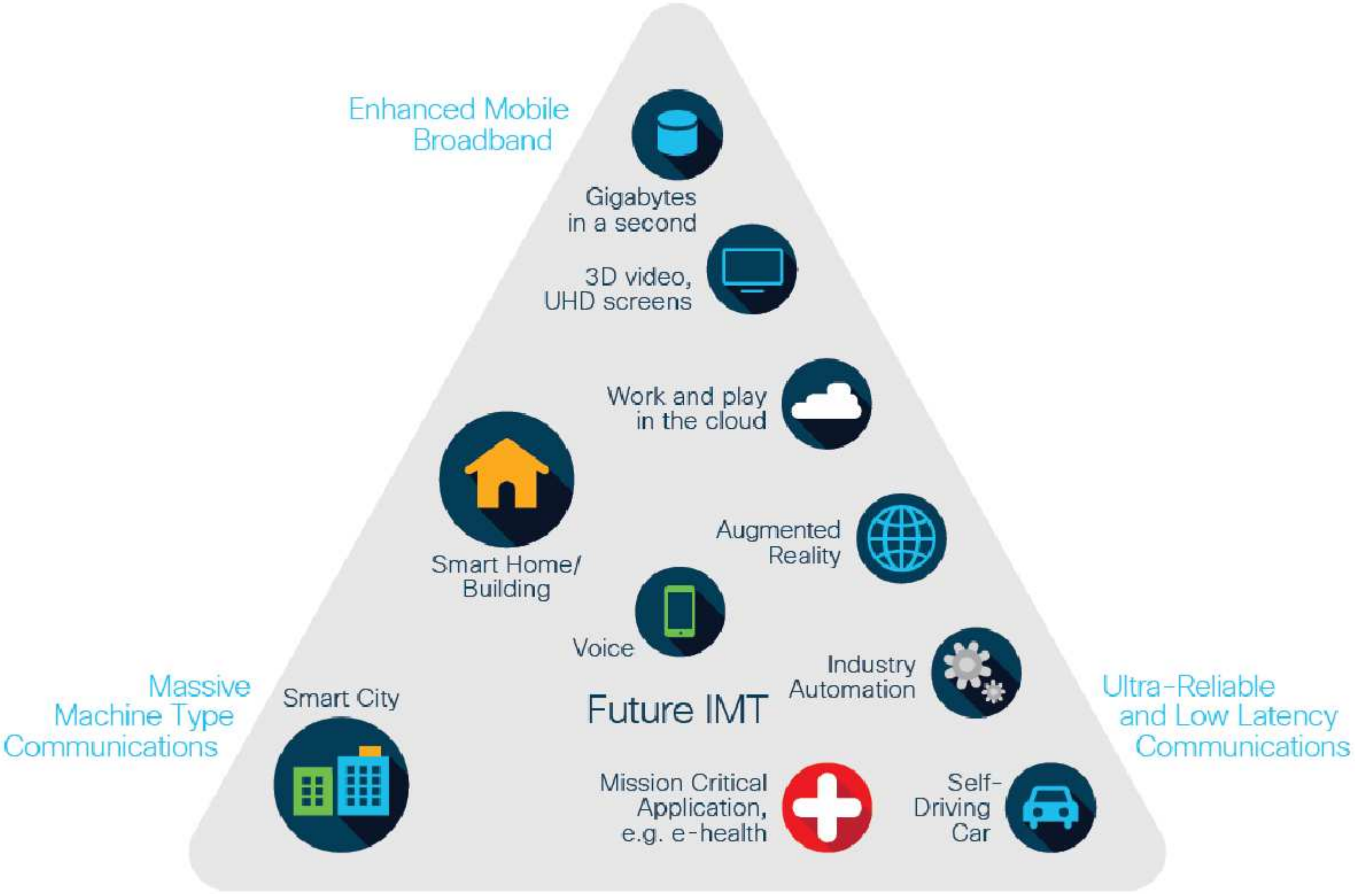

Enhanced Mobile Broadband
Gigabytes in a second
3D video, UHD screens
Work and play in the cloud
Smart Home/ Building
Augmented Reality
Voice
Industry Automation
Massive Machine Type Communications
Smart City
Future IMT
Ultra-Reliable and Low Latency Communications
Mission Critical Application, e.g. e-health
Self-Driving Car

Figure 1-13. 5G Use cases [13]

With the capabilities of 5G, we can envision a wide variety of use cases. 5G is scalable and was designed to support several use cases and scenarios

5G has been designed to support three main service requirements or use cases, described below and illustrated in Figure 1-13.

1. **eMBB: enhanced mobile broadband**

   This use case will address the limitations of LTE by providing a higher data capacity and offering more traffic capability. These applications are used where a faster connection is imperative to support the desired performance or need. Highly interactive applications cover a range of use cases, from streaming and high-definition video to 360-degree video and other data and video-intensive cases. For example, these include gaming apps, professional flight simulation and training, and precision medical applications. In addition, 5G will provide enhanced mobile broadband services to support the users' needs in public transit, as well as new applications, such as virtual reality (VR) and augmented reality (AR) applications. [14]

2. **Critical Communication, CC /URLLC: ultra-reliable low-latency communication**

   These require strict capabilities such as low latency, reliability, availability and instantaneous communication. Examples include the wireless control of industrial manufacturing or production processes, remote medical surgery, distribution automation in a smart grid, transportation safety, etc. [15]

3. **mMTC: massive machine-type communication (IoT)**

This involves wireless communication between machines without direct human interaction. 5G will be able to connect massive numbers of embedded sensors placed virtually anywhere seamlessly. This encompasses services characterized by many connected devices typically transmitting a relatively low volume of non-delay sensitive data. However, the critical challenge is that devices are usually required to be low-cost; and have a very long battery lifetime. Critical examples for this service type would be logistics applications (e.g., tracking tagged objects), smart metering, or agricultural applications where small, low‐cost and low‐power sensors are sprinkled over large areas to measure ground humidity, fertility, etc. [15]**.** Massive machine-type communications (mMTC) and the massive internet of things (IoT) both imply the aggregation and analysis of data from large numbers of connected devices [16]**.**

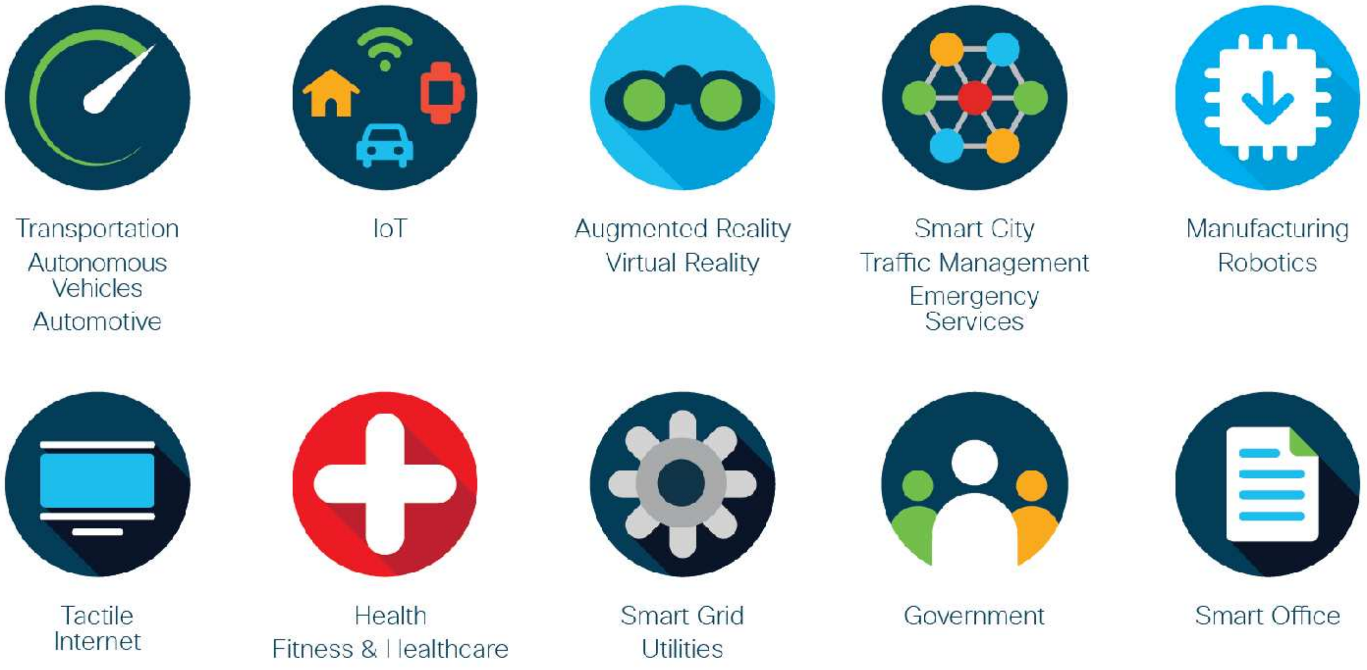


Figure 1-14. Some more use cases of 5G [13]

As shown in Figure 1-14, new vertical use cases are the main driver for deploying 5G networks. These use cases are anticipated to depend on network slicing, network function virtualization, Software-defined networking and Cloud Computing technologies. [17]

1.5.3 The 5G System Architecture

In the SA deployment option, the 5G System (5GS) is composed of the User Equipment, the Access Network (including the “New Radio” or NR) and the Core Network (5GC or 5GCN).

The 5GC architecture relies on a so-called "Service-Based Architecture" (SBA) framework, where the architectural elements are known as "Network Functions" (NFs) rather than "traditional" Network Entities. Via interfaces of a common framework, any given NF offers services to all authorized NFs and any "consumers" permitted to use these provided services. Such an SBA approach offers modularity and reusability. [18]

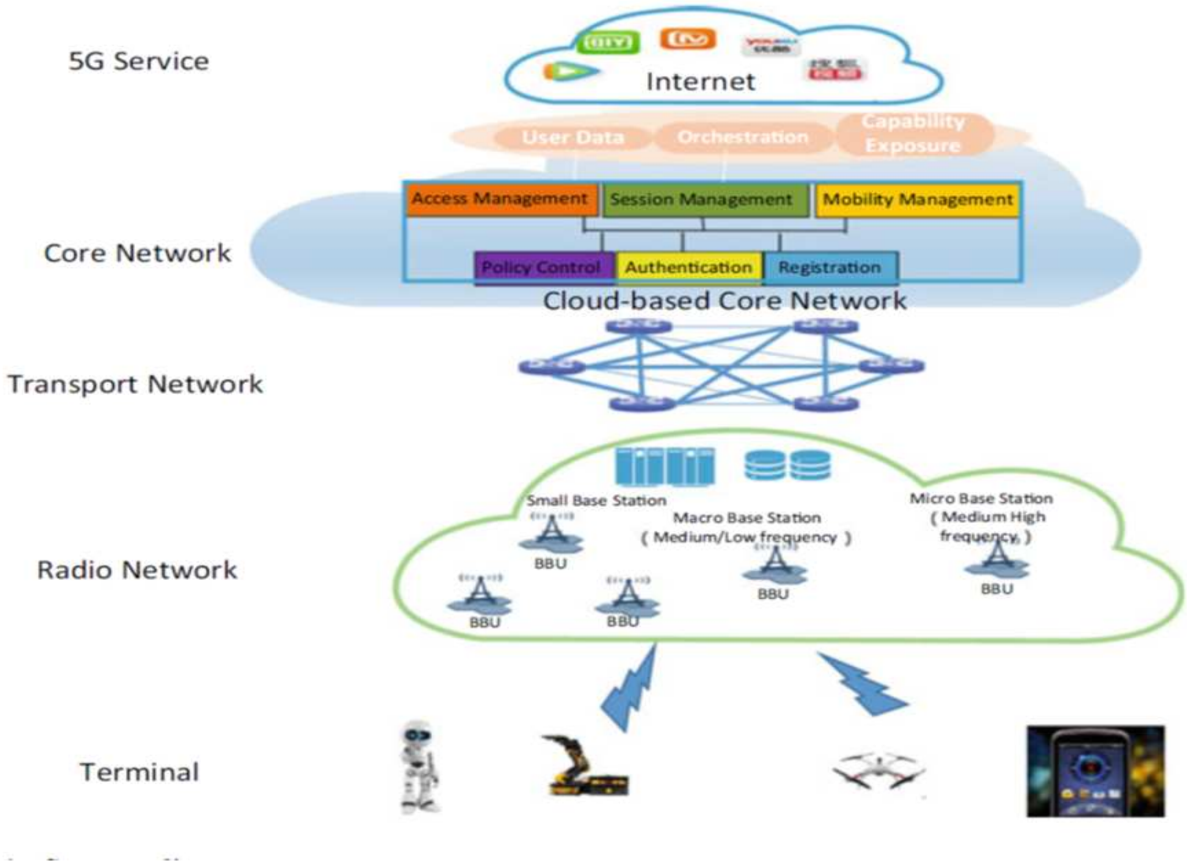


Figure 1-15. 5G overall system architecture [2]

1.5.4 THE 5G RAN ARCHITECTURE

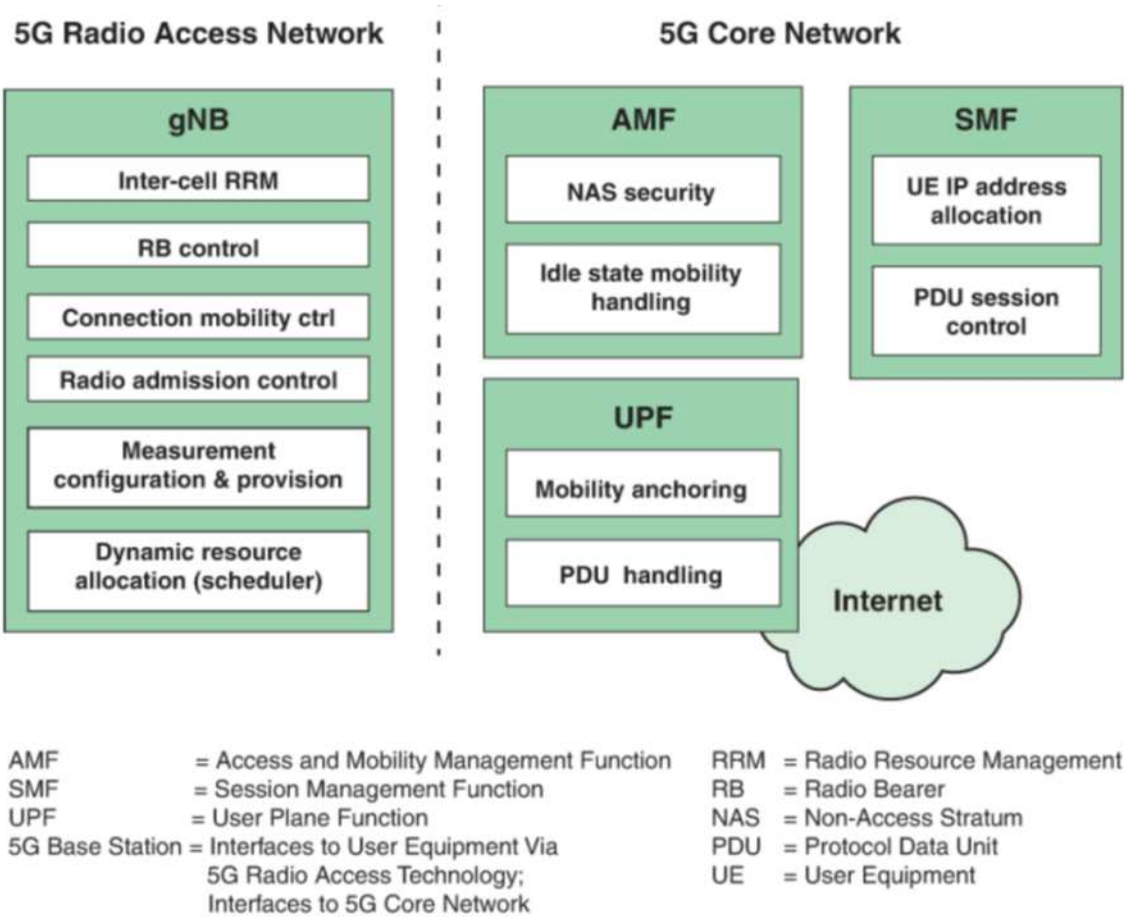

5G Radio Access Network
5G Core Network
gNB
Inter-cell RRM
RB control
Connection mobility ctrl
Radio admission control
Measurement configuration & provision
Dynamic resource allocation (scheduler)
AMF
NAS security
Idle state mobility handling
SMF
UE IP address allocation
PDU session control
UPF
Mobility anchoring
PDU handling
Internet
AMF = Access and Mobility Management Function
SMF = Session Management Function
UPF = User Plane Function
5G Base Station = Interfaces to User Equipment Via 5G Radio Access Technology; Interfaces to 5G Core Network
RRM = Radio Resource Management
RB = Radio Bearer
NAS = Non-Access Stratum
PDU = Protocol Data Unit
UE = User Equipment

Figure 1-16. SPLIT NG-RAN and 5G Core network [7]

The various functions of the gNB have been elaborated upon according to [7]

- Inter-cell radio resource management: The UE can detect neighbour cells, request information about the best-serving cell, and support handover decision-making by providing measurement feedback.
- Radio bearer control (RBC): for configuration (e.g. security), initializing and maintaining the radio bearer (RB) on uplink and downlink with different quality of service (QoS). Radio bearer refers to a transmission path of definite capacity, delay, bit error rate, and other parameters.

- Connection mobility control (CMC): Functions in UE idle and connected modes. In idle mode, the UE does not have an established connection until it is started. Then, the UE is switched on to establish a link in connection mode. The CMC performs cell selection and reselection in idle mode. The connected method involves handover procedures triggered based on the outcome of CMC algorithms.
- Radio admission control (RAC): performs new radio bearer admission request admission or rejection. It is done to optimize resource usage while maintaining the QoS of existing user connections.
- Measurement configuration and provision: provisioning and designing of UE radio source management procedures such as cell selection and reselection and requesting measurement reports to improve scheduling.

1.5.5 DEPLOYMENT OPTIONS FOR 5G

One unique capability of 5G is its forward compatibility meaning it can support future technology unavailable today. Another key characteristic of 5G is that the 5G Access Network can connect to a new 5G Core Network and the 4G (LTE) Core Network. This is known as the NSA architecture, while the 5G AN connected to a 5G CN is called the SA architecture. Thus, two deployment options are available for 5G Network infrastructure, which is gradually evolving from a non-standalone architecture (NSA) that works with the support of 4G LTE resources to a standalone architecture with 5G operating independently. There will also be a need to develop new devices that will support the high-speed capability of 5G. Network hardware and software updates will also have to be done to accommodate the expansion and evolution of 5G. The full deployment will be a relatively slower process in the commercial, industrial and government spaces than in the consumer space. [18]

1.5.5.1 THE 5G NON-STANDALONE ARCHITECTURE (NSA)

The NSA architecture is a temporary or initial step toward full-scale 5G deployment, where the 5G Access Network is connected to the 4G Core Network.

In the Non-Standalone (NSA) architecture, the 5G Radio Access Network (AN) and its New Radio (NR) interface are designed and implemented with an LTE and EPC infrastructure Core Network (a 4G Radio and 4G Core). The 4G services can leverage this design’s 5G New Radio (lower latency, etc.) capacity. The NSA is also known as "E-UTRA-NR Dual Connectivity (EN-DC)" or "Architecture Option 3". [18]

The two main types of nodes in the RAN are;

- gNB: Provides 5G user plane and control plane protocol terminations toward the UE.

- ng-eNB: Provides 4G (E-UTRA) user plane and control plane protocol terminations toward the UE and connects via the NG interface to the 5G core. 5G networks can support a UE that uses the 4G air interface. The UE uses 5G protocols to interact with the 5G core network. The eNB can work in EN-DC (dual connectivity between E-UTRA of 4G and New Radio of 5G) configuration; so that the eNB can communicate with the en-gNB through the known X2 interface (X2-C/X2-U), the same used to interconnect two eNBs. In this case, the eNB becomes the Master Node (MN), while the en-gNB acts as the Secondary Node (SN). The two Base Stations, eNB and en-gNB, can provide the User Plane protocols, while the eNB provides the Control Plane function for communication with the UE. Option 3-NSA expands the traffic capacity but has no support for network slicing functions. Thus, URLLC cannot be implemented. [1]

The Non-standalone is a gradual deployment process or steps towards full 5G deployment and operation. Here we are harnessing the existing 4G infrastructure to provide 5G services. The non-standalone architecture helps operators already running 4G LTE to augment their network with 5G, leveraging its high-speed capability for their mobile networks without having to change their existing infrastructure completely.

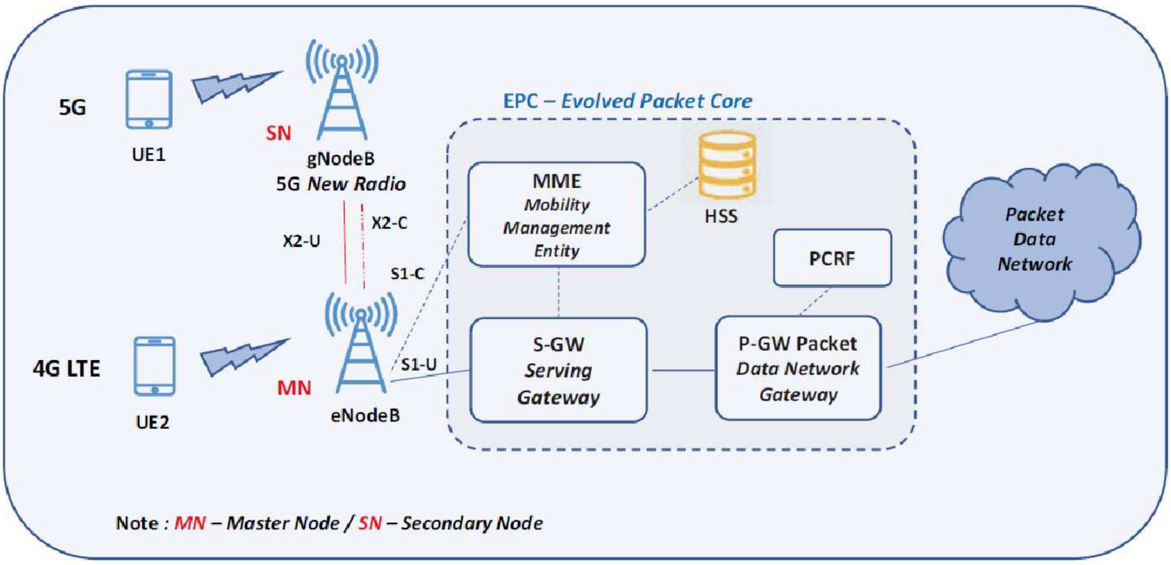


Figure 1-17. 5G Non-standalone architecture [1]

1.5.5.2 5G STANDALONE ARCHITECTURE

In 5G standalone architecture, the New Radio (NR) is connected to the 5G Core Network. This configuration; supports the complete set of 5G Phase 1 services.

The SA architecture does not need any part of the 4G infrastructure. It is a full-scale deployment of 5G infrastructure.

Option 2-SA is used to serve 5G deployment from scratch, especially where there is no existing 4G LTE infrastructure or legacy technology to leverage on. The entire installation is done according to 5G SA standards. This system can provide all expected services, including traditional services and the three fundamental types of new services – eMBB, mMTC, and URLLC. [1]

Option 3

In Option 5 – SA, the LTE Base Station is connected to the 5G Core (5GC) through the NG-C and NG-U interfaces (Next Generation Control Plane and User Plane). In this option, the EPC is replaced by the 5GC. For this to be possible, the eNodeB software must be modified and upgraded; to be able to interoperate with 5GC in such a way as to become an ng-eNB (next generation – eNB). This will allow the provision of differentiated services of Network Slicing. Still, the system will not be able to benefit from the advantages of the 5G NR (5G New Radio) air interface, such as multiple Numerology, already explained before and which provides greater spectral efficiency. The ng-eNB is, therefore, still an LTE Base Station. Therefore, this is an alternative that is unlikely to be applied." [1]

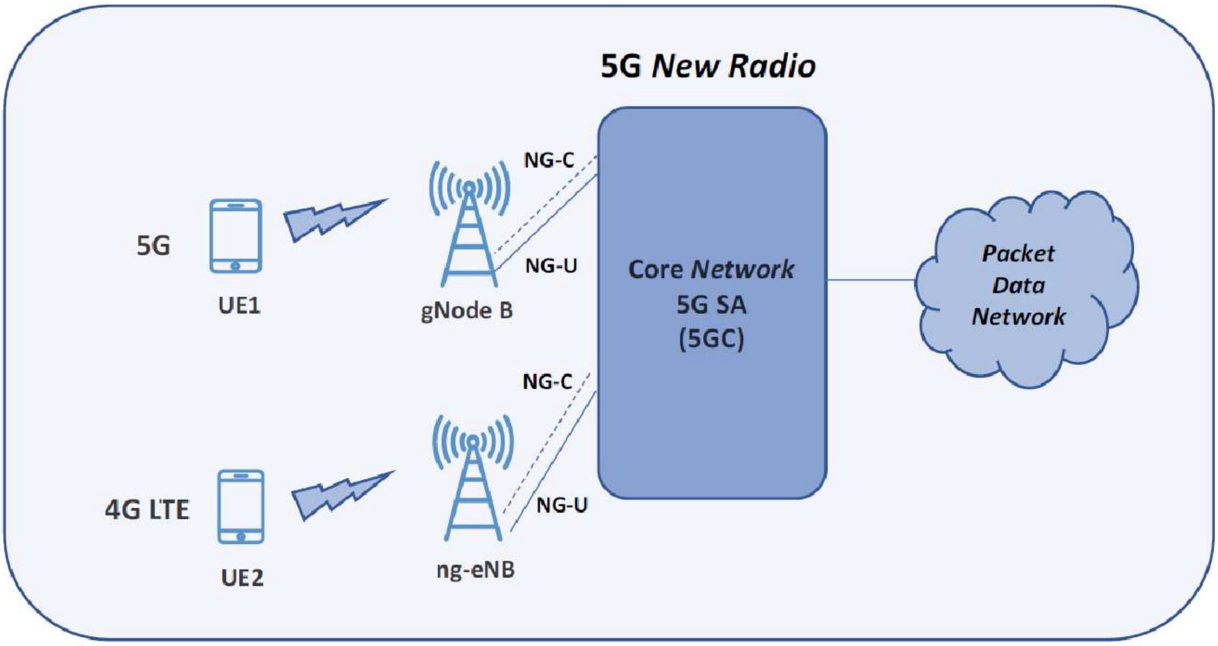


Fig 5G final standalone architecture [1]

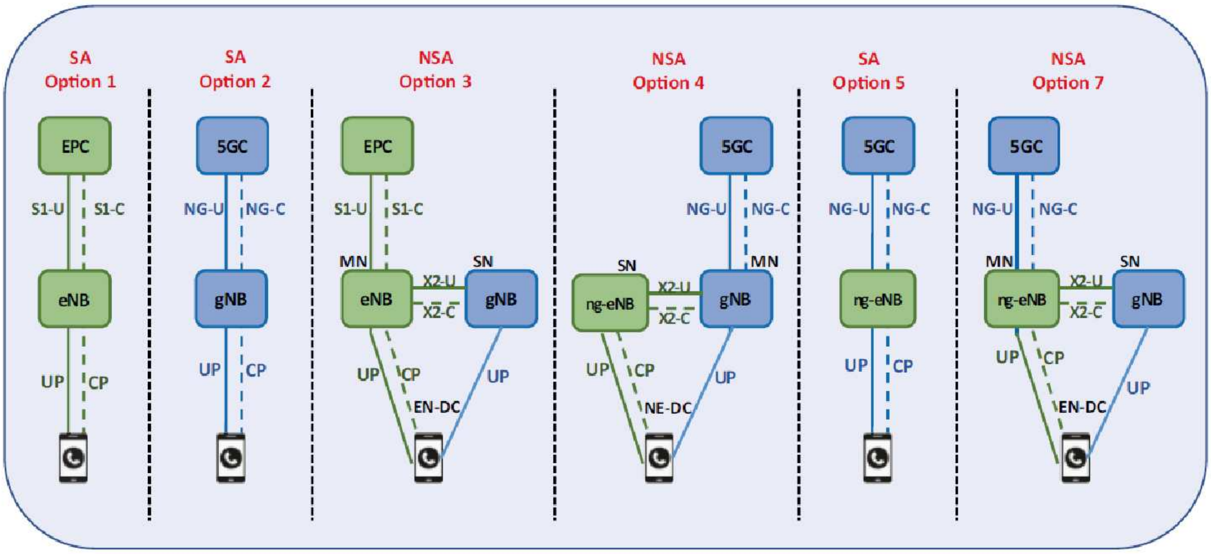
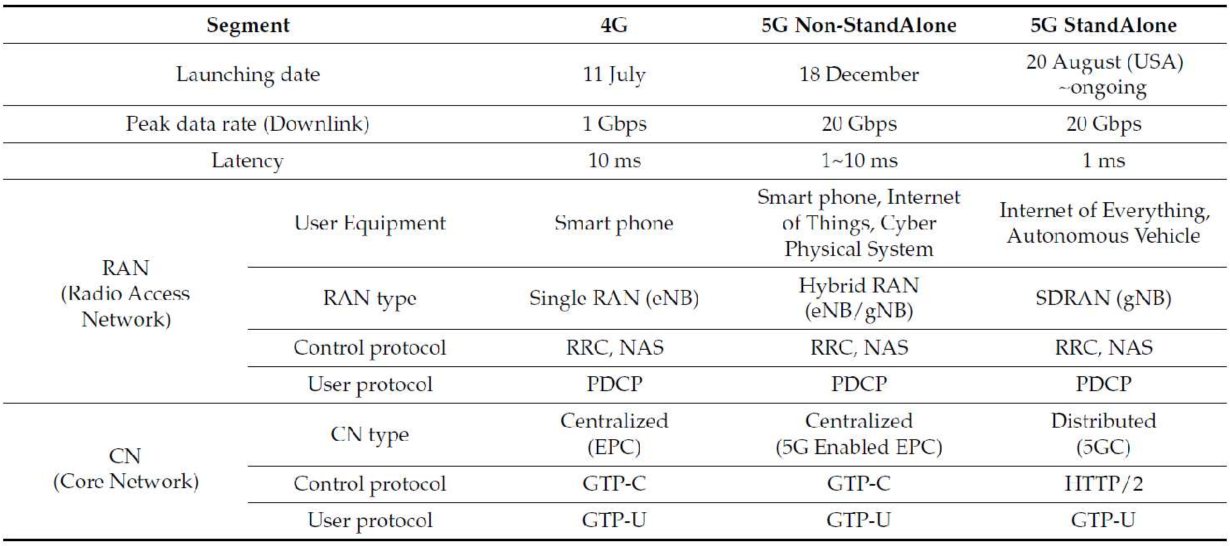


Fig Options envisaged for redesigning LTE and 5G networks [1]

1.5.6 How 4G mobile technologies compare to 5G technologies.

The table below compares 4G, 5G NSA and 5G SA technologies.

Table 1-5. A comparison of 4G and 5G technologies [19]



Segment		4G	5G Non-StandAlone	5G StandAlone
Launching date		11 July	18 December	20 August (USA) ~ongoing
Peak data rate (Downlink)		1 Gbps	20 Gbps	20 Gbps
Latency		10 ms	1~10 ms	1 ms
RAN (Radio Access Network)	User Equipment	Smart phone	Smart phone, Internet of Things, Cyber Physical System	Internet of Everything, Autonomous Vehicle
	RAN type	Single RAN (eNB)	Hybrid RAN (eNB/gNB)	SDRAN (gNB)
	Control protocol	RRC, NAS	RRC, NAS	RRC, NAS
	User protocol	PDCP	PDCP	PDCP
CN (Core Network)	CN type	Centralized (EPC)	Centralized (5G Enabled EPC)	Distributed (5GC)
	Control protocol	GTP-C	GTP-C	HTTP/2
	User protocol	GTP-U	GTP-U	GTP-U

1.5.7 Key enabling technologies of 5G core networks

These key technologies of the 5G core network, according to [2], have been designed to satisfy four main principles and concepts. These are;

(1) IT-based: A cloud computing infrastructure, the representative IT technology, such as functional software, computing and data separation, are introduced. It is transformed from proprietary network devices to cloud "Network Function."

(2) Internet-based: A more flexible network design has been introduced. In addition, the service-based architecture and the new core network protocol architecture based on HTTP/2.0 Internet protocol are introduced.

(3) Minimal-based: the minimal architecture and functional design; to improve the performance of data forwarding and the flexibility of network control as much as possible.

(4) Service-based: A more personalized service design to meet various needs of different vertical industries, and the core of its technology is to realize "Network as a Service." Through network

slicing, edge computing, low latency significant connection and so on, the transformation of the network from universal service to personalized and customized service is realized. [2]

1.5.7.1 Software-Defined Networking

While NFV separates the hardware and the software in a communication network, SDN is implemented by separating the data plane functions from the control plane functions to facilitate network management. It is a new approach to designing, building, and operating large-scale networks based on programming the forwarding decisions in routers and switches via software from a central server. Thus, devices such as routers, switches, LANs, etc., have data and control functions separated by an application programming interface (API). It allows for flexibility and programmability of the 5G network design. [7]

It was initially deployed in the campus and data network centers.

Since most network hardware devices encompassed data and control functions, it took much work to modify the network infrastructure and operation to large-scale addition of end systems, virtual machines and networks.

By introducing SDN technology to 5G, Network operators can easily introduce new services while quickly responding to the network's rapid resource changes and QoS requirements. This used to be more time-consuming. With the Open Flow design of SDN, network operators do not need to configure individual network components like routers and switches from different vendors anytime there is an upgrade or implementation of a new design or network behaviour.

Being adopted by 5G, SDN will enable a more agile and flexible core network architecture. In addition, the SDN's programmability and openness; will help mobile operators shorten the life cycle of introducing their new services and innovation into markets. By control and data plane separation, the network infrastructure can be constructed on demand and the basis of service requirements (network-as-a-service), thus improving resource efficiency. It is worth noting that the SDN concept can also be used in the RAN domain, where the SDN controller could control and schedule the radio resources for base stations, thus improving the spectrum efficiency as well as mobility management [9]

There are three logical planes [9]

1) Application plane: consists of applications for various network functionalities such as network management, QoS management and security services, etc.

2) Control plane: is the logically centralized network control platform running the Network Operating System (NOS), having a global view of the network resources and stats and providing hardware abstractions to the applications in the application plane.

3) Data plane: also called the data plane, which consists of the data forwarding elements acting on the instructions of the control plane for dealing with the data packets or traffic flows.

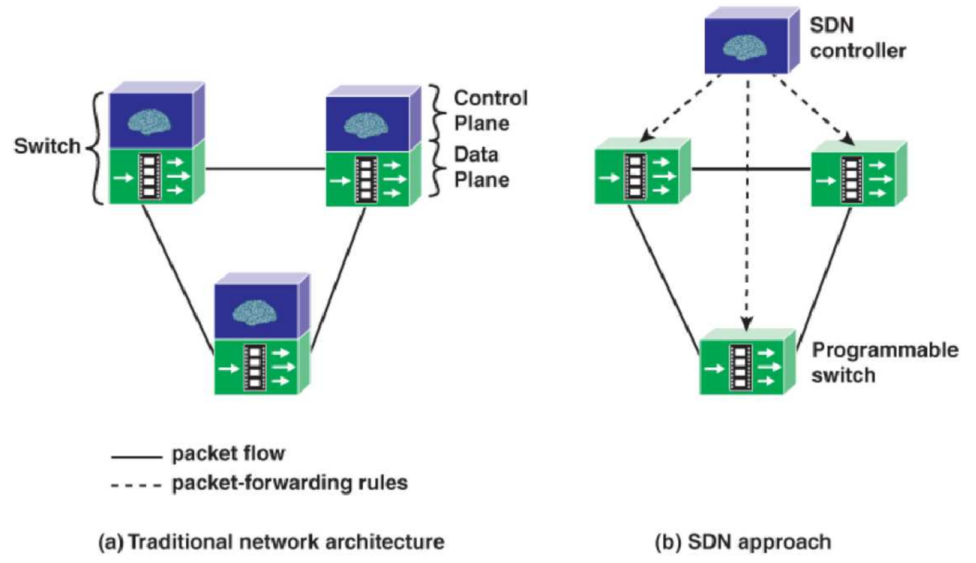


Figure 1-18. Traditional networking approach vs the SDN approach [7]

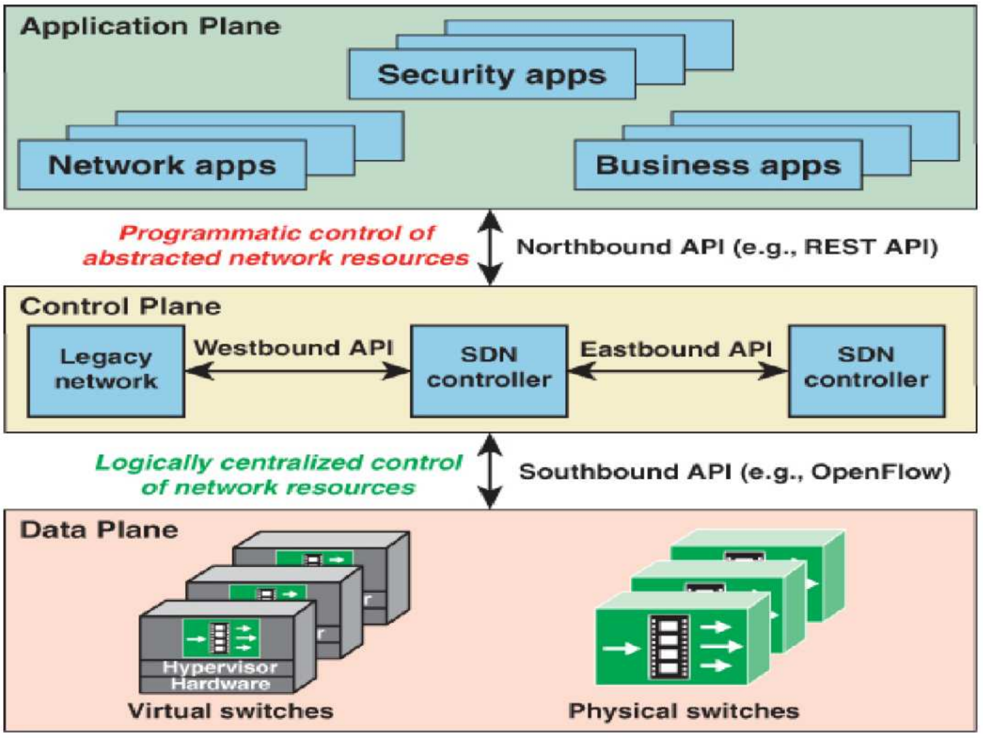


Figure 1-19. A simple architecture of software-defined networking [7]

1.5.7.2 Network Function Virtualization

In Network Function Virtualization, the network functions of the 5G core network are substituted with virtual machines in place of hardware components. In other words, storage and network functions are virtualized by implementing these functions in software and running them on virtual machines. Virtualization encompasses a variety of technologies for managing computing resources by providing a software translation layer between the software and the physical hardware. Virtualization turns physical resources into logical or virtual resources. Virtualization enables users, applications, and management software operating above the abstraction layer to manage and use resources without needing to be aware of the physical details of the underlying resources. The NFV approach moves away from dependence on a variety of hardware platforms to the use of a small number of standardized platform types, with virtualization techniques used to provide the needed network functionality. [7]

The issue with hardware components is that there are ever-increasing varieties of hardware coupled with their software. Suppose there is a new service or function; there has to be network reconfiguration or onsite installation of new systems to satisfy the unique service requirements, which may require more physical space or construction of sites and extra trained human resources. In addition, the telecom world is a fast-evolving digital world with fast innovation cycles that require greater flexibility and dynamism than hardware-based appliances allow. A hard-wired network with single-function boxes is tedious to maintain, slow to evolve and prevents service providers from offering dynamic services.

Fully automated, virtualized network functions allow networks to be agile and capable of responding automatically to the needs of the traffic and services running over them. [20]

By implementing network functions and components with software, NFV brings several benefits: the hardware is generic and cheaper; the software is easier, faster and more affordable to write and upgrade; the software and hardware can be uncoupled from one another, with the two procured from different vendors; and it is easier for new vendors to enter the market. [16]

The design of NFV allows for easier and faster setup, modification and tear down of individual network functions in a management system. [16]

The significant components of an NFV architecture are the virtualized network functions (VNFs), NFV Infrastructure (NFVI) and NFV management and orchestration (MANO).

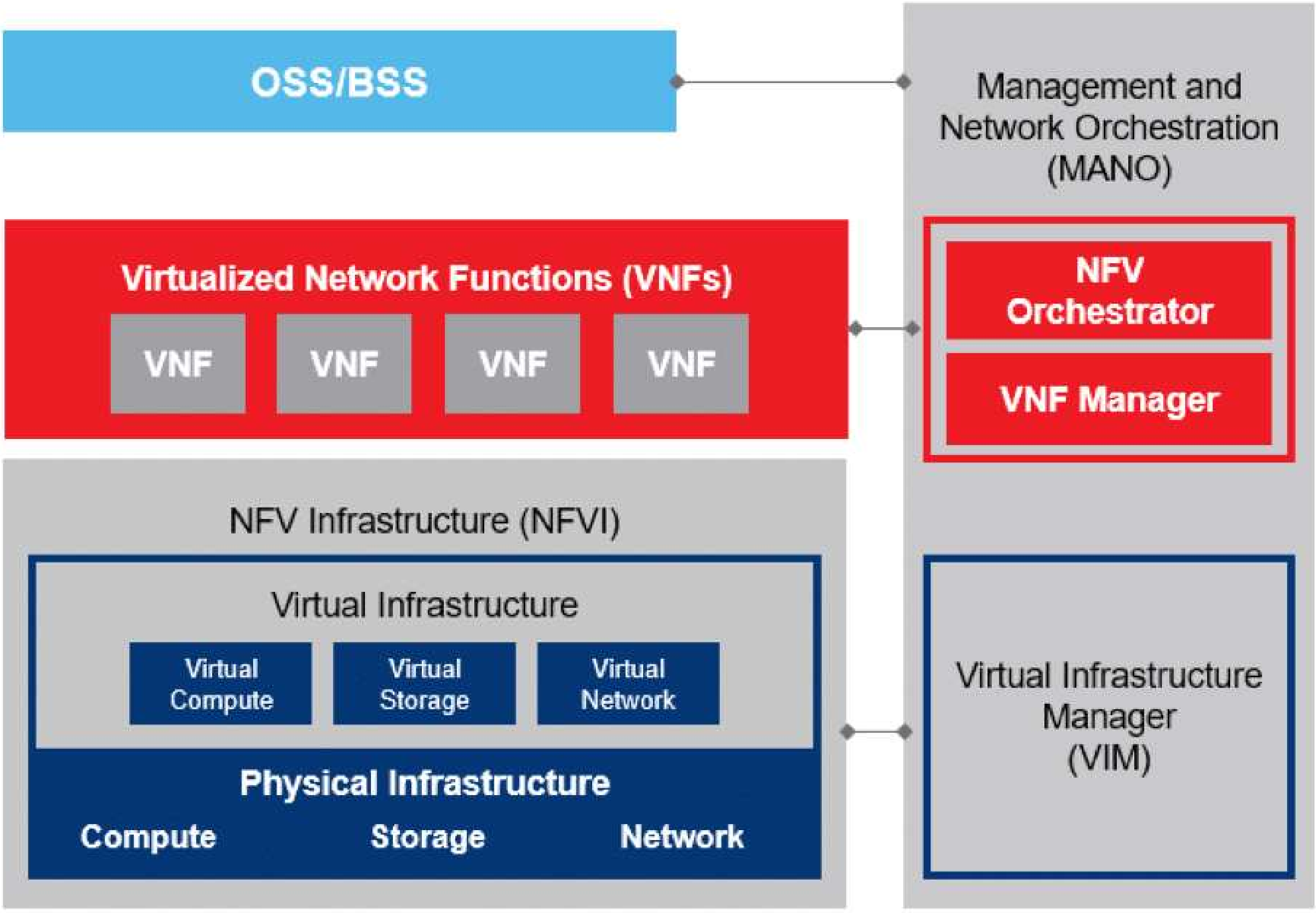


Figure 1-20. An NFV architecture [21]

Network Functions Virtualization Infrastructure

It provides the virtualization and abstracts hardware resources to be logically partitioned and provisioned to support VNFs. [21]

Virtual Network Functions

These functions are run on one or more virtual machines on the hardware networking infrastructure. VNFs can be routers, switches, SD-WAN, firewalls and a growing number of other network services now available as software from vendors like Cisco, Juniper Networks and Palo Alto Networks.

With an NFV architecture, VNFs are deployed based on demand, drastically reducing the deployment delays associated with traditional network hardware and the need for on-site technical skills when remotely deployed. In addition, VNFs provide the agility to quickly determine or adjust to dynamic network performance or expansion demands in hybrid and multi-cloud environments. [21]

NFV Management and Network Orchestration

NFV management and network orchestration (MANO) is a framework developed by a European Telecommunications Standards Institute (ETSI) working group. From initial to quotidian operations, NFV MANO coordinates resources—the NFVI and VNFs—running in a virtualized data center, including computing, networking, storage and virtual machines (VM). NFV MANO uses templates for standard VNFs that allow network architects to select the right NFVI resources to be deployed. [21]

NFV MANO is comprised of three functional areas:

NFV Orchestrator for VNF onboarding, lifecycle management, global resource management and validation and authorization of NFVI resource requests.

VNF Manager controls VNF lifecycle management of instances, providing a coordination and adaptation role for NFVI and Element/Network Management Systems configuration and event reporting.

Virtual Infrastructure Manager controls and manages the NFVI compute, storage and network resources.

1.5.7.3 Cloud Computing

Cloud computing enables the movement from a hardware-based mode to a software-based mode. With the cloud, instead of having resources located on-site or on-premises, these resources are hosted on the internet virtually. The resources are put together in resource centers called data centers. [22]

We can expand and reduce cloud resources according to our specific service network requirements. For example, when server resources are needed for a particular task, they could be allocated to that task for a given time and reallocated elsewhere for another purpose instead of having several pieces of hardware with their resources for each task.

C With its anticipated benefits, cloud computing has been considered an ideal solution for re-designing the current RAN architecture and a key enabler for NFV and SDN technologies Such as on-demand and elastic provisioning of services and resources over the internet; cloud computing has made it one of the critical enablers for designing 5G core networks. In this case, 5G core network functions will be realized as virtual machines or containers controlled by the cloud manager.

Some of the significant benefits of cloud computing are that;

It is agile, with accessible innovation and implementation of new technologies faster than traditional methods. You can quickly scale up or scale down depending on the needs of the network;

Cloud computing ensures that there is rapid elasticity with the fact that you can allocate the resources that are needed at a given time or for a specific task or service;

With the cloud, we can have a measured on-demand service where resources are controlled, managed and optimized based on the client’s needs. In addition, new services can quickly be deployed without physically working on hardware devices.

Cloud computing can help mobile operators reduce their capital and optimize operational expenses. Cloud computing was created to virtualize the commodity IT hardware, and many cloud technologies such as OpenStack1 or VMware2 are serving as the resource backend for virtual network functions.

Though centralizing all resources will ease the management and provisioning process, it will result in a long delay in end‑to‑end communication, which may not be suitable for some of the newly defined 5G services.

Therefore, combining cloud computing and other computing technologies, such as mobile cloud computing, fog and edge computing, will be a significant step in efficiently deploying mobile networks like 5G. (Madhusanka Liyanage, 2018) [23]

Types of Cloud computing

Infrastructure as a Service (IaaS)

Made up of the basic building blocks for cloud IT. It typically provides access to networking features, computers (virtual or on dedicated hardware), and data storage space. IaaS offers the best flexibility management and controls over your IT resources. [24]

Platform as a Service (PaaS)

PaaS removes the need to manage underlying infrastructure (usually hardware and operating systems) and allows you to focus on deploying and managing your applications. This helps you be more efficient as you don’t need to worry about resource procurement, capacity planning, software maintenance, patching, or any other undifferentiated heavy lifting involved in your application. [24]

Software as a Service (SaaS)

SaaS provides you with a complete product that is run and managed by the service provider. In most cases, people referring to SaaS refer to end-user applications (such as web-based email). With a SaaS offering, you don't have to consider how the service is maintained or the underlying infrastructure is managed. You only need to consider how to use that particular software [24].

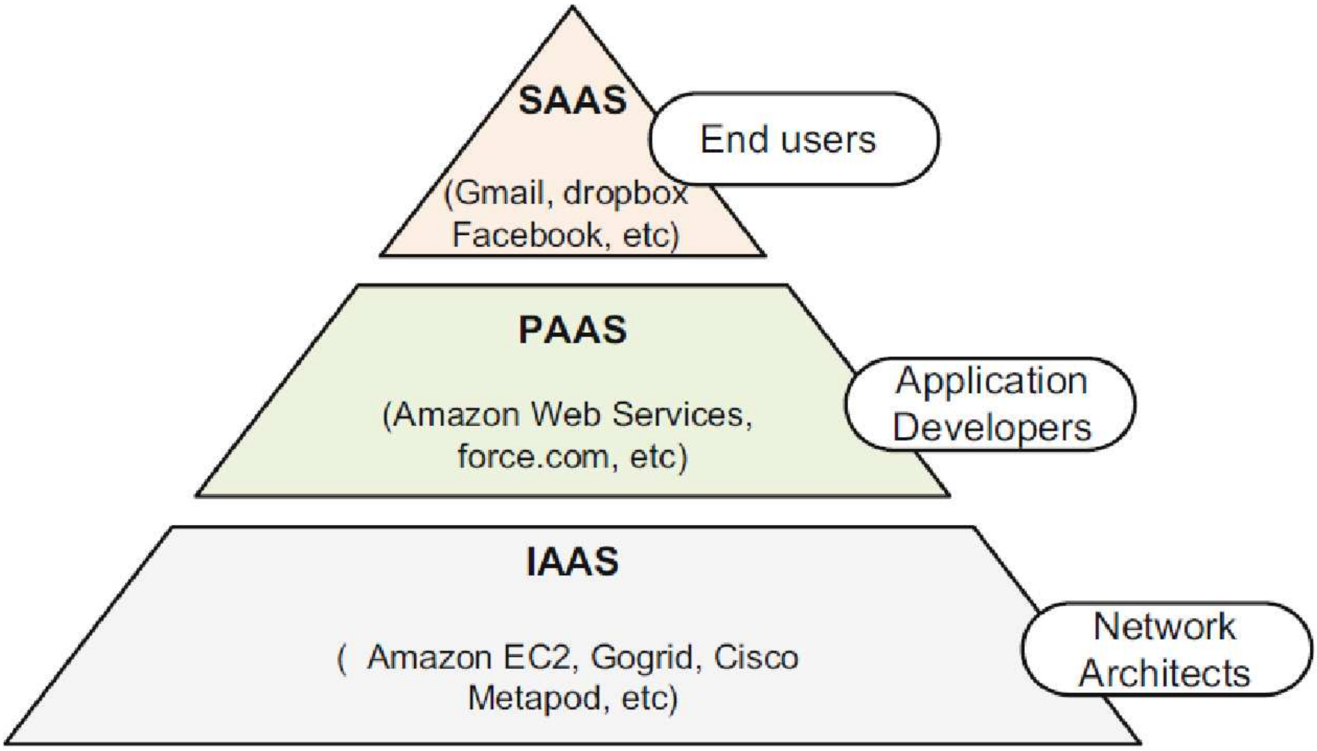


Figure 1-21. Cloud computing models [25]

1.5.7.4 Network Slicing

Network Slicing is a crucial enabler of 5G core networks because we can cater to various applications with diverse network requirements. Furthermore, the physical infrastructure can be divided into multiple virtual networks through Network Function Virtualisation (NFV) and Software Defined Networks.

Network slicing is one crucial technology which gives a 5G network the unique capability of logically sharing or slicing network resources while providing specific network capabilities and characteristics to serve a defined business need of a customer. This unique capability of slicing the 5G core network is revolutionary in providing an enabling network environment for different market verticals, e.g., automotive, healthcare, critical infrastructure, etc. Network slices can be dedicated to these other purposes, ensuring a specific application or service gets priority access to capacity and delivery or isolating traffic to particular users or device classes. Each slice of these networks can have its logical topology, security rules and performance characteristics -- within limits imposed by the underlying physical networks, thereby maximizing network resources while ensuring service flexibility. [26] [27].

Network slicing will be tackled in detail in Chapter 2.

1.5.7.5 Multi-access Edge Computing (MEC)

MEC uses the Control and user plane separation (CUPS) architecture of 5G placing the user plane function and applications closer to the network edge, thereby enabling use cases requiring low latency services. Control Plane User Plane Separation (CUPS) architecture allows the distribution of the core elements, leading to greater function utilization efficiencies and placing the network functions into the network where most appropriate given service constraints of an SLA, for example. CUPS also allows efficient network function scalability based on traffic dimensioning [13].

5G enables operators to host latency-critical services closer to the end user through the Multi-access Edge Computing (MEC) capability. MEC helps to reduce the traffic load on the transport network and meet the Service Level Agreements (SLA) of an ultra-low-latency service by local hosting. (Rahim Tafazolli, 2021). Multi-Access Edge Computing (MEC) transforms how data is processed and stored by moving some core network functions closer to the end user at the network edge rather than relying on a central location that may be hundreds of miles away. However, introducing untrusted 5G components into the MEC could expose core network elements to risks presented by software and hardware vulnerabilities, counterfeit parts, and component flaws caused by poor manufacturing processes or maintenance procedures [28].

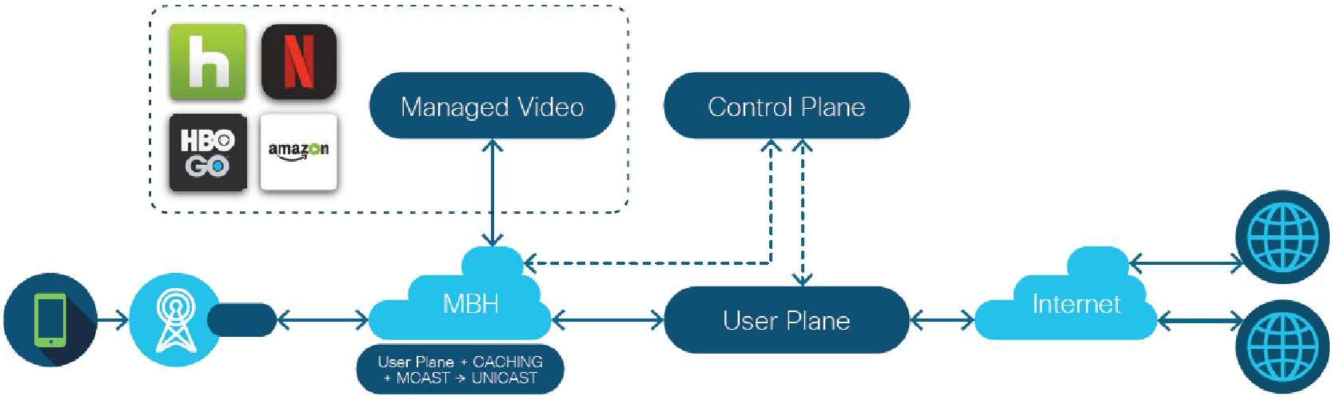


Figure 1-22. An MEC architecture [13]

Some other key enabling technologies of 5G include,

1.5.7.6 Device-to-Device Communication

Device-to-device (D2D) communication, another key enabling technology of 5G, has many benefits ranging from coverage expansion, power management, spectrum efficiency, improving user data rate and capacity and communication latency with the reuse of radio resources which allows network functions to devices that also provide services such as safety, traffic offloading, and location-based proximity services. D2D communication does not depend on the base station for functionality. It involves direct communication among two or more cellular users. In other words, a D2D pair can directly communicate without routing its traffic through the central base station. D2D communication allows Peer to Peer or direct device-to-device communication eliminating base station or IP-based oriented connectivity. D2D communication can use an unlicensed ISM band or licensed cellular band (as in LTE Direct), which can provide security services and resource management from the cellular network [29] [25].

According to [29], D2D communication is divided into different types based on intervention with network control from infrastructure, such as autonomous, network-controlled, and network-assisted D2D.

• In autonomous D2D, devices in the network establish links and communicate with each other in a fully distributed manner similar to ad-hoc; each device head handles all network functions identically to self-organizing networks.

• In network-controlled D2D, all devices are allowed for data communication only when the network is fully centralized,

- In network-assisted D2D, infrastructure supports network functions such as security, synchronization, and link management are supported by infrastructure.

D2D communication is said to have some challenges of interference management, resource allocation, and security, specifically where D2D and Heterogeneous networks coexist.

Implementing D2D in the existing cellular communication causes challenges with interference management and efficient distribution of resources.

These challenges could be mitigated by applying interference avoidance schemes like conflict graphs to reduce interference; and game theory for the contention to access the licensed spectrum. [25]

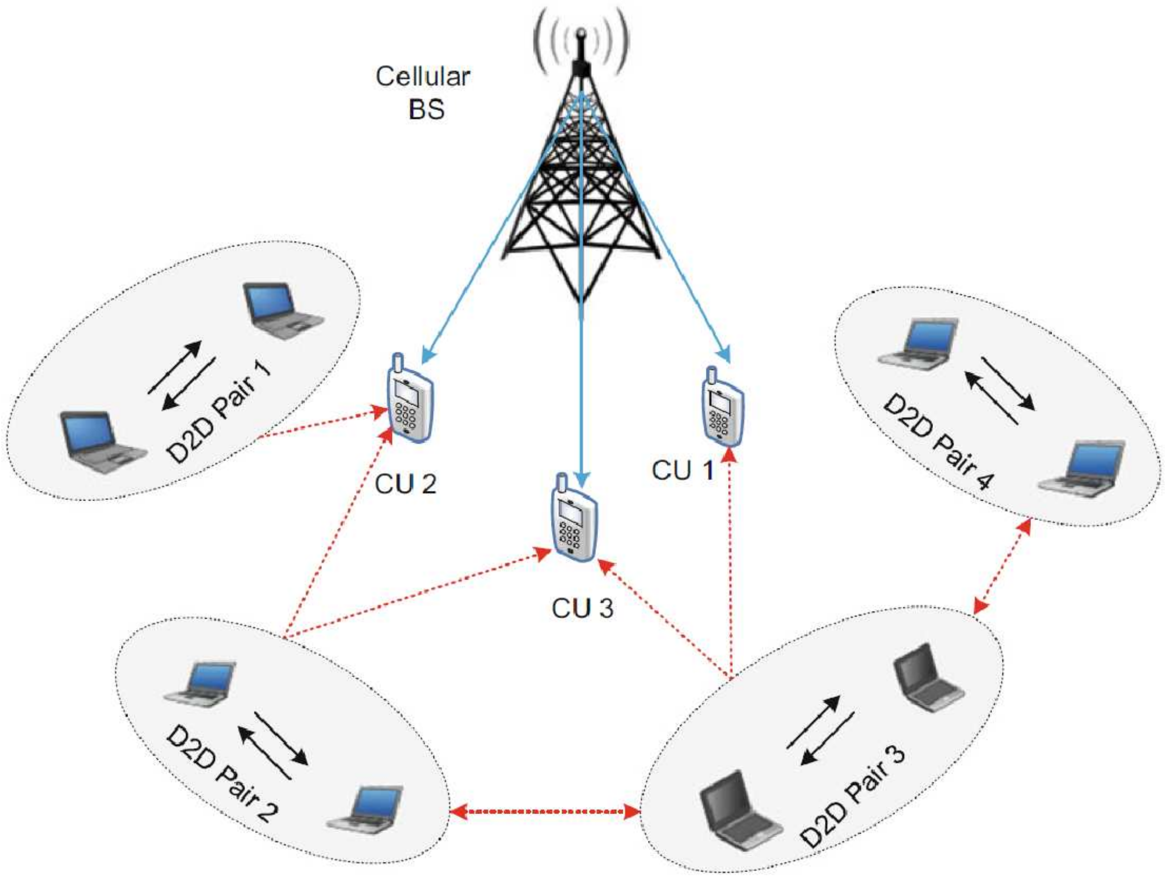


Figure 1-23. A downlink D2D communication network, according to [25]

1.5.7.7 Machine-to-Machine (M2M) Communication

Machine-to-machine communications require stringent latency attributes for efficient and successful usage. As a result, it is anticipated to increase in demand on the spectrum and the Internet significantly. Here, devices communicate not on behalf of humans but to address some need or function of the equipment supported by these networks. This might include automation in a factory setting, sensors detecting smog in a city, meters responding to queries, or network devices exchanging control information. For transmitting a small size of sensed data with time constraints, M2M communication is used.

According to spectrum resources, two types of Random Access Technologies (RAT), namely Lower Powered Wide Area Networks (LPWAN) and Cellular IoT, are used. The network capacity must be adequate to handle many simultaneous connections in a large coverage area. Data aggregation and offloading are used in M2M communication to enhance energy efficiency and communication. M2M Communication gives intelligent machines the autonomy to generate data, process, and transfer operations. Deployment using actuators, sensors, devices, and objects that works independently with trim or without human assistance is known as M2M communication. For example, sensors are used to record the occupancy of car parking spaces in real-time. In M2M communication in real time, researchers provided various mechanisms for supporting small data bursts to reduce the consumption of power and to avoid network congestion [29, pp. 7-8] [8]

1.5.7.8 Millimetre Wave (mm-Wave)

Mm-Wave in 5G is emerging as a crucial technology for next-generation mobile networks, significantly increasing network capacity and user experiences. Mm-Wave bands have been utilized for large bandwidths 30–300 GHz (1–10 mm wavelength), which supports Gigabit wireless services such as ultra-high-definition TV and high-speed internet access. The mm-Wave suffers from more path loss than microwaves, i.e., the power is reduced to low due to increased frequency. Secondly, the mm-Wave signal suffers from high absorption losses because of the rain and atmosphere. These are suited for line-of-sight communication, and if there are significant obstacles in their path, mm waves are highly vulnerable to blockages. To handle narrow beams effectively, mm-Wave systems contain high directional antennas in many arrays, which are also appropriate for short-range communication. 5G systems use both mm-Wave spectrum and microwave due to the limited spatial coverage of mm-Wave. Therefore, handling mm-Wave and microwave base stations and user equipment use separate signal processing components. Mm Wave signals are used in cellular access if the base stations are densely installed, enabling the highest data rates. The need for a high cost of transmitters and receivers and very high path loss is practically limited. In 5G, many small cells are overlaid on macro cells, each containing its base station. Interconnecting each with a fibre cable becomes much more expensive. Hence, the network can be organized with mm-Wave, which will be cost-effective. Mm-wave is used for high-speed WLAN and WPAN indoor services in macro cells. [29]

1.5.7.9 Massive Input/Massive Output (MIMO)

Short millimetre waves do not travel well through obstacles, so to improve communication, LTE started using MIMO (Multiple Input Multiple Output) antennas, which, together with OFDM technology, allowed to remarkably improve communication performance and minimize the probability of transmission errors by using space diversity and STC (Space Time Coding) [1]

This is an antenna system where the transmitter and receiver implement multiple antennas. The MIMO antenna architecture has become a crucial technology which enables high-speed wireless networks, including IEEE 802.11 Wi-Fi LANs and 4G and 5G, to improve wireless systems in terms of capacity, range, and reliability. The MIMO is based on the transmitter and receiver antenna beamforming, further improving performance and limiting interferences. Massive MIMO provides more sophisticated beamforming and management capability and narrower antenna patterns. Not only is beamforming valuable for high frequencies, but it can also form an essential base for many low-frequency scenarios to extend coverage and provide higher data rates. The MIMO antenna system also implements hundreds of antenna elements in a dynamic pattern to increase performance. To put things into perspective, the traffic volume of a cell operating with a 64-element Massive MIMO antenna with 100 MHz bandwidth can reach more than 300 Gbps or exceed 3000 Gbps in exceptional cases. This volume of data traffic will continue to grow dramatically as more massive MIMO antennas are used for 5G deployment. [1] [7] [3]

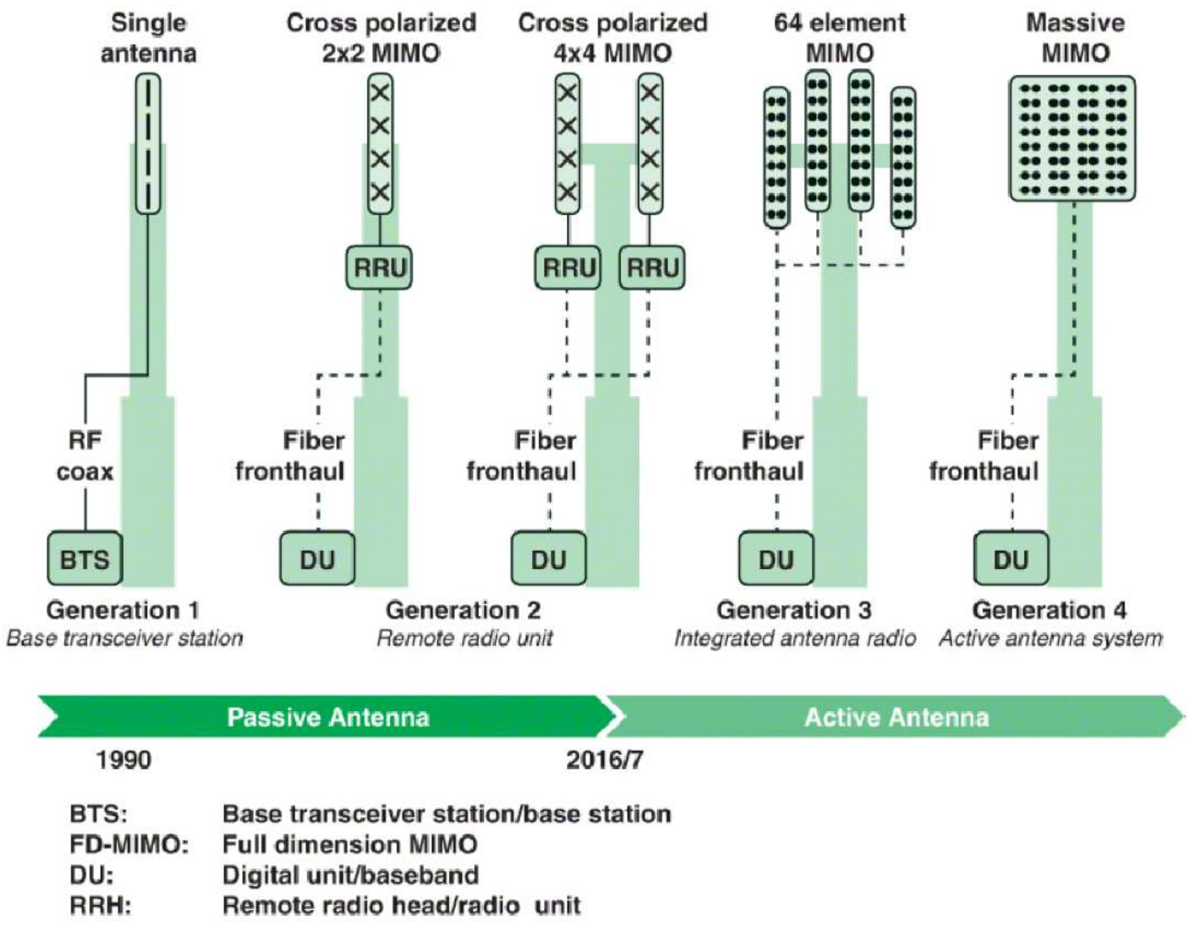


Figure 1-24. Evolution of antennas before Massive MIMO was implemented in 4G/5G [7]

1.5.7.10 V2X Communication

The advantage of higher traffic information systems, autonomous cars, and more reliable safety services used for the development of technology for vehicles with low latency, higher data rate, and reliability is known as vehicle-to-everything (V2X) communication which includes vehicle-to-pedestrian (V2P), vehicle-to-infrastructure (V2I) and vehicle-to-vehicle (V2V) communication. For enabling V2X communication, D2D communication for cellular networks is more suitable because D2D provides a long transmission range and short end-to-end latency. With high mobility, it can support road safety services, and all nodes move at the highest speeds in vehicular networks. The characteristics of V2X communication are as follows [29]

E2E delay: 10-100 ms

- Reliability: 10−5
- Positioning accuracy: 30 cm
- Data rate: 10-40 Mbps

1.5.7.11 Full-Duplex and Green Communication

To double the data rate, Full-duplex communication uses the same frequency band. As a result, it allows uplink and downlink communication simultaneously, whereas the traditional transceivers use half-duplex. In addition, various algorithms are used to suppress self-interference due to outgoing signals from a local antenna, to achieve acceptable reception quality.

Green communication reduces energy consumption in which base stations consume a lot of energy. If a large number of base stations are used, then there is a severe issue for the environment. Energy efficiency can be improved by reducing radio frequency transmit power. Transmitting reference signals between a base station and user equipment only is necessary rather than in every subframe reduces energy expenditure. Introducing a duty cycle mechanism by shutting down some low-traffic base stations is another way to reduce energy consumption. Service providers can also use green energy sources such as solar or wind energy. Base stations can also provide coordination by transferring the excess stored energy from one base station to the other, which has lower power. [29]

1.5.8 5G Security Architecture

The 5G security architecture was designed to tackle security threats that had been identified with previous architectures.

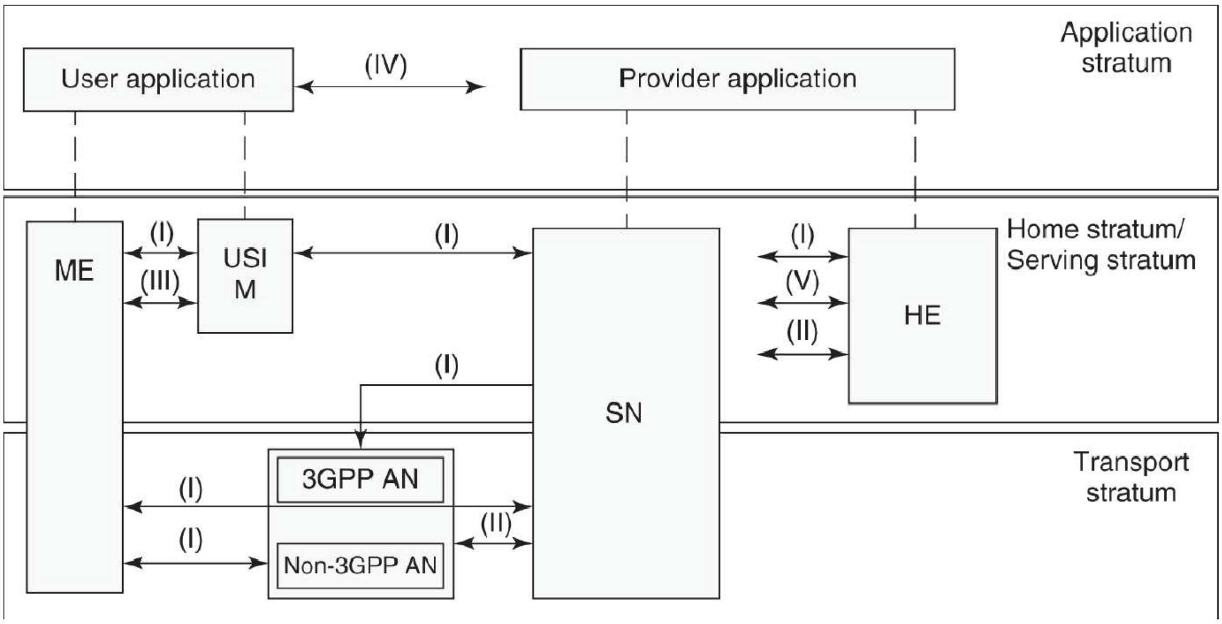


Figure 1-25. 5G security architecture as defined by 3GPP release 15 [18]

Table 1-6 Below are the descriptions of the various components based on [18] [30]

Component	**Name**	**Function**
(I)	Network Access Security	It enables the UE to authenticate securely and access network services. It also provides security for 3GPP and non-3GPP access technologies.
(II)	Network Domain Security	Enables the exchange of data between the signalling and user plane.
(III)	User Domain Security	Creates secure access to the UE.
(IV)	Application Domain Security	User and provider domain applications can securely exchange messages.
(V)	Service-Based Architecture (SBA) Domain Security	Provides security features for network element registration, discovery, and authorization, and security for service-based interfaces.
(VI)	Visibility and configuration of security	Includes security features that inform users whether security features are in operation or not.

1.5.9 Security Challenges in 5G networks

Here, we look at the various security challenges that 5G networks face.

1.5.9.1 5G Security vulnerabilities

Below are multiple vulnerabilities and threats on a 5G network as described by [31]

Interoperability with 2G-4G Networks

5G will extend its level of interoperability with previous generations of mobile networks, which means it will be exposed to vulnerabilities in Diameter signalling and SS7 functions.

Issues related to data protection and privacy

Man-in-the-Middle (MITM) attack in a 5G network where an attacker accesses personal data using International Mobile Subscriber Identity (IMSI)-catchers or cellular rogue base stations mimicking genuine mobile network operator equipment.

Possibility of rerouting sensitive data

The 5G core network SBA itself could make the 5G network vulnerable to Internet Protocol (IP) attacks such as Distributed Denial of Service (DDoS). Similarly, network hijacking, which involves redirecting confidential data through an intruder's network, could be another attack.

Collision of Politics and Technology

Government entities can impact 5G security when producing hardware for cellular networks. For instance, various countries have new regulations that ban the use of 5G infrastructure equipment procured from Chinese companies (Huawei and ZTE), citing concerns over possible surveillance by the Chinese government.

Network Slicing and Cyberattacks

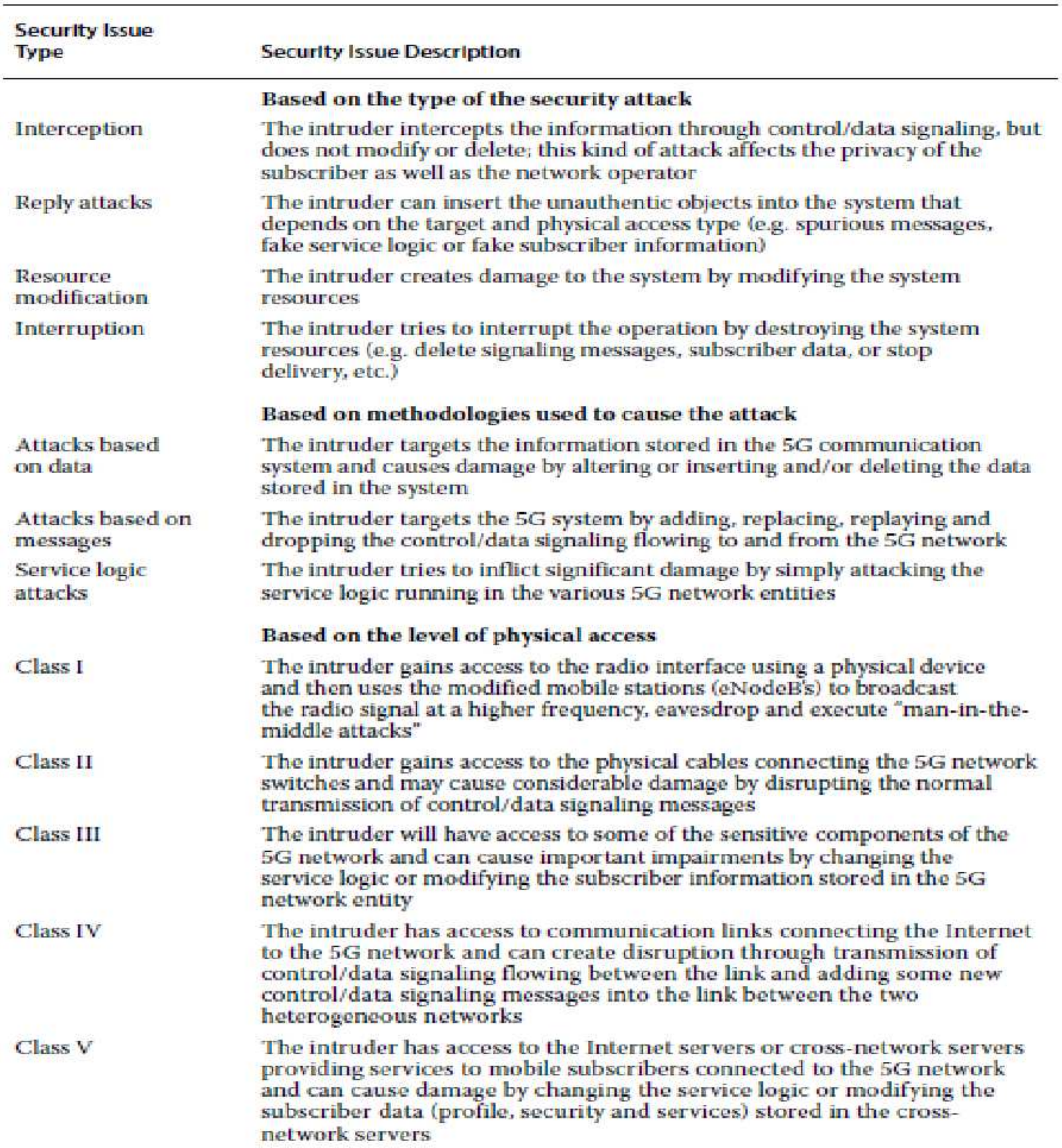
Network slicing is a 5G SA core network function (defined by 3GPP) that can logically separate network resources. The facility empowers a cellular network operator to create multiple independent and logical (virtual) networks on shared access. However, concerns about security risks in how a perpetrator could compromise a network slice to monopolize resources for compute-intensive activities have been raised despite the benefits.

Other Challenges are:

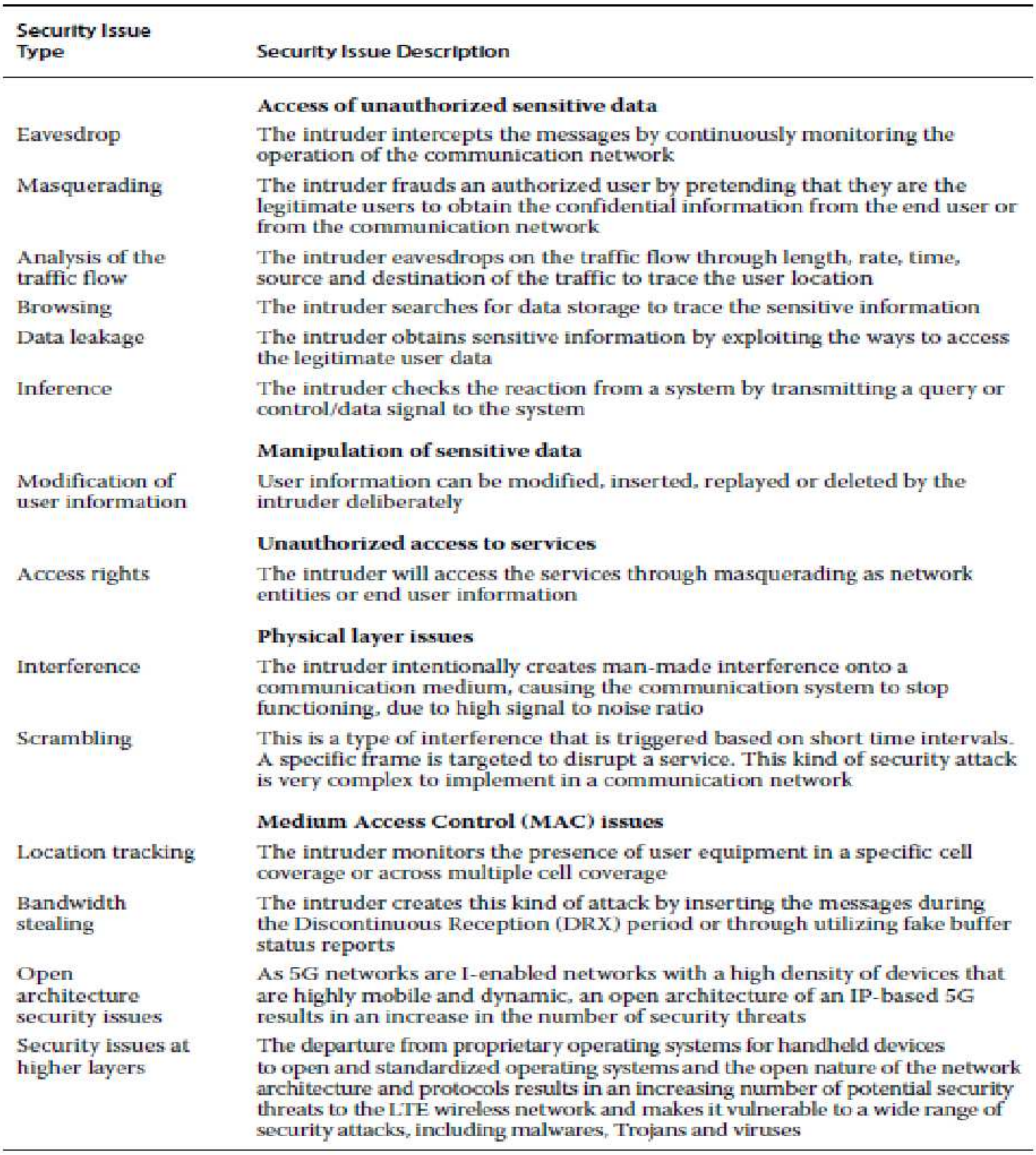
- Flash network traffic: High number of end-user devices and new things (IoT).
- Security of radio interfaces: Radio interface encryption keys are sent over insecure channels.
- User plane integrity: No cryptographic integrity protection for the user data plane.
- Mandated security in the network: Service-driven constraints on the security architecture lead to the optional use of security measures.
- Roaming security: User-security parameters are not updated with roaming from one operator network to another, leading to security compromises with roaming.
- Denial of Service (DoS) attacks on the infrastructure: Visible nature of network control elements and unencrypted control channels.
- Signalling storms: Distributed control systems requiring coordination, e.g. Non-Access Stratum (NAS) layer of Third Generation Partnership Project (3GPP) protocols.
- DoS attacks on end-user devices: No security measures for operating systems, applications, and configuration data on user devices.

The tables below show various 5G security issues and their description.

Table 1-7. 5G security issues and their description [9]



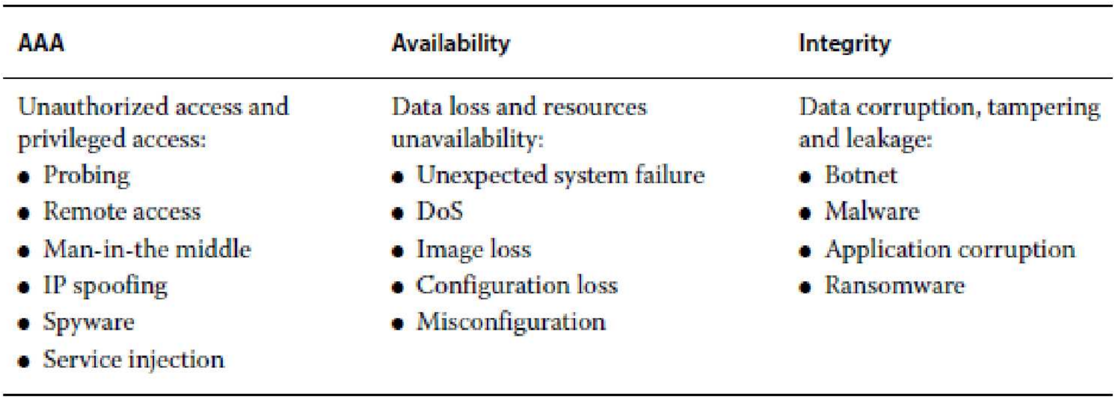
Security Issue Type	Security Issue Description
	Based on the type of the security attack
Interception	The intruder intercepts the information through control/data signaling, but does not modify or delete; this kind of attack affects the privacy of the subscriber as well as the network operator
Reply attacks	The intruder can insert the unauthentic objects into the system that depends on the target and physical access type (e.g. spurious messages, fake service logic or fake subscriber information)
Resource modification	The intruder creates damage to the system by modifying the system resources
Interruption	The intruder tries to interrupt the operation by destroying the system resources (e.g. delete signaling messages, subscriber data, or stop delivery, etc.)
	Based on methodologies used to cause the attack
Attacks based on data	The intruder targets the information stored in the 5G communication system and causes damage by altering or inserting and/or deleting the data stored in the system
Attacks based on messages	The intruder targets the 5G system by adding, replacing, replaying and dropping the control/data signaling flowing to and from the 5G network
Service logic attacks	The intruder tries to inflict significant damage by simply attacking the service logic running in the various 5G network entities
	Based on the level of physical access
Class I	The intruder gains access to the radio interface using a physical device and then uses the modified mobile stations (eNodeB's) to broadcast the radio signal at a higher frequency, eavesdrop and execute "man-in-the-middle attacks"
Class II	The intruder gains access to the physical cables connecting the 5G network switches and may cause considerable damage by disrupting the normal transmission of control/data signaling messages
Class III	The intruder will have access to some of the sensitive components of the 5G network and can cause important impairments by changing the service logic or modifying the subscriber information stored in the 5G network entity
Class IV	The intruder has access to communication links connecting the Internet to the 5G network and can create disruption through transmission of control/data signaling flowing between the link and adding some new control/data signaling messages into the link between the two heterogeneous networks
Class V	The intruder has access to the Internet servers or cross-network servers providing services to mobile subscribers connected to the 5G network and can cause damage by changing the service logic or modifying the subscriber data (profile, security and services) stored in the cross-network servers



Security Issue Type	Security Issue Description
	Access of unauthorized sensitive data
Eavesdrop	The intruder intercepts the messages by continuously monitoring the operation of the communication network
Masquerading	The intruder frauds an authorized user by pretending that they are the legitimate users to obtain the confidential information from the end user or from the communication network
Analysis of the traffic flow	The intruder eavesdrops on the traffic flow through length, rate, time, source and destination of the traffic to trace the user location
Browsing	The intruder searches for data storage to trace the sensitive information
Data leakage	The intruder obtains sensitive information by exploiting the ways to access the legitimate user data
Inference	The intruder checks the reaction from a system by transmitting a query or control/data signal to the system
	Manipulation of sensitive data
Modification of user information	User information can be modified, inserted, replayed or deleted by the intruder deliberately
	Unauthorized access to services
Access rights	The intruder will access the services through masquerading as network entities or end user information
	Physical layer issues
Interference	The intruder intentionally creates man-made interference onto a communication medium, causing the communication system to stop functioning, due to high signal to noise ratio
Scrambling	This is a type of interference that is triggered based on short time intervals. A specific frame is targeted to disrupt a service. This kind of security attack is very complex to implement in a communication network
	Medium Access Control (MAC) issues
Location tracking	The intruder monitors the presence of user equipment in a specific cell coverage or across multiple cell coverage
Bandwidth stealing	The intruder creates this kind of attack by inserting the messages during the Discontinuous Reception (DRX) period or through utilizing fake buffer status reports
Open architecture security issues	As 5G networks are I-enabled networks with a high density of devices that are highly mobile and dynamic, an open architecture of an IP-based 5G results in an increase in the number of security threats
Security issues at higher layers	The departure from proprietary operating systems for handheld devices to open and standardized operating systems and the open nature of the network architecture and protocols results in an increasing number of potential security threats to the LTE wireless network and makes it vulnerable to a wide range of security attacks, including malwares, Trojans and viruses

1.5.9.2 5G threat classification and description

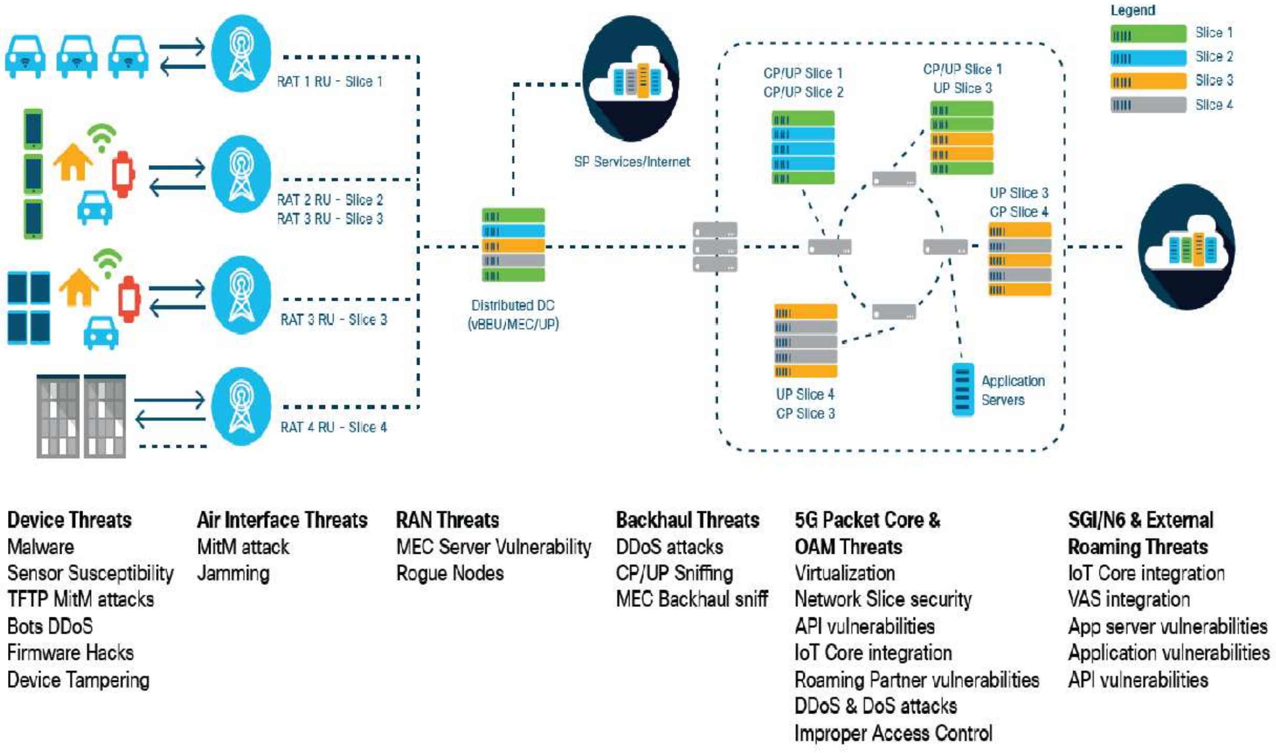
Table 1-8. Classification of threats in 5G [9]



AAA	Availability	Integrity
Unauthorized access and privileged access: • Probing • Remote access • Man-in-the middle • IP spoofing • Spyware • Service injection	Data loss and resources unavailability: • Unexpected system failure • DoS • Image loss • Configuration loss • Misconfiguration	Data corruption, tampering and leakage: • Botnet • Malware • Application corruption • Ransomware

1.5.10 The 5G threat landscape

This figure displays the 5G architecture and the significant threats each part of the 5G network is vulnerable to



Device Threats
Malware
Sensor Susceptibility
TFTP MitM attacks
Bots DDoS
Firmware Hacks
Device Tampering

Air Interface Threats
MitM attack
Jamming

RAN Threats
MEC Server Vulnerability
Rogue Nodes

Backhaul Threats
DDoS attacks
CP/UP Sniffing
MEC Backhaul sniff

5G Packet Core & OAM Threats
Virtualization
Network Slice security
API vulnerabilities
IoT Core integration
Roaming Partner vulnerabilities
DDoS & DoS attacks
Improper Access Control

SGI/N6 & External Roaming Threats
IoT Core integration
VAS integration
App server vulnerabilities
Application vulnerabilities
API vulnerabilities

Figure 1-26. The 5G architecture threat surface, according to [13]

1.6 5G Core Network Architecture

The plan with 5G core architecture is to deliver the whole network as a service. The 5G core network is re-designed based on a service-oriented architecture by parsing everything into complex functions and sub-functions. For example, the Mobility Management Entity (MME) functionality has been redistributed into precise families of mobility and session management network functions. Functionalities offered by 4G MME, such as registration, reachability, mobility management and connection management services, are provided by a new 5G general network function called Access and Mobility Management Function (AMF).

Additionally, session establishment and session management, formerly part of the MME, are now services offered by a new network function called the Session Management Function (SMF). Furthermore, packet routing and forwarding functions (currently performed by the SGW and PGW in 4G) are now realized as services rendered through a new network function called the User Plane Function (UPF). This is achieved with 5G core technologies such as SDN and NFV, which are software-based solutions. With this granular approach, more resilient networks may be realized. [32]

Networks must serve devices and applications with varying traffic profiles. As such, it is essential to accommodate the needs of applications and allocate network resources based on these diverse requirements. The 5G network flexibly gives its resources, based on rules defined in software, for optimal service. This flexibility is achieved with the help of software-defined networking and network function virtualization. [33]

1.6.1 Network interfaces in the 5G Core

NG1: A reference point between the UE and the Access and Mobility Management functions

NG2: A reference point between the gNB and the Access and Mobility Management functions

NG3: A reference point between the gNB and the User plane function (UPF)

NG4: A reference point between the Session Management function (SMF) and the User plane function (UPF)

NG5: A reference point between the Policy Function (PCF) and an Application Function (AF)

NG6: A reference point between the User Plane function (UPF) and a Data Network (DN)

NG7: A reference point between the Session Management function (SMF) and the Policy Control function (PCF)

NG8: A reference point between Unified Data Management and AMF

NG9: A reference point between two Core User plane functions (UPFs)

NG10: A reference point between UDM and SMF

NG11: A reference point between Access and Mobility Management functions (AMF) and Session Management function (SMF)

NG12: A reference point between Access and Mobility Management Functions (AMF) and Authentication Server Function (AUSF)

NG13: A reference point between UDM and Authentication Server Function (AUSF)

NG14: A reference point between two Access and Mobility Management functions (AMF)

NG15: A reference point between the PCF and the AMF in case of a non-roaming scenario, V-PCF and AMF in case of a roaming scenario

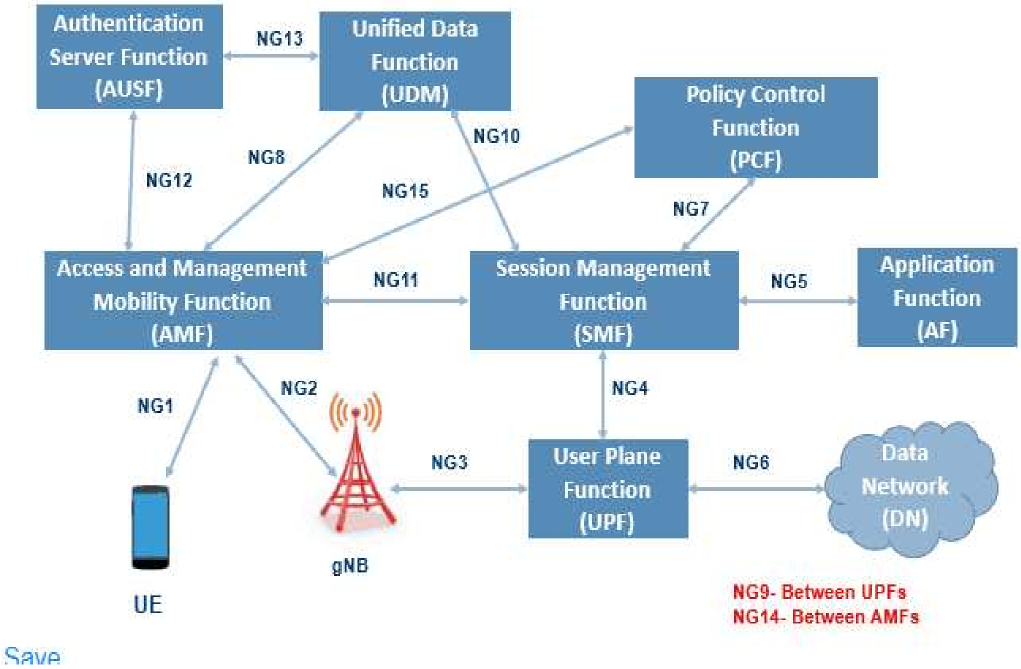

Authentication Server Function (AUSF)
NG13
Unified Data Function (UDM)
Policy Control Function (PCF)
NG12
NG8
NG10
NG15
NG7
Access and Management Mobility Function (AMF)
NG11
Session Management Function (SMF)
NG5
Application Function (AF)
NG1
NG2
NG4
NG3
User Plane Function (UPF)
NG6
Data Network (DN)
UE
gNB
NG9- Between UPFs
NG14- Between AMFs
Save

Figure 1-27. 5G core network architecture showing the various interfaces connecting the 5G network functions [34]

1.6.2 Service-Based Architecture of the 5G core network

The 5G architecture supports two types of interaction between network functions: interface-based and service-based. The first type demonstrates the interaction between network function services described as point-to-point interaction (for example, the N11 interface). This interface-centric approach is well-known from previous generations of networks. On the other hand, service-based architecture is a new way to address the mobile network architecture. Therefore, it also includes interface-based elements, as seen in Figure 1-28

In the upper part of the diagrams, network elements are connected by a single bus, with which an authorized control plane (CP) network function can access the services of another NF. [35]

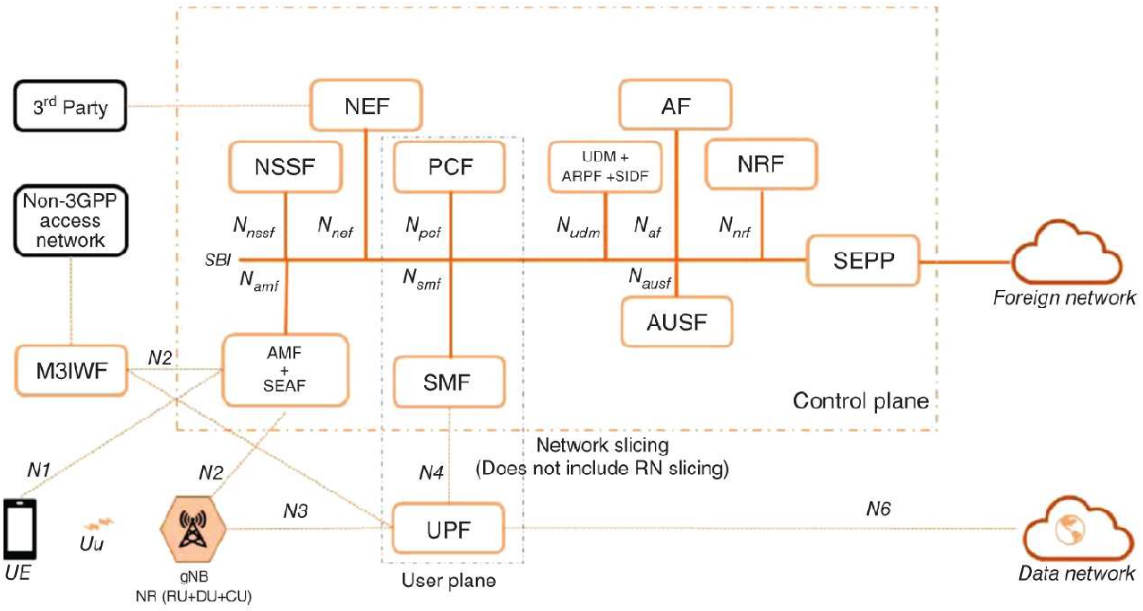


Figure 1-28. 5G-CN SBA representation as specified by 3rd Generation Partnership Project (3GPP) [30]

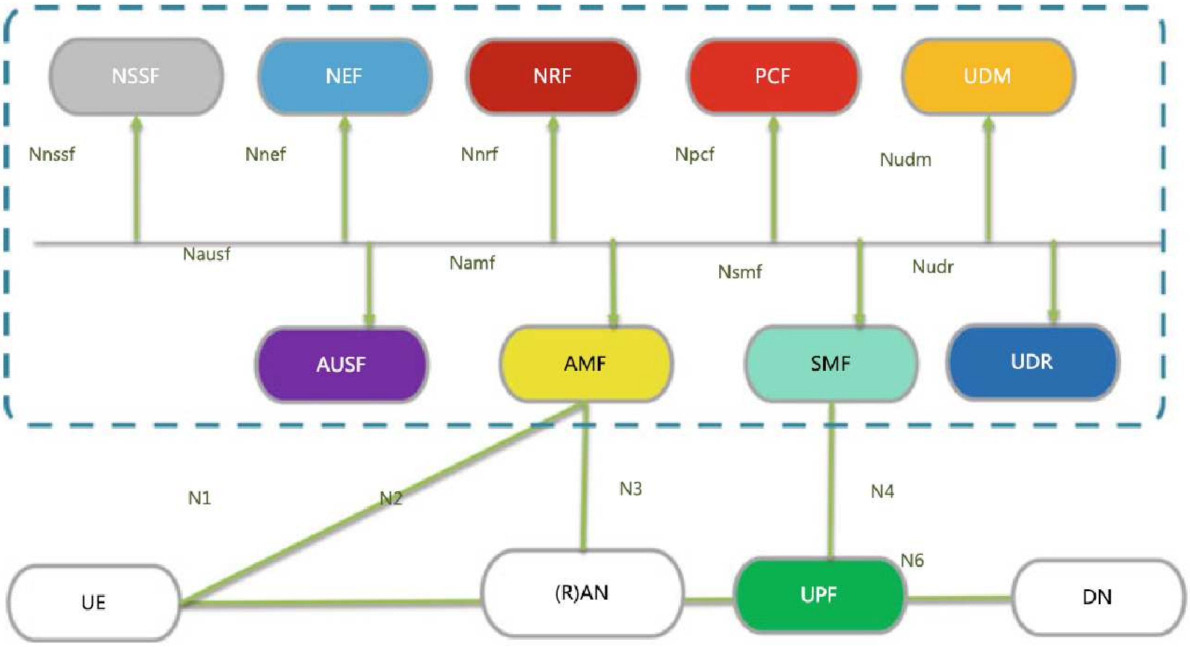


Figure 1-29. 5G core network architecture (Service-Based) [2]

The SBA uses the Representational State Transfer (REST) design model adopted by 3GPP to support communication between the distributed applications and functions instead of the GTP and diameter protocols used in 4G. The REST API implements the SBA with control plane communications via RESTful APIs using HTTPv2 methods. With this, network functions can be virtualized and provide services, using the standard

HTTP/2 Internet protocol and REST API-based Service Based Interfaces (SBI), to other network functions, external parties or market verticals.

In the SBA, the SMF and AF depend on the PCF for communication to create and authorize policies. The SBA uses the service communication proxy (SCP) for network internal communication. Some of its functions include indirect communication between network functions, authorization of the NF service consumer to access the NF service provider API, forwarding and routing of messages to destination NFs and some control functions to a next hop SCP, or for roaming purposes, load balancing, monitoring and overload protection. [17] [11]

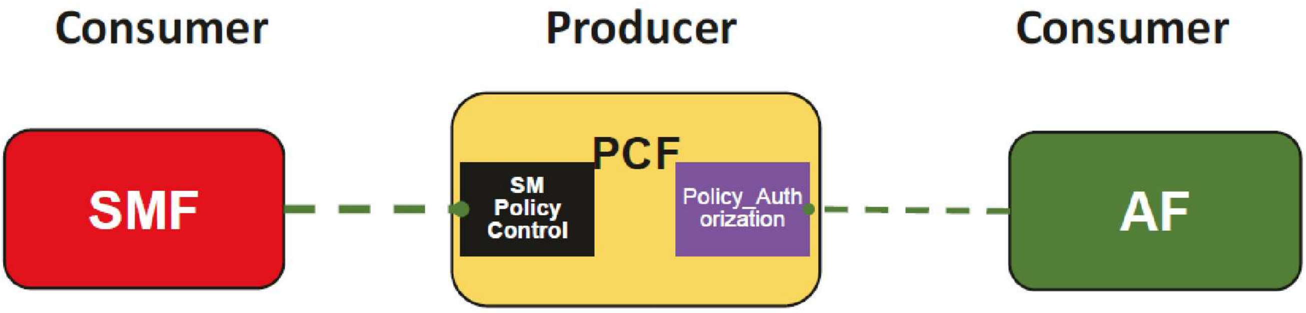


Figure 1-30. SBA communication [11]

Suppose the consumer receives an NF producer set instead of one producer through the discovery process; the consumer contacts the **service communication proxy (SCP)** and provides the NF set obtained via NRF. The SCP uses its communication logic to select the right producer from the set of NF producers. SCP has load-balancing logic and knows the best producer to serve the consumer. Also, the discovery function can be delegated to the SCP, where the consumer contacts the SCP for a service request. This communication process is displayed in Figure 1-31. [11]

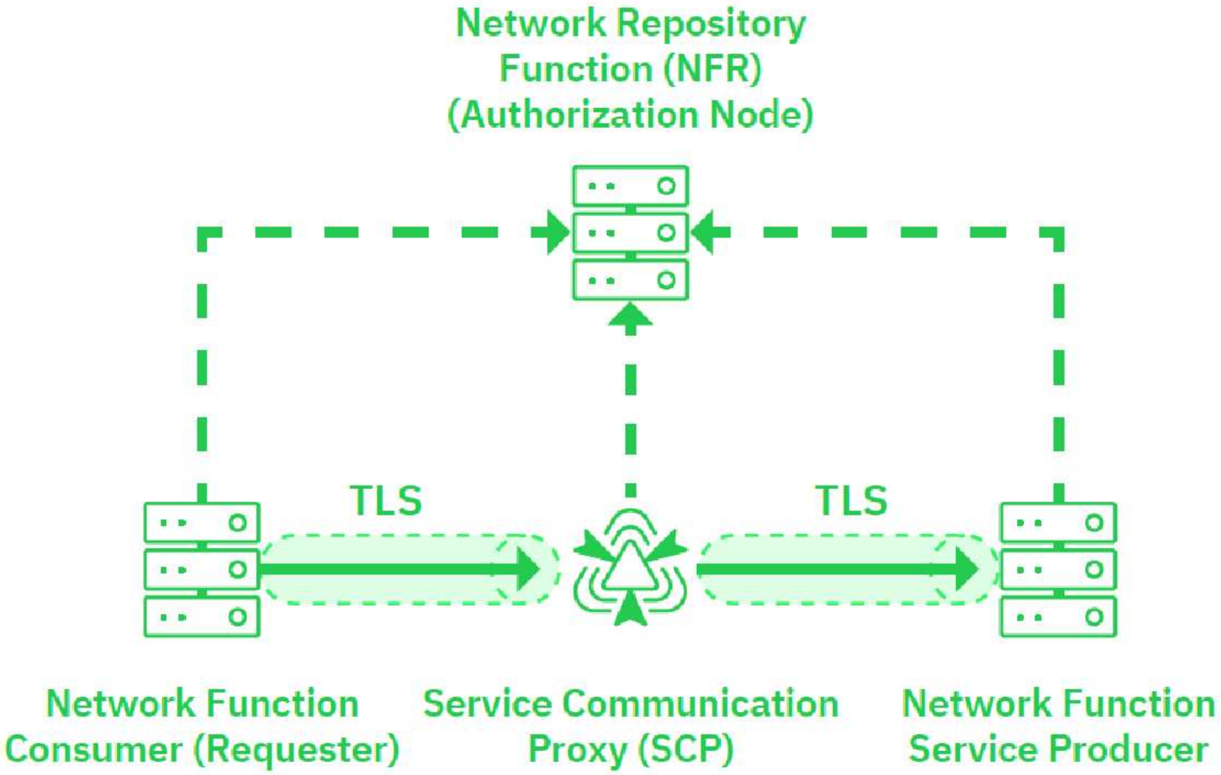

Network Repository Function (NFR) (Authorization Node)
TLS
TLS
Network Function Consumer (Requester)
Service Communication Proxy (SCP)
Network Function Service Producer

Figure 1-31. The Service communication proxy [17].

1.6.3 5G Core Network Functions.

Access and Mobility Function.

Unlike in 4G, where we have an MME, which handles mobility and session management, we have an AMF specifically for Mobility management.

The AMF is responsible for access authorization and also checking subscriptions for roaming rights, inter- and intra-system handovers, control and execution of paging and retransmissions to the GNB, support of network slicing and is used heavily to query the network slice selection function (NSSF) for proper slice selection. [11]

Essentially performs registration, reachability, and mobility management tasks.

Authentication Server Function

Performs authentication between UE and the network. The AMF initiates the UE authentication by invoking the AUSF. The AUSF selects an authentication method and performs UE authentication procedures. [7]

Session Management Function

Provides connectivity (i.e., PDU session) for UE and control of the user plane for that connectivity (e.g., selection/reselection of user plane network functions and user path, enforcement of policies including QoS policy and charging policy). [7]

User Plane Function

Performs traffic routing and forwarding, PDU session tunnel management, and QoS enforcement. The PDU session tunnels are used between the access network and UPFs, and between different UPFs as user plane data transport for PDU sessions. [7]

Policy Control Function

Controls and manages policy rules, including QoS enforcement, charging, and traffic routing. The PCF enables end-to-end QoS (e.g., maximum bit rate, guaranteed bit rate, priority level) at the appropriate granularity (e.g., per UE, flow, and PDU session).

Unified Data Management

Responsible for access authorization and subscription management. Works with the AMF and AUSF by storing UE subscription data

Network Slice Selection Function

It is responsible for slice selection. For example, the AMF uses it to select a particular "slice" for a particular use case.

Selects appropriate network slice instances for UE. When UE requests registration with the network, the AMF sends a network slice selection request to the NSSF with the preferred network slice selection information. The NSSF responds with a message including the list of appropriate network slice instances for the UE.

Network Repository Function

Assists in the discovery and selection of required network functions (NFs). Each NF instance registers itself when instantiated and updates its status (i.e., activation/deactivation) so that the NRF can maintain information about the available network function instances. Generally, each network slice instance has its own NRF, at least logically. In certain cases, such as when the network slice instances are in the same administrative domain, multiple network slice instances can share a single NFR instance. [7]. It maintains the SCP profile of available SCP instances and supports SCP discovery by SCP instances [11]

Network Exposure Function

Exposes capabilities of network functions and network slices as a service to third parties. To expose the capabilities, NEF stores the capability information and provides it upon capability discovery request.

The NEF will receive information from other network functions based on capabilities exposed by other network functions. The NEF will then store the received information as "structured data" using a standardized interface to a UDR**.** [7]

Unified Data Repository (UDR) UDR supports the following functionality: the storage and retrieval of subscription data by the UDM; the storage and retrieval of policy data by the PCF; the storage and retrieval of structured data for exposure; application data (including Packet Flow Descriptions (PFDs) for application detection and AF request information for multiple UEs), by the NEF.

Unstructured Data Storage Function (UDSF)

It supports storage and information retrieval as unstructured data by any NF.

Security Edge Protection Proxy (SEPP) SEPP is a non-transparent proxy and supports the following functionality: Message filtering and policing on inter-PLMN control plane interfaces and topology hiding.

Nausf, Nnrf, Nudm, Nnef, Namf, Nmssf, Nsmf, Npcf, Naf

These are service-based interfaces exhibited by 5G Core Control-plane functions.

N1 Reference point between the UE and the AMF. N2 Reference point between the RAN and the AMF. N3 Reference point between the RAN and the UPF. N6 Reference point between the UPF and a Data Network

Related technologies include LTE-advanced radio access networks (RANs), massive MIMO (Maximo), millimetre wave (mmWave), artificial intelligence (AI), software-defined networking (SDN), edge computing, network function virtualization (NFV), the internet of things (IoT), cloud computing, and network slicing.

The 5G core network security architecture was designed to tackle security threats that had been identified with previous architectures.

1.6.4 Security functions and procedures defined in 5G core network security

Table 1-9. Security functions and procedures defined in 5G core network security [30]

Authentication credential repository and Processing function (ARPF)	It generates authentication vectors for the authentication server function. It is a home network entity.
Authentication server function (AUSF)	A home network entity which authenticates user equipment and provides keying material for the serving network.
Subscription identifier de-concealing function (SIDF)	A home network entity that decrypts users' permanent identifiers, which are encrypted with public key cryptography in 5G to increase the privacy of users' location.
Security anchor function (SEAF)	A serving network entity that re-authenticates devices that are moving to a different access network. Minimizes signalling costs with the home network.
Security edge protection proxy (SEPP)	A firewall that filters inter-operator network traffic (particularly between home and serving network domains).
Extensible Authentication Protocol and Key Agreement (EAP-AKA) and 5G AKA authentication methods.	A form of authentication runs over a 3GPP access network that can also provide keys to establish security between the UE and a Non-3GPP Interworking Function (N3IWF) used in untrusted non-3GPP access. It is a secure identity management for identifying and authenticating subscribers through the 5G authentication and key agreement (5G AKA) protocol and the extensible authentication protocol (EAP) framework. A key feature of EAP is the flexible way different authentication protocols and credential types can be used without affecting intermediate nodes.

Security Edge Protection Proxy (SEPP)

It helps protect the 5G core network from attacks from a roaming network and serves as a security zone or firewall between two different networks [17]. It provides end-to-end confidentiality and integrity protection between two 5G networks. [11]

Some of the functions of SEPP include protecting interfaces used for roaming purposes, accounting for considerations on performance and overhead, preventing replay attacks, covering algorithm negotiation and prevention of bidding down attacks and accounting for operational aspects of key management. [32]

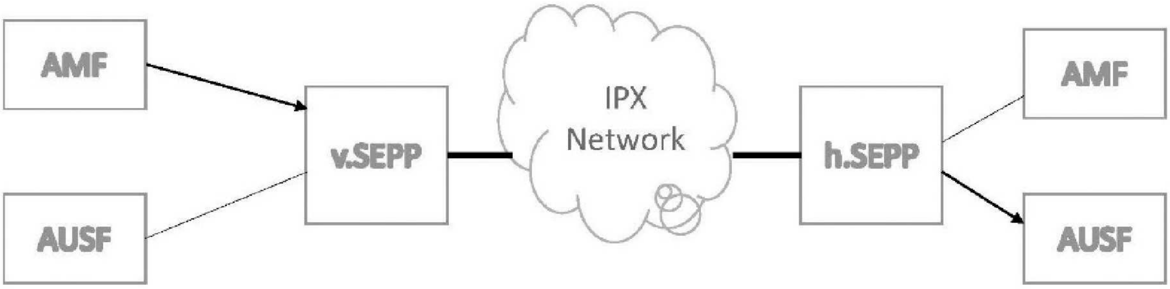


Figure 1-32. The Security edge protection proxy [32]

2 CHAPTER 2: 5G core network slicing: architecture, benefits and security flaws

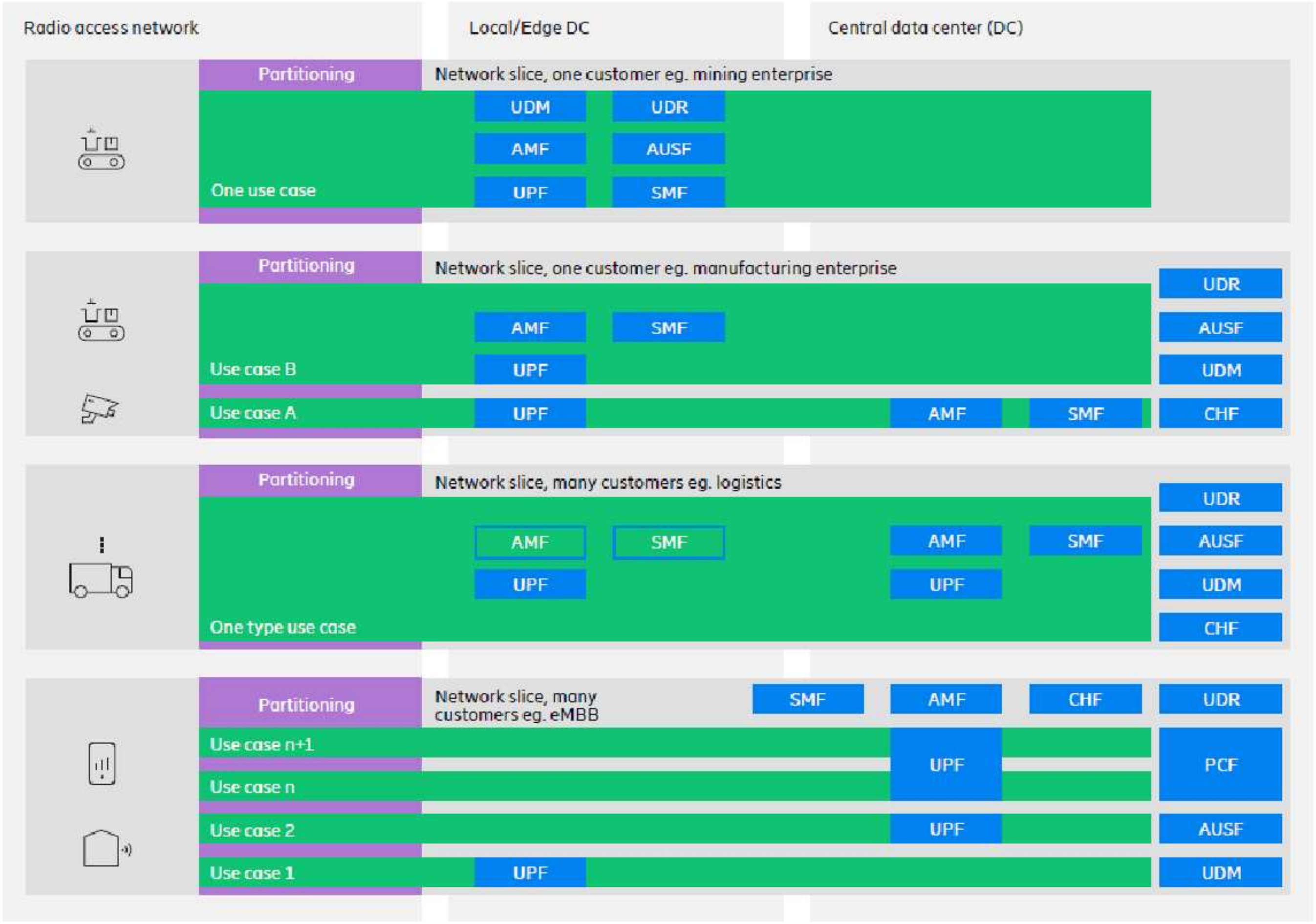


Figure 2-1. 5G network slicing catering for different and unique service needs [36]

2.1 What is Network Slicing?

Network slicing is a key enabling technology of the 5G network. It allows MNOs, CSPs, etc., to create dedicated virtual logical networks on their physical infrastructure. These virtual logical networks can have their own QoS and other requirements to meet a defined business need of a consumer or enterprise or to deliver services to various unique use cases or service needs. Network slicing involves slice creation, slice isolation and slice management. Network slicing will be implemented on the RAN, Core and transport network. These areas are where 5G network slicing could be implemented. We will first look at a detailed breakdown of End-to-end network slicing, then narrow down our focus on 5G core network slicing. These areas have been broken down into a bit of detail below;

2.2 E2E Network Slicing

Each slice, as shown in the Figure below, is designed as an end-to-end network slice made up of several sub-slices:

- A RAN (sub) slice
- A transport (sub) slice connecting the RAN slice to the core slice
- A core (sub) slice
- A second transport (sub) slice connecting elements within the core network
- The RAN controller or RAN NSSMF manages the RAN slice.
- The transport slice is managed by the transport slice controller or transport NSSMF.
- The core controller or core NSSMF manages the core slice.

Each of the slices in the Figure below is managed by a domain-specific orchestrator/controller called the Network Slice Subnet Management Function (NSSMF). This is responsible for creating, maintaining, terminating and implementing a north-bound interface. A Network Slice Management Function (NSMF) combines or cascades the sub-slices to develop an end-to-end slice. The NSSMF receives Communication from the NSMF via the north-bound interface [37]

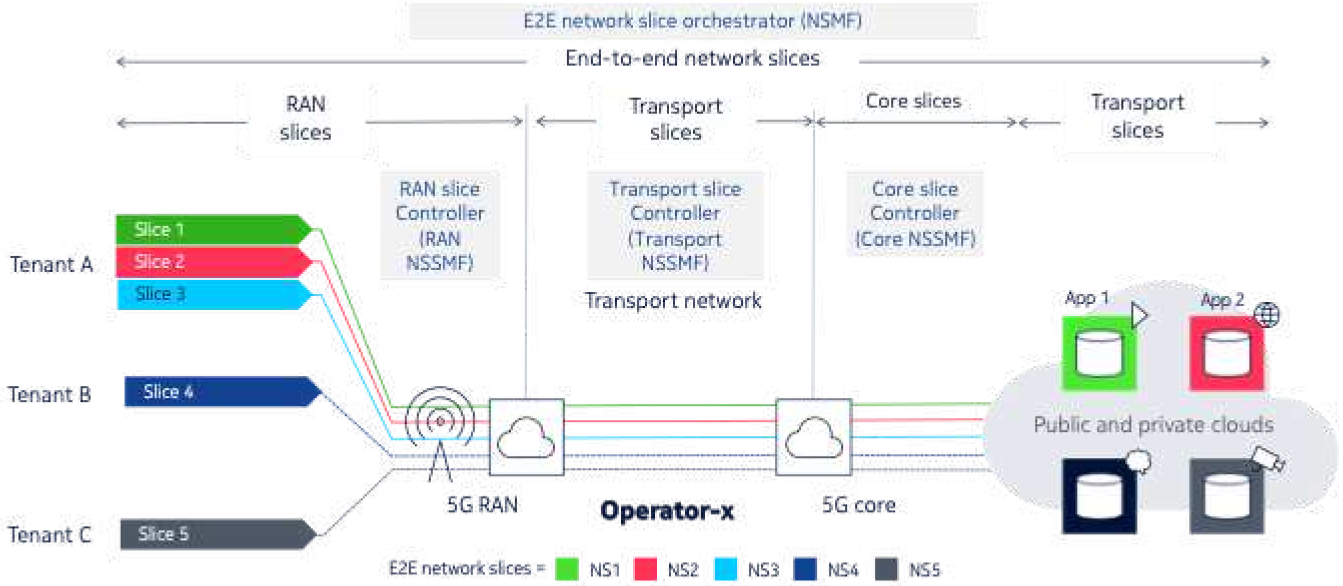

E2E network slice orchestrator (NSMF)
End-to-end network slices
RAN slices
Transport slices
Core slices
Transport slices
RAN slice Controller (RAN NSSMF)
Transport slice Controller (Transport NSSMF)
Core slice Controller (Core NSSMF)
Transport network
Tenant A
Slice 1
Slice 2
Slice 3
Tenant B
Slice 4
Tenant C
Slice 5
App 1
App 2
Public and private clouds
5G RAN
Operator-x
5G core
E2E network slices = NS1 NS2 NS3 NS4 NS5

Figure 2-2. End to End network slicing concept by [37]

2.2.1 5G Network Slicing in the RAN

This involves slicing specific areas of the spectrum or specific subcarriers. Multiple slices can also be on the same frequency. Radio resources can be allocated and prioritized for multiple slices while offering different service needs or QoS. One or more E2E network slices or users can be represented by a RAN partition [36]

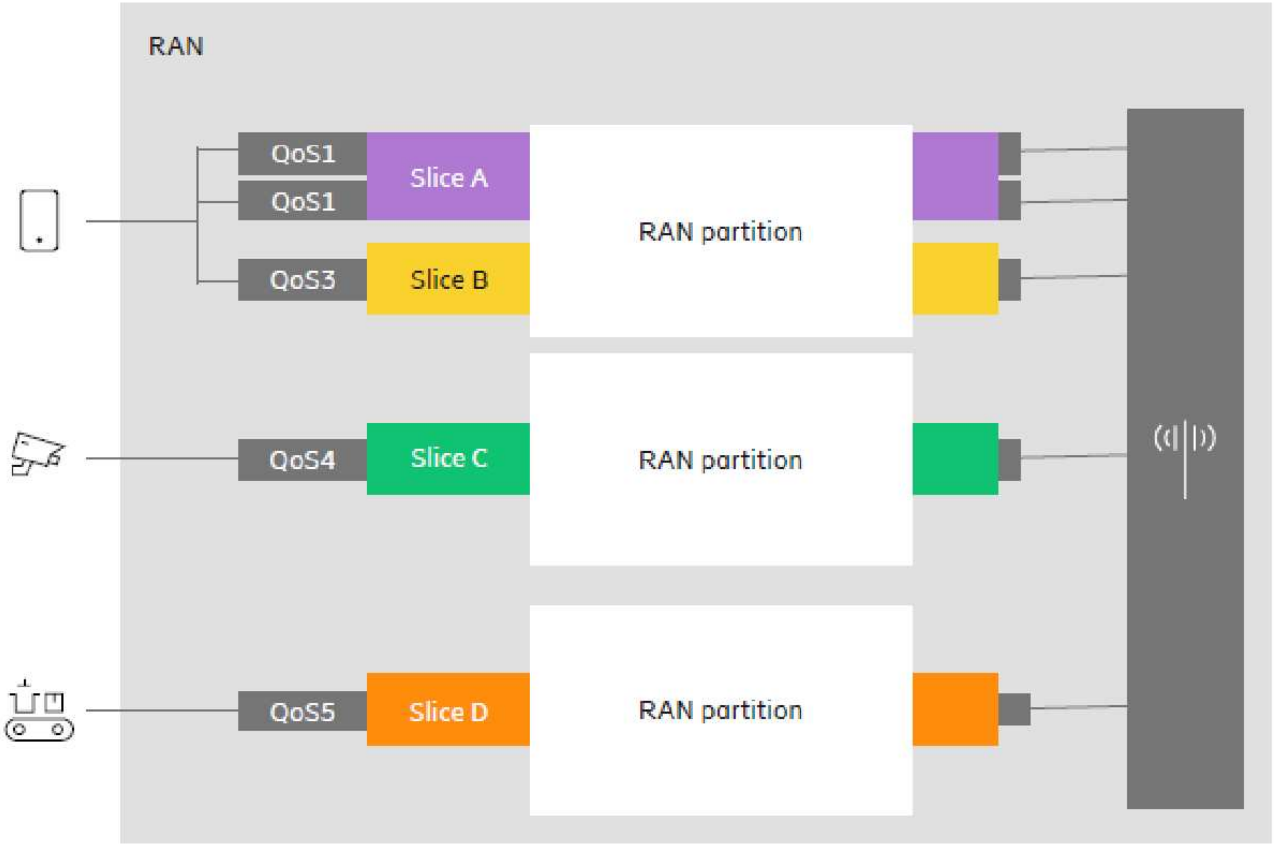


Figure 2-3. Slicing in the RAN [36]

2.2.2 5G Network Slicing in the Transport

This can be done by dynamically defining different routes through the backbone network. There is also the implementation of segment routing to split the transport network into several small segments that can be linked to define an E2E path. Traffic flows for individuals or a group of slices can be mapped into separated transport services in the transport network, and specific segments can be allocated to different slices. [38] [36]

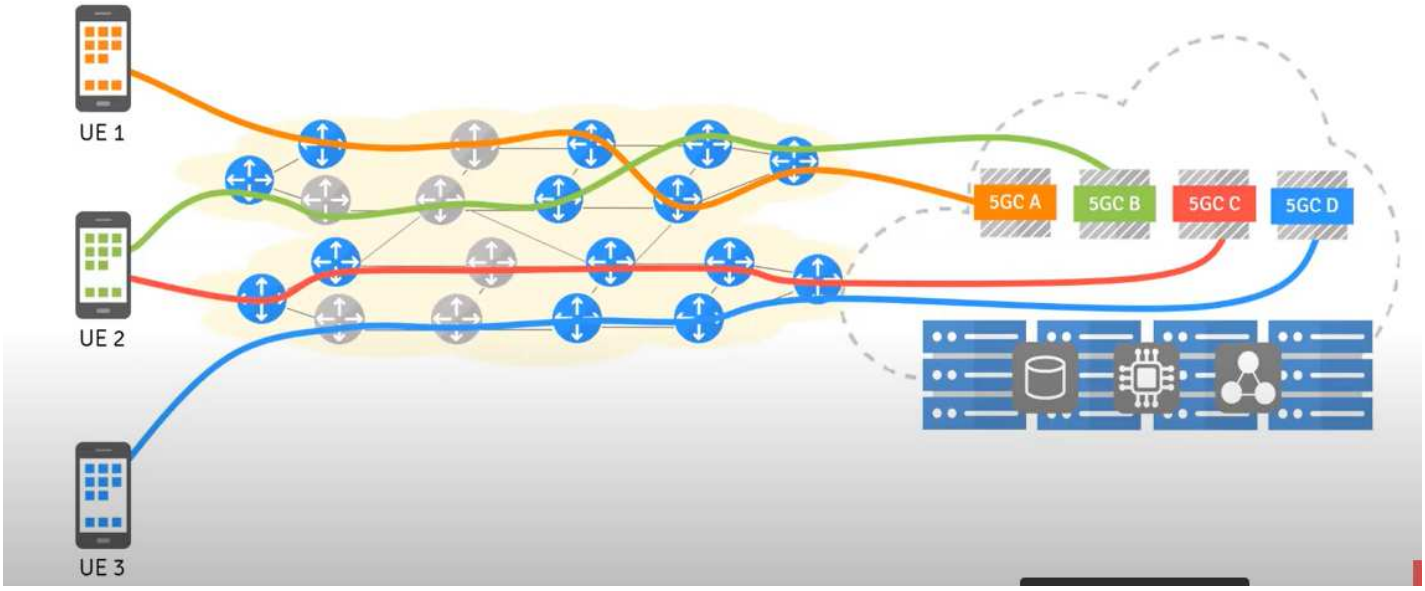


Figure 2-4. A depiction of network slicing in the transport network [38]

2.2.3 Network Slicing of the 5G Core

The framework for 5G core network slicing is more mature and well spelled out than that of the RAN and Transport. Therefore, this project scope is limited to network slicing in the core network and its security flaws.

A slice in the core network consists of a group of Network Functions (NFs) that support that slice. Those network functions can be exclusively assigned to that slice or be shared among different slices. The network functions can be virtual or physical. A physical node may host several network functions in 5G. A shared network function can provide services to several slices. [39, 39]

CN slices are designed with the logical separation between CP and UP functions and the corresponding NFs implemented as VNFs. As a result, certain NFs can be common to multiple slices, whereas some are customized for specific slices. [30]

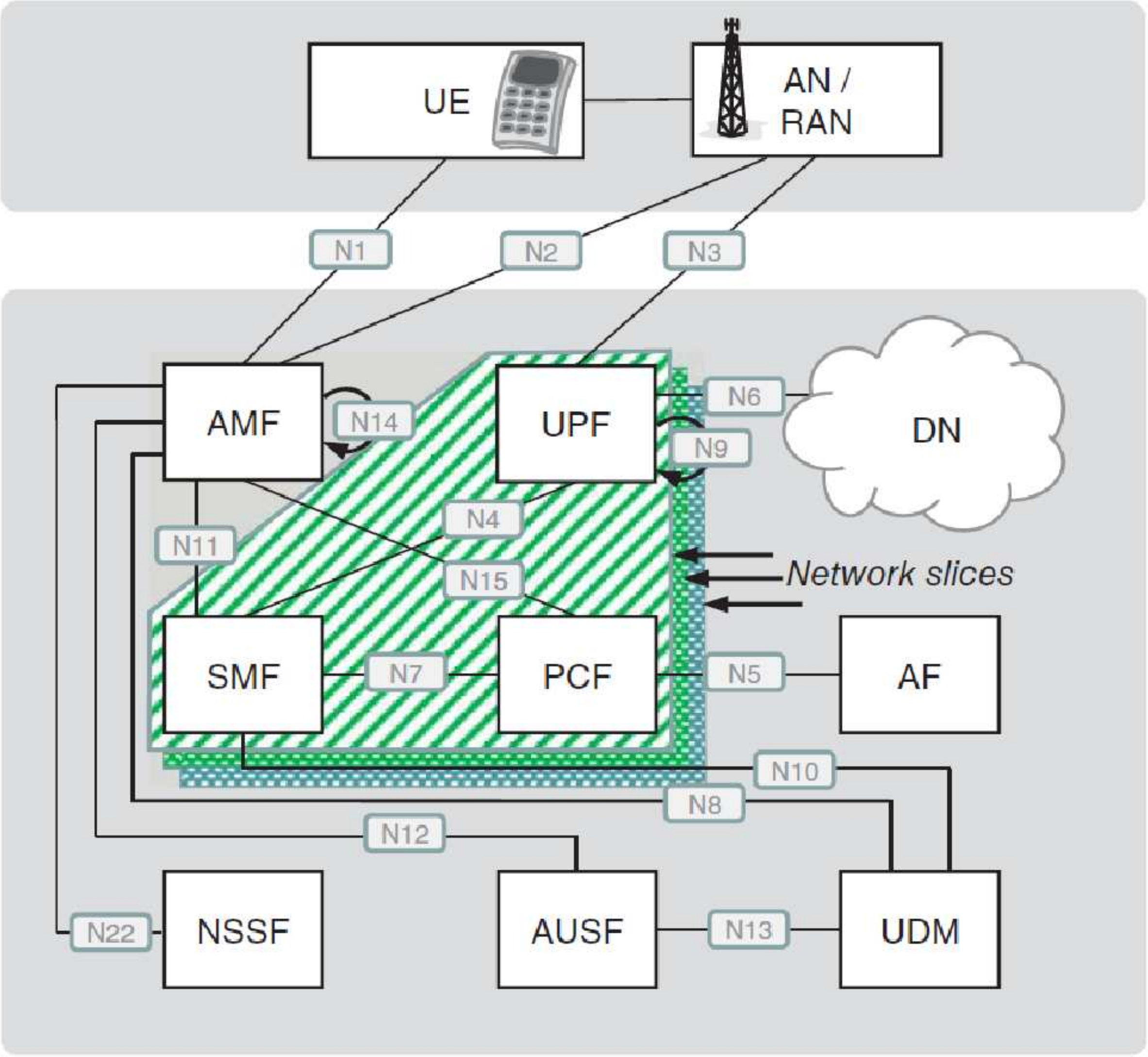


Figure 2-5. 5G core network slicing, according to [3]

5G network slicing is a breakthrough in providing a more personalized and optimized network service environment. With network slicing, MNOs can logically share network resources with ^{third} parties, UEs and other network users by breaking down their physical network into several virtual logical networks called slices, each serving a different purpose or providing a particular or a variety of network functions and QoS. Over the years, throughout the evolution of networks from 1G to 4G, networks were developed as a one size fits all type of service. Still, with 5G, we can offer services specific to the requirements of different market verticals to ensure effective and efficient management of resources and also design a more secure network for these various market verticals or use cases. With just one physical network, it is hard to meet the needs of multiple services. It is more challenging to optimize the network to meet their needs considering how fast-advancing technology has become with various market verticals and use cases requiring particular kinds of quality of service specific to their mode of operation. By doing this, we can optimize each network slice to meet the performance requirements or needs of a particular user or service. With the service-based architectural design of 5G networks, which 3GPP proposed, this is highly possible and easy to implement with the key enablers of 5G core network slicing, which are NFV and SDN technologies. [38]

Network Slicing is cost-effective because, through NFV and SDN technologies, we can use commonly shared hardware.

2.3 What are the Benefits we derive from 5G network slicing?

According to [3], these are some benefits we get from network slicing as a key enabler of 5G service-based network architecture.

Network slicing provides the possibility of isolating network resources for certain services. For example, the 5G system can execute resource provisioning to assign a particular set of logical network nodes to a third party.

The network slice concept allows the differentiation of the level of the provisioned security per service type, which different slices take care of. This is a remarkable difference from the previous mobile generations because before, the applied security would be the same for all the users (devices), regardless of the service type.

Although this has not been a limitation, it could be more optimal too. The level of applied security in networks before the 5G may be unnecessarily strong for some instances, such as simple IoT communications. Also, the uniformly applied security may be considered weak for other service types, such as Critical Communication.

Furthermore, the resource utilization is optimal as the control plane of the 5G can be managed by the cloud, ensuring flexible and interrupt-persistent functioning via geo-redundant configuration. The user plane nodes, in turn, can work on services within dedicated slices. Moreover, the slice can be instantiated, updated, and deleted according to the NFV concept, making it possible to differentiate the QoS fluently.

2.4 3GPP Slice and Service types

In TS 23.501, 3GPP has defined some standard slice types based on the quality of service offered though more may be added in the future. These slice types focus on some use cases of 5G. [17]

These slice types are;

1. **massive Machine Type Communication (mMTC)**

   This slice-type scenario relates to deployments of many connected devices, which typically transmit a relatively small amount of data, such as sensors and utility meters.

2. **enhanced Mobile Broadband (eMBB)**

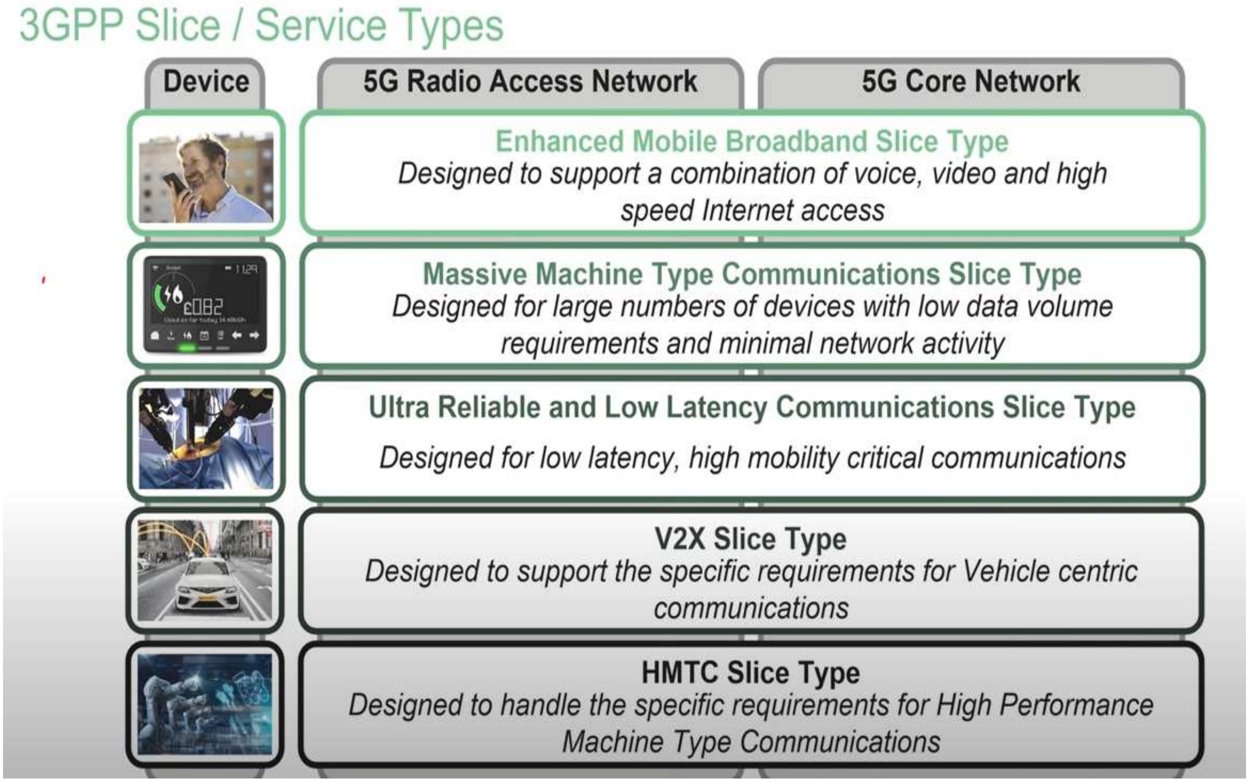
   This slice type covers improvement of the current mobile broadband services, which are typically human‑centric in terms of user-experienced data rates, traffic volume, coverage, and seamless mobility compared to services delivered in today’s system [9]

3. **Ultra-Reliable Low Latency Communications (URLLC)**

   This usage scenario is about the capability to provide a given service with stringent requirements in terms of ultra‑low latency, ultra‑high reliability and high availability, as well as high throughput

4. **Vehicle to everything (V2X)**

   This is focused on vehicle communications and connected and self-driving cars.



3GPP Slice / Service Types

Device | 5G Radio Access Network | 5G Core Network

Enhanced Mobile Broadband Slice Type
Designed to support a combination of voice, video and high speed Internet access

Massive Machine Type Communications Slice Type
Designed for large numbers of devices with low data volume requirements and minimal network activity

Ultra Reliable and Low Latency Communications Slice Type
Designed for low latency, high mobility critical communications

V2X Slice Type
Designed to support the specific requirements for Vehicle centric communications

HMTC Slice Type
Designed to handle the specific requirements for High Performance Machine Type Communications

Figure 2-6. 3GPP Slice/ Service types [40]

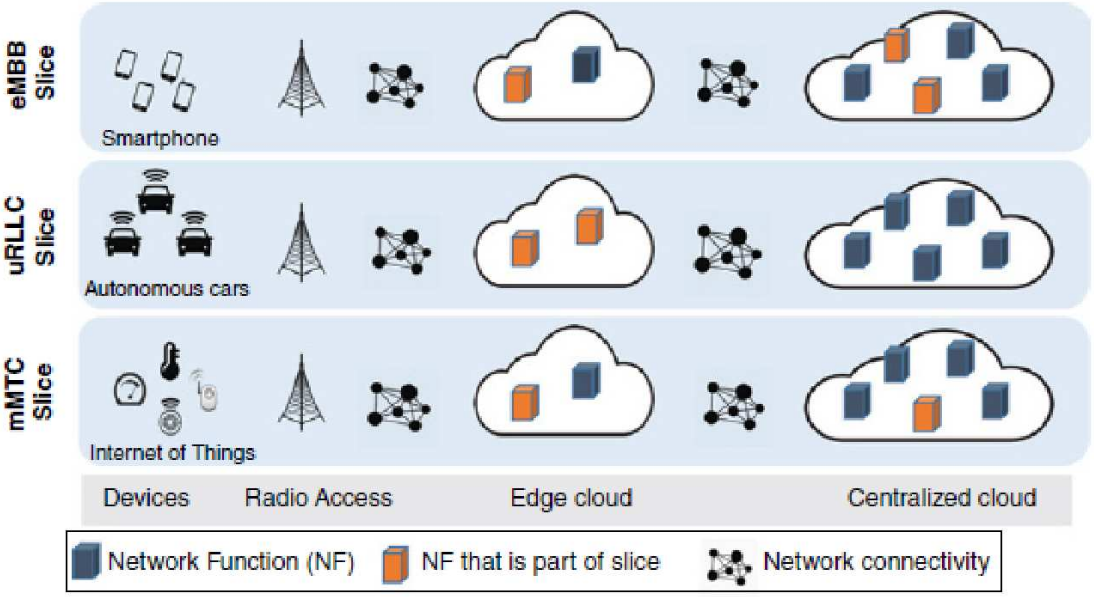


Figure 2-7. An Example of network slicing [9]

2.5 5G Core network slicing enablers and the cloud

A cloud-native architecture promotes software services broken down into compact, more convenient constituent software, achieved using a microservice architecture. 5G-CN adopts a cloud-native architecture so that each part of such micro technologies can be individually scaled, reconfigured, and upgraded. The actual cloud-native core network is a unique feature of 5G. [30] 5G core network functions will be realized as virtual machines or containers controlled by the cloud manager. [9]

The 5G core network slicing architecture will depend on Network-as-a-Service and Infrastructure-as-a-service (IaaS) with various network and security services to improve operational efficiency, scalability and reliability of network slices. With cloud implementation, different customer segments can effectively and efficiently be supported. [41]

The demanding service and network requirements in 5G required a fundamental change in the core networks compared to the previous generations. The EPC in 5G is thus based on recent technological developments such as cloud computing, SDN, NFV, and slicing. The core network components in 5G are mostly network functions (NFs) implemented in software and deployed in cloud platforms that can scale up and down dynamically based on service requirements. Therefore, the 5GCN architecture definitions call it a "service-based architecture (SBA)" framework (3rd Generation Partnership Project (3GPP) 2019), where the architectural elements are "NFs" rather than traditional network entities. All NFs are connected via interfaces to enable the services of one NF to be accessed by other authorized NFs or any authorized consumers. [30]

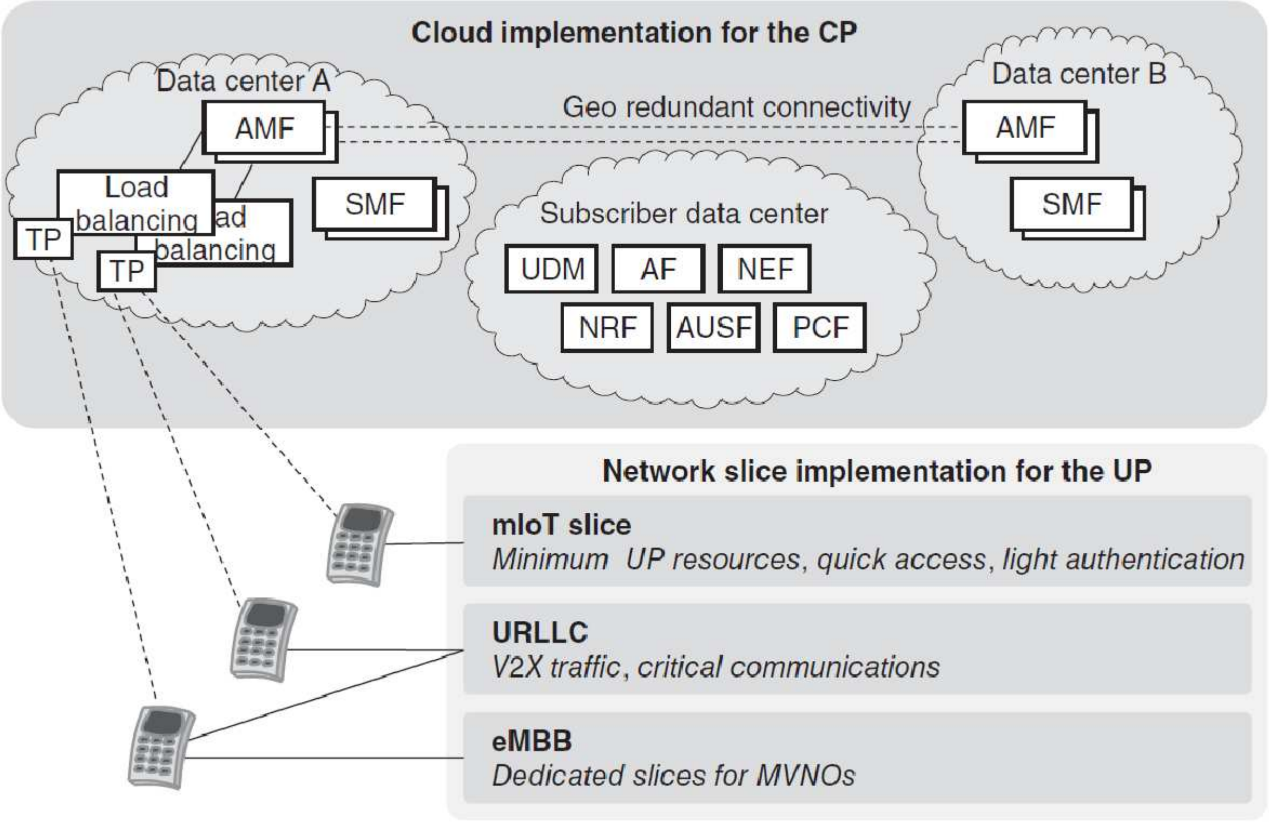

Cloud implementation for the CP
Data center A
AMF
Load balancing
ad balancing
TP
TP
SMF
Geo redundant connectivity
Data center B
AMF
SMF
Subscriber data center
UDM
AF
NEF
NRF
AUSF
PCF
Network slice implementation for the UP
mIoT slice
Minimum UP resources, quick access, light authentication
URLLC
V2X traffic, critical communications
eMBB
Dedicated slices for MVNOs

Figure 2-8. Example of the network slice set and cloud implementation [3]

We can see the implementation for the Control plane (CP) and the User plane (UP)

The objective of network slicing won't be significantly achieved without a cloud-native architecture made possible with key enabling technologies like SDN and NFV.

Software-defined Networking (SDN) minimizes the limitations of having a hardware design. With software, we have a more centralized control plane via an API. With this approach, the network services can become compatible with many types of the underlying hardware, and they can be offered and utilized regardless of the connected hardware. We can now avoid having dedicated network functions such as the policy control function, which used to be standalone in earlier networks. So with SDN, our control plane functions, such as PCF, can perform tasks on one standard hardware while sharing with other network function elements.

The SDN concept's benefits are optimized bandwidth and enhanced latency performance. Furthermore, the rerouting of the data flows occurs practically in real-time via the SDN, considerably improving the prevention of network outages and thus contributing to the development of high-availability services. [3]

With control and data plane separation, the network infrastructure can be constructed on demand and the basis of service requirements (network‐as‐a‐service), thus improving resource efficiency. [9]

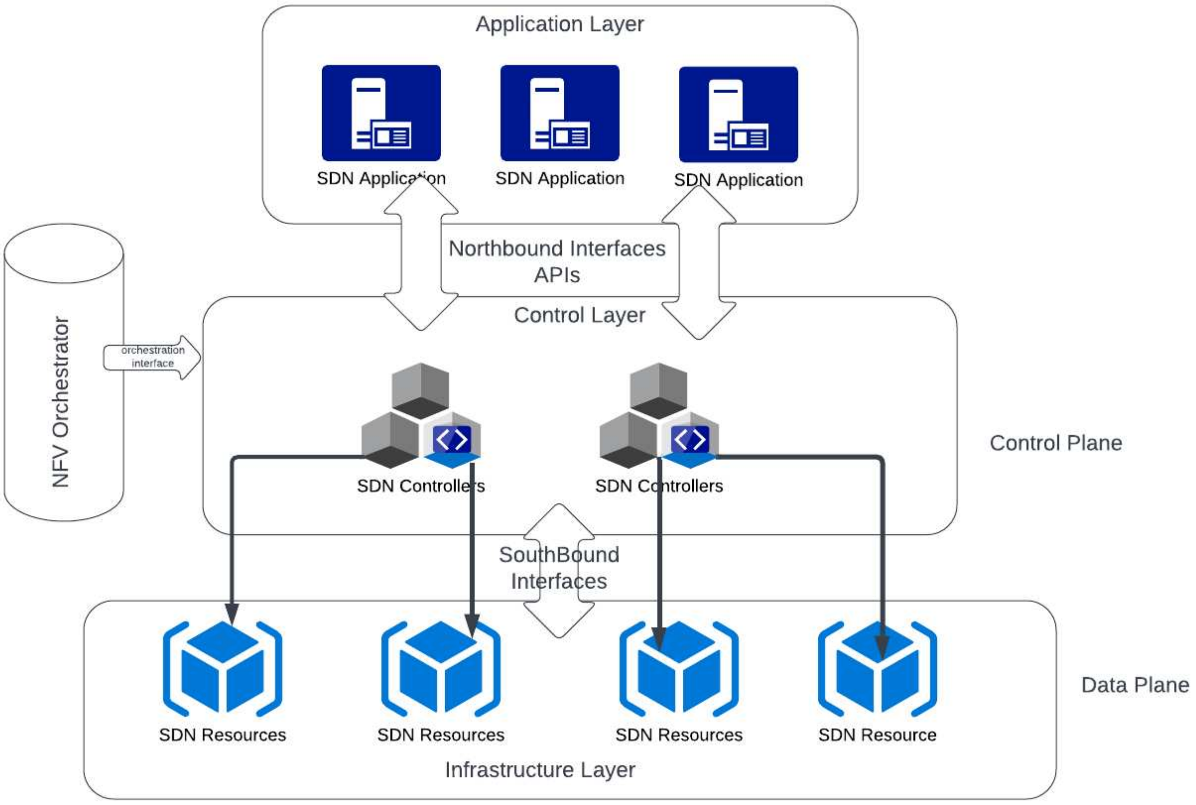


Figure 2-9. An SDN architecture

Network Function Virtualization

It simply separates hardware from software with the aid of a Virtual manager (VM). The VM in NFV helps create a network function or allocate a resource as and when required. Evolving network requirements have led to a demand for a flexible response approach to controlling traffic flows within a network or on the Internet [7]. With NFV MNOs don't have to spend on standalone elements that are mainly proprietary hardware architectures. Using NFV, network functions can be deployed faster than just dedicated hardware and software. NFV is more scalable in providing the required performance of various network functions. This makes it ideal for network slicing because we can easily create logical virtual networks over one physical infrastructure [3]. It offers a more agile and automated method to deploy and manage widely distributed network infrastructure and resources.

2.6 REQUIREMENTS OF NETWORK SLICING

According to [25], Network Slicing should be tailored to these three main requirements for more effective and secure operation.

Slice Isolation

Slice isolation guarantees the performance of a slice with the security of different market vertical or customer slices ensured by enabling the feature that one slice is independent of the other. Isolation also makes it possible to restrict one slice from accessing or modifying the other slices. Network function virtualization has the inherent capability of ensuring proper or essential isolation of slices with the additional power of imposing limits on the network resources usage. This feature guarantees the performance of different slices by fair sharing of the resources.

According to [39], NG.116 divides isolation into two main types.

Physical Isolation- involves Process and thread isolation, Physical memory isolation, and Physical network isolation.

Logical Isolation:

- Virtual resource isolation –a network slice has access to a specific range of resources that do not overlap with other network slices (e.g. virtual machine isolation)

- Network isolation - the network function is dedicated to the slice and that vertical customer, but virtual resources are shared

- Tenant/Service Isolation – the vertical customer data are isolated from other verticals, but virtual resources and network functions are shared.

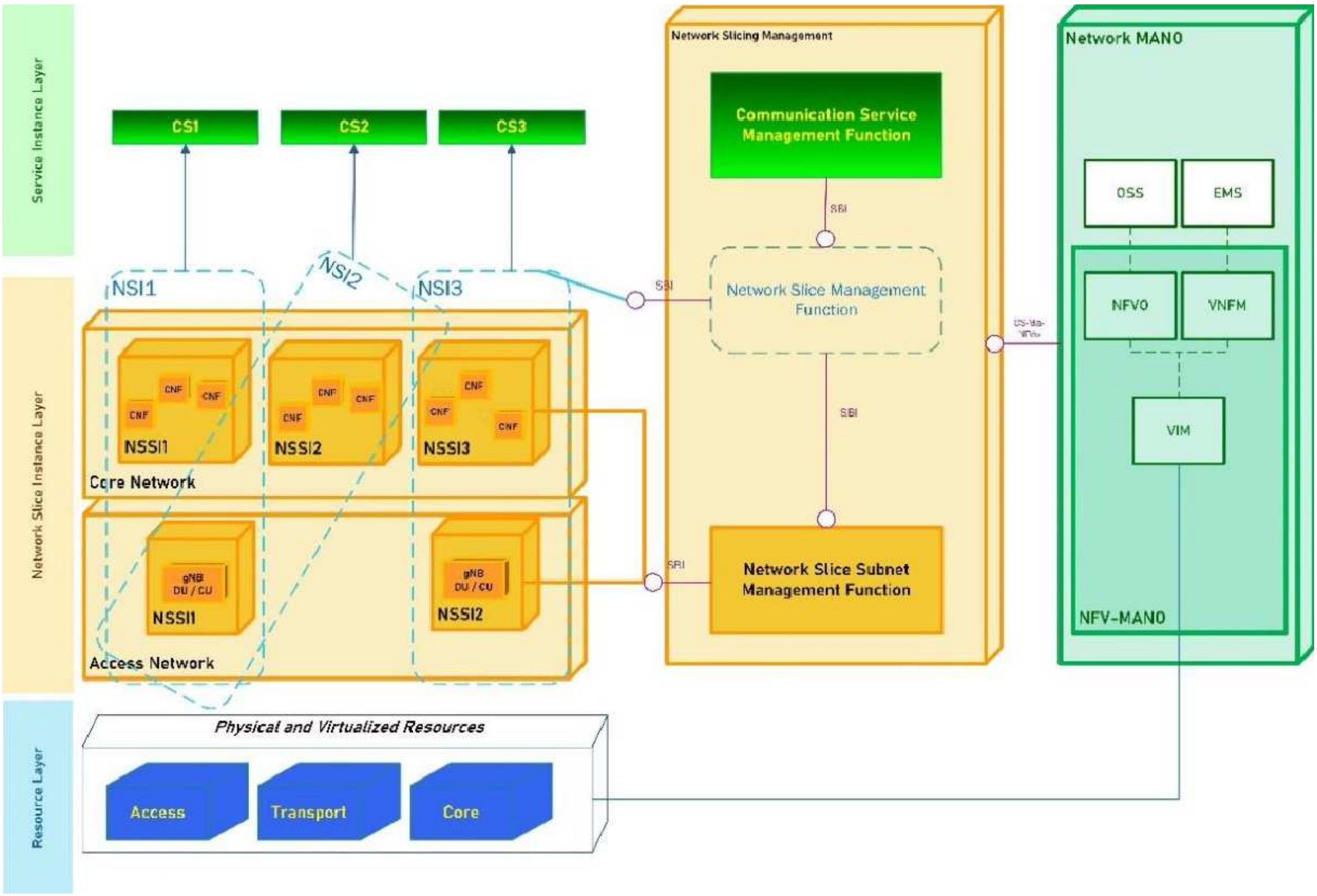
Elasticity

Elasticity allows the dynamic alteration of resources allocated to different customers or vertical market slices to utilize resources effectively. Elasticity can be realized by relocating the virtual network functions, scaling up/down the allocated resources, and reprogramming the control and data elements functionalities. The main challenge in implementing elasticity is the policy for inter-slice negotiation so that the performance of different slices remains unaffected.

End-to-End Customization

Customization ensures that allocated shared resources to different market verticals are utilized effectively. The key enablers of customization are network function virtualization, software-defined networking, and network orchestration can be exploited to provide resource allocation in a more flexible or agile way.

2.7 5G CORE NETWORK SLICING ARCHITECTURE



Service Instance Layer
Network Slice Instance Layer
Resource Layer
CS1
CS2
CS3
NSI1
NSI2
NSI3
CNF
NSSI1
NSSI2
NSSI3
Core Network
gNB DU / CU
NSSI1
NSSI2
Access Network
Physical and Virtualized Resources
Access
Transport
Core
Network Slicing Management
Communication Service Management Function
SBI
Network Slice Management Function
SBI
SBI
SBI
Network Slice Subnet Management Function
Network MANO
OSS
EMS
NFVO
VNFM
VIM
NFV-MANO

Figure 2-10. Network slicing architecture with NFV MANO [42]

Service Instance Layer (Communication Services) The Service Instance Layer represents the services (end-user or business services) to be supported. A Service instance represents each communication service. These services can be provided by the network operator or by third parties. Therefore, it can mean an operator service or a third-party-provided service.

Communication Service Management Function (CSMF) handles the translation of the communication service-related requirement to network slice-related requirements, including the number of users allowed, uplink and downlink transmission rates, latency, and jitter. It is mainly the customer-facing part of the functional block. It also manages service subscriptions and cancellations. The CSMF communicates with the Network Slice Management Function (NSMF). [41]

Network Slice Management Function (NSMF) This function is responsible for the management (including lifecycle) of NSIs across multiple domains. It derives network slice subnet-related requirements from the network slice-related requirements. NSMF communicates with the NSSMF and the CSMF.

Network Slice Subnet Management Function (NSSMF) manages and orchestrates Network Slice Subnet Instances.

Network Slice Instance (NSI) The Network Slice is a logical network that provides specific capabilities and characteristics. The Network Slice Instance represents a set of network functions and the associated resources (e.g. compute, storage and networking resources) supporting the network slice.

Network Slice Subnet Instance (NSSI) The network slice subnet represents a set of network functions (including their corresponding resources) that form part or whole parts of a network slice. The set of network functions ensures the management of each collection of network functions is managed independently of the network slice.

Network Functions (NF) Core Network Functions (CNF) Access Network Functions (gNB) a network slice instance (NSI) contains Network Functions such as the Core Network Control Plane and User Plane Network Functions in the Home Network and the Access Functions in the serving network. Release 16 of the 3GPP specification includes improved interworking with the LTE Evolved Packet Core (EPC).

NFV MANO includes NFV Orchestrator (NFVO), VNF manager (VNFM) and Virtualised infrastructure manager (VIM). The MANO is the management and network orchestration system that orchestrates and manages network components and software elements according to business processes. The objects produced by MANO mainly include a telecom operation support system (OSS/BSS), virtualized network function (VNF) and virtualized infrastructure (NFVI)

Element Management System (EMS) Element Management is responsible for FCAPS management of network functions used in the network slice instance.

Operations Support System (OSS) functions to manage and orchestrate systems, including legacy ones. They may have complete end-to-end visibility of services in an operator's network. Resources layer Network functions run software components on top of hardware infrastructure. Virtualization enables an agile, automated environment where network, compute, and storage services can be scaled up or down if required. Many resources can now be hosted as software services and dynamically initialized in different network segments.

Capability Exposure Platform

This platform provides standard application programming interfaces and a self-management portal; it also connects to the CSMF, NSMF, business support system (BSS), operations support system (OSS), network exposure function (NEF), multi-access edge computing (MEC) and other such functions and nodes that integrate and orchestrate their network capabilities. [41]

Management Functions Service-Based Interface (SBI) The management of the 3GPP network is provided by management services. Management Services offer services via standardized service interfaces composed of individually specified components. Os-Ma-nfvo The Os-Ma-nfvo reference point can interact with 3GPP slicing-related management functions and NFV-MANO. To properly interface with NFV-MANO, the NSMF and NSSMF consume the NFV MANO interface, exposed in the Os-Ma-nfvo, Ve-Vnfm-em and Ve-Vnfm-vnf reference points (the last two not displayed in the Figure due to graphical limitations). [42]

Management and Orchestration [9]

With the advancement of 5G networks, due to the diversity of use cases, services, and the number of network slices created with different resource requirements, it has become imperative that the management and orchestration (MANO) of the network is well designed and made secure to handle all the overhead and manage the whole 5G network infrastructure. It will be mainly responsible for fault management, configuration, accounting, performance, and security. More importantly, MANO will be in charge of lifecycle management and provisioning the network resources for the end‐to‐end connectivity of network slices in a dynamic, automated, and efficient manner. As illustrated in Figure 2.11, the end‐to‐end management and orchestration role will be multi‐domain, multi‐operators, and multi‐technology spanning from the infrastructure to the application (service) layer and from the RAN to the core of the network.

“MANO systems allow a mobile network operator to create end-to-end network slices and operate them over their entire lifecycle as a customer orders them.” [41]

In 2014, the NFV MANO working group in the European Telecommunications Standards Institute (ETSI) specified an architectural framework showing the main components, functionalities, and operations. In the meantime, several efforts have implemented the NFV MANO concept as open-source platforms, such as OpenStack Tacker, OpenBaton, OSM, Open‐O, etc. These open-source platforms are being applied to today’s 4G core network and integrated into deploying the 5G network architecture.

However, due to the diversity of network resources from RAN to the core, the current NFV MANO framework should be extended to manage virtualized network functions, resources, and physical nodes. In addition, dynamically managing and orchestrating the network services and slices would also be challenging.

The MANO System in Figure 2-11 [41] supports slice design and creation, activation, deactivation, and termination across the Radio Access Network (RAN), core network, and transport network domains.

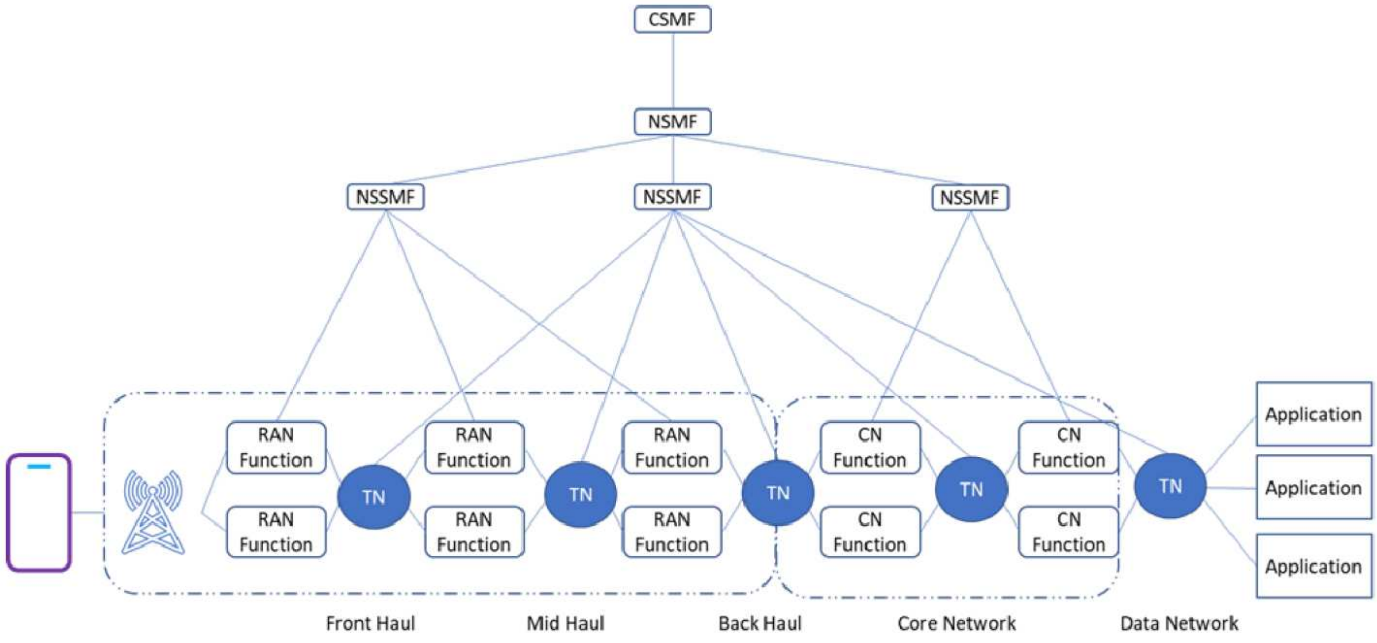


Figure 2-11. Architecture Diagram for End-to-End Network Slicing with crucial management functions [41]

2.7.1 HOW A NETWORK SLICE INSTANCE IS CREATED AND ITS LIFE CYCLE

We will look at how a slice is created, run and decommissioned to subsequently understand the security flaws that may be available at each stage as we progress.

Network slicing needs a complete framework for handling the life cycle management of slices to gain the support and confidence of enterprise customers and individual users

During the preparation and whole Lifecycle management process, the customer can provide its requirements using APIs from which the customer gets information on how the Network Slices perform and can modify its conditions to adapt to the needs of the customer. [43]

A network slice lifecycle consists of the following phases as illustrated in the diagram in Fig:

Preparation phase

In this phase, the network environment is prepared, and other necessary preparations are done as required to create an NSI. The NSI does not exist only during the preparation phase. It includes creating the network slice templates, slice design, creation of the network environments, onboarding the templates to support the lifecycle of the NSIs, and other preparation activities. During the preparation phase, created templates can also be verified. During this phase, some NSIs may meet the customer's requirement or create a new template to meet the new customer's needs. If this is the case, the new template created may be added to a repository of network slice templates so that the preparation phase can be skipped for any new customer with similar requirements. This phase also includes uploading required information, e.g., the designed templates, into the production system, validation of, e.g., templates and virtual machines (VM) images, and everything the orchestration system needs in the next step. [11] [43]

Instantiation, configuration, and activation phase

This phase involves the creation of the network slice and allocating and configuring of resources to meet the requirements of the Network slice. In the activation step, the NSI is made active by actions like traffic diversion, activation of the database, and so on. Network slice instantiation, design, configuration, and activation can include instantiation, configuration, and activation of other shared and non-shared NFs [43] [11]

Runtime or Operational phase

The runtime phase includes activation, supervision, performance reporting (e.g., for KPI monitoring), resource capacity planning, monitoring and related modification activities like upgrades, reconfiguration, NF association and disassociation with an NSI, and so on.

Some of the supported runtime operations are;

— Activate an NSI: The NSI is activated to support communication services. This may trigger the activation of the corresponding NSSIs as well.

— Resource planning and modification of an NSI: Resource capacity planning includes any actions that calculate resource usage based on an NSI provisioning and performance monitoring and generates modification policies resulting from the calculation. Performance can also be monitored according to agreed

KPIs. During the NSI modification, the NSI is reconfigured, and several workflows, such as a change in NSI capacity and topology, can be triggered. For example, NSI modification can be started by a difference in the network slice requirement or change in the communication service requirements or can result from assurance action from the NSI monitoring automatically. In addition, NSI modification may trigger corresponding NSSI modification.

— Deactivate an NSI: Here, the communication service provided by the network slice is stopped or rendered inactive. Before modifying an NSI, there might be a need to deactivate the NSI to perform the necessary changes, followed by activation of the NSI. NSI deactivation triggers NSSI deactivation to halt corresponding NSSIs not used by other NSI(s). [43] [11]

Decommissioning phase

If required, the decommissioning phase includes decommissioning non-shared resources and removing the NSI-specific configuration from the shared resources. The network slice manager (NSM) demolishes the NSI-specific configuration from the shared constituents (Rahim Tafazolli, 2021). The dedicated resources of the NSI can now be reassigned. After the decommissioning phase, the NSI is terminated and no more exists. [11] [43]

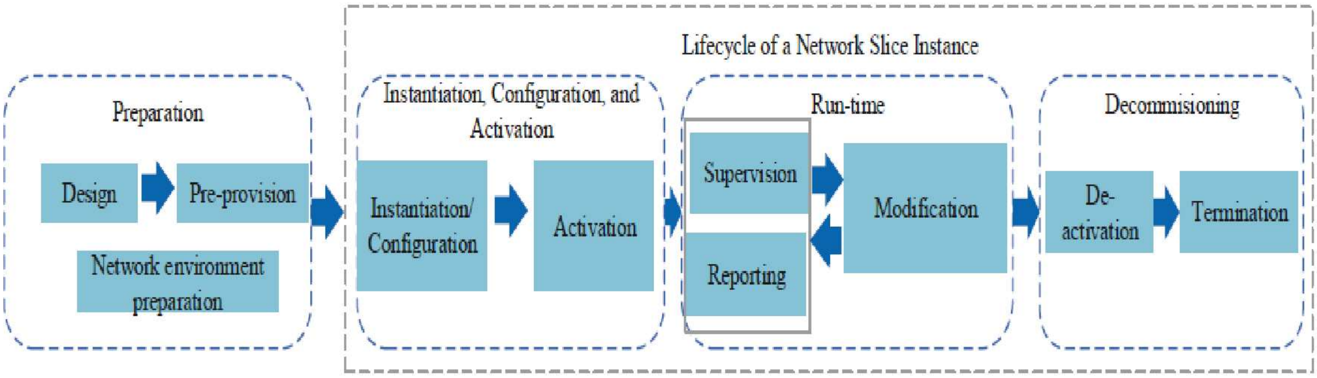

Lifecycle of a Network Slice Instance
Preparation
Design
Pre-provision
Network environment preparation
Instantiation, Configuration, and Activation
Instantiation/ Configuration
Activation
Run-time
Supervision
Reporting
Modification
Decommisioning
De-activation
Termination

Figure 2-12. Lifecycle phases of an NSI ➤ 3GPP spec 23.801 [11]

A network slice has a RAN segment, a transport network segment, a Core network segment, some services and an appropriate and secure life cycle management [38].

For this, we will use a case study of the city of Leduc, which comprises several towers from different MNOs. But we will use Telus as an MNO for this.

But before we can proceed, it is imperative to understand Network slice subscription and how a User's Equipment accesses network slice services.

But before we can proceed, it is imperative to understand Network slice subscription and how a User's Equipment accesses network slice services.

2.7.2 Network Slice Subscription from a user perspective

We need to understand these terms before we can proceed to explain this.

NSSAI: Network Slice Selection Assistance Information. This means, is it eMBB? Is it URLLC? And is it company X? It is a list of slices, and many different NSSAIs are configured, allowed, requested, and so on.

S-NSSAI: Single Network Slice Selection Assistance Information. Within a list of NSSAI, we can have this which identifies one specific slice. Inside the core network, the S-NSSAI is used for traffic differentiation and QoS aspects, but also authorization, policy enforcement and potentially for routing. That one specific slice may have SST values and, in some instances, a slice differentiator (SD) [17] [38]

Table 2-1. Use cases and their respective SST values

Use Case Scenario	**SST Value**
massive Machine Type Communication (mMTC)	1
enhanced Mobile Broadband (eMBB)	2
Ultra-Reliable Low Latency Communications (URLLC)	3
Vehicle to X (V2X)	4

DNN: Data Network Name

SST: Slice and Service Type. This is a predefined value for eMBB or mMTC, etc.

SD: Slice Differentiator

An MNO may offer the same slice type to different verticals, e.g. eMBB for streaming providers or mMTC for various IoT service providers. The MNO can choose, if it wants, to populate the SD and what value to put there. An MNO can also use non-standard S-NSSAI how it pleases, such as using their non-3GPP defined SST value or an MNO-specific and self-defined SST and SD.

Network slicing is managed in the subscription profile for the user. So we may have slicing information for the subscriber, in the subscription profile, for example, Slice 1=eMBB. Within that, we can allocate several DNNs, e.g., internet, IMS, etc., so that when a subscriber access any of those data networks will be within the eMBB slice.

It is also possible to have another slice, for instance, one for banking, e.g. Royal Bank of Canada (RBC), where the subscriber has VPN access to this company's network. So, we can take the DNN of RBC and we will allocate it to that particular eMBB slice.

Within each data network, we can provide QoS at the DNN and PDU session QoS flow level, just like with the APN in 4G LTE.

The User profile has all the slicing information, and the device uses this when it first accesses the network to tell the network, "These are the slices I would like to have access to."

Since devices are allowed the liberty to roam and also since not all network slices are available in all geographical areas, when the UE wants to register in a 5G core network, the AMF will look at the requested slices from the device or UE by looking up what is in the user profile to know what is available in its particular registration area or PLMN. The AMF then identifies a list of allowed slices for this specific UE that may not match everything in the User's profile.

An example is when a subscriber is to roam to another network and tries to access a slice peculiar to its home network, it may need access to, say, the TELUS TV slice because it is operator specific.

The AMF determines the exact slices available to the device in a specific registration area. [38]

2.7.3 How can user equipment (UE) access a network slice?

For network slicing, we must understand how a UE connects to a slice, accesses services of several slices or interacts with the 5G core to access a network slice. This is beautifully described based on 3GPP TS 29.531 and [17]

3GPP TS 29.531 defines a Network Slice Selection Function (NSSF) in its architecture, containing the NSSAI. Which is a list of various slice identities called S-NSSAIs. A typical vertical use case is that a UE would have access to a specific data network, such as a private or factory network. The subscription information of the UEs for each slice would also contain different Data Network Names (DNNs), which identify the data networks the UEs are allowed to access and which would belong to "their" slices.

There are two ways to authenticate and authorize UE access to a slice:

a. Simple slice access, which is done during the registration of the UE

This is performed for "normal" network authentication, the slice identities (S-NSSAI), and UE roaming.

b. Slice-specific access, which requires an extra authentication step (e.g. for Intranet access of private networks)

This employs the Extensible Authentication Protocol (EAP), the extra authentication needed for a UE to access a specific vertical use case slice or private network. This caters to the additional security authorization and authentication a slice needs to provide to the UE by running an authentication server before it can access a particular slice.

2.7.3.1 The Simple Slice Access Process

Access to a slice is part of the standard registration procedure in 5G, defined in 3GPP TS 23.502 and described by [17].

1. "The UE sends to the RAN in the UE registration request a list of S-NSSAI (ordered NSSAI) and potentially a mapping of requested NSSAI, which assists, in case of roaming, to find the correct slice for the UE. However, the RAN network is not aware of the subscription data for this UE.

2. The RAN performs an initial AMF selection based on the information provided by the UE. This selection can be based on a potential AMF address provided by the UE or based on Radio Access Technology (RAT) and requested NSSAI. The RAN may also have a local configuration when insufficient information is provided or the data is invalid.

3. The RAN sends the registration request to the initial AMF. This message also contains the requested NSSAI and the mapping if it was provided.

4. The initial AMF must validate if the user is allowed to access those S-NSSAI. Then, it contacts the UDM to request UE's Slice Selection Subscription data for that.

5. The initial UDM (after potentially having fetched the UE’s Slice Selection Subscription data from the UDR) provides the requested data to the initial AMF.

6. The AMF now has the data to cross-check if the UE is allowed to access the slices he requested. In addition, the AMF now knows which slices the UE is subscribed to from the data provided by the UDM.

7. The initial AMF might only be able to serve some of the S-NSSAIs from the NSSAI request that the UE is subscribed to. In that case, it sends a network slice selection request to the NSSF. It contains, among other parameters, the requested NSSAI, mapping, and subscribed S-NSSAI. The NSSF validates the request according to TS 23.501 5.15.5.2.1 option B. based on the provided information, tracking area, potential roaming scenario, and configuration. The NSSF may now need to contact the NRF to discover the target AMFs for this UE.

8. The NSSF contacts the NRF to request a list of candidate AMF(s) and includes the S-NSSAI it deems suitable according to the procedure in TS 23.501 that the candidate AMF needs to support.

9. NRF discovers the suitable AMF instances and returns a candidate list of AMF(s) to the NSSF.

10. The NSSF returns to the initial AMF the allowed NSSAI, optionally mapping the allowed NSSAI and the target AMF set or, based on configuration, the list of candidate AMF(s). For example, there can be more NSSAI lists if there is a second access type and slice-specific NRF information for the case we described in Figure 5.

11. The initial AMF received the candidate AMF. If it does not have the candidate AMF instance address stored, it needs to contact the NRF for discovery. The initial AMF now has two possibilities: either it can redirect the UE to the new target AMF or inform the target AMF to take care of the UE. This decision is based on local configuration and subscription information. In Figure 8, the initial AMF decides to reroute the message via the RAN. For this, it sends a Reroute NAS message to the RAN, which includes the information needed for the new target AMF, and the complete registration request.

12. The RAN sends the Initial UE message to the new target AMF. It indicates the route due to the slicing information from the NSSF via the initial AMF in the previous message.

13. The new AMF that serves the UE-requested slice now continues with the standard Registration procedure according to TS 23.502. “ [17]

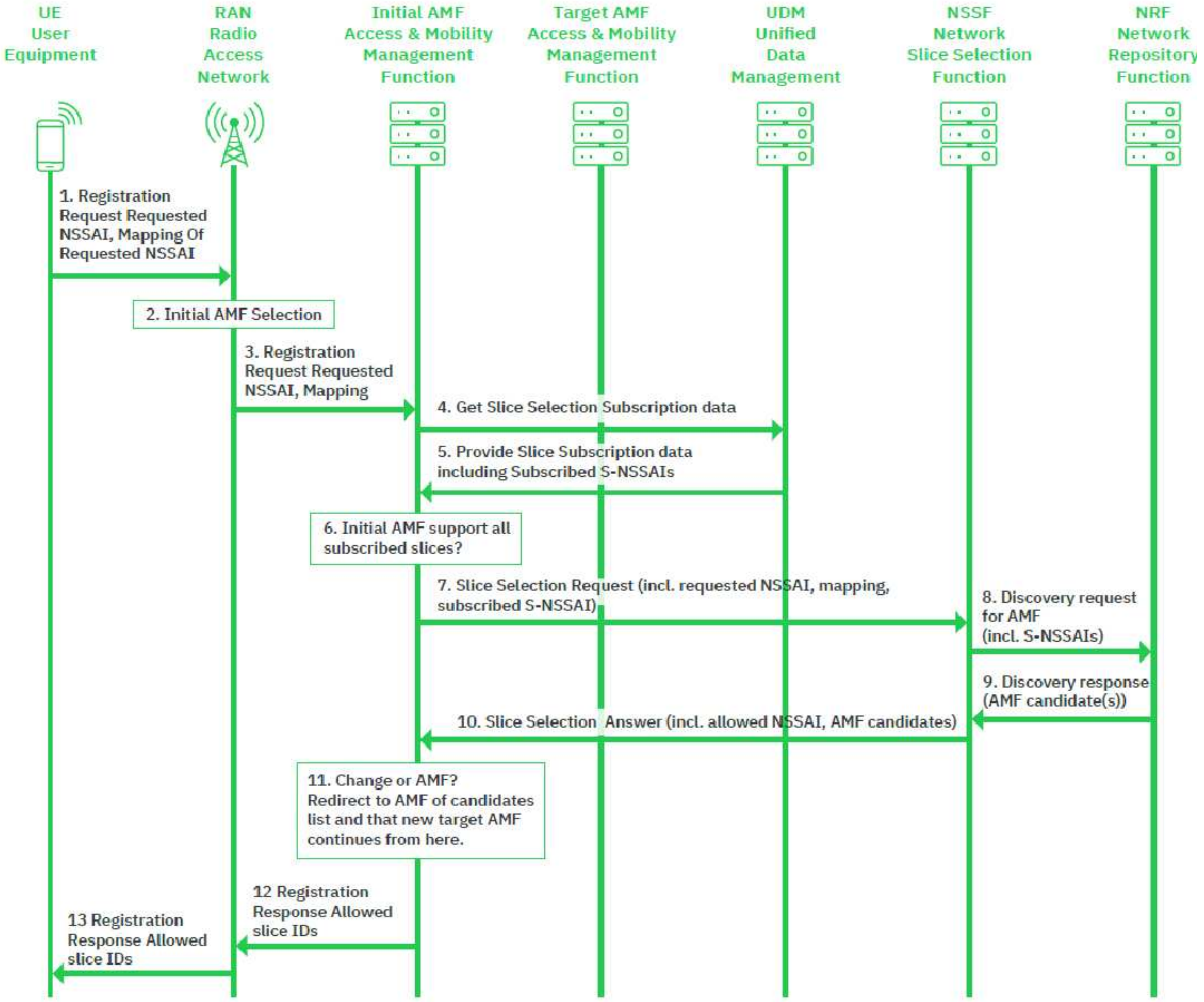

UE
User
Equipment
RAN
Radio
Access
Network
Initial AMF
Access & Mobility
Management
Function
Target AMF
Access & Mobility
Management
Function
UDM
Unified
Data
Management
NSSF
Network
Slice Selection
Function
NRF
Network
Repository
Function
1. Registration Request Requested NSSAI, Mapping Of Requested NSSAI
2. Initial AMF Selection
3. Registration Request Requested NSSAI, Mapping
4. Get Slice Selection Subscription data
5. Provide Slice Subscription data including Subscribed S-NSSAIs
6. Initial AMF support all subscribed slices?
7. Slice Selection Request (incl. requested NSSAI, mapping, subscribed S-NSSAI)
8. Discovery request for AMF (incl. S-NSSAIs)
9. Discovery response (AMF candidate(s))
10. Slice Selection Answer (incl. allowed NSSAI, AMF candidates)
11. Change or AMF? Redirect to AMF of candidates list and that new target AMF continues from here.
12 Registration Response Allowed slice IDs
13 Registration Response Allowed slice IDs

Figure 2-13. Representation of a UE Slice access security process [17]

2.7.3.2 The Slice-Specific Authentication Process

“If the slice use case requires an additional level of security for the UE to access the slice, there is the possibility that other additional authentication of the UE may be required. This requirement is part of the subscription information of the UE.

During the registration procedure of the UE, the slice identity S-NSSAI is used to trigger the usage of the EAP framework (as defined in TS 33.501 with an AAA Server (AAA-S) which can be hosted by the MNO or by a third party, such as a company’s AAA-Server for intranet access iii. The UE needs to support the Network Slice Specific Authentication & Authorization (NSSAA) feature in the UE to perform slice-specific authentication with the AAA server, which was only introduced in 3GPP Release 16.

When the UE connects to the network and the registration procedure is triggered in the AMF (step 3 Figure 8 above). The AMF decides that the network slice-specific authentication and authorization are required for that slice based on the S-NSSAI slice identity. The AMF then takes the role of the EAP Authenticator when communicating with the AAA Server. The NSSAAF (Network Slice Specific Authentication & Authorization

Function) performs the interworking of any AAA protocol and the protocol supported by the AAA Server. The NSSAAF is an interworking function to plug in external authentication servers to the SBA of the MNO. Here we again see a server (i.e. the AAA Server for additional authentication) connected to the core network, which could be a third-party node.

As a summary of section 2, the core network relies on a list of S-NSSAI slice identities concerning the UE identity, which is matched against policies stored in the UDR or UDM. This information prevents unauthorized access from the UE to a slice. But one thing that becomes clear from this is that there are no means to validate if a network function is presenting the correct S-NSSAI slice identity, as the network functions assume that the check between UE identity and S-NSSAI slice identity was performed when the UE was connected. These kinds of missing checks can lead to the security challenge." [17]

2.8 A case study of how Network slicing can be implemented in the City of Leduc

Experts have defined and are already experimenting with numerous use cases of 5G networks. These use cases come from all major industries worldwide, including manufacturing, healthcare, telecommunications, energy, TV and media, transportation, and other infrastructures**.** [9]

A use case of how network slicing can be implemented on a 5G core network such as Telus, which provides different slices to different 3rd parties

With this model, we are using the City of Leduc as an example. Telus uses its core network to provide network slices to different vertical markets or third parties such as RBC, Telus TV, Alberta Health Services and EPCOR.

Per Figure 2-14, the Telus core network provides slices of its core network for various purposes and different QoS characteristics.

A slice in the Telus core network consists of a group of Network Functions (NFs) that support that slice. Those network functions can be assigned only to one particular slice or shared among those different slices. The network functions can be virtual or physical. A physical node may host several network functions in 5G. A shared network function can provide services to several slices; the slices would not be separated on the transport and IP layer.

Some of these network functions might belong to someone other than the hosting MNO, but a third party, so a third party, has access to the core network. Here's an example: a host MNO, Telus, hosting four slices, Slice 1 for Telus TV streaming, Slice 2 for RBC banking, slice 3 for EPCOR (Water and power services), and slice 4 for Alberta Health Services. Slice 1 would have dedicated network functions like a Session Management Function (SMF) for session management, Policy Control Function (PCF) for policy control etc. But this Slice 1 would not only consist of the network functions in the Slice 1 "box" but also the UDR (Unified Data Repository), AMF (Access Management Function) and SEPP from the shared network functions "box." The network functions in the shared box would be available to be used by the other slices and the hosting MNO.

Also, similarly, for Slice 2, RBC Banking would consist of the functions of the Slice 2 "box," and additionally, the shared functions UDR, SEPP and AMF would also be available to Slice 2 as well as the other slices.

A visual representation of the different "boxes" is shown in the Figure. This approach to not-shared shared and slice-dedicated network functions allows for maximum flexibility. The Slice 1 for Telus TV may want to offer low latency and fast streaming content near the users and may take ownership of the UPF function. Also, Telus TV may not want to handle the AMF themselves and may be run by

Telus, as the hosting MNO, may want to enlarge its business but still need to give third parties full access to their network functions, so it is not sharing some of the functions of the network with the third parties in Slices 2, 3 and 4. But Telus may decide to share the AMF function, which manages the radio access and mobility for technical and some convenience reasons and efficiency.

According to [17], all those network functions in the different “boxes” (i.e. the slices, shared and not-shared network parts) are connected to the Service Based Architecture (SBA) of the 5G core network and its interfaces. This is because all those distinct network functions need to exchange signalling messages. In contrast, some may require some form of oversight of another to operate more efficiently.

Two slices may want to communicate with each other so that they would use the common SBA.

An example of such inter-slice Communication is the EPCOR slice having to communicate with the RBC mobile banking slice for subscription payment and other financial purposes.

The MNO manages the network functions in the slice, but the partner might manage the content provided to the user. A typical business case for such an architecture would be a sports stadium company operating the slice for their visitors. The content streamed to the user is coming from an advertisement company.

Network slicing allows the flexible ramping up of network functions, fast and dynamic service deployment, the logical separation of network functions and grouping of network functions. It also enables the support of different use cases and business models. For example, if a slice owner wants to run their own 5G user database or connect their local data serving network to the 5G network. This targeted vertical support and flexibility is impossible in 4G and other legacy networks. But from the design, much consideration must be made about how these slices will be allocated and the various security flaws the various NFs are exposed to.

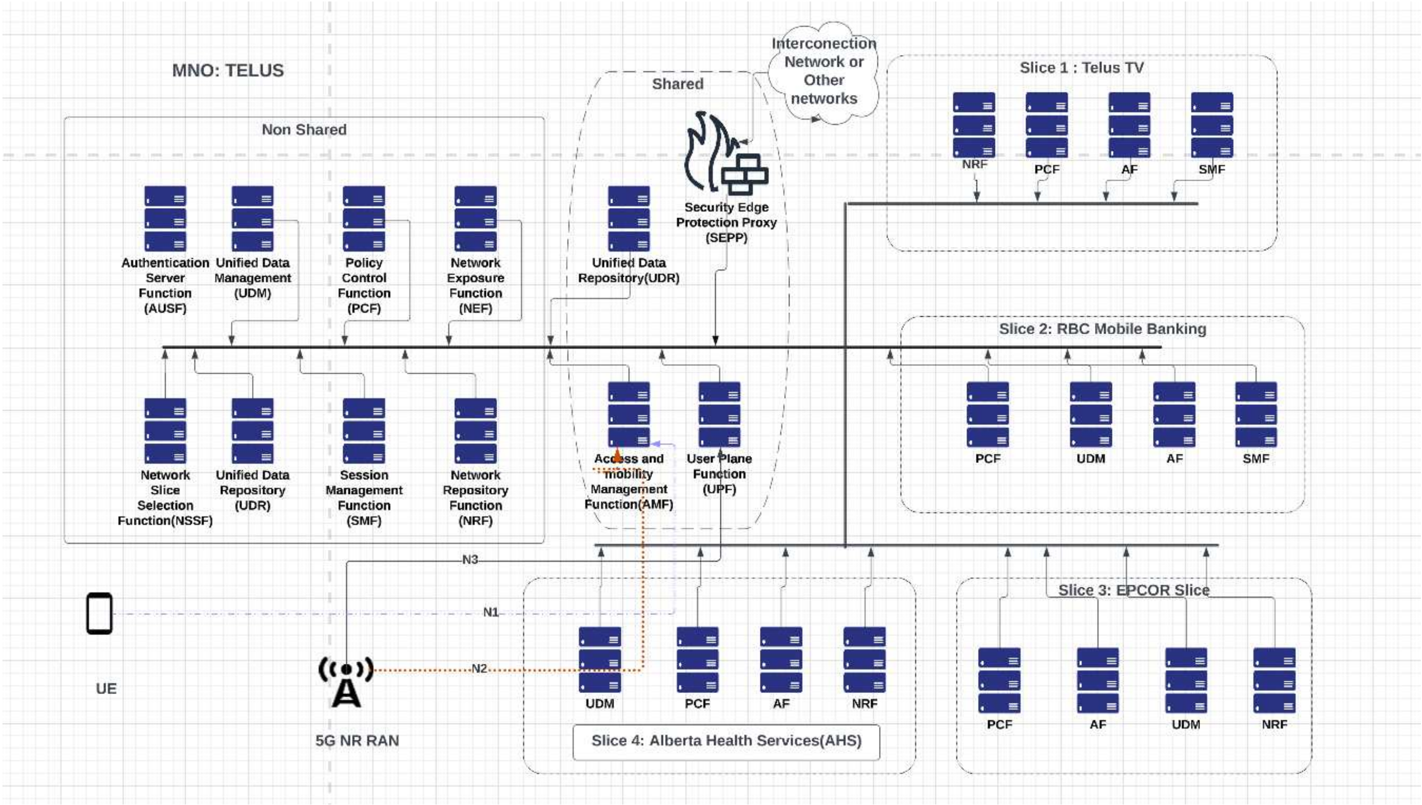

MNO: TELUS
Non Shared
Authentication Server Function (AUSF)
Unified Data Management (UDM)
Policy Control Function (PCF)
Network Exposure Function (NEF)
Network Slice Selection Function(NSSF)
Unified Data Repository (UDR)
Session Management Function (SMF)
Network Repository Function (NRF)
Shared
Unified Data Repository(UDR)
Security Edge Protection Proxy (SEPP)
Interconection Network or Other networks
Access and mobility Management Function(AMF)
User Plane Function (UPF)
Slice 1 : Telus TV
NRF
PCF
AF
SMF
Slice 2: RBC Mobile Banking
PCF
UDM
AF
SMF
Slice 3: EPCOR Slice
PCF
AF
UDM
NRF
Slice 4: Alberta Health Services(AHS)
UDM
PCF
AF
NRF
N3
N1
N2
UE
5G NR RAN

Figure 2-14. A depiction of network slicing of Telus's core network [17]

2.9 SECURITY FLAWS IN 5G CORE NETWORK SLICING

Security in network slicing is a critical problem to be addressed because of resource sharing among slices. Network slices serving different types of services may have different levels of security policy requirements. Therefore, while designing network-slicing security protocols, it is necessary to consider the impact on other slices and the entire network systems. In addition, security issues become more complex when network slicing is implemented on the multi-domain infrastructure. Security policy coordination mechanisms among different domain infrastructures need to be designed. [44]

The most common security threat that a 5G core network is vulnerable to is DDoS attacks which could lead to several slices going down or subscribers being unable to access the services of some network slices. Some could be critical slices like V2X, and some for hospitals and other essential market verticals.

In the early stages of deployment for many operators, there will be a combination of 4G, 5G NSA, and 5G SA networks, all co-existing and with significant coverage overlaps. It becomes essential for the operators to have proper planning as far as devices. The provisioning is considered to ensure that the end-users (UEs) are mapped according to their capabilities and provisioning on the network. [11]

A relationship between various 5G threat vectors and network slicing threats was established beautifully by [41] to explore the highest level of threats to which 5G core network slicing is vulnerable, as displayed in Table 2-2. **A comparison of 5G network slicing threat vectors with relation to network slicing** below.

Table 2-2. A comparison of 5G network slicing threat vectors with relation to network slicing [41]

DoS attack on the signalling plane	DoS on centralized control elements	H
Hijacking attacks	Attacks on SDN hypervisor controller	L
Unauthorized access	Unauthorized access through low-power access points	L
Configuration attacks	Attacks that take advantage of misconfigured system controls	H
Saturation attacks	Ping-pong behaviour in access points and MME due to service saturation	M
Penetration attacks	The malware attack that exposes subscriber info	M
User identity theft	Breaking into user information databases and stealing user credentials	M
Man-in-the-Middle attack	Accessing unencrypted channels or network links and acting as a relay in communications between 2 parties	H
TCP Level attacks	TCP Session or SYN Flooding in gateways, routers	M

Key Exposure	Compromise of the authentication and key agreement	L
Session replay attack	Session keys in a non-3GPP access	M
IP spoofing	Control channels	M
Scanning attacks	Radio interface interference	L
IMSI caching attacks	Roaming and User Equipment (UE)	M
Jamming attacks	Wireless channels	L
Channel prediction attacks	Radio interfaces	L
Active eavesdropping	Control channel	L
Passive eavesdropping	Eavesdropping on the control channel (i.e., inter-Virtual Network Function (VNF) data) can reveal slice configurations and users, enabling hijacking and other attacks.	L
NAS signalling storms	The attack against UE traffic and signalling messages to the core network	M
Traffic bursts by IoT	Saturation of GTP endpoints	M

Based on [41] and deducing from table Table 2-2, it was determined that there are three significant threats that 5G core network slicing is vulnerable to, and they are;

a) **Denial of Service/ Distributed Denial of Service attacks:** These attacks are meant to compromise the availability of a network slice, impacting communication and data access.

b) **Man-in-the-Middle attacks:** meant to impact a network slice's confidentiality, integrity, and availability. With this attack, the attacker can change the content of communication messages, cause misinformation or disinformation, and intercept or get access to information or data between two network entities or endpoints to satisfy a malicious intent.

c) **Configuration attacks:** where malicious actors may try to exploit or take advantage of misconfigurations and other configuration gaps to launch an attack on system controls. Attackers may also change security settings or turn off the security monitoring mechanism.

Some more network slicing security vulnerabilities/flaws are displayed in the table below.

Table 2-3. Some Network slice security flaws were also identified according to a white paper by ENISA [42]**.**

Vulnerabilities in the implementation of NS security functionalities	Vulnerabilities in network segment negotiation procedures could give room to malicious attacks, e.g. man-in-the-middle (MitM) attacks that could modify and downgrade slice capabilities.
Service-Based Vulnerabilities in Network Slicing Management	A malicious party can gain access to an insecure management interface, or if it could replay or modify a valid message, it would be able to spoof a genuine network manager to compromise slice security.
Improper protection of Data and Information	Adequate security controls are needed to protect sensitive data stored, processed and transferred by NSI. Relevant vulnerabilities include: • Improper protection of Network Slice Instance supervision/reporting data; • Lack of/ineffective tamper-proofing of Network Slice Subnet Template (NSST).
Vulnerable mechanisms for authentication and authorization in Network Slicing Management	• Improper slice-specific authentication mechanisms; • Lack of protection of NSSAI and home control; • Lack of protection of the User ID and credentials.
Improper hardening of network slicing components	• Unnecessary or insecure services/protocols; • Unrestricted reachability of services; • Presence of unused software/functions/components; • Unrestricted remote login for privileged users; • Excessive file-system authorization privileges; • Vulnerable OS configuration; • Vulnerable Web server configuration; • Improper separation of traffic.
Virtualization vulnerabilities of relevant network slicing components	• Vulnerabilities in virtualization of OS layer; • Container vulnerabilities; • Vulnerabilities in function virtualization.
Insufficient or improper monitoring mechanism of Network Slice Instance (NSI)	• Insufficient/inadequate logging and auditing across the NSI lifecycle; • Improper protection of security event log files; • Improper isolation of monitoring capabilities and data; • Improper or insufficient end-to-end monitoring capabilities for NSI.

2.9.1 5G NSA Security Flaws

Most MNOs are still deploying 5G on top of their existing 4G infrastructure. NSA configures the core network with an LTE-based EPC and uses eNB resulting in inherent LTE security threats. Existing LTE-based security threats can be exploited to expose the security vulnerabilities for 5G NSA networks. [19]. 5G networks will be used alongside 4G, even with 3G and 2G networks. We must also remember that different operators and countries will move from 4G to 5G at their speeds. Mobile operators must use 5G security to connect with previous-generation networks. [33]. As we gradually transition to an entire 5G core, we will still have to deal with existing 3G/4G security problems; and the risks associated with equipment from untrusted suppliers. The early and current 5G NSA networks must use the LTE control plane protocols and the LTE Evolved Packet Core (EPC) network. The initial launches of 5G NSA are set to provide services for only Enhanced Mobile Broadband (eMBB). This means any threats and vulnerabilities encountered in the 4G LTE networks will still pose some risks to the 5G NSA. Even if we have an entire 5G core implemented or 5G SA with network slicing, there will still be networks that will still be employing the use of the 5G NSA though the 5G SA will leverage neither the same LTE control plane protocols nor the LTE EPC network. [45]

Some common attacks are known to occur in networks before 5G SA could be implemented, especially in instances where MNOs are still deploying 5G SA (a whole 5g core with Network slicing implemented) and 5G NSA networks. These attacks are;

Downgrade Attacks (2G/3G)
Here, an attacker manipulated a phone’s network connection to downgrade to legacy networks, giving attackers or hackers access to security loopholes or weaker security controls in 3G and 4G services. Additionally, these malicious actors could perform man-in-the-middle attacks (MitM) or more passive attacks like eavesdropping to uncover sensitive information. [45]

Man-in-the-Middle Attacks
Integrity Protection security algorithms do not adequately protect over-the-air user plane traffic. A customer's message and communication flow could be intercepted in the middle between the UE and the server. An adversary could manipulate the customer's message and communication flow between the UE and the server. This attack is possible if the customer's Communication is protected by end-to-end security encryption protocols (e.g. SSL, TLS, IPSec, VPN, etc.). Almost all corporate, business and social media communications (e.g. Corporate VPN, Banking, Facebook, Twitter, etc.) are protected by end-to-end encryption protocols. [45]

According to [19], security threats in 5G NSA networks exploit LTE vulnerabilities, including information leaks, target-type user denial of service (DoS), target-type network device DoS, voice eavesdropping and unauthorized data use.

The significant attacks we see on 5G NSA are downgrade attacks, aka version rollback or bidding down attacks described earlier, where an attacker tries to make a connection or protocol drop to a less secure or older version.

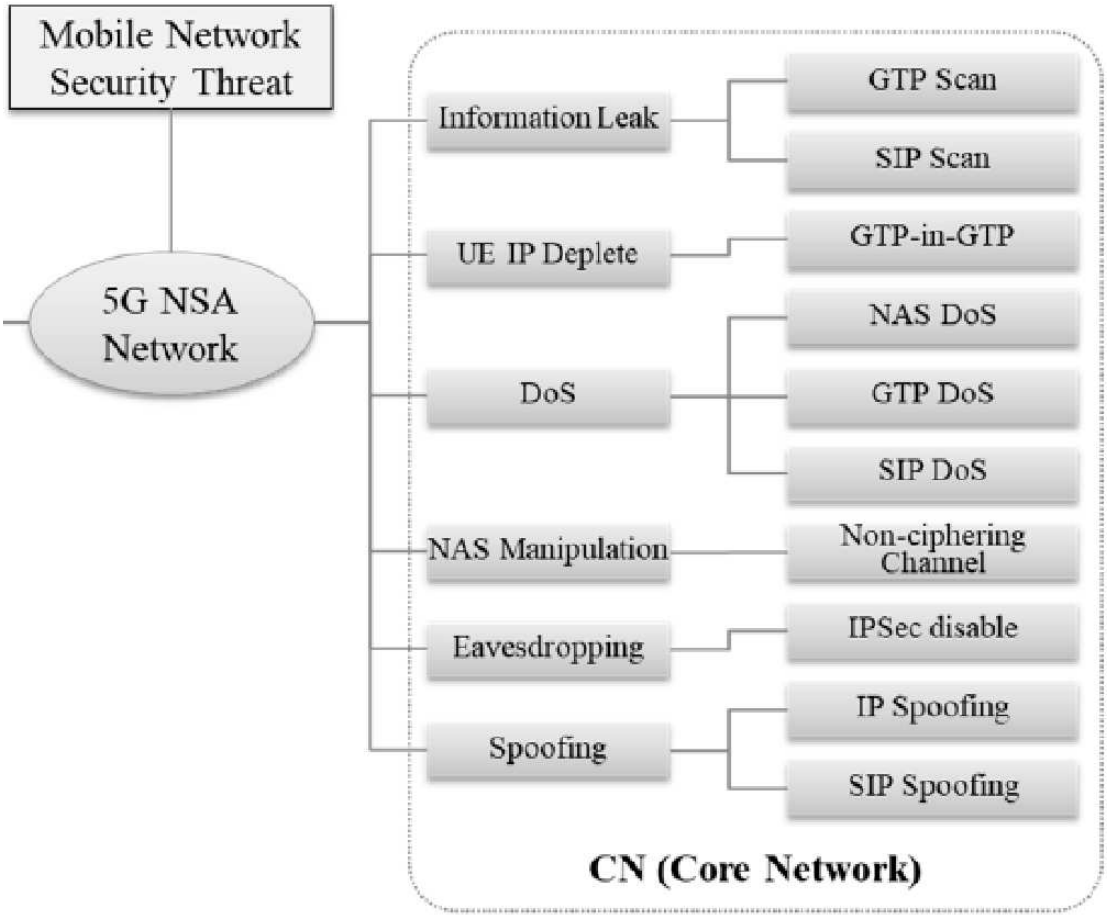


Figure 2-15. Attack tree of the 5G NSA core network [19]**.**

[3] Described six Types of 5G Core Network Security Threats which will have an impact on network slicing

“1. **Information Leak**: Information on 5G NSA core networks can be divided into information on EPC equipment processing data and information on IMS equipment to provide various services. Because EPC equipment communicates using GTP protocol and IMS equipment communicates using session initiation protocol (SIP) protocol, the attacker can select a protocol suitable for the desired information. GTP protocol is divided into GTP-C, used between core network equipment, and GTP-U, which delivers data traffic in the user terminal through a tunnel between the base station and PGW. To find out the IP information of the EPC equipment, the attacker can use a packet injection method that loads an echo request, GTP-C message for health check between core network equipment, on the data payload to send. A packet is created when running Android Debug Bridge (ADB) command in the Android terminal using Packit. When sending the packet to the IP band identified through Tracert in tethering status, the GTP-C packet is injected and transmitted to the mobile communication network. PGW checks this and sends an echo response, where the attacker can identify that the source IP of that message is PGW IP.

2. **IP Depletion**: GTP-in-GTP is the packet injection method described earlier to provoke an information leak threat. The attacker can deplete IP Pools allocated to terminals in the core network similarly. While GTP-C echo request that plays the role of ping is used to acquire IP for core network equipment, GTP-C Create Session Request is injected and sent to the core network to allocate the IP to the terminal. The attacker can increase the terminal number in the create session request sequentially, so PGW assigns multiple IPs. If PGW gives all available IPs, make session requests from regular terminals would be rejected, and all terminals accessing that core network could not communicate.

3. **DoS**: An attacker can send an attach-request message continuously to access the 5G NSA network by configuring multiple terminals as botnets and repeating airplane mode on/off. This may cause excessive traffic load on a specific mobile carrier's core network. For example, one attack request can create a maximum of eight GTP-C messages, which brings eight times the amount of traffic to the CN function in the core network in proportion to one malicious manipulation done by the attacker.

4. **NAS Manipulation**: in NAS protocol messages for signalling between terminals and core network, attach request messages used in the initial attaching step do not guarantee their ciphering or integrity. Therefore, an attacker can install a rogue base station near the victim to steal and manipulate those messages.

In particular, attached request messages have a UE network capability field that can set ciphering or integrity for all data received or transmitted by the terminal. An attacker can manipulate values in EEA, a field to convey ciphering algorithm selected by the terminal and EIA, a field to share the integrity verification algorithm chosen by the terminal, within the UE network capability field. 3GPP technical specification (TS.) 33.401 defines the essential use of integrity verification algorithm in terminals but describes the selective use of ciphering algorithm. The test results conducted by Ruhr University in Germany in 2019 on five European countries and 12 carriers showed that four of 12 need to allow the use of integrity that must be used.

Eavesdropping: Voice communication on a 5G network uses an IMS network and initiates sessions through SIP protocol according to the 3GPP standard. Therefore, Security in SIP protocol is critical, mainly through internet protocol security (IPSec) security associations (SAs). However, IPSec SAs are also selectively done by 5G network operators. Supporting voice-over LTE (VoLTE) does not mean supporting all IPSec because of its significant impact on terminal performance. The Samsung Galaxy S10 model, a recently released 5G terminal, also supports IPSec, but there is a problem in which the setting in question can be turned off through a secret menu. If an attacker can remotely access the victim's hidden menu and change the IPSec setting, the victim's call will communicate without ciphering. If the EEA field is altered through the NAS manipulation described above and the NAS ciphering algorithm is not used, wireless Communication in the AS section is non-ciphered, either. In this situation, an attacker can sniff wireless traffic as a man-in-the-middle (MitM) and eavesdrop on the unencrypted victim's voice traffic as it is.

Spoofing: IP spoofing is a typical network attack. Suppose an attacker changes the IP of data traffic transmitted from every 5G network to the victim's IP and sends the data traffic; its responses are all delivered to the victim, which can cause invalid charging and even DoS. Additionally, SIP or MMS spoofing can be abused for voice phishing. When the "from" header that indicates the outgoing number in the SIP packet header is falsified, the incoming terminal displays that falsified number" [19]

2.9.2 5G SA SECURITY FLAWS

Slice Life Cycle Security Flaws

The Preparation phase

During this phase, the network slice template (NEST) is poorly designed, e.g., with various design flaws, without up-to-date security patches, an attacker could target it by injecting malware into the template. Once an attacker has access to its creation in this phase, they can influence how a slice is created and how further or subsequent slices should behave. This could compromise the integrity of the entire network slice template and put every network entity at risk. [46]

Instantiation, configuration, and activation phase

An attacker can target the API in this phase to create fake slices or change the configuration of slices for some malicious agenda. In this phase, a natural point of attack is the API, whose compromise would permit an adversary to interfere in the installation, configuration, or activation of a slice. [46]

The Runtime or Operational phase:

This phase is exposed to the widest threats, including Denial of Service (DoS), performance attacks, data exposure, and privacy breaks. Besides, management-related threats, such as unauthorized changes in the configuration, persist at run time, and new threats, such as the deactivation of a slice, appear. The API remains a central point of attack in this phase, along with the services that consume the slice. [46]

Decommissioning phase

The main threat during and even after the deactivation of slices consists of exposing sensitive data improperly handled during decommissioning. A second threat is to consume resources improperly freed to mount a DoS attack. [46]

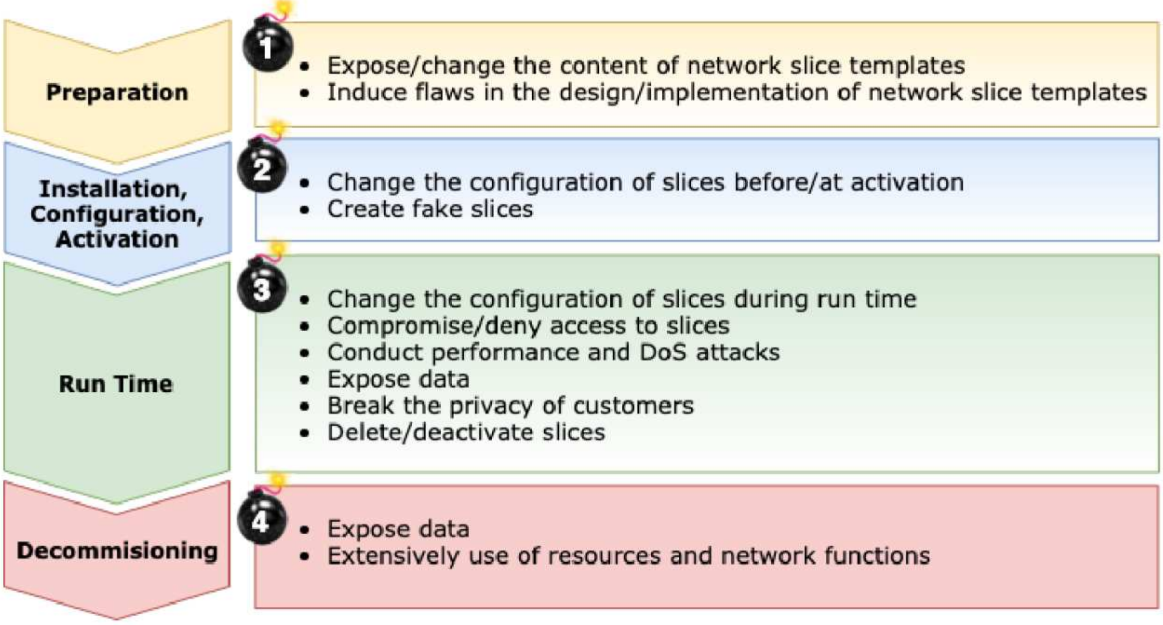

Preparation
1
• Expose/change the content of network slice templates
• Induce flaws in the design/implementation of network slice templates
Installation, Configuration, Activation
2
• Change the configuration of slices before/at activation
• Create fake slices
Run Time
3
• Change the configuration of slices during run time
• Compromise/deny access to slices
• Conduct performance and DoS attacks
• Expose data
• Break the privacy of customers
• Delete/deactivate slices
Decommisioning
4
• Expose data
• Extensively use of resources and network functions

Figure 2-16. A visual representation of threats at each life cycle phase of a network slice [46]

2.9.3 Security flaws as a result of SDN/NFV implementation in Network slicing

SDN and NFV technologies enable network slicing, facilitating the smooth creation of a network slice. Still, these technologies also have some flaws that make 5G core network slicing vulnerable to attacks.

2.9.3.1 Security flaws as a result of NFV implementation in Network slicing

Virtualization brings some new attack surfaces with known vulnerabilities in virtualization environments. If the hypervisor is compromised, other vulnerabilities can arise exponentially. Potential security issues associated with NFVI, considering possible attack scenarios such as VM escape attack, attack on the hypervisor management interface, denial of service (DoS) and DNS amplification attack. [47]

Virtual machines could be created, deleted or moved around the network. Once an attacker has access to this level of power and control, much damage could be done to the core network and how it operates and functions.

The ease with which these network functions or services can be migrated from one point to another or from one resource to another is worrying since the number of services or virtual functions will grow. A primary concern is related to the manual configurations of the virtual systems or VNFs as they grow, which can lead to potential security breaches due to the increased complexity of the growth of the systems. Similarly, the increased number of VNFs is also a significant concern for unauthorized data access, traffic eavesdropping, and theft of services.

Some security concerns for implementing NFV are described below;

Physical security controls are ineffective: Virtualizing network components increases their vulnerability to new attacks compared to the physical equipment at a data center.

Malware is difficult to isolate and contain: It is easier for malware to travel among virtual components running off of one virtual machine than between hardware components that can be isolated or physically separated.

Network traffic needs to be more transparent: Traditional traffic monitoring tools have spotted potentially harmful anomalies within network traffic travelling east-west between virtual machines, so NFV requires more fine-grained security solutions.

Complex layers require multiple forms of Security: Network functions virtualization environments are inherently complex, with various layers that are hard to secure with blanket security policies. [48]

Reduced isolation. With NFV, most components can communicate directly, at least on a physical level. On traditional networks, they are physically isolated. [33]

Risk of sharing resources. Several non-related components can draw on the same hardware resources, impacting each other's performance. Attack on any virtual function can affect other virtual machines running on the same physical server. [33]

Access control issues. How can credentials and access keys be distributed between functions to prevent intruders' access? [33]

Attacks targeting the Management Interface and API

Security challenges are related to web/API vulnerabilities, account compromise, privileged user access, unauthorized access, unauthorized data/packet, inspection/Modification of data, and compromise of MANO components. Improper enforcement of security policies or improper updating of policy rules and data access procedures, allowing attackers to gain access to the NFV MANO module and further perform unauthorized control for all operations. [47]

Without securing these network slice management interfaces, attackers may gain access to the management interface. That access allows attackers to create network slice instances requiring significant network resources or many network slice instances. As a result, the network resources are exhausted, leading to Denial of Service (DoS) attacks. Attackers could incite fraudulent activities, like false charging, by replaying management messages. An attacker may also eavesdrop on the transmission of supervision and reporting data and extract sensitive information to execute attacks of running network slice instances. [32]

Attacks on the hypervisor

A hypervisor is capable of creating and executing multiple guest operating systems. Furthermore, it controls the necessary CPU scheduling and memory partitioning for various systems making it the leading entity in the hypervisor-based virtualized eco-system.

Thus, the hypervisor can be targeted for several attacks, such as exploiting the host operating system to damage the isolation of a network slice, DoS attack on VMs, and VM hopping attacks. [10]

Attacks as a result of the dynamicity of the creation of VMs and VNFs

VMs of VNFs can be created, deleted and moved around a network with great ease, as explained earlier. This raises security concerns since tracking a malicious virtual machine or VNF would be much more complex. This dynamism could also lead to configuration errors since network slicing will involve several VNFs interacting with each other leading to complex configuration commands, which could lead to misconfigurations creating loopholes that expose slices and other core network functions to attackers. Also, tracking malicious devices and developing a trust system between hypervisors, VMs, or management modules will be further complicated. [10]

2.9.3.2 SECURITY FLAWS OR VULNERABILITIES IN THE USE OF SDN TECHNOLOGY

The ability to control a network by software and centralization of network intelligence in network controllers centralizing the network control and ‘softwarizing’ network function opens new security challenges. For example, centralized control will be a good choice for Denial of Service (DoS) attacks, and exposing the critical APIs to unintended software can render the whole network down.

We will address the various security flaws in the three logical layers of SDN as described by [9]

Application Layer

The main security challenges that applications can pose to the network will be due to the availability of open APIs in network equipment, the trust relationship between the controller and the applications (mainly third-party applications and authentication and authorization of applications to change or modify the network behaviour.

Table 2-4. Security flaws in the application layer of SDN [9]

Lack of authentication and authorization	There are no effective mechanisms for the authentication and authorization of applications, which is more threatening in the case of many third-party applications.
Insertion of fraudulent Rules	Malicious applications can generate false flow rules.
Lack of access control and accountability	A problem for the management plane and illegal usage of network resources

Controller Layer

The SDN control plane (e.g. OpenFlow controller) is the centralized decision-making or the central nervous system of the SDN System. Hence, due to its unique role, the controller can be the most targeted or exploited for exposing the network or carrying out malicious activities.

Malicious applications or attackers can acquire network information from the controller if there are no robust authentication and authorization mechanisms implemented or if they are improperly designed and implemented. [9]

Table 2-5. Security flaws in the control layer [9]

DoS, DDoS attack	Due to the visible nature of the control plane, Dos and DDoS attacks are possible.
Unauthorized controller access	No effective mechanisms to obligate access control for applications.
Scalability or availability	Centralizing intelligence in one entity will present scalability and availability challenges.

<u>Data Plane layer</u>

Since the switches are dumb by taking intelligence to the control plane, it will be impossible to differentiate genuine flows from malicious ones. Therefore, the compromised switch can be used for attacks against other switches and the controller. Furthermore, the data plane depends on the control plane's security. If the controller's security is compromised so that it does not provide instructions for the incoming flows, the data plane will be practically offline. [9]

Table 2-6. Security flaws in the data plane layer [9]

Fraudulent flow rules	The data plane is dumb and hence more susceptible to fraudulent flow rules and the limited capability of the switch to buffer legitimate TCP/UDP flows.
Flooding attacks	OpenFlow switch flow tables can store a finite or limited number of flow rules.
Controller hijacking or compromise	Data Plane depends on the control plane, making its security dependent on controller security.

2.9.4 MAJOR NETWORK SLICING SECURITY FLAWS LEADING TO DoS/DDoS Attacks, etc.

Exploring the security flaws due to network slicing, which attackers can exploit to cause DDoS attacks

DoS and Distributed DoS (DDoS) attacks can circumvent the operation of critical infrastructure such as energy, health, transportation, and telecommunication networks. DoS attacks are usually designed to exhaust the targeted devices' physical and logical resources. This threat will be more severe due to the possibility of attacks from machines that are geographically dispersed and are in huge numbers (compromised IoT) [10]

DoS and DDoS attacks originating from large sets of connected devices pose a real threat to 5G networks. These attacks can be against the network infrastructure or the end user devices. Attacks against the infrastructure are designed to deplete the network operator infrastructure resources serving users and devices. [9]

DoS attacks on 5G network infrastructure likely target the resources related to connectivity and bandwidth at promised service levels. Hence, the focus can be on the following areas:

1) The signalling plane needed for authentication, connectivity and bandwidth assignment, and mobility of 5G users;

2) User plane needed to support two-way Communication of devices;

3) Management plane that supports the configuration of network elements that support signalling and user planes;

4) support systems that perform user/devices billing;

5) Radio resources providing access to user devices; and

6) Physical and logical resources supporting network clouds.

[9]

2.9.4.1 How DDoS attacks can be implemented due to improper isolation between slices

The lack of proper isolation between the network slices (Inter-Slice Isolation) and improper isolation between the components of the same slice (Intra-Slice Isolation) can lead to a compromised slice gaining access to the resources and data of another slice due to some network functions being shared among slices.

As shown in the Figure 2-17, still based on our case study, we could have a scenario where an attacker breaches an IoT slice through an IoT device that has been compromised or infected with malware. Furthermore, his threat could be migrated between the slices due to the vulnerability of the IoT device. As a result, an attacker could gain access to another slice, like the RBC banking slice, through some shared network functions.

In the scenario from Figure 2-17, the malware may be able to deplete the resources of the RBC Banking slice or multiple slices, causing DoS (Denial of Service) to an actual subscriber trying to access services. An attacker may also exhaust resources common to multiple slices, causing a denial of service or depletion of resources in other slices. This leads to severe degradation in the offered network services. Just by an attacker gaining access to one slice, say the IoT slice, the RBC banking slice could be targeted, causing several clients to be denied access to mobile banking services.

The resources of the MNO running these slices could also be depleted by this attack leading to large-scale unavailability of services of multiple slices. [32]

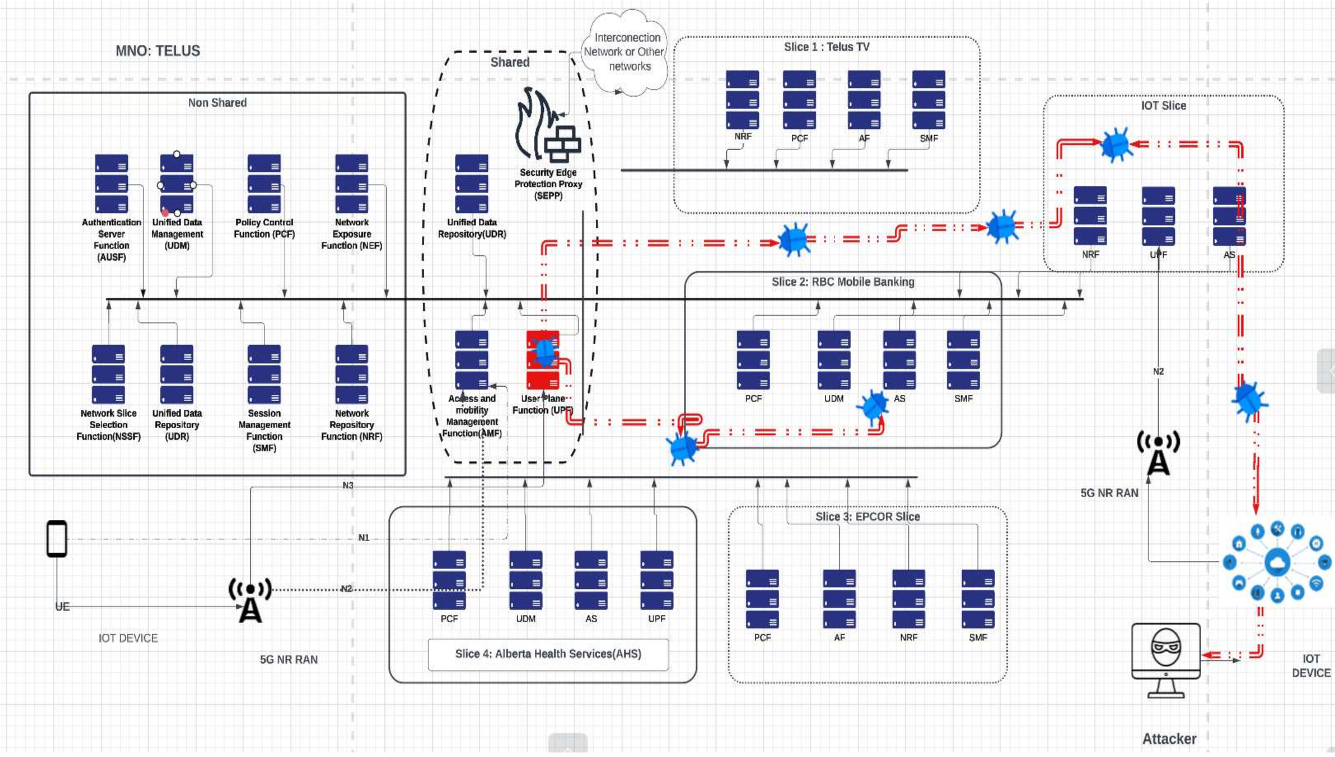

MNO: TELUS
Non Shared
Authentication Server Function (AUSF)
Unified Data Management (UDM)
Policy Control Function (PCF)
Network Exposure Function (NEF)
Network Slice Selection Function(NSSF)
Unified Data Repository (UDR)
Session Management Function (SMF)
Network Repository Function (NRF)
Shared
Security Edge Protection Proxy (SEPP)
Unified Data Repository(UDR)
Access and mobility Management Function(AMF)
User Plane Function (UPF)
Interconection Network or Other networks
Slice 1 : Telus TV
NRF
PCF
AF
SMF
IOT Slice
NRF
UPF
AS
Slice 2: RBC Mobile Banking
PCF
UDM
AS
SMF
N2
5G NR RAN
N3
N1
N2
UE
IOT DEVICE
5G NR RAN
PCF
UDM
AS
UPF
Slice 4: Alberta Health Services(AHS)
Slice 3: EPCOR Slice
PCF
AF
NRF
SMF
IOT DEVICE
Attacker

Figure 2-17. DoS/DDoS attack depiction [32].

2.9.4.2 Denial of Service attack with the use of overload control indicator

This attack could be launched due to a 3GPP overload control indicator flaw.

Dr. Silke Holtmann explained how a DoS attack could be launched using overload control, "The 3rd Generation Partnership Project (3GPP) has a feature of an overload control indicator. It is important that all the network functions can talk to each other, so a feature to avoid overload is important. An overload control indicator is like a do not disturb feature whereby one network function tells another function, do not disturb me for a while; I am busy right now. In 5G, this overload indication is a special header you can put on top of any message. But the information in the header is not cross-checked with the sender's identity. Theoretically, I could put a do not disturb sign for you, and since there is no cross-checking, you would not be contacted with notifications or other messages when a 'do not disturb' request is active. The network functions would not talk to each other for some time; they would use cached data. If you launched another attack during that time, it gives you some cover." [49]

A white paper published by AdaptiveMobile explained further how this attack could be implemented. The security flaw discovered pertained to how elements within the 5G messages in the HTTP header of 5G messages are validated when they are exchanged between network functions. Based on Dr. Silke Holtmann's explanation, the 3GPP TS 29.500 overload control indicator, which can indicate overload between two or more network functions during normal service operations, can be used to initiate a DoS attack. The Overload Control Indicator (OCI) allows one network function to indicate overload conditions to another network function across network boundaries. Such an indication can be put into a service request or other types of messages. The receiving network function is then alerted that the sender is suffering an overload, and measures can be taken to reduce the load. [17] [49]

Figure 2-18 shows what ensues if a rogue slice, Slice 2, wants to run a DoS Attack against another slice, Slice 1

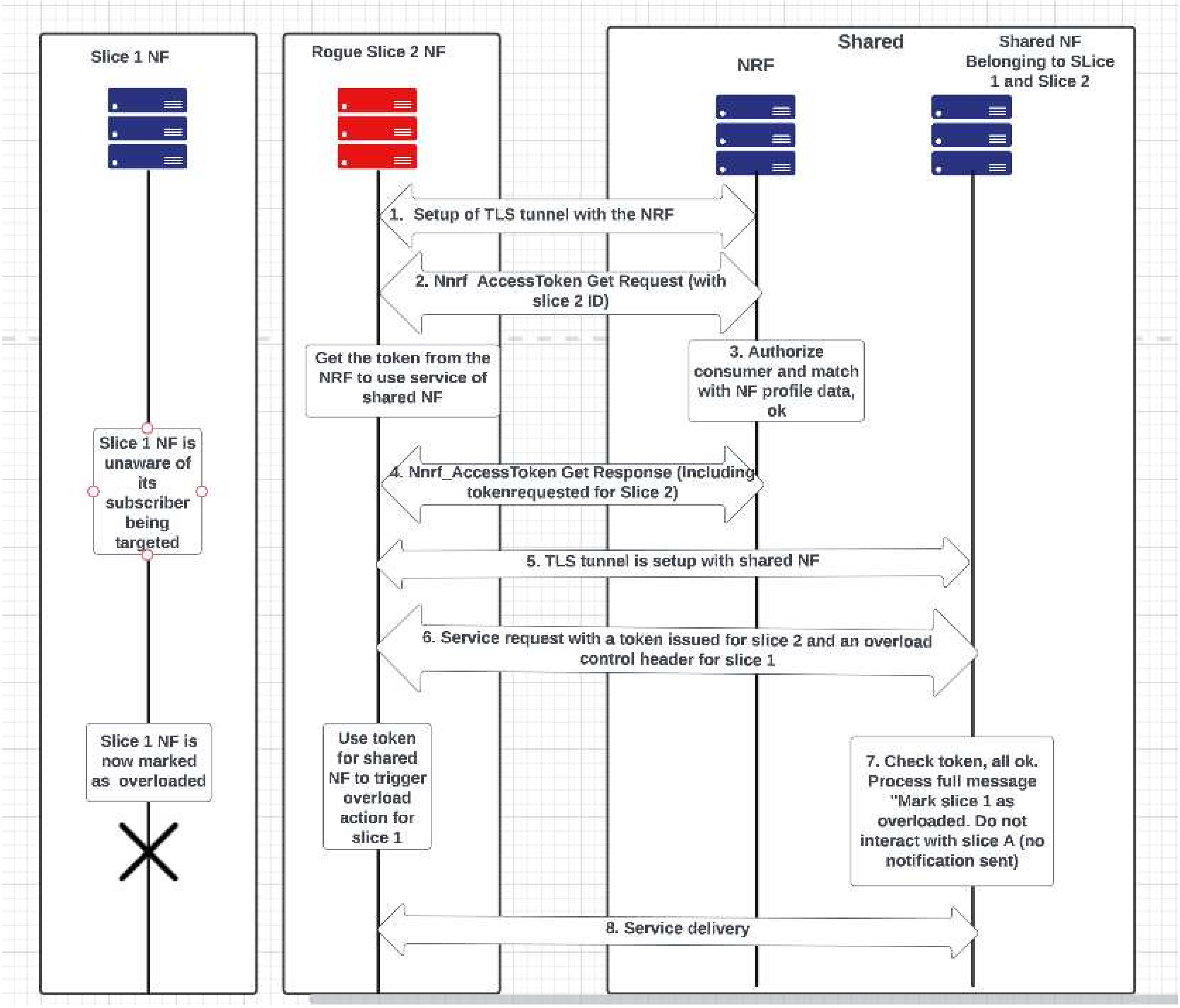

Slice 1 NF
Rogue Slice 2 NF
Shared
NRF
Shared NF Belonging to SLice 1 and Slice 2
1. Setup of TLS tunnel with the NRF
2. Nnrf AccessToken Get Request (with slice 2 ID)
Get the token from the NRF to use service of shared NF
3. Authorize consumer and match with NF profile data, ok
Slice 1 NF is unaware of its subscriber being targeted
4. Nnrf_AccessToken Get Response (Including tokenrequested for Slice 2)
5. TLS tunnel is setup with shared NF
6. Service request with a token issued for slice 2 and an overload control header for slice 1
Slice 1 NF is now marked as overloaded
Use token for shared NF to trigger overload action for slice 1
7. Check token, all ok. Process full message "Mark slice 1 as overloaded. Do not interact with slice A (no notification sent)
8. Service delivery

Figure 2-18. Denial of Service attack using overload control [49]

3GPP did not specify a requirement to validate if the slice identity provided in the 3GPP-SbiOci header matches the slice identity in the token for the service API usage. The slice identity, instance ID or similar information in the token needs to be specified in 3GPP as it requires the usage of the additional scope field in the AuthenticationTokenClaims, which is not defined in detail and would not provide interoperability between network functions from different vendors. This kind of mismatching could lead to misuse of the overload control features of 3GPP, which could result in partial network delays or outages. [17]

2.9.5 SIDE CHANNEL ATTACKS DUE TO THE CLOUD-BASED ARCHITECTURE

Due to the cloud-based architecture, 5GC (5G Core) has all the functions virtualized that provide the added flexibility required for network slicing. However, this leads to another threat vector. Side channel attacks and improper isolation between network slices lead to data exfiltration.

Suppose an attacker can observe or influence how code runs in functions in one slice, say, the Telus TV slice; he may be able to affect the running of code in functions in the RBC banking slice or extract information about the running of code in the RBC banking slice. This may allow side-channel attacks –

particularly timing attacks – that extract information about cryptographic keys or other secrets in the RBC banking slice.

This is critical in sensitive parts of the Telus TV and RBC banking slice. Even the Telus MNO, such as billing, charging and subscriber authentication layers, can be compromised, leading to critical information falling into the hands of an unknown attacker.

As shown in Figure, if the slices and the components within the slice are not adequately isolated, the attacker could access other slice components using the infected device or endpoint in another slice.

And if slices share common hardware platforms, the attacker can observe the crypto code running on one slice and affect the security functions running on another. [30]

Cloud computing systems are more vulnerable to attacks due to their centralized nature. For this reason, practical security algorithms should be designed to avoid the onslaught of malicious users.

[32]

2.9.6 THE 5G SERVICE-BASED CORE NETWORK AND ITS SECURITY FLAWS

The shift to IP protocol used in the SBA architecture for the control and user planes in all network functions has introduced new vulnerabilities and flaws in the 5G core network slicing.

There is ease of communication between network functions with the current service-based architecture as opposed to 4G, where there was some form of isolation between the network functions. [17]

The service-based architecture uses the HTTP/2 protocol and REST API for interaction between all services. Also, the 5G SBA is facilitated by HTTP-based web interfaces and an open API which serves internal communication service functions and data access for service providers of vertical industries. Since there is more knowledge about this web-based API by attackers, there could be exploitation. Open API can also cause the exposure of API functions like SCEF and NEF to the outside network. [50]

Aside from HTTP/2, 5G SBA uses the GTP-U and PFCP protocols. The GTP-U protocol is used to carry user plane (UP) traffic or data from the RAN to the UPF via the N3 interface. The N4 interface also uses the PFCP protocol for Communication between the control plane and user plane on the SMF and UPF since all traffic control and management functions have moved to the SMF. [35]

CSPs or MNOs operating 5G networks worry about whether they can detect the connection of IP traffic transmitted from the user plane path, various types of IoT DDoS traffic passing through 5G networks, DDoS attacks through virtualized network slicing, and abnormal traffic in numerous edge networks. [50]

[35] tried to show through a test how the attacker can leverage the PCFP and HTTP/2 protocol used in the SBA to exploit their various flaws. The Packet Forwarding Control Protocol (PFCP) is used on the N4 interface between the control and the user planes. The security analysis found several potential attack scenarios against various established subscriber sessions.

2.9.7 Denial of service to subscribers due to exploitation of vulnerabilities in the PFCP protocol used in the core network

The SMF uses the PFCP protocol in the 5G core network to create a session on the User plane function (UPF) to manage the GTP channel, ensuring the subscriber gets internet service. Testing was done by [35] on the N4 interface (where the PFCP protocol interacts with the control plane and user plane of the SMF and UPF) based on three procedures by which the PFCP protocol functions, which are Session Establishment, Modification and Deletion. This test revealed more flaws or vulnerabilities that could impact network slices' secure establishment and operation.

2.9.7.1 Denial of Service to a subscriber session by an attacker sending a session deletion request.

Here, an attacker sends a session deletion request to the UPF. As a result, the subscriber loses packet data transmission from a network slice or service it is trying to access, though it remains connected to the network.

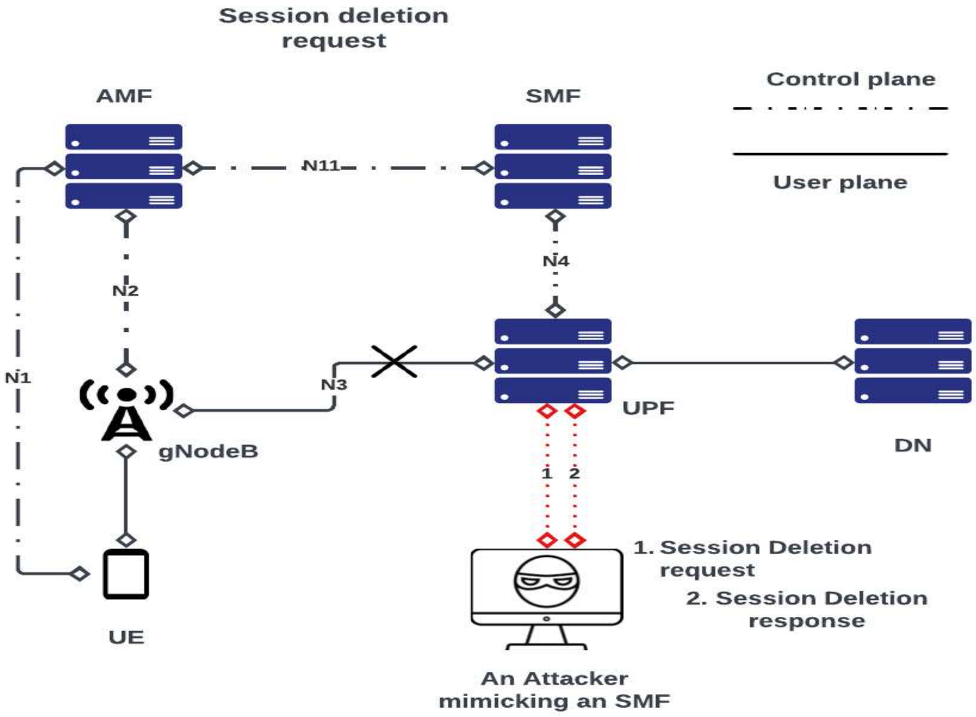
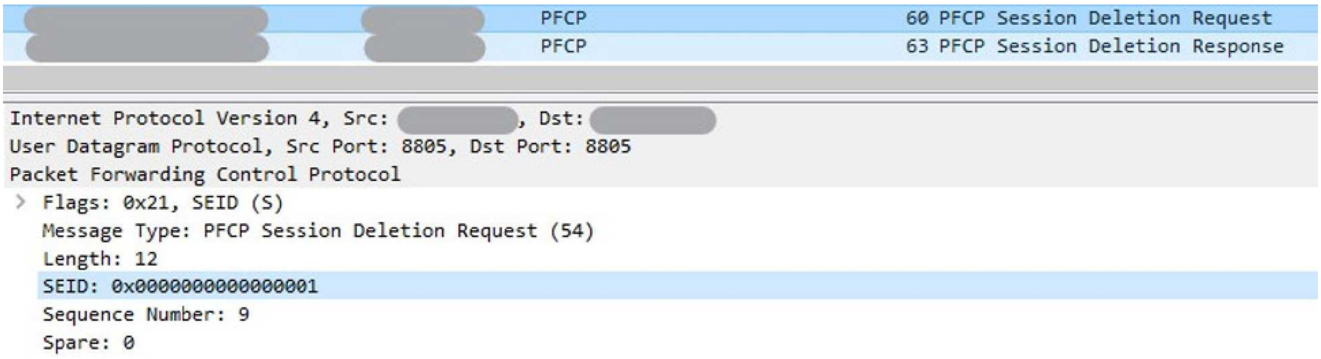

Session deletion request
AMF
SMF
Control plane
User plane
N11
N4
N2
N1
N3
gNodeB
UPF
DN
1
2
UE
1. Session Deletion request
2. Session Deletion response
An Attacker mimicking an SMF

Figure 2-19. Session deletion request scenario according to [35].



```
[redacted]    [redacted]    PFCP        60 PFCP Session Deletion Request
[redacted]    [redacted]    PFCP        63 PFCP Session Deletion Response

Internet Protocol Version 4, Src: [redacted], Dst: [redacted]
User Datagram Protocol, Src Port: 8805, Dst Port: 8805
Packet Forwarding Control Protocol
 > Flags: 0x21, SEID (S)
   Message Type: PFCP Session Deletion Request (54)
   Length: 12
   SEID: 0x0000000000000001
   Sequence Number: 9
   Spare: 0
```

Figure 2-20. Test showing the deletion of the subscriber session.

2.9.7.2 Denial of service and MitM attacks via Session Modification Request

An Attacker sent a session modification request containing a 'DROP' flag resulting in the deletion of important information on the UPF, such as the IP address of the base station or slice instance. Thus, the subscriber or UE's connection is severed, preventing access to the internet or the network slice it is trying to access. The attacker can also operate as a man in the middle and alter information the subscriber or UE has requested by redirecting traffic from the UPF to an attacker-controlled resource.

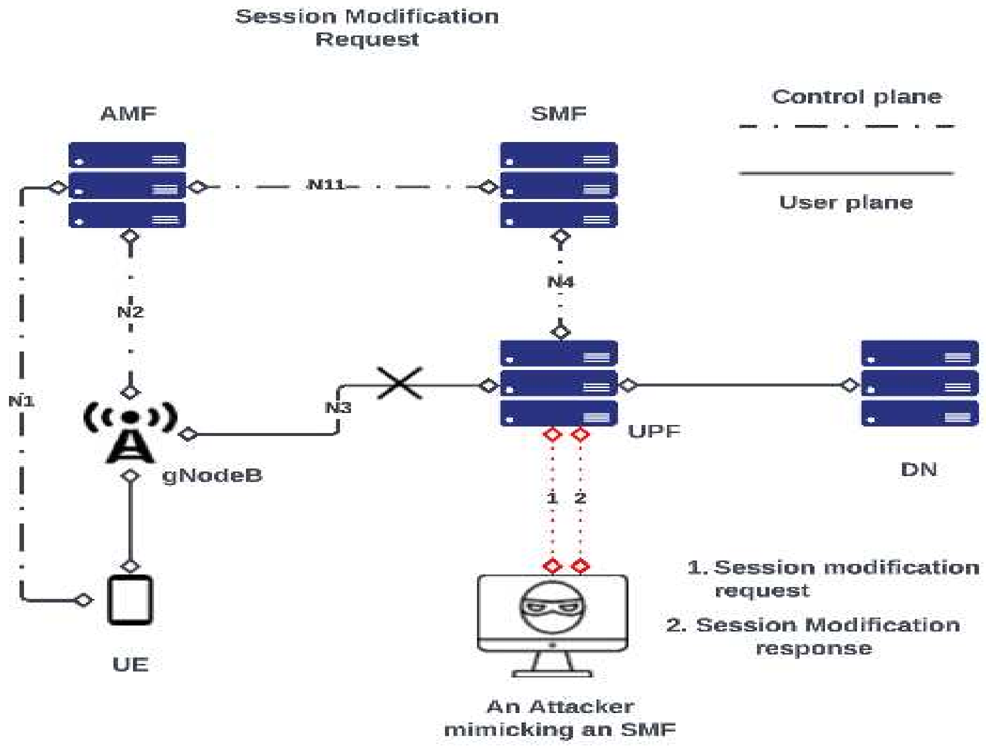

Session Modification Request
AMF
SMF
Control plane
User plane
N11
N4
N2
N1
N3
UPF
gNodeB
DN
1 2
UE
1. Session modification request
2. Session Modification response
An Attacker mimicking an SMF

Figure 2-21. Denial of service attack scenario [35]

2.9.8 How DDoS attacks can be implemented on network slices due to a vulnerability in NRF and HTTP/2 protocol

In this section of the report by [35], the authors considered the Network Repository Function and subscriber authentication vulnerabilities.

The Network Repository Function registers new network functions and stores profiles. It also receives requests to discover available NFs that meet specific criteria.

This was clearly described by (A Slice in Time: Slicing security in 5G Core Networks White Paper, 2021) offers three primary services:

1. Managing the network functions and network function instances by registering, updating and de-registering their profiles in the NRF.

2. Network Function Discovery, where other network functions can detect the services offered by other network functions or instances within a network. It is also be used by an NF to query another NRF in another network

3. It offers an OAuth2 Authorization service, allowing a network function to access the services of another network function through tokens.

Based on [18] requirements, the network function and the NRF authenticate each other with TLS or IPSec. One issue discovered is that the security was provided on a much lower level of the OSI model but needs to provide authentication for individual sensitive information components.

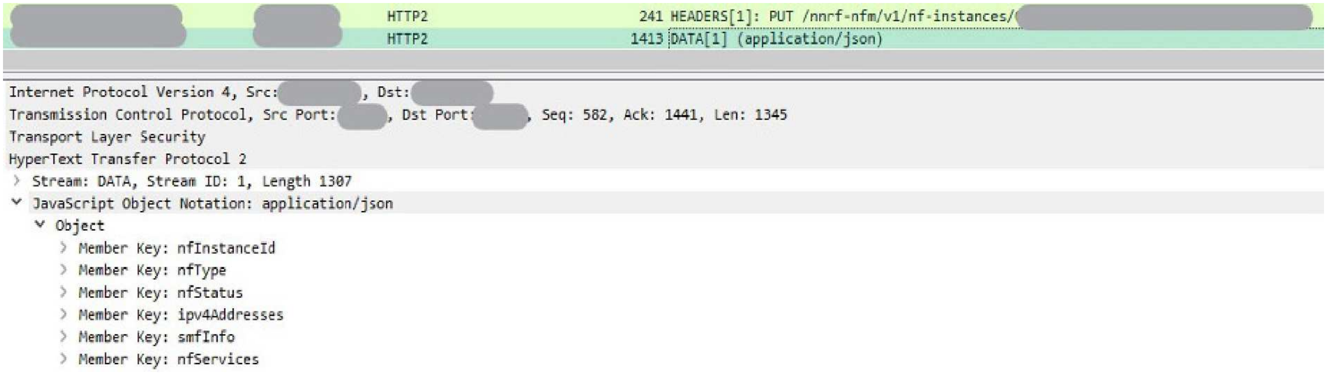
Also, a test was conducted by [35] to determine how secure the three procedures of Registering a new NF, Obtaining the NF profile and Deleting the NF profile

The researchers found that "none of the components verify the TLS certificate when connecting" [34]. Thus, there was no verification made on the TLS certificates by the various components during the connection process. Thus there was no process for service authorization on the NRF

The multiple scenarios during the management of the network function can be seen below:

2.9.8.1 During the registration of a new NF

The interface description and the NF's unique number are stored in an Instance ID header, as shown in the Figure below.



```
[redacted]          [redacted]     HTTP2             241 HEADERS[1]: PUT /nnrf-nfm/v1/nf-instances/[redacted]
[redacted]          [redacted]     HTTP2            1413 DATA[1] (application/json)

Internet Protocol Version 4, Src:[redacted], Dst:[redacted]
Transmission Control Protocol, Src Port:[redacted], Dst Port:[redacted], Seq: 582, Ack: 1441, Len: 1345
Transport Layer Security
HyperText Transfer Protocol 2
 > Stream: DATA, Stream ID: 1, Length 1307
 v JavaScript Object Notation: application/json
   v Object
     > Member Key: nfInstanceId
     > Member Key: nfType
     > Member Key: nfStatus
     > Member Key: ipv4Addresses
     > Member Key: smfInfo
     > Member Key: nfServices
```

Figure 2-22. Registration of a new NF [18] [35]

According to the test by [35], An attacker can try to register the same network function to serve a UE or subscriber, making the UE believe it is accessing a legitimate NF in a particular slice instance or network slice, thus giving the attacker the ability to access sensitive information.

2.9.8.2 Impersonation attacks by an attacker obtaining the NF profile of a slice

An attacker who obtains the profile of a network function will be able to use it in subsequent attacks that require indicating the Instance ID in the request body.

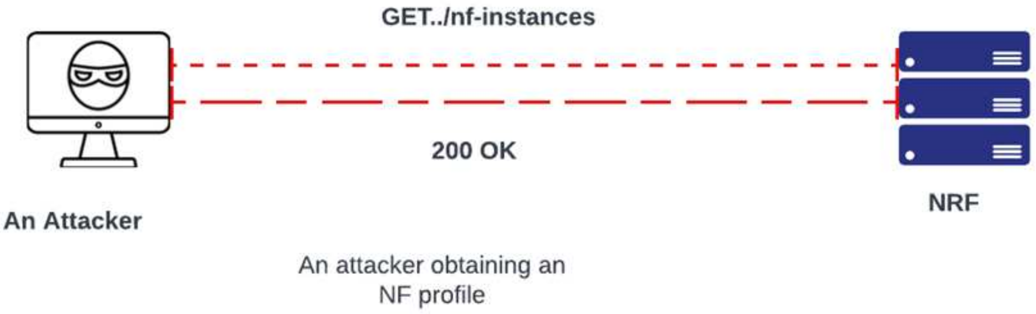


Figure 2-23. An attacker obtaining the NF profile of the network function of a slice based on [35]

Attackers can then impersonate any network service for other NFs and obtain profile data, such as authentication status, current location, and subscriber settings for network access.

2.9.8.3 Deleting the NF profile of an NF in a slice or multiple slices by an attacker due to a lack of restrictions set by the NRF

An attacker can quickly gain access to various NF profiles of several slices and deactivate or delete them, causing several subscribers to lose services or access to some network functions in a slice or an entire network slice they are accessing. This will cause financial loss to an MNO and eventually lead to a loss of trust and confidence.

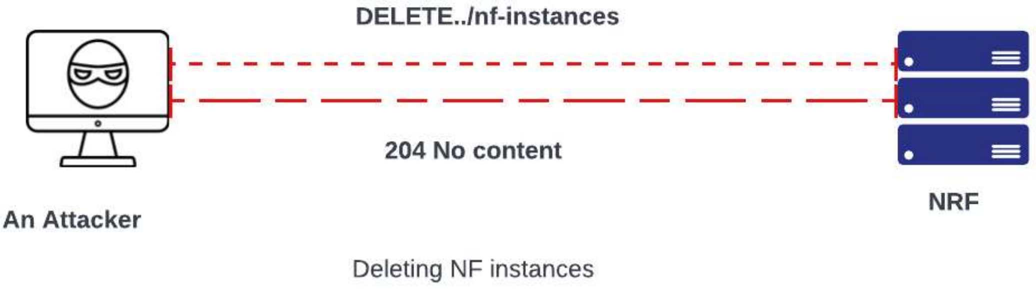


Figure 2-24. Deleting an NF profile [35]

2.9.9 SECURITY FLAWS AS A RESULT OF 5G COMPLEXITY LEADING TO MISCONFIGURATIONS.

Increasing the number of slices on a 5G network may lead to more configuration errors and even deterioration of operator awareness, adversely impacting security overall.

As the configuration burden and the number of parameters increase, the higher probability of a security slip up. This may be especially true when 5G network infrastructure is built jointly by several operators or when a single 5G network is shared by several virtual mobile operators [33].

Also, the progress from 4G to 5G represents a massive increase in protocol complexity. [17] Analyzed the protocol structure of 4G and 5G core network protocols to compare complexity. The study showed that,

- there are 4.7 times as many types of commands (messages) that can be sent inter-MNO over 5G, compared to 4G

- Over 3.4 times as many information elements (attributes) sent inter-MNO over 5G, compared to 4G

The study revealed that 5G is several times more complex. This has obvious security implications because these commands need to be inspected and many elements approved to prevent illegal or unwanted activity from being sent to the network. This becomes more difficult the more commands and information element types are received.

2.9.10 Security Flaws in the MANO

According to the MANO abuse model in the white paper [51]

An attacker targets the MANO component remotely with valid account credentials to be able to access the 5G core network. The attacker then alters the firewall settings to stop or evade all the security controls to access a network slice, exposing the various NFs to the malicious entity. Once this attacker has access, critical NFs could be targeted. A DoS attack to deplete the resources is issued, resulting in all services of all user equipment trying to access the particular slice getting denied.

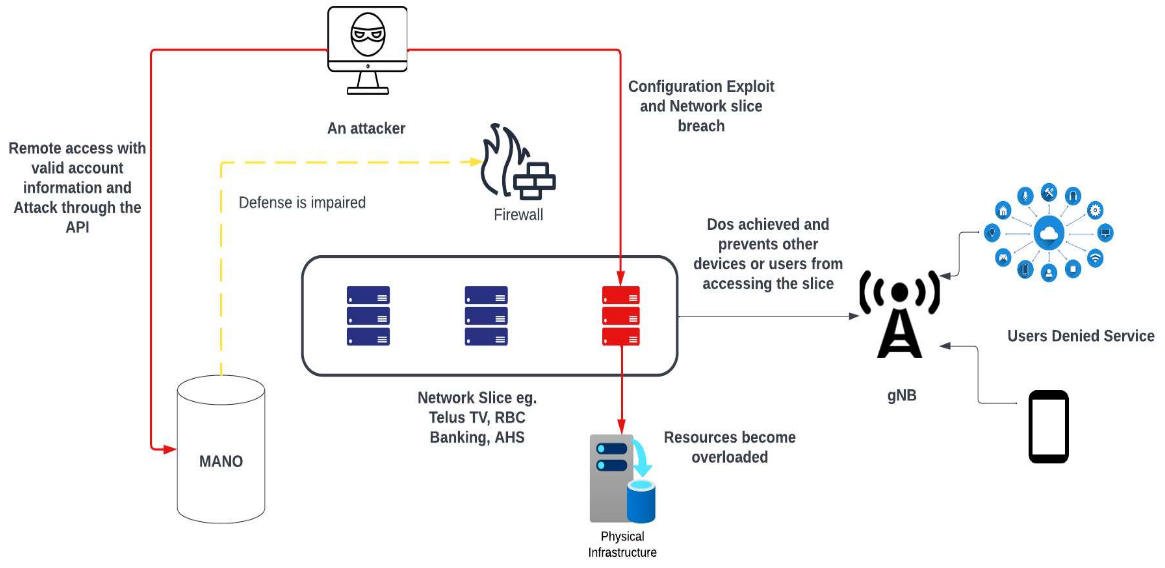

An attacker
Remote access with valid account information and Attack through the API
Configuration Exploit and Network slice breach
Defense is impaired
Firewall
Dos achieved and prevents other devices or users from accessing the slice
Users Denied Service
gNB
Network Slice eg. Telus TV, RBC Banking, AHS
MANO
Resources become overloaded
Physical Infrastructure

Figure 2-25. MANO abuse attack scenario [51]

2.9.11 Information Leakage due to the lack of security zones in the SBA architecture

According to [17], One of the major security flaws discovered is the need for more security zones within the SBA architecture, which has a significant challenge of proper configuration and how authorization works on just the slice and service level. According to their test, the authorization token is limited to services, network functions and the identity of a slice. Requests usually contain several information elements or factors to provide services. Thus, verification between the information elements and network slice identity is imperative in ensuring that a malicious or rogue slice or actor does not trick the NRF into providing information about another slice.

An attack was simulated to show this flaw in the diagram, and this was redesigned to match our case study.

According to the simulation in Figure 2-26,

In the 1st step, an AMF authorizes retrieval of the location information of the UE. Then, a compromised and malicious network function belonging to slice 2, e.g. the RBC banking slice in our case study, sets up a TLS connection with the NRF.

This malicious network function then sends a request to use the services of a particular network function commonly shared between Slice 2 and slice 1, e.g. the Telus TV slice. Since the AMF is shared between slices 1 and 2, the NRF is tricked into generating a token in the 3rd step and sends this token to the compromised NF belonging to slice 2, the RBC Banking Slice, in the 4th step. It then creates a request to the shared network function that has the token for its slice, which is slice 2, but the identity of the UE from the actual service request is from that of slice 1, the Telus TV slice in the 5th step. The shared NRF consequently assumes that all authorization is done and passes the token as valid. The gap here is that the shared NF did not validate in the 6th step if the UE belonged to RBC Banking (Slice 2) but verified the correctness of the NRF token issued for the RBC Banking Slice (Slice 2). Finally, the user's location is revealed in the 7th step based on the assumption that the request is valid. Looking at the entire process, no verification was made for the UE to ensure that the service request sent from the compromised or malicious network function should be available to the RBC banking slice (Slice 2).

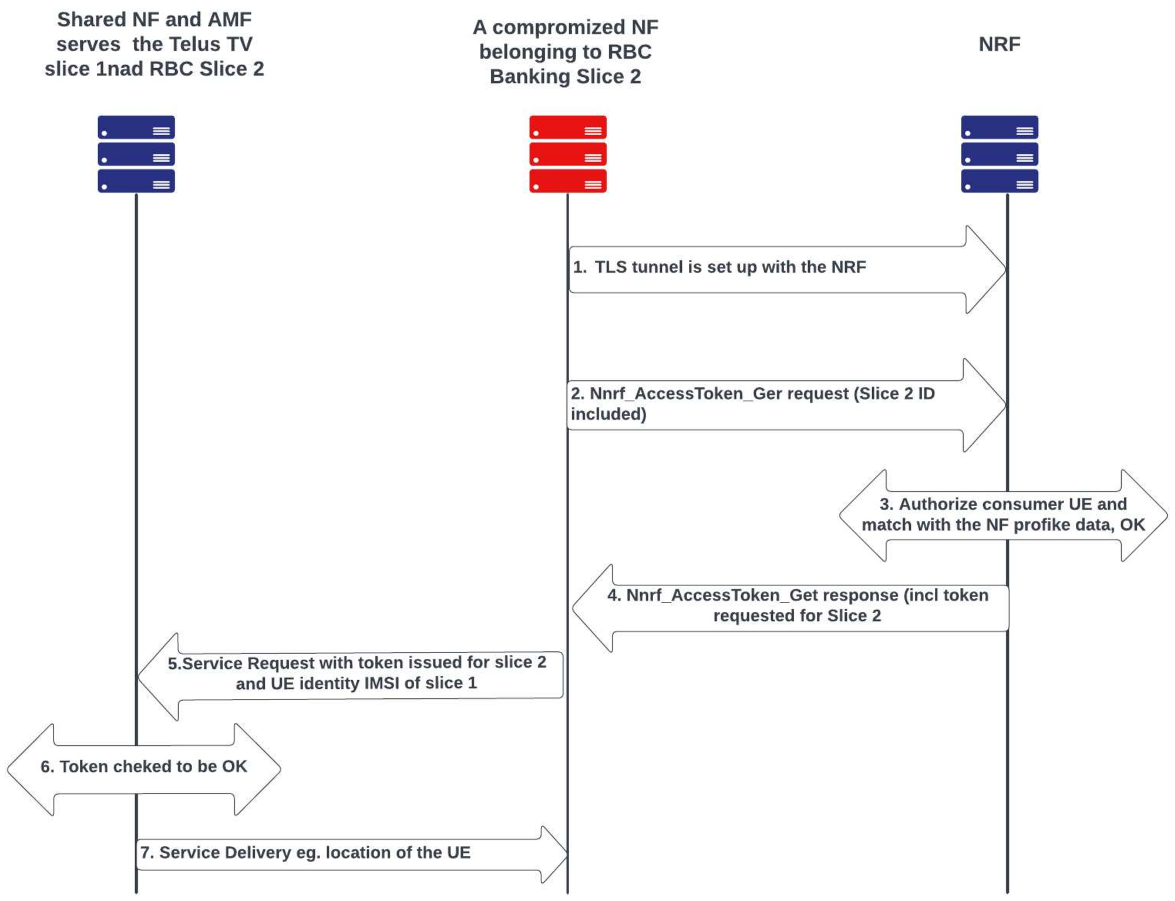

Shared NF and AMF serves the Telus TV slice 1nad RBC Slice 2
A compromized NF belonging to RBC Banking Slice 2
NRF
1. TLS tunnel is set up with the NRF
2. Nnrf_AccessToken_Ger request (Slice 2 ID included)
3. Authorize consumer UE and match with the NF profike data, OK
4. Nnrf_AccessToken_Get response (incl token requested for Slice 2
5.Service Request with token issued for slice 2 and UE identity IMSI of slice 1
6. Token cheked to be OK
7. Service Delivery eg. location of the UE

Figure 2-26 Leakage of information between shared and dedicated NFs due to a compromised NF [17]

[52]**_Spelt out some 5G core network slicing flaws that must be addressed before full-scale implementation can be realized.**

2.9.12 Lack of robust authentication mechanisms leading malicious attacks against Network Slice Manager or Host (physical) platforms within an operator network

The network slice manager is responsible for the dynamic creation and destruction of network slice instances while mapping them to host platforms. The network slice manager must be able to verify if a particular host platform should be allowed to run a specific slice or a set of slices. Likewise, the host platforms should also be able to verify that the service that the Slice manager is offering is from a legitimate slice manager and not from a compromised one. Suppose these gaps are not closed, and there are no special authorization, authentication and access techniques designed, there will be an abuse of these flaws by malicious actors to impersonate or perpetuate man-in-the-middle attacks or DoS attacks that will impact both the operator network and customers trying to access a network slice. Also, critical information about slices could be exposed to malicious actors. [52]

2.9.13 Different security protocols or policies in different slices

"If different slices offer different services, then those services may have different performance constraints and security requirements. For instance: The service in one slice may require extremely low latency, which constrains the security protocol in some way (e.g. affecting key derivation on service setup or key management on inter-cell handover). The service in one slice may require an extremely long device battery life, which constrains the security protocol in some other way (e.g. how often re-authentication is performed). The service in one slice may be very privacy-sensitive, requiring unusually intensive security procedures (e.g. frequently reallocating temporary identities). The fact that some aspect of security is constrained in one slice shouldn't mean that it has to be similarly constrained in all slices. It is natural, therefore, to expect that security mechanisms will vary somewhat between slices – different "tuning" of security protocols (such as frequency of re-authentication) or possibly even different protocols. However, where security varies between slices, we need to consider how well those slices are isolated from each other. If someone can attack the "lower security slice, "can they also impact the "higher security slice?" We also need to consider the network security as a whole: if someone can attack a "lower security slice," can they impact the whole network?" [52]

2.9.14 Exhaustion of security resources in other slices

Suppose an attacker wishes to do something malicious to Slice 2, such as the RBC Banking Slice. Normally, Slice 2 would run its standard security protocols and checks, preventing the attack. But now, suppose the attacker can exhaust resources in Slice 3, i.e., The EPCOR Slice, in a targeted and well-timed way, with the result that Slice 2 is short of resources and unable to run its standard security protocol; The attacker can now proceed to their earlier objective of accessing Slice 2 (RBC Banking Slice). [52]

3 Chapter 3: Security recommendations, techniques and solutions for the security flaws of 5G core network slicing

3.1 Network Slice Isolation

Isolation is one of the best techniques that can be used to protect and secure network slices from DoS and DDoS attacks which are common in network slice implementation. By isolating slices, we can provide slice-specific security deployments based on the requirements of each network slice. Isolation can be conducted between network slices, network functions, users, market verticals, etc. [46]

Isolation of slices can also be done physically and logically. So, we can have dedicated resources or separation of highly critical slices like the Alberta Health Services or RBC slice to ensure that such crucial information like health and customer banking information does not interfere with that of other less critical slices or less secure slices. These slices could have separate physical hardware to ensure a high level of security due to how essential they are in nature.

Where physical isolation is impossible, logical techniques and mechanisms must be implemented.
Some technologies like firewalls, gateways, and hypervisors can be used to ensure isolation between slices. The creation of security trust zones is also a feasible isolation measure. Some enabling techniques include physical resource block scheduling, slice scheduling, and traffic shaping. (NSA, 2022) Some slices based on implementation needs have to communicate with each other, which must also be considered in the isolation techniques used. (NENCIONI2, 2020)

By isolating network slices, we can also effectively control inter-network slicing communications and ensure much-needed security visibility, monitoring, authentication and encryption.
In addition to network slice isolation, multi-layer isolation could also reduce the attack surface and lessen the impact. Examples of multi-layer isolation include NFVI boundary isolation, isolation of MANO system, security domain isolation, service instance isolation, VNF isolation, etc. Various technologies and software/hardware cryptography must be adapted to the desired isolation levels. The technologies include different software, hardware, and cryptographic mechanisms. These implementations may cover a range of options, including managed containers, hypervisor-managed virtual machines, and VPNs. [32]
In the paper, Towards Secure Slicing: Using Slice Isolation to Mitigate DDoS Attacks on 5G Core Network Slices by (Sattar & Matrawy, 2019), a test was conducted where network isolation was used to tackle the challenging problem of Distributed Denial-of-Service attacks in 5G network slicing. They used a mathematical model that can provide on-demand slice isolation and guarantee end-to-end delay for 5G core network slices. Finally, they evaluated their work with results proving that the sequestration of network slices and other critical entities in the network could mitigate Distributed Denial-of-Service attacks and increase the availability of the slices.

3.2 Secure communication between slices and between slices and UEs.

Attacks or faults occurring in one slice must not impact other slices. Moreover, each slice must have independent security functions that prevent unauthorized entities from having read or write access to slice-specific configuration/management/accounting information and be able to record any of these attempts, whether authorized or not.
Separate authentication of a UE accessing multiple slices at once. Suppose a UE can access various slices simultaneously, and those slices have different security levels concerning network access. The operator policy should request the UE to authenticate separately for each slice. Otherwise, the UE may establish itself to the 'lower security slice' and thus be allowed access to the 'higher security slice.' [52]
Also, a security protocol should not be accepted for one slice if it is considered weak; network configuration and security parameters (such as authentication frequency) should not be set lower in any one slice than would be regarded as acceptably secure for the network as a whole.

3.3 Security Recommendations for SDN and NFV implementation in 5G core network slicing

Some Security recommendations for NFV include**;**

- A mitigation technique for networked DoS attacks in virtualized systems is the implementation of firewall proxies; this will require an ACK to be received on the client side before an attacker's request can be forwarded [10].
- Distributed VNFs can be deployed, increasing the system's scalability and availability to resolve DoS and DDoS attacks.
- Constantly validating security characteristics of VNFs
- Identity and access management for controlling who can use and manage VNF.
- Security monitoring – i.e. correlating and analyzing data collected from the user data and management and control planes – is an essential enabler for automated security orchestration. It is also necessary for detecting intrusions against the NFV framework and functions.
- Function isolation depends on the operating system, hypervisor level mechanisms and physical controls.

- Communication security protocols are needed to protect the authenticity and confidentiality of communication between the framework elements.
- Firewalling, zoning and topology hiding solutions are needed to protect the virtualization framework against external threats.
- Trusted computing technologies, such as secure boot and attestation, are needed to ensure and attest the integrity and trustworthiness of the software running in shared hardware and NFs [30].

Some security recommendations for SDN implementation are;

- Increasing the security of the control plane through increasing the control plane resources, distributing control functions among multiple nodes, and enhancing the access control procedures
- The SDN controller should not be used without appropriate authentication and authorization checks.
- Separating different users from each other and reserving resources so that DoS attacks performed in one data flow do not affect other users. This can also ensure that one user is not able to see the traffic flow of another user
- Thorough verification of SDN applications before they are granted access to network configurations through the control plane.

3.4 Resource capping to protect against DoS/DDoS attacks on Network slices

An attacker can exhaust resources in one or multiple slices and cause service degradation in that particular slice. Sometimes resources can be common to various slices, which DoS attacks can deplete. While allocating the network resources for security to individual slices, they should be given with minimum level of resources, or resources have to be capped. An attack on one slice may cause the exhaustion of resources in another slice, launching a DoS attack. These attacks can be avoided by running ring-fencing resource allocation policies in the security protocols
among slices. [30] [52]

3.5 Network slice managers and host platform security

Network Slice Managers and host platforms should support mutual authentication. Network Slice Managers should authenticate the hosts before activating a slice instance. Host platforms should also certify the Slice Manager before loading a slice instance and running on the physical hardware. In addition, mutual authentication between slice managers should be a requirement where multiple slice managers are involved in instantiating an end-to-end network slice. Alternatively, Network Slice Managers should authenticate the host platforms in the networks they control on which they want to load network slice instances. Hosts platforms should also authenticate the Network Slice Manager before allowing the instance of the network slice to be loaded. [32] [52]

3.6 Security Zoning

AdaptiveMobile [17] suggests potential attack paths that exploit the lack of cross-validation between various layers. Based on their study, a pure IP layer firewall or general transport layer security solution gave a false sense of security because its controls can be bypassed on the signalling layer, cannot provide a total or more holistic approach to protecting network slices and lacks the required understanding of the interaction between layers, such as whether the slice identity in the actual signalling layer request matches the transport layer, or if a UE identity belongs to a slice or not.

They proposed various points where based on the analysis of these challenges, some key points where signalling security filters could be installed between zones:

- Between roaming partners and hosting network

- Between not-shared and shared network functions
- Between third-party and shared network functions
- Between 5G and legacy network elements
- Between the core network and external partners
- For inter-slice communication between slices

By relating this to the Telus MNO scenario, based on the [17] design and illustration in Figure 3-1, we could have a secured and filtered security zone for inter-slice communication, e.g. The Telus TV and RBC Banking slice.

Also, network functions (NFs) are shared between the slices and the host MNO, in this case, the Telus MNO. Those NFs are potentially vulnerable and must be protected from attacks on the signalling and information element level through fine-grained filtering. This zone would also belong to the SEPP for slices which require roaming support. The SEPP would offer protection of the slices from attacks from the roaming network.

The network functions dedicated to the host MNO and shared with slices would require their security zone, which must be protected on the signalling layer from malicious requests.

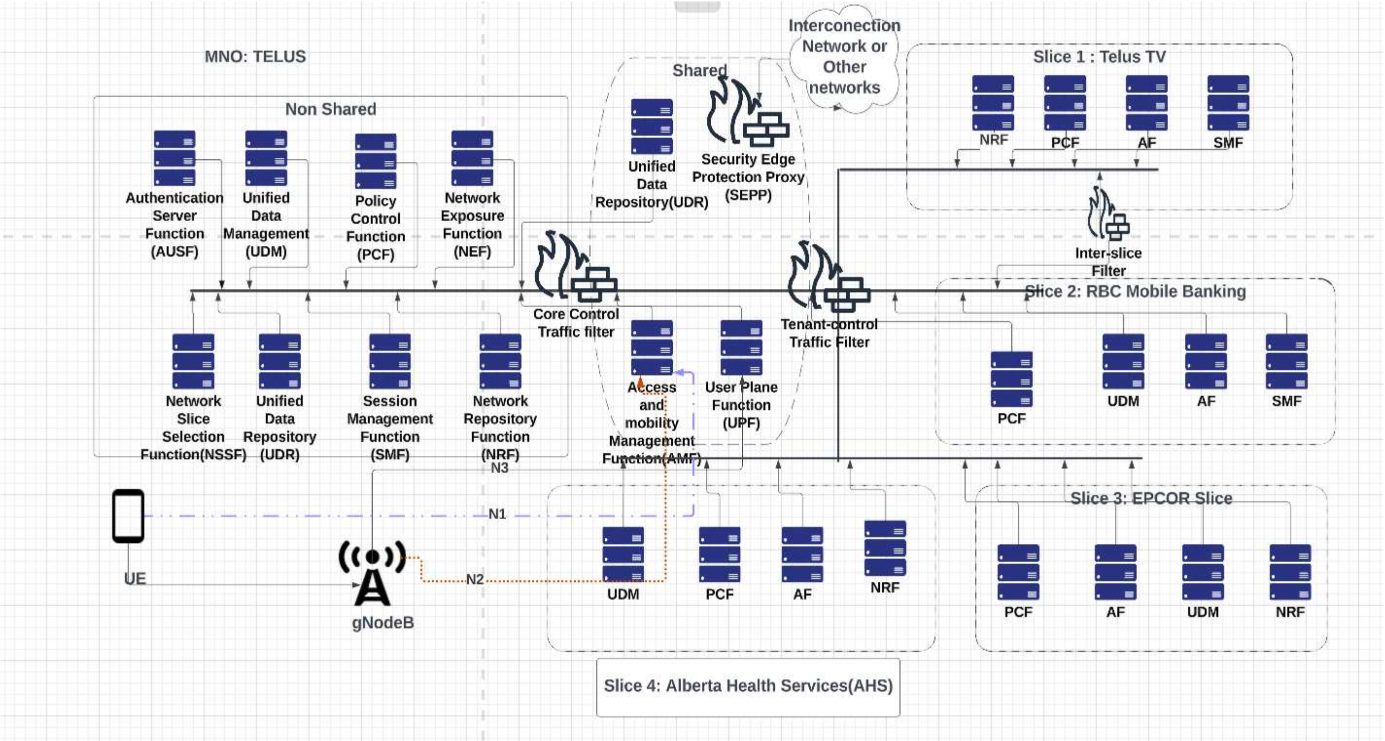

MNO: TELUS
Non Shared
Authentication Server Function (AUSF)
Unified Data Management (UDM)
Policy Control Function (PCF)
Network Exposure Function (NEF)
Shared
Unified Data Repository(UDR)
Security Edge Protection Proxy (SEPP)
Interconection Network or Other networks
Slice 1 : Telus TV
NRF
PCF
AF
SMF
Inter-slice Filter
Core Control Traffic filter
Tenant-control Traffic Filter
Slice 2: RBC Mobile Banking
PCF
UDM
AF
SMF
Network Slice Selection Function(NSSF)
Unified Data Repository (UDR)
Session Management Function (SMF)
Network Repository Function (NRF)
Access and mobility Management Function(AMF)
User Plane Function (UPF)
N3
N1
N2
UE
gNodeB
UDM
PCF
AF
NRF
Slice 3: EPCOR Slice
PCF
AF
UDM
NRF
Slice 4: Alberta Health Services(AHS)

Figure 3-1. Security Zoning [17]

Security zoning provides a more holistic approach to protecting various layers of the network core from malicious attacks or attacks that could extend to other zones on the network.

Also, the enhanced filtering and validation approach combines information from different layers and protocols and integrates external threat information. This kind of filtering and validation approach allows the division of the network into security zones, thus safeguarding the 5G core network. Furthermore, the cross-correlation of attack information between that security network function maximizes 5G network protection against sophisticated attackers and allows better mitigations and faster detection while minimizing false alarms. [49]

3.7 Zero trust security Implementation

The zero-trust approach is one of the best approaches to network security which follows the ideology of “Never trust but always verify.” This approach will help CSPs, MNOs, and other organizations interested in

implementing network slicing succeed in reducing the threat surface and impact of attacks and help prevent specific vulnerabilities that persist. With the Zero Trust architecture, all logged-in users are constantly verified, and all requests and interactions with the network are validated before access is granted. If adequately implemented, every stage of digital interaction can be secured through strong authentication and authorization methods. It can help reduce the possibility of attacks like MitM and configuration attacks in network slicing. [41] [53]
According to Sentinel One [53], "Historically, most corporate applications and solutions that store corporate data were protected behind the corporate network. However, this has changed dramatically with the adoption of cloud applications and the mobile workforce. Today, many applications or storage solutions that were unthinkable to be accessible outside the corporate network are hosted on cloud-native solutions and are accessible from virtually anywhere. For this reason, the old perimeter that security professionals would set and protect no longer exists, and perimeter-based security models are obsolete."

This has motivated the development of a more modern security model like the Zero Trust security to reduce and cut down the threat surface of 5G networks, particularly core network slicing. If this model is implemented well, most CSPs and MNOs looking to take advantage of network slicing can create a flexible and robust security model for their networks and various network slice implementations, restoring the confidence of consumers and enterprises.

Legacy and traditional networks were mainly designed to stop vulnerabilities and threats outside the network. Thus trust was given implicitly to threats that managed to get through the incoming firewalls. Unfortunately, this led to many vulnerabilities and exposed many networks to threats. 5G network slicing cannot implement this same model since the impact will be devastating considering the amount of information and data that an intruder will have access to once they can get through a CSP's or MNO's network.
Zero trust is modelled never trust but always verify incoming requests to connect to a network and whatever device is already inside the network. All identities and endpoints are always verified.

With the Zero Trust model, if some suspicious behaviour or activity is identified, various actions can be triggered, such as segregating and isolating the interactions, terminating access credentials, or initiating multi-factor authentication (MFA). [41]
Some Zero Trust techniques include MFA, Access control and Encryption.

Sentinel One [53] beautifully describes the various guiding principles of the zero trust model and compares the architecture of legacy systems to the zero trust architecture in terms of how each handles threats or potential threats through simulations or scenarios as shown in Figures …..

3.7.1 Principles of Zero trust

Never trust, always verify- Treat every user, endpoint, application or workload, and data flow as untrusted. Assuming there is a security breach with the opinion that a threat is already in the network environment and ensuring all resources and network elements are protected. Deny by default and scrutinize all users, endpoints, data flows, and access requests.

Verify explicitly by taking control and dictating access to all resources consistently and securely using multiple trust signals for contextual access decisions.

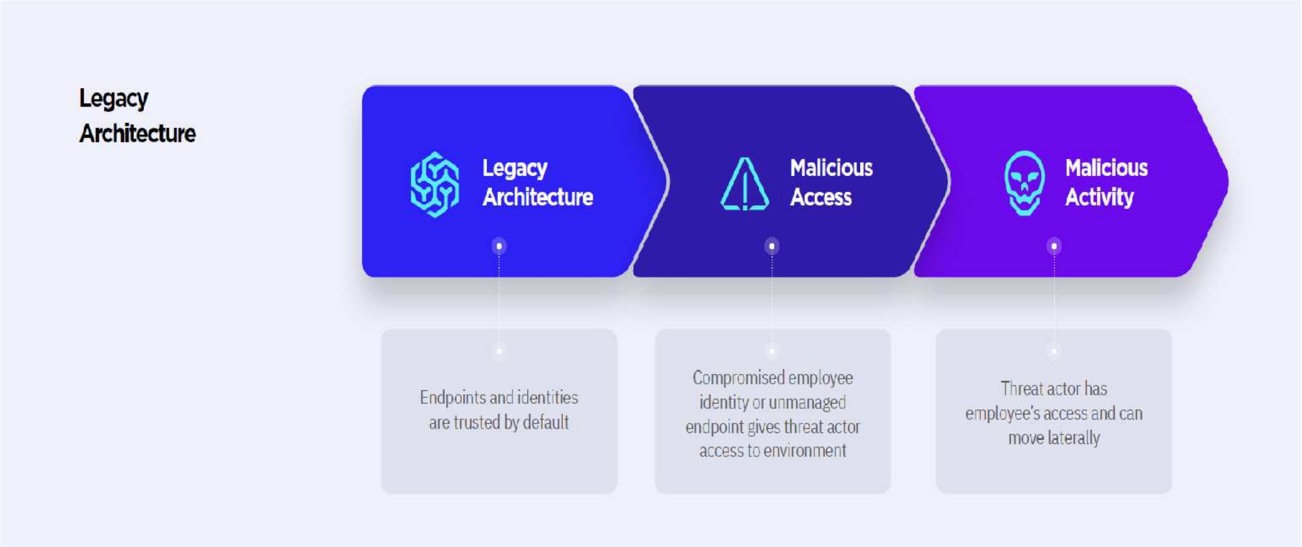

Legacy Architecture
Legacy Architecture
Malicious Access
Malicious Activity
Endpoints and identities are trusted by default
Compromised employee identity or unmanaged endpoint gives threat actor access to environment
Threat actor has employee's access and can move laterally

Figure 3-2. Attack scenario on legacy security architectures [53]

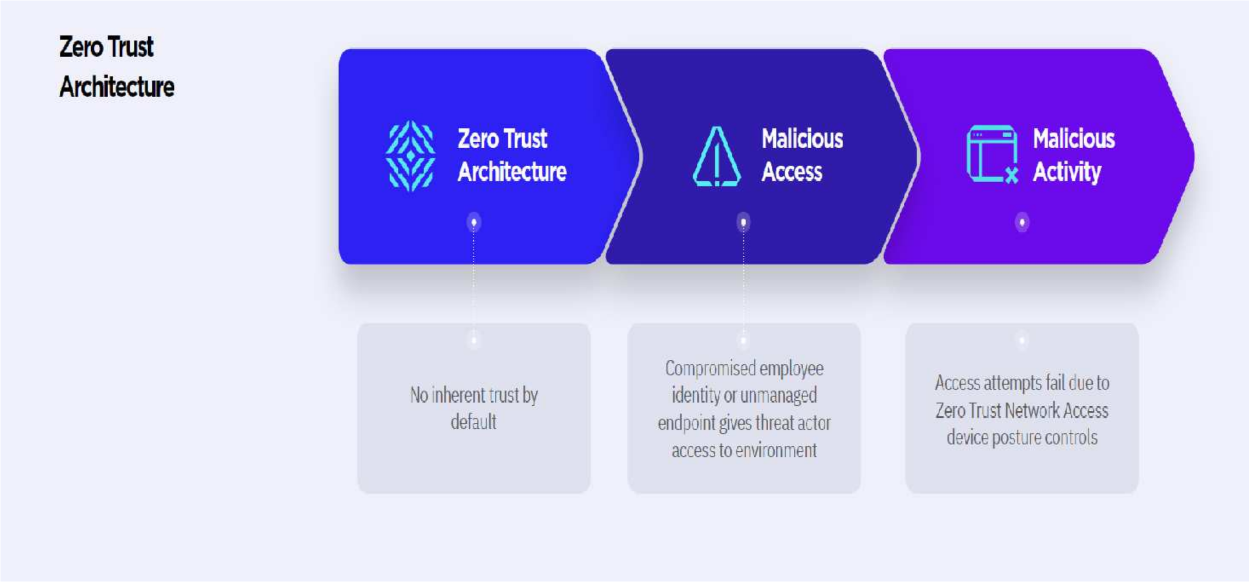

Zero Trust Architecture
Zero Trust Architecture
Malicious Access
Malicious Activity
No inherent trust by default
Compromised employee identity or unmanaged endpoint gives threat actor access to environment
Access attempts fail due to Zero Trust Network Access device posture controls

Figure 3-3. Attack Scenario on Zero Trust Architecture [53]

John Kindervag, former VP & Principal Analyst at Forrester and creator of the Zero Trust methodology, suggested a five-step deployment guide for Zero Trust his methods can be related to and used as a framework to implement a more secure 5G core network slicing design for various MNOs, CSPs and enterprises

1. **Defining Your Protect Surface**
   Attackers always try to find an open attack surface they can easily exploit. But, in reality, you can’t secure the 5G core network even if organizations try their best to secure every attack surface. So, it is necessary that organizations, such as CSPs and MNOs, as well as consumers of a network slice or network service, define the protected surface, including critical data, applications, assets, and network slices and data services.
2. **Mapping the Transaction Flows**

As they transition from network perimeter-based security to modern architectures, most organizations are aware of their network and know how to protect it. What changes is the fact that organizations implementing network slicing need analytical insights into data and services within the network? How are critical data from various network slices accessed? How can anomalies such as that of network slices be detected?

3. **Architecting the Environment**

There is no such thing as an architectural blueprint that fits all organizations worldwide. This statement remains true as organizations embrace a Zero Trust architecture. ZTN designs are unique per organization. Your protected surface defines them. Ideally, you want to bring security controls as close as possible to your protected surface by defining micro-perimeters and ensuring that access requests are always verified based on the health state of the entity requesting the access.

4. **Create the Zero Trust Policy**

Determine the Zero Trust policies by answering who, what, when, where, why, and how you should get access to 5G core network slice resources and services. This framework will go a long way in helping various MNOs and CSPs design a framework for 5G core network slicing zero trust security model to ensure continuous protection of network slices

5. **Monitor and Maintain the Environment**

The final step is gathering telemetry, leveraging autonomous solutions to perform analytics, detecting anomalies, and automatically responding based on the defined zero trust policies. Continuous monitoring of all network slices and testing for vulnerabilities is essential

3.8 Secure Network slice management, orchestration and network monitoring techniques.

Network Slice Managers and host platforms should support mutual authentication. Network Slice Managers should authenticate the hosts before activating a slice instance. Host platforms should also authenticate the Slice Manager before loading a slice instance and running on the physical hardware. In addition, mutual authentication between slice managers should be a requirement where multiple slice managers are involved in instantiating an end-to-end network slice. Alternatively, Network Slice Managers should authenticate the host platforms in the networks they control on which they want to load network slice instances. Hosts platforms should also authenticate the Network Slice Manager before allowing the instance of the network slice to be loaded. [32] [52]
In addition, security for the network slice life cycle phases must be considered.

According to [46], security must be enforced in all four phases because a vulnerability in one step can introduce vulnerabilities in other steps. Some of the recommendations made include;

- All network slices and templates should be well-logged with unique tracking mechanisms, documented and audited with security monitoring in real-time so that suspicious slices can easily be identified.
- In the preparation phase of a slice, security-by-design principles must be incorporated
- Proper authentication of all network slice templates and their respective sources with particular confidentiality and integrity verification mechanisms during transmission and storage. Isolation should be secured at slice creation, monitored, and updated during the run-time.
- The legal agreement between entities using and managing slices for how APIs should be secured and handled based on access and operation rights. The consequences of exposure to traffic and other confidential data should be spelled out.
- Destruction of highly sensitive data and reallocating resources or temporary assigning of available resources and channels to prevent access to them by attackers. Information that has to be stored should be well encrypted using unique cryptographic mechanisms.

3.9 Future Research recommendations and solutions to ensure 5G Network Slicing Security

According to the report by [30], some recommendations were made on how network slicing can be protected. Most of the work on protection and security mechanisms for the 5G network slicing architecture considers authentication protocols from the user end or for inter-slice communication. Still, more needs to be discussed on customizing security architecture to meet current security concerns and for the network slicing security architecture to withstand changing plans and points of attack on the 5G core network.
Some solutions were suggested, and they are explained as follows;

3.9.1 Implementation of AI technologies

Attacks on the 5G core network and network slices will evolve. We will be witnessing more automated, coordinated and well-planned attacks on the security of mobile networks, which may be difficult to manage or prevent. If this fight is to be won, the 5G core network architecture, including the network slicing architecture, must implement more modern techniques to counter the various threats to which network slicing is bound to be exposed. In addition, sophisticated and intelligent security solutions are required to counter these attacks. AI can be used to design smart security solutions for network slices, the 5G core network, and the 5G network. AI algorithms and models (e.g. Markov models, neural networks, genetic algorithms, and machine learning techniques) can be used to find configuration errors, security vulnerabilities, and threats to reduce human intervention and to be able to match these advanced attacks.

3.9.2 Micro-Segmentation

With SDN, network slice security can be more robust by micro-segmentation to isolate traffic flow related to different applications, users and network slices. This was initially used as a protection scheme for data centers which used to have just one firewall where once there is a breach, a malicious entity has access to everything. With micro-segmentation introduced, single points of failure were eliminated. Using this mechanism with network slicing will mean more specific points of isolation, access control and security policies will be defined or designed.

3.9.3 Single or separate network slices for security

A single or separate network slice or network slices can be implemented as security slices to minimize overhead to other security mechanisms.
If this is achieved, security-related communication such as authentication messages, firewall updates, and security policy updates can be transported or transmitted over this slice or set of slices. Also, network monitoring and security incident handling systems cryptographic service, authentication and access control, security auditing and security service life-cycle management can be run on top of this security slice to ensure the proper operation and security of the 5G core network. With a separate security slice, there can be the flexible deployment of security resources, which can be scaled up or down to counter threats more agilely.

3.9.4 Automation and orchestration of Security

Due to how dynamic and rapidly changing the mobile networks are now and since it's not bound to slow down with 5G core network slicing is the future of communication and services, it is essential to reduce as much human management of security as much as possible due since it is going to be more cumbersome and less feasible as the world progresses to keep track of all security requirements. With an automated security

orchestrator, deployment, configuration, maintenance, monitoring, life-cycle management, and all other security functions in a ‘softwarized’ mobile network can be handled relatively easily.
Though the European Telecommunications Standards Institute (ETSI)-Industry Specification Groups (ISGs) group has already defined the security orchestrator for NFV systems, it will also trickle down to network-slicing security mechanisms.

3.9.5 The Security by Design (SbD) principle

With this method for designing more secure systems and networks, security flaws and all possible vulnerabilities are considered in the design of a particular software system. The current SbD approaches can offer benefits such as establishing reliable operation of controls and enabling continuous and real-time auditing. The SbD approach can be used on two occasions in Network slice systems. First, it can be used to design slicing systems (including relevant VNFs) with a secure foundation. Also, the SbD approach can be used during the slice creation process for different network services. With this method, the entire network slicing architecture can be redesigned to counter all possible or known attacks or security concerns related to network slicing.

3.9.6 The Security-as-a-service option

In providing network slices to various market verticals, e.g. Automotive, agriculture, health etc., the service provider cannot guarantee total security protection for the slices of these customers. Also, these market verticals may need more expertise to manage their slices and protect their information securely. Thus with security-as-a-service (SaaS), security service providers or providers with security expertise can offer security services for customers, enterprises, and various market verticals. These security services could include authentication, security monitoring, intrusion detection, penetration testing, and security event management. The SaaS concept can offer a much easier integration route for the network slices of these market verticals.

4 Conclusion

In this report, we researched various security flaws associated with network slicing. During this research, what was discovered is that there is still more to be studied, tested and analyzed if we are to fully implement network slicing on a large scale involving big industry players and consumers. To operate network slices with robust security and maximum information protection, all stakeholders in the 5G market have to come together and build a system based on the best security standards.
Furthermore, isolation of network slices is of paramount interest if we want to design a 5G core network that is secure and resistant to DoS/DDoS attacks which is the most prevalent attack 5G core network slicing is vulnerable to. In cases where slices have inter-slice communication, specific security protocols must be designed and adhered to before this communication can be authorized.
Also, NFV and SDN technologies have to be made more secure though they have made network slicing possible while increasing the threat surface of the 5G core but of the entire 5G system architecture.
All that we have researched and discussed regarding network slicing proves that there are still too many security shortcomings, some of which were carried on from legacy architectures and still have to be resolved. Therefore, we still need a deep dive into 5G core network slicing security to effectively beef up the security features of 5G core network slicing while looking at what could be added as a feature based on what threats currently exist and what possible vulnerabilities and flaws each implementation could have. This will require all stakeholders to come together regularly to define robust security measures with continuous monitoring, updates and adjustments before network slicing can be trusted as a secure crucial enabling technology.

5 Glossary

1G	First Generation
2G	Second Generation
3G	Third Generation
3GPP	Third Generation Partnership Project
4G	Fourth Generation
5G	Fifth Generation
5GC	Fifth Generation Core Network
5GGUTI	5G-Global Unique Temporary Identifier
AAA	Authentication, Authorization, and Accounting
AF	Application Function
AI	Artificial Intelligence
AMF	Access & Mobility Management Function
AMPS	Advanced Mobile Phone System
AN	Access Network
API	Application Programming Interface
APN	Access Point Name
AR	Augmented Reality
ARPF	Authentication Credential Repository and Processing Function
AuC	Authentication Centre
AUSF	Authentication Server Function
BSC	Base Station Controller
BSS	Business Support Systems
CDMA	Code Division Multiple Access
CHF	Charging Function
CN	Core Network
CS	Circuit Switched
CSMF	Communication Service Management Function
CSP	Communication Service Provider
CU	Centralized Unit
CUPS	Control and User Plane Separation
DDoS	Distributed Denial of Service
DHCP	Dynamic Host Configuration Protocol
DN	Data Network
DNN	Data Network Name
DoS	Denial of Service
DTU	Data Transfer Unit
DU	Distributed Unit
EAP-AKA	Extensible Authentication Protocol- Authentication and Key Agreement
EDGE	Enhanced Data for Global Evolution
EIR	Equipment Identity Register
eMBB	Enhanced Mobile Broadband
EMS	Element Management System
EPC	Evolved Packet Core
ETSI	European telecommunications Standards Institute
E-UTRAN	Evolved-UMTS Terrestrial Radio Access Network
FCC	Federal Communications Commission
FCC	Federal Communications Commission
FDD	Frequency Division Duplex
FDD	Frequency Division Duplex

FDD	Frequency Division Duplex
FDMA	Frequency Division Multiple Access
g-NodeB	Next-generation NodeB
GPRS	General Packet Radio Service
GSM	Global System for Mobiles
GSM	Global System for Mobile Communication
GSMA	Global System for Mobile Communications Alliance
GTP	GPRS Tunneling Protocol
HLR	Home Location Register
HSDPA	High-Speed Downlink Packet Access
HSPA	High-Speed Packet Access
HSS	Home Subscriber Service
HTTP	Hypertext Transfer Protocol
IaaS	Infrastructure-as-a-Service
IEEE	Institute of Electrical and Electronics Engineers
IMEI	International Mobile Equipment Identity
IMS	IP Multimedia Subsystem
IMSI	International Mobile Subscriber Identity
IMT-2000	International Mobile Telecom System-2000
IMTS	Improved Mobile Telephone System
IoT	Internet of Things
ITU-R	International Telecommunication Union Recommendations
JTACS	Japanese Total Access Communication System
LPWAN	Low Power Wide Area Network
LTE	Long Term Evolution
MANO	Management and Orchestration
MBMS	Multimedia Broadcast Multicast Services
MEC	Multi-access Edge Computing
MIMO	Multiple-Input Multiple-Output
MitM	Man in the Middle
MM	Mobility Management
MME	Mobility Management Entity
mMTC	Massive Machine-Type Communications
MNO	Mobile Network Operator
MSC	Mobile Switching Center
N3IWF	Non-3GPP Interworking Function
NAS	Non-Access Stratum
NEF	Network Exposure Function
NFV	Network Function Virtualization
NGMN	Next generation Mobile Networks
NR	New Radio
NRF	Network Repository Function
NSA	Non-Standalone
NSI	Network Slice Instance
NSMF	Network Slice Management Functions
NSSAAF	Network Slice-Specific Authentication & Authorization Function
NSSAI	Network Slice Selection Assistance Information
NSSF	Network Slice Selection Function
NSSI	Network Slice Subnet Instance
NSSMF	Network Slice Subnet Management Function
OCI	Overload Control Indicator

OFDMA	Orthogonal Frequency Division Multiple Access
OSS	Operational Support Systems
PaaS	Platform-as-a-Service
PCF	Policy Control Function
PCFP	Packet Forwarding Control Protocol
PCRF	Policy and Charging Rule Function
PDU	Protocol Data Unit
PGW	Packet Network Data Gateway
PLMN	Public Land Mobile Network
PSTN	Public Switched Telephone Network
QAM	Quadrature Amplitude Modulation
QoS	Quality of Service
RAT	Radio Access Technology
REST	Representational State Transfer
RNC	Radio Network Controller
SA	Standalone
SaaS	Software-as-a-Service
SAE	System Architecture Evolution
SBA	Service-Based Architecture
SbD	Service by Design
SCP	Service communication Proxy
SD	Slice Differentiator
SDN	Software Defined Networking
SEAF	Security Anchor Function
SEPP	Security edge Protection Proxy
SGW	Serving Gateway
SIDF	Subscriber Identity De-concealing Function
SIP	Session Initiation Protocol
SLA	Service Level Agreement
SMF	Session Management Function
SMS	Short Messaging Service
S-NSSAI	Single Network Slice Selection Function
SSH	Secure Socket Shell
SST	Slice/Service Type
SUCI	Subscriber Concealed Identifier
TACS	Total Access Communications System
TCP	Transmission Control Protocol
TCP/IP	Transmission Control Protocol/Internet Protocol
TD-SCDMA	Time Division-Synchronous Code Division Multiple Access
TLS	Transport Layer Security
UDM	Unified Data Management
UDR	User Data Repository
UDR	Unified Data Repository
UE	User Equipment
UMTS	Universal Mobile Telecommunications System
UPF	User Plane Function
URLLC	Ultra-Reliable Low Latency Communication
V2X	Vehicle to Everything
VLAN	Virtual Local Area Network
VM	Virtual Machine
VNF	Virtual Network Function
VNF	Virtual Network Function

VR	Virtual Reality
WCDMA	Wideband Code Division Multiple Access
WiMAX	Worldwide Interoperability for Microwave Access

6 References

[1] É. A. d. S. José Luiz Frauendorf, The Architectural and Technological Revolution of 5G, https://doi.org/10.1007/978-3-031-10650-7_7, Springer Nature Switzerland AG 2023.

[2] Z. Li, X. Z. Wang and Tongxu, 5G+ How 5G Change the Society, Springer Nature Singapore Pte Ltd. 2021, 2021.

[3] J. T. J. Penttinen, 5G Explained: Security and Deployment of Advanced Mobile Communications, John Wiley & Sons Ltd, 2019.

[4] N. Boudriga, Security of mobile communications, Auerbach Publications, 2010.

[5] ". Barakovic, E. Kurtovic, O. Bozanovic, A. Mirojevic, S. Ljevakovic, A. Jokic, M. Peranovic, J. B. Husic", E. Kurtovic, O. Bozanovic, A. Mirojevic, S. Ljevakovic, A. Jokic, M. Peranovic and Jasmina, "Security issues in wireless networks: An overview," vol. BIHTEL.2016.7775732, no. 7775699, 2016.

[6] "EVOLUTION OF COMMUNICATION - FROM 1G TO 4G & 5G," 23 July 2016. [Online]. Available: https://youtu.be/2nsEAw_SirQ.

[7] D. W. Stallings, 5G Wireless: A Comprehensive Introduction, Copyright © 2021 Pearson Education, Inc., 2021.

[8] J. Rodriguez, Fundamentals of 5G Mobile Networks, John Wiley & Sons, Ltd., 2015.

[9] I. A. A. Madhusanka Liyanage, A Comprehensive Guide to 5G Security, © 2018 John Wiley & Sons Ltd, 2018.

[10] S. S. T. K. J. O. A. G. a. M. Y. S. M. I. Ijaz Ahmad, "Security for 5G and Beyond," vol. VOL. 21, no. No. 4, 2019.

[11] R. S. Shetty, 5G Mobile Core Network- Design, Deployment, Automation, https://doi.org/10.1007/978-1-4842-6473-7, Apress, 2021.

[12] Rajiv, "Evolution of wireless technologies 1G to 5G in mobile communication," 01 August 2022. [Online]. Available: https://www.rfpage.com/evolution-of-wireless-technologies-1g-to-5g-in-mobile-communication/. [Accessed 11 October 2022].

[13] M. G. a. P. Nair, "Cisco White paper (Public) 5G Security Innovation," © 2018 Cisco, 2018.

[14] Digital International, "Planning your journey to 5G," www.digi.com, Hopkins, 2021.

[15] Ö. B. O. Q. M. B. Patrick Marsch, 5G System Design: Architectural and Functional Considerations and Long Term Research, John Wiley & Sons Ltd, 2018.

[16] C. Cox, An introduction to 5G : the new radio, 5G network and beyond, © 2021 John Wiley & Sons Ltd, 2021.

[17] AdaptiveMobile, "A Slice in Time: Slicing Security in 5G Core Networks White Paper," AdaptiveMobile Security www.adaptivemobile.com, 2021.

[18] 3GPP, "http://www.3gpp.org/release-15," 3GPP, 09 2019. [Online]. Available: http://www.3gpp.org/release-15. [Accessed 30 0 2022].

[19] D. K. Y. P. H. C. D. K. a. S. K. Park Seongmin, "5G Security Threat Assessment in Real Networks," MDPI , 17 August 2021. [Online]. Available: https://doi.org/10.3390/s21165524. [Accessed 28 October 2022].

[20] ETSI, "https://www.etsi.org/technologies/nfv," ETSI, [Online]. Available: https://www.etsi.org/technologies/nfv. [Accessed 03 10 2022].

[21] Andrea Leonhardt, "Defining The Elements of NFV Architectures," 17 October 2019. [Online]. Available: https://blog.equinix.com/blog/2019/10/17/networking-for-nerds-defining-the-elements-of-nfv-architectures/.

[22] G. Pujolle, Software Networks: Virtualization, SDN, 5G and Security, vol. 1, ISTE Ltd and John Wiley & Sons, 2020.

[23] W. Stallings, Foundations of Modern Networking: SDN, NFV, QoE, IoT, and Cloud, Pearson Education, Inc, 2016.

[24] "What is cloud computing?," Amazon Web Services, 2022. [Online]. Available: https://aws.amazon.com/what-is-cloud-computing/. [Accessed 08 November 2022].

[25] L. U. K. N. H. T. C. S. H. S. M. Ahsan Kazmi, Network Slicing for 5G Networks and Beyond, Springer Nature Switzerland AG 2019, 2019.

[26] C. M. D. S. M. C. S. V. S. N. S. O. H. S. T. B. H. D. B. D. A. a. H. B. P. Rost1, "Network Slicing to Enable Scalability and Flexibility in 5G Mobile Net-works," no. 10.1109/MCOM.2017.1600920, 2017.

[27] John Burke, Nemertes Research, "https://www.techtarget.com/whatis/definition/network-slicing," May 2022. [Online]. Available: www.techtarget.com.

[28] CISA, "POTENTIAL THREAT VECTORS to 5G INFRASTRUCTURE," CISA, 2021.

[29] S. S. e. al, "Key Enabling Technologies of 5G Wireless Mobile," no. ICCIEA 2020 Journal of Physics: Conference Series, 202

[30] C.-L. W. P. C. M. L. Rahim Tafazolli, The Wiley 5G Ref Security, 2021 John Wiley & Sons, Ltd, 2021.

[31] Akash Tripathi, Alan Weissberger, "5G Security explained: 3GPP 5G core network SBA and Security Mechanisms," IEEE Communication Society, 01 January 2022. [Online]. Available: https://techblog.comsoc.org/2022/01/01/5g-network-security-threats-and-3gpp-security-mechanisms/. [Accessed 01 November 2022].

[32] 5. Americas, "The Evolution of Security in 5G: A Slice of Mobile Threats," 2019.

[33] Positive Technologies, "5G SECURITY ISSUES," Positive Technologies-positive-tech.com, 2019.

[34] Techplayon, "5G Reference Network Architecture," 3 May 2017. [Online]. Available: https://www.techplayon.com/5g-reference-network-architecture/.

[35] Technologies, Positive, "5G Standalone core security research," 2020.

[36] Ericsson, "The essential building blocks of E2E network slicing," Ericsson, 2021.

[37] Paresh Khatri, "Realizing transport-layer slicing using segment routing," 11 August 2022. [Online]. Available: https://blog.apnic.net/2022/08/11/realizing-transport-layer-slicing-using-segment-routing/.

[38] A. S. Inc, "Network Slicing | Webinar," 2020.

[39] AdaptiveMobile, "https://info.adaptivemobile.com/network-slicing-security?hsLang=en#download," 2021 AdaptiveMobile, 2021. [Online]. Available: https://info.adaptivemobile.com/network-slicing-security?hsLang=en#download.

[40] Mpirical, "5G Network Slicing Defined | Mpirical," 20 June 2022. [Online]. Available: https://www.youtube.com/watch?v=SEL-9-P9J8A.

[41] C. I. S. C. D. NSA, "ESF Potential Threats to 5G Network Slicing," 2022.

[42] ENISA, "https://www.enisa.europa.eu/publications/enisa-threat-landscape-report-for-5g-networks," ENISA, 14 December 202 [Online]. Available: https://www.enisa.europa.eu/publications/enisa-threat-landscape-report-for-5g-networks. [Accessed October 2022].

[43] S. ROMMER, P. HEDMAN, M. OLSSON, L. FRID, S. SULTANA and C. MULLIGAN, 5G CORE NETWORKS: POWERING DIGITALIZATION, Academic Press, 2020.

[44] M. S. H. A. C. D. B. L. G. Xin Li, "Network Slicing for 5G: Challenges and Opportunities," 2022.

[45] 5G Americas, "5G-Americas White paper: Security considerations for the 5G era," 2020.

[46] R. F. O. a. G. NENCIONI2, "5G Network Slicing: A Security Overview," *IEEE Access,* 2020.

[47] E. U. A. f. C. Security, "ENISA Threat Landscape for 5G Networks Report," ENISA https://www.enisa.europa.eu/publications/enisa-threat-landscape-report-for-5g-networks, 2020.

[48] vmware, "What is network functions virtualization?," 2023. [Online]. Available: https://www.vmware.com/topics/glossary/content/network-functions-virtualization-nfv.html#:~:text=An%20NFV%20architect
re%20consists%20of,compute%2C%20storage%20and%20network%20resources..

[49] Máirín OSullivan, "DoS Attack Vulnerability - Denial of Service Attacks on 5G Networks," 19 May 2021. [Online]. Available: https://blog.adaptivemobile.com/5g-network-slicing-vulnerability-denial-of-service-attacks.

[50] H. Kim, "5G core network security issues and attack classification from a protocol perspective," 2020.

[51] R. Pell, S. Moschoyiannis, E. Panaousis and R. Heartfield, "TOWARDS DYNAMIC THREAT MODELLING IN 5G CORE BASED ON MITRE ATTACK," vol. arXiv:2108.11206v3 [cs.CR] 1 Sep 2021, 2021.

[52] NGMN Alliance, "5G security recommendations Package #2: Network Slicing," *NGMN 5G Security - Network Slicing,* pp. 2-12, 27 April 2016.

[53] S. One, "Moving_to_an_Endpoint-Centric_Zero_Trust Security_Model_with_SentinelOne_10142021," sentinelone.com, 2021

[54] 1. M. S. H. A. C. F. I. D. B. L. G. S. Xin Li, "Network Slicing for 5G: Challenges and Opportunities," *IEEE Internet Computin*
Vol. 21, Issue 5, August 2017, , pp. pp. 20-27, 2017.

[55] P. D.-I. Ulrich, 5G: An Introduction to the 5th Generation Mobile Networks, © 2021 Walter de Gruyter GmbH, Berlin/München/Boston, 2021.

[56] H. Technologies, "https://carrier.huawei.com/~/media/CNBG/Downloads/Program/5g_nework_architecture_whitepaper_en.pdf," Huawei Technologies Company Ltd, [Online]. Available: https://carrier.huawei.com/~/media/CNBG/Downloads/Program/5g_nework_architecture_whitepaper_en.pdf. [Accessed 01 10
2022].

[57] D. Sattar and A. Matrawy, "Towards Secure Slicing: Using Slice Isolation to mitigate DDos attacks on 5G Core Network Slices," *2019 IEEE Conference on Communications and Network Security (CNS),* 2019.

[58] S. M. A. K. •. L. U. Khan, Network Slicing for 5G and Beyond Networks, Springer, 2019.

[59] . J. Ji, "Security Risks of 5G Core Network Introduced by New Technology," NSFocus, 08 March 2022. [Online]. Available: https://nsfocusglobal.com/security-risks-of-5g-core-network-introduced-by-new-technology/#:~:text=The%20security%20risks
%20faced%20by,attack%20the%20SDN%20data%20surface.. [Accessed 27 October 2022].

[60] M. Hemmings, 5G Networks: Background, issues and security, Nova Science Publishers, 2021.

[61] R. T. D. S. M. a. B. G.-N. C. Dimitrios Schinianakis, "Security Considerations in 5G Networks: A Slice-Aware Trust Zone," *IEEE,* 2019.

[62] "Google delivers 5G network slicing capabilities for enterprises," 02 November 2021. [Online]. Available: https://cloud.google.com/blog/topics/telecommunications/5g-network-slicing-with-google-android-enterprise-and-cloud.

[63] P. Technologies, "5G Standalone core security research_A4.ENG.0003.03," no. A4.ENG.0003.03, 2020.