Research on security threats posed by legacy RATs (Radio access technologies) in 5G networks.

MINT-709
Capstone Project Report

Presented by

Liton Kumar Das

University of Alberta
Master of Science in Internetworking
Department of Electrical and Computer Engineering
Edmonton, Canada

Supervisor
Sandeep Kaur

ABSTRACT

Due to its impact expected on the economy and society, the fifth generation of mobile telecommunications (5G) is one of the most important innovations of our time. From a conceptual perspective, 5G technology promises to deliver low latency, high speed, and more reliable connections to new generations of autonomous systems and edge-type devices, covering massive and critical machine-type communications. 5G has initiated its full development to satisfy an increasing demand for mobile data traffic and big data bandwidth. Centralized data processing, collaborative radio, real-time cloud infrastructure, and cloud radio access network (C-RAN), along with their excellent advantages, are being sought by more and more operators to meet end-user requirements.

This research paper aims to study the concept of security threats on RAT technology in 5G networks. Its historical background and architecture through the history have evolved. The evolution of networks from 1G to 5G is also discussed as an analysis zed how faster data is required for faster networks.

Proposed 5G architectures are designed to close security gaps from previous iterations of cellular networks, but the pervasive nature of 5G introduces new security challenges outside the traditional space. In addition, 5G's attractive, transformative services will likely introduce threat vectors not yet seen or experienced. This paper will examine how 5G differs from other wireless architectures and what possible threats, vulnerabilities, and attacks are. Security considerations will discuss various software, virtualization, automation, orchestration, and Radio Access Network (RAN) considerations. Finally, zero-Trust security and several other techniques will be addressed to mitigate the threats, and various recommendations will be proposed for protection.

As an excellent mobile wireless network architecture, compared with traditional RAN, C-RAN has incomparable advantages in terms of low power consumption, reduced base station (BS) numbers, and economic capital and operating expenditure. It can also improve network capacity and BS utilization rate. However, C-RAN technology in 5G mobile communication networks must deal with several security challenges relating to virtual network functions and software-defined networking. This research project aims to study and present a comprehensive evaluation of existing security studies in the field of C-RAN and corresponding security threats and attacks. Furthermore, we indicate open solutions to research issues and propose future research trends.

ACKNOWLEDGEMENT

I am grateful to express my gratitude to my mentor, **Ms. Sandeep Kaur**, who guided me throughout the project and provided invaluable motivation and suggestions to expand my knowledge base. In addition, she offered insight and freedom to work on my project while ensuring that I stayed on track and did not stray from my project's core.

I also want to convey my earnest appreciation to my professor, **Dr. Mike MacGregor,** for providing me with assistance and such a great opportunity.

Finally, I thank my wife, **Shibani Rani Paul**, and my respected parents for supporting and motivating me and ensuring my focus on the target. I also thank my classmates, professors, and the University of Alberta for assisting and supporting me in achieving this goal whenever feasible.

The contents of this document reflect the research, analysis, and conclusions of my last six months' effort based on thousands of online techno, official basis documents, research papers, surveys, technical specifications, and so many things. I acknowledge that this document and the information contained herein are for informational purposes only due to my degree project requirement. I assume there is no error or false information, as per my best knowledge. This document is subject to revision or removal at any time with my Capstone Project supervisor's recommendation. I am not liable for and, at this moment, disclaim any direct, indirect, punitive, special, incidental, consequential, or exemplary devastation arising out of or in connection with the use of this document and any information contained in this document.

Table of Contents

List of Figures

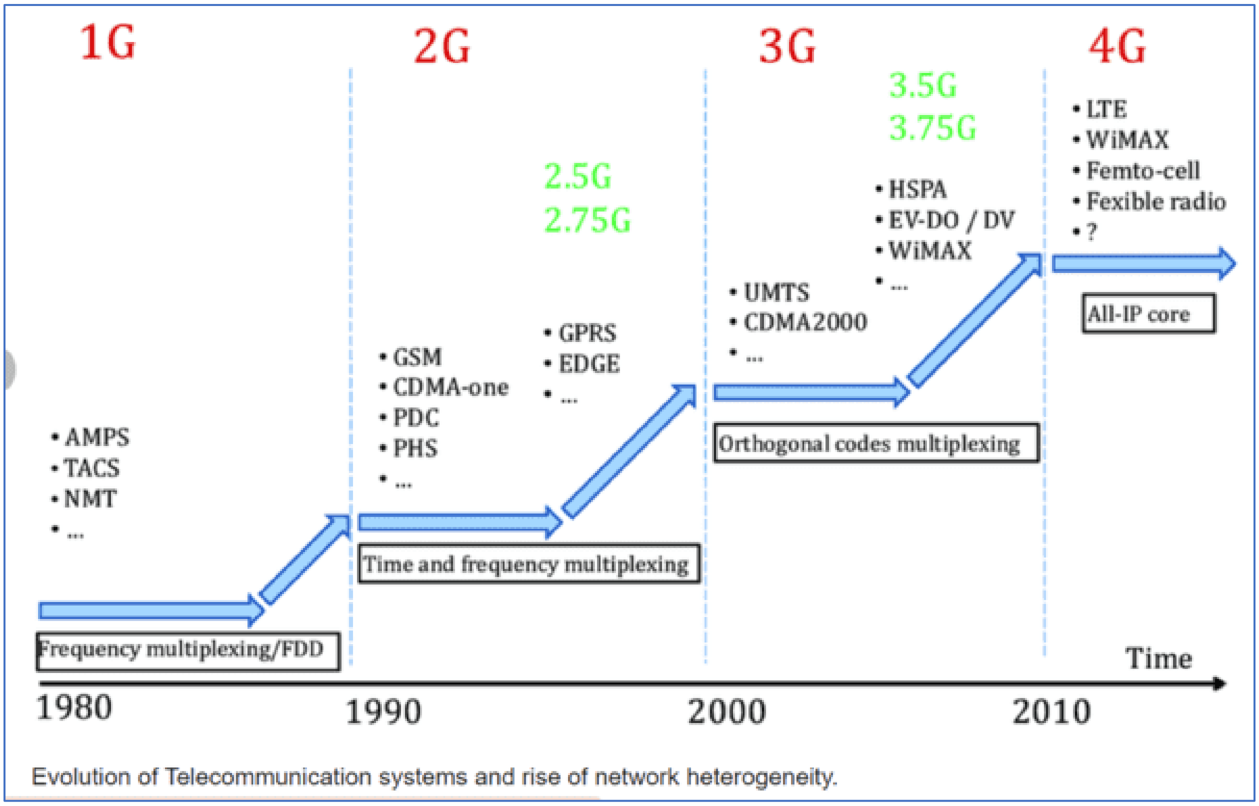
List of Tables

SECTION 1. INTRODUCTION

Wireless communication (or just **wireless**) transfers between two or more points without using an electrical conductor, optical fiber, or other continuous guided medium. A radio wave is the most used wireless technology. With the radio wave technological concept, intended distances might be short, like a few meters for Bluetooth or millions of kilometers for deep-space radio communications. Wireless operations permit mobile and interplanetary communications services; implementing these services with wires is nearly impossible or impractical with cables. The term is mainly used in the telecommunications industry to denote the telecommunications systems (e.g., radio transmitters and receivers, remote controls, etc.) that use some form of energy (e.g., radio waves and acoustic energy) to transfer information without the use of wires. [1]

FIGURE 1: THE EVOLUTION OF MOBILE HANDSET AS PER TIME

During the past three decades, the world has seen significant changes in the telecommunications industry. There have been some remarkable aspects to the rapid growth in wireless communications, as seen by the significant expansion in mobile systems. Recent Internet technology advancements have increased network traffic considerably, resulting in the rapid growth of data rates. This phenomenon has also impacted mobile systems, resulting in the extraordinary growth of the mobile internet. [3]

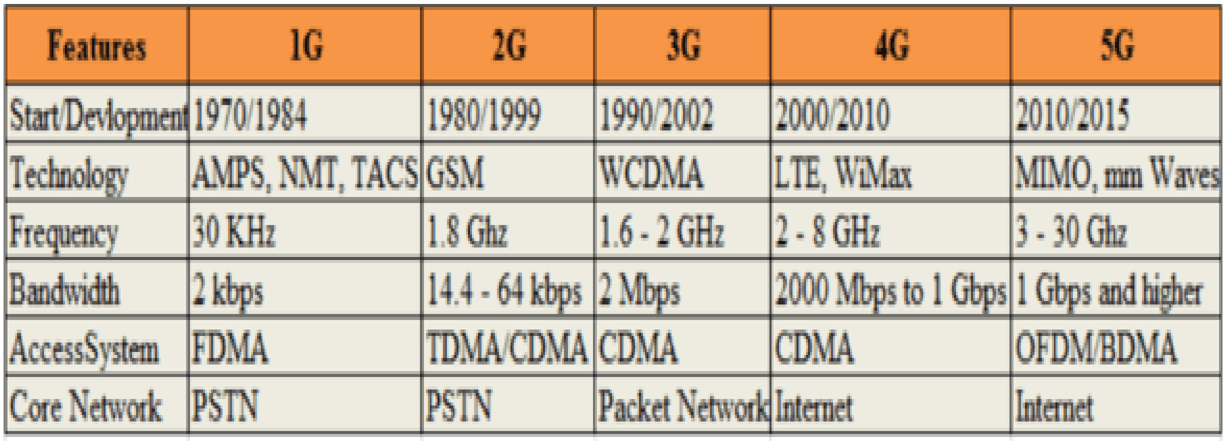


Evolution of Telecommunication systems and rise of network heterogeneity.

FIGURE 2: EVOLUTION OF TELECOMMUNICATION SYSTEMS [4]

1.1 Evolution of Telecom Wireless Technology

Wireless communication aims to provide high-quality, reliable communication like wired communication (optical fiber), and each new generation of services represents a big step in that direction. This evolution journey started in 1979 from 1G, which is still happening in 5G. Generations must meet standards to use the "G" terminology officially.



Features	1G	2G	3G	4G	5G
Start/Devlopment	1970/1984	1980/1999	1990/2002	2000/2010	2010/2015
Technology	AMPS, NMT, TACS	GSM	WCDMA	LTE, WiMax	MIMO, mm Waves
Frequency	30 KHz	1.8 Ghz	1.6 - 2 GHz	2 - 8 GHz	3 - 30 Ghz
Bandwidth	2 kbps	14.4 - 64 kbps	2 Mbps	2000 Mbps to 1 Gbps	1 Gbps and higher
AccessSystem	FDMA	TDMA/CDMA	CDMA	CDMA	OFDM/BDMA
Core Network	PSTN	PSTN	Packet Network	Internet	Internet

TABLE 1: FEATURES OF FREQUENCIES

The "G" stands for "GENERATION." While we connect to the internet, our internet speed depends upon the signal strength denoted in alphabets such as 2G, 3G, 4G, etc., right next to the handset signal bar on our mobile home screen. Each Generation describes a set of telephone network standards and details the technological employment of a particular mobile phone system. The speed increases with time while technology changes. For example, 1G offers 2.4 kbps, 2G offers 64 Kbps based on GSM, 3G offers 144 kbps-2 Mbps, and 4G offers 100 Mbps-1 Gbps based on LTE technology. Until the second Generation (2G) was released, 1G was not called to identify wireless technology. A significant technological jump when wireless networks went from analog to digital.[2]

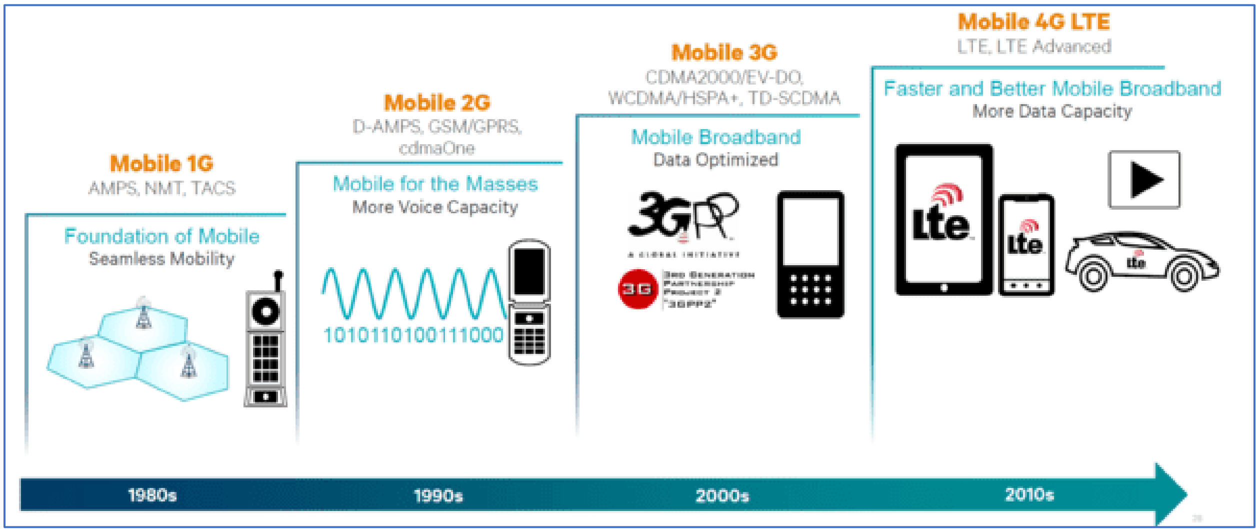


FIGURE 3: TECHNOLOGY UPDATES WITH TIME [5]

GSM:

1.2 1G Mobile Communication System: Analogue Voice Technology

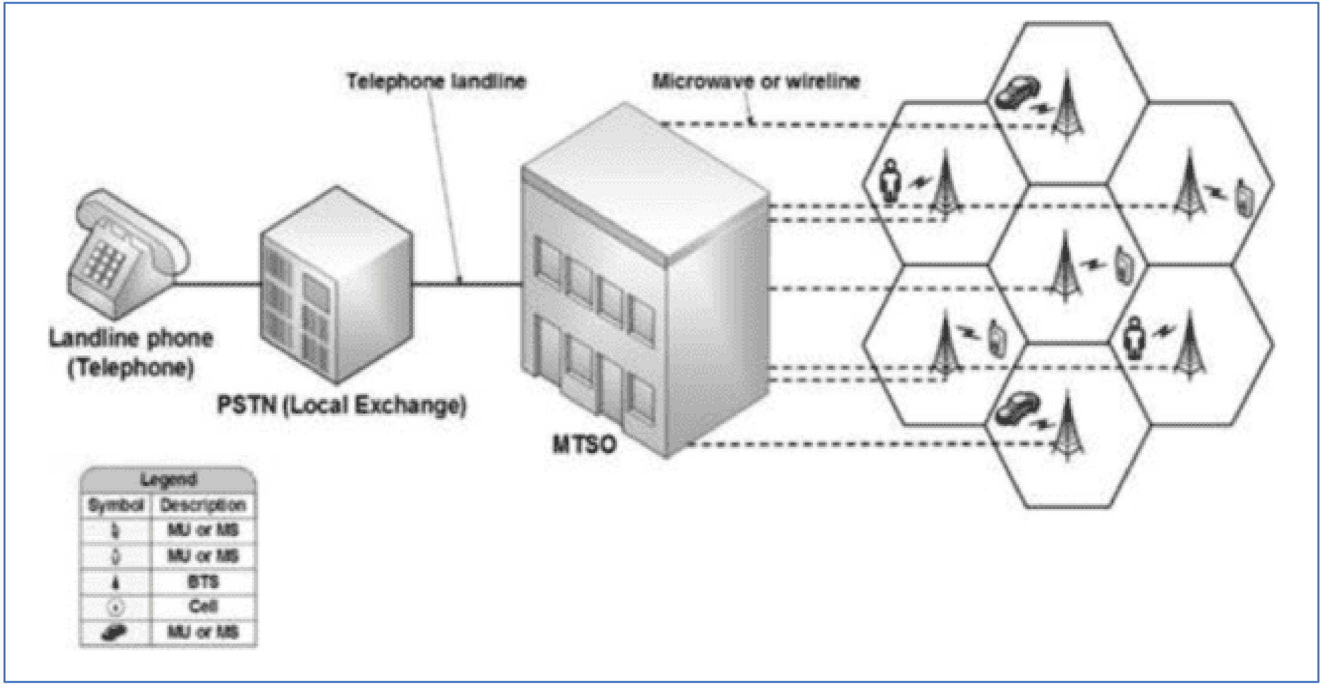


FIGURE 4: ANALOG VOICE TECHNOLOGY

1.2.1 1G: AMPS Architecture [6]

1G denotes the "First Generation" of wireless cellular technology. These analog mobile telecommunications standards were introduced in the 1980s and superseded by 2G. The original concept for a cellular mobile phone network was first published in December 1947 by Douglas H. Ring in a Bell Labs memoranda entitled “Mobile Telephony – Wide Area Coverage.” However, it was not until 1973 that it became a practical reality when Dr. Martin Cooper and his team at Motorola produced the first working cell phone. These first mobiles and their networks used the Advanced Mobile Phone Service standard, entirely based on non-encrypted analog transmission. The Advanced Mobile Phone Service standard, the Total Access Communication System (TACS), was modified for use in the UK.

Nippon Telegraph and Telephone (NTT) launched the first commercial cellular network in Japan in 1979, initially in the metropolitan area of Tokyo. The first phone used in this network was called TZ-801, built by Panasonic. Within five years, the NTT network expanded to cover the whole population of Japan and stood as the first nationwide 1G/cellular network.

The Nordic countries were the pioneers in wireless technologies in the pre-cellular era. These countries designed the NMT standard, first launched in Sweden in 1981. NMT introduced the first mobile phone network to feature international roaming. In 1983, the first 1G cellular network was established in the United States, Chicago-based Ameritech using the Motorola DynaTAC mobile phone.

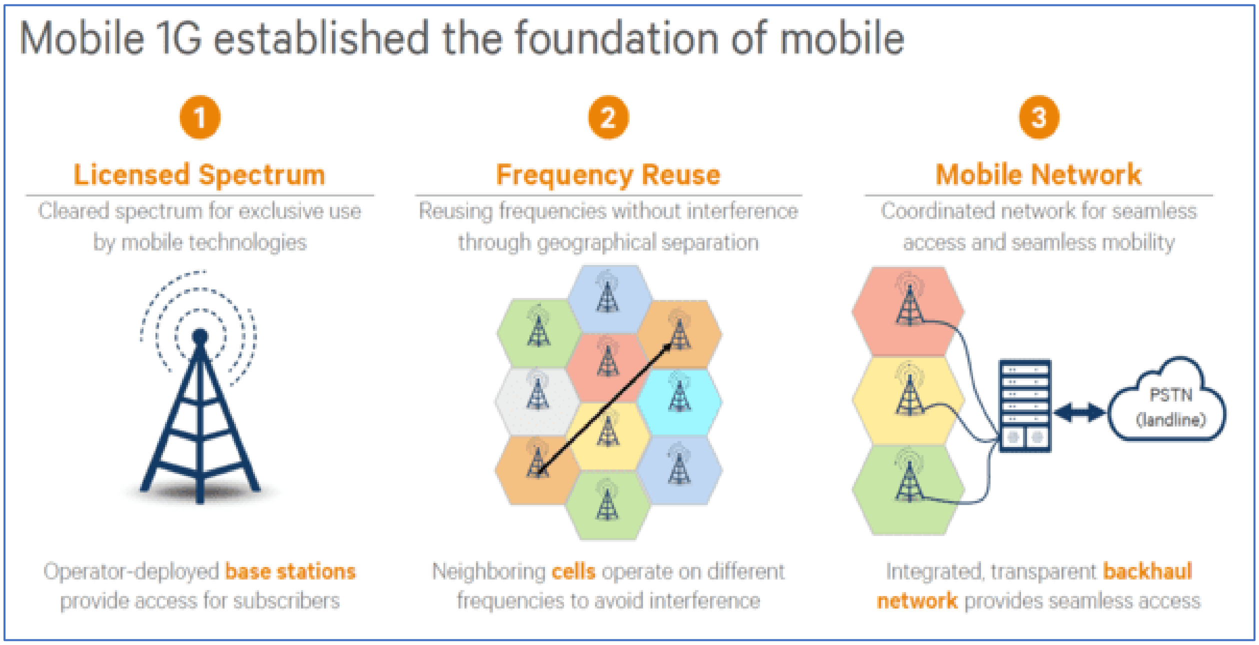
After Japan, the earliest commercial cellular networks were launched in 1981 in Sweden, Norway, and Saudi Arabia, followed by Denmark, Finland, and Spain in 1982, the U.S. in 1983, and Hong Kong, South Korea, Austria, and Canada in 1984.

1G Standards:

Analog cellular technologies that were used were:

- Advanced Mobile Phone System (AMPS)
- Nordic Mobile Telephone (NMT)
- Total Access Communications System (TACS) was developed in the United Kingdom
- C-450 was developed in West Germany and adopted in Portugal and South Africa
- Radiocom 2000 in France
- RTMI in Italy
- MCS-L1 and MCS-L2 (developed by NTT) in Japan
- JTACS (a variant of TACS operated by Daini Denden Planning, Inc. (DDI)) in Japan

Features:



Mobile 1G established the foundation of mobile

1 **Licensed Spectrum**
Cleared spectrum for exclusive use by mobile technologies
Operator-deployed **base stations** provide access for subscribers

2 **Frequency Reuse**
Reusing frequencies without interference through geographical separation
Neighboring **cells** operate on different frequencies to avoid interference

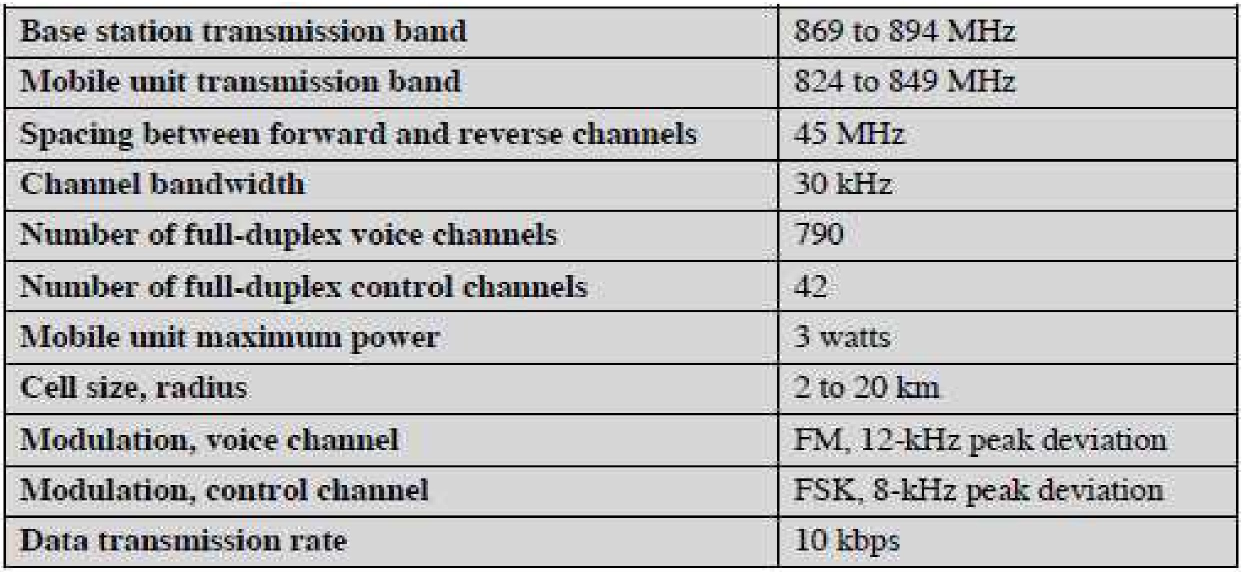
3 **Mobile Network**
Coordinated network for seamless access and seamless mobility
Integrated, transparent **backhaul network** provides seamless access

FIGURE 5: FEATURES OF 1G [5]

1.2.2 Advanced Mobile Phone System (AMPS) Attributes:

Two 25-MHz bands allocated to AMPS:

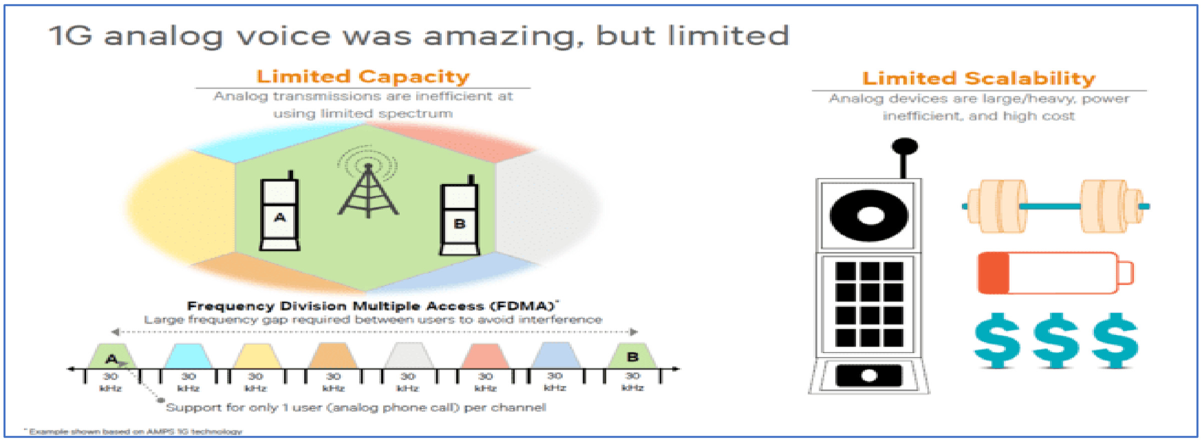
- One for transmission from the base to the mobile unit (DL: 869-894)
- One for transmission from the mobile unit to the base station (UL: 824-849)
- Each band is split into two 12.5 MHz bands to encourage competition (for two operators)
- FDMA and Frequency-Reuse are exploited
- The channels are spaced 30 kHz apart, which allows a total of 416 channels per operator.
- Twenty-one channels are allocated for control, leaving 395 to carry calls.
- Reusing frequency, higher capacity can be achieved
- Traffic channels are analog and use an FM modulation technique



Base station transmission band	869 to 894 MHz
Mobile unit transmission band	824 to 849 MHz
Spacing between forward and reverse channels	45 MHz
Channel bandwidth	30 kHz
Number of full-duplex voice channels	790
Number of full-duplex control channels	42
Mobile unit maximum power	3 watts
Cell size, radius	2 to 20 km
Modulation, voice channel	FM, 12-kHz peak deviation
Modulation, control channel	FSK, 8-kHz peak deviation
Data transmission rate	10 kbps

TABLE 2: AMPS ATTRIBUTES

First Generation Systems Major Limitations:
o Limited Capacity: couldn't cope with the increase in subscribers
o poor handoffs
o Poor voice quality
o High cost of handsets and required large phone size
o Lack of security
o Incompatibility between systems in different countries (no roaming)
o Each country developed its design, incompatible with everyone else's equipment and operation. [8]



1G analog voice was amazing, but limited

Limited Capacity

Analog transmissions are inefficient at using limited spectrum

Frequency Division Multiple Access (FDMA)*

Large frequency gap required between users to avoid interference

Support for only 1 user (analog phone call) per channel

Limited Scalability

Analog devices are large/heavy, power inefficient, and high cost

* Example shown based on AMPS 1G technology

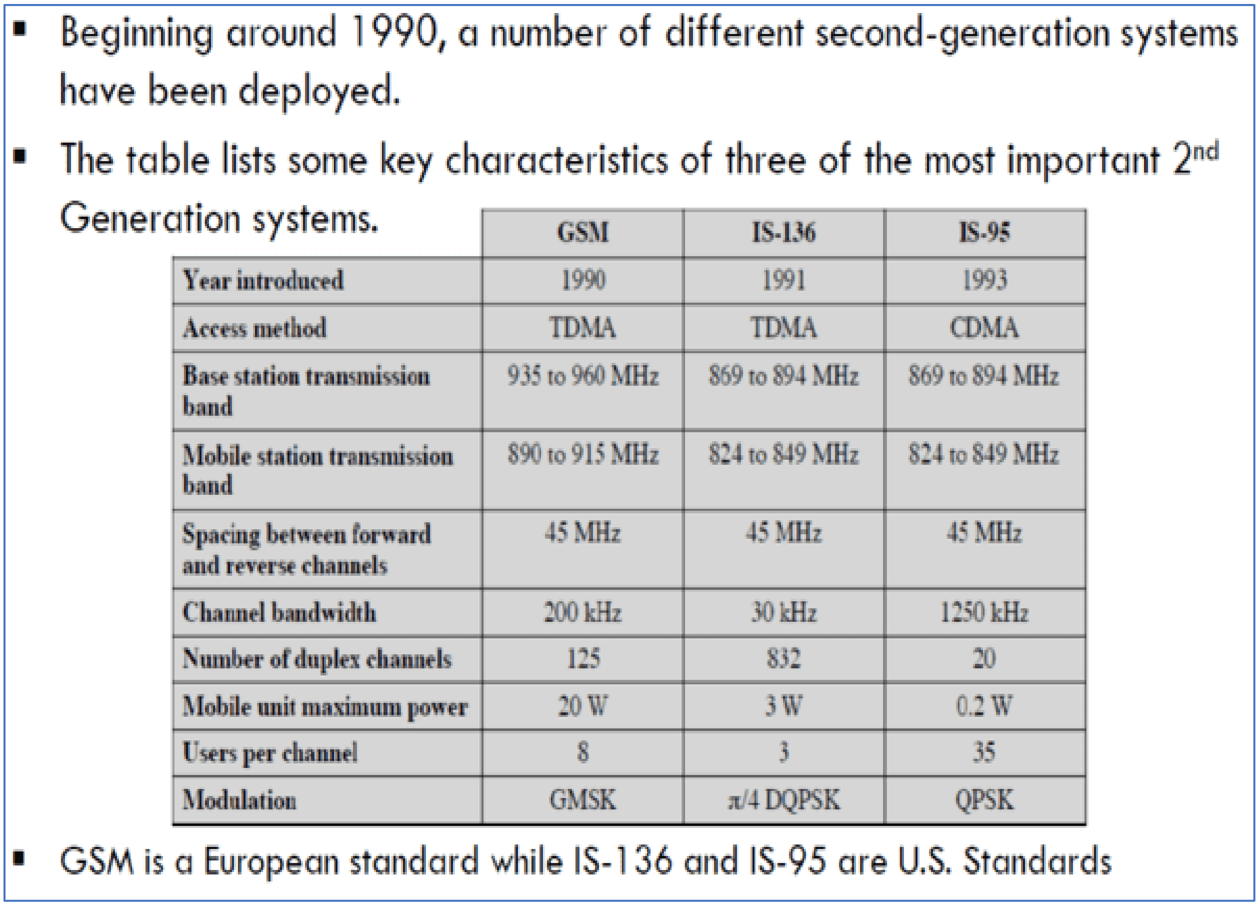
FIGURE 6: 1G PERSPECTIVE [5]

1.3 Second-Generation (2G): Going Digital with Pan-European GSM

2G is a short notation for the **second-generation cellular network**, a group of technology standards deployed for cellular networks. Radiolinja (now part of Elisa Oyj) commercially launched 2G on the GSM standard in Finland in 1991. While 1G radio signals networks are analog, the signal is converted to digital on 2G technology. However, cellular radio towers are connected digitally with both systems to the rest of the mobile network system.

Three primary advantages of 2G networks over their 1G predecessors were:

1. Digitally encrypted conversations between the mobile phone and the cellular base station.
2. The use of efficient the radio frequency spectrum enables more users per frequency band.
3. Mobile data services, starting with SMS text messages and expanding to Services (MMS). [9]



- Beginning around 1990, a number of different second-generation systems have been deployed.
- The table lists some key characteristics of three of the most important 2nd Generation systems.

	GSM	IS-136	IS-95
Year introduced	1990	1991	1993
Access method	TDMA	TDMA	CDMA
Base station transmission band	935 to 960 MHz	869 to 894 MHz	869 to 894 MHz
Mobile station transmission band	890 to 915 MHz	824 to 849 MHz	824 to 849 MHz
Spacing between forward and reverse channels	45 MHz	45 MHz	45 MHz
Channel bandwidth	200 kHz	30 kHz	1250 kHz
Number of duplex channels	125	832	20
Mobile unit maximum power	20 W	3 W	0.2 W
Users per channel	8	3	35
Modulation	GMSK	π/4 DQPSK	QPSK

- GSM is a European standard while IS-136 and IS-95 are U.S. Standards

FIGURE 7: ATTRIBUTES OF 2G [10]

Time-division multiple access (TDMA)-based GSM standard was commonly used for 2G technology in most countries outside Japan and North America. Digital AMPS (IS-54 and IS-136) and CDMA One (IS-95) were the central systems in North America. In Japan, another Personal Handy-phone System (PHS), like the Personal Digital Cellular (PDC), also existed.

1.3.1 GSM

In 1982 the Conference of European Posts and Telegraphs (CEPT) formed a group called the Global System for Mobile Communications (GSM) to study and develop a pan-European public land mobile system.

The proposed system had to meet specific criteria:
o Good subjective speech quality
o Low terminal and service cost
o Support for international roaming
o Ability to support handheld terminals
o Support for a range of new services and facilities
o Spectral efficiency
o ISDN (Integrated Services Digital Network) compatibility

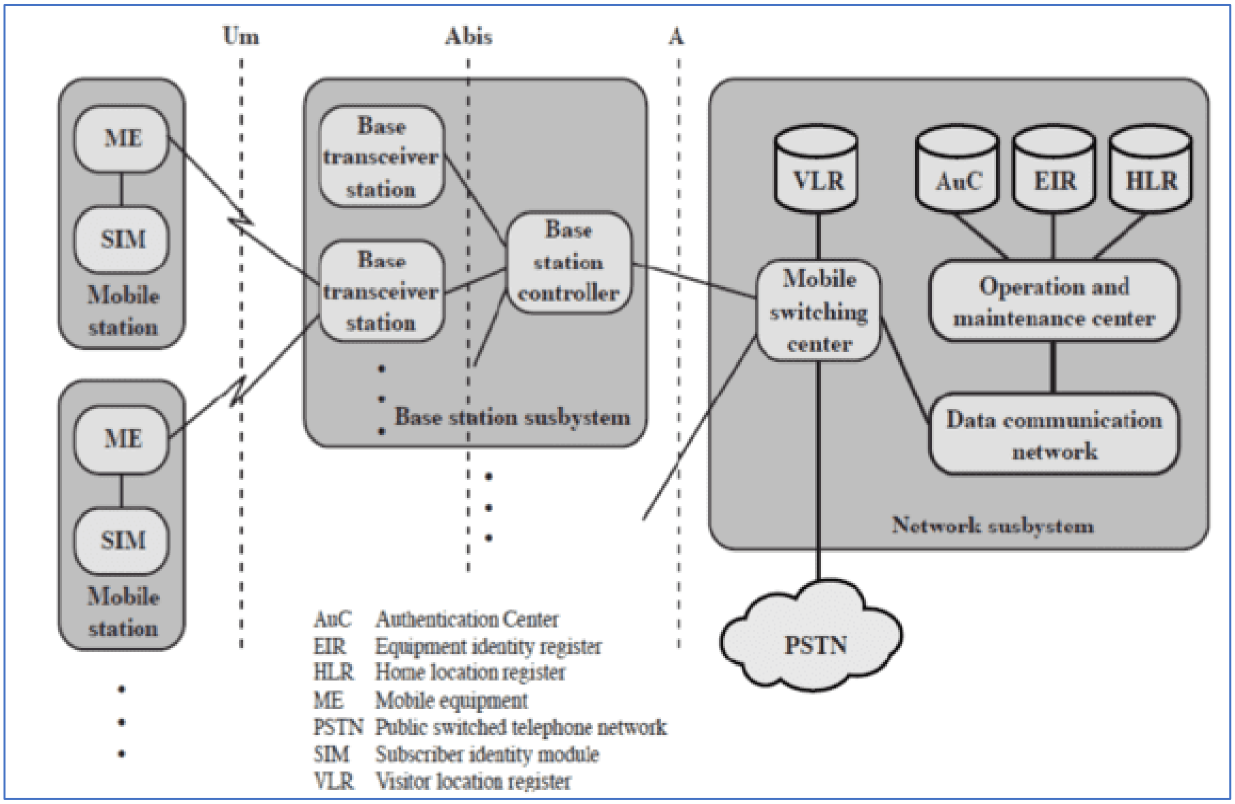
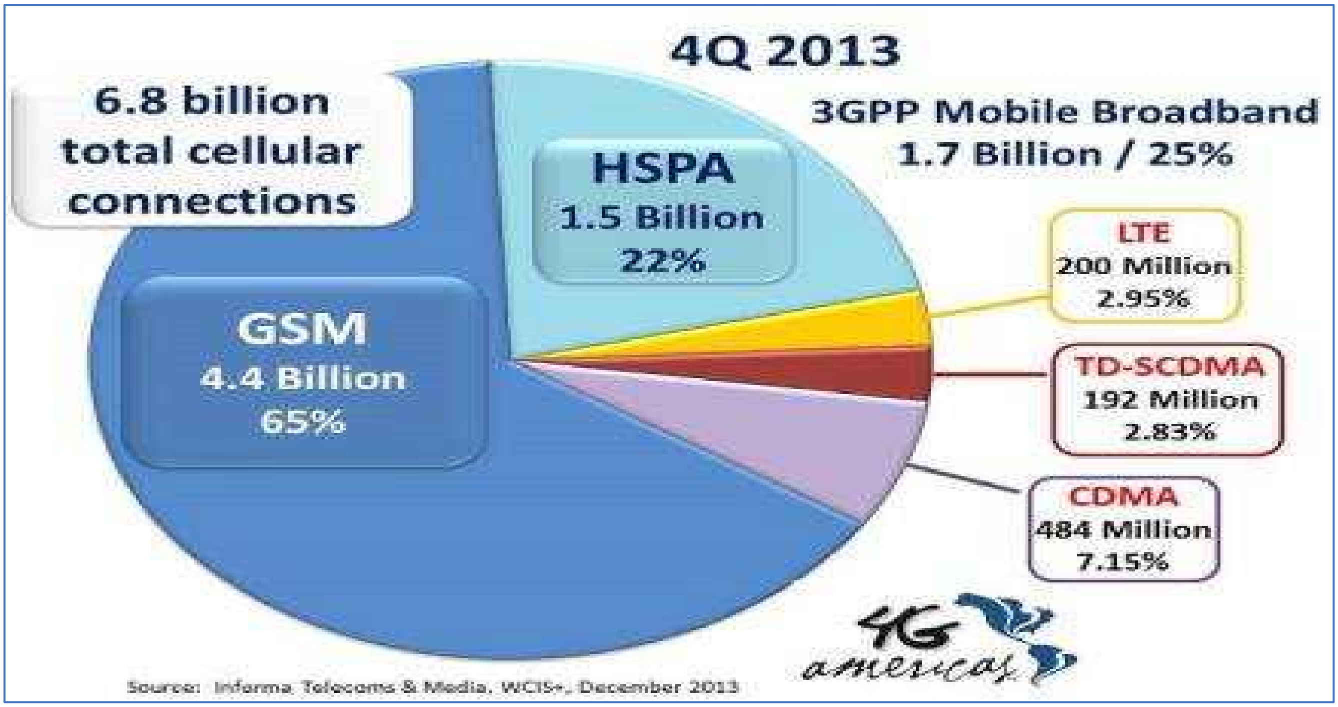

Um
Abis
A
ME
SIM
Mobile station
ME
SIM
Mobile station
Base transceiver station
Base transceiver station
Base station controller
Base station susbsystem
VLR
AuC
EIR
HLR
Mobile switching center
Operation and maintenance center
Data communication network
Network susbystem
PSTN
AuC Authentication Center
EIR Equipment identity register
HLR Home location register
ME Mobile equipment
PSTN Public switched telephone network
SIM Subscriber identity module
VLR Visitor location register

FIGURE 8: 2G GSM NETWORK ARCHITECTURE [10]

In 1989, GSM responsibility was handed over to the European Telecommunication Standards Institute (ETSI). The acronym GSM now stands for Global System for Mobile communications. Phase-I of the GSM specifications was published in 1990, and the commercial service was started in mid-1991. In 2010, the GSM Association estimated that technologies defined in the GSM standard provide 80% of the global mobile market. Moreover, it encompasses more than 5 billion people, and more than 212 countries and territories, making GSM the most omnipresent of the many standards for cellular networks.

By the end of 2013:
o 65 % of the cellular subscription are GSM subscriptions
o 4.4 billion GSM subscribers



Source: Informa Telecoms & Media, WCIS+, December 2013

FIGURE 9: INITIAL PREDICTION OF GSM SUBSCRIPTION [10]

GSM Specifications

- GSM Communication

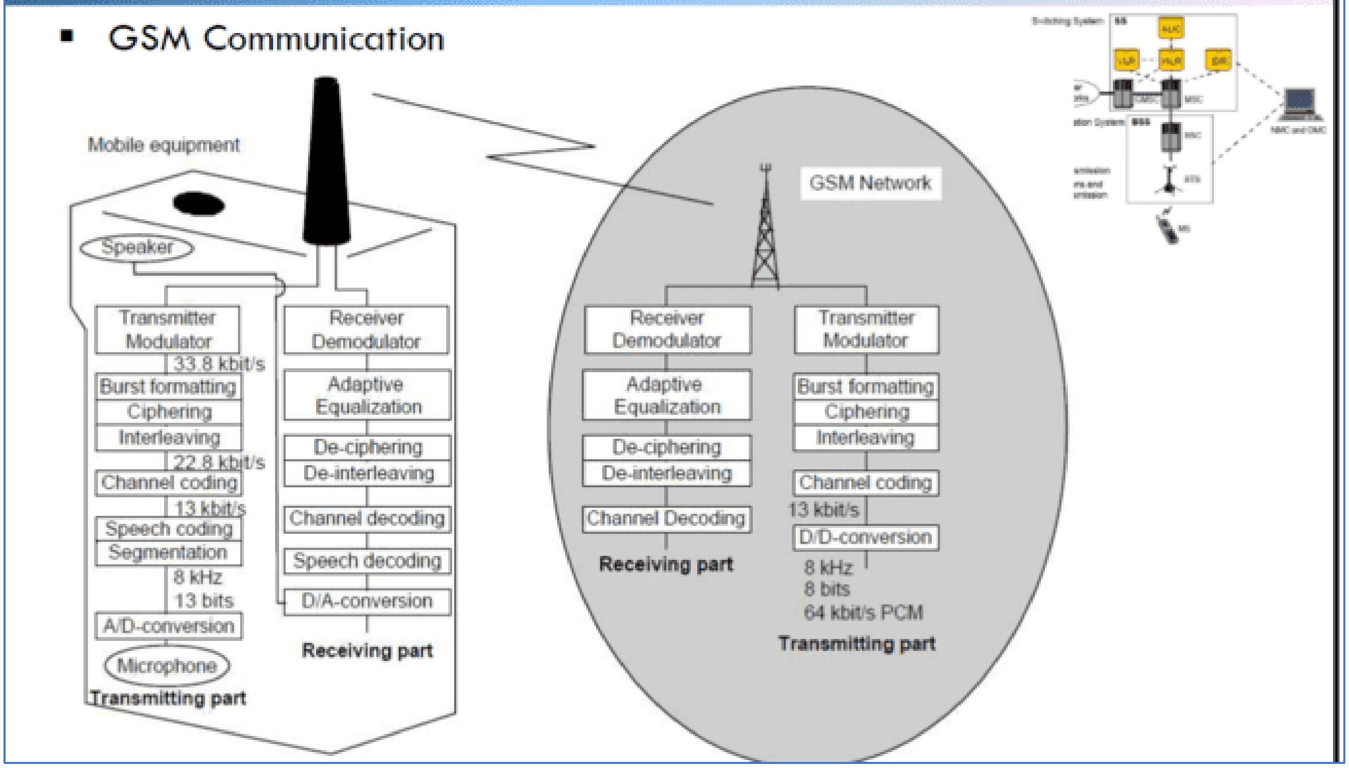


FIGURE 10: GSM SPECIFICATION [10]

The original frequency band specified for GSM was 900MHz. However, as GSM has grown worldwide, it has expanded to operate at four **leading** frequency bands: 900, 1800, 1900, and 850 MHz. In Europe, Africa, and the Middle East, the bands 900MHz and 1800 MHz are commonly used.

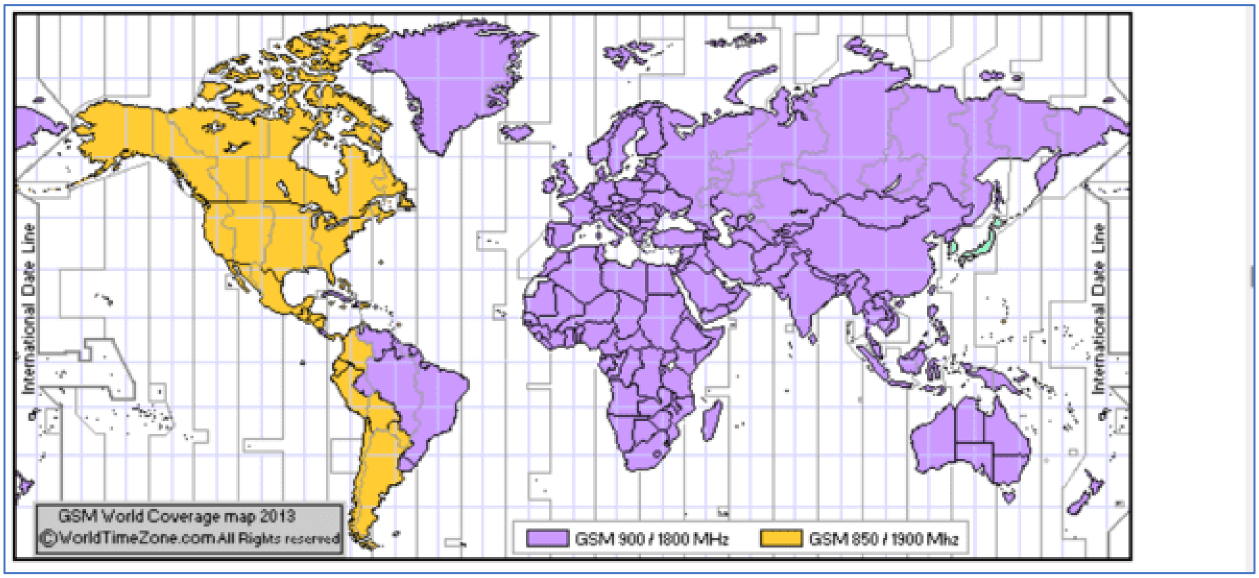


FIGURE 11: GSM WORD COVERAGE MAP-2013 [10]

GSM Evolution

GSM uses a single radio time slot to deliver 9.6 kbit/s data transmission. This is known as Circuit Switched Data (CSD). With data communications and the internet, there has been an increasing demand for improved data handling capabilities. As a result, GSM evolved to provide higher data rates for supporting data applications such as Multimedia messages, Internet browsing, Chatting, Corporate LAN Access, etc. GSM evolved to HSCSD, then to GPRS, then to EDGE➔

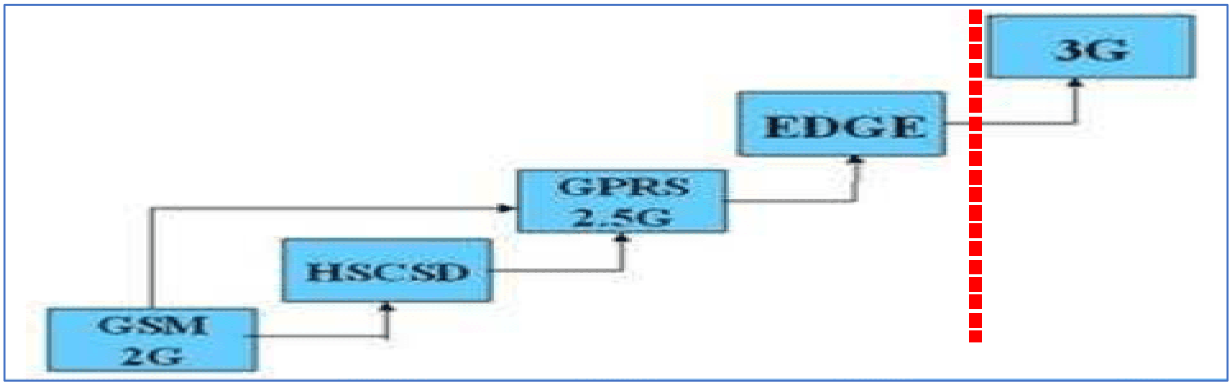


FIGURE 12: GSM EVOLUTION

1.3.2 **High-Speed Circuit Switched Data (HSCSD)**

Based on CSD but designed to provide higher data rates (up to 14.4 Kbps) using more efficient channel coding. In the best-case scenario, CSD can put through 14.4 kbit/s in a single time slot. It can use multiple time slots at the same time. Using the maximum of four-time places can increase the maximum transfer rate by up to 57.6 kbps. HSCSD requires the time slots to be entirely reserved for the user. The network is often configured to allow regular voice calls to take preference over additional time slots for HSCSD users. As a result, the user is typically stimulated for HSCSD at a higher rate than a traditional phone call (the number of allocated time slots) for the total period, and the user has an active connection.

1.3.3 General Packet Radio Service (GPRS)—2.5G

General Packet Radio Service (**GPRS**) is a global mobile communications (GSM) system's packet-oriented mobile data standard on 2G and 3G cellular technology. 2G cellular technology combined with GPRS is described as 2.5G. The European Telecommunications Standards Institute (ETSI) established GPRS to replace the earlier CDPD and its i-mode packet-switched cellular technology. It is now governed by the 3rd Generation Partnership Project (3GPP).

GPRS is a best-effort service, implying variable throughput and latency that depend on the number of other users sharing the service concurrently. The PCU is responsible for assigning channels to the different GPRS MSs. It also manages the transfer of user data packets between MSs and the SGSN..

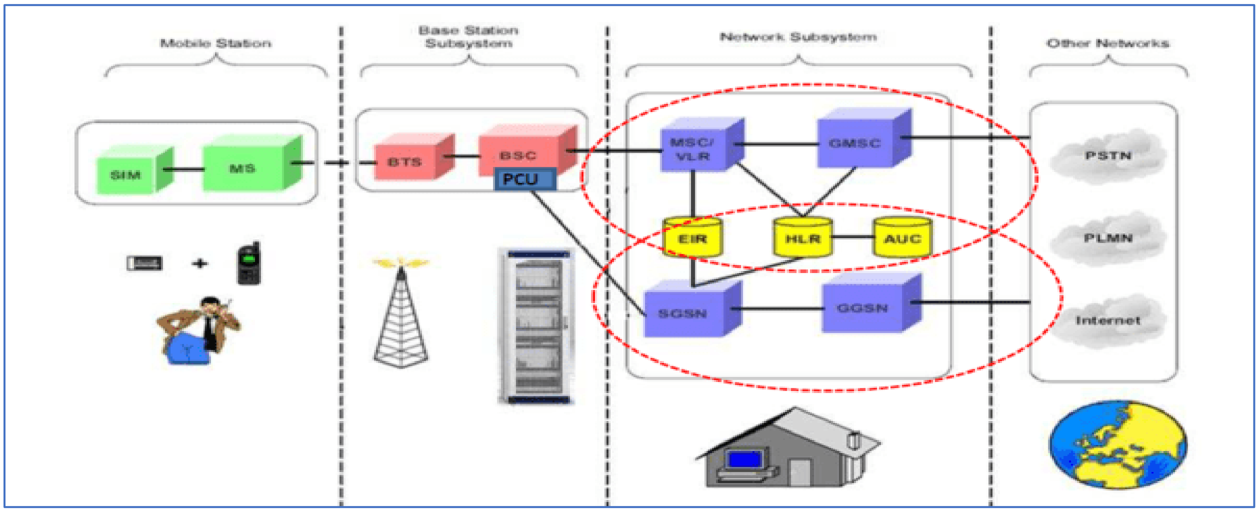


FIGURE 13: GPRS ARCHITECTURE [GENERIC]

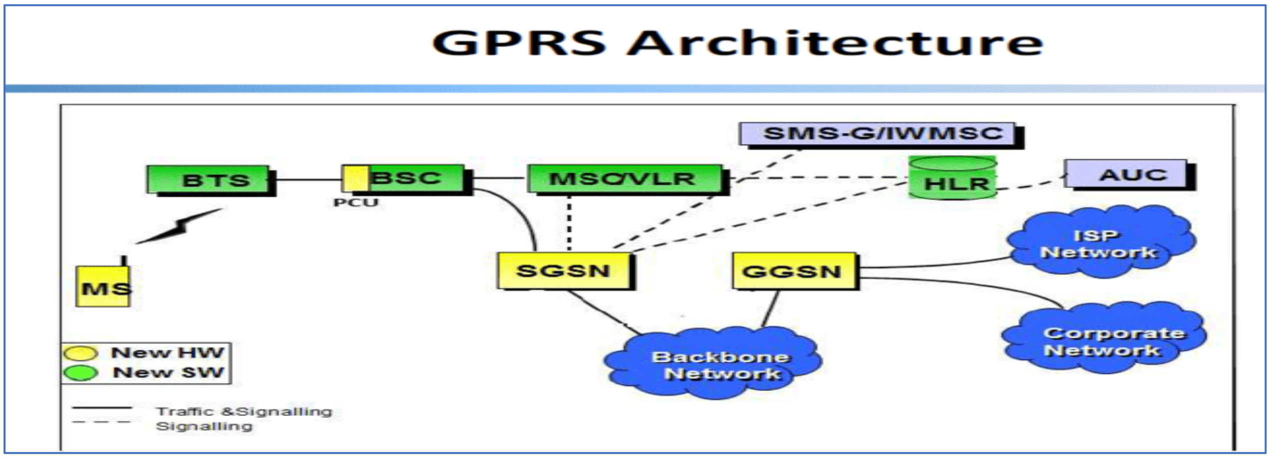


FIGURE 14: GPRS ARCHITECTURE CONSIDERING NEW HW AND SW

Two new nodes are integrated into the GPRS backbone to ensure the functionality:
o The Serving GPRS Support Node (SGSN): provides packet routing to and from the SGSN service area and is responsible for authentication, registration, and collecting information for charging for the air interface.
o The Gateway GPRS Support Node (GGSN) makes up the interfaces towards the external IP networks.

A new Traffic Channel for data (PDTCH: Packet Data Traffic Channel) is used in this architecture. The data rate per PDTCH (per one TDMA slot) depends on the channel coding used. Four coding schemes (CS1 to CS4) are used in GPRS:

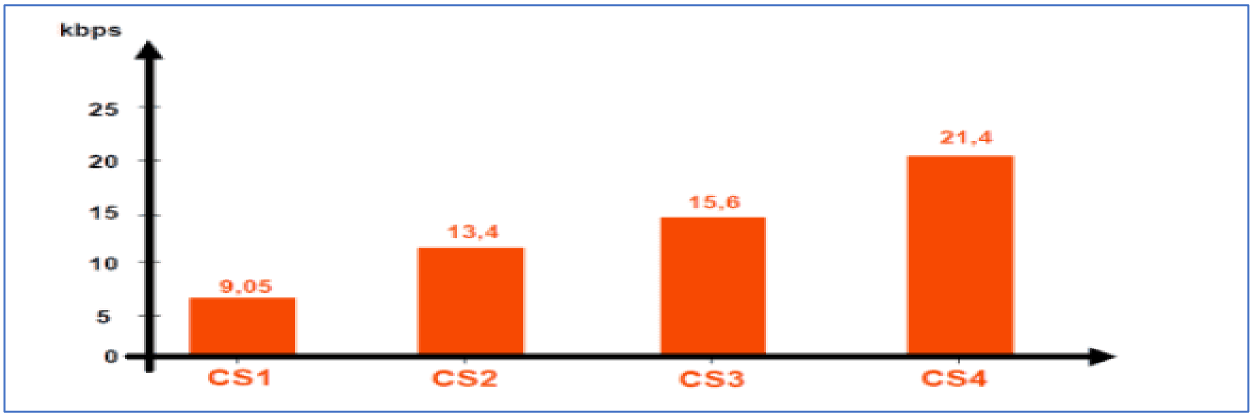


FIGURE 15: GPRS CODING SCHEME [10]

The highest data rate per one PDTCH is 21.4 kbps. It may combine up to 8 PDTCHS, so the max GPRS data rate: is 8*21.4= 171.2 kbps.

1.3.4 Enhanced Data Rate for GSM Evolution (EDGE)

Enhanced Data rates for GSM Evolution-EDGE are also denoted as Enhanced GPRS-EGPRS, IMT Single Carrier IMT-SC, or Enhanced Data rates for Global Evolution). It is a digital mobile phone technology that allows improved data transmission rates as a backward-compatible extension of GSM. EDGE is considered a pre-3G radio technology and is part of ITU's 3G definition. EDGE was initially deployed on GSM networks in the United States in 2003 by Cingular (now AT&T).

3GPP also standardizes EDGE as part of the GSM family. By introducing sophisticated coding and transmitting data methods, EDGE delivers higher bit rates per radio channel, resulting in a threefold increase in capacity and performance compared with an ordinary GSM/GPRS connection. As a result, EDGE can be used for any packet-switched application, such as an Internet connection.[16]

EDGE Data Rate:

In EDGE, the Modulation and Coding Schemes MCS-1 to MCS-9 take the place of the Coding Schemes of GPRS and specify which modulation scheme is used, GMSK or 8PSK. MCS-1 through MCS-4 use GMSK and perform similarly (but not equal) to GPRS, while MCS-5 through MCS-9 use 8PSK. The highest data rate per one-time slot is 59.2 Kbps. It may combine up to 8-time slots, so the max EDGE data-rate: is 8*59.2 = 473.6 kbps. 2G cellular technology combined with EDGE is described as 2.75G

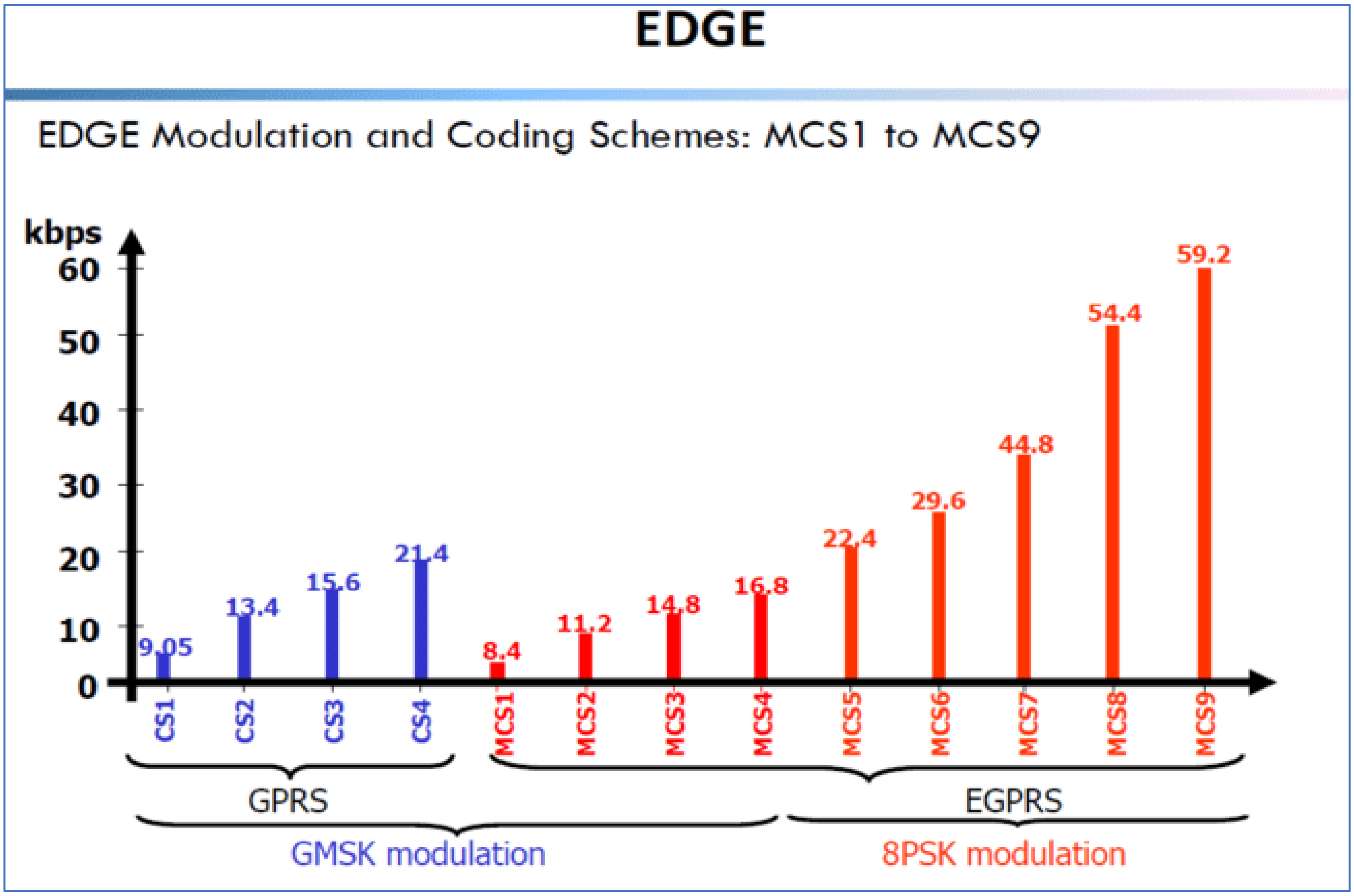
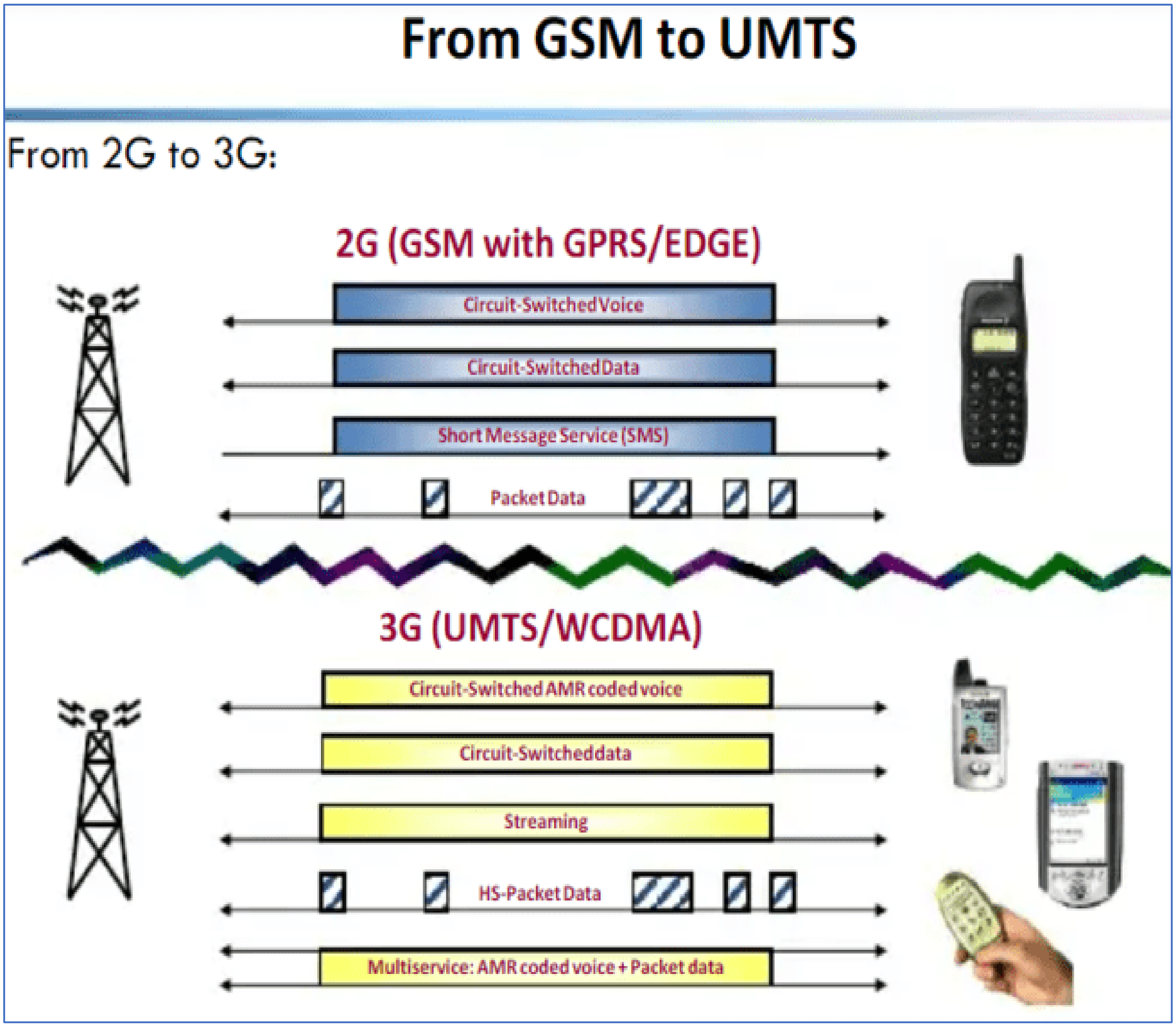


FIGURE 16: COMPARISON BETWEEN GPRS AND EDGE MODULATION [10]

1.4 Third Generation (3G)

The wireless mobile telecommunications technology of the third generation, denoted as **3G** offers faster data transfer and better voice quality. Mobile devices and mobile telecommunications use 3G services and networks that satisfy the International Mobile Telecommunications (IMT) 2000 specifications by the International Telecommunication Union (ITU). 3G finds applications in wireless voice telephony, video calls, mobile TV, mobile Internet access, and fixed wireless Internet access. The first commercial launch of 3G was by NTT DoCoMo in Japan in October 2001.



From GSM to UMTS

From 2G to 3G:

FIGURE 17: BASIC COMPARISON 2G AND 3G TECHNOLOGY [10]

3G telecommunication networks support services with a data transfer rate of 144 kbit/s. Later 3G released 3.5G and 3.75G, which ensure mobile broadband access with several Mbit/s to smartphones and mobile modems in laptops and computers. This provides wireless voice and video calls, mobile Internet access, fixed wireless Internet access, and mobile TV technologies. [11]

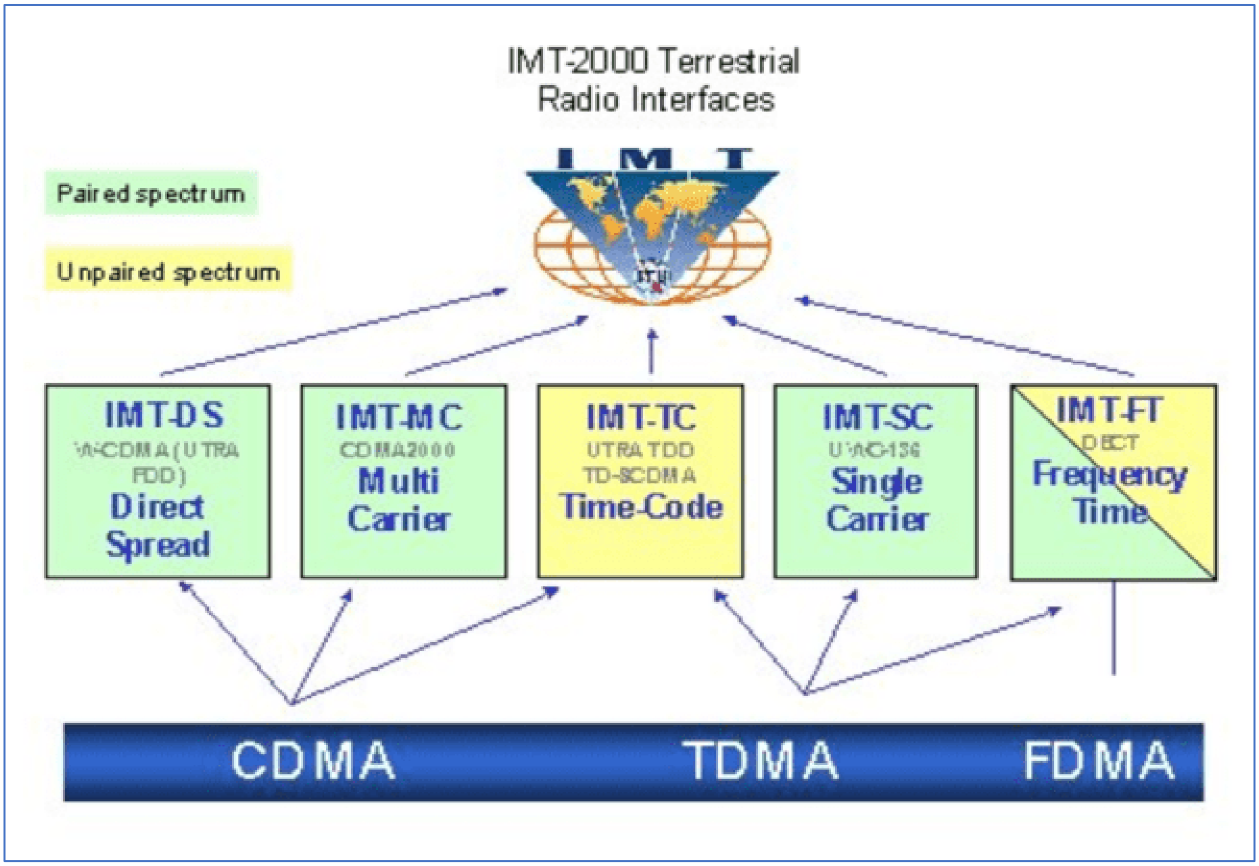


FIGURE 18: IMT-2000 ARCHITECTURE [15]

1.4.1 UMTS (Universal Mobile Telecommunications System)

3G branded standards:

- The UMTS - Universal Mobile Telecommunications System standardized by 3GPP in 2001 was initially used in Europe, Japan, and China with GSM 2G system infrastructure with different radio interfaces. Cell phones are generally UMTS and GSM hybrids. Several radio interfaces are offered and share the same infrastructure:

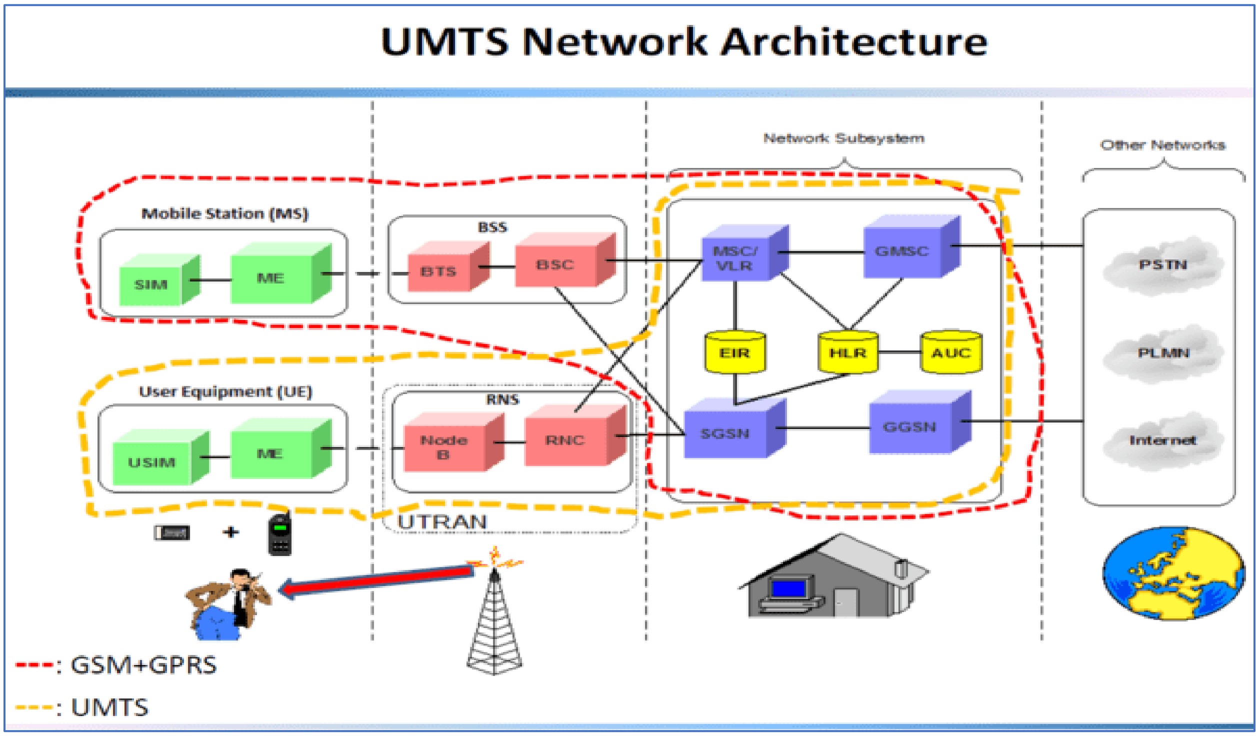
- W-CDMA (Wideband Code Division Multiple Access) is the original and most widespread radio interface.

- The TD-SCDMA radio interface was commercialized in China in 2009.

- HSPA+, the latest release of UMTS, can provide peak data rates of 56 Mbit/s in the downlink in theory and 22 Mbit/s in the uplink.

- The CDMA2000 system was first offered in 2002, standardized by 3GPP2, shared infrastructure with the IS-95 2G standard, and was primarily used in North America and South Korea. As a result, cell phones are typically hybrids of CDMA2000 and IS-95.

- DECT cordless phones and Mobile WiMAX standards are approved as 3G standards by ITU and formally fulfill the IMT-2000 requirements.



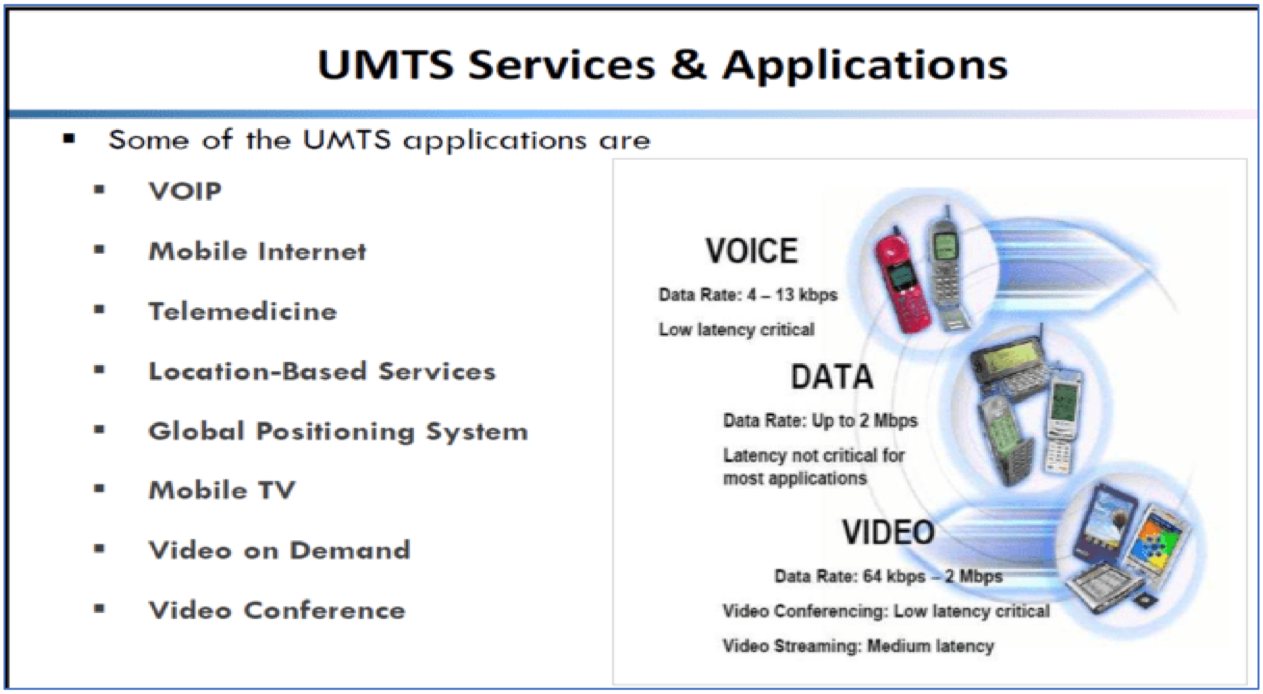
UMTS Network Architecture

---: GSM+GPRS

---: UMTS

FIGURE 19: UMTS NETWORK ARCHITECTURE [10]

UMTS Applications:



UMTS Services & Applications

- Some of the UMTS applications are
  - VOIP
  - Mobile Internet
  - Telemedicine
  - Location-Based Services
  - Global Positioning System
  - Mobile TV
  - Video on Demand
  - Video Conference

VOICE

Data Rate: 4 – 13 kbps

Low latency critical

DATA

Data Rate: Up to 2 Mbps

Latency not critical for most applications

VIDEO

Data Rate: 64 kbps – 2 Mbps

Video Conferencing: Low latency critical

Video Streaming: Medium latency

FIGURE 20: UMTS SERVICE AND APPLICATION

1.4.2 W-CDMA (UTRA-FDD)

W-CDMA (WCDMA; Wideband Code-Division Multiple Access), UTRA-FDD, UMTS-FDD, or IMT-2000 CDMA Direct Spread is a 3G mobile telecommunications network's standard air interface standard. In addition, it provisions conventional cellular voice, text, and MMS services. It can carry data at high speeds, permitting mobile operators to serve higher bandwidth applications, including broadband Internet access and video streaming.

With 5 MHz, W-CDMA uses the DS-CDMA wide channel access method. In contrast, the CDMA2000 system uses 1.25 MHz channels for each direction of communication. W-CDMA systems are widely criticized for their extensive spectrum usage and delayed deployment in countries that acted slowly in allocating new frequencies specifically for 3G services (such as the United States).

1.4.3 High-Speed Packet Access (HSPA) 3.5 G

High-Speed Packet Access (HSPA) is an amalgamation of two mobile protocols, High-Speed Downlink Packet Access (HSDPA) and High-Speed Uplink Packet Access (HSUPA) improves the performance of existing 3G mobile telecommunication networks using the WCDMA protocols. The maximum theoretical DL rate is 14.4 Mbit/s with reduced delay. The ultimate theoretical UL rate achieved is 5.76 Mbit/s.

1.4.4 Evolved High-Speed Packet Access (HSPA+) 3.75G

Evolved High-Speed Packet Access, HSPA (Plus), HSPA+, or HSPAP, is a standard for wireless broadband telecommunication. It is the second phase of HSPA introduced in 3GPP release seven and improved more in later 3GPP releases. HSPA+ confirms data rates of up to 42.2 Mbit/s. In addition, it introduces antenna array technologies such as beamforming and multiple-input multiple-output communications (MIMO). Beamforming focuses the transmitted power of an antenna in a beam toward the user's direction. MIMO uses multiple antennas on the sending and receiving sides. HSPA+ is an evolution of HSPA that advances the existing 3G network and accommodates a method for telecom operators to migrate towards 4G speeds that are more comparable to the initially available speeds of newer LTE networks without deploying a new radio interface.

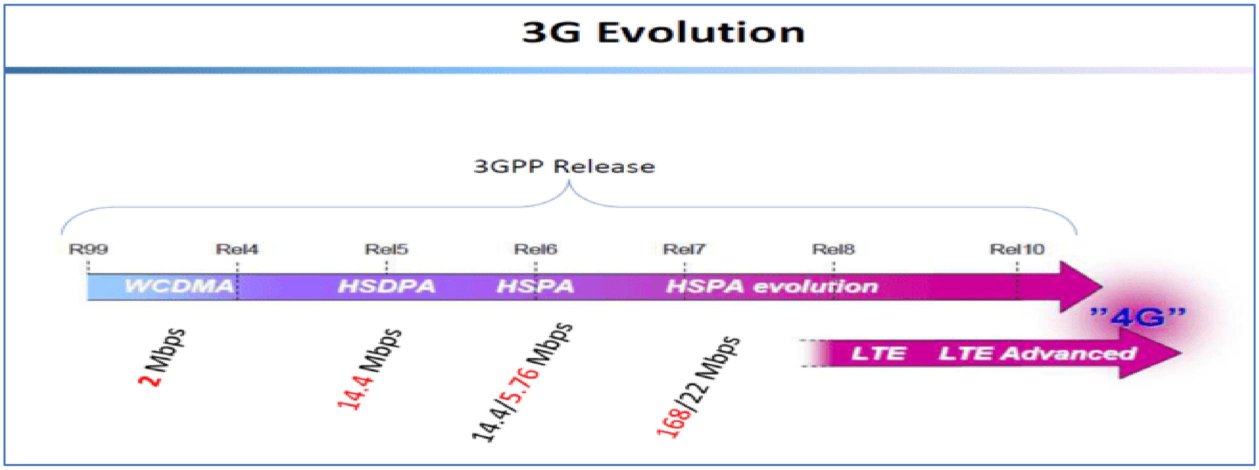
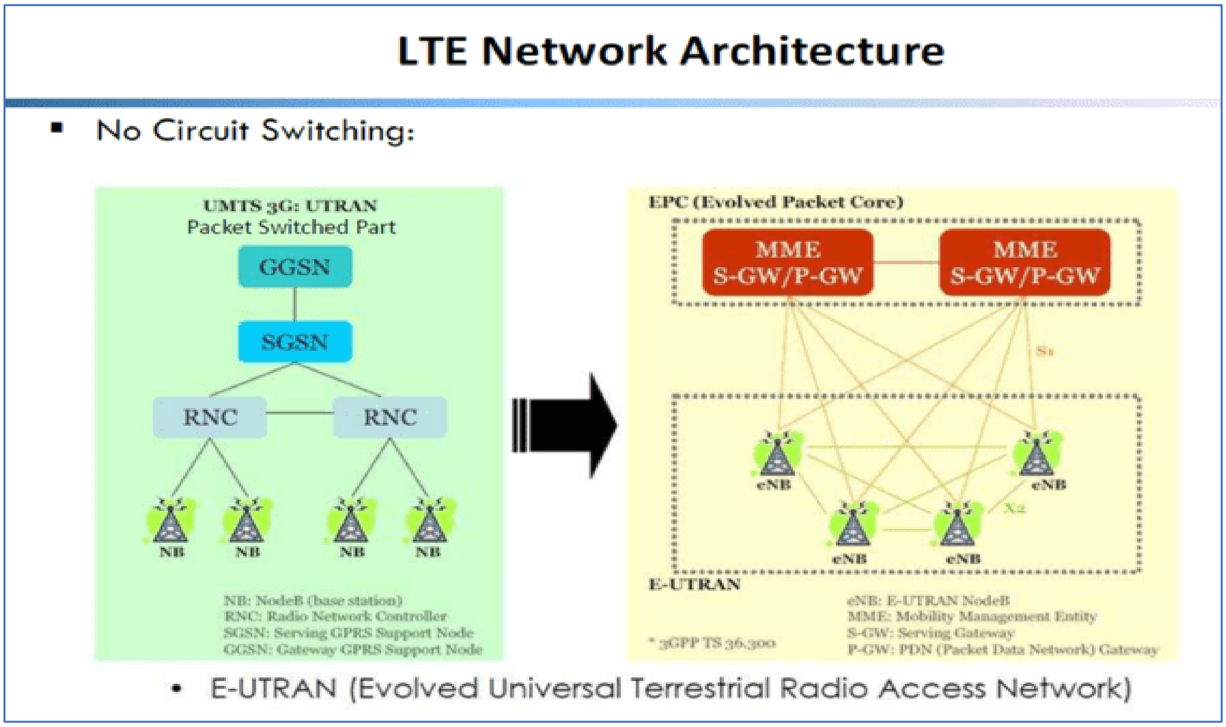


FIGURE 21: 3G EVOLUTION CONCERNING TOP DATA SPEED

1.4.5 pre-4G (3.9G) LTE

LTE - Long Term Evolution is a standard for wireless broadband communication based on the GSM/EDGE and UMTS/HSPA standards for mobile devices and data terminals. Following a different radio interface and core network enhancements improves those standards' capacity and speed. LTE is the upgrade path for GSM/UMTS and CDMA2000 network carriers. However, the frequencies of LTE differ from country to country, and only multi-band phones can use LTE in all countries where it is supported.

The 3GPP (3rd Generation Partnership Project) developed the standard and specified the Release 8 document series, with minor enhancements described in Release 9. As a result, LTE is also mentioned as 3.95G and has been advertised as "4G LTE" and "Advanced 4G,". However, it only meets the 3GPP Release 8 and 9 document series specification for LTE Advanced for 4G wireless service.



LTE Network Architecture

- No Circuit Switching:

NB: NodeB (base station)
RNC: Radio Network Controller
SGSN: Serving GPRS Support Node
GGSN: Gateway GPRS Support Node

eNB: E-UTRAN NodeB
MME: Mobility Management Entity
S-GW: Serving Gateway
P-GW: PDN (Packet Data Network) Gateway

- E-UTRAN (Evolved Universal Terrestrial Radio Access Network)

FIGURE 22: LTE NETWORK ARCHITECTURE

Features:

- Peak download speed rates are 299.6 Mbit/s, whereas upload rates are up to 75.4 Mbit/s depending on the user equipment category.
- For the downlink, OFDMA - Orthogonal frequency division multiple access is used, and for the uplink, Single-carrier FDMA is used to conserve power.
- Support FDD and TDD communication systems and half-duplex FDD with the same radio access technology.
- Increased spectrum flexibility for standardized: 1.4 MHz, 3 MHz, 5 MHz, 10 MHz, 15 MHz, and 20 MHz wide cells.

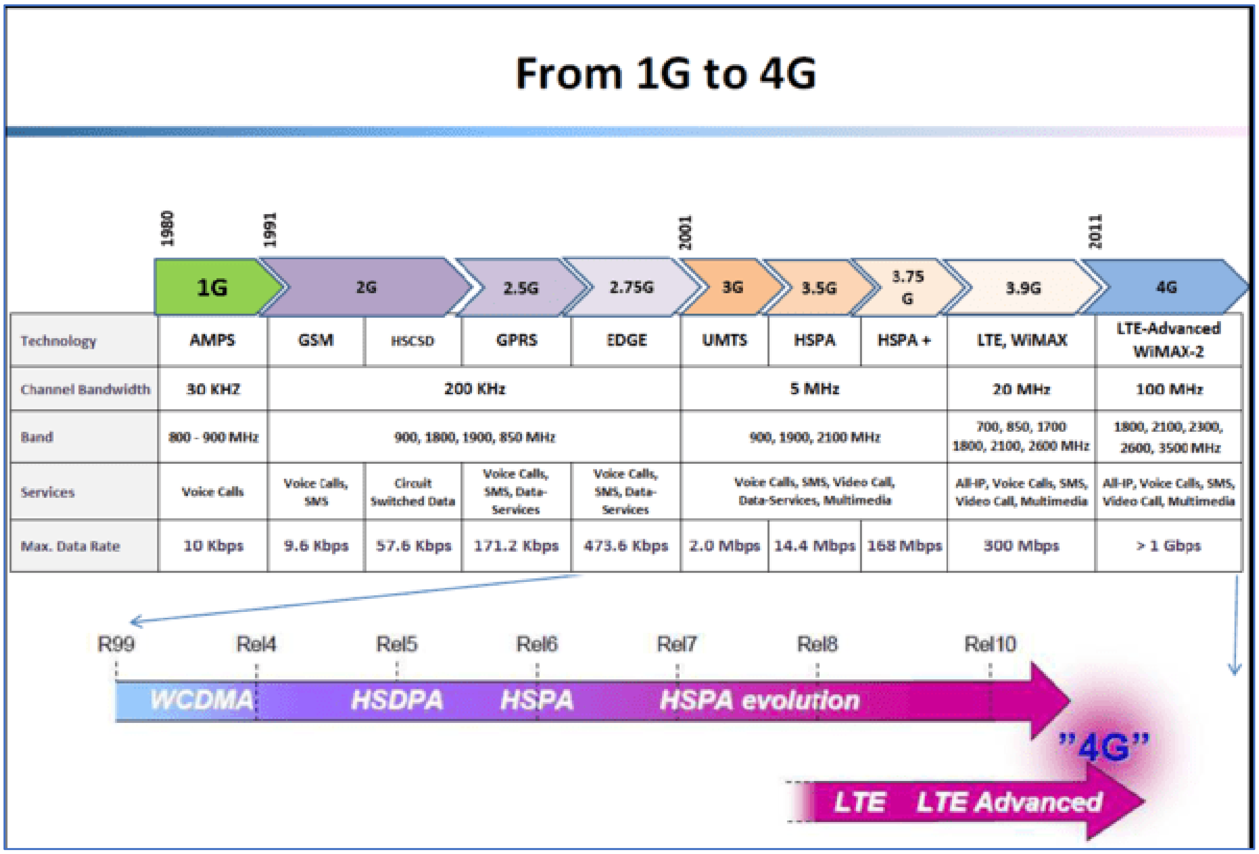
- Support cell sizes from tens of meters radius (femto and picocells) up to 100 km (62 miles) radius macrocells.
- Support 200 active data clients (connected users) in every 5 MHz cell.
- Uplink and downlink Carrier aggregation.
- Packet-switched radio interface. [12]

1.5 Fourth Generation 4G

4G is the fourth-generation of broadband cellular networks, succeeding 3G and preceding 5G. A 4G system must fulfill the capabilities defined by ITU in IMT Advanced. Some potential applications are amended mobile web access, IP telephony, video conferencing, gaming services, high-definition mobile TV, 3D television, etc.

In December 2010, the ITU extended its definition of 4G to initialize Long Term Evolution (LTE), Evolved High-Speed Packet Access (HSPA+), and Worldwide Interoperability for Microwave Access (WiMAX).

Each Generation of wireless cellular technology has ensured increased bandwidth and network capacity. 4G users receive speeds of up to 100 Mbit/s whereas 3G promises a peak speed of 14 Mbit/s. As of 2021, 4G technology occupies 58% of the worldwide mobile telecommunication market.


From 1G to 4G

	1G	2G		2.5G	2.75G	3G	3.5G	3.75 G	3.9G	4G
Technology	AMPS	GSM	HSCSD	GPRS	EDGE	UMTS	HSPA	HSPA +	LTE, WiMAX	LTE-Advanced WiMAX-2
Channel Bandwidth	30 KHZ	200 KHz				5 MHz			20 MHz	100 MHz
Band	800 - 900 MHz	900, 1800, 1900, 850 MHz				900, 1900, 2100 MHz			700, 850, 1700 1800, 2100, 2600 MHz	1800, 2100, 2300, 2600, 3500 MHz
Services	Voice Calls	Voice Calls, SMS	Circuit Switched Data	Voice Calls, SMS, Data-Services	Voice Calls, SMS, Data-Services	Voice Calls, SMS, Video Call, Data-Services, Multimedia			All-IP, Voice Calls, SMS, Video Call, Multimedia	All-IP, Voice Calls, SMS, Video Call, Multimedia
Max. Data Rate	10 Kbps	9.6 Kbps	57.6 Kbps	171.2 Kbps	473.6 Kbps	2.0 Mbps	14.4 Mbps	168 Mbps	300 Mbps	> 1 Gbps

FIGURE 23: GENERATION-TO-GENERATION ATTRIBUTES COMPARISON

System Standards

IMT-2000 compliant 4G standards:
As of October 2010, ITU-R Working Party 5D approved two industry-developed technologies (LTE Advanced and Wireless MAN-Advanced) for inclusion in the ITU's International Mobile Telecommunications Advanced program (IMT-Advanced program), which is focused on global communication systems.

1.5.1 4.5G LTE Advanced

LTE Advanced (LTE+) is a cellular communication standard and a significant enhancement of the Long Term Evolution (LTE) standard. The LTE+ format was first proposed in Japan by NTT DoCoMo and has been adopted as the international standard. In December 2009, Sweden and Norway first commercially launched it, followed by the United States and Japan in 2010. It is also called IMT Advanced (4.5 G) as defined by the International Telecommunication Union.

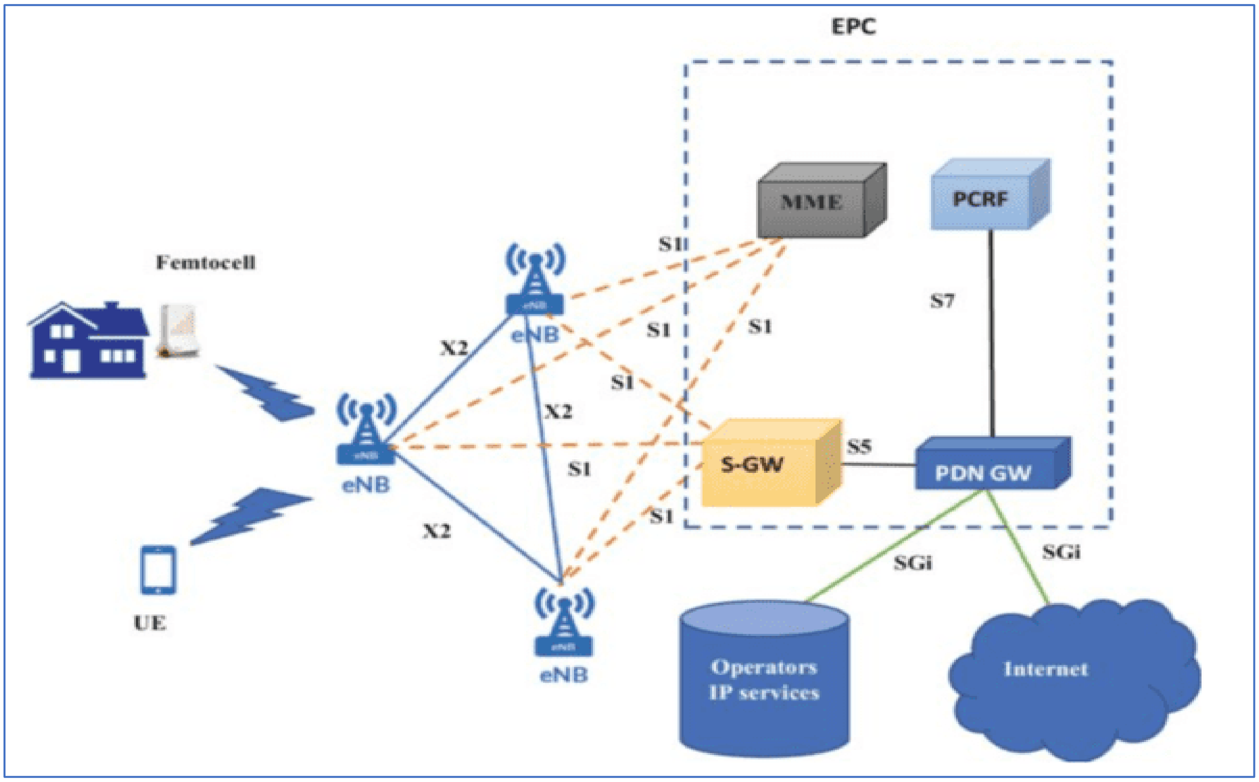


FIGURE 24: LTE ADVANCED ARCHITECTURE [13]

One of the critical LTE Advanced benefits is the capability to grant advantage of advanced topology networks and optimized heterogeneous networks with a mix of macro-cells with low power nodes such as picocells, femtocells, and new relay nodes. as a result, LTE Advanced further improves the capacity and coverage and ensures user fairness. LTE Advanced **also** introduces multicarrier to ultra-wide bandwidth, which supports up to 100 MHz of the spectrum with very high data rates.

Many proposals have been studied for LTE Advanced (LTE-A) technologies in the research phase. Therefore, the recommendations could roughly be categorized into the:

- Provision base stations for relay node
- Coordinated multipoint (CoMP) reception and transmission
- UE solutions for SU-MIMO and diversity MIMO, commonly called 2x2 MIMO, support Dual TX antenna.
- Scalable system bandwidth ranges from 20 MHz to 100 MHz
- Carrier aggregation confirms contiguous and non-contiguous spectrum allocations
- Local area optimization of air interface
- Nomadic / Local Area network and mobility solutions
- Flexible spectrum usage
- Cognitive radio
- Automatic and autonomous network configuration and operation
- Support of autonomous network and device tests tied to network optimization and management
- Precoding and forward error correction are enhanced
- Interference suppression and management
- FDD allocates asymmetric bandwidth assignment
- Uplink uses hybrid OFDMA and SC-FDMA
- Supports SONs, Self Organizing Networks features and methodologies

LTE-Advanced and WiMAX 2 use up to 8x8 MIMO and 128-QAM in the downlink direction. For example, 100 MHz aggregated bandwidth, LTE-Advanced provides almost 3.3 Gbit peak download rates per base station sector under ideal conditions. The distributed and collaborative, innovative antenna technologies of advanced network architectures provide several years' road map of commercial enhancements. In 2019, the Global mobile Suppliers Association (GSA) reported that 304 commercially launched LTE-Advanced networks were deployed in 134 countries. For tests, trials, deployments, or commercial service provision, 335 operators are investing in LTE-Advanced in 141 countries. [12]

1.5.2 4.9G LTE Advanced Pro

LTE Advanced Pro (LTE-A Pro, also known as 4.9G, Pre-5G) is a name for 3GPP releases 13 and 14. It is an evolution of the LTE Advanced (LTE-A) cellular standard supporting data rates over 3 Gbit/s using 32-carrier aggregation. It also introduces the concept of License Assisted Access, which allows sharing of licensed and unlicensed spectrum. Additionally, it incorporates several new technologies associated with 5G, such as 256-QAM, Massive MIMO, LTE-Unlicensed, and LTE IoT, that facilitated early migration of existing networks to enhancements promised with the full 5G standard. [12]

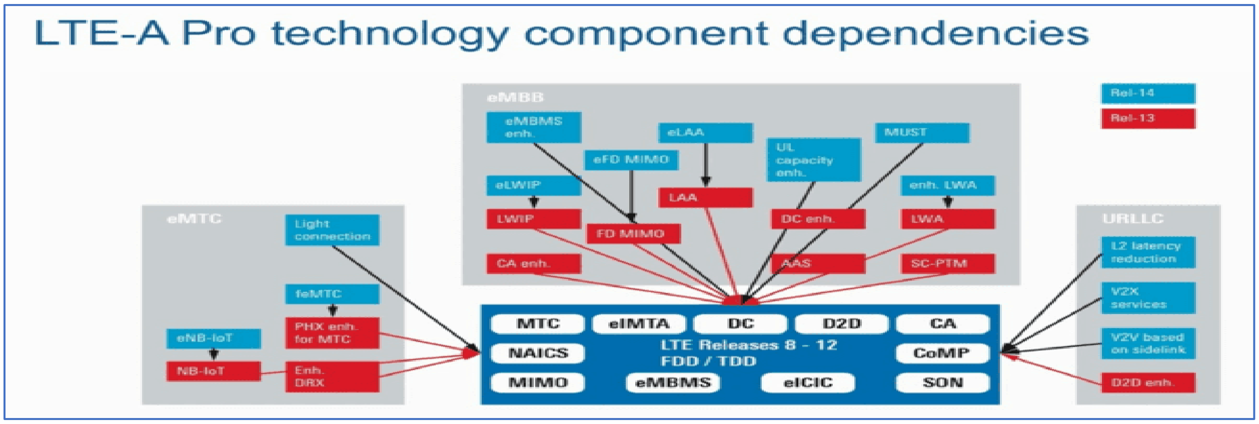


FIGURE 25: LTE-A PRO TECHNOLOGY ARCHITECTURE [14]

SECTION 2. BIRD'S EYE VIEW OF 5G TECHNOLOGY

5G:

The word '5G' stands for the fifth generation and the latest mobile wireless standard based on the IEEE 802.11ac standard of broadband technology. It represents a significant step forward regarding speed, capacity, and performance. For example, the new networks have higher download speeds of up to 10 gig/sec (Gbit/s).

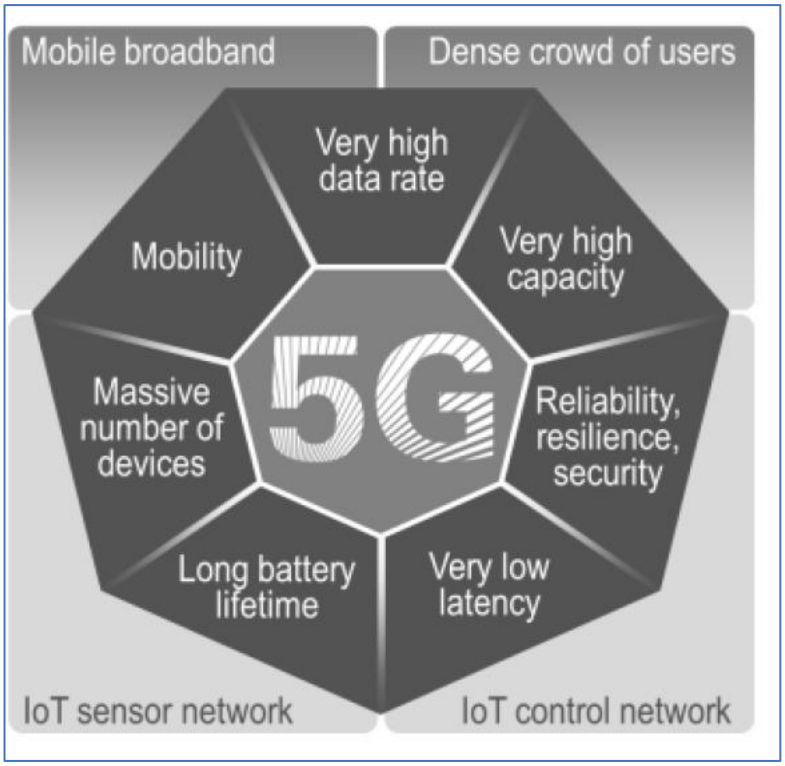


FIGURE 26: WHY 5G

With the 5G technology, users can expect significantly faster download and upload speeds, lower latency, or time for transmitting data. This makes 5G ideal for applications that require high-bandwidth and low-latency connections, such as streaming video, online gaming, and virtual reality.

The industry consortium 3rd Generation Partnership Project (**3GPP**) defines the standard of "5G" as any system using 5G NR (5G New Radio) software – and this definition came into general use by late 2018. Cellular phone companies began deploying 5G worldwide in 2019. It is the planned successor to the 4G networks, which provide connectivity to most current cell phones. This technology divides the geographical service areas into small cells. The 5G wireless devices within a service cell area communicate via radio waves via a cellular base station where specific frequency channels are assigned via fixed antenna. The base stations, named nodes, are integrated with the switching centers in the telephone network architecture. In contrast, routers are connected for Internet access by a high-bandwidth optical fiber or wireless backhaul connections. Mobile devices moving from one service cell to another are seamlessly handed off like other cellular networks.

There are endless benefits yet to explore for 5G technology, and it is believed that the future will change by adopting 5G technology. The main advantages of 5G are its ability to support a much larger number of devices than previous generations of cellular technology. This makes it ideal for applications that require high levels of connectivity, such as smart cities, connected vehicles, and the Internet of Things (**IoT**). 5G also has the potential to revolutionize a wide range of industries, from healthcare and education to entertainment and retail.

2.1 Introduction to 5G Architecture

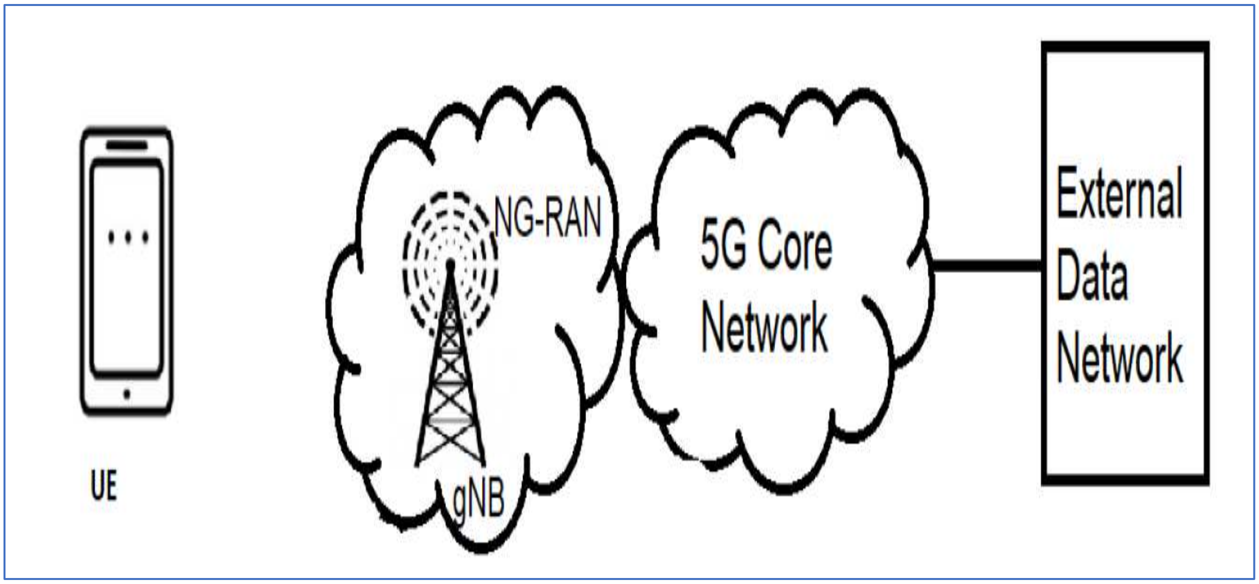


FIGURE 27: 5G GENERAL ARCHITECTURE

The original purpose of 3GPP (1998) was to initialize the Technical Specifications and Reports for a 3G Mobile System based on evolved GSM radio access technologies and core networks. The technologies support Frequency Division Duplex (FDD) and Time Division Duplex (TDD) modes.

The scope was subsequently amended to include the maintenance and development of the Technical Specifications and Reports for evolved 3GPP technologies beyond 3G.

The 5G and LTE-Advanced ecosystem will allow for global network evolution at the appropriate pace for the market need and local readiness. In addition, the 3GPP model maximizes its compatibility with legacy 3GPP infrastructure and equipment, delivering the promise of a ubiquitous end-to-end ecosystem that can support a growing number of use cases.

Research on 5G services and their technical requirements has been performed by the International Telecommunication Union-Radiocommunication Sector (ITU-R), the 3rd Generation Partnership Project (3GPP), and the Next Generation Mobile Networks-NGMN Alliance. In the ITU-R Working Party (WP) 5D, 5G is defined by International Mobile Telecommunications-2020 (IMT-2020), and various 5G services are presented in a vision document.

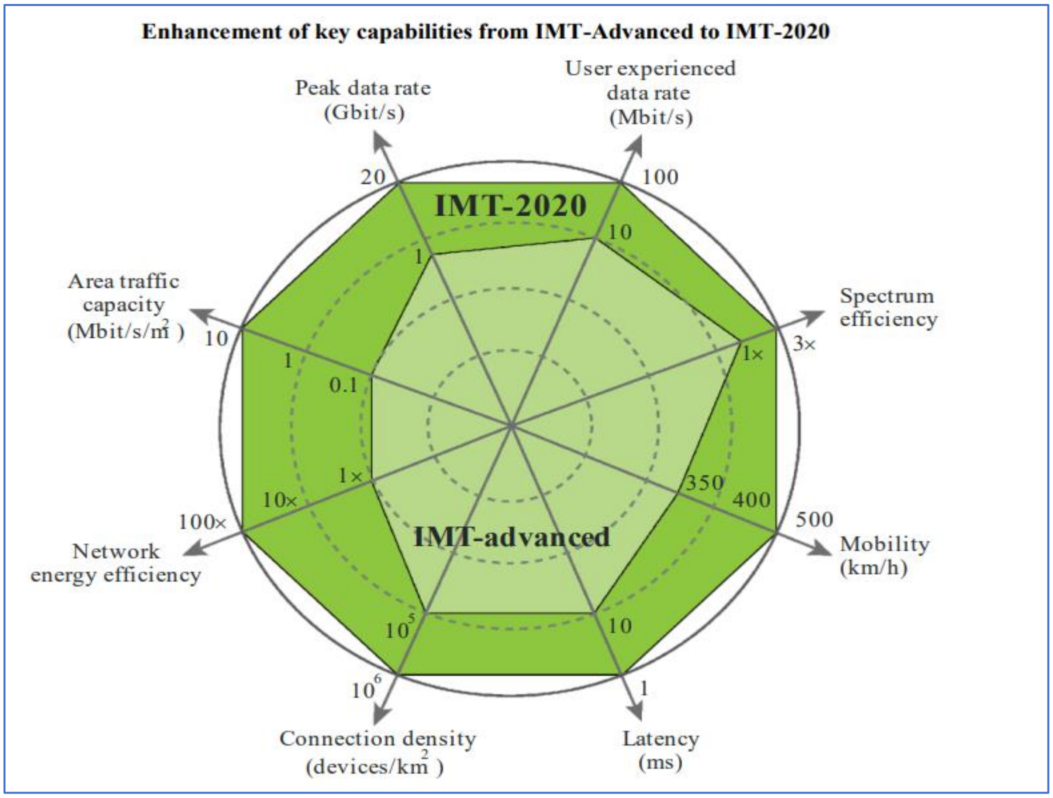


FIGURE 28: IMT VISION, ITU-R M.2083-0 [21]

2019 was a big year for 5G since it saw the main drop of Release 15, the first Release to directly address 5G operation. But Release 15 is just the beginning. Here is the information about releases:

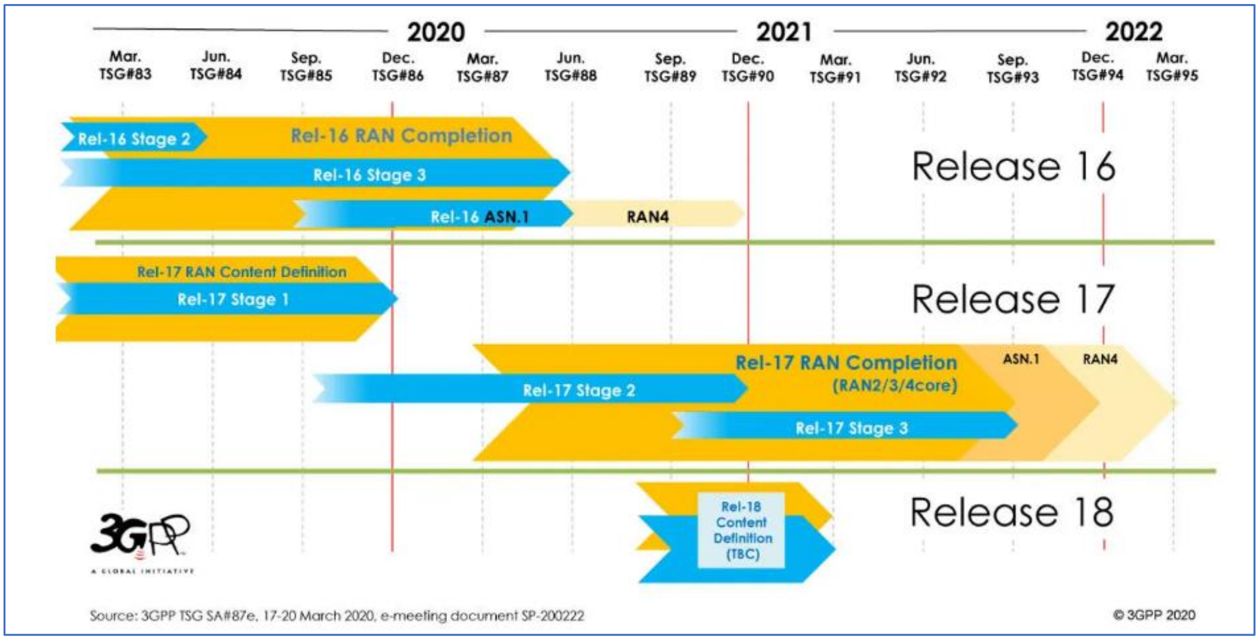
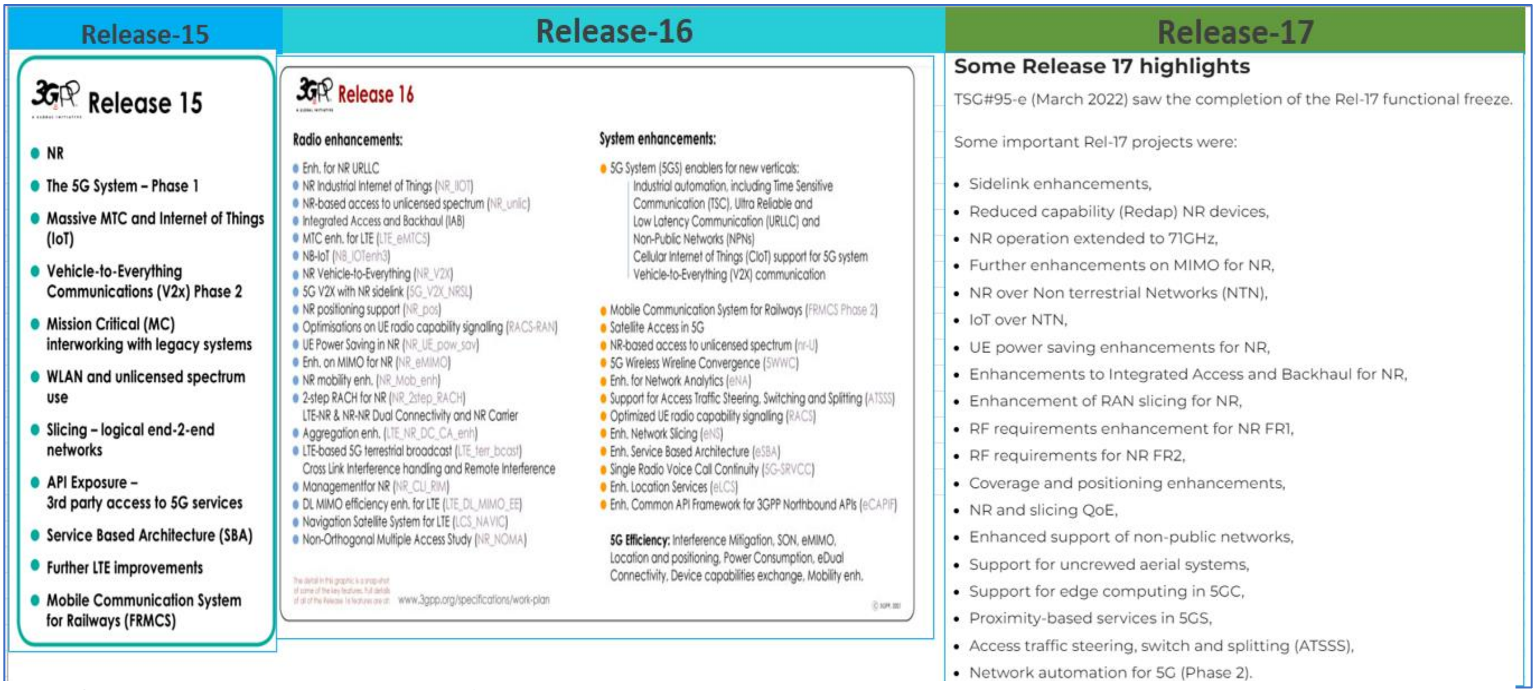


FIGURE 29: TIMETABLES FOR FUTURE 5G-RELATED ARCHITECTURE RELEASES AS OF APRIL 2021

Image Source: 3GPP.org



Release-15

3GPP Release 15

- NR
- The 5G System – Phase 1
- Massive MTC and Internet of Things (IoT)
- Vehicle-to-Everything Communications (V2x) Phase 2
- Mission Critical (MC) interworking with legacy systems
- WLAN and unlicensed spectrum use
- Slicing – logical end-2-end networks
- API Exposure – 3rd party access to 5G services
- Service Based Architecture (SBA)
- Further LTE improvements
- Mobile Communication System for Railways (FRMCS)

Release-16

3GPP Release 16

Radio enhancements:

- Enh. for NR URLLC
- NR Industrial Internet of Things (NR_IIOT)
- NR-based access to unlicensed spectrum (NR_unlic)
- Integrated Access and Backhaul (IAB)
- MTC enh. for LTE (LTE_eMTC5)
- NB-IoT (NB_IOTenh3)
- NR Vehicle-to-Everything (NR_V2X)
- 5G V2X with NR sidelink (5G_V2X_NRSL)
- NR positioning support (NR_pos)
- Optimisations on UE radio capability signalling (RACS-RAN)
- UE Power Saving in NR (NR_UE_pow_sav)
- Enh. on MIMO for NR (NR_eMIMO)
- NR mobility enh. (NR_Mob_enh)
- 2-step RACH for NR (NR_2step_RACH)
  LTE-NR & NR-NR Dual Connectivity and NR Carrier
- Aggregation enh. (LTE_NR_DC_CA_enh)
- LTE-based 5G terrestrial broadcast (LTE_terr_bcast)
  Cross Link Interference handling and Remote Interference
- Managementfor NR (NR_CLI_RIM)
- DL MIMO efficiency enh. for LTE (LTE_DL_MIMO_EE)
- Navigation Satellite System for LTE (LCS_NAVIC)
- Non-Orthogonal Multiple Access Study (NR_NOMA)

System enhancements:

- 5G System (5GS) enablers for new verticals:
  Industrial automation, including Time Sensitive Communication (TSC), Ultra Reliable and Low Latency Communication (URLLC) and Non-Public Networks (NPNs)
  Cellular Internet of Things (CIoT) support for 5G system
  Vehicle-to-Everything (V2X) communication
- Mobile Communication System for Railways (FRMCS Phase 2)
- Satellite Access in 5G
- NR-based access to unlicensed spectrum (nr-U)
- 5G Wireless Wireline Convergence (5WWC)
- Enh. for Network Analytics (eNA)
- Support for Access Traffic Steering, Switching and Splitting (ATSSS)
- Optimized UE radio capability signalling (RACS)
- Enh. Network Slicing (eNS)
- Enh. Service Based Architecture (eSBA)
- Single Radio Voice Call Continuity (5G-SRVCC)
- Enh. Location Services (eLCS)
- Enh. Common API Framework for 3GPP Northbound APIs (eCAPIF)

5G Efficiency: Interference Mitigation, SON, eMIMO, Location and positioning, Power Consumption, eDual Connectivity, Device capabilities exchange, Mobility enh.

The detail in this graphic is a snap-shot of some of the key features. Full details of all of the Release 16 features are at: www.3gpp.org/specifications/work-plan

© 3GPP 2021

Release-17

Some Release 17 highlights

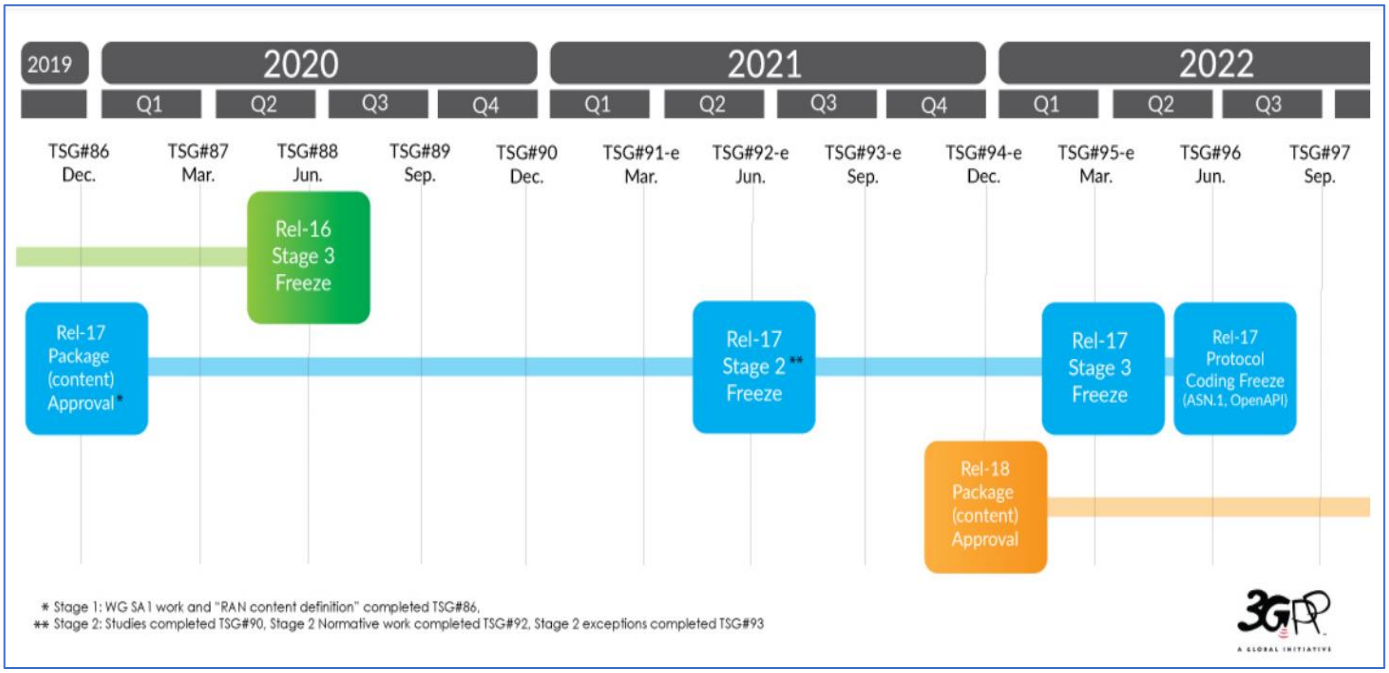
TSG#95-e (March 2022) saw the completion of the Rel-17 functional freeze.

Some important Rel-17 projects were:

- Sidelink enhancements,
- Reduced capability (Redap) NR devices,
- NR operation extended to 71GHz,
- Further enhancements on MIMO for NR,
- NR over Non terrestrial Networks (NTN),
- IoT over NTN,
- UE power saving enhancements for NR,
- Enhancements to Integrated Access and Backhaul for NR,
- Enhancement of RAN slicing for NR,
- RF requirements enhancement for NR FR1,
- RF requirements for NR FR2,
- Coverage and positioning enhancements,
- NR and slicing QoE,
- Enhanced support of non-public networks,
- Support for uncrewed aerial systems,
- Support for edge computing in 5GC,
- Proximity-based services in 5GS,
- Access traffic steering, switch and splitting (ATSSS),
- Network automation for 5G (Phase 2).

FIGURE 31: DIFFERENT RELEASE FEATURES COMPARISON

Image Source: 3GPP.org



* Stage 1: WG SA1 work and "RAN content definition" completed TSG#86,
** Stage 2: Studies completed TSG#90, Stage 2 Normative work completed TSG#92, Stage 2 exceptions completed TSG#93

FIGURE 30: RELEASE 17 TIMELINE

Note: The release 18 discussion is ongoing, and we hope to freeze it sometime in Q2 2024.

2.2 5G Use Cases (23)

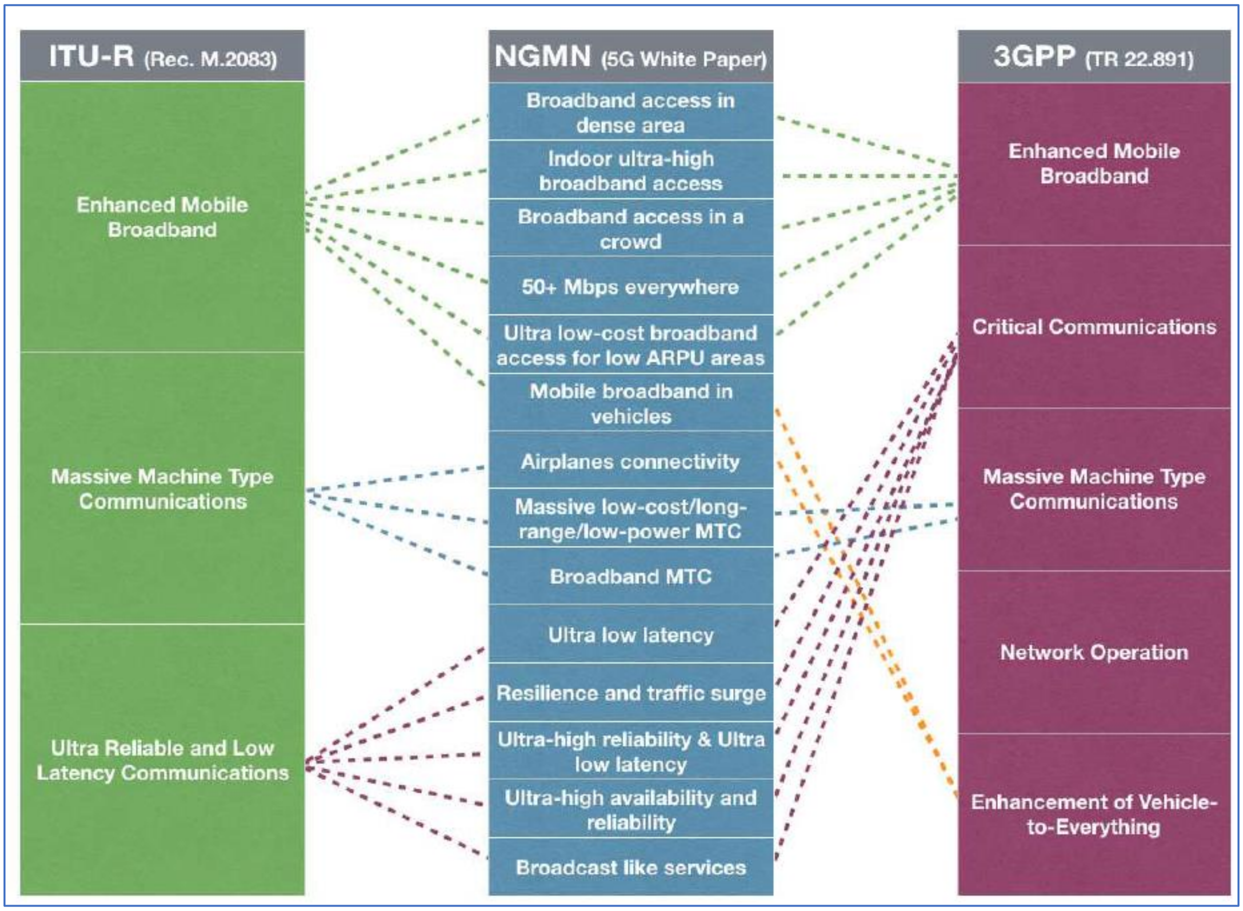

ITU-R (Rec. M.2083)
Enhanced Mobile Broadband
Massive Machine Type Communications
Ultra Reliable and Low Latency Communications
NGMN (5G White Paper)
Broadband access in dense area
Indoor ultra-high broadband access
Broadband access in a crowd
50+ Mbps everywhere
Ultra low-cost broadband access for low ARPU areas
Mobile broadband in vehicles
Airplanes connectivity
Massive low-cost/long-range/low-power MTC
Broadband MTC
Ultra low latency
Resilience and traffic surge
Ultra-high reliability & Ultra low latency
Ultra-high availability and reliability
Broadcast like services
3GPP (TR 22.891)
Enhanced Mobile Broadband
Critical Communications
Massive Machine Type Communications
Network Operation
Enhancement of Vehicle-to-Everything

FIGURE 32: COMPARISONS OF DIVERSE 5G SERVICES AMONG ITU-R), NGMN AND 3GPP [19]

According to the network design and architecture description, different Use Cases are defined for 5G Networks. 3GPP defined these Use Cases as part of its New Services and Markets Technology Enabler (SMARTER) project. The objective behind SMARTER was to develop high-level use cases and identify which features and functionalities are required to enable them. The process started in 2015 and resulted in over 70 use cases, initially grouped into five categories, which have been reduced to three. The three sets of Use Cases are as follows.

- Enhanced mobile broadband (eMBB)
- Ultra-reliable low latency communication (URLLC).
- Machine Type Communications (MTC).

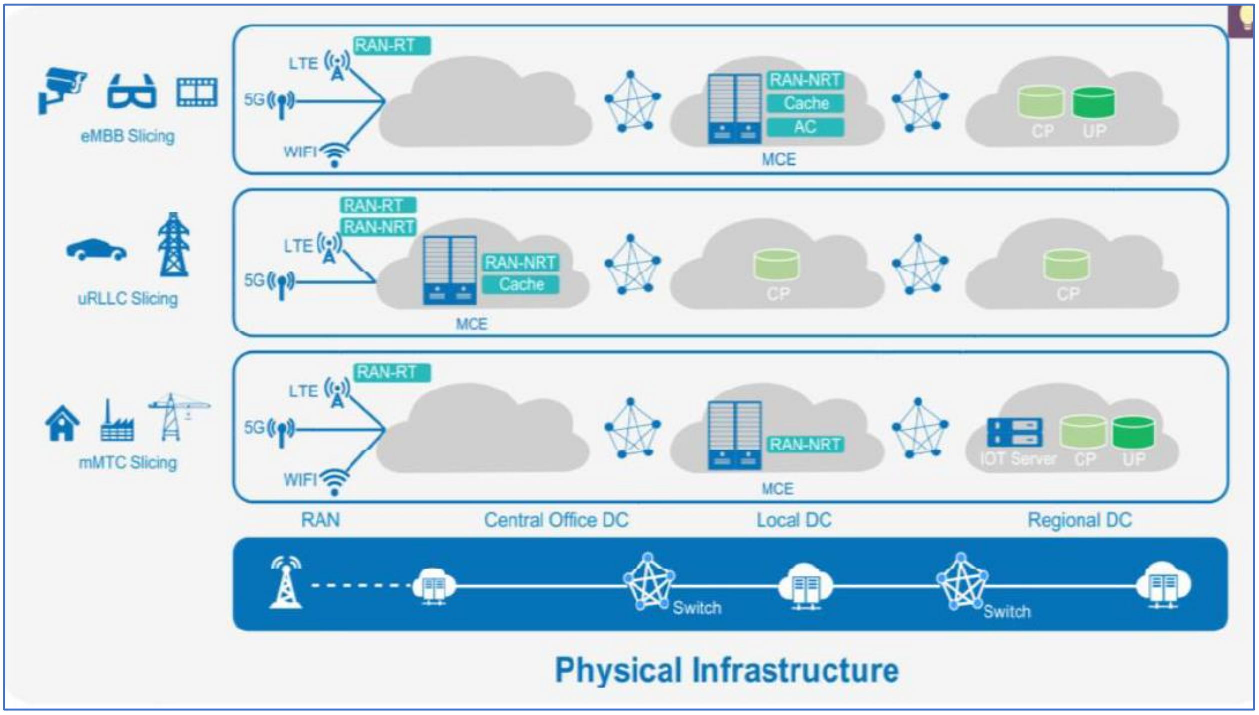
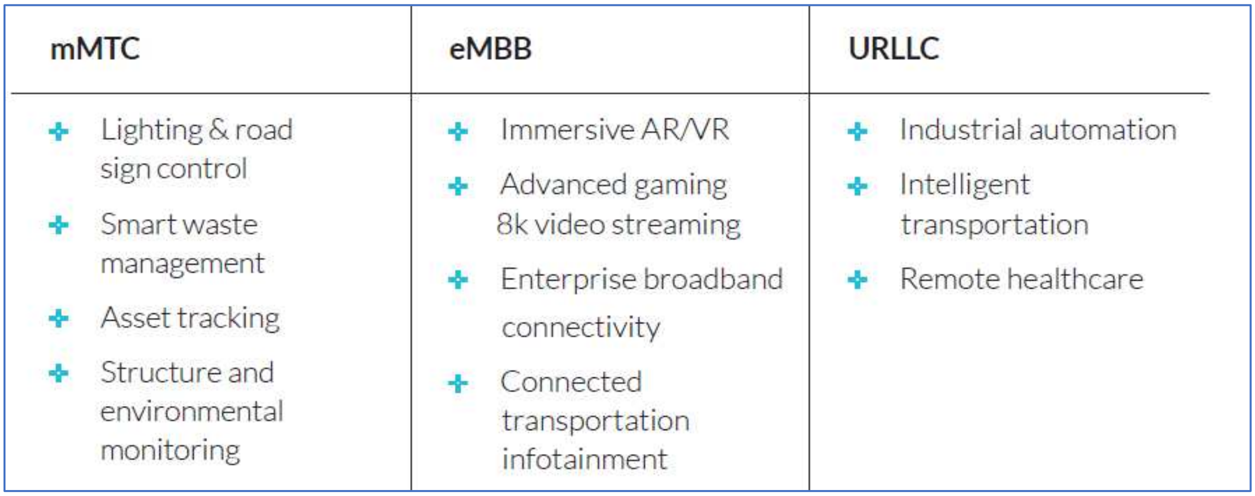


FIGURE 33: PHYSICAL INFRASTRUCTURE [25]

Some sample 3GPP use cases comparison: (24)



mMTC	eMBB	URLLC
✚ Lighting & road sign control ✚ Smart waste management ✚ Asset tracking ✚ Structure and environmental monitoring	✚ Immersive AR/VR ✚ Advanced gaming 8k video streaming ✚ Enterprise broadband connectivity ✚ Connected transportation infotainment	✚ Industrial automation ✚ Intelligent transportation ✚ Remote healthcare

TABLE 3: 3GPP USE CASE COMPARISON

The various features added to each Release address different aspects of these three categories. The specific use cases already in use today or close to arriving are addressed in earlier Releases. In contrast, use cases that are farther in the future are handled in later Releases. It’s all part of the ongoing evolution of 5G.

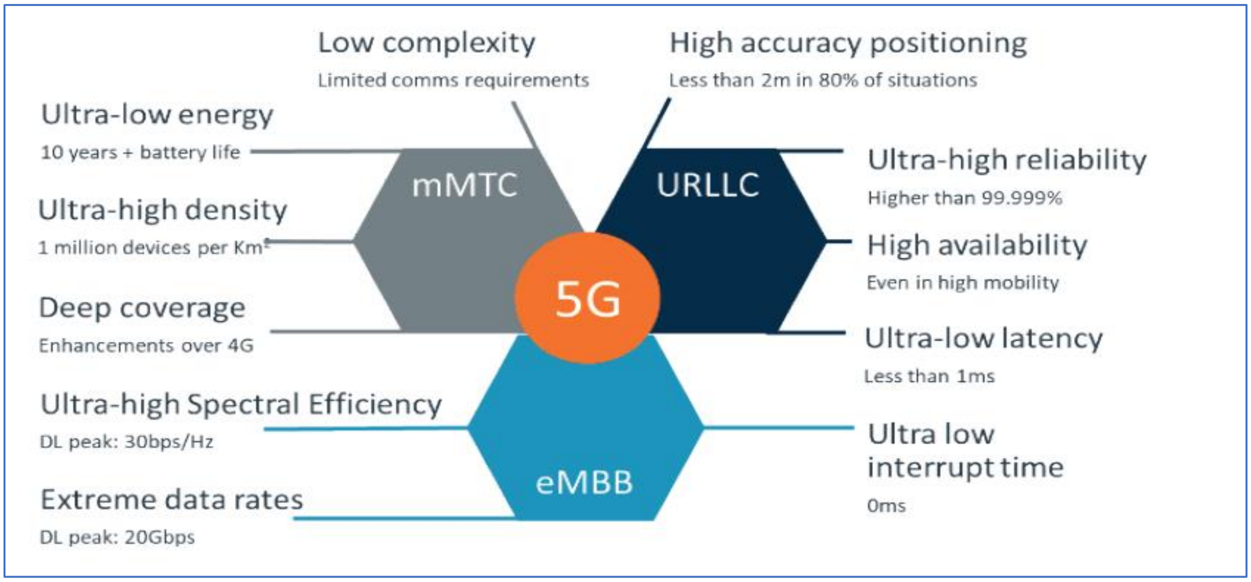


FIGURE 34: 5G ATTRIBUTES FOR DIFFERENT FEATURES [24]

<u>Use cases linked to various scenarios</u>:

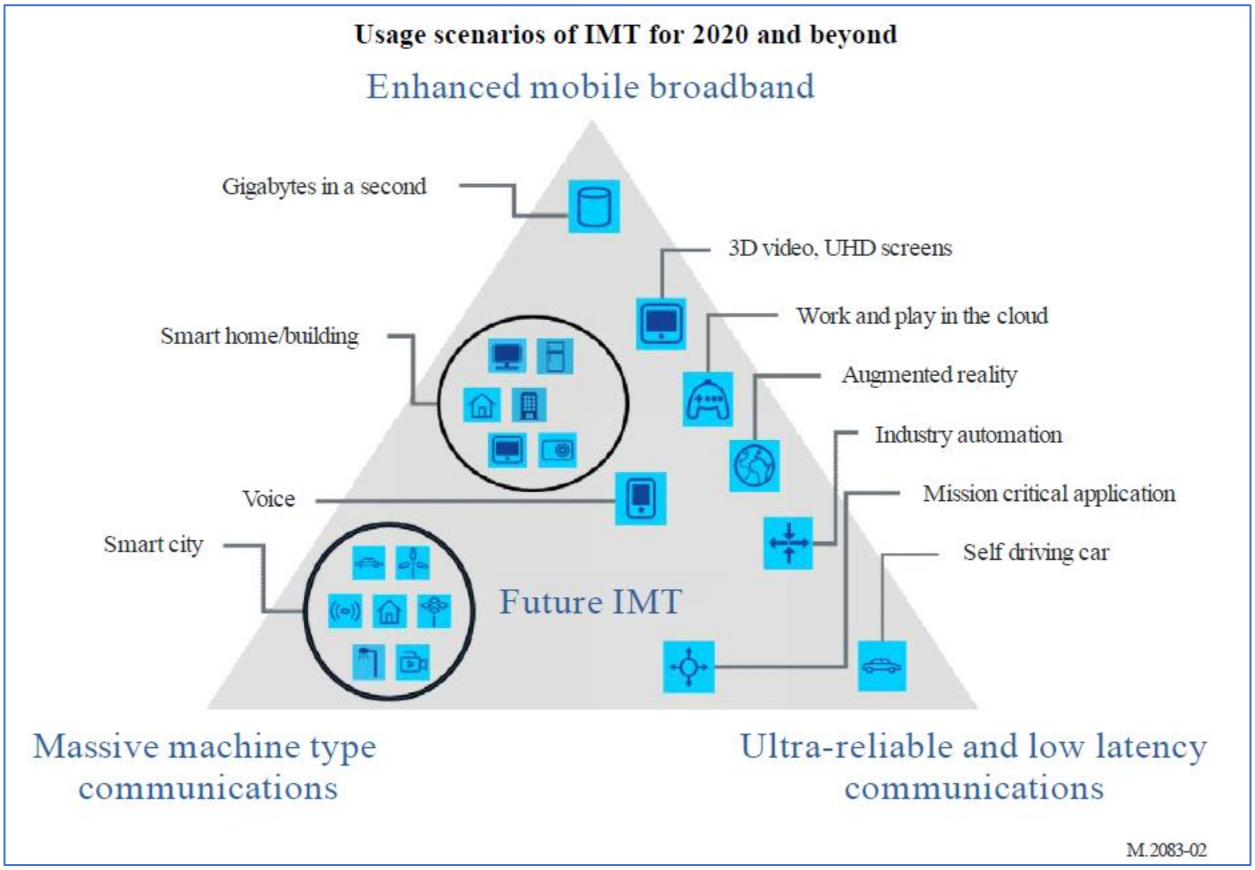

Usage scenarios of IMT for 2020 and beyond
Enhanced mobile broadband
Gigabytes in a second
3D video, UHD screens
Work and play in the cloud
Smart home/building
Augmented reality
Industry automation
Voice
Mission critical application
Smart city
Self driving car
Future IMT
Massive machine type communications
Ultra-reliable and low latency communications
M.2083-02

FIGURE 35: PROJECTION OF INTERRELATED 5G SERVICE TIERS [21]

Image Source: International Telecommunications Union [ITU]

Additional use cases are expected to emerge, which need to be foreseen. For future IMT, flexibility will be necessary to adapt to new use cases with a wide range of requirements. Depending on the circumstances and the different needs in different countries, future IMT systems should be highly modular so that not all features must be implemented in all networks.

2.3 5G Deployment: Non-Standalone (NSA) and Standalone (SA) mode

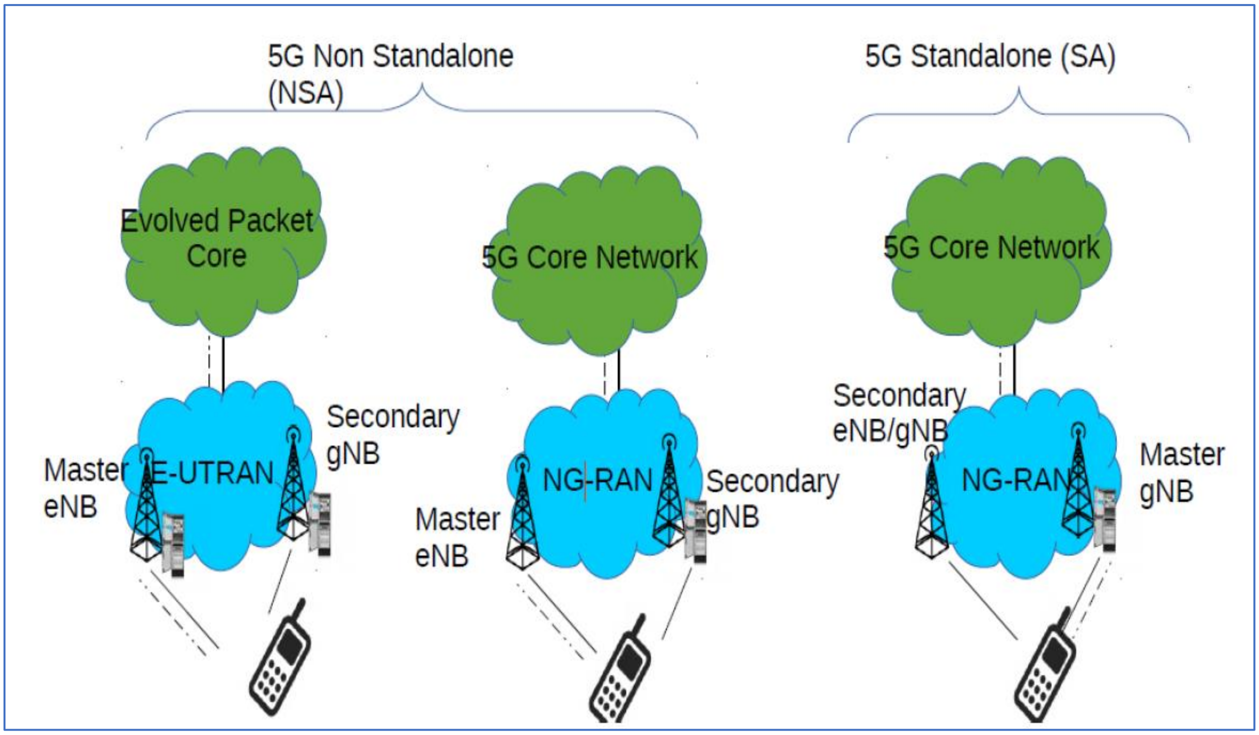


FIGURE 36: NSA AND SA 5G DEPLOYMENT

Two deployment options are defined for 5G:

- The "**Non-Stand Alone**" NSA architecture, where the New Radio (NR) interface and 5G Radio Access Network (AN) are used in conjunction with the existing LTE and EPC infrastructure's Core Network (respectively 4G Radio and 4G Core). Thus, making the NR technology supportable without any network replacement. The 4G services are supported via the configuration, but they allow the capacities offered by the 5G New Radio (lower latency, etc.). Hence, the NSA is also known as "E-UTRA-NR Dual Connectivity (EN-DC)."

- The "**Stand-Alone**" (SA) architecture, where the New Radio NR is connected to the 5G Core Network. Only in this configuration is the complete 5G Phase 1 service support.

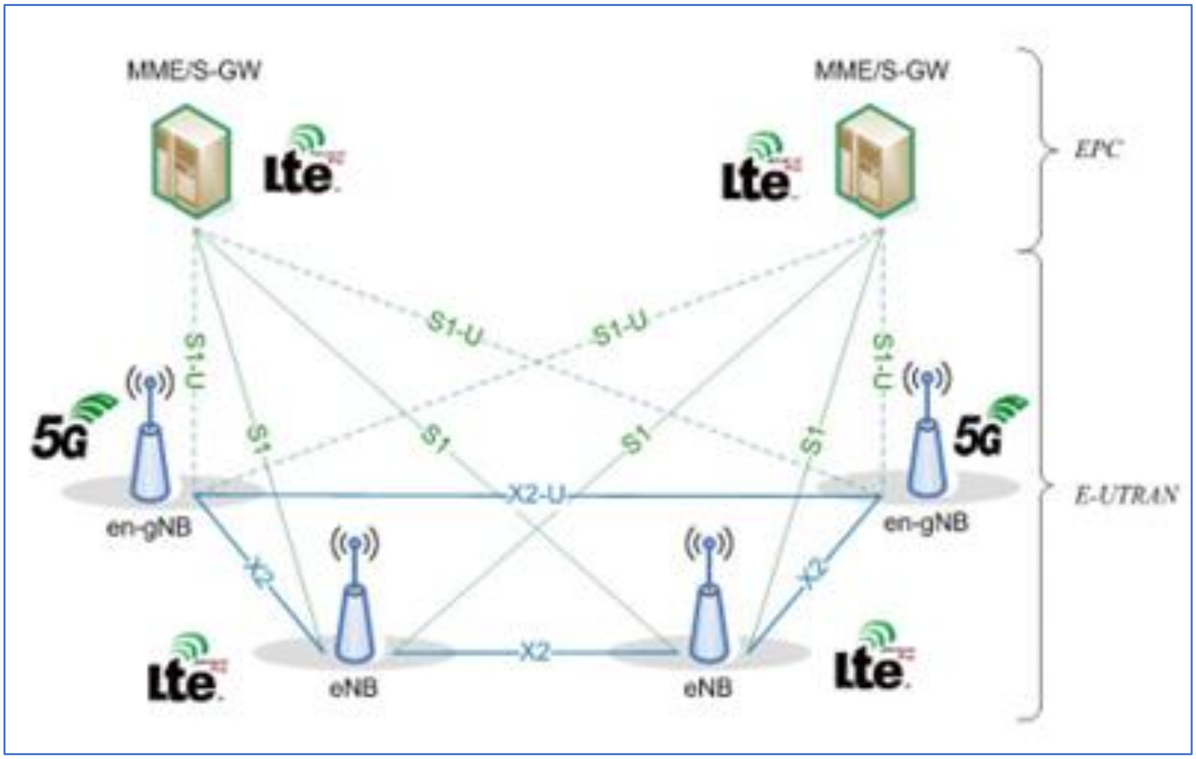


FIGURE 37: THE NSA ARCHITECTURE [26]

Image Source: 3GPP

2.3.1 The NSA Architecture

The NSA architecture is an interim step towards a "complete 5G" deployment, where the 5G Access Network AN is connected to the 4G Core Network. In the NSA architecture, the (5G) NR base station (logical node "en-gNB") relates to the (4G) LTE base station (logical node "eNB") via the X2 interface. The X2 interface was introduced before Release 15 to connect two eNBs. Therefore, release 15 also supports connecting an eNB and en-gNB to sss NSA.

The NSA offers dual connectivity via the 4G AN (E-UTRA) and the 5G AN (NR). It is thus also called "EN-DC" for "E-UTRAN and NR Dual Connectivity."

In EN-DC, the 4G's eNB is the Master Node (MN), while the 5G's en-gNB is the Secondary Node (SN).

2.3.2 The SA Architecture

The standalone (SA) mode of 5G NR denotes using 5G cells for signaling and information transfer as the "full 5G deployment", not needing any part of a 4G network to operate. It involves the new 5G Packet Core architecture instead of depending on the 4G Evolved Packet Core EPC to allow the deployment of 5G without the LTE network.

The NR base stations (logical node "gNB") connect via the Xn interface, and the Access Network (called the "NG-RAN for SA architecture") relates to the 5GC network using the NG interface. It is expected to have lower cost, better efficiency, and assist the development of new use cases.

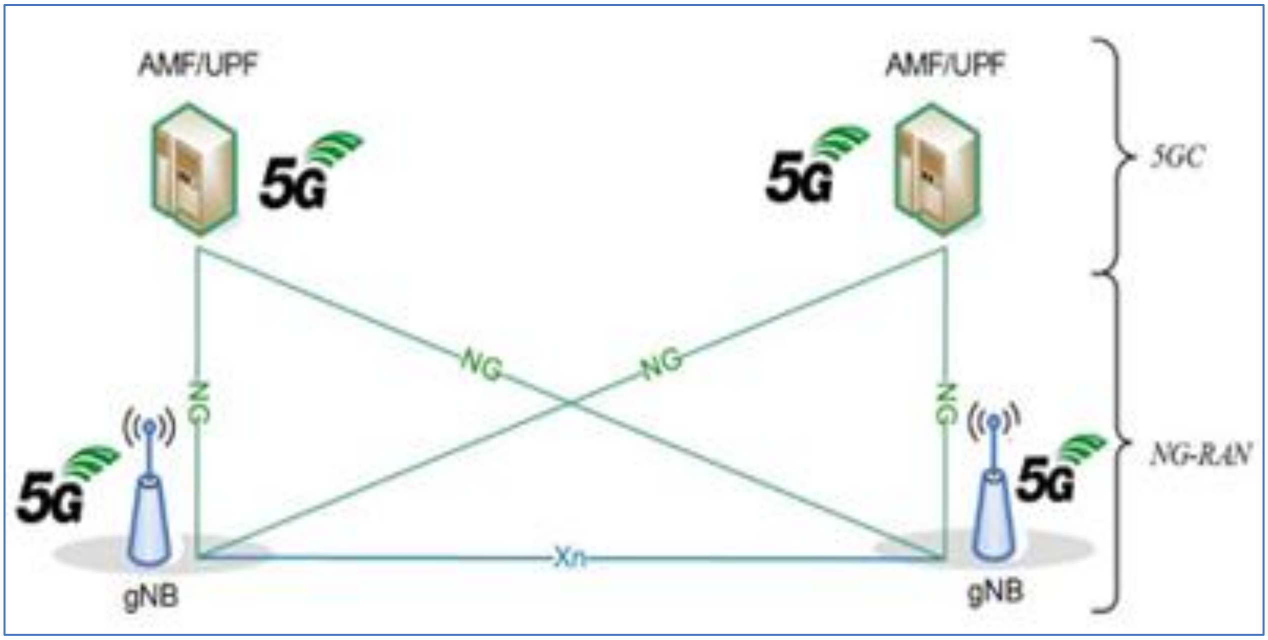


FIGURE 38: THE SA ARCHITECTURE [26]

Image Source: 3GPP

2.4 Key Features of 5G RAN [17], [25], [27]

5G is relatively new and covers an extensive list of features and operating requirements; it can take time to know what to expect from the development effort for those features. 5G NR includes significant advances over LTE, each with specific benefits.

2.4.1 Millimeter Waves

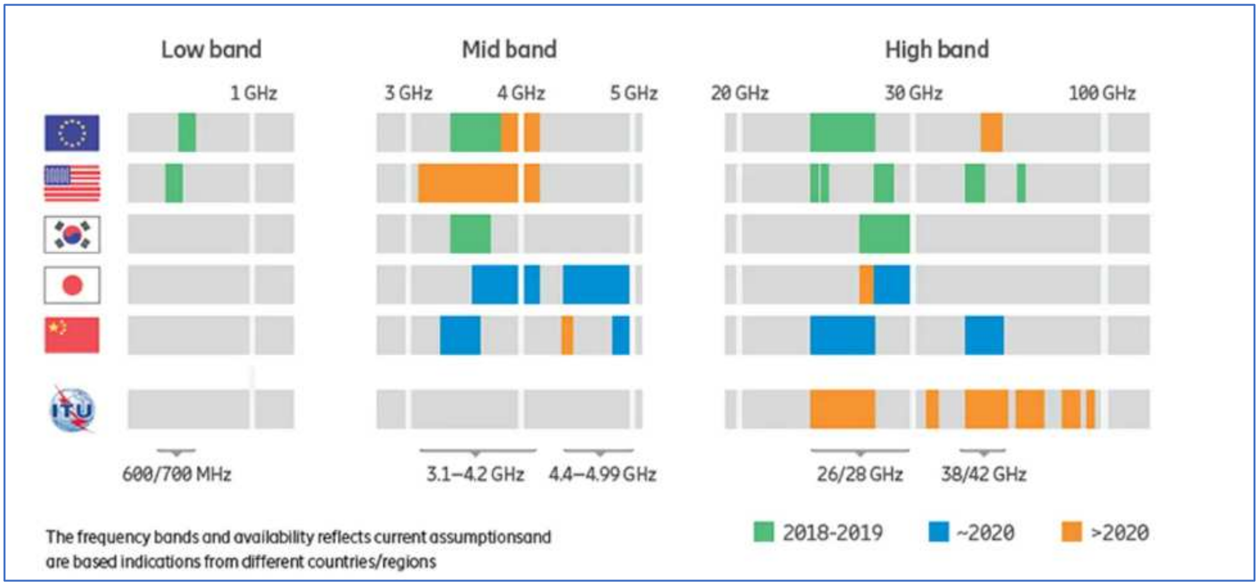


FIGURE 39: AVAILABLE FREQUENCY BANDS CONSIDERING DIFFERENT COUNTRIES

Image Source: Ericsson

Different spectrum bands are available for 5G NR worldwide on different timescales. New Radio (NR) is defined by 3GPP for 5G for the air interface, and the technical specification is divided into two frequency bands, FR1 (below 6 GHz) and FR2 (24–54 GHz).

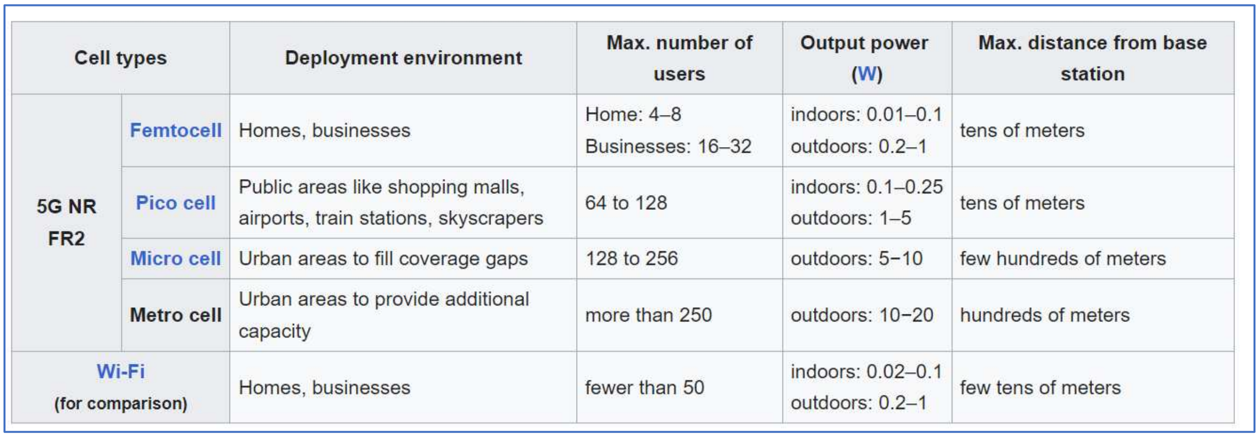
Frequency range 1 [FR1] (< 6 GHz)

It is also categorized as sub-6; the max. Channel-bandwidth introduced for FR1 is 100 MHz because of the scarcity of continuous spectrum in this crowded frequency range.   In this range, the band most widely used for 5G is 3.3–4.2 GHz.

Some parties mentioned the term "mid-band" frequency to refer to the higher part of the frequency range, which was not used in previous generations of telecommunication.

Frequency range 2 [FR 2] (24–71 GHz)

The Channel-bandwidth assigned for FR2 is 50 MHz minimum, and the maximum is 400 MHz, where the two-channel aggregation is provisioned in 3GPP Release 15. The higher frequency supports the higher data transfer speeds. Signaling frequency range with wavelengths between 4 and 12 mm are called millimeter waves**,** which is the most discussed characteristic of the 5G.



Cell types		Deployment environment	Max. number of users	Output power (W)	Max. distance from base station
5G NR FR2	Femtocell	Homes, businesses	Home: 4–8 Businesses: 16–32	indoors: 0.01–0.1 outdoors: 0.2–1	tens of meters
	Pico cell	Public areas like shopping malls, airports, train stations, skyscrapers	64 to 128	indoors: 0.1–0.25 outdoors: 1–5	tens of meters
	Micro cell	Urban areas to fill coverage gaps	128 to 256	outdoors: 5−10	few hundreds of meters
	Metro cell	Urban areas to provide additional capacity	more than 250	outdoors: 10−20	hundreds of meters
Wi-Fi (for comparison)		Homes, businesses	fewer than 50	indoors: 0.02–0.1 outdoors: 0.2–1	few tens of meters

TABLE 4: INFORMATION TABLE FOR 5G NR FR2

Table Source: Wikipedia

5G technology in the 24 GHz range or above using higher frequencies than 4G. As a result, some 5G signals cannot travel considerable distances (over a few hundred meters), unlike 4G or lower frequency 5G signals. It needs to place 5G base stations every few hundred meters for higher frequency bands to fulfill this. Also, the 5G signals of higher frequency cannot penetrate solid objects, such as walls, cars, trees, and humans, because of the nature of higher-frequency electromagnetic waves. Due to that, the 5G networks will likely augment the traditional cellular towers with another modern technology called small cells.

2.4.2 Small Cells [17], [25]

Small cells, which are low-powered cellular radio access nodes, operate in licensed and unlicensed spectrums. It involves a range of 10 meters to a few kilometers. Small cells are critical as radio waves can't travel long distances with higher frequencies. Small cells are small base stations that require minimal power to operate, and these miniature base stations can be placed every 250 meters throughout the cities.

The signal drop is the most problematic aspect, and to prevent signal drop, carriers could install thousands of these base stations where the network is very dense, like cities. Those base stations act like a relay team for receiving signals from other base stations and will send data to users anywhere.

There can be many causes of signal drops, like attenuation and densely populated buildings, weather, and absorption. Traditional cell networks must rely on increasing numbers of base stations for achieving 5G performance will require an even more excellent infrastructure. For small cells, antennas' requirements are also smaller than the traditional antennas if they transmit tiny milometers waves. The small cells make it easier to stick cells on light poles and top buildings. The radically different network structure should provide more targeted and efficient spectrum use so that more bandwidth and frequency can be used for other services. Many stations can reuse frequencies in different serving areas to connect with devices to serve another customer.

2.4.3 Modulation and framing [27]

5G modulation and framing is an advancement from existing ideas but a significant one. Like LTE, 5G NR uses OFDM (orthogonal frequency division multiplexing) as its underlying modulation scheme. OFDM combines multiple subchannels within a channel and is known to be robust against interference and efficient in using frequencies. It is also highly flexible, and different subcarriers can be included to increase a channel capacity or reduce numbers to ensure many lower-power and lower-bandwidth options.

5G NR can select subcarrier spacing from 15kHz to 240kHz, with a maximum of 3300 subcarriers simultaneously on one channel. However, channels could be at most 400MHz wide. The standard is frequency agnostic, and any subcarrier configuration can be used on any band. The mid-and low-band frequencies below 6GHz have markedly different channel and noise characteristics and different maximum bandwidths to the high-band allocations, so they will use 15 to 60kHz channel spacing. In contrast, the high band will use 60 to 120kHz. There are no 5G band allocations between 6GHz and 24.25GHz; the standard allows optimal OFDM configuration to match any future expansion.

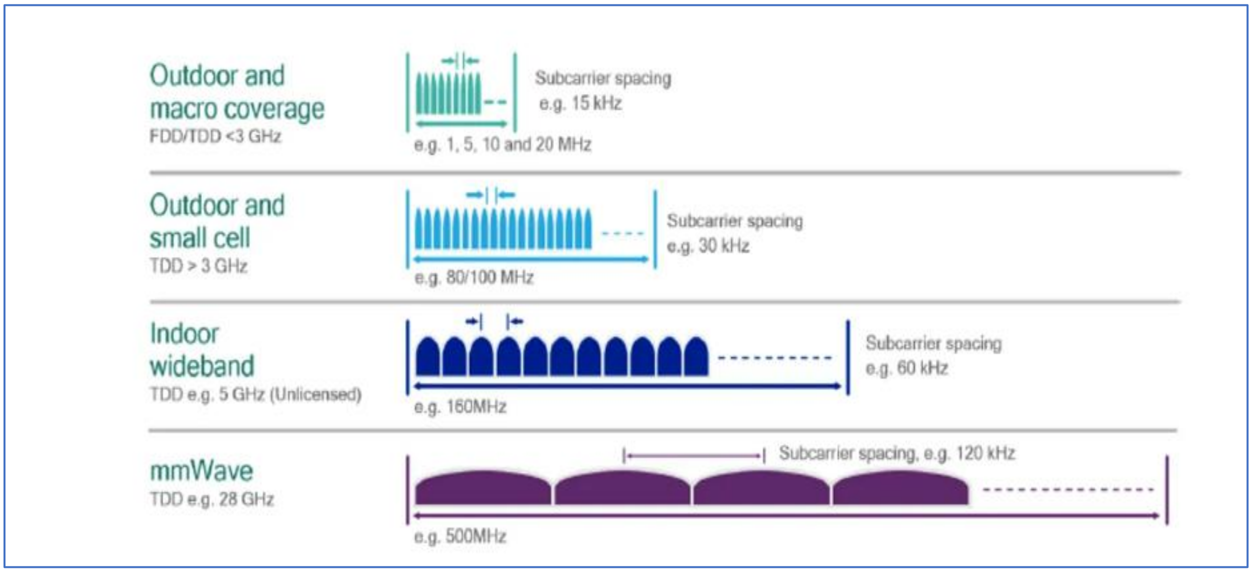

Outdoor and macro coverage
FDD/TDD <3 GHz
Subcarrier spacing e.g. 15 kHz
e.g. 1, 5, 10 and 20 MHz
Outdoor and small cell
TDD > 3 GHz
Subcarrier spacing e.g. 30 kHz
e.g. 80/100 MHz
Indoor wideband
TDD e.g. 5 GHz (Unlicensed)
Subcarrier spacing e.g. 60 kHz
e.g. 160MHz
mmWave
TDD e.g. 28 GHz
Subcarrier spacing, e.g. 120 kHz
e.g. 500MHz

FIGURE 40: 5G ODFM USAGE MODELS, CHANNEL BANDWIDTHS, AND SUBCARRIER SPACING.

Image Source: Qualcomm

All the devices on 5G NR do not support each available bandwidths, which is a change from LTE. Furthermore, 5G NR supports adaptive bandwidth, forcing devices to move to a low-bandwidth and low-power configuration. It only allows higher bandwidths whenever required. This creates the opportunity for low-average power devices that still deliver high performance. The 5G NR specification denotes these different configurations as 'bandwidth parts,' Theoretically, a device can simultaneously support multiple bandwidth parts on the same channel.

Each data is divided into ten milliseconds frames within a subchannel, subdivided into ten 1ms subframes. Each frame consists of 10 ten subframes of 1ms duration. One set of frames for the uplink (0-4) and one for the downlink (5-9) on a carrier.

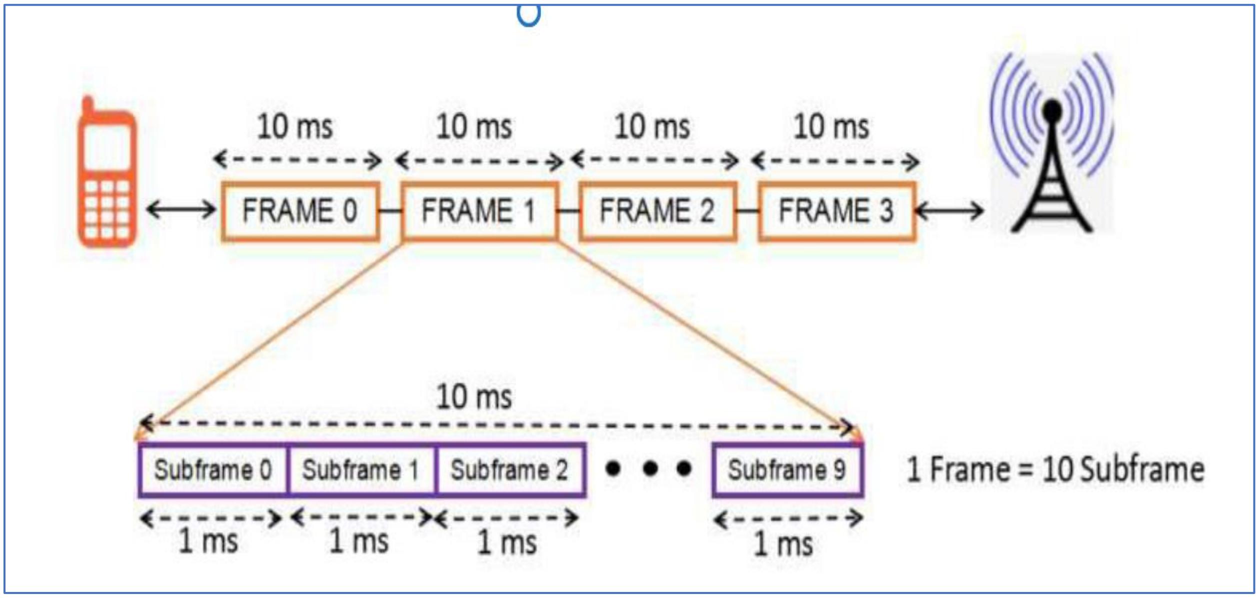


FIGURE 41: FRAME STRUCTURE 5G NR [28]

Those subframes are divided into slots of 14 OFDM symbols separately. Thus, wider bandwidth subchannels include more OFDM symbols/sec; each slot becomes shorter, but the basic frame structure remains the same. At the lowest 15kHz subcarrier spacing, the frames are identical to LTE, simplifying compatibility.

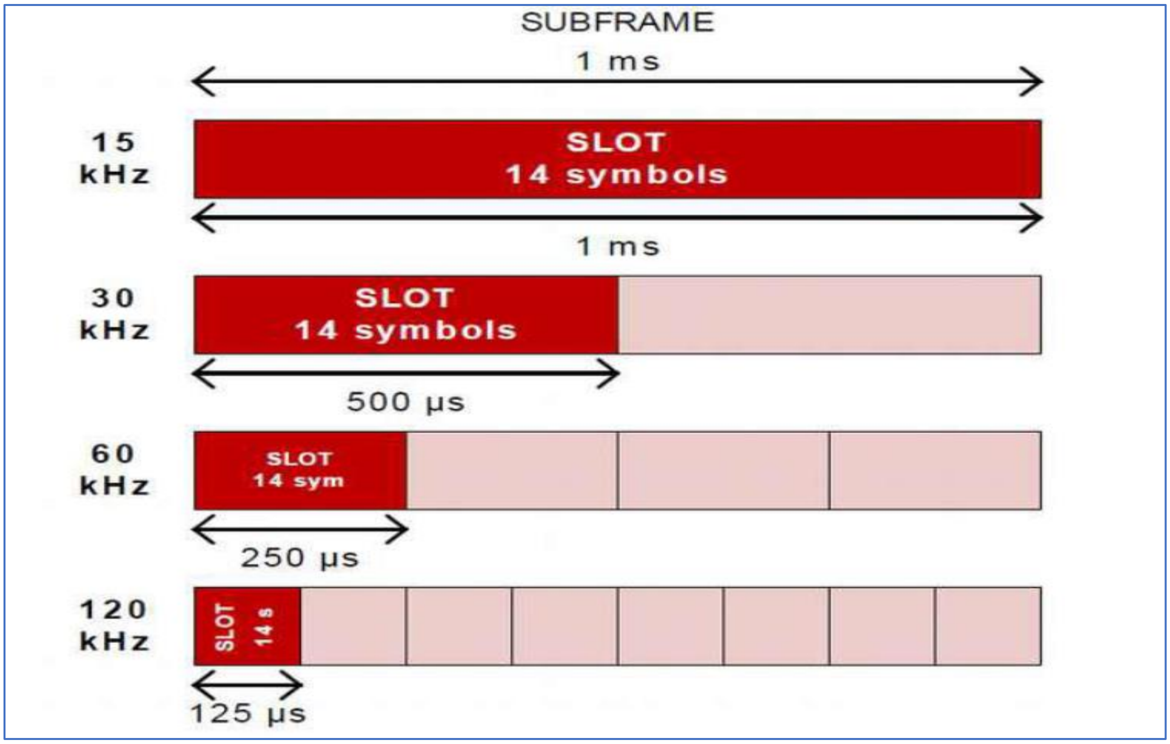


FIGURE 42: RELATIONSHIP BETWEEN SLOT AND SUBFRAME

LTE allocates bandwidth to different devices by slot, but 5G NR has a mechanism called 'mini slots' for transmission to start within a slot effectively. This is particularly useful for the high bands, which can have giant OFDM symbols. Thus, a relatively short message improves both channel reuse and latency. Another potential advantage is that 5G expands to an unlicensed spectrum, usually with a rule to prevent interference 'listen before use.' If a channel occurs quietly, the ability to start a transmission without waiting for a slot boundary decreases the chance of another device grabbing the channel.

5G NR follows the low-latency adaptations to start data transmission after a channel is granted or restriction for processing delay. This is accomplished in the higher network layers by modifying header structures so that processing can begin without the complete packet information. The radio receives essential information from reference and downlink control signals at the physical layer instead of deriving it from the symbol stream.

2.4.4 Massive MIMO [25], [29]

MIMO systems use multiple antennas at a wireless communication system's transmitter and receiver ends. Without changing the bandwidth requirements for multiplexing, spatial dimension is used in multiple antennas.

One of the advantages of MIMO is its flexibility. It can improve multiple connections' reliability, capacity, and spectral efficiency (bits per second per hertz). The maximum amount of error-free digital data can be transmitted over a communication channel that can be calculated below formula by the Shannon-Hartley theorem and Shannon-Hartley law:

- $C \approx W.n.log2(1+SNR)$ [Shannon-Hartely Law]

$C = Capacity$
$W = Spectrum$
$n = Antenna$

- **Capacity = Channels *BW* log2(1 + S/N)** [Shannon-Hartely Theorem]

S/N = Signal-to-noise ratio

Massive MIMO (multiple-input and multiple-output) antennas increase capacity density and sector throughput using large numbers of antennas. There are two specifications: Single User MIMO and Multi-user MIMO (MU-MIMO).

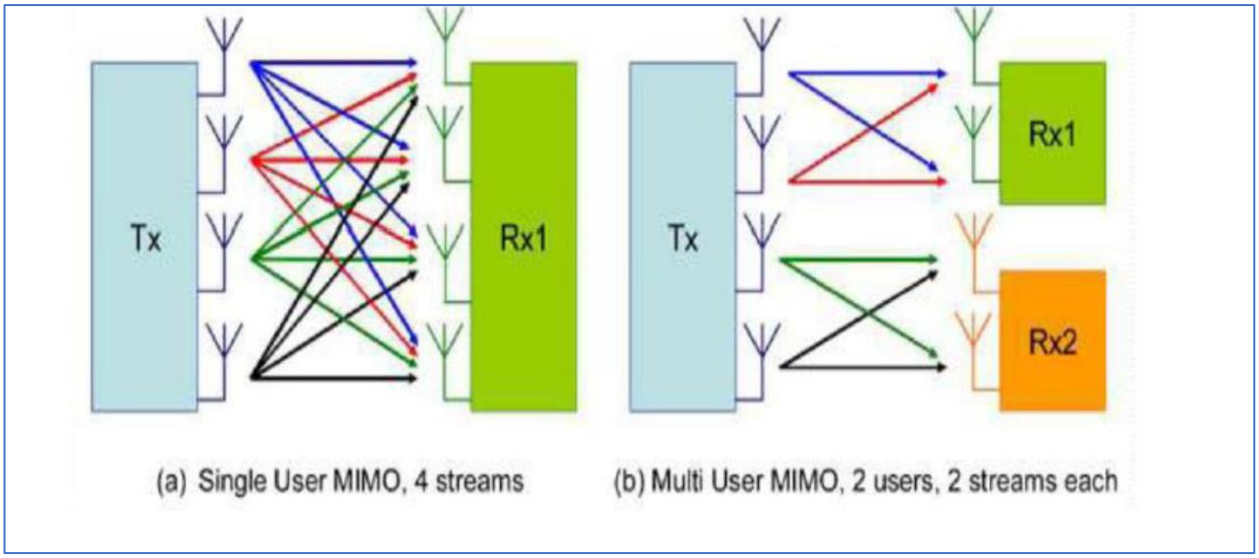

Tx
Rx1
Tx
Rx1
Rx2
(a) Single User MIMO, 4 streams
(b) Multi User MIMO, 2 users, 2 streams each

FIGURE 43: SU-MIMO AND MU-MIMO

SU-MIMO vs. MU-MIMO

In Single User MIMO, both the transceiver and receiver have multiple antennas. As a result, various data streams are transmitted simultaneously using the same resources (time/frequency) to double or quadruple the peak throughput.

In MU-MIMO, multiple data streams are transmitted simultaneously by a base station using the same resources (time/frequency), one per receiver. Hence there is an increase in the total throughput.

In general, more antennas equal better performance. But more antennas also require more extensive arrays that draw more power. In addition, in some places, service providers deploy radio links with very tight constraints, so finding the right solution means weighing trade-offs. For example, the performance gain is often be worth it for in-building coverage; however, for outdoor or street-level coverage, maybe not.

Under Massive MIMO, streams are sent from all the antennas to all the users simultaneously, within the same time-frequency resource block. The different streams are designed to minimize interference between signals from other users.

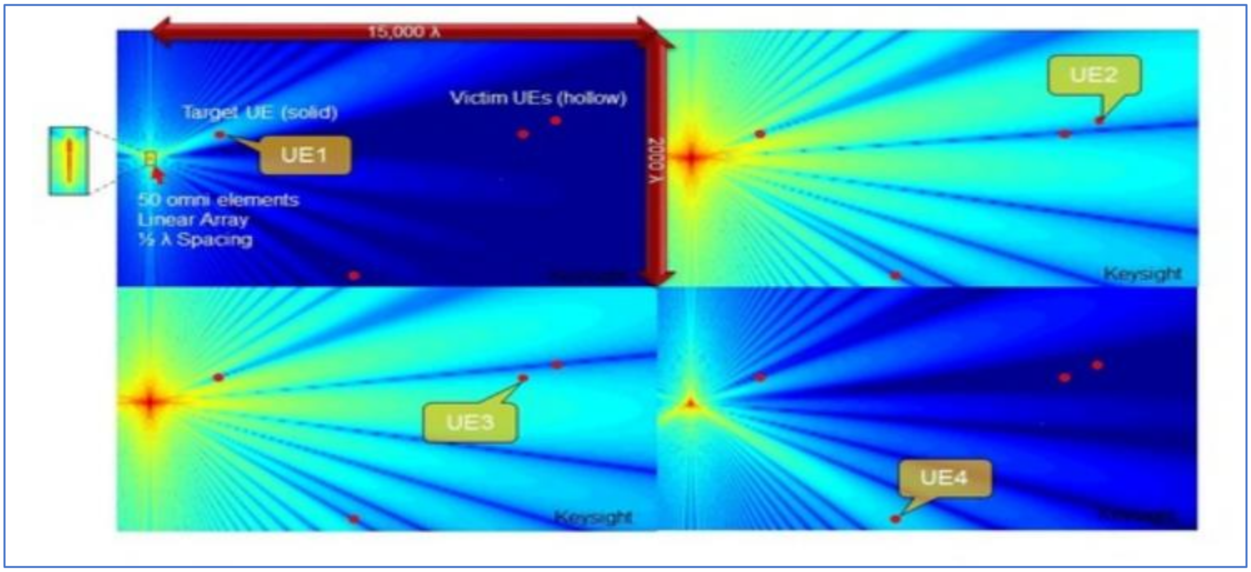


FIGURE 44: VISUALIZING A MASSIVE MIMO BASE STATION TRANSMISSION TO FOUR USERS WITHIN THE SAME RESOURCE BLOCK. [29]

The above fig. 44 shows how a linear antenna array with omnidirectional elements would transmit to four users using the same resource block. Because each user's terminal sends pilot tones to the base station, it knows what the channel between itself and each of them looks like and adjusts the simultaneously transmitted signals for the best aggregate result.

The image at the top left of the graphic above shows that the base station has designed a relatively low power signal for user 1; since they are close by, this ensures the signal does not overlap with any other users. However, at the top right, the signal for user two needs to be optimized, so much power is being wasted to deliver the same signal to user two as to user 1.

The image at the bottom left shows the signal for user 3, which has been designed, so it produces nulls at the locations of the other users. The fourth image, bottom right, shows the radiated signal for user 4, similar to using one but uses more transmission power to cover the extra distance from the base station.

The critical function of the algorithm that designs each signal is to drive enough signal to reach the specific user as efficiently as possible while creating nulls at the other users' positions. It should also ensure that when all the radiation patterns are added together, the only signal arriving at each user's location is its own.

2.4.5 Beamforming

5G New Radio NR has a much more advanced concept of beamforming than LTE. Beamforming is a cellular base station-based traffic-signaling system that initiates the most efficient data-delivery route to a user. It also reduces interference for nearby users in the process. 5G NR extends this to control channels while increasing the overall precision and adaptability for operation under different conditions. This is accomplished by combining elements in an antenna array so that signals at some angles encounter constructive interference while others experience destructive interference. To achieve spatial selectivity, beamforming can be used at both the transmitting and receiving ends.

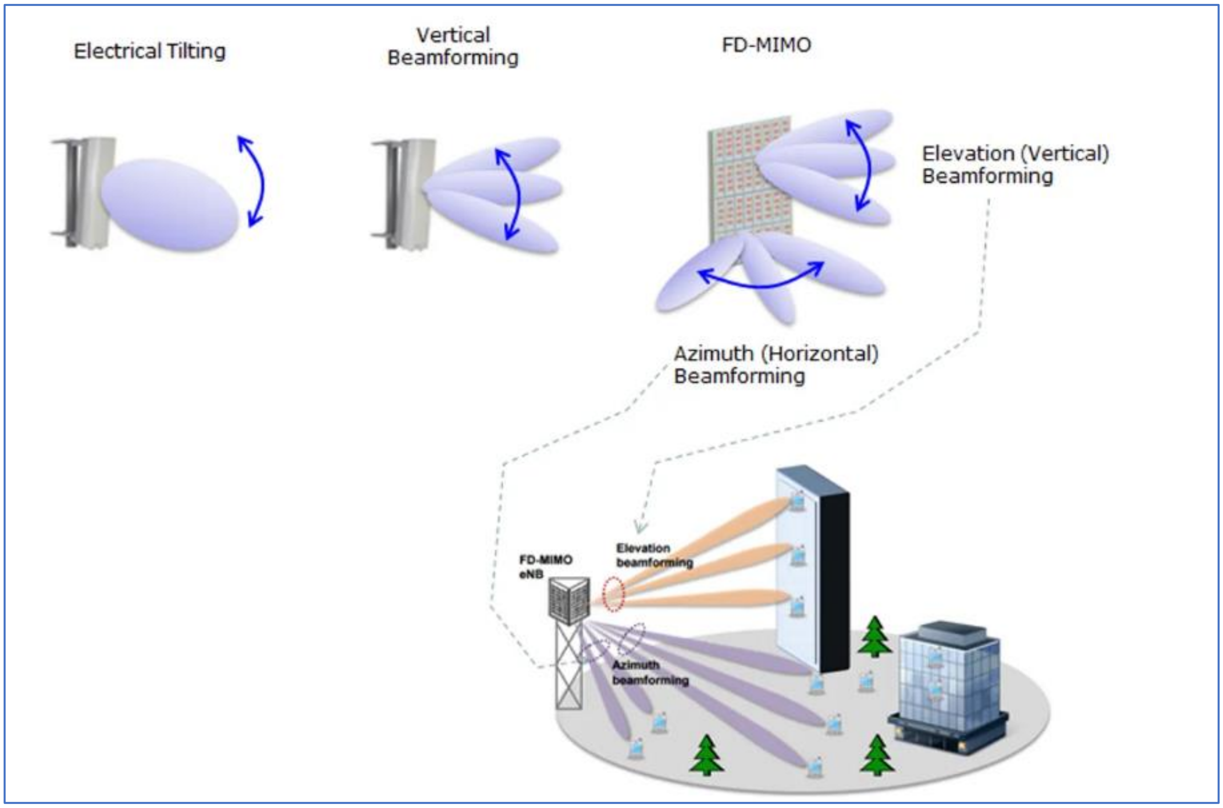


FIGURE 45: WITH FD-MIMO, THE ANTENNA SYSTEM FORMS BEAMS IN BOTH HORIZONTAL AND VERTICAL DIRECTIONS. [27]

Image Source: Sharetechnote.com

Beamforming can be implemented in diverse ways in 5G networks. At the high bands, beamforming will primarily increase range by energy focus. Beamforming will be a crucial part of MIMO for the mid and low bands below 6GHz, where the attenuation is less of a problem. Beamforming helps massive MIMO arrays that will make more efficient use of the spectrum and around them. In telecommunication, there is always the problem of interference on a network in 5G; the interference also causes issues in overcoming this problem.

Beamforming can reduce interference when it simultaneously transmits more information from many antennas. First, massive MIMO base stations and their signal processing algorithms plot the best route through the air to the transmission for the user. That will help to reduce the interference in the transmitted signal. Then, they can initialize individual data packets in many directions and bounce them from buildings and other objects in a precisely coordinated pattern. The packet's movements and arrival time can allow many users and antennas to exchange much more information at once by beamforming on a massive MIMO array. As the millimeter waves have problems with absorption and distraction from buildings, the beamforming addresses how it can help by focusing a signal on a concentrated beam that points only in a user's direction rather than broadcasting in many directions simultaneously. This approach will strengthen the signal chances to arrive intact and reduce interference for everyone else.

SECTION 3. SURGERY OF 5G TECHNOLOGY

3.1 5G Reference Architectures

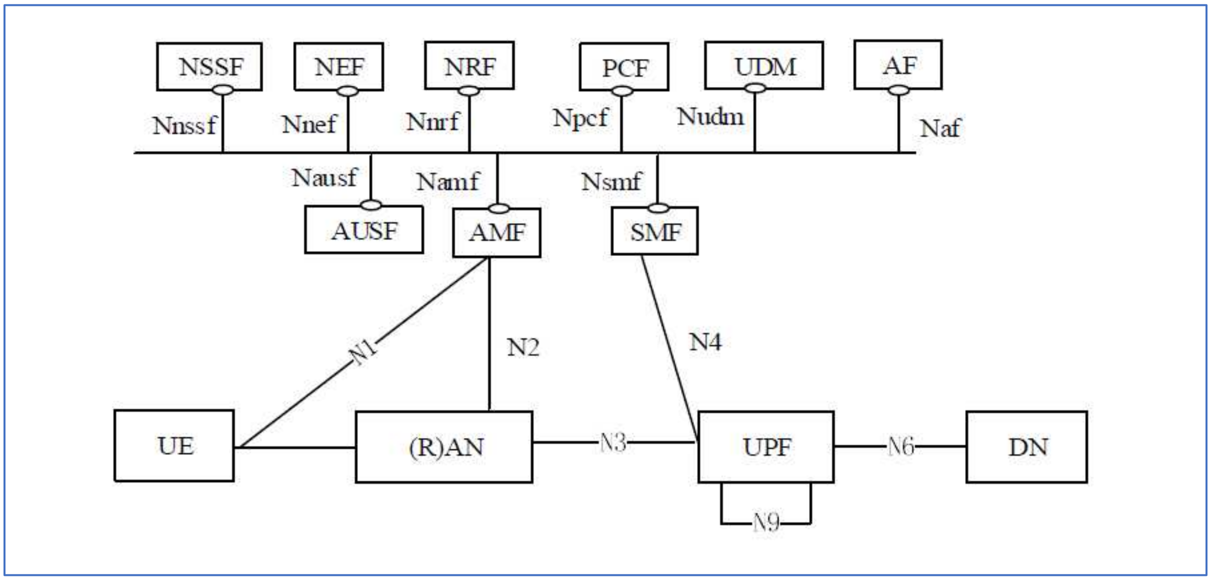


FIGURE 46: SYSTEM ARCHITECTURE FOR THE 5G SYSTEM (5GS) (TS 23.501) [26]

The figure above shows the User Plane, the Network Functions (NFs), and the elements transporting user data at the bottom level. In contrast, the upper part of the figure shows all the essential NFs within the signaling plane.

Network nodes and their functions [30]:

3.1.1 Next-Gen Node Base Station (gNB)

Following functionalities:

- Radio Transmission/Reception
- Digital Signal Processing
- Encryption and data compression
- Process Access Stratum Signaling
- Relay Non-Access Stratum signaling to Core.
- Radio Resource Management
- Communication with Core Network and nearby base stations

3.1.2 Core Access and Mobility Management Function (AMF)

Following functionalities:

- Mobility Management
- Registration management.
- Connection management.
- Reachability management.
- Termination of RAN control plane interface (N2) and NAS (N1) ciphering and integrity protection.
- Lawful intercept
- Access Authentication/Authorization

3.1.3 User Plane Function (UPF)

Following functionalities:

- Packet routing & forwarding.
- Packet inspection.
- Policy rule enforcement.
- Lawful intercept (User Plane).
- Traffic reporting and accounting
- QoS management for the user plane.
- Anchors point for Intra/Inter-RAT mobility.
- Transport level packet marking in the uplink/downlink.

3.1.4 Session Management Control Function (SMF)

Following functionalities:

- Session Management
- Allocation of IP for UE & management
- Control and Selection of User Plane function
- Termination of interfaces (Policy control)
- Lawful intercept
- Termination of Session Management
- Downlink Data Notification
- Roaming functionality
- Charging data collection
- Support of charging interface

3.1.5 Data Network (DN)

Provides operator services, Internet access, or other services.

3.1.6 Authentication Server Function (AUSF)

Perform the UE authentication process.

3.1.7 Unified Data Management (UDM)

Following functionalities:

- Access authorization based on subscription data.
- De-concealment of privacy-protected subscription identifier (SUCI).
- Subscription and SMS management.
- Generation of 3GPP, AKA Authentication Credentials.
- Include Authentication Credential Repository and Processing Function (ARPF).

3.1.8 Policy Control Function (PCF)

Following functionalities:

- Accesses subscription information
- Policy rules to control plane function to enforce them.

3.1.9 Application Function (AF)

Following functionalities:

- Requests dynamic policies and charging control.
- Communication with the core network to request a packet flow.
- IMS Node ordering voice call

3.1.10 NRF (Network Repository Function)

Following functionalities:

- Maintains profiles for Network Functions
- Receive Network Function Discovery Request
- Location and Identification of stored data/information.

3.1.11 NSSF (Network Slice Selection Function)

Following functionalities:

- Selection of set for Network Slice instances serving the UE.
- Determining the Allowed/Configured NSSAI

3.1.12 NEF (Network Exposure Function)

Following functionalities:

- Exposes the capabilities of 5G core network functions to an external AF.
- Acts as the middleman for information exchanged between 5GC and AF.
- The received information is stored as structured data.

3.1.13 Service-Based Interfaces

The system architecture includes the following Service-based interfaces:
Namf: Service-based interface exhibited by AMF.
Nsmf: Service-based interface displayed by SMF.
Nnef: Service-based interface shown by NEF.
Npcf: Service-based interface displayed by PCF.
Nudm: Service-based interface displayed by UDM.
Naf: Service-based interface displayed by AF.
Nnrf: Service-based interface displayed by NRF.
Nnssf: Service-based interface displayed by NSSF.
Nausf: Service-based interface displayed by AUSF.

3.1.14 Reference Points

The System Architecture contains the following reference points:
N1: A Reference point between the UE and the AMF.
N2: A Reference point between the RAN and the AMF.
N3: A Reference point between the RAN and the UPF.
N4: A Reference point between the SMF and the UPF.
N6: A Reference point between the UPF and a Data Network.

3.2 5G Network Slicing [25], [32-33]

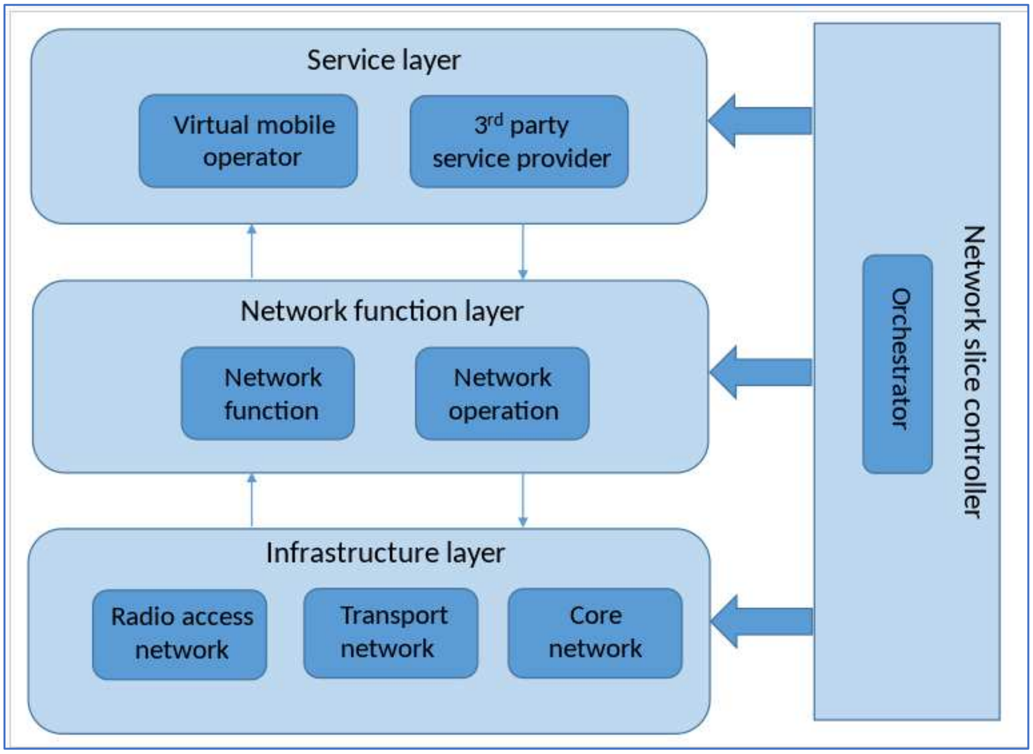


FIGURE 47: GENERIC 5G NETWORK SLICING FRAMEWORK [32]

Image Source: Wikipedia

5G network slicing is an architecture of the mobile network that enables multiplexing virtualized and independent logical networks established on the same physical network infrastructure. Each slice is an isolated end-to-end network tailored to fulfill the diverse requirements of a particular application request.

For this reason, this technology confirms a central role in supporting 5G mobile networks designed to embrace many services with very different service level requirements-SLR. The service-oriented view of the network leverages the concepts of network function virtualization (NFV and software-defined networking (SDN) that permits the deployment of scalable and flexible network slices on top of a shared network infrastructure.

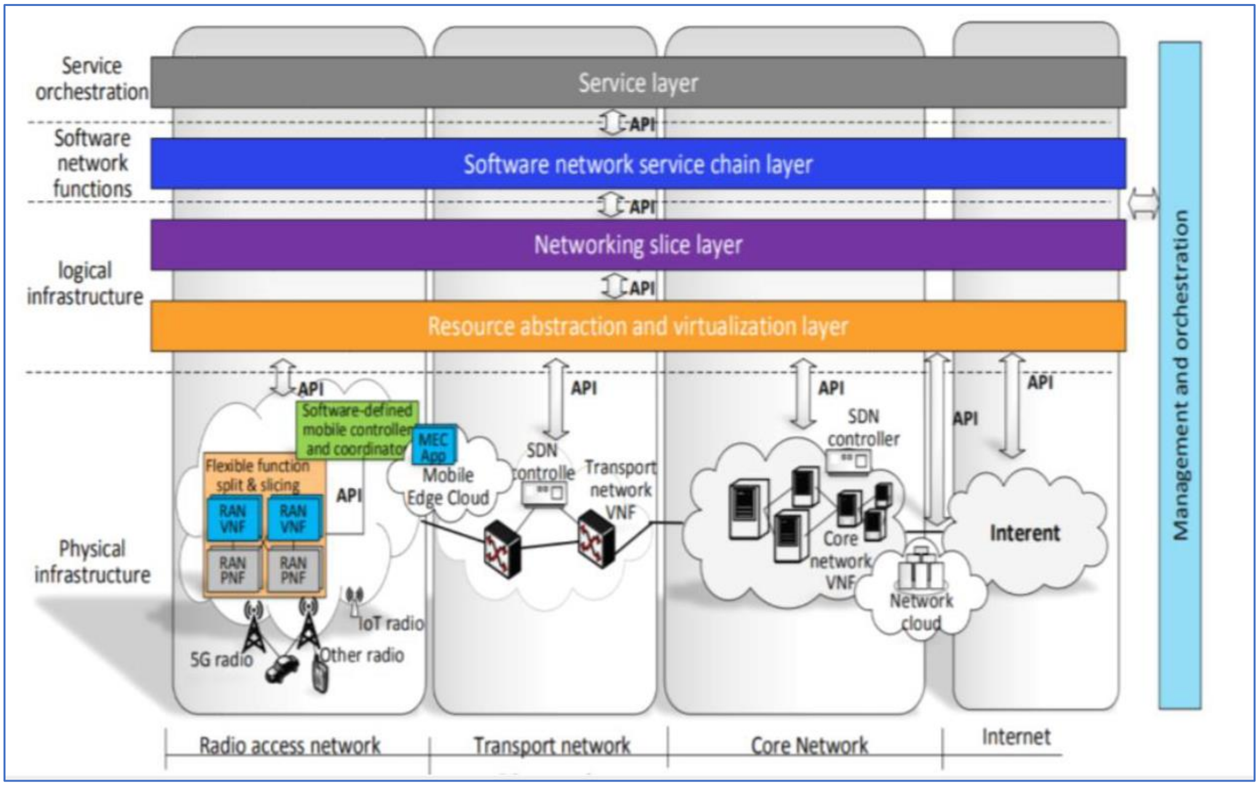


FIGURE 48: 5G NETWORK SEGMENT [25]

For a critical IoT use case, the characteristic of network slices is low latency, high bandwidth, and ultra-reliability, whereas for a massive IoT use case, higher latency and lower bandwidth. Furthermore, the convergence of OSS and BSS ensures automated business and operational processes to manage the network slices and maximize revenues efficiently. With programmable and flexible 5G networks, including advanced AI (Artificial Intelligence) & Service Level Agreement (SLA) driven orchestration, the required business network functions can be created with flexibility, quickly deployed, and automatically managed throughout the life cycle.

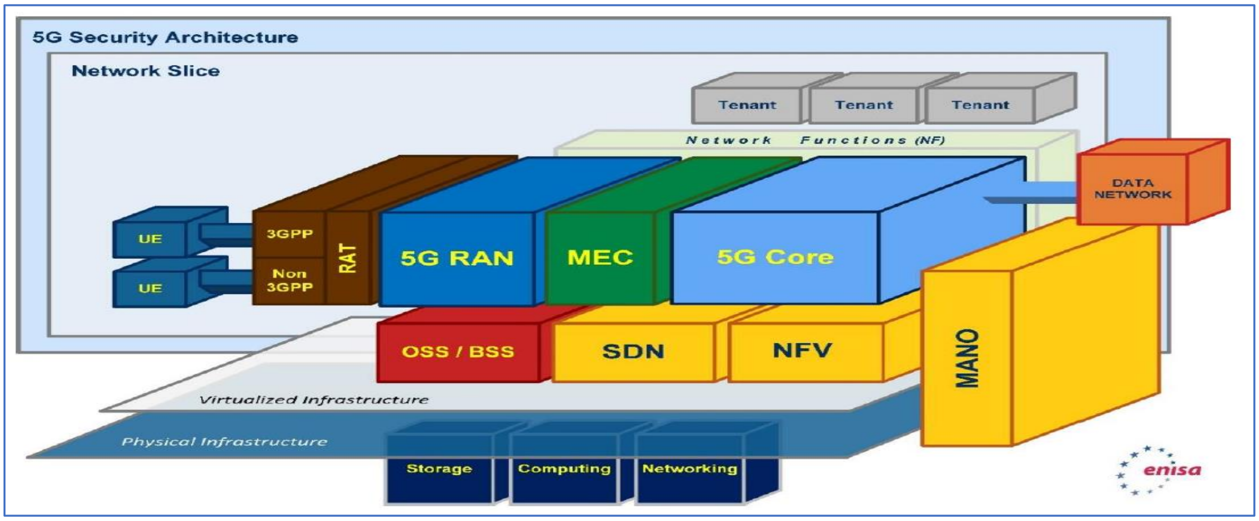


FIGURE 49: 5G HIGH-LEVEL TECHNICAL ARCHITECTURE

Image Source: ENISA

A network slice can be devoted to one enterprise customer or shared by multiple tenants. For example, a slice may consist of dedicated radio, transport, and core resources, including a reliable user plane function at the edge. Another slice shares radio & transport resources between tenants but provides dedicated core network functions per tenant. Finally, a network slice comprises accurate and shared resources, e.g., processing power, storage, and bandwidth, and has isolation from the other network slices.

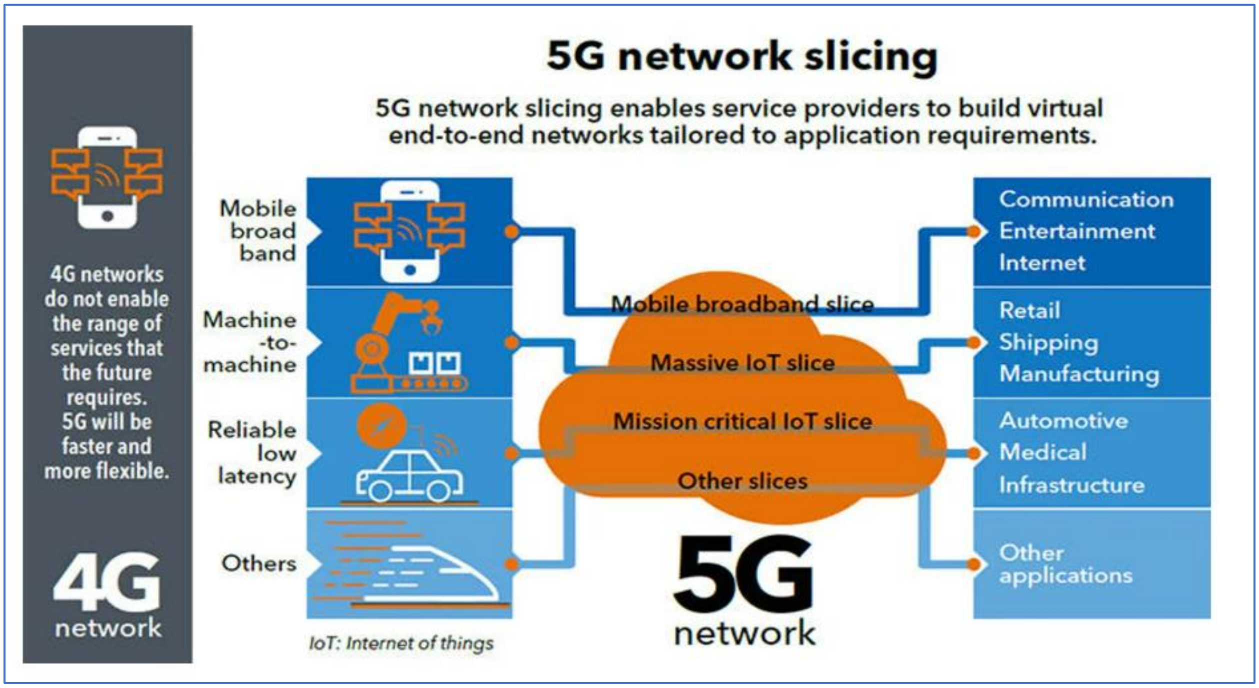


FIGURE 50: 5G NETWORK SLICING

IMAGE SOURCE: ITUNEWS

Network slicing adds an extra dimension to the NFV domain and allows multiple network functions to operate simultaneously on a shared infrastructure. This allows a single network operator to manage multiple virtual networks (which will likely require different latency, throughput, availability, etc.), each assigned to an additional 5G function and, therefore, to enable the vast array of 5G use cases.

Different 5G verticals have different focuses, meaning network slicing will be part of the fundamental design of any 5G network. This model allows for optimizing every aspect of a system, reducing costs, maximizing resources, and granting a greater degree of freedom and flexibility to network operators managing numerous networks. With networks ‘sliced’ into independently operated, controlled, and customized sections, operators can trial potential new 5G services easily and bring them to market rapidly.

3.3 5G Radio Protocol Stack [26], [34]

A protocol stack is defined in TS 23.501 ETSI doc [18] for communications between several of these NFs, and secondary ones. A few highlights of some of the main ones are:

3.3.1 Control plane: the UE-to-AMF and UE-to-SMF protocol stack

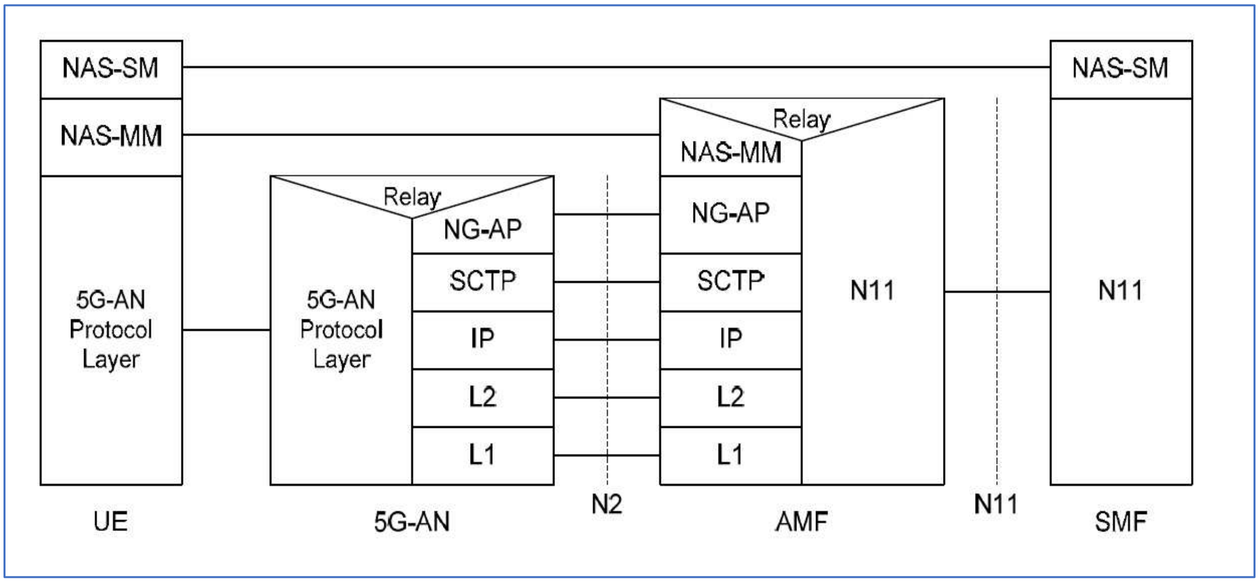


FIGURE 51: CONTROL PLANE PROTOCOL STACK AMONG THE UE, THE 5G-ACCESS NETWORK, THE AMF, AND THE SMF [34]

Source: ETSI TS 23.501, section 8.2

NAS-SM: supports the processing of Session Management between the UE and the SMF. It helps the user plane PDU Session initialization, modification, and release. The AMF transferred and transparent it. It is defined in the 'Non-Access Stratum (NAS) protocol for 5G System (5GS) in Stage 3' (TS 24.501) ETSI doc.

NAS-MM: supports registration management functionality, connection management functionality, and user plane connection activation and deactivation. It is also responsible for ciphering and integrity protection of NAS signaling. 5G NAS protocol is described in TS 24.501 ETSI doc.

5G-AN Protocol layer: This set of protocols and layers depends on the 5G Access Network. For NG-RAN, the radio protocol between the UE and the NG-RAN node (eNodeB and gNodeB, respectively) is categorized in the E-UTRA & E-UTRAN. The reference specifications are TS 36.300 and TS 38.300 ETSI doc.

NG Application Protocol (NG-AP): Application Layer Protocol exists between the 5G-AN node and the AMF. NG-AP is defined in TS 38.413 ETSI doc.

Stream Control Transmission Protocol (SCTP): This protocol confirms the delivery of signaling messages between AMF and 5G-AN node (N2). SCTP is defined in IETF RFC 4960 doc.

Note: 5G-AN and SMF have a direct communication called N2 SM information. This is the subset of NG-AP information, and the AMF transparently transmits between the 5G-AN and the SMF. This includes the NG-AP messages and the N11-related messages.

3.3.2 User plane: the UE-to-AMF and UE-to-SMF protocol stack

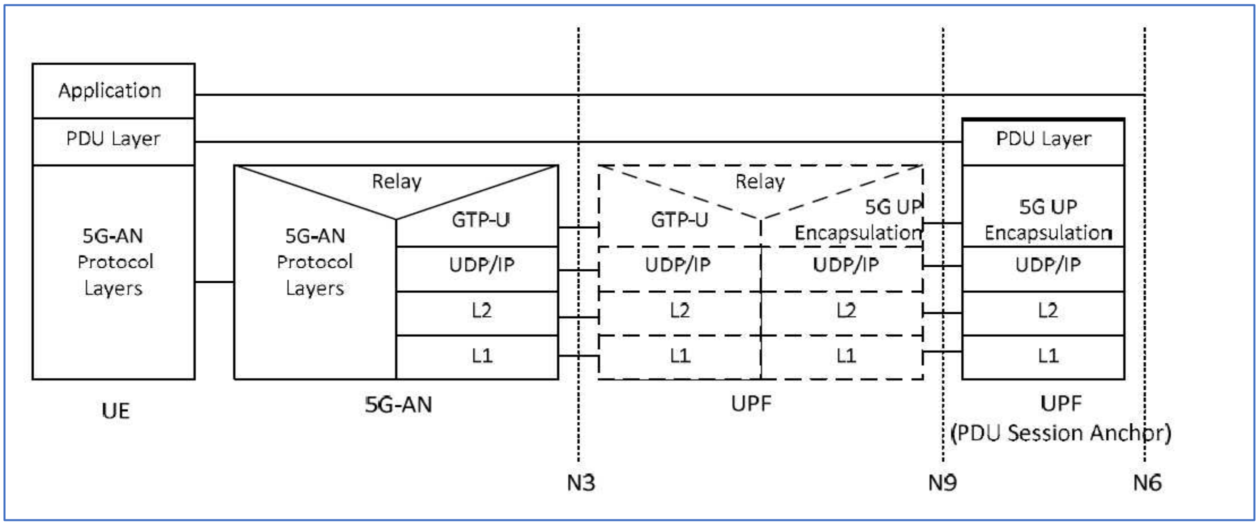


FIGURE 52: USER PLANE PROTOCOL STACK BETWEEN THE UE, THE 5G-AN, AND THE UPF [34]

Source: ETSI TS 23.501

PDU layer: This layer relates to the PDU transported between the UE and the DN via the PDU Session. If the PDU Session carried IPv4 or IPv6 or IPv4v6, it transported to IPv4 or IPv6 or both packets, respectively. When the PDU Session holds Ethernet, it relates to the Ethernet frames.

GPRS Tunnelling Protocol GTP U): This protocol for the user panel handles tunneling user data over N3 (between the 5G-Access Network node and the UPF) and N9 (between different UPFs of the 5G Core) in the backbone network (TS 29.281 ETSI doc. GTP shall encapsulate all end-user PDUs. It ensures encapsulation on a per PDU Session level. This layer also carries the marking associated with a QoS Flow, and this protocol is also used on the N4 interface as defined in TS 29.244 ETSI doc.

5G-AN protocol stack: This set of protocols and layers depends on the Access Network (AN). The 5G-AN is a 3GPP NG-RAN, and these protocols/layers are defined in TS 38.401 ETSI doc. In addition, this radio protocol between the UE and the 5G-Access Network node (eNodeB and gNodeB) is specified in TS 36.300 and TS 38.300 ETSI doc.

UDP/IP: These are the backbone network protocols.

SECTION 4. ANALYZING THE LATEST 5G TECHNOLOGY (RAN)

4.1 RAN Architecture [23]

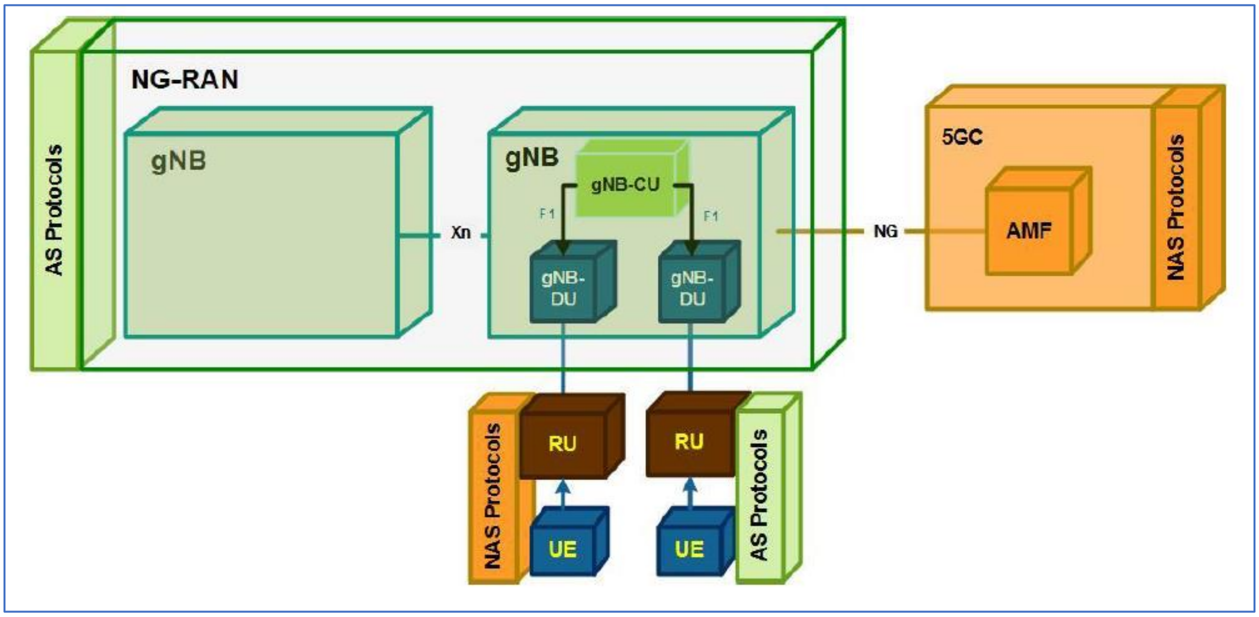


FIGURE 53: RAN ARCHITECTURE [23]

Image source: ENISA

5G-PPP and the latest 3GPP specifications on NG-RAN describe the baseline architecture. It identifies the main innovation, the split of the F1 interface into a Centralized Unit (CU) and Distributed Unit (DU), with a Service Data Adaptation Protocol (SDAP). The SDAP architecture includes a Packet Data Conversion Protocol (PDCP) in the CU and an Air Radio Link Control (ARLC) in the DU. All this is based on IP transport on a TNL/Ethernet network, very similar to the mobile backhaul of today. Another critical aspect of the NG-RAN is the ability to provide small-cell coverage to multiple operators 'as-a-service' in a two-tier architecture. These tiers support the 5G use cases mentioned, providing low latency services and high processing power.

The elements of the RAN architecture are as follows:

4.1.1. User Equipment (UE)

The user equipment is any device users use to communicate within the 5G infrastructure. Besides a SIM, the user equipment may be home appliances (e.g., computers, IoT devices, etc.).

4.1.2 Radio Unit (RU)

It is an element connecting user equipment with the operator network.

4.1.3 gNB

The next generation Node/Base Station is a node outlining the NR user plane and control plane protocol terminations towards the UE-the connection established via the NG interface to the 5GC.

4.1.4 gNB Distributed Unit (gNB-DU)

gNB-DU is a logical node hosting RLC, MAC, and Physical layers of the gNB or en-gNB, and its execution is gNB-DU is a logical node hosting RLC, MAC, and Physical layers of the gNB or en-gNB and its execution is partly controlled by gNB-CU. One gNB-DU supports one or multiple cells. One gNB-DU only supportss one cell. The gNB-DU terminates the F1 interface connected with the gNB-CU.

4.1.5 gNB Central Unit (gNB-CU)

gNB-Central Unit (CU) is a logical node hosting RRC, SDAP, and PDCP protocols of the gNB or RRC and PDCP protocols of the en-gNB that controls the operation of one or more gNB-DUs. The gNB-CU terminates the F1 interface connected with the gNB-DU.

4.1.6 Access and Mobility Management Function (AMF)

AMF is a Network Function (NF). It includes some or all the following functionalities:

- Termination of RAN CP interface.
- Termination of NAS.
- NAS ciphering and integrity protection.
- Registration management.
- Connection management.
- Reachability management.
- Mobility Management.
- Lawful intercept.
- Transport for SM messages between UE and SMF.
- Transparent proxy for routing SM messages.
- Access authentication; access authorization.
- Transport for SMS messages between UE and SMSF; security anchor functionality (SEAF)
- Location services management; transport for Location Services messages between UE and LMF and between RAN and LMF.
- EPS Bearer ID allocation for interworking with EPS:
- UE mobility event notification.

4.1.7 F1

The logical interface with the F1 Application Protocol is defined in ETSI TS 138 473 doc.

4.1.8 Xn

Xn is a network interface between NG-RAN nodes; 3GPP TS 38.420 ETSI doc specifies the Xn interface’s general aspects and principles.

4.1.9 NG interface

ETSI defines the NG interface as an element that logically separates signaling and data transport network.

4.1.10 Non-Access Stratum (NAS)

NAS is a functional layer in the protocol stack between UE and Core Network. NAS protocol for 5G System is defined in 3GPP TS 24.501 ETSI doc.

4.1.11 Access Stratum (AS)

AS is a functional layer in the protocol stack between UE and RAN responsible for transporting data over the wireless connection and managing radio resources.

4.2 NFV Architecture [23], [35-36]

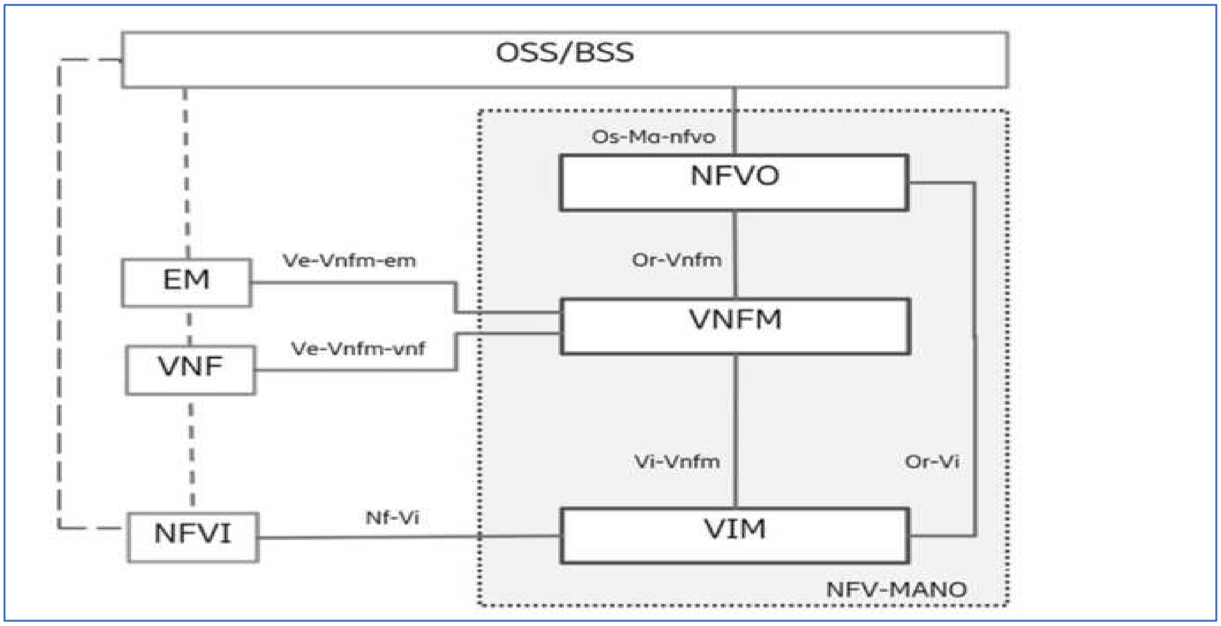


FIGURE 54: GENERIC NFV-MANO ARCHITECTURAL FRAMEWORK

Image Source: ETSI

Network functions virtualization (NFV) is a concept of network architecture that leverages IT virtualization technologies to virtualize entire classes of network node functions. Like building blocks, it connects or chains together to initiate and provide communication services.

This concept relies upon traditional server-virtualization techniques such as those used in enterprise IT. However, a VNF is implemented within one or more virtual machines (VM) or containers running different software and processes. Instead of having custom hardware appliances, it operates on top of commercial off-the-shelf (COTS) high-volume servers, switches, spines, storage equipment, or cloud computing infrastructure. For each building block (network function), thereby, can avoid vendor lock-in status.

NFV introduces a new concept for service providers to accelerate the deployment of new network services to support their revenue and growth plans. It translates to using standard IT virtualization technologies applied to deploying Network Functions, aiming at a faster provision of new network services.

The NFV consists of **three main components:**

1. **Virtualized network functions** (VNFs) are software-implemented network functions that can be employed on a network functions virtualization infrastructure (NFVI).
2. **Network functions virtualization infrastructure** (NFVI) is all the hardware and software components that make the environment where NFVs are deployed. The NFV infrastructure can span several locations. Therefore, the network connectivity between these locations is considered part of the NFV infrastructure.
3. **Network functions, virtualization management, and orchestration** is the architectural framework (NFV-MANO Architectural Framework) and the combination of all functional blocks, data repositories, and reference points and interfaces. These available blocks exchange information to manage and orchestrate VNFs and NFVI.

The building block builds the NFV platform with the NFVI and the NFV-MANO. The NFVI role comprises virtual and physical processing, storage resources, and virtualization software. Its NFV-MANO role consists of VNF and NFVI managers and virtualization software enabling a hardware controller. In addition, the NFV platform implements carrier-grade features to manage and observe the platform components, recover from failures, and provide adequate security.

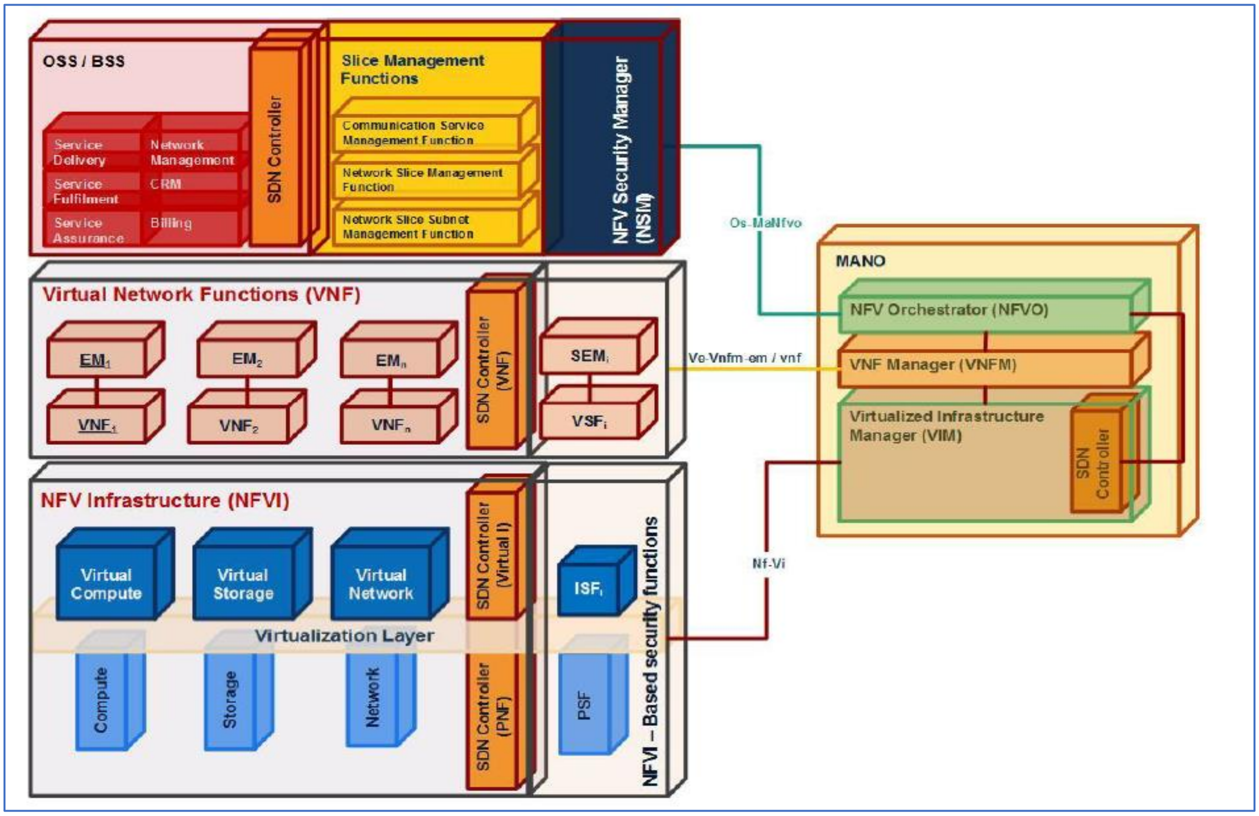


FIGURE 55: NFV ARCHITECTURE ZOOM-IN [23]

Image Source: ENISA

The elements of the NFV architecture are as follows:

4.2.1 Operations Support System/Business Support System (OSS/BSS)

OSS/BSS functions manage and orchestrate systems, including legacy ones. They may have complete end-to-end visibility of services offered by legacy network functions in an operator's network. Processes covered by OSS/BSS include Network Management, Service delivery/fulfillment/assurance, Customer Relationship Management, and Billing.

4.2.2 Virtualised Network Function (VNF)

A VNF is virtualizing a network function in a legacy non-virtualized network. ETSI GS NFV 001 lists use cases and examples of target network functions (NFs) for virtualization. The functional behavior and state of an NF are mainly independent of whether the NF is virtualized or not. The dynamic behavior and the external operational interfaces of a Physical Network Function (PNF) and a VNF are expected to be the same.

4.2.3 Element Management (EM)

Element Management is responsible for the **FCAPS** feature:

- **Fault management** for the network functions provided by the VNF.
- **Configuration management** for the network functions provided by the VNF.
- **Accounting** for the usage of VNF functions.
- Collecting **performance measurement** results for the functions provided by the VNF.
- **Security management** for the VNF functions.

4.2.4 NFV Infrastructure (NFVI)

The NFV Infrastructure corresponds to all hardware and software components that build up the environment in which VNFs are deployed, managed, and executed. The NFV Infrastructure can span several locations, i.e., places where NFVI-PoPs are operated. The connectivity between these locations is part of the NFV Infrastructure. From the VNF's perspective, the virtualization layer and the hardware resources look like a single entity providing the VNF with desired virtualized resources.

4.2.5 Hardware Resources

In NFV, the physical hardware resources include computing, storage, and network that provide processing, storage, and connectivity to VNFs through the virtualization layer (e.g., hypervisor). Computing hardware is assumed to be COTS as opposed to purpose-built hardware. In addition, storage resources can be differentiated between shared network attached storage (NAS) and storage on the server itself. Therefore, computing and storage resources are commonly pooled. Finally, network resources comprise switching functions, e.g., routers and wired or wireless links.

4.2.6 Virtualisation Layer and Virtualised Resources

The virtualization layer abstracts the hardware resources and decouples the VNF software from the underlying hardware, thus ensuring a hardware-independent lifecycle for the VNFs. In short, the virtualization layer is responsible for the following:

- Enabling the software that implements the VNF to use the underlying virtualized infrastructure.
- Providing virtualized resources to the VNF to execute the latter.

4.2.7 Virtualised Infrastructure Manager (VIM)

Virtualized infrastructure management comprises the functionalities used to control and manage the interaction of a VNF with computing, storage, and network resources under its authority, as well as their virtualization. The Virtualised Infrastructure Manager performs resource and operations management according to the list of hardware resources specified in the architecture. Therefore, multiple Virtualised Infrastructure Manager instances may be deployed.

4.2.8 NFV Orchestrator

The NFV Orchestrator orchestrates and manages NFV infrastructure and software resources and realizes network services on NFVI.

4.2.9 VNF Manager

VNF Manager is responsible for VNF lifecycle management (e.g., instantiation, update, query, scaling, termination). Multiple VNF Managers may be deployed; a VNF Manager may be deployed for each VNF, or a VNF Manager may serve various VNFs.

4.2.10 Os-Ma-nfvo

This reference point is used for exchanges between OSS/BSS and NFV Orchestrator and supports the following:

- Network Service Descriptor and VNF package management.
- Network Service instance lifecycle management
- VNF lifecycle management
- Policy management and enforcement for Network Service instances, VNF instances, and NFVI resources
- Querying relevant Network Service instance and VNF instance information from the OSS/BSS.
- Forwarding of events, accounting and usage records, and performance measurement results regarding Network Service instances, VNF instances, and NFVI resources to OSS/BSS, as well as information about the associations between those instances and NFVI resources.

4.2.11Ve-Vnfm-em

This reference point is used for exchanges between EM and VNF Manager and supports the following functions:
*VNF instantiation * VNF instance query * VNF instance update * VNF instance scaling out-in and up-down * VNF instance termination * Forwarding configuration and events from the EM to the VNFM * Forwarding of configuration and events regarding the VNF from the VNFM to the EM.

Note: This reference point is only used if the EM knows about virtualization.

4.2.12 Ve-Vnfm-vnf

This reference point is used for exchanges between VNF and VNF Manager and supports the following:
* VNF instantiation *VNF instance query * VNF instance update * VNF instance scaling out-in and up-down * VNF instance termination * Forwarding of configuration and events from the VNF to the VNFM * Forwarding of configuration, events, etc. regarding VNF, from the VNFM to the VNF * Verification that the VNF is still alive/functional.

4.2.13 NFVI - Virtualised Infrastructure Manager (Nf-Vi)

This reference point is used for the following:

- The specific assignment of virtualized resources in response to resource allocation requests
- Forwarding of virtualized resources state information
- Hardware resource configuration and state information (e.g., events) exchange.

4.2.14 NFV Security Manager (NSM)

NSM is the logical functional block for overall security management, e.g., on behalf of network services. In cooperation with MANO blocks dedicated to managing the virtualized network, the policy-driven NSM is specialized to manage the security of a network service over its entire lifecycle. It covers the following functionalities:

- Security Policy Planning, designs and optimizes security policies for specific targets of protection (e.g., network services).
- Security Policy Enforcement & Validation automates the deployment and supports lifecycle management of security functions as defined in the design phase, then configure security policies on the security functions. In addition, during the lifetime of a network service, the validation and re-configuration/remediation of associated security policies are supported, also in an automated manner.
- NFVI Security Manager (ISM)

4.2.15 NFVI Security Manager (ISM)

NFVI Security Manager is the logical function dedicated to security management in the NFVI layer. It builds and manages the security in NFVI to support NSM requests for managing the security of network services in a higher layer.

4.2.16 Security Element Manager (SEM)

SEM refers to Element Manager managing Security Functions.

4.2.17 Virtual Security Function (VSF)

This element is a type of VNF running on top of NFVI with tailored security functionality (e.g., firewall, IDS/IPS, virtualized security monitoring functions like vFEP and vTap). VSFs are mainly required to protect the other VNFs, which constitute a network service. VSF is managed by either dedicated VNFM or generic VNFM concerning its lifecycle.

4.2.18 NFVI-based Security Function (ISF)

This element is a security function provided by the NFV Infrastructure. It includes virtualized security appliances or software security features (e.g., hypervisor-based firewalls) and hardware-based security appliances/modules/features (e.g., Hardware Security Modules, Crypto Accelerators, or Trusted Platform Modules).

4.2.19 Physical Security Function (PSF)

This element is a conventionally realized security function in the physical part of the hybrid network. Even if a telco, the Network Services run on top) as a whole. PSF is part of the non-virtualized traditional network and is not maintained by the NFVI provider. Hence it is managed by the SEM instead of the VIM.

4.3 SDN Architecture [23], [36-40]

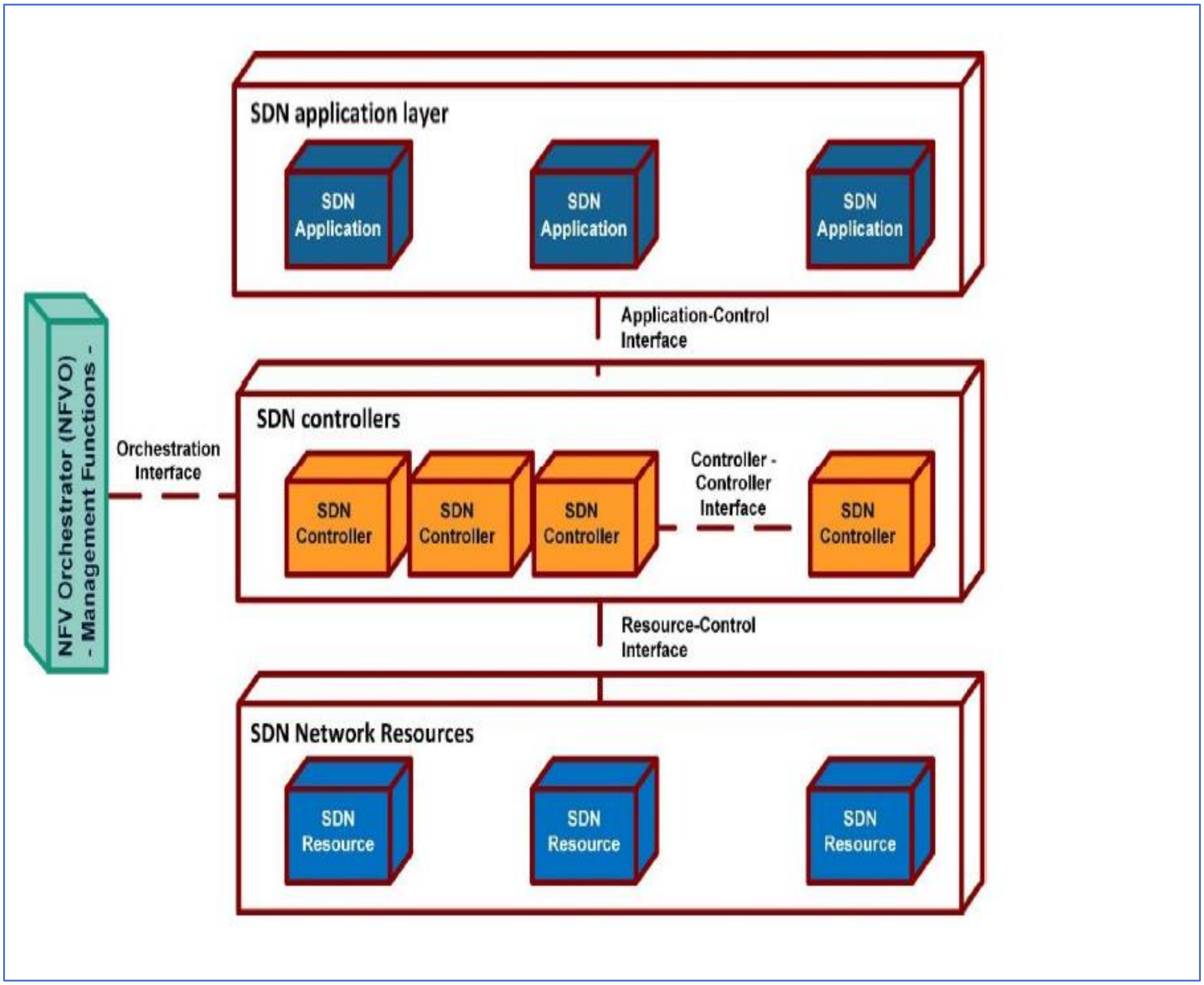


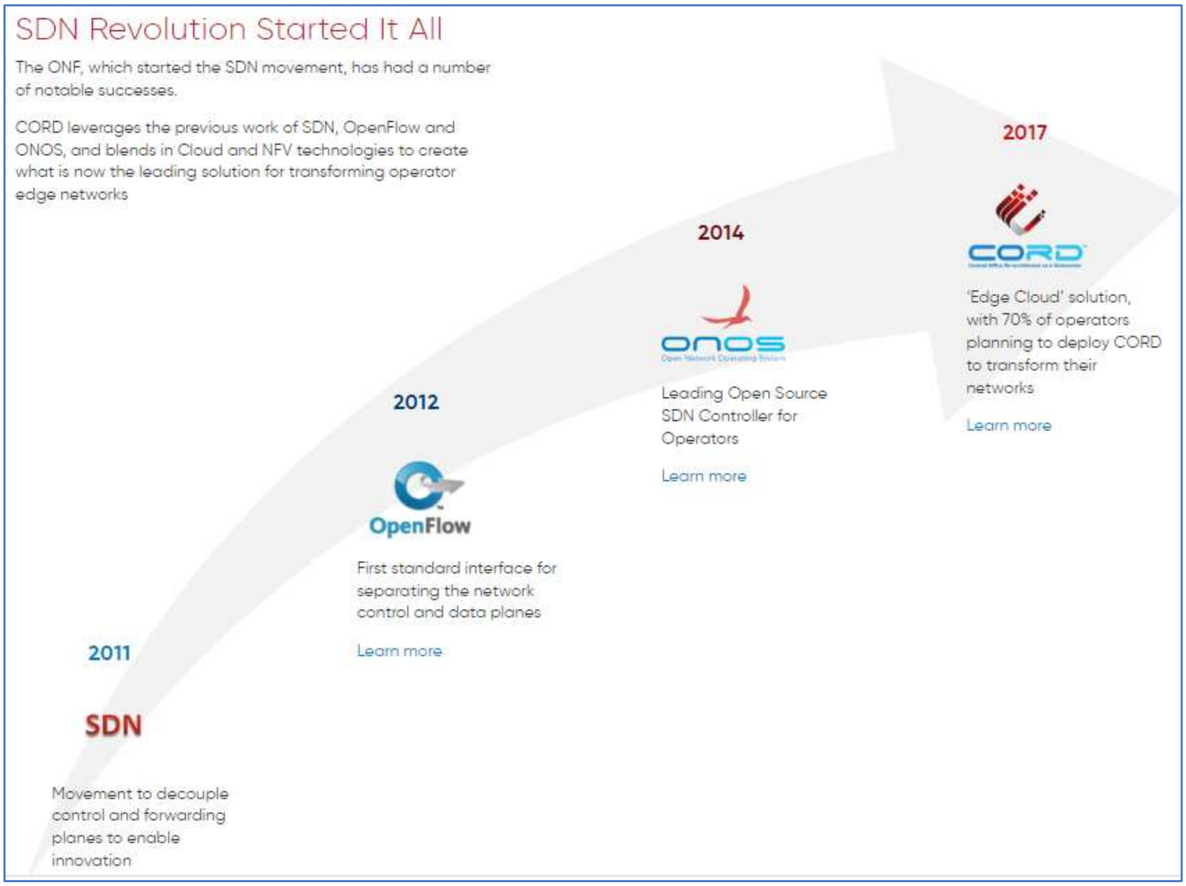
FIGURE 56: GENERIC SDN ARCHITECTURE ZOOM-IN [23]

Image Source: ENISA

Software-Defined Networking (SDN) is an emerging dynamic, manageable, cost-effective, and adaptable architecture, making it ideal for today's applications' high-bandwidth, dynamic nature. The architecture decouples the network control and forwarding functions. It enables the network control part to become directly programmable and the underlying infrastructure part to be abstracted for applications and network services.

The SDN technology is commonly associated with the OpenFlow protocol to determine the path of network packets across network switches for remote communication with network plane elements.

4.3.1 SDM Revolution and Architecture Comparison



SDN Revolution Started It All

The ONF, which started the SDN movement, has had a number of notable successes.

CORD leverages the previous work of SDN, OpenFlow and ONOS, and blends in Cloud and NFV technologies to create what is now the leading solution for transforming operator edge networks

2011

SDN

Movement to decouple control and forwarding planes to enable innovation

2012

OpenFlow

First standard interface for separating the network control and data planes

Learn more

2014

ONOS

Leading Open Source SDN Controller for Operators

Learn more

2017

CORD

'Edge Cloud' solution, with 70% of operators planning to deploy CORD to transform their networks

Learn more

FIGURE 57: SDN REVOLUTION [38]

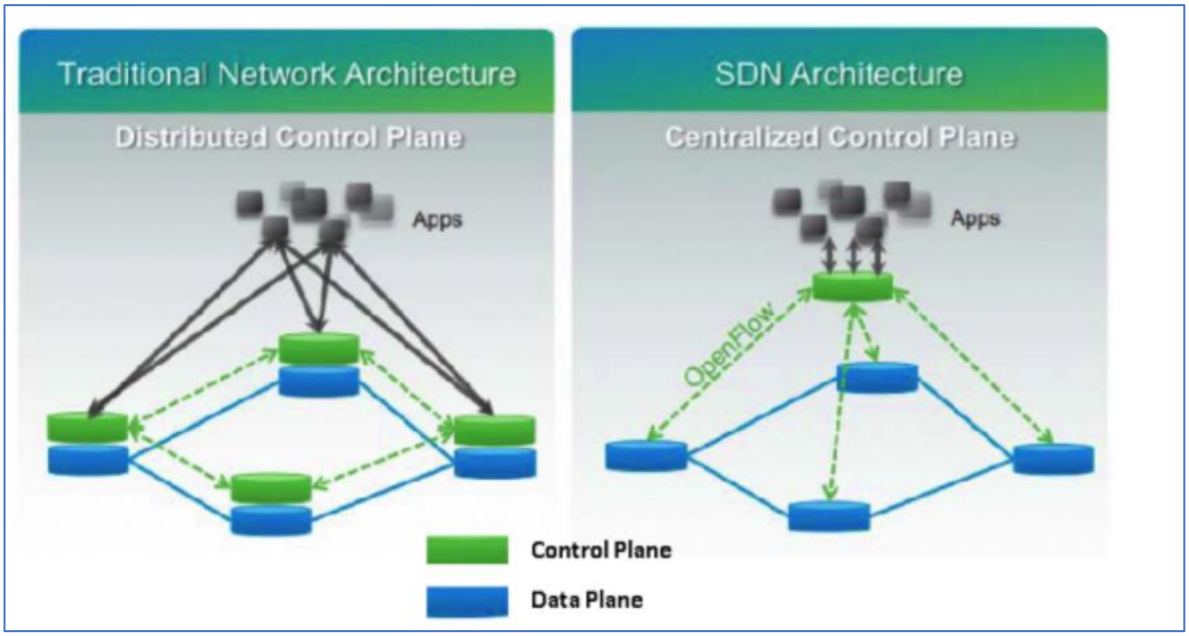

Traditional Network Architecture
Distributed Control Plane
Apps
SDN Architecture
Centralized Control Plane
Apps
OpenFlow
Control Plane
Data Plane

FIGURE 58: TRADITIONAL NETWORK VS. SDN ARCHITECTURE [41]

SDN is the solution to addressing the static architecture of traditional networks. SDN tries to centralize network intelligence in one network component by splitting the forwarding process of network packets, i.e., the data plane, from the routing process, i.e., the control plane. The control plane includes one or more controllers, considered the SDN network's brain, where natural intelligence is incorporated.

The SDN architecture is:

- ***Directly Programmable*****:** Network control is directly programmable as it is decoupled from forwarding functions.

- ***Agile*****:** Abstracting control panel from forwarding panel lets administrators dynamically accommodate network-wide traffic flow to meet changing needs.

- ***Centrally Managed*****:** Network intelligence is centralized in software-based SDN controllers that sustain a global view of the network, which connects to policy engines and applications as a single logical switch.

- ***Programmatically Configured*****:** With dynamic and automated SDN programs, network engineers can configure, manage, secure, and optimize network resources very quickly. It is possible as the programs do not depend on proprietary software.

- ***Open Standards-Based and Vendor-Neutral*****:** The architecture simplifies network design and operation as SDN controllers provide the instructions. When implemented through open standards, there is no dependency on multiple vendor-specific devices and protocols.

4.3.2 SDM Layers

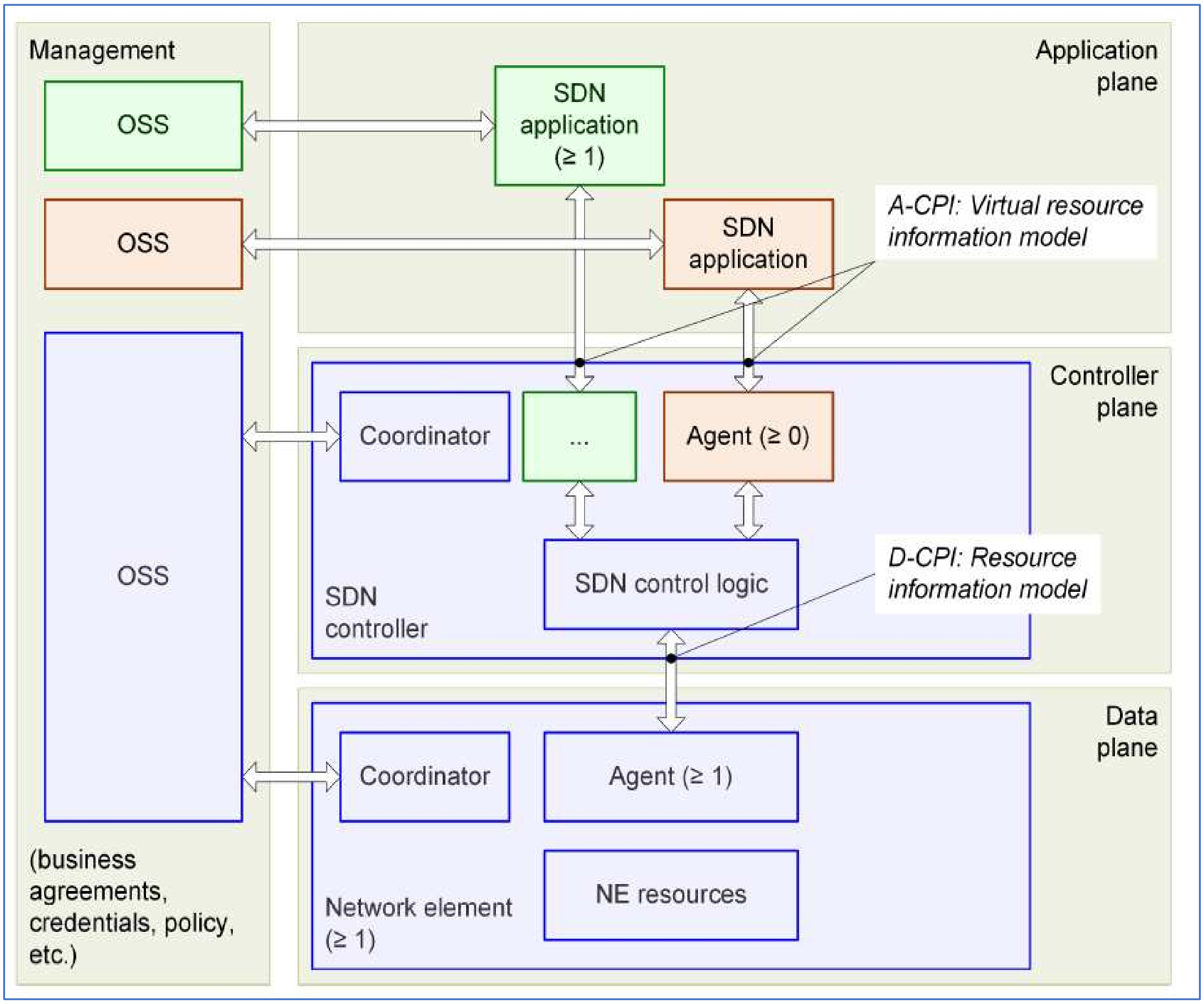


FIGURE 59: SDM REFERENCE MODEL [37]

The SDN Architecture comprises three layers:

• The **Data Plane** contains network elements, which expose their capabilities toward the control layer (Controller Plane) via the data-controller plane interface (D-CPI).

• In the **Controller Plane**, the SDN controller translates the applications' requirements and exerts more granular control over the network elements while providing relevant information to the SDN applications. Services are offered to applications via the application-controller plane interface (A-CPI, often called NBI) through an Information model instance derived from the underlying resources, management-installed policy, and local or externally available support functions. An SDN controller may orchestrate competing application demands for limited network resources.

• SDN applications reside in the **Application Plane** and communicate their network requirements toward the Controller Plane via the A-CPI.

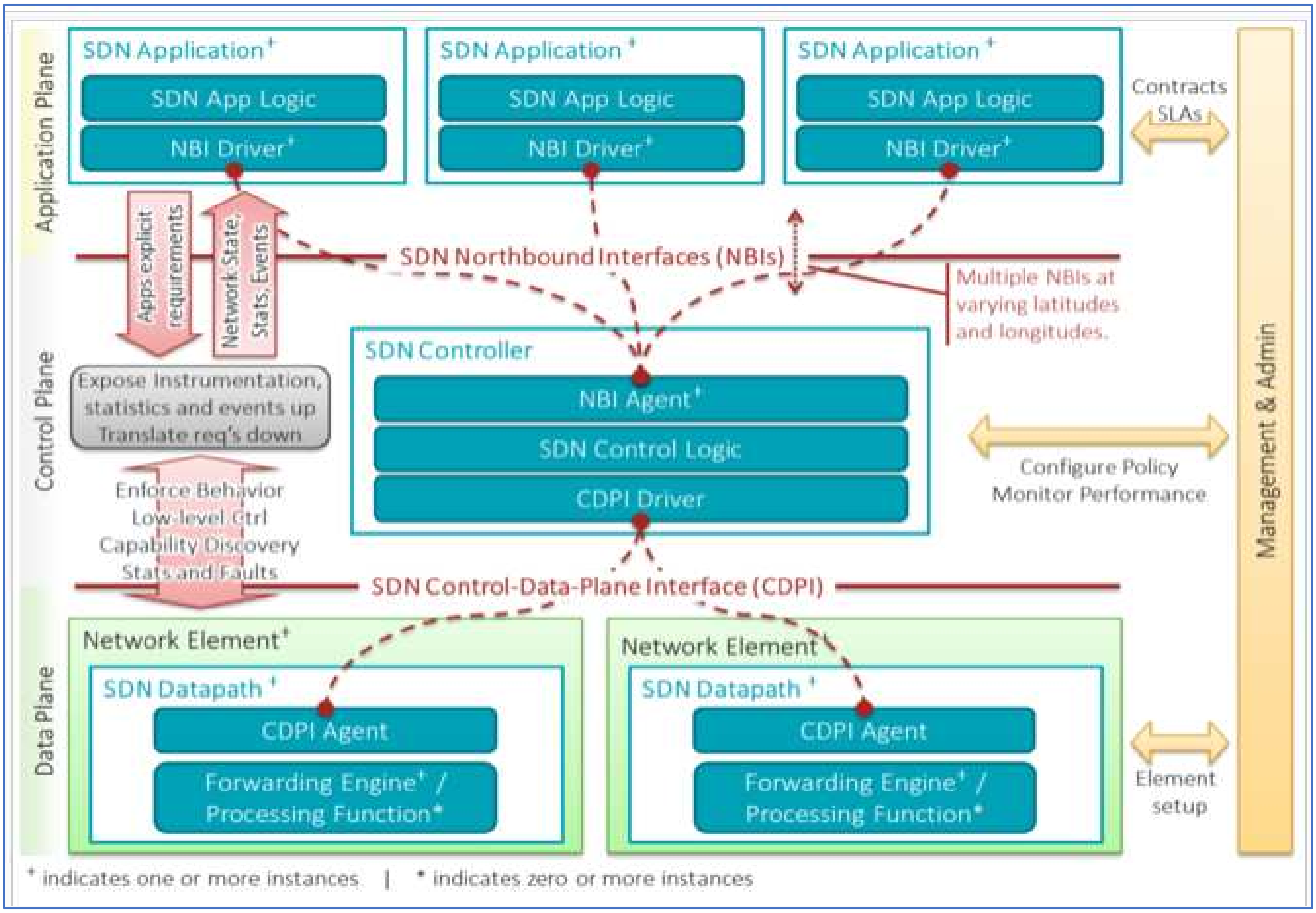


FIGURE 60: A HIGH-LEVEL OVERVIEW OF THE SOFTWARE-DEFINED NETWORKING ARCHITECTURE [40]

Image Source: Wikipedia

4.3.3 SDM Elements

The elements of the SDN architecture are as follows:

SDN Application

SDN Applications explicitly, directly, and programmatically communicate with the network requirements and take necessary actions to meet desired network behavior to the SDN Controller via a northbound interface (NBI). In addition, it may consume an abstract view of the network for internal decision-making purposes. An SDN Application includes one SDN Application Logic and one or more NBI Drivers. In addition, SDN Applications may expose another layer of network control, thus offering one or more higher-level NBIs through respective NBI agents.

Multiple case scenarios might be envisioned, for the position of the SDN applications in the NFV architectural framework, such as:

• the network hardware might be a physical appliance talking to an SDN controller or a complete solution including multiple SDN components, such as an SDN controller + SDN application.

• the VIM might be an application interfacing with an SDN controller in the NFVI - for instance, OpenStack Neutron as a VIM interfacing with an SDN controller in the NFVI.
• the SDN application might be a VNF talking to an SDN controller, being Virtualised or not. For instance, a PCRF VNF might speak to an SDN controller for some policy management for traffic steering.
• the SDN application might be an element manager interfacing with an SDN controller to collect some metrics or configure some parameters, and
• the SDN application might be interfacing with an SDN controller, for instance, in the OSS-BSS for tenant SDN service definitions.

SDN Controller

The SDN Controller is a centralized entity in charge of:

(i) ensure translation of the requirements from the SDN Application layer beneath to the SDN Datapath and

(ii) provide the SDN Applications with a localized network view, including statistics and events.

An SDN Controller includes the SDN Control Logic, one or more NBI Agents, and the Control to Data-Plane Interface (CDPI) driver.

SDN Datapath

The SDN Datapath is a network device that describes the visibility and control over its forwarding and data processing functions. This logical representation may comprehend all or a subset of the physical substrate resources. An SDN Datapath encompasses a CDPI agent, a set of traffic forwarding engines, and more traffic processing functions. These engines and procedures include simple forwarding functions among the Datapath's external interfaces, internal traffic processing, and termination functions. SDN Datapath may also be contained in a single physical network element where a combination of communications resources is managed as a unit. Multiple physical network elements can also be involved with An SDN Datapath functionality.

SDN Control to Data-Plane Interface (CDPI)

The SDN CDPI is the interface classified between an SDN Datapath and an SDN Controller, which provides:

(i) programmatic control of involved all forwarding operations,

(ii) advertisement capabilities,

(iii) statistical reporting, and

(iv) notification of events.

One value of SDN defines to meet the expectation that the CDPI is implemented in an open, vendor-neutral, and interoperable way.

SDN Northbound Interfaces (NBI)

SDN NBIs are interfaces between SDN Applications and SDN Controllers and typically provide abstract network views and enable direct expression of network behavior and requirements. This may occur at any level of abstraction (latitude) and across different sets of functionalities (longitude). One value of SDN lies in the expectation that these interfaces are implemented in an open, vendor-neutral, and interoperable way.

SDN resources

Multiple scenarios might be envisaged for the actual location of SDN resources:

- physical switch or router.
- virtual switch or router.
- e-switch, software-based SDN-enabled switch in a server NIC and
- switch or router as a Virtual network function (VNF).

4.3.4 Relationship to NFV

NFV Network Function Virtualization is a technical concept that Accompaniments SDN. So, NFV is not contingent on SDN or SDN concepts. Instead, NFV deallocates software from hardware to allow dynamic operation and flexible network deployment. NFV is customarily deployed on typical hardware-based servers' solutions to run network services software versions. These software-based services in an NFV environment are called Virtual Network Functions (VNF). NFV aims to accelerate service innovation and provisioning using standard IT virtualization technologies via an SDN-NFV hybrid program to provide high efficiency, elastic and scalable capabilities.

In addition, SDN provides the agility of controlling generic forwarding devices such as routers and switches using SDN controllers. On the other hand, NFV agility is programmed by using virtualized servers for network applications. Therefore, deploying a virtualized network function VNF as a standalone entity mode is possible using existing networking and orchestration paradigms. However, there are some intrinsic benefits to implementing SDN concepts and managing an NFV infrastructure. Therefore, multivendor platforms that incorporate SDN and NFV in concerted ecosystems are being defined while looking at the management and orchestration of VNFs. However, centralization has drawbacks regarding security, scalability, and elasticity, which is SDN’s central issue.

SECTION 5. CYBERSECURITY IN 5G TECHNOLOGY

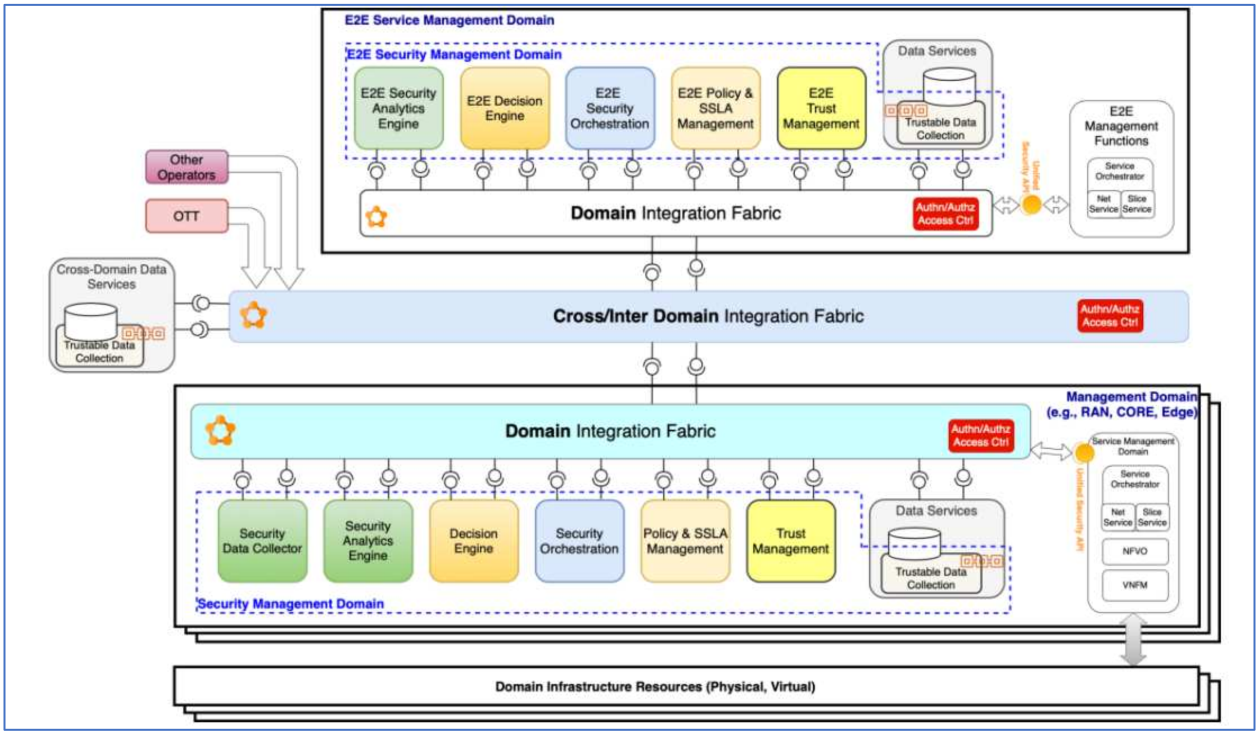

E2E Service Management Domain
E2E Security Management Domain
E2E Security Analytics Engine
E2E Decision Engine
E2E Security Orchestration
E2E Policy & SSLA Management
E2E Trust Management
Data Services
Trustable Data Collection
Unified Security API
E2E Management Functions
Service Orchestrator
Net Service
Slice Service
Domain Integration Fabric
Authn/Authz Access Ctrl
Other Operators
OTT
Cross-Domain Data Services
Trustable Data Collection
Cross/Inter Domain Integration Fabric
Authn/Authz Access Ctrl
Management Domain (e.g., RAN, CORE, Edge)
Domain Integration Fabric
Authn/Authz Access Ctrl
Service Management Domain
Service Orchestrator
Net Service
Slice Service
NFVO
VNFM
Unified Security API
Security Data Collector
Security Analytics Engine
Decision Engine
Security Orchestration
Policy & SSLA Management
Trust Management
Data Services
Trustable Data Collection
Security Management Domain
Domain Infrastructure Resources (Physical, Virtual)

FIGURE 61: SECURITY FRAMEWORK HIGH-LEVEL ARCHITECTURE

Image Source: 5g-ppp

Cybersecurity is protecting networks, devices, and data from unauthorized access or illegal use and ensuring confidentiality, integrity, and availability of information. While the transition to 5G presents a wealth of opportunities and capabilities, it also introduces new vulnerabilities and threats. For example, 5G technologies enable the potential for billions of connected network devices, supporting a wealth of new capabilities and innovation. However, these devices and infrastructure capabilities, such as cellular towers, beamforming transmission, small cells, and mobile devices, allow malicious actors to expose vulnerabilities across increased threat vectors. For example, suppose network devices were compromised through a network layer or exploit. In that case, malicious actors could obtain unauthorized access to the 5G network, potentially disrupting operations and enabling critical data's interception, manipulation, and destruction.

In 2014 ETSI created TC CYBER with the responsibility for the standardization of cyber security and for providing a center of relevant expertise. Growing dependence on networked digital systems has increased the variety and quantity of cyber threats. Moreover, the different methods governing secure communication in the various Member States of the European Union and beyond Europe sometimes need help to assess the respective risks and ensure adequate security. Therefore, building on its world-leading expertise in protecting Information and Communications Technologies (ICT), ETSI set up a new cyber security committee (TC CYBER) to meet the growing demand for standards to protect the Internet and the communications and business it carries.

5.1 Security Architecture (SA) [23]

The 5G security architecture consists of various network functions (NF) and components responsible for securing end-to-end communications, providing Authentication, and other security functions. In addition, the 5G security architecture consists of components of various other architectures acting thus horizontally across all different architectures. In particular, security functions secure the access of users within the radio access network (RAN), they cover security functions in the core network and perimeter entities (edge computing), and they provide security functions in the Network Function Virtualisation (NFV). Finally, a set of elements covers security management functions, audits, and analytics.

The 5G security architecture's detailed structure is shown in Figure 62.

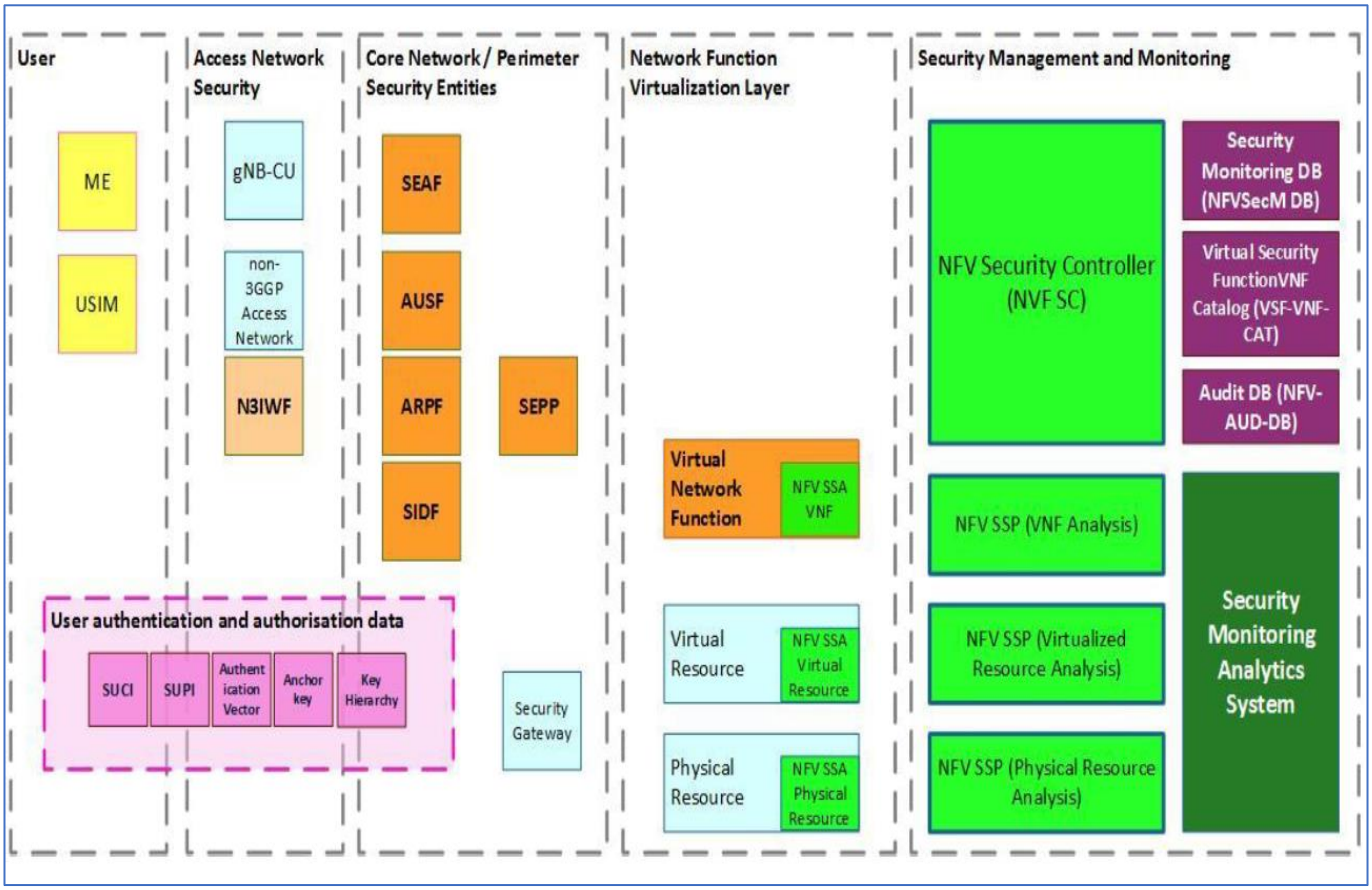

User
Access Network Security
Core Network / Perimeter Security Entities
Network Function Virtualization Layer
Security Management and Monitoring
ME
USIM
gNB-CU
non-3GGP Access Network
N3IWF
SEAF
AUSF
ARPF
SIDF
SEPP
Security Gateway
User authentication and authorisation data
SUCI
SUPI
Authentication Vector
Anchor key
Key Hierarchy
Virtual Network Function
NFV SSA VNF
Virtual Resource
NFV SSA Virtual Resource
Physical Resource
NFV SSA Physical Resource
NFV Security Controller (NVF SC)
Security Monitoring DB (NFVSecM DB)
Virtual Security FunctionVNF Catalog (VSF-VNF-CAT)
Audit DB (NFV-AUD-DB)
NFV SSP (VNF Analysis)
NFV SSP (Virtualized Resource Analysis)
NFV SSP (Physical Resource Analysis)
Security Monitoring Analytics System

FIGURE 62: 5G SECURITY ARCHITECTURE ZOOM-IN

Image Source: ENISA

A very brief description of these elements of the 5G security architecture is as follows:

5.1.1 Mobile Equipment (ME)

ME stands for all kinds of mobile equipment that can be connected to the 5G network. For example, ME can be sensors, IoT components, connected autonomous systems, eHealth devices, etc.

5.1.2 Universal Subscriber Identity Module (USIM)

USIM is the SIM card of 5G. It is a platform for securing access and communication in 5G. It is the only security module mentioned in the 3GPP specification.

5.1.3 5G Node Base Station Central Unit (gNB-CU)

3GPPP has formulated some security requirements for gNB-CU. These requirements increase the security properties of gNB and – when implemented - are relevant to the security architecture.

5.1.4 Non-3GPP Access Network

Security for non-3GPP access to the 5G Core network is achieved using IKEv2 as defined in RFC 7296 to set up one or more IPsec ESP security associations. The UE takes the role of the IKE initiator (or client), and the N3IW accepts the part of the IKE responder (or server).

5.1.5 Security Anchor Function (SEAF)

The *SEAF* will create a unified anchor key KSEAF (typical for all accesses) for the primary Authentication that the UE and the serving network can use to protect the subsequent communication.

5.1.6 Authentication Server Function (AUSF)

The Authentication server function (AUSF) shall handle authentication requests for both 3GPP access and non-3GPP access. The AUSF should provide SUPI to the VPLMN only after authentication confirmation if VPLMN sends an authentication request to SUCI. In addition, the AUSF shall inform the UDM that a successful or unsuccessful authentication of a subscriber has occurred.

5.1.7 Authentication Credential Repository and Processing Function (ARPF)

ARPF selects an authentication method based on subscriber identity and configured policy and computes the authentication data and keying materials.

5.1.8 Subscription Identifier De-concealing Function (SIDF)

The SIDF is responsible for the de-concealment of the Subscription Concealed Identifier (SUCI) and shall fulfill the following requirements:

- The SIDF shall be a service offered by UDM.
- The SIDF shall resolve the SUPI from the SUCI based on the protection scheme used to generate the SUCI.

5.1.9 Security Edge Protection Proxy (SEPP)

The 5G System architecture introduces a Security Edge Protection Proxy (SEPP) as the entity sitting at the perimeter of the mobile network. The SEPP shall act as a non-transparent proxy node.

5.1.10 NFV Security Services Agent (SSA)

The NFV SSA exists in both the NFVI domain and in VNF domain. NFV SSA in the VNF domain may exist as a separate VSF or within a VNF. The NFV SSA is responsible for securely receiving the Security Monitoring policy and implementing the same.

5.1.11 NFV Security Controller (NSC)

The NFV SC may interface with other security systems (e.g., Security Analytics), security databases, and other policy engines. The NFV SC orchestrates the system's broad security policies. The NFV SC acts as a trusted 3rd party that resides independently.

An NFV SC manages NFV SSAs (like VSFs) to keep them consistent according to the policy specified. SC also facilitates secure bootstrapping of SSAs (like VSFs), working instances of SSAs, fast pairing up with SSA's VNFMs and EMs, personalize the SSAs, policy management, integrity assertion, credential management, facilitating clustering of multiple SSAs into a distributed appliance, monitoring of SSAs for failure and remediation.

5.1.12 NFV Security Services Provider (SSP)

The NFV SSP is located within the VIM and VNFM. It is responsible for security monitoring policy orchestration received from the Security Controller (NFV SC) and interacting with the various VIM/VNFM components to implement the policy across multiple systems comprising the NFVI/VNF. Furthermore, NFV SSP is also responsible for receiving the telemetry data from various NFV SSAs, and optionally making some analysis based on this data.

5.1.13 NFV Security Monitoring Database

The NFV SecM-DB is a secure database of security data for deploying NFV system-wide Security Monitoring. This includes Security Monitoring policy and configurations, security credentials for facilitating secure communications between the various Security Monitoring components, and credentials for safe storage of telemetry, including tenant-specific security policies.

5.1.14 SA/VSF Catalogue Database (VSF-NVNF-CAT)

The NFV VSF-VNF-CAT is a repository for Security Services Agents like the Virtual Security Functions (VSF) VNFs. The catalog can add and remove SSAs (VSF) packages and images and includes a VSF VNFD containing metadata and information about that VSF VNF. Once the SSA (VSF) package or instance is added to the catalog, it becomes available for orchestration.

5.1.15 Audit DB

The NFV AUD-DB is a secure database consisting of security audit information.

5.1.16 Security Monitoring Analytics System

The Security Monitoring Analytics system securely receives Security Monitoring telemetry from across the NFV systems, including the MANO and all the NFVIs that may be geographically distributed. The analytics system applies advanced machine learning techniques on the telemetry to perform advanced detection of security anomalies and emerging threats. This system also can trigger remediation actions through the NFV SC.

5.1.17 Subscription Concealed Identifier (SUCI)

A one-time use subscription identifier contains the Scheme-Output and additional non-concealed information for home network routing and protection scheme usage.

5.1.18 Authentication Vector

A vector consisting of RAND, authentication Token (AUTN), and Hash eXpected RESponse (HXRES).

5.1.19 Anchor Key

The security key KSEAF provided during Authentication was used to derivate subsequent security keys.

5.1.20 Key Hierarchy

Hierarchy of cryptographic key derived from Anchor Key. (Defined in ETSI TS 133 501 section 6.2.). It includes the following keys: KAUSF, KSEAF, KAMF, KNASint, KNASenc, KN3IWF, KgNB, KRRCint, KRRCenc, KUPint, and KUPenc.

5.2 5G System Model [45]

The 5G technology ecosystem is complex and still evolving. Any 5G cybersecurity assessment approach must extend and grow as new 5G standards, threat vectors, deployment features, and policies are introduced. Moreover, a shared understanding and lexicon are needed to identify and describe critical elements, functions, relationships, and processes. To investigate the need for a 5G security evaluation process, it is necessary to:

- Identify a 5G system model to serve as the lexicon for 5G and delineate the 5G attack surface for the investigation.
- Describe 5G and how it differs from prior generations of cellular technology.
- Understand—at a high level—the threats that could impact a 5G-enabled system.

A 5G system includes User Equipment (UE), 5G RAN, and 5G Core and may consist of Multi-Access Edge Computing (MEC) and Network Slicing. Because of the extensive use of virtualization in 5G and the critical need for secure network orchestration and management, these subsystems are also included in the 5G system model shown in Figure 63. Each top-level 5G system component may be represented as a single subsystem configuration or a collection of component configurations. In addition, 5G networks likely will include multiple service providers (e.g., cloud, communications, or application service providers for MEC or Core), thereby introducing additional complexity to the security risk assessment.

Recognizing this facet of 5G, Figure 63 shows that each 5G subsystem has a set of attributes (e.g., whether the 5G system uses SA or NSA architecture or whether UE can be updated/patched, authenticated, or centrally managed or who owns the RAN and whether it is virtualized). This 5G system decomposition enables unique functions, and system attributes to be applied to system components, serves as the reference model for the 5G security evaluation process investigation, and facilitates the identification of 5G threats.

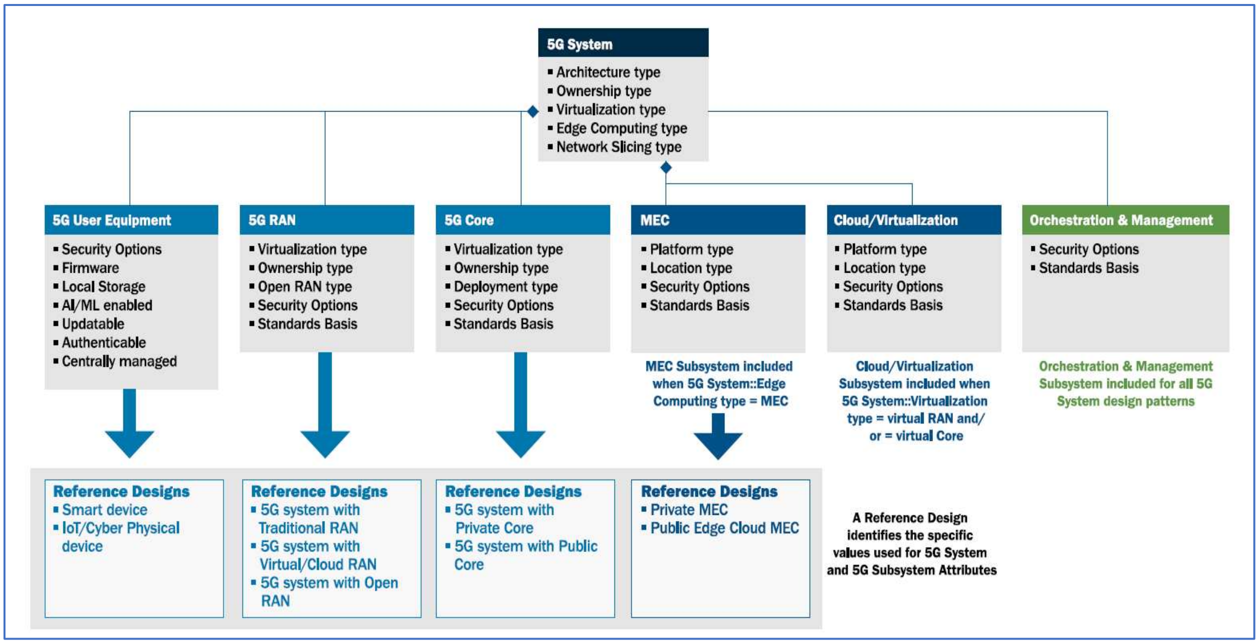


FIGURE 63: 5G SYSTEM MODEL FOR SECURITY EVALUATION PROCESS [45]

Image Source: CISA

The middle tier (i.e., 5G UE, RAN, Core, etc.) depicted in the above figure is already discussed in earlier sections. The lowest tier in the above figure shows ***reference designs,*** which represent a standard implementation of a 5G subsystem (i.e., when traditional RAN, virtual RAN, and O-RAN are reference designs for the RAN subsystem). In addition to the attack surface for each 5G subsystem, a reference design may introduce additional threats, such as different interfaces for O-RAN that do not exist in traditional RAN. This level of abstraction should facilitate risk assessments by multiple domain subject matter experts and engagement by non-technical stakeholders and decision-makers. A few examples of 5G reference designs are presented in Figure 64.

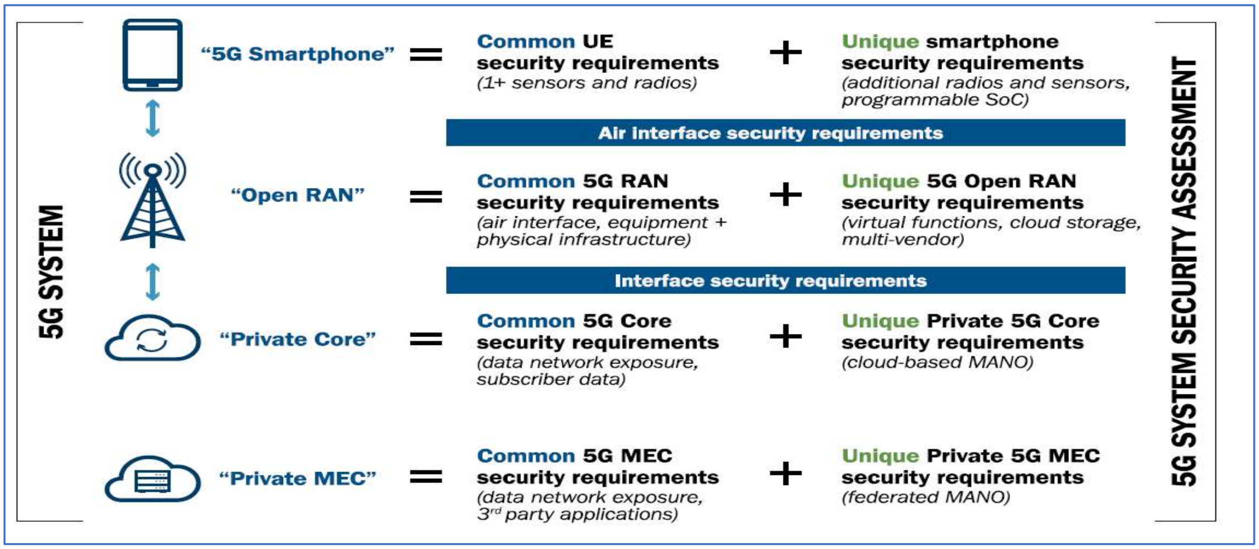

5G SYSTEM
"5G Smartphone" =
Common UE security requirements
(1+ sensors and radios)
+
Unique smartphone security requirements
(additional radios and sensors, programmable SoC)
Air interface security requirements
"Open RAN" =
Common 5G RAN security requirements
(air interface, equipment + physical infrastructure)
+
Unique 5G Open RAN security requirements
(virtual functions, cloud storage, multi-vendor)
Interface security requirements
"Private Core" =
Common 5G Core security requirements
(data network exposure, subscriber data)
+
Unique Private 5G Core security requirements
(cloud-based MANO)
"Private MEC" =
Common 5G MEC security requirements
(data network exposure, 3rd party applications)
+
Unique Private 5G MEC security requirements
(federated MANO)
5G SYSTEM SECURITY ASSESSMENT

FIGURE 64: TOP-LEVEL 5G SYSTEM REFERENCE DESIGN [45]

Image Source: **CISA [TLP: WHITE]**

The above four example reference designs in Figure 64 may be combined to represent a single 5G system. One representation each for a 5G UE, RAN, Core, and MEC solution shows how security requirements for a "generic" subsystem (e.g., generic UE) can be extended to include additional security requirements introduced by the internal "unique" solution architecture (e.g., hardware, software, and interfaces), as well as functions and attributes defined for the given reference design (e.g., additional sensors, radios, and computing capacity for a smart device). These requirements are critical inputs to the proposed 5G security evaluation process. As an added benefit, these reference designs may be reused and modified to encapsulate unique functional and operational details of a future 5G system application or use case.

5.3 5G Security and Threat Landscape [45]

Compared to 4G cellular network technology, 5G will serve more devices of varying types and use cases. In addition, 5G introduces new features and services, including the following:

o NR with enhanced capabilities, increased spectrum, spectrum sharing, and low-, mid-, and high-band frequencies.

o Cell densification to serve large numbers of users and new techniques such as beamforming to direct the wireless communication channel at users and reduce interference.

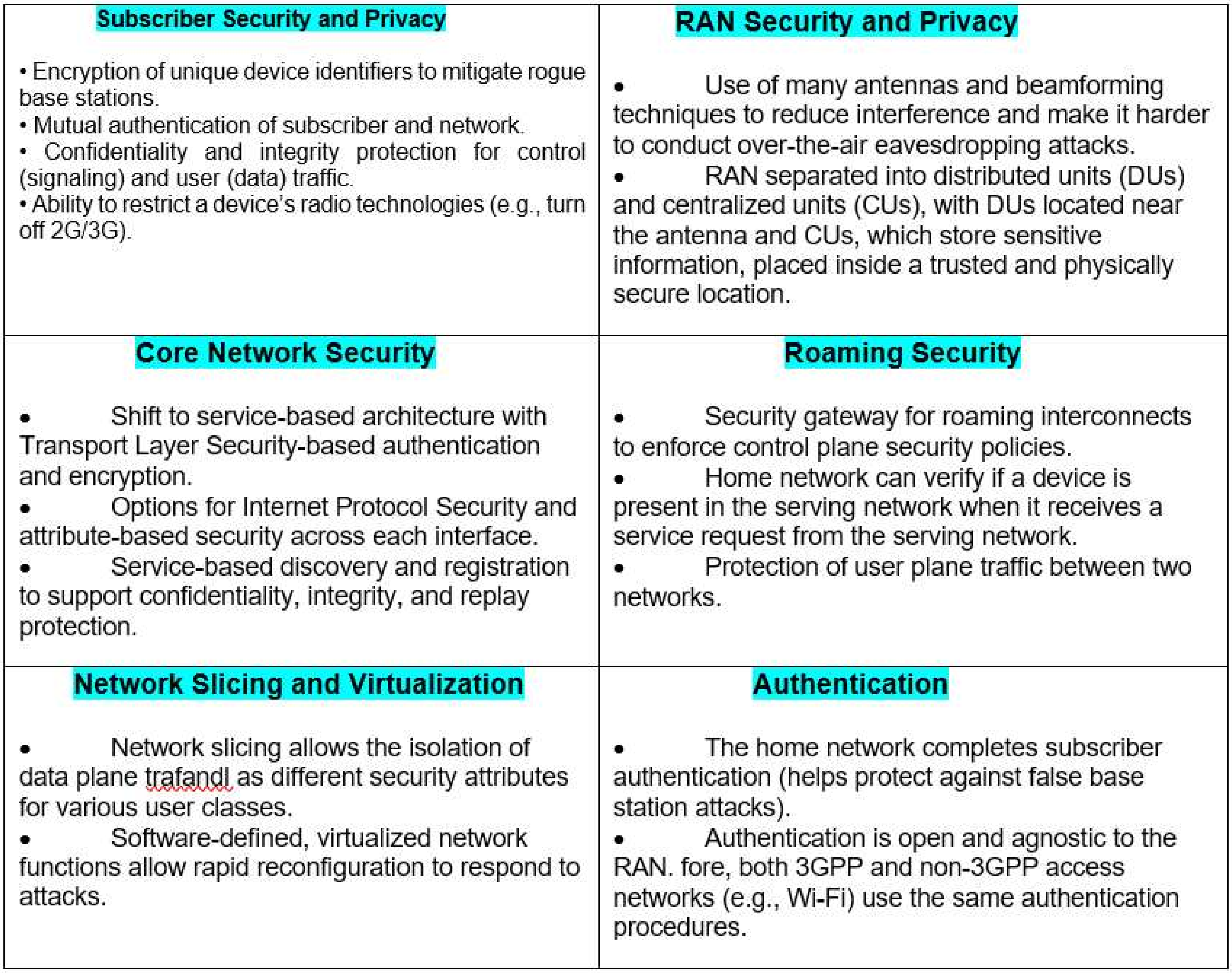
o MEC, which typically moves centralized applications closer to the network to reduce latency, sustain high data transfer rates, and ingest high volumes of data.

o Network slicing creates multiple virtual networks that provide different quality of service levels over shared physical infrastructure.

o Virtualization of the RAN and the 5G Core to dynamically scale network functions.

5.3.1 Security Improvements

The Third Generation Partnership Project (3GPP), the central standards-development organization for 5G, has built many security improvements into 5G, summarized in Table 5.



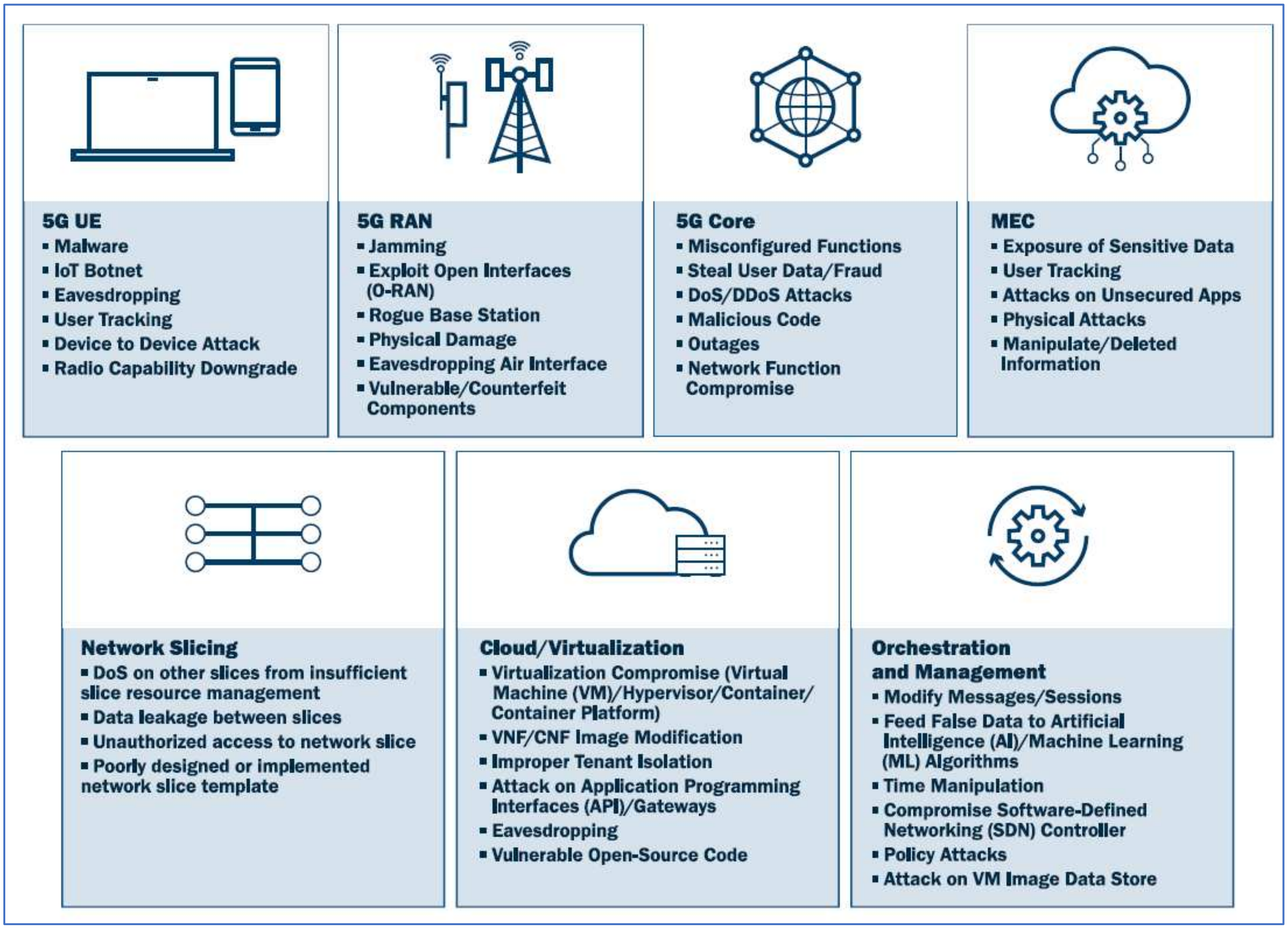
Subscriber Security and Privacy • Encryption of unique device identifiers to mitigate rogue base stations. • Mutual authentication of subscriber and network. • Confidentiality and integrity protection for control (signaling) and user (data) traffic. • Ability to restrict a device's radio technologies (e.g., turn off 2G/3G).	**RAN Security and Privacy** • Use of many antennas and beamforming techniques to reduce interference and make it harder to conduct over-the-air eavesdropping attacks. • RAN separated into distributed units (DUs) and centralized units (CUs), with DUs located near the antenna and CUs, which store sensitive information, placed inside a trusted and physically secure location.
Core Network Security • Shift to service-based architecture with Transport Layer Security-based authentication and encryption. • Options for Internet Protocol Security and attribute-based security across each interface. • Service-based discovery and registration to support confidentiality, integrity, and replay protection.	**Roaming Security** • Security gateway for roaming interconnects to enforce control plane security policies. • Home network can verify if a device is present in the serving network when it receives a service request from the serving network. • Protection of user plane traffic between two networks.
Network Slicing and Virtualization • Network slicing allows the isolation of data plane trafandl as different security attributes for various user classes. • Software-defined, virtualized network functions allow rapid reconfiguration to respond to attacks.	**Authentication** • The home network completes subscriber authentication (helps protect against false base station attacks). • Authentication is open and agnostic to the RAN. fore, both 3GPP and non-3GPP access networks (e.g., Wi-Fi) use the same authentication procedures.

TABLE 5: 3GPP SECURITY IMPROVEMENTS

5.3.2 Threat Landscape

A key input to any security risk assessment is threat analysis. The 5G system model supports the depiction of the attack surface for the investigation. In addition, there are numerous threat frameworks such as those offered by MITRE ATT&CK, the European Union Agency for Cybersecurity's (ENISA) 5G Threat Landscape, the Threat Modeling Framework for Mobile Communication Systems, 3GPP's Security Assurance Specifications (SCAS) and Technical Specification (TS) 33.501 publications released by the Federal Communications Commission (FCC) Communications Security, Reliability, and Interoperability Council VII (CSRIC), 5G Enablers for Network and System Security and Resilience (ENSURE), and the GSM Association's (GSMA) Security Manual.

According to these resources and threat analyses conducted by 3GPP and a paper on potential 5G threat vectors below, Figure 65 shows some of the threats to the 5G subsystems extracted from these sources. Some threats, such as eavesdropping, theft of user data, or user location tracking, may impact the integrity and confidentiality of user data and service availability to individual users. Other threats may impact local or regional network, application, or service availability (e.g., denial of service [DoS] or Distributed DoS [DDoS] attacks, misconfigured or compromised virtualization platforms or network functions, vulnerable components [supply chain threats], physical attacks on edge computing components), with follow-on effects on the confidentiality, integrity, and availability of 5G services and applications for enterprises relying on 5G for their missions.



5G UE
- Malware
- IoT Botnet
- Eavesdropping
- User Tracking
- Device to Device Attack
- Radio Capability Downgrade

5G RAN
- Jamming
- Exploit Open Interfaces (O-RAN)
- Rogue Base Station
- Physical Damage
- Eavesdropping Air Interface
- Vulnerable/Counterfeit Components

5G Core
- Misconfigured Functions
- Steal User Data/Fraud
- DoS/DDoS Attacks
- Malicious Code
- Outages
- Network Function Compromise

MEC
- Exposure of Sensitive Data
- User Tracking
- Attacks on Unsecured Apps
- Physical Attacks
- Manipulate/Deleted Information

Network Slicing
- DoS on other slices from insufficient slice resource management
- Data leakage between slices
- Unauthorized access to network slice
- Poorly designed or implemented network slice template

Cloud/Virtualization
- Virtualization Compromise (Virtual Machine (VM)/Hypervisor/Container/ Container Platform)
- VNF/CNF Image Modification
- Improper Tenant Isolation
- Attack on Application Programming Interfaces (API)/Gateways
- Eavesdropping
- Vulnerable Open-Source Code

Orchestration and Management
- Modify Messages/Sessions
- Feed False Data to Artificial Intelligence (AI)/Machine Learning (ML) Algorithms
- Time Manipulation
- Compromise Software-Defined Networking (SDN) Controller
- Policy Attacks
- Attack on VM Image Data Store

FIGURE 65: THREATS TO 5G SUBSYSTEMS [45]

Image source: **CISA**

Understanding these threats helps enterprises to prioritize security activities and identify the security capabilities needed to mitigate threats relevant to the 5G systems and subsystems within their 5G-enabled system boundary. The threat categories are:

- ❖ **General Cybersecurity Threats**
  These threats affect ***all 5G subsystems***, including misconfigurations, human error, failure to properly harden software and hardware, adversary lateral movement, information spillage, and general unauthorized access attacks. For example, attackers could exploit the misconfiguration of components or failure to properly set hardware or software to reconfigure 5G elements, steer traffic to an attacker, or steal data.

- **Virtualization Threats**
  Threats to a virtual machine (VM) and container service platforms impact the *5G Core, RAN, MEC, Network Slicing, Virtualization, Orchestration, and Management*. Threats include DoS, VM/container escape, side-channel attacks, and cloud service consumer misconfigurations. For example, extreme resource consumption by one tenant in a multi-tenant virtualization environment can create a DoS event for adjacent tenant systems. Such an event can prevent or seriously degrade mission functionality. Similarly, colocation attacks such as VM/container escape or side-channel attacks can put neighboring compute workloads at risk for resource deprivation, lateral movement, and compromise of data confidentiality, integrity, or availability. Finally, a side-channel attack on 5G RAN or Core functions can bypass user account permissions, virtualization boundaries, or protected memory regions, exposing sensitive information.

- **Network and Management Interface Threats**
  These threats impact network, management, and over-the-air interfaces of *all 5G subsystems****, including DoS, jamming, eavesdropping, address spoofing, traffic/message tampering, system/protocol discovery***, improper tenant traffic isolation, and access control attacks. Over-the-air interface threats are located between the UE and the RAN, where radio jamming techniques can cause interference that could prevent access by the UE or cause loss of 5G service. In addition, core network functions that are virtualized/containerized are deployed as tenants on shared cloud infrastructure, where improper isolation of traffic between tenants can expose those virtual environments to unauthorized access or loss of information confidentiality (e.g., subscriber data, network configurations, etc.).

- **Application and Service Threats**
  Threats associated with delivering 5G applications and services impact *all 5G subsystems*, including malware and malicious code injection, DoS and DDoS, Application Programming Interface (API) manipulation, exploitation of software vulnerabilities, and access control attacks. For example, Es such as smartphones is vulnerable to exploiting applications and malicious code that can expose private data to threat actors. Additionally, unprotected, or vulnerable APIs at the MEC could lead to unauthorized applications and information at the MEC and facilitate further attacks from within the network.

- **Rogue Elements**
  Threats from rogue *UE*, rogue base stations or Radio Units in the *RAN,* and rogue network hosts or spoofed components in the *MEC* can be used to attack the 5G system. Rogue base stations, for example, can use jamming to force UE to use the rogue base station and then capture user information and location. In contrast, rogue or malicious components in the MEC can compromise MEC applications to delete, alter, or steal data.

- **Privacy Threats**
  Threats to *UE* as well as systems in the *RAN* and *5G Core* involved in the handling, sharing, storage, and communication of user and user-associated information in a 5G network include eavesdropping, user and device identifier and location tracking, and user, protocol, and system spoofing attacks. For example, an attacker could monitor the air interface between the RAN and UE device to extract an unprotected unique device identifier and track the device user. At the same time, unauthorized access to subscriber data stored in the 5G Core could be used for identity theft or telecom fraud.

- **Environmental and Physical Threats**

Vulnerabilities and weaknesses in environmental and physical access control systems, natural disasters, and power outages impact the *RAN, 5G Core, MEC, and Virtualization* subsystems. Physical access to ports, equipment, and devices; natural disasters; electromagnetic pulse; and loss of power are predominant concerns. For example, small cells positioned on lamp posts in the RAN could be subjected to physical theft or damage. In contrast, a power outage or natural disaster could damage/render inaccessible RAN nodes or parts of the 5G Core.

- **Supply Chain Threats**

Threats can occur during provisioning, acquiring, and incorporating software, firmware, and hardware components into *UE, RAN, 5G Core, and Virtualization* subsystems. Threats include vulnerable or malicious component insertion, vulnerable or malicious open-source components, and attacks on vulnerable hardware, firmware, or operating systems. For example, malicious code injection into standard code repositories used to build system software for production release can profoundly impact operations, especially if the affected systems have access to privileged user systems such as those employed for identity and access management or network health and configuration management. Additionally, including firmware/hardware components of unknown provenance or security posture (e.g., in UE or the RAN) can introduce malicious or counterfeit parts into these subsystems, creating the potential for exposure of sensitive user and network data to adversaries.

- **Artificial Intelligence/Machine Language (AI/ML) Threats**

Threats to data integrity, confidentiality, and availability of *UE* (e.g., gateways for IoT or cyber-physical devices), the *RAN*, and the *Orchestration and Management* subsystems. These threats impact AI/ML software and systems and the network elements and services that rely on the data's accuracy, timeliness, and trustworthiness for decisions based on AI/ML, such as dynamic allocation of network functions. For example, corruption in the analytic function code used to execute algorithms or insert false or tainted data into the AI/ML algorithms can degrade network operation and potentially impact human safety (e.g., in autonomous vehicles or intelligent city traffic management).

SECTION 6. ANALYSING SECURITY CHALLENGES [47]

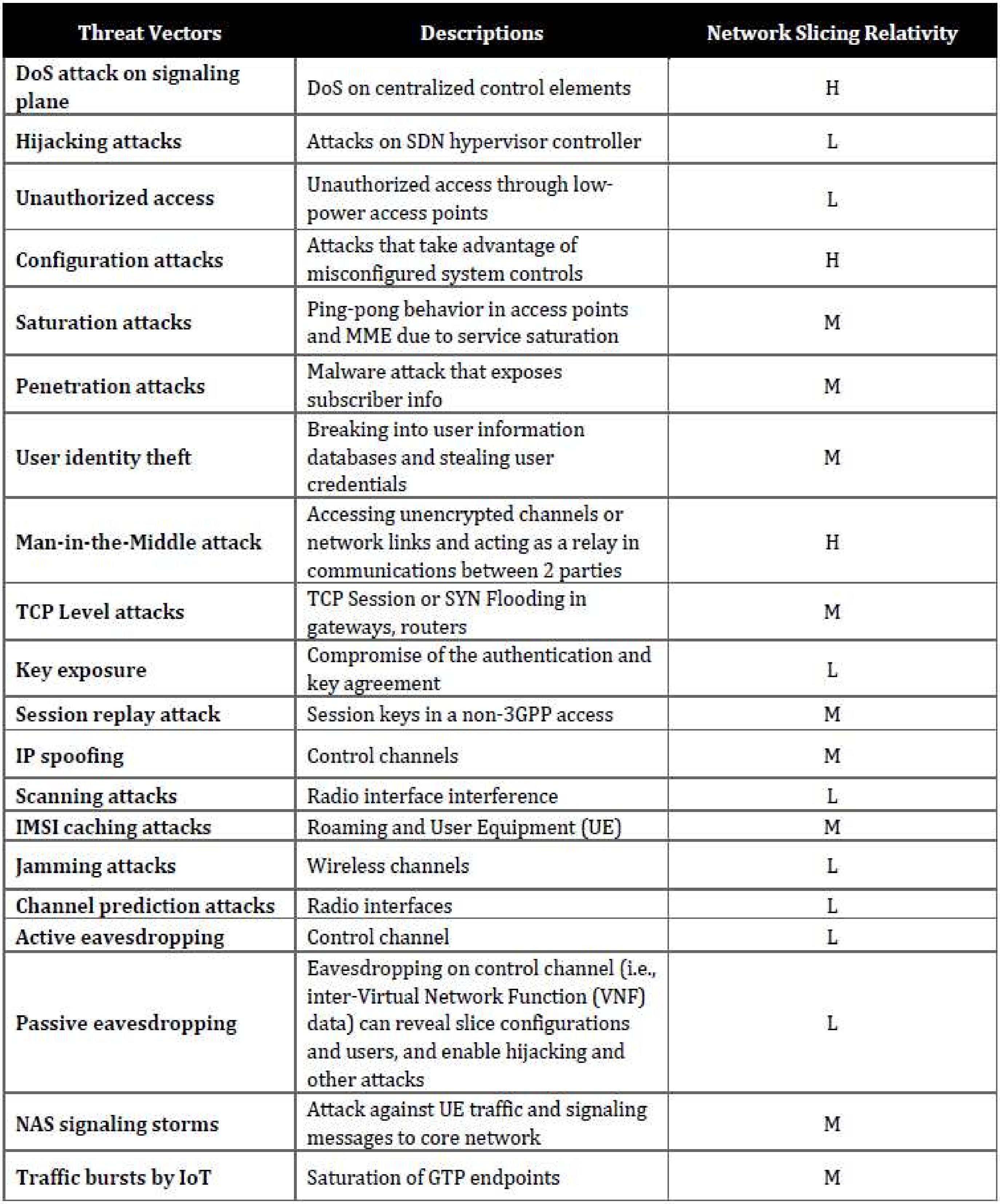
The National Security Agency-NSA, the Cybersecurity, Infrastructure Security Agency-CISA, and the Enduring Security Framework-ESF have analyzed the *Potential Threat Vectors to 5G Infrastructure based on factors such as* spanning policy & standards, supply chain, system architecture, etc. Although all are important to assess threats for network slicing, the core concern is systems architecture. As a result, the following sub-threats were identified within 5G systems architecture, as ascertained by ESF.

- Software Configuration
- Network Security
- Network Slicing
- Legacy Communications Infrastructure
- Multi-Access Edge Computing
- Spectrum Sharing
- Software Defined Networking

6.1 Network Slicing Threat Vectors

Network Slicing allows users to be legitimated for only one network area, allowing data and security isolation. However, network slicing can be challenging, considering managing and slicing a complex network. Though the network operators build the 5G network in their own specified way, there are no precise specifications for network operators to develop and employ security for network slicing. Improper network slice management might allow malicious actors to permit data from different network slices or deny access to prioritized users.

Extensive analysis was performed to correlate the 5G threat vectors and a network slicing threat to develop a high, medium, and low relativity assessment. This information is reflected in below Table 6.



Threat Vectors	Descriptions	Network Slicing Relativity
DoS attack on signaling 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 behavior in access points and MME due to service saturation	M
Penetration attacks	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 control channel (i.e., inter-Virtual Network Function (VNF) data) can reveal slice configurations and users, and enable hijacking and other attacks	L
NAS signaling storms	Attack against UE traffic and signaling messages to core network	M
Traffic bursts by IoT	Saturation of GTP endpoints	M

TABLE 6: POTENTIAL 5G THREAT VECTORS [47]

Among these threat vectors, three were measured as having a high level: denial-of-service (DoS), Man-in-the-Middle (MitM) attacks, and configuration attacks.

•**Denial of Service (DoS) Attacks**
O DoS attacks primarily affect a network slice's availability, resulting in compromised or unavailable communication services under these types of attacks.
O DoS attacks target the signaling plane to attack.
O Multiple types of threats may lead to a denial of service, such as flooding, amplification, signaling storm, and saturation attacks.
O An attack combining multiple vectors may lead to a distributed DoS (DDoS) attack.

•Man-in-the-Middle Attacks
O MitM attacks can cause a broad range of adverse effects on a network slice's confidentiality, integrity, and availability.
O MitM attacks infer that the adversary relays and alters between two endpoints while communicating. This type of attack could be devastating as the malicious actor modifies the contents of the messages via misinformation and disinformation.
O Confidentiality can also be violated as the middle person diverts and exposes information or data.

•Configuration Attacks
O Configuration attacks also have a broad range of adverse effects on a network slice's confidentiality, integrity, and availability.
O These attacks are when malicious actors exploit configured system controls. They may also include security features inadvertently turned off or system monitoring services being disabled.

Although not identified in Table 6, Network Function Virtualization (NFV) is another aspect of network slicing that presents increased potential risks. These include but are not limited to the following:

•The additional microservices and functionalities required by NFV make integration and testing activities more difficult, resulting in new attack surfaces.
•In a virtualized architecture, it will be more challenging to detect and recognize the types of traffic crossing these networks and mitigate against any new threats.
•Malicious actors may compromise a network slice by gaining access via the physical components from another slice due to a lack of isolation. This type of compromise may result in data spills.

Network Slicing Threat impact "In Real Life."
Network slicing will be pivotal in emerging technology such as autonomous vehicles. However, the slice is used to connect to and communicate with the autonomous vehicle, presenting the opportunity for a malicious actor to conduct an International Mobile Subscriber Identity (IMSI) caching attack, degrading the performance, reliability, and robustness of the network services. Additionally, the actor can use IMSI caching to expose the autonomous vehicle's geolocation and information about the cargo, such as sender, destination, and traffic routes.

From here, the actor can launch a DoS attack on the network signaling plane to cause distraction between the autonomous vehicle and its authorized controller. In addition, assuming the malicious actor has access to the subscriber identity, the actor can also separately launch a configuration attack to tamper with the security features and virtual network function (VNF) policies. This can compromise authentication and authorization policies, thereby granting illicit access to the network slice. In addition, this type of attack can allow the malicious actor to expand unauthorized access to other network slices if the security between the slices is not adequate.

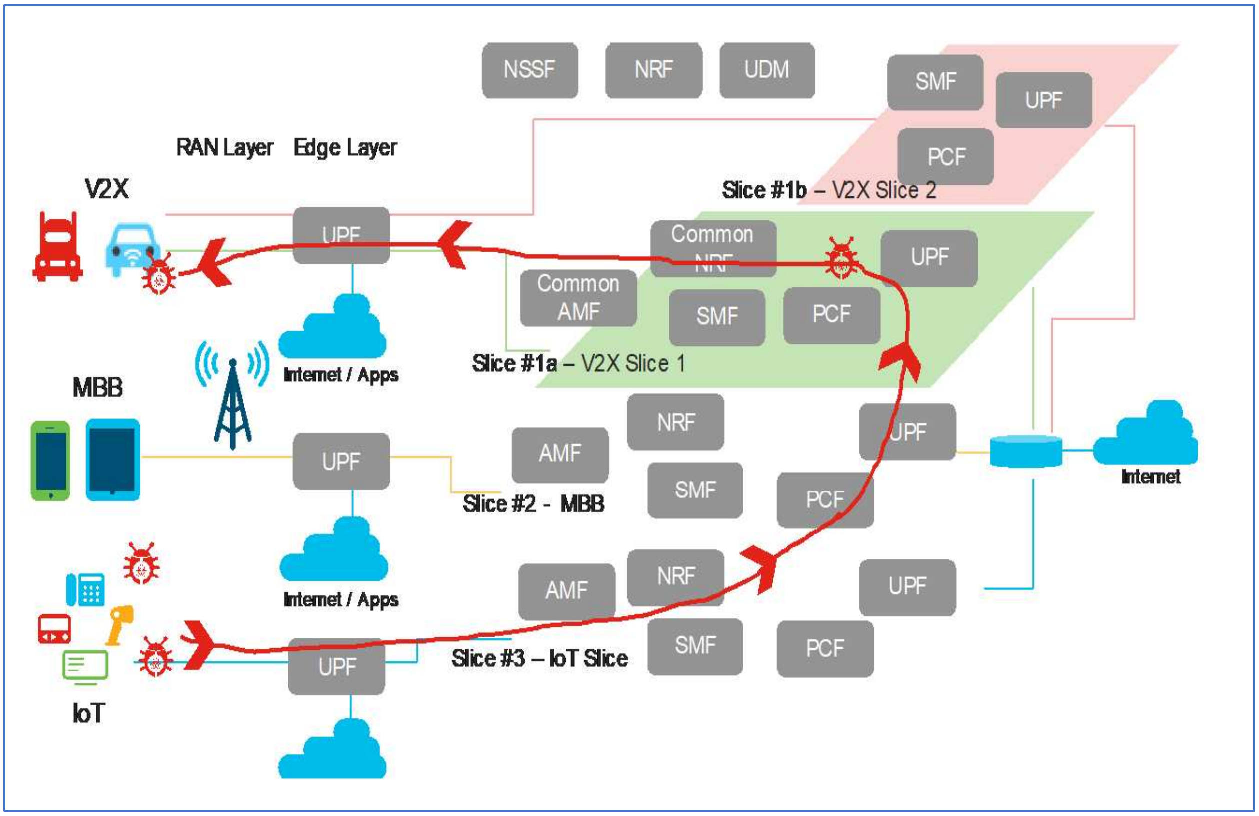


FIGURE 66: 5G SEGMENTATION THREAT IN MOTION [48]

Image Source: 5G Americas

6.2 Access Network Threats [23]

6.2.1 Abuse of Spectrum Resources

The illegal use of these resources, due to the dynamic allocation/ reallocation of the same, may allow the occupation of a specific idle spectrum band by imitating the characteristics of a legitimately licensed unit and causing interference in radio frequencies. Furthermore, this illegal spectrum occupation may also induce a network node to reject spectrum resources requested by unlicensed teams - due to the apparent lack of idle resources – thus blocking someone out of the core network.

6.2.2 Address Resolution Protocol (ARP) Poisoning

This kind of attack is also called ARP cache spoofing: a technique by which an attacker sends spoofed ARP messages onto the network. Generally, the aim is to acquaintance the attacker's MAC address, including the IP address from another host, such as the default gateway, printer, NTP server, etc. This means forwarding continuous traffic from that IP address instead of the attacker.

6.2.3 Fake Access Network Node

Classified as a nefarious activity, this threat considers the compromise of a base station (gnB) by masquerading as legitimate, facilitating different types of attacks such as man-in-the-middle or network traffic manipulation. Additionally, the threat considers tampering with the communication between the mobile user equipment (UE) and the network to initiate other malicious actions.

6.2.4 Flooding Attack

This threat involves flooding radio interfaces with requests. Flooding occurs through data transmission that can exhaust component resources and lead to a reduction or complete shutdown of the radio frequency provided by the component.

6.2.5 IMSI Catching Attacks

This threat relates to cellular paging protocols that a malicious actor can exploit in the vicinity of a victim to associate the victim's soft identity (e.g., phone number, Twitter handle) with its paging occasion. For example, through an attack dubbed '$mathsf{ToRPEDO}$,' a malicious actor can verify a victim's coarse-grained location information, inject fabricated paging messages, and mount denial-of-service attacks.

6.2.6 Jamming the Radio Frequency

Classified as a nefarious activity/abuse of assets, this threat refers to an intentional disruption/interference of the network radio frequency (NRF), causing the core network (and related services) to become unreachable for affected users. The threat also refers to the unavailability of the transport layer when using radio-based networks and interference with the geo-positioning system (GPS).

6.2.7 MAC Spoofing

It is a technique for changing a networked device's factory-assigned Media Access Control (MAC) network interface address. The hard-coded MAC address on a network interface controller (NIC) generally cannot be changed. However, many drivers permit the MAC address to be changed. Additionally, some tools force an operating system to believe that the user has chosen to change the MAC address in NIC. This masking process of a MAC address is known as MAC spoofing. MAC spoofing entails changing a computer's identity to conduct an attack.

6.2.8 Manipulation of Access Network Configuration Data

This threat involves compromising an access network element (e.g., base stations) to forge configuration data and launch other attacks (e.g., DoS).

6.2.9 Radio Interference

A threat in which the offender seeks to make a network resource unavailable to its deliberate users by temporarily or indefinitely interfering or disrupting the Radio Access Network service. Introducing compromised 5G devices in a radio access network will present a more substantial DoS threat.

6.2.10 Radio traffic Manipulation

This threat considers the manipulation of network traffic at the base station level. A man-in-the-middle attack can be launched based on a rogue base station when a malicious actor masquerades its Base Transceiver Station (BTS) as a real network's BTS. This threat is still considered valid due to backward compatibility with previous generations of mobile technology. Other associated threats follow:

- Traffic redirecting

6.2.11 Session Hijacking

This threat is classified as nefarious activity or abuse of assets and relates to attacks on open-air interfaces. The threat considers the theft of legitimate authenticated conversation session ID by a malicious actor to control the whole session of specific traffic to conduct other types of attacks.

6.2.12 Signalling Fraud

One of the areas of concern is the international signaling interconnection between networks which may be misused for fraud (e.g., false charging). Another example is the threat of greedy mobile nodes that transmit fake incumbent signals and force all other users to vacate a specific band (spectrum hole) to acquire its exclusive use.

6.2.13 Signalling Storms

'Signaling storms' are launched by malware or apps in the mobile networks, which overload the bandwidth at the cell, the backbone signaling servers, and Cloud servers and may also deplete the battery power of mobile devices. In addition, signaling storms will be more challenging due to the excessive connectivity of UEs, small base stations, and high user mobility.

6.3 Generic Threats [23]

6.3.1 Data Breach, Leak, Theft Destruction, and Manipulation of Information

This includes the theft of personal information through unauthorized access to the systems and network, unauthorized access to and possible publication of personally identifiable information/biometric/medical (privacy breach), confidential company information (intellectual property, commercial and financial data), or government/state-related information (classified information). In addition, the theft, breach, or leak of other data types, such as user credentials, encryption keys, network security logs, software configuration, etc., may also help malicious actors conduct attacks.

6.3.2 Eavesdropping

Eavesdropping is a threat in which the perpetrator seeks to tamper with the application and communication layers from the various 5G network elements (SDN controller, network function, edge node, virtualization orchestrator). It includes eavesdropping on subscriber data, confidential information, system time, subscriber location, electronic messages, and data signal relayed over the network. In addition, the threat actor monitors, spies, and eavesdrops on Nation-State citizens and organizations to track the location or access sensitive information.

6.3.3 Exploitation of Software and Hardware Vulnerabilities

This type of threat enables a malicious actor to take advantage of unknown (to the vendor and user) or unpatched software or hardware flaws to perform an attack. Examples include the exploitation of known hardware and software flaws such as meltdown, specter, and buffer overflow. It also provides for the exploitation of other known vulnerabilities related to previous generations of mobile telecommunications and older signaling protocols such as SS7 (Signalling System 7) and Diameter.

6.3.4 Malicious Code or Software

The threat includes installing and distributing malicious software or implanting specific code or software inside a product or updates. Examples of malicious software include malware, ransomware, virus, worms, trojans, SQL injections, rogue security software, rogueware, and careware. An example of malicious software in the 5G context considers using an unauthorized VNF that could abusively install and register itself into the core network to expose malicious APIs.

6.3.5 Compromised Supply Chain, Vendor, and Service Providers

This threat considers the intentional insertion by a vendor into the product of concealed hardware, malicious software, and software flaws. It also considers the implementation of uncontrolled software updates, manipulation of functionalities, and the inclusion of functions to bypass audit mechanisms, backdoors, and undocumented testing features left in the production version, among others.

This threat also relates to activities performed by untrustworthy third-party personnel during product testing, maintenance, configuration, and operation. For example, third-party personnel can access the network management facilities (locally and via remote interface) to perform maintenance activities and provide technical support. This privileged access to the network's operation, administration, and management (OAM) provides an advantage to untrustworthy third parties' personnel to access various types of data such as (subscribers, system and network configuration, and telemetry data).

6.3.6 Targeted Threats

Highly sophisticated attacks or advanced persistence threats may target sensitive information, e.g., state secrets, industrial secrets, intellectual property, or the availability of sensitive and critical services.

6.3.7 Exploiting Flaws in Security, Management, and Operational Procedures

Not directly related to 5G, this threat will become relevant when dealing with the complexity of the technology and the need to introduce operational procedures to the management of the Network. This threat includes but is not limited to the exploitation of flaws in the active and security control of the network, configuration, update, and patch management of the software. Additionally, the errors from the lack of poorly designed operating and safety procedures may have consequences on the integrity and availability of the network.

6.3.8 Abuse of Authentication

This threat may affect network entry points, such as user equipment (mobile devices and IoT), operation and management interfaces, roaming, and vertical services. This threat includes the theft of user credentials, brute force of user accounts, password cracking, masking the user identity, and impairment of an IoT grouping authentication as techniques used by threat actors to abuse the 5G authentication systems.

6.3.9 Identity Theft or Spoofing

This threat may materialize when a malicious actor successfully determines the identity of a legitimate entity and then masquerades to launch further attacks. Identity spoofing is a threat that can affect any software component or human agent. In this attack, the attacker spoofs the identity of a legitimate controller and interacts with the network functions controlled by the legitimate controller (i.e., elements of the data plane) to trigger several other types of attacks (instigate network flows, divert traffic, etc.). In addition, social engineering and brute force user account/password cracking may also be used to spoof or steal user credentials.

6.4 List of Threats

As per ENISA Threat Landscape for 5G Networks document, the below-listed threats are identified.

6.4.1 Nefarious Activity/ Abuse of Assets (NAA)

Threats	Potential Impact	Affected Assets	
Manipulation of network configuration/data forging - Routing tables manipulation - Falsification of configuration data - DNS manipulation - Manipulation of access network and radio technology configuration data - Exploitation of misconfigured or poorly configured systems/networks - Registration of malicious network functions	- Information integrity - Information destruction - Service unavailability	- SDN, NFV, MANO - RAN, RAT	- System configuration data - Network configuration data - Security configuration data - Business services
		- SDN, NFV, MANO	
Exploitation of software, hardware vulnerabilities		- RAN, RAT	
- Zero-day exploits		- MEC	- Subscribers' data
- Abuse of edge open application programming interfaces (APIs)	- Information integrity	- API	- Application data
- Application programming interface (API) exploitation	- Information destruction	- Physical infrastructure	- Security data
	- Service unavailability	- Business applications	- Network data
		- Security controls	- Business services
		- Cloud, virtualisation	
Denial of service (DoS) - Distributed denial of service (DDoS) - Flooding of core network components - Flooding of base stations - Amplification attacks - MAC layer attacks - Jamming of the network radio - Edge node overload - Authentication traffic spikes	- Service unavailability - Outage	- SDN, NFV - RAN, RAT - MEC - CLOUD	- Network services - Business services

Remote access exploitation	- System integrity	- SDN, NFV, MANO - CLOUD	- Network services
Malicious code/software - Injection attacks (SQL, XSS) - Virus - Malware - Rootkits - Rogueware - Worms/trojan - Botnet - Ransomware	- Service unavailability - Information integrity - Information destruction - Other software asset integrity - Other software asset destruction	- Data network - Business applications - Security controls - Cloud, virtualisation	- Subscribers' data - Application data - Security data - Network data - Business services - Network services
Abuse of remote access to the network	- Information integrity - System integrity	- SDN, NFV - RAN, RAT	- Subscribers' data - Application data - Security data - Network data
Abuse of information leakage - Theft and/or leakage from network traffic - Theft and/or leakage of data from cloud computing - Abuse on security data from audit tools - Theft/breach of security keys	- Information integrity - Information destruction - Information confidentiality	- Data storage/repository - - Subscribers' data - Cryptographic keys - Monitoring data - User subscription profile data	

Abuse of authentication			- Subscribers' data
- Authentication traffic spikes	- Information integrity	- Security data	- Application data
- Abuse of user authentication/authorization data by third parties' personnel	- Information destruction	- Network service	- Security data
	- Service unavailability		- Network data
Lawful interception function abuse	- Information integrity - Information destruction	- Subscribers' data - User subscription profile data	
Manipulation of hardware and software			
- Manipulation of hardware equipment			
- Manipulation of the network resources orchestrator		- Cloud data center equipment	
- Memory scraping		- User equipment	
- MAC spoofing	- Service unavailability	- Radio access/units	- Subscribers' data
- Side channels attacks	- Information integrity	- Light data centers	- Network services
- Fake access network node	- Information destruction	- SDN, MANO, NF	
- False or rogue MEC gateway		- RAN, RAT	
- UICC format exploitation		- Virtualisation	
- User equipment compromising			
Data breach, leak, theft and manipulation of information	- Information integrity - Information destruction - Information confidentiality	- Subscribers' data - Subscriber geo locations - Financial data - Commercial data, IP - Configuration data - Service data - Network data	

Unauthorized activities/network intrusions	- Information integrity		- Network services
- IMSI catching attacks	- System integrity	- User equipment	- Business services
- Lateral movement			
Identity fraud/account or service	- Service unavailability	- User subscription profile data	
- Identity theft	- Information destruction	- Subscribers' data	
- Identity spoofing	- Information integrity		
		- RAT	
Spectrum sensing	- Service unavailability	- Radio access units	
Compromised supply chain, vendor and service providers - Threat from third parties' personnel accessing MNO's facilities	- Service unavailability - Information integrity - Information destruction	- SDN, NFV, MANO - RAN, RAT - MEC - API - Physical infrastructure - Business applications - Security controls - Cloud, virtualization	- Network services - Business services
Abuse of virtualization mechanisms			
- Network virtualization bypassing	- Service unavailability	- Virtualization	
- Virtualized host abuse	- Information integrity	- SDN, NFV, MANO	- Network services
- Virtual machine manipulation	- Information destruction	- Cloud	- Business services
- Data center threats			
- Abuse of cloud computational resources			
Signaling threats	- Service unavailability	- RAT	- Network services
- Signaling storms	- Information integrity	- Radio access units	- Business services
- Signaling fraud	- Information destruction	- Protocols	

TABLE 7: NEFARIOUS ACTIVITY/ ABUSE OF ASSETS (NAA)

6.4.2 Eavesdropping/Interception / Hijacking (EIH)

Threats	Potential Impact	Affected Assets	
Nation-state espionage	- Information integrity - Information Confidentiality	- Subscribers' data - Subscriber geo locations	
Corporate espionage	- Information integrity - Information Confidentiality	- Financial data - Commercial data - IP	
Traffic sniffing	- Information integrity - Information Confidentiality	- Data traffic - Subscribers' data - Subscriber geolocation	
Manipulation of network traffic, network reconnaissance, and information gathering - Radio network traffic manipulation - Malicious diversion of traffic - Traffic redirecting - Abuse of roaming interconnections	- Information integrity - Information Confidentiality	- Data traffic - Subscribers' data - Subscriber geo locations	
Man in the middle/ Session hijacking	- Information integrity - Information Confidentiality	- Data traffic - Subscribers' data - Subscriber geo locations	
Interception of information	- Information integrity - Information Confidentiality	- Data traffic - Subscribers' data - Subscriber geo locations	

TABLE 8: EAVESDROPPING/ INTERCEPTION/ HIJACKING (EIH)

6.4.3 Physical Attacks (PA)

Threats	Potential Impact	Affected Assets	
	- Service unavailability	- Radio access units	
Sabotage of network infrastructure (radio access, edge servers, etc.)	- Information Destruction - Information integrity	- ICT equipment - Light data center - Cloud data center	- Network services - Business services
	- Service unavailability	- Radio access units	
Vandalism of network infrastructure (radio access, edge servers, etc.)	- Information Destruction - Information integrity	- ICT equipment - Light data center - Cloud data center	- Network services - Business services
Theft of physical assets	- Service unavailability - Information Destruction - Information integrity	- Radio access units - ICT equipment - Light data center - Cloud data center	- Network services - Business services
The terrorist attack against network infrastructure	- Service unavailability - Information Destruction - Information integrity	- Radio access units - ICT equipment - Light data center - Cloud data center	- Network services - Business services
Fraud by MNO employees	- Service unavailability - Information Destruction - Information integrity	- Radio access units - ICT equipment - Light data center - Cloud data center	- Network services - Business services
Unauthorized physical access to based stations in shared locations	- Service unavailability - Information Destruction - Information integrity	- RAT - Radio access units	- Network services - Business services
Misconfigured or poorly configured systems/networks	- Service unavailability - Information integrity	- Management process - Policies	- SDN, NFV, MANO, API - RAN, RAT, MEC - Physical infrastructure

TABLE 9: PHYSICAL ATTACKS (PA)

6.4.4 Unintentional damages (accidental) (UD)

Threats	Potential Impact	Affected Assets	
Inadequate designs and planning or lack of adaption			- SDN, NFV, MANO
- Outdated system or network from the lack of update or patch management - Errors from the lack of configuration change management - Poorly design network and system architecture	- Service unavailability - Information integrity	- Management processes - Policies - Human assets	- RAN, RAT - MEC - API - Physical infrastructure - Business applications - Security controls - Cloud, virtualisation
			- SDN, NFV, MANO
		- Management processes	- RAN, RAT
Erroneous use or administration of the network, systems and devices	- Service unavailability	- Policies	- MEC, UE, API
	- Information integrity	- Human assets	- Physical infrastructure
			- Business applications
			- Security controls
			- Cloud, virtualisation
Information leakage/sharing due to human error	- Information integrity - Information confidentiality	- Data storage/repository - Management processes - Policies - Legal - Human assets	- Subscribers’ data - Application data - Security data - Network data
	- Information integrity	- Management processes	- Subscribers’ data
Data loss from unintentional deletion	- Information confidentiality	- Policies	- Application data
		- Human assets	- Security data
			- Network data

Table 10: Unintentional Damages (Accidental) (UD)

6.4.5 Failures or Malfunctions (FM)

Threats	Potential Impact	Affected Assets	
Failure of the network, devices or systems	- Service unavailability - Information destruction - Information integrity	- Cloud data center - User equipment - RAT, Radio unit - Light data center	- Network services - Business services
Failure or disruption of communication link	- Service unavailability - Information destruction - Information integrity	- Cloud data center	- Network services - Business services
Failure or disruption of main power supply	- Service unavailability - Information destruction - Information integrity	- Cloud data center	- Network services - Business services
Failure or disruption from service providers (supply chain)	- Service unavailability - Information destruction - Information integrity	- Network services - Business services	
Malfunction of equipment (devices or systems)	- Service unavailability - Information destruction - Information integrity	- Radio access units - ICT equipment - Light data center - Cloud data center	- Network services - Business services

TABLE 11: FAILURES OR MALFUNCTIONS (FM)

6.4.6 Outages (OUT)

Threats	Potential Impact	Affected Assets	
Loss of resources - Human resources - Physical resources	- service unavailability - Information Destruction - Information integrity	- human assets - Legal	- network services - Business services
Support services	- Service unavailability - Information Destruction - Information integrity	- human assets - Managementprocesses - Policies - Legal	- Network services - Business services
Data network (access)	- service unavailability - Information Destruction - Information integrity	- Cloud data center	- network services - Business services
Power supply	- service unavailability - Information Destruction - Information integrity	- Cloud data center	- network services - Business services

TABLE 12: OUTAGES (OUT)

6.4.7 Disasters (DIS)

Threats	Potential Impact	Affected Assets	
Natural disasters - Earthquakes - Landslides	- service unavailability - Information Destruction - Information integrity	- radio access units - ICT equipment - Light data center - Cloud data center	- Network services - Business services
Environmental disaster - Floods, storms - Pollution, dust, corrosion - Fires, **heavy winds** - Unfavorable climatic conditions	- Service unavailability - Information Destruction - Information integrity	- radio access units - ICT equipment - Light data center - Cloud data center	- Network services - Business services

TABLE 13: DISASTERS (DIS)

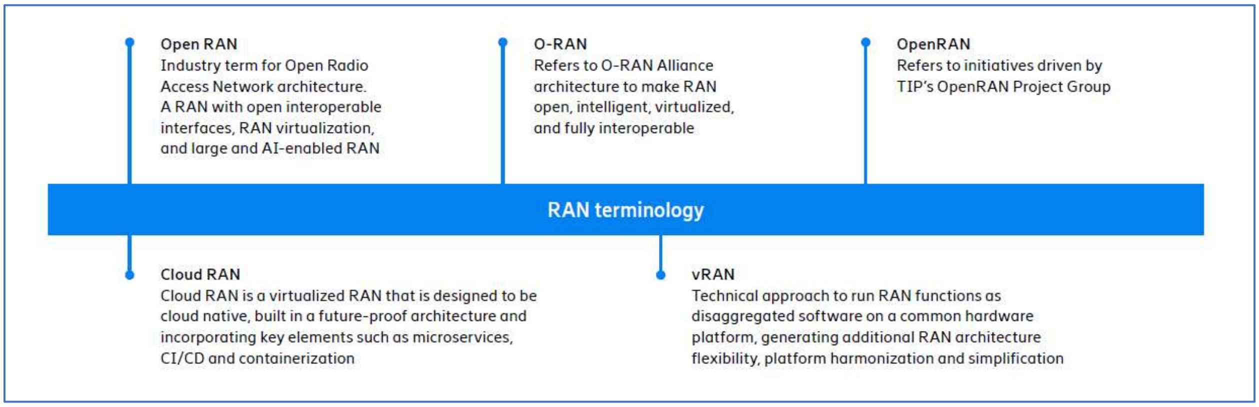
6.4.8 Legal (LEG)

Threats	Potential Impact	Affected Assets
Breach of service level agreement (SLA)	- Services unavailable - Information Destruction - Information Destruction & Integrity	- Network services - Business services
Breach of legislation	- service unavailability - Information Destruction - Information integrity	- network services - Business services
and	- service unavailability - Information Destruction - Information integrity	- network services - Business services

TABLE 14: LEGAL (LEG)

SECTION 7. ANALYZING POSSIBLE SOLUTIONS

7.1 RAN Terminology Comparison



Open RAN
Industry term for Open Radio Access Network architecture. A RAN with open interoperable interfaces, RAN virtualization, and large and AI-enabled RAN

O-RAN
Refers to O-RAN Alliance architecture to make RAN open, intelligent, virtualized, and fully interoperable

OpenRAN
Refers to initiatives driven by TIP's OpenRAN Project Group

Cloud RAN
Cloud RAN is a virtualized RAN that is designed to be cloud native, built in a future-proof architecture and incorporating key elements such as microservices, CI/CD and containerization

vRAN
Technical approach to run RAN functions as disaggregated software on a common hardware platform, generating additional RAN architecture flexibility, platform harmonization and simplification

FIGURE 67: RAN TERMINOLOGY

Image Source: Ericsson

The 5th generation (5G) of cellular communications guarantees faster and more trustworthy communications, with high-bandwidth and real-time capabilities that will offer immense potential by enabling new use cases. However, the future of 5G is more than just the following high-speed mobile network. The use of cloud computing, Artificial Intelligence - AI/ Machine Language - ML, Augmented Reality - AR, Virtual Reality - VR, and the connected billions of devices will push the borders of wireless communications. As a result, Mobile Network Operators (MNO) are looking for ways to adopt open, virtualized, and cloud-based Radio Access Networks (RAN) that will allow them to achieve greater network flexibility, reliability, and the ability to quickly implement new service types as 5G use cases are discovered. To realize these 5G benefits, MNOs are moving away from traditional, proprietary RANs that use purpose-built hardware and software to open hardware and software-based ecosystem called Open RAN and Cloud-based Radio Access Systems called C-RAN.

7.1.1 O-RAN

Comparing these two RAN technologies, Open RAN has an expanded attack surface due to its additional functions, interfaces, and cloud deployment models. Open RAN deployments in public and hybrid clouds also have a raised threat surface due to increased risk from internal threats and APTs, as shown in Figure 68. ENISA's NIS Cooperative Report on Open RAN Cybersecurity identified 5G Core and Open RAN as having increased risk of internal threats in the cloud due to:

- greater dependency on cloud service providers
- lack of defined security roles across stakeholders
- resource sharing with other tenants
- broader use of open-source software
- use of insecure third-party hardware

Deployment in an MNO's on-premises private cloud can reduce risks from these threats. However, the German Federal Office of Information Security (BSI) Open RAN Risk Analysis [12] identified high risks in O-

RAN deployments due to the optional use of critical security controls, specification of weak protocols and cipher suites, assumptions of internal trust, missing protections from denial-of-service attacks, and lack of cloud security controls for O-Cloud.

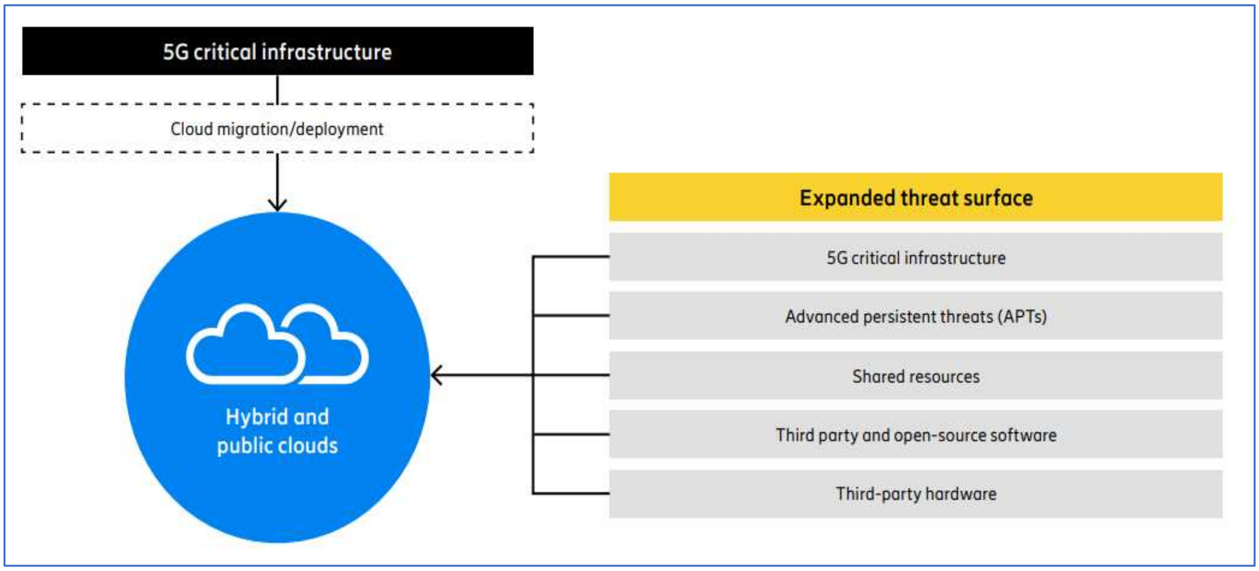

5G critical infrastructure
Cloud migration/deployment
Hybrid and public clouds
Expanded threat surface
5G critical infrastructure
Advanced persistent threats (APTs)
Shared resources
Third party and open-source software
Third-party hardware

FIGURE 68: EXPANDED THREAT SURFACE FOR 5G CLOUD DEPLOYMENTS

Image Source: Ericsson

7.1.2 C-RAN

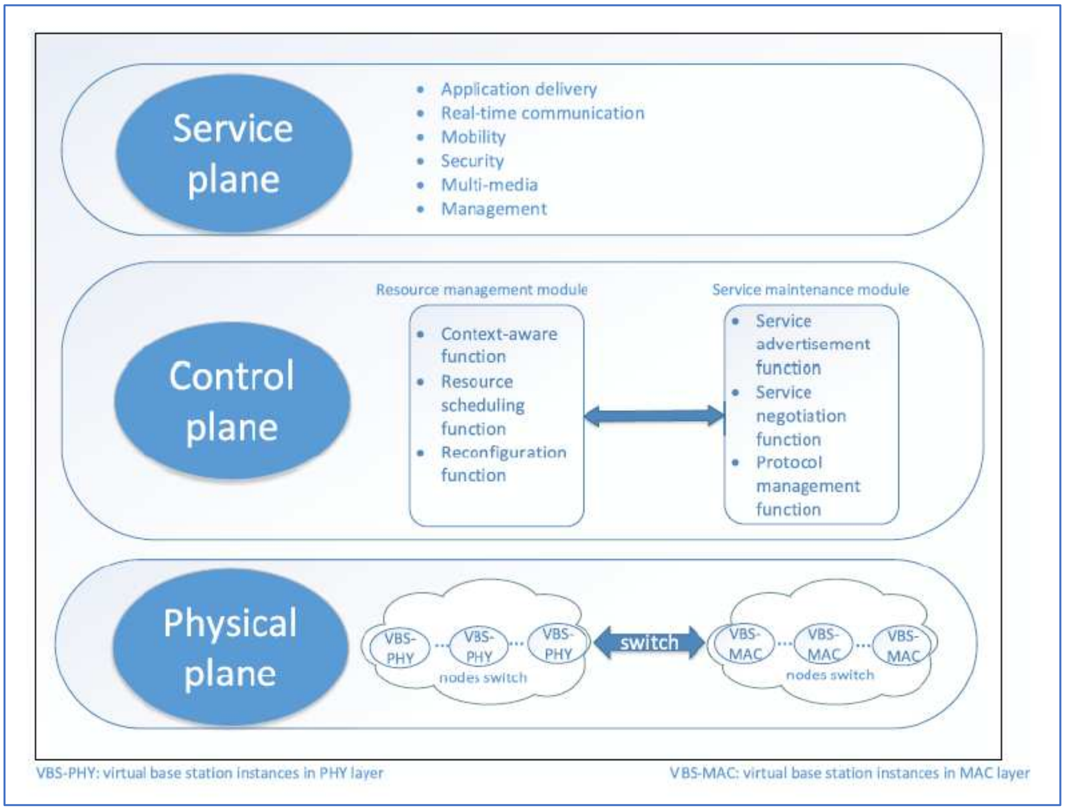

Service plane
Application delivery
Real-time communication
Mobility
Security
Multi-media
Management
Control plane
Resource management module
Context-aware function
Resource scheduling function
Reconfiguration function
Service maintenance module
Service advertisement function
Service negotiation function
Protocol management function
Physical plane
VBS-PHY
VBS-PHY
VBS-PHY
nodes switch
switch
VBS-MAC
VBS-MAC
VBS-MAC
nodes switch
VBS-PHY: virtual base station instances in PHY layer
VBS-MAC: virtual base station instances in MAC layer

FIGURE 69: C-RAN LOGICAL ARCHITECTURE [49]

C-RAN includes the physical, control, and service planes, as shown in Figure 69. It focuses on service-oriented cloud architecture, commerce, and personal resource scheduling and management. Based on the collaboration between the virtualized BBUs pool and RRHs, C-RAN has fewer network delays than other cellular networks. According to an LTE protocol stack, there are L1, L2, and L3 layers in C-RAN. Among them, L1 is the physical layer (PHY), which mainly provides a data transmission service to the higher layers, channel coding, rate matching, and Multiple Input Multiple Output (MIMO) technologies. L2 is the layer responsible for Media Access Control (MAC), Radio Link Control (RLC), and Packet Data Convergence Protocol (PDCR), which mainly provides data link control. L3 is the Radio Resource Control (RRC) layer, mainly providing signaling and radio resource control. In this section, we review and discuss existing solutions to resist the threats and attacks in the C-RAN communication system based on the C-RAN logical structure.

7.2: Solution of Security Threats and Vulnerabilities in C-RAN [49]

As we can see in Figure 69, the physical plane is mainly responsible for performing virtualized resource allocation, node switch (e.g., signal transmission and processing), and baseband pool interconnection based on the channel decoding technology, multi-point processing, Fast Fourier Transform (FFT), and so on. The safety of the physical plane is a foundation that guarantees a secure and reliable C-RAN system. Therefore, this plane has been a focus of security concern. The current work mainly focuses on overcoming the following attacks and threats.

7.2.1 Solutions to Overcome Threats Related to Physical Plane Security

i) Eavesdropping Attack

Massive MIMO technology has drawn operators' attention and will be integrated into 5G network architecture. It is one of the critical techniques of the C-RAN physical layer. Eavesdropping attack is a common problem in all RANs. Massive Multiple-Input Multiple-Output systems to physical security introduced two schemes for detecting the eavesdropping attack to prevent BS and channel estimation from passive eavesdropping and active attacks. In the first scheme, a legitimate user can generate an additional random phase-shift keying sequence, and the BS can effectively detect the eavesdropping attack through the received sequence. In the second scheme, the beamformer is adopted to detect the eavesdropping attack. Benefiting from the same beamformer (between a BS and an initial user), the BS transmits a pilot based on a received signal to the initial user. The initial user can compare this pilot with a previously agreed value (between the BS and the initial user). This value will change when the eavesdropper forges and modifies the original information to BS.

ii) Jamming Attack

It is also called a DoS attack, which means a malicious node interferes with other network nodes' radio frequencies by sending out white noise or useless network traffic signals. The most common threats in wireless sensor networks are spot jamming, sweep jamming, barrage jamming, and deceptive jamming. The jamming attacks are summarized as four possible jamming goals:
(1) through an immediate DoS attack to block user access to the radio network nodes.
(2) occupying most of the spectrum and leaving a small portion of the spectrum to degrade core network functions.
(3) learning the core network's defense strategy to achieve the next attack.
(4) herding of a jammer by attacking a radio network node concert with other malicious jammers.

iii) Impersonation Attack

In the traditional radio network architecture, there are two main types of impersonation attacks: Cognitive Radio (CR) node impersonation attack and primary user impersonation attack. In the first type, assuming that a CR node is attacked, it can cooperate with other attacked nodes and provide false information (e.g., idle spectrum and user geographical location) to a normal node. Furthermore, it may even refuse services to achieve selfish aims or damage the core network. However, due to its open nature, any CR nodes, illegitimate or not, can access the core network functions in the C-RAN architecture.

7.2.2 Solutions to Overcome Threats Related to Control Plane Security

The control layer of C-RAN is divided into two modules: the service maintenance module and the resource management module. The resource management module is responsible for resource allocation and distribution with context awareness. The service maintenance module contains service advertisement, negotiation functions, and protocol management (e.g., Quality-of-Service management, common control channel, spectrum resource allocation, MAC, network layer protocol management, etc.). Physical plane security is the precondition to guarantee a secure and reliable C-RAN system. However, control plane security is the core of C-RAN security. Current research in control plane security focuses on the following aspects.

1) Network and Mac Layer Protocol Attacks

Network layer protocol attacks include IP spoofing, hijacking, and Smurf attacks. Targeting the MAC layer attacks, previous literature often proposes novel cognitive radio MAC protocols to improve radio nodes' cognitive ability and security in a distributed cognitive radio architecture. The application scenarios, features, advantages, and disadvantages of the standard cognitive radio MAC protocols and divided into three classes: random access protocols, time-slotted protocols, and hybrid ones. An opportunistic spectrum MAC protocol was proposed to protect MAC layer security. This protocol can adaptively and dynamically seek and utilize available spectrum bands of licensed and unlicensed spectra. Different users can access and share these resources. Licensed and unlicensed users can mutually cooperate. However, to ensure that different users can communicate with each other, this protocol depends on a trusted third authoritative party to divide the available spectrum into a secure common control channel and multiple secure data channels.

2) Common Control Channel Threats

A common control channel differs from a band channel in that the former uses a predefined frequency channel to send or receive information (e.g., collaborative processing requests, spectrum resource state, channel negotiation information, etc.), which is very important for operators. A common control channel faces three threats: (1) MAC spoofing since most current cognitive radio networks lack a model that can authenticate data integrity spread to every node; (2) extended DoS threats; (3) jamming attacks.

In DoS attacks, attackers can exploit the control channel saturation problem to attack the common control channel and impair its functions (e.g., resource allocation function). Furthermore, regarding selfish misbehaviors, a selfish CR node impairs the standard channel negotiation process by disrupting data packet forwarding, which causes false channel information (e.g., about channel availability).

3) IEEE 802.22 Specific Threats

In 2006, IEEE 802.22 was designed as the first MAC layer confidentiality and authentication standard IEEE 802.22 added new air interfaces based on the Wireless Area Network (WAN). The common threats that IEEE 802.22 faces include DoS attacks, replay attacks, special jamming attacks, PUEAs, and wireless microphone beacons. Besides, a secure sub-layer based on the IEEE 802.22 standard consists of an encapsulation protocol and a privacy-preserving key management protocol. However, the secure sub-layer needs an effective solution to generate, manage and distribute related keys.

4) Radio Spectrum Resource Threats

Compared to the traditional radio wireless network's one-to-one architecture, the C-RAN architecture uses distributed RRHs and centralized virtual BBU pool management, which is more vulnerable regarding spectrum security. For example, a malicious user or node selfishly uses unauthorized spectrum resources to induce much traffic and occupy bandwidth or exploits this to generate a DoS attack on others.

Through maximizing the various modules of C-RAN (e.g., virtual BBU pool, user groups, RRHs, and transmit beamforming, etc.), experts propose two algorithms. One is a dynamic user-centric scheduling algorithm for solving the imbalance between users' traffic and their non-uniform geographical locations. The other is the transmit beamformer optimization algorithm to achieve an optimal allocation between each user to maximize QoS and each RRH's maximized capacity load. By applying both algorithms, the security performance and utility of the C-RAN system can be improved with sound QoS. However, this approach needs to collect user personal information but does not consider user privacy.

5) SSDF Attack

Among radio spectrum resource threats, the most widely researched one is an SSDF attack, in which malicious users disturb the accuracy of collaborative spectrum sensing and resource allocation by sending error observations in a CRN environment. In the C-RAN architecture, to a certain extent, the centralized virtual BBU pool defends against this attack by uniformly observing and processing the spectrum signals that remote RRHs sense. However, the SSDF attack seriously affects the system spectrum resource allocation balance, especially for the virtual BBUs pool. The cooperative spectrum sensing, and resource allocation technique is a standard method to prevent attacks.

Several factors, such as signal-to-noise ratio, signal-to-interference ratio, the number of secondary users, sample correlation, etc., for reducing secondary user interferences in a collaborative spectrum sensing process. The weighted Sequential Probability Ratio Test (WSPRT) is a very effective way to prevent the SSDF attack.

7.2.3 Solutions to Overcome Threats Related to Service Plane Security

The service plane of the C-RAN architecture is a cloud platform that directly interacts with the users or service providers. For example, with the service plane, end users only consider the QoS problem but need to be made aware of who is the service provider. The service provider must only meet users' requirements regardless of their identities. Recently, the service plane's safety has attracted increasing attention due to its importance. In C-RAN, the service layer should prevent the cloud infrastructure and the virtual BBUs pool from invasion and provide security functionalities such as identity authentication, access control, etc. Recent security research in the service layer focuses on overcoming the following attacks and threats.

1) Transport and Application Layer Protocol Attacks

The application delivery service mainly involves relevant protocols in the transport and application layers. The transport or application layers attacks include TCP/UDP flooding attacks, sequence number prediction attacks, SQL injection, FTP bounce attacks and SMTP attacks, and so on. This is like the traditional wireless network. Thus, the existing solution can be applied to resolve this issue.

2) Cloud Computing Security Threats

The most significant difference between traditional RAN and C-RAN is that cloud computing is applied in C-RAN. Thus, it is essential to consider cloud computing security problems. For example, a security risk could occur when multiple base stations share a resource (e.g., service, hardware, data storage, etc.) over the cloud. In addition, the C-RAN architecture applies to cloud computing-related technology (e.g., virtualization technology, cloud storage, real-time data analysis, process, etc.), which brings new security threats and challenges.

This report summarized the opportunities, solutions, and progress of cloud security and privacy research in recent years, such as data storage and management security, access control, trust management, etc. The threats and security challenges of the cloud system and defined the basic requirements for building a secure and trustworthy cloud system:

a) outsourcing security that the cloud provider shall be trustworthy by providing trust and privacy protection, and they should ensure the confidentiality and integrity of the outsourced data.
b) multi-tenancy security that the shared cloud platform should ensure the safety of resource allocation in a virtualized environment.
c) massive data and intense computation security that it is necessary to design new strategies and protocols to satisfy massive data and low computation.

Cloud Security Alliance (CSA) proposed nine security threats to cloud computing. For C-RAN, the following security threats should be seriously considered:

Data loss and leakage, shared technology issues, abuse and nefarious use of cloud services, and Distributed Denial of Service (DDoS) attacks. One example attack is a hacker who can steal other virtualized machines' private keys from one virtualized machine. Besides, virtualized BBUs are responsible for handling cloud services, user data, spectrum allocation, and so on based on hardware resources. Therefore, once the virtualized BBUs pool is attacked, the core network performance will be significantly influenced, which may lead to severe damage and economic loss.

3) Virtualization Threats

One of the leading technologies applied in cloud computing is virtualization. In the C-RAN architecture, virtualized BSs pool security is essential for the overall network architecture. For current common virtualization attacks (e.g., tampering with guest or host machine, virtual machine covert channel, virtual machine-based rootkits, and Virtual Machine Manager (VMM) attacks), they summarized four defense methods: virtual machine-based intrusion detection, virtual machine-based kernel protection, virtual machine-based access control, and virtual machine-based trusted computing.

4) Privacy Threats

The privacy of users is easily attacked. For example, in C-RAN application scenarios, idle spectrum resources are expected to be allocated to users based on their geographic locations. In this process, users' private information (e.g., personal affairs, personal information, personal domain, etc.) may be leaked to unauthorized parties. Thus, mobile user privacy should be considered, especially when a user is served by a cloud computing service that cannot be fully trusted.

5) Other Security Threats
For C-RAN, some studies explored the cloud platform itself (e.g., OpenStack, cloud stack, etc.) to improve the security of the whole architecture. In this attack, the attackers hack into a computer node and get the administrator privileges of the virtual machine deployed on the node. As a result, they can steal all tenant tokens and the administrator rights of the whole platform. To resist this attack, the experts proposed a secure platform that supports freely designing a security policy to ensure secure interaction between different components and nodes.

7.3 Security Requirements of C-RAN [49]

The C-RAN system should satisfy the relevant security requirements to resist various threats and attacks. These requirements are also measured to compare existing security solutions (as shown in Table 15) and attempt to find open issues for directing future research trends. Finally, some security requirements are discussed in terms of cloud computing services.

7.3.1 Access Control to Resources (AC)

This is the most basic security requirement that a C-RAN system should fulfill. The system should forbid unauthorized users from accessing resources or services anytime and anywhere. It is an effective solution to fight against PUEAs (Primary User Emulation Attacks), privacy intrusions, and cognitive radio node impersonation attacks.

7.3.2 Robustness (Rb)

The C-RAN system should ensure not only the robustness of software or hardware resources but also guarantee the robustness of the cognitive radio channel for meeting the QoS of communication services required by users. For example, in some scenarios, the robustness of spectrum sensing should be enhanced when some sensing nodes easily malfunction. In addition, robustness is essential for overcoming the security threats caused by jamming, DoS, or DDoS attacks.

7.3.3 Confidentiality, Integrity, and Availability (CnInA)

No matter which kind of framework, one-to-one architecture or novel C-RAN architecture, confidentiality, integrity, and availability is commonly considered as three basic security properties. Integrity means that the system, the components of the system, and the data or information transmitted in the system are complete. Any data, such as user data and spectrum resources, should be confidential and available. In the C-RAN system, confidentiality requires data, no matter whether signal processing results, required cloud computing services, or user data uploaded to the virtualized BBUs pool, should have exclusiveness. Only authorized users can access or use these data. Integrity requires that the data associated with cloud computing is complete, adequate, and accurate and cannot be illegally manipulated, corrupted, tampered with, or forged. Finally, the availability of the C-RAN requires continuous and punctual data or services which is not interrupted or delayed.

7.3.4 Authentication (Au)

Authentication is a very effective way to overcome CR node and primary user impersonation attacks. By applying an authentication mechanism, the C-RAN system can verify who performs what, thus making it possible to detect fake CR nodes and malicious users. Moreover, discussing a new authentication mechanism to support authentication across domains and collaboration among multiple mobile operators to resist potential security threats when switching or accessing CRN is essential.

7.3.5 Privacy (Pr)

In C-RAN, the privacy of operators and end users should be considered. The privacy of end users can be divided into data privacy, identity privacy, and personal information privacy. Most Communication services are to gather data and personal information about end-users themselves, which may reveal information sensitive to their privacy. In addition, adversaries would further extract more personal information about workers, such as location information, trajectory, and preference.

7.3.6 Trustworthiness (Tr)

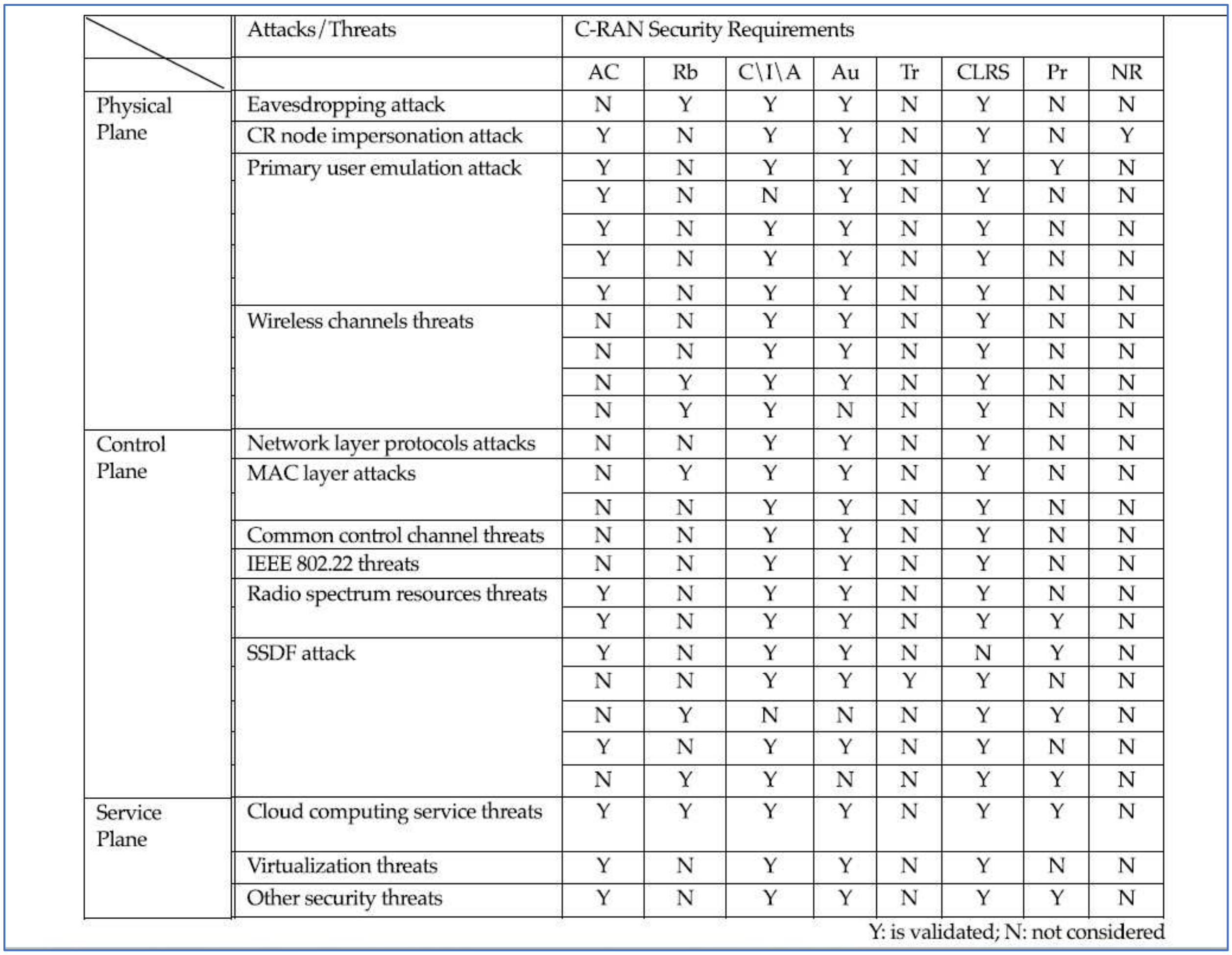
Compared to a traditional cellular network, the C-RAN communication environment is highly scalable, open, and heterogeneous. Many C-RAN usage scenarios are accomplished effectively through cooperation among mobile operators. Therefore, a trust management mechanism becomes crucial for trustworthy collaboration among the operators. Furthermore, overcoming virtualization or MAC layer-related threats may be a practical solution.

7.3.7 Compliance with Local Regulatory Standards (CLRS)

The C-RAN system should be designed to meet the regulatory standards of its local operator, which is a prerequisite for establishing a communication system. Furthermore, when the C-RAN architecture is deployed in a public place, it should meet all relevant security standards and requirements made by the Trans European Trunked Radio (TETRA) or the Association of Public-Safety Communications Officials (APCO).

7.3.8 Non-Repudiation (NR)

It is also called accountability. This is because the C-RAN system can verify any user’s actions, and this kind of action cannot be denied. Therefore, it is an effective solution to overcome the threats caused by impersonation attacks and radio spectrum attacks.



	Attacks/Threats	C-RAN Security Requirements							
		AC	Rb	C\I\A	Au	Tr	CLRS	Pr	NR
Physical Plane	Eavesdropping attack	N	Y	Y	Y	N	Y	N	N
	CR node impersonation attack	Y	N	Y	Y	N	Y	N	Y
	Primary user emulation attack	Y	N	Y	Y	N	Y	Y	N
		Y	N	N	Y	N	Y	N	N
		Y	N	Y	Y	N	Y	N	N
		Y	N	Y	Y	N	Y	N	N
		Y	N	Y	Y	N	Y	N	N
	Wireless channels threats	N	N	Y	Y	N	Y	N	N
		N	N	Y	Y	N	Y	N	N
		N	Y	Y	Y	N	Y	N	N
		N	Y	Y	N	N	Y	N	N
Control Plane	Network layer protocols attacks	N	N	Y	Y	N	Y	N	N
	MAC layer attacks	N	Y	Y	Y	N	Y	N	N
		N	N	Y	Y	N	Y	N	N
	Common control channel threats	N	N	Y	Y	N	Y	N	N
	IEEE 802.22 threats	N	N	Y	Y	N	Y	N	N
	Radio spectrum resources threats	Y	N	Y	Y	N	Y	N	N
		Y	N	Y	Y	N	Y	Y	N
	SSDF attack	Y	N	Y	Y	N	N	Y	N
		N	N	Y	Y	Y	Y	N	N
		N	Y	N	N	N	Y	Y	N
		Y	N	Y	Y	N	Y	N	N
		N	Y	Y	N	N	Y	Y	N
Service Plane	Cloud computing service threats	Y	Y	Y	Y	N	Y	Y	N
	Virtualization threats	Y	N	Y	Y	N	Y	N	N
	Other security threats	Y	N	Y	Y	N	Y	Y	N

Y: is validated; N: not considered

TABLE 15: C-RAN SECURITY THREAT AND REQUIREMENTS [49]

7.4 SDN Security and Solutions [52]

Software-defined networking (SDN) is an attractive incentive for operators due to keeping promises of agility, simplified control, and real-time programmability with advances in virtualization technologies. The aim is for this architecture to serve as a secure complement to cloud computing and to ensure that networks are protected from attack by malicious intruders.

Traditional Network architecture has reached the ability point to adapt to the dynamic environments enabled by virtualization technologies. SDN raises the level of system abstraction, opens the network programmability door, increases the speed of operations, and simplifies by separating the control plane from the data plane.

7.4.1 SDN Benefits and Vulnerabilities.

SDN accelerates the integration of security appliances into networks and implements a control panel directly rather than adding separate appliances or instantiated within multiple NEs. In addition, SDN's centralized management enables the events within the entire network to collect and aggregate, resulting in a broader, more coherent, and more accurate image of the network's status, making security strategies easier to enforce and vigilance.

Implementing the security mechanisms directly on top of the controller or steering traffic at a run time enables it dynamically by adding taps and sensors at various places in the network. This makes network monitoring more effective with an accurate picture of its status, and the network can detect more attacks, and the number of false positives reported can be reduced. For example, suppose a tap indicates to the SDNC (SDN Controller) that a botnet has hijacked any device. In that case, the SDNC can steer the potentially offending traffic to ids for analysis and monitoring. If the ids deem the traffic malicious, the SDNC can filter it and instruct the first hop aggressively.

To facilitate the collection of network-status information and enable automatic detection and resolution of any security breach, SDN is ideal for necessary integration into the network and SOC-Service Operation Centers. Unfortunately, the highly distinctive feature set of SDN also enables a larger attack surface than traditional networks – an issue documented in several recently published research papers.

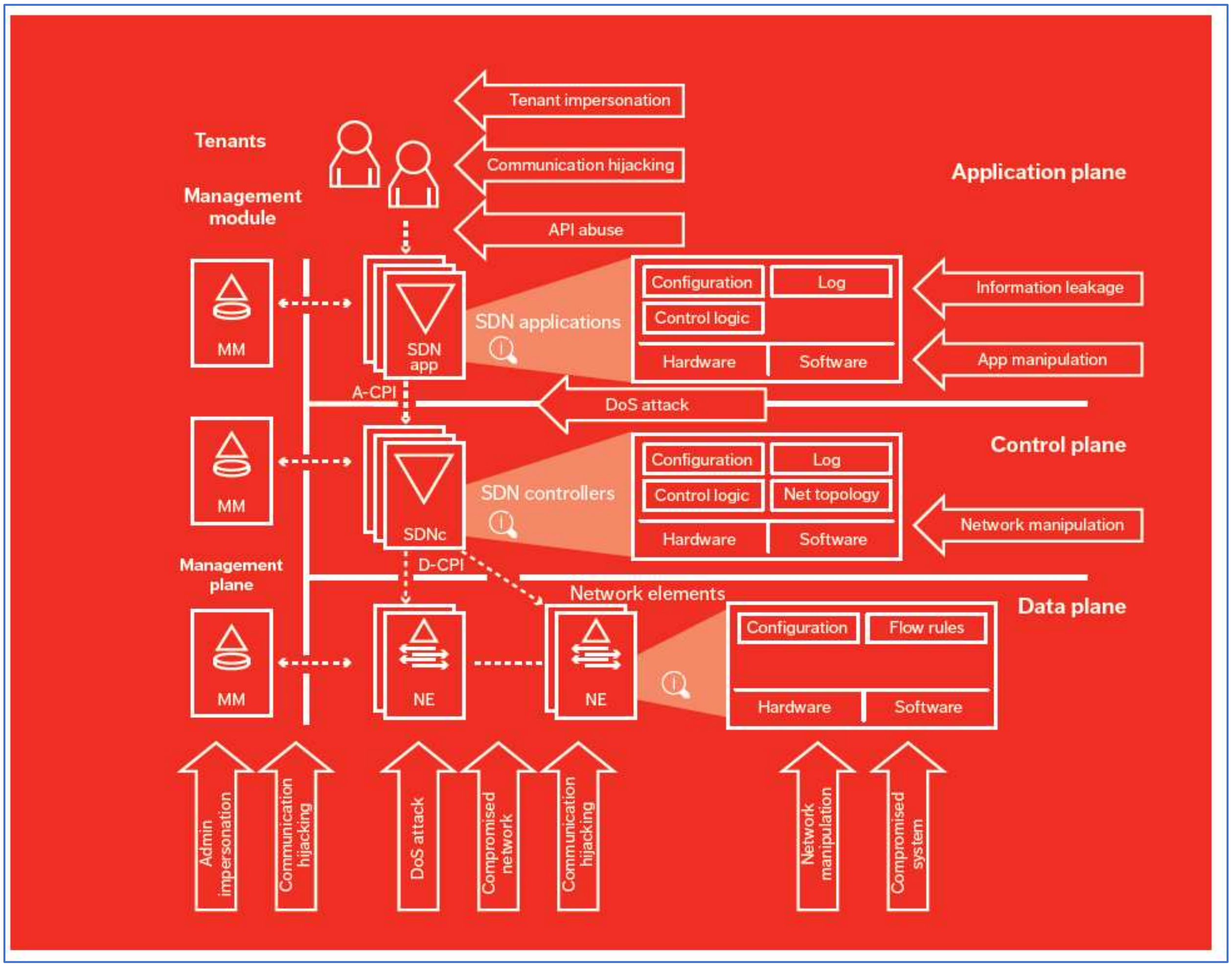


FIGURE 70: POTENTIAL VULNERABILITIES OF SDN ARCHITECTURE [52]

Image Source: Ericsson

7.4.2 SDN Security Solutions

While it sounds like several hurdles to overcome, the programmability and centralized management brought about by SDN enable a much greater autonomy to alleviate any security breaches.

a) Centralized Network Management

The NEs are monitored and managed individually in traditional networks. However, network management has become cumbersome without standard protocols interacting with all NEs irrespective of their vendor or generation. The SDN approach enables coordinated monitoring and control of network forwarding policies among the integrated NEs, which results in a more compliant management process.

Though the SDN control plane has a bottleneck risk, its outline of the entire network makes it dynamically competent to mitigate any reported incident. For example, a DDOS attack can be detected by quickly mitigated by isolating the suspect traffic, networks, or hosts. In addition, the traditional DDOS appliances, which generally carry only a local network view, the centralized elements of SDN possess an extensive network topology view and performance, enabling the SDN to be an ideal candidate for vigorously enforcing a coherent security posture.

b) Resilient Control Plane

The three significant elements of SDN are SDN apps, SDNC, and NEs. Since network control is centralized, all communication within the control plane must be treated as critical. Therefore, an outage from a successful attack may have an undesired effect on business stability. For example, the entire network and its tenants may be affected if the SDNC is prevented from taking critical action to mitigate a DOS attack. To avoid this, the control plane needs more resiliencies built into it.

An effective way to improve the resilience of the centralized control plane and protect against the spread of DDOS control-plane attacks to the whole network is to apply a rate limit on the NEs side regarding the consumption of resources and bandwidth, such as CPU load, memory usage, and API calls.
Resilience can be enhanced through the appropriate dedication of the resource – where the SDNC authenticates each resource request and checks requests against the control policies of the authorization.

c) **Strong Authentication and Authorization**

Authentication and authorization are used to identify an unknown source and determine access privileges. These processes can safeguard networks from certain types of attack if they are implemented appropriately, such as:

〉〉 **provision of false feedback to the system -**fooling the system into imagining it is under attack, resulting in the unnecessary deployment of countermeasures, which consumes resources and inevitably leads to sub-optimal use;

〉〉 m**odification of a valid on-path request** – results in a direct attack that amends the behavior of the network;

〉〉 **forwarding traffic that is not implied to be forwarded;**

〉〉 **not forwarding traffic that should be subverting network isolation.**

〉〉 **gaining control access to any component** – depicting the entire network as untrustworthy**.**

Encryption is one of the ways of preventing control data from being leaked. But, even with integrity protection, more than encryption is needed to protect against man-in-the-middle type attacks. So, all the communication within the control plane must be mutually authenticated. Security protocols like TLS and IPSec provide mutual authentication and replay attack protection, confidentiality, and integrity protection.

By enforcing strict authorization and accountability processes, destructions can be limited, and reliable forensics traces can be provided. Role-based access control is a generally used approach for administering an application's permitted actions. Roles can be classified on a host, user, or application basis.

d) Multi-Tenancy

Where networks are built using SDN techniques, the same physical network can be shared among several tenants, which can, in turn, manage their virtual networks. Multi-tenancy permits for better utilization of network resources and lowering the total ownership cost. SDN shortens the time to respond to changing situations by automatically scaling resources for tenants. To maintain an acceptable security level, tenants should not be able to interfere with each other's networks. They need not even be aware that they share network resources with others.

Tenant isolation (separating one tenant's resources and actions from another) is an essential feature of SDN framework security.

e) **Control Plane Isolation**

Isolation is one of the ways to avoid the actions of one tenant from others. This is a critical business attribute that must be vigorously enforced. SDNC orchestrates tenant isolation and implements the specific forwarding rules through SDN NEs. While providing secure isolation lies with the SDNC, tenants also play an essential role in sharing that process.

The network provides isolation primarily on the link layer. If a tenant involves with weak network security, information disclosure may arise, resulting in a breach of isolation at higher layers. A rogue SDN app, for example, with privileges that span beyond the isolation borders, may affect overall network security by driving traffic to a third party by over- or under-billing (theft of service) or by dropping traffic (DOS). The SDN control plane's centralized nature further accentuates such attacks' impact. Consequently, providing isolation can only be partially offloaded onto the SDN network.

f) **Data Plane Isolation**

Tenants running a business on a virtual network using SDN may experience the same kind of network-based attacks as traditional networks. However, the impact of such an attack may be forced on some or even all these tenants due to the shared networking infrastructure. In addition, this new risk may have a commercial effect; nobody wants to launch a business in front of a known or perceived troublemaker, or one inclined to an attack.

So, flows associated with each particular tenant must remain isolated for the data plane. Isolation may be executed logically through the overlay networks and imposed within the NEs. Tagging the ownership of traffic spawned by each tenant, for example, the traffic can be transferred over a shared infrastructure whenever it has been encapsulated (tagged). Tunnels labeled are then forwarded for that tenant to the virtual network. Many complementary and alternative techniques are available for this encapsulation type, including GRE, MPLS, and IPsec.

Including logical isolation, traffic may be encoded with particular tenant keys. This guarantees that in the case of a valid encapsulation violation, the data traffic endures isolation and results in information leakage. Isolation issues need to be settled while directing resource consumption in mind. While traffic isolation can assist the data leakage, shared resource usage also requires resource isolation. As the issue overloads the underlying network equipment, a forwarding loop within one tenant may impact all tenants. To neutralize this problem, the SDNC must enforce resource isolation and use rate-limiting measures to minimize a tenant's impact on the network.

g) **Programmability**

One significant benefit of SDN is programmability: the capability to configure a network efficiently, securely, and promptly. SDN programmability occurs in varying levels of complexity and abstraction. For example, programmability facilitates NEs to be dynamically reprogrammed to forward data flows according to their abilities and higher-level policies in the network. Alternatively, SDN apps allow tenants to programmatically issue run-time requirements to the network. All requests are combined by the SDNC, which fulfills higher-level demands from the capabilities available at the lower levels.

The primary benefit of programmability is flexible control for networks composed using the SDN architecture approach. The capability to control a network and apply changes promptly increases the network’s level of agility. In addition, such flexibility can manufacture the network more secure, as it is constantly screened and outlined to mitigate malicious behavior in more or less real-time. However, the shortcoming of the flexibility delivered by programmability is the substantial impact it has on security.

h) **Configuration Coherency**

Allowing tenants to grant programmatic changes to the network adjusts to changing conditions and increases network agility. In practical terms, programmability can reduce the time it takes to set up a customer alliance network from days or months to hours or minutes.

The programmability also eliminates the need for manual configuration, which is inclined to error. The result of the automatic reconfiguration of networks is viable, providing the SDNC with a global network view and enabling it to execute sanity checking and regression testing.

Another type of conflict occurs due to the complex virtual network topologies and the difficulty of maintaining a coherent security policy across a network. Special care is required for traffic that requires forwarding to security equipment for monitoring purposes. As the traffic can be transmitted over different paths, methods must be implemented to cover all the traffic. Consequently, monitoring is necessary on all paths.

i) Dynamicity

The dynamic and reactive nature of networks constructed using the SDN approach opens new prospects for fighting network attacks. Some techniques include automated network reconfigurations, honeypot forwarding, and black hole routing. Service chaining is yet another technique that utilizes SDN properties and can be used to monitor for malicious payload and trigger mitigating actions.

A network built using SDN techniques can do lower-layer analysis based on parameters such as data rate, source, and packet size. At the same time, the tenant can provide higher-layer analysis based on protocols, transport ports, and payload fingerprints. Once suspicious behavior has been detected, the network can use its programmability features to analyze the situation in more detail or trigger mitigating actions.

7.5 NFV Security and Solution [55]

An essential part of the 5G network is the underlying NFV and software-defined networking (SDN) technologies, which help address the diverse needs of emerging applications (for example, higher data rates and lower latency) and enable economies of scale. Network function virtualization (NFV) yields numerous benefits, particularly the possibility of a cost-efficient transition of telco hardware functionalities on the software platform to break the vendor lock-in problem. However, these benefits come at the price of some security flaws. Indeed, with NFV, virtual mobile networks become vulnerable to several security threats. Nevertheless, these threats can be leveraged using some available mitigation techniques and other emerging solutions.

SDN and NFV are complementary and bring significant advantages when used together. For example, NFV can get the benefits of virtualizing SDN controllers and thus allowing dynamic mobility of SDN controllers to desired locations. On the other hand, SDN can bring value to NFV, allowing dynamic network connectivity by programming the network to be optimal based on network traffic monitoring and analysis. Some practical examples of VNF are vRouters, vFirewalls, virtual content delivery servers, vIPS/vIDS, vDNS servers, and virtual VPN servers.

7.5.1 Security Risks Associated with NFV [55]

This section will review the security challenges that threaten NFVI and explain how these security attacks can be carried out on NFVI. Based on the severity of these security attacks, some best security practices will be discussed to cope with these attacks.

VNFs run over virtual resources such as VMs. Therefore, the security threats associated with VNFs combine those on physical networking and virtualization technologies where NFV specific threats emerge when the two sets intersect. The following will discuss the potential security risks associated with NFVI, considering some possible attack scenarios.

a) **Isolation Failure Risk**

This attack can impose significant risk once an attacker breaks into a hypervisor by compromising some VNFs running over it. This is called a VM escape attack and is depicted in Fig 71.

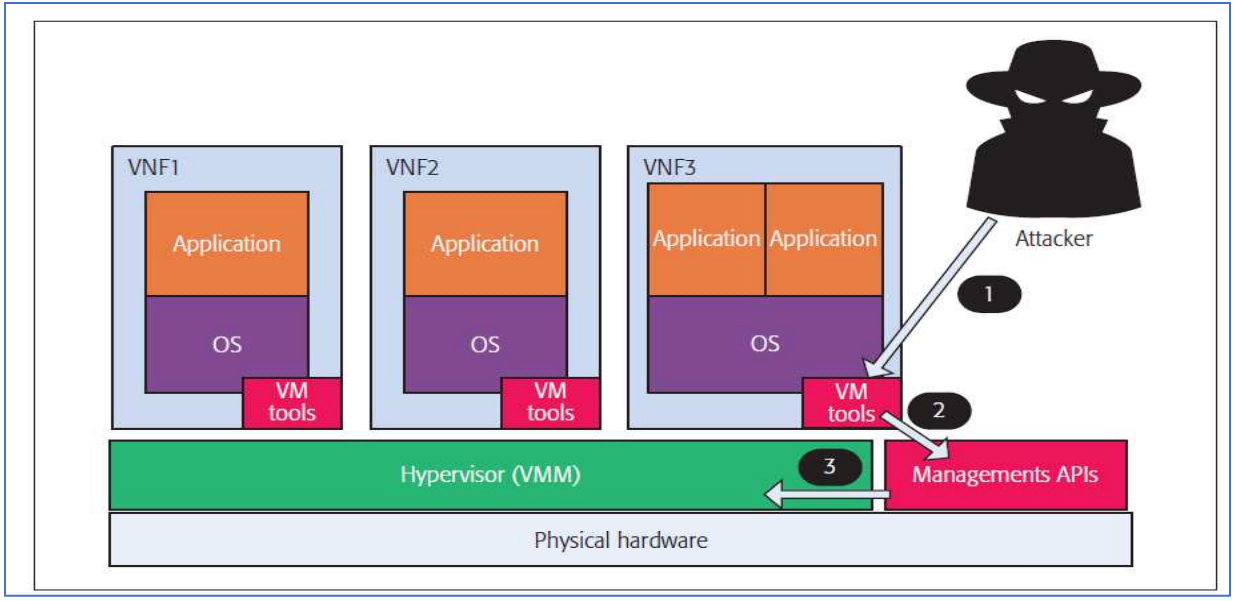


FIGURE 71: A VM ESCAPE ATTACK SCENARIO. [55]

In this attack scenario, the attacker first compromises one VNF by gaining access to its operating system (step 1 in Fig). Next, the attacker gains access to the hypervisor management API using tools and VNF network connectivity with the cloud management network (step 2 in Fig). Then, the attacker breaks into the hypervisor to cause a significant impact (step 3 in Fig). These attacks are possible due to the improper isolation between hypervisors and VNFs. A practical example of this attack could be launched by an application running in a VNF and sending crafted network packets to exploit heap overflow with a compromised virtualization process and resulting in the execution of arbitrary code on the hypervisor to gain access to the host.

In another attack scenario, a VNF may orchestrate other VNFs, which can be achieved by granting the VNF API access to the virtualization infrastructure to instantiate new VNFs. The API can be misused by an attacker who can break in by compromising the VNF and gaining full access to all infrastructure resources.

b) Network Topology Validation and Implementation Failure

Using NFV, virtual networking components (e.g., virtual routers and virtual networks) can be easily created. However, quick and dynamic service decisions can result in human error when a virtual router is designed to interconnect virtual networks without any firewall. Furthermore, virtual network appliances' dynamicity and connectivity can lead to improper separation between the network and its subnets compared to physical network appliance deployments. Using the VM as mentioned above escape attack, an attacker can compromise virtual firewalls to restrict firewall functionality while allowing enough access to carry out the attack. In a similar attack scenario, an attacker may acquire knowledge about a multi-site network infrastructure using the elastic nature of NFVI. Effectively, an attacker can trigger the VNF instantiation or migration in another NFVI point of presence with lower security protection (i.e., without any IDS/IPS/deep packet inspection (DPI) capabilities).

c) Regulatory Compliance Failure

Attacks aiming to place and migrate workload outside the legal boundaries were only possible with traditional infrastructure. However, using NFV, violation of regulatory policies and laws becomes possible by moving one VNF from a legal location to another illegal location, as depicted in Fig. 72.

The consequences of violating regulatory policies can be in the form of a complete banning of service and exerting a financial penalty, which may be the original intention of the attacker to harm the service provider. One possible attack scenario can be when an attacker exploits the insecure VNF API to dump personal data records from the database to violate user privacy.

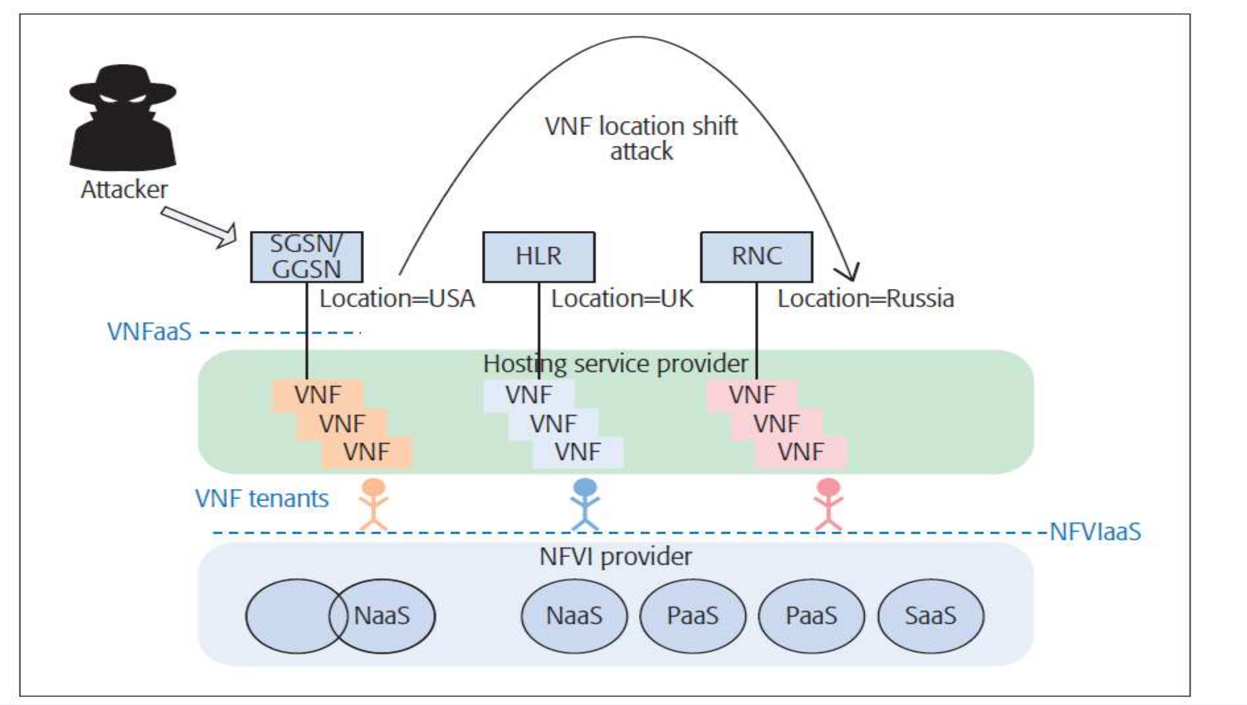


FIGURE 72: VNF LOCATION SHIFT ATTACK [55]

d) Denial of Service Protection Failure

DoS attacks may be directed to virtual networks or VNFs' public interfaces to exhaust network resources and impact service availability. A considerable volume of traffic from a compromised VNF can be generated and sent to other VNFs executing on the same hypervisor or even different hypervisors. Similarly, some VNF applications can consume high CPU, hard disk, and memory resources to exhaust the hypervisor. In this vein, Fig. 73 depicts one practical scenario of a DNS amplification attack.

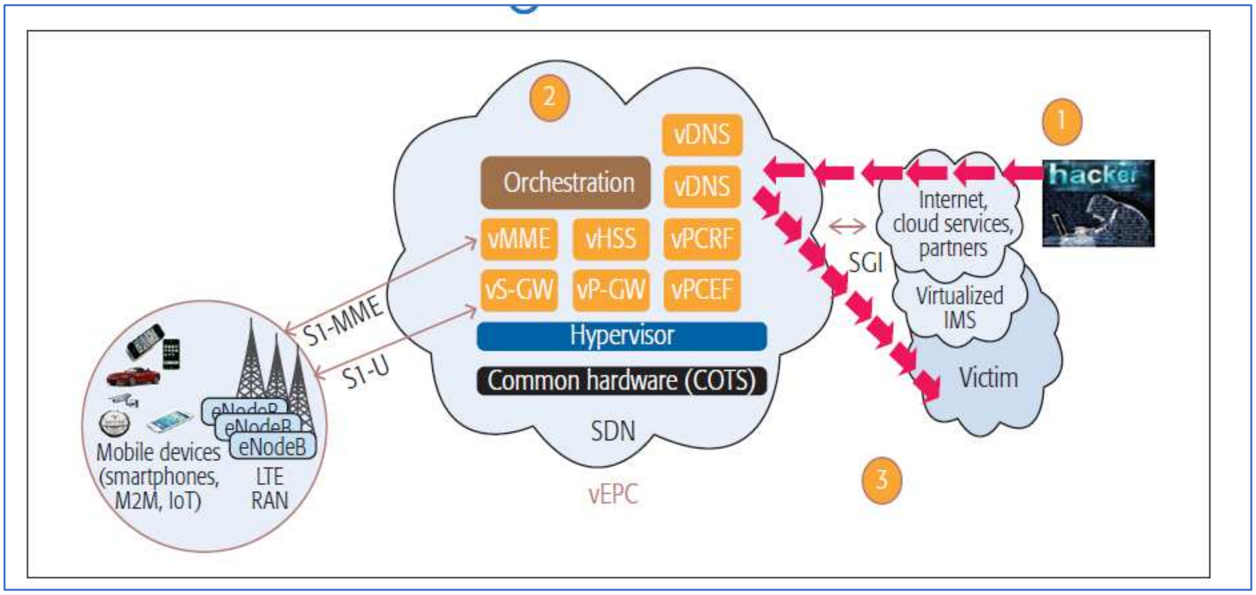


FIGURE 73: DNS AMPLIFICATION ATTACK. [55]

In this scenario, an NFVI infrastructure hosts a virtual DNS server as a component of a virtual evolved packet core (vEPC). The NFVI orchestrator can deploy additional virtual DNS servers if the traffic load increases. An attacker may spoof several victims' IP addresses and launch many malicious DNS queries using the spoofed IP addresses (step 1 in Fig.). In response to such an attack, the orchestrator will instantiate new VMs to scale out the vDNS function to accommodate more queries (step 2 in Fig.). Accordingly, multiple recursive DNS servers will respond to the victims that will ultimately receive amplified DNS query responses (step 3 in Fig.), which can result in service disruption or unavailability.

e) Security Logs Troubleshooting Failure

In this security attack, compromised VNFs can generate massive logs on the hypervisor, making it challenging to analyze logs from other VNFs, especially when the initial entries in the log files are deleted. Furthermore, there is also a risk when the infrastructure logs are leaked, which enables cross-relating logs from one VNF operator with another to extract sensitive information.

f) Malicious Insider

These risks are classified as internal security risks caused by the vicious actions of internal administrators. For example, in one attack scenario, a malicious administrator takes the memory dump of a user’s VM. Since the evil administrator has root access to the hypervisor and uses a search operation, he can extract the user ID, passwords, and SSH keys from the memory dump, violating user privacy and data confidentiality. In a second attack scenario, an internal attacker may extract a user’s data from the hard-drive volume managed by the cloud storage devices. To execute this attack, the attacker first creates a backup copy of the VM drive and then uses open-source tools, such as kpartx and vgscan, to extract sensitive data.

7.5.2 NFV Best Security Practices [55]

The security of VNFs through a security orchestrator in correspondence with the European Telecommunications Standards Institute (ETSI) network virtualization function architecture is presented in the figure below. The proposed architecture ensures virtual functions in multi-tenant environments and secures the physical entities of the telecommunications network.

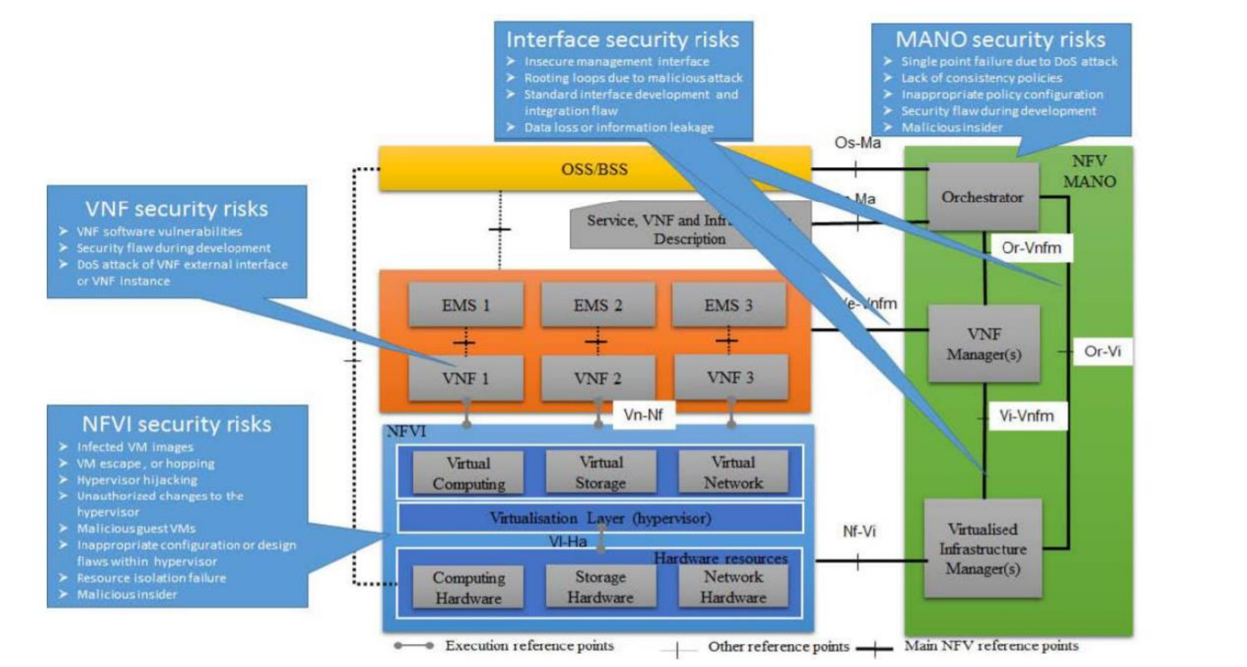
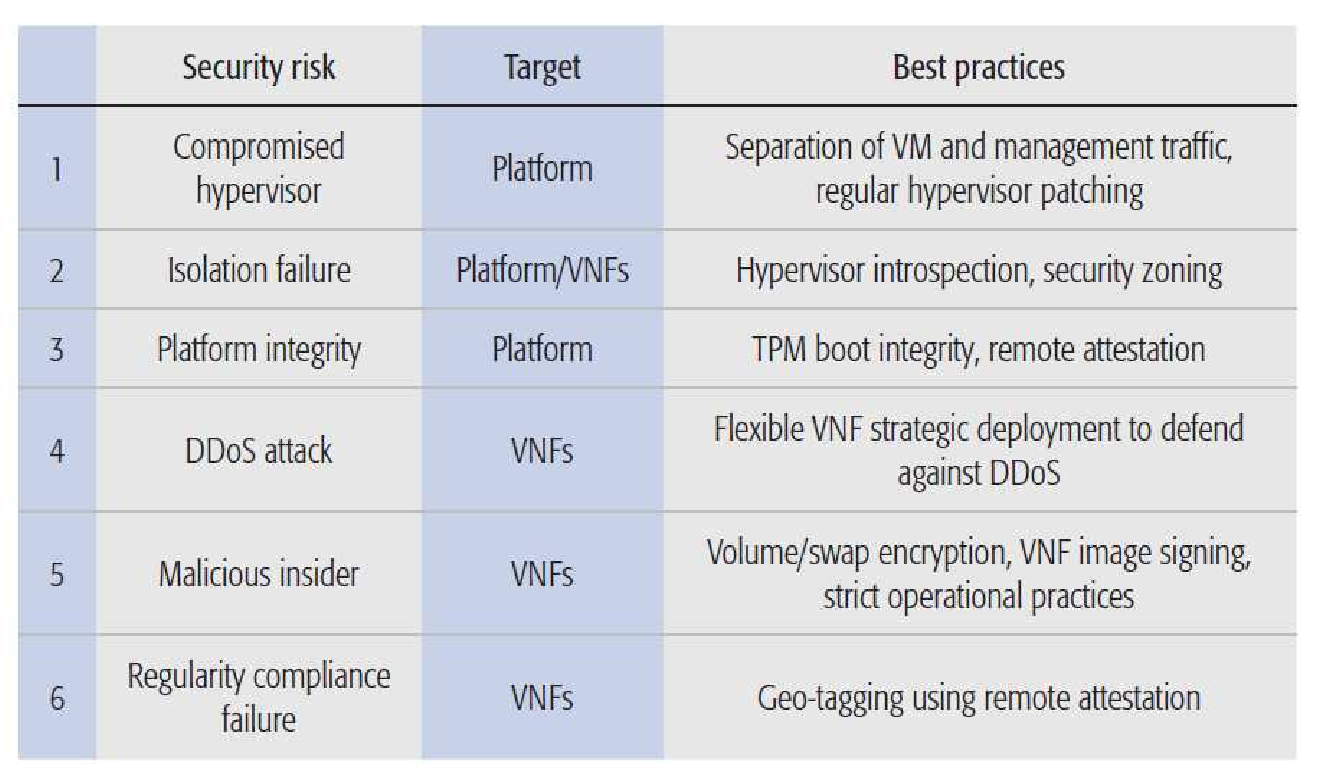

Interface security risks
➤ Insecure management interface
➤ Rooting loops due to malicious attack
➤ Standard interface development and integration flaw
➤ Data loss or information leakage
MANO security risks
➤ Single point failure due to DoS attack
➤ Lack of consistency policies
➤ Inappropriate policy configuration
➤ Security flaw during development
➤ Malicious insider
VNF security risks
➤ VNF software vulnerabilities
➤ Security flaw during development
➤ DoS attack of VNF external interface or VNF instance
NFVI security risks
➤ Infected VM images
➤ VM escape , or hopping
➤ Hypervisor hijacking
➤ Unauthorized changes to the hypervisor
➤ Malicious guest VMs
➤ Inappropriate configuration or design flaws within hypervisor
➤ Resource isolation failure
➤ Malicious insider
OSS/BSS
Service, VNF and Infr
Description
Os-Ma
Ma
NFV MANO
Orchestrator
Or-Vnfm
Ve-Vnfm
VNF Manager(s)
Or-Vi
Vi-Vnfm
Virtualised Infrastructure Manager(s)
EMS 1
EMS 2
EMS 3
VNF 1
VNF 2
VNF 3
Vn-Nf
NFVI
Virtual Computing
Virtual Storage
Virtual Network
Virtualisation Layer (hypervisor)
Vl-Ha
Hardware resources
Computing Hardware
Storage Hardware
Network Hardware
Nf-Vi
Execution reference points
Other reference points
Main NFV reference points

FIGURE 74: SECURITY ORCHESTRATOR NFV AND RISKS [54]

This section will shed light on best security practices that should be followed to achieve reasonably better security protection against the earlier-mentioned threats in an NFV environment. It should be noted that these practices do not guarantee the foolproof security of NFVI but will provide better resiliency against these threats.



	Security risk	Target	Best practices
1	Compromised hypervisor	Platform	Separation of VM and management traffic, regular hypervisor patching
2	Isolation failure	Platform/VNFs	Hypervisor introspection, security zoning
3	Platform integrity	Platform	TPM boot integrity, remote attestation
4	DDoS attack	VNFs	Flexible VNF strategic deployment to defend against DDoS
5	Malicious insider	VNFs	Volume/swap encryption, VNF image signing, strict operational practices
6	Regularity compliance failure	VNFs	Geo-tagging using remote attestation

Table 16: NFVI Security Risks and Best Practices [55]

i) Boot Integrity Measurement Leveraging TPM

Using a trusted platform module (TPM) as a hardware root of trust, the measurement of system-sensitive components such as platform firmware, BIOS, bootloader, OS kernel, and other system components can be securely stored and verified. However, the platform measurement can only be taken when the system is reset or rebooted; there is no way to write the new platform measurement in TPM during the system run-time. Instead, TPM's launch control policy (LCP) or the remote attestation server can validate the platform measurements.

ii) Hypervisor And Virtual Network Security

The hypervisor enables virtualization between underlying hardware and VMs. Virtual networks in the cloud use SDN to allow connectivity among VMs and outside networks. Security of these elements is a must to protect the whole infrastructure. One security best practice is keeping the hypervisor up to date by regularly applying the released security patches. Failure to do that would result in exposure to security risks in the future. Another best practice is to disable all services that are not in use. For example, SSH and remote access services may only sometimes be needed; therefore, enabling these services only when needed would be a good idea. Finally, cloud administrators are the gatekeepers of the whole infrastructure, and their accounts are the keys. It should be mandated to secure admin accounts by applying a solid password policy and strictly following an organization's security guidelines.

iii) Security Zoning

Separating VM traffic and management traffic is a good practice to prevent a VM from impacting other VMs or hosts. This will prevent attacks by VMs tearing into the management infrastructure. It is also a good idea to divide the VLAN traffic into groups and disable all other VLANs that are not in use. Likewise, VMs of similar functionalities can be grouped into specific zones, and their traffic should be isolated. Each zone can be protected using access control policies and a dedicated firewall based on its needed security level. One example of such zones is a demilitarized zone (DMZ).

iv) Linux Kernel Security

In virtualized platforms, the kernel of the host systems is an essential component that provides isolation between the applications. The SELinux module, developed by the National Security Agency (NSA), is implemented in Kernel and provides robust isolation between the tenants when virtualization technology is used over the host. Secure virtualization (sVirt) is a new form of SELinux developed to integrate mandatory access control security with Linux-based hypervisors. sVirt provides isolation between VM processes and data files. Beyond these tools, other kernel hardening tools can help secure the Linux kernel. A notable example is hidepd, which can be used to prevent unauthorized users from seeing other users' process information. Another example is GRSecurity, which protects against attacks on corrupted memory [49].

v) Hypervisor Introspection

Hypervisor introspection can scrutinize software running inside VMs to find abnormal activities. It acts as a host-based IDS that has access to the states of all VMs so that the rootkit and boot kit inside VMs cannot hide easily. Using introspection capabilities, the hypervisor's functionalities are enhanced, enabling it, among other things, to monitor network traffic, access files in storage, and read memory execution. Hypervisor introspection APIs are powerful tools for deep VM analysis and potentially increase VM security. However, they can also be used as an exploit that allows breaking and bypassing the isolation between VMs and the hypervisor. LibVMI is the library for hypervisor introspection for various platforms, implemented in C language with Python bindings. It gives the hypervisor the means to perform deep inspection of VMs (e.g., memory checking, vCPU register inspection, and recording trapping events) [53].

vi) Encrypting VNF Volume/Swap Areas

Virtual volume disks associated with VNFs may contain sensitive data. Therefore, they need to be protected. The best practice to secure the VNF volume is safely encrypting and storing the cryptographic keys. The TPM module can also be used to store these keys securely. In addition, the hypervisor should be configured to securely wipe out the virtual volume disks in the event a VNF is crashed or intentionally destroyed to prevent it from unauthorized access. VM swapping is a memory management technique used to move memory segments from the main memory to the disk, which is used as a secondary memory to increase system performance in case the system runs out of memory. These transferred memory segments can contain sensitive information such as passwords and certificates. Therefore, they can be stored on the disk and remain persistent after a reboot. This enables an attack scenario whereby a VM swap is copied and investigated to retrieve helpful information. One way to avoid this attack is to encrypt VM swap areas. Linux-based tools such as dm-crypt can be used for this purpose.

vii) VNF Image Signing

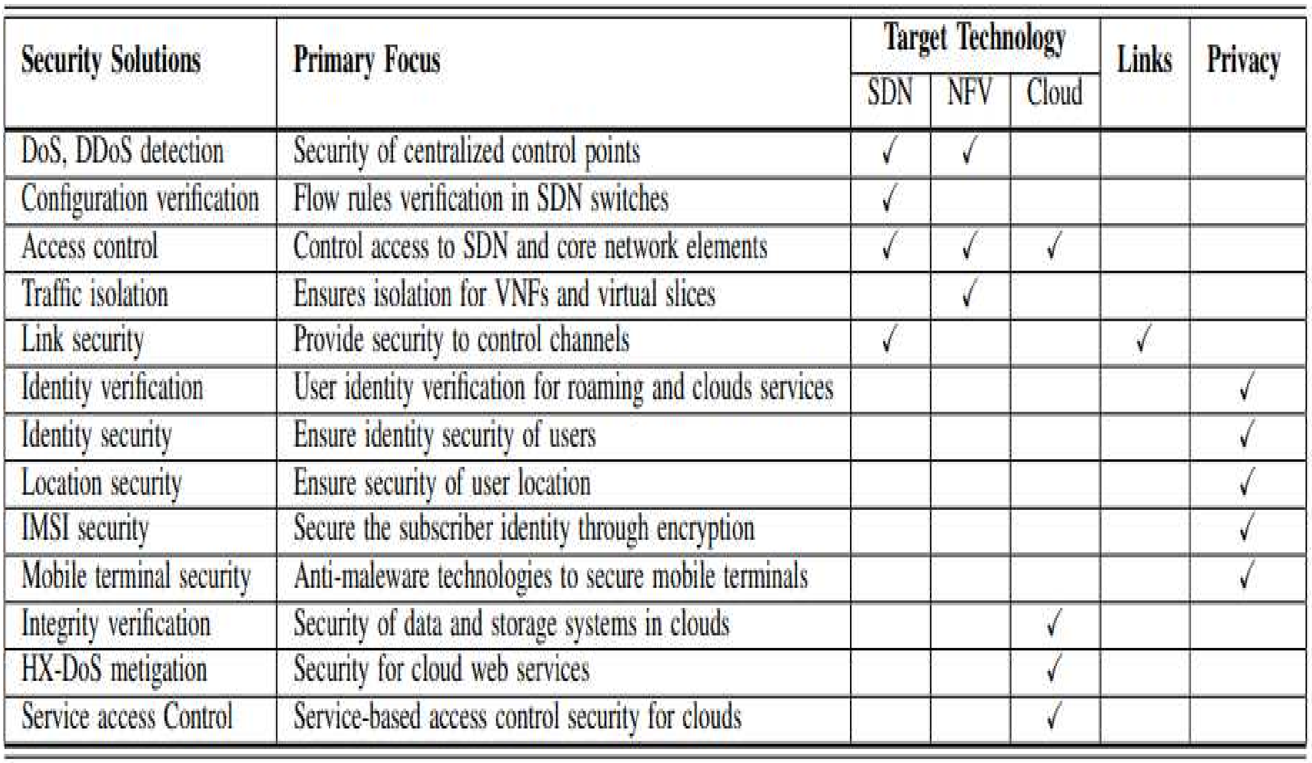
It is easy to tamper with VNF images. It requires only a few seconds to insert some malware into a VNF image file while uploaded to an image database or transferred from an image database to a compute node. Luckily, VNF images can be cryptographically signed and verified during launch time. This can be achieved by setting up some signing authority and modifying the hypervisor configuration to verify an image's signature before they are launched.

viii) Security Management and Orchestration

One best practice consists of designing an NFV orchestrator incorporating the security and trust requirements of the NFVI. The orchestration and management of security functions require integration by enabling interaction among the security orchestrator, the VNF manager, and the element management systems (EMS). This type of protection can be achieved by setting scaling boundaries in the VNFD or network service descriptor (NSD), for example, and having the NFVO enforce these restrictions to protect from attacks such as a DNS amplification attack.

ix) Remote Attestation

The remote attestation technique can be used to verify the trust status of an NFV platform remotely. As mentioned, the concept is based on boot integrity measurement leveraging TPM. Remote attestation can be provided as a service and may be used by either the platform owner or a consumer to verify if the platform has been booted in a trusted manner. Practical implementations of the remote attestation service include the open cloud integrity tool (openCIT), an open-source software hosted on GitHub.



Security Solutions	Primary Focus	Target Technology			Links	Privacy
		SDN	NFV	Cloud		
DoS, DDoS detection	Security of centralized control points	✓	✓			
Configuration verification	Flow rules verification in SDN switches	✓				
Access control	Control access to SDN and core network elements	✓	✓	✓		
Traffic isolation	Ensures isolation for VNFs and virtual slices		✓			
Link security	Provide security to control channels	✓			✓	
Identity verification	User identity verification for roaming and clouds services					✓
Identity security	Ensure identity security of users					✓
Location security	Ensure security of user location					✓
IMSI security	Secure the subscriber identity through encryption					✓
Mobile terminal security	Anti-maleware technologies to secure mobile terminals					✓
Integrity verification	Security of data and storage systems in clouds			✓		
HX-DoS metigation	Security for cloud web services			✓		
Service access Control	Service-based access control security for clouds			✓		

Table 17: Potential Security Solutions for Targeted Threats [56]

SECTION 8. STRATEGIES FOR NETWORK SLICING AND SECURITY [23], [45]

As discussed in the previous sections, ensuring confidentiality, integrity, and availability for 5G network slicing protection is paramount. In addition, more advanced mitigations include Zero Trust Architecture (ZTA) requirements, Multi-Layer Security (MLS), Cross-Domain Solutions (CDS), Post-Quantum Cryptography (PQC), and Isolation.

8.1 End-to-End Network Slice Security (UE, RAN, Core, Orchestration)

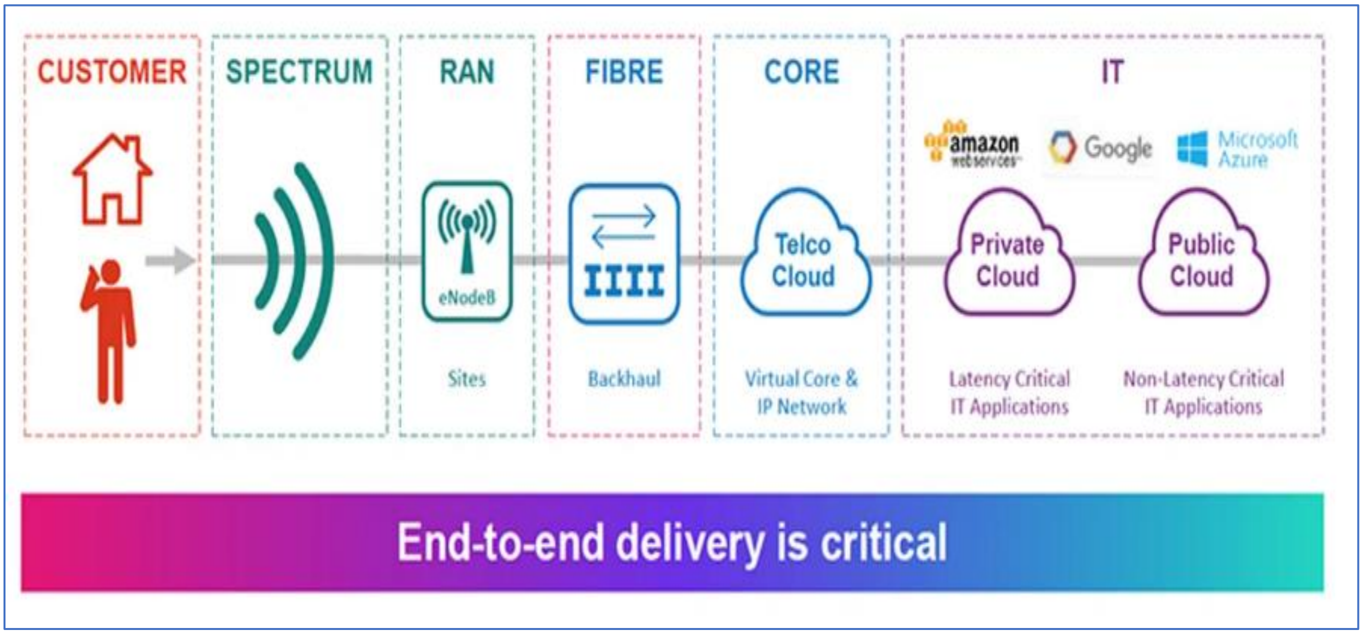


FIGURE 75: END-TO-END DELIVERY

Image Source: Three UK

Network slicing is introduced to cellular technology in 5G and is defined as a logical network that provides specific network capabilities and characteristics. The instance of a Network Slice is a set of Network Function instances and the required resources (i.e., compute, storage, and networking resources) that form a deployed Network Slice. A Network Slice instance contains at least one virtual network function.

An end-to-end slice is from the UE via the RAN, transport network (e.g., backhaul), and the core network. The security capabilities for each end-to-end network slice segment are different. Management and orchestration are also associated with provisioning the end-to-end slice (including preparation with network slice template, commissioning, operation, and decommissioning). There needs to be more guidance to address how all these segments work together as a cohesive slice with all the required security capabilities; inter- and intra-slice isolation must be addressed.

8.2 Zero Trust Architecture (ZTA)

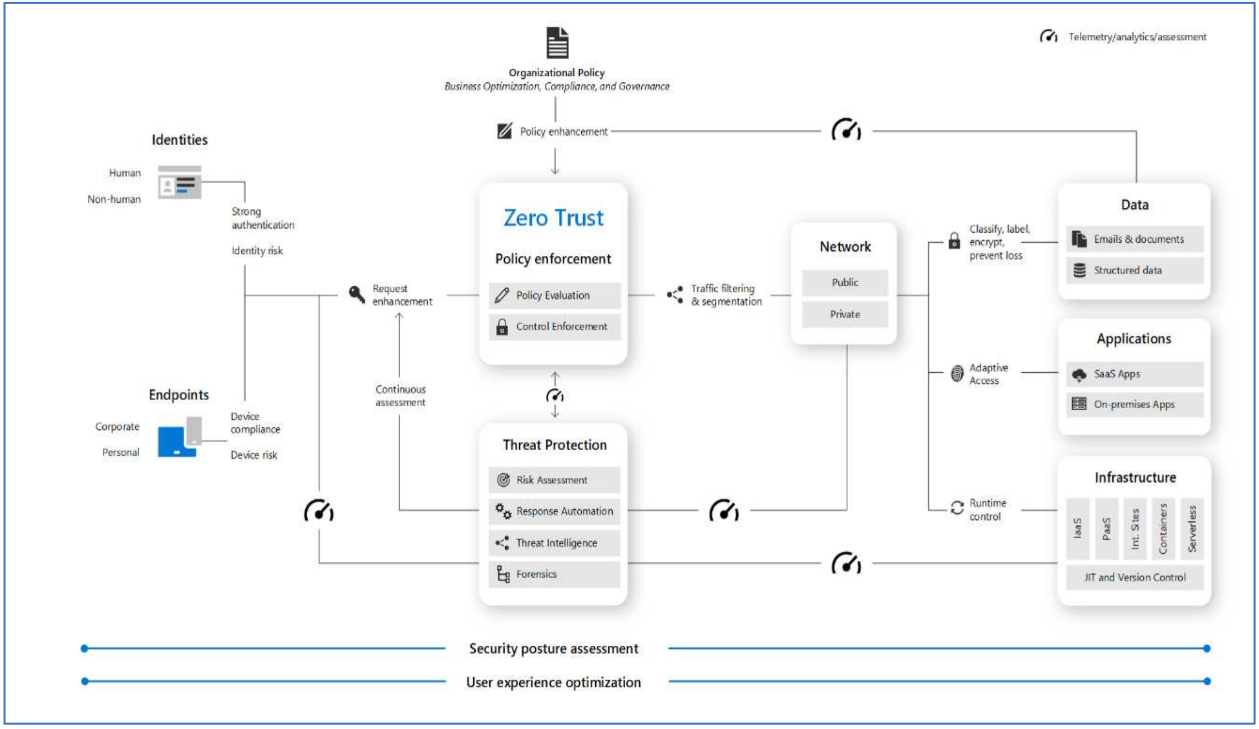


FIGURE 76: ZERO TRUST ARCHITECTURE [59]

Image Source: Microsoft

ZTA is based on the principle of “never trust, always verify,” which directly relates to the confidentiality and integrity of a given system. It focuses on eliminating implicit trust and continuously validating every stage of a digital interaction through solid authentication and authorization methods. If implemented appropriately, it can drastically reduce the possibility of attacks like MitM and configuration attacks in network slicing.

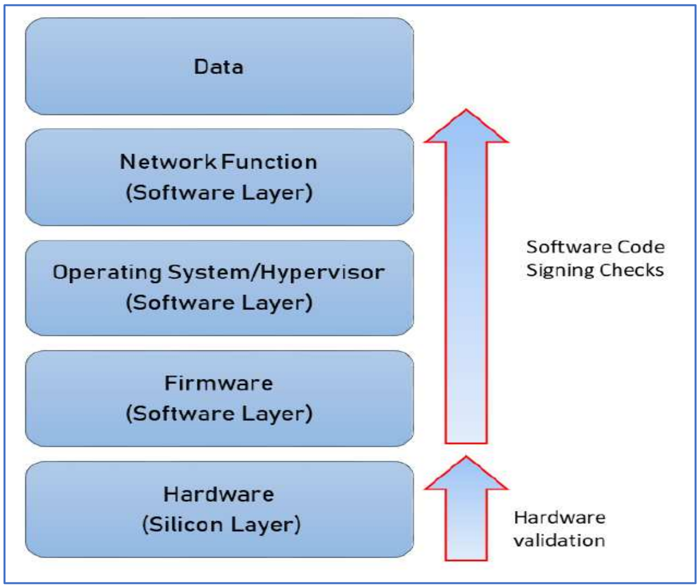


FIGURE 77: ZERO-TRUST VALIDATION CHECKS AT BOTH THE HARDWARE AND SOFTWARE LAYERS. [58]

Various actions can be triggered upon detecting anomalous behavior, including segregating, and isolating the interactions, terminating access credentials, or initiating multi-factor authentication (MFA). In addition, the tenets of ZTA aim to reduce the exposure of resources to malicious actors by leveraging network segmentation. This can prevent incidents and minimize lateral movement.
Common ZTA techniques:

- MFA
- Encryption
- Access control

8.3 Multi-Layer Security

ZTA is also a prerequisite for enabling Multi-Layered Security (MLS). With the proper access control and authentication, MLS permits access to users with different levels of access while preventing users from obtaining access to information for which they lack authorization. In addition, MLS protects itself from subversion and has robust mechanisms to separate information domains based on trustworthiness.

8.4 Cross-Domain Solutions

Cross-Domain Solutions (CDS) is a set of mechanisms with a controlled interface that controls the flow of information and enforces the security policies among interconnected information systems that ensure the ability to manually and automatically access and transfer data between different security domains3. CDS is intended for high-value networks where assurance-based security measures, such as firewalls, Security Information and Event Management, and Intrusion Detection Systems, are insufficient to ensure the trusted domain's security.

8.5 Post-Quantum Cryptography

Another way to mitigate risks further is the use of advanced encryption techniques. A future quantum computer, if built, would be capable of undermining the widely deployed public key algorithms currently used for asymmetric key exchanges and digital signatures. New quantum-safe solutions will eventually need to be adopted in critical infrastructure to mitigate this potential vulnerability. Post-quantum cryptography (PQC) algorithms are under evaluation and will be standardized for the relevant technologies, including 5G. PQC is a possible solution to enhance data protection when a higher level of security is required. 5G vendors and operators must incorporate updated standards to mitigate potential future risks as they become available.

8.6 Isolation

Isolation is another cyber technique to enhance security in 5G network slicing. The level and strength of isolation may vary depending on slicing requirements and usage. In some network slicing instances, there may be a security requirement for strict slice isolation, but other slicing instances may require communication between slices. Thus, isolation may be performed in many ways, and it involves a set of properties chosen according to implementation needs. Isolation could be based on a sandbox, virtual machine, operating system, or hardware and physical. Some enabling technologies include physical resource block scheduling, slice scheduling, and traffic shaping.

Slice isolation includes the following aspects:

- **Isolation of traffic**: the slices should ensure that the data flow of one slice does not move to another.
- **Isolation of bandwidth**: slices should not use any bandwidth assigned to other slices.
- **Isolation of processing**: while all virtual slices use the same physical resources, independent processing of packets is required.
- **Isolation of storage**: slice-related data should be stored separately from data used by another slice.

8.7 Security as a Service (SECaaS) [57]

The next generation of 5G wireless access systems aims to improve the Quality of Experience (QoE) for critical services and provide high availability, low latency, elasticity, and increased security. SDN, NFV, and Network Slicing are the concepts and technologies that have emerged as 5G enablers. These technologies can be leveraged to provide Security as a Service (SECaaS) by deploying Security Virtualized Network Functions (VNFs) within different slices and ensuring optimal resource provisioning to reduce Operational Expenditures (OPEX) while ensuring the provisioning of the Service Level Agreement (SLA). Furthermore, proper resource allocation is crucial as a malfunctioning security VNF can compromise the network; therefore, a predictive auto-scaling function, implementing application-specific policies, needs to be deployed along with the monitoring and flow control mechanisms.

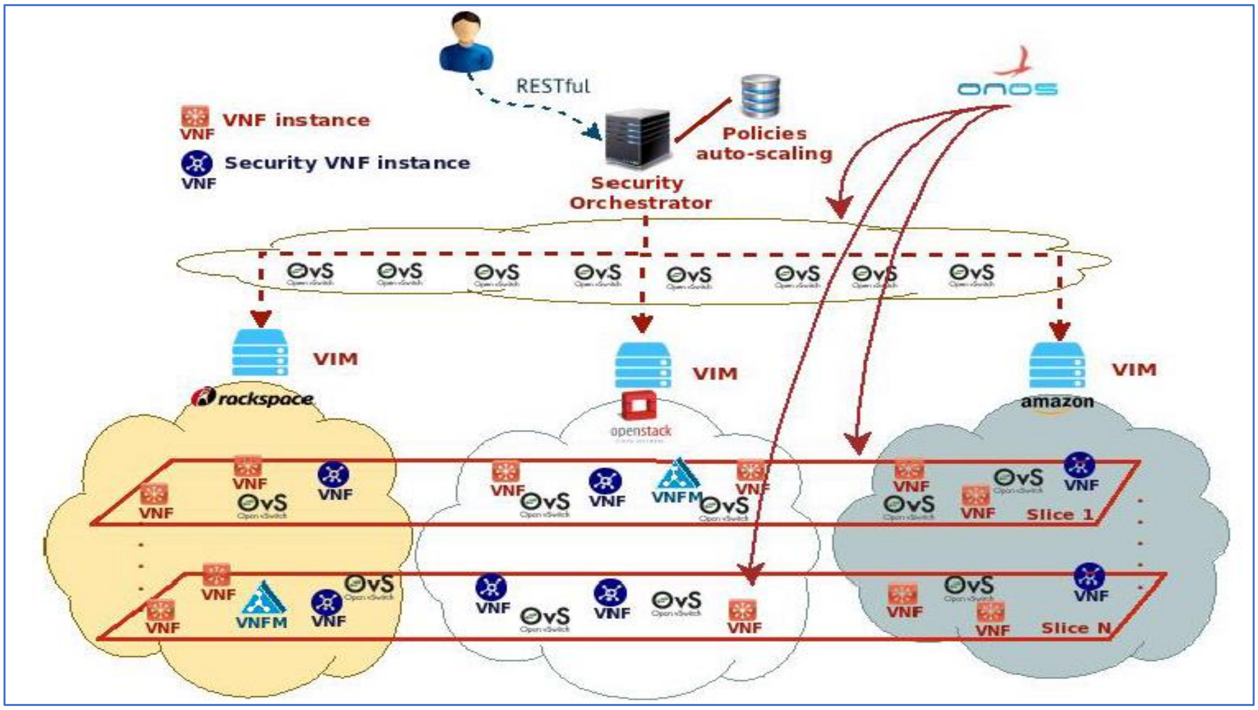


FIGURE 78: SECURITY AS A SERVICE ARCHITECTURE CONSIDERATION [57]

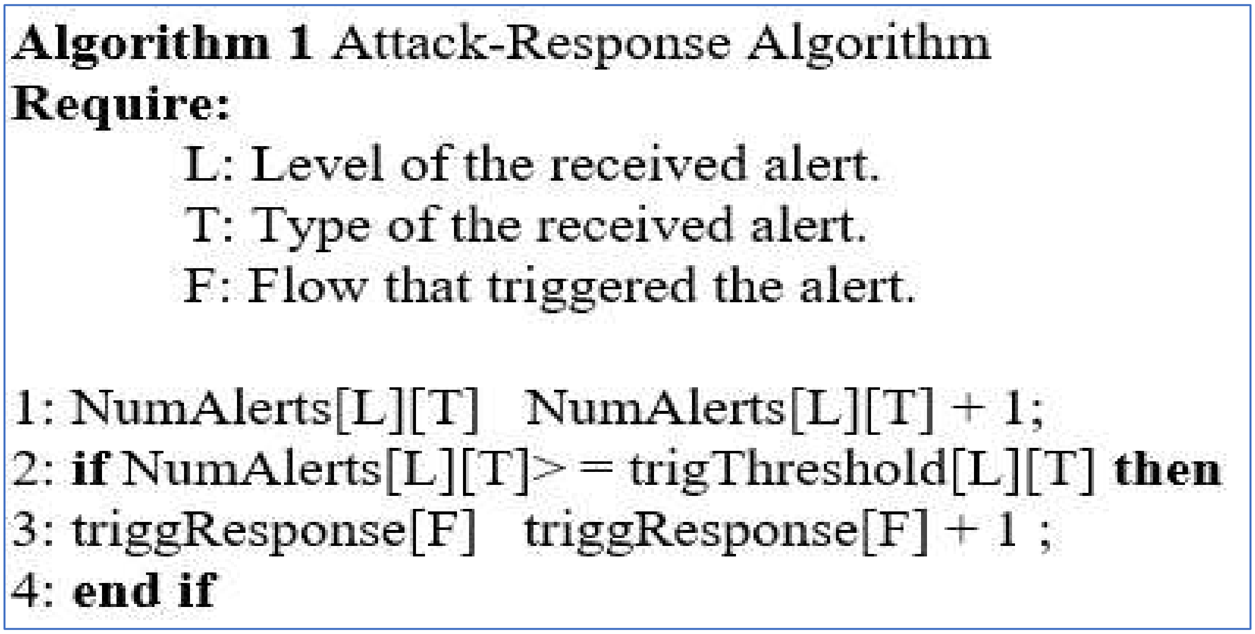
Network slicing is mainly based on SDN, NFV, and cloud computing. ETSI NFV has defined a reference architecture for efficiently enabling NFV orchestration and VNF management. The NFV technology will allow elasticity and flexibility for creating different slices across multiple domains. Meanwhile, the SDN technology will enable the programmability of other Open Virtual Switches (OVS) and SDN-enabled switches to ensure the connectivity between different VNFs in the same network slice.

Creating different VNFs in different slices would create more vulnerability in different VNFs compared to the static network. Therefore, ensuring security within the same slice can be a challenging problem. Fig. 78 shows an overview of the proposed architecture enabling SECaaS in an inter-domain platform. This architecture will deploy and manage security VNFs, including IDS/IPS and Deep Packet Inspection (DPI). Furthermore, the proposed architecture framework aims to ensure elasticity by dynamically deploying security VNF instances, monitoring their performance, and performing predictive auto-scaling based on pre-defined policies and metrics.

8.7.1 Security as a service with SDN

A solution must consider NFV resource management, workload mobility, VNF placement, and VNF security. ONOS acts as an SDN controller to enable SECaaS in the architecture and use the intent concept to create per-flow point-to-point intents to route traffic to specific firewall or IPS instances. The intents aggregate the output traffic of a firewall instance and forward it to the correct node/VM in the secure network.

The ONOS controller can deploy multi-points to single-point intents to forward traffic to its original destination and mirror it to a specific IDS instance. Besides, when an IDS detects a malicious flow, it generates an alert. It sends it to the security orchestrator, which will then consider the number of received signals, as well as the level of severity to instruct the controller to either stop the malicious flow temporarily or permanently or constrain its bandwidth to avoid overloading the network while maintaining a certain level of service. Furthermore, SDN's capabilities are mandatory to enable auto-scaling support, i.e., guaranteeing that a complete traffic analysis may be supported even during an attack that would overload the current slice security configuration. However, scaling out an IDS instance requires splitting the incoming traffic between the new and existing instances. Additionally, network flows must be managed to ensure security isolation between slices.



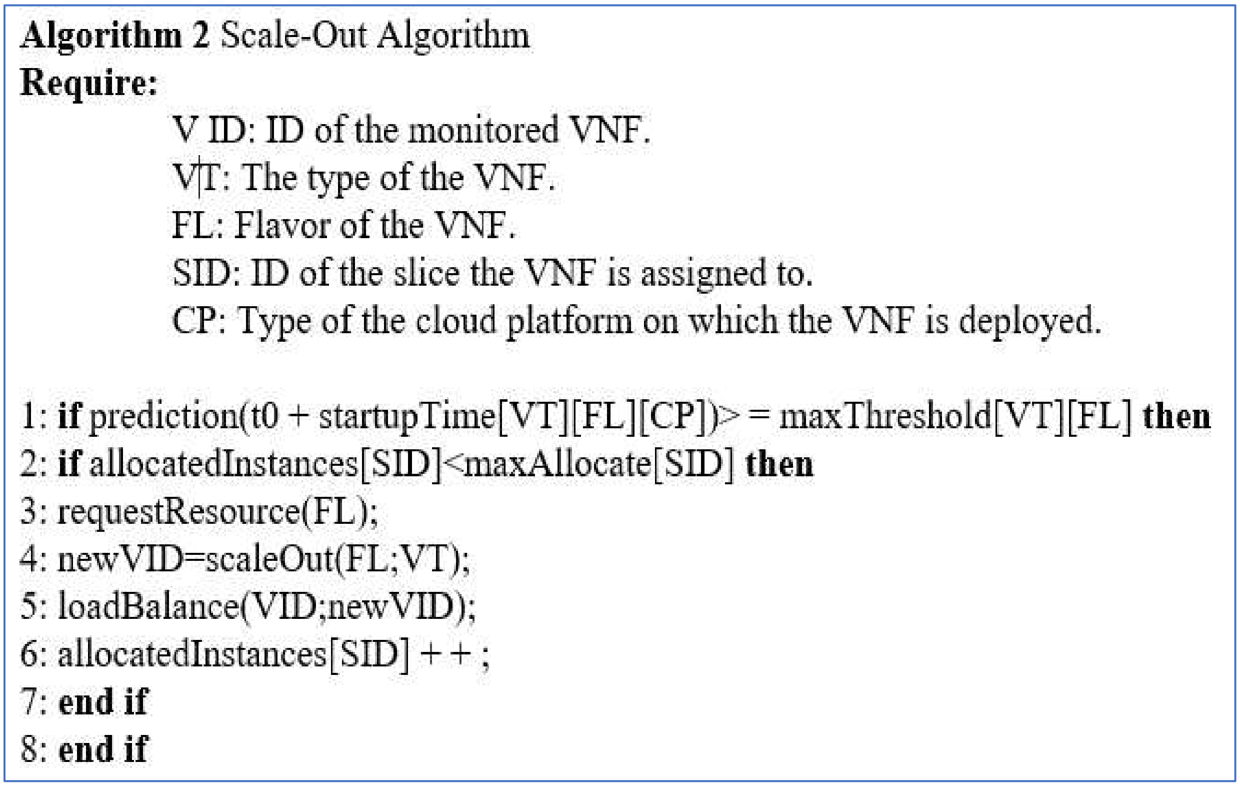
```
Algorithm 1 Attack-Response Algorithm
Require:
        L: Level of the received alert.
        T: Type of the received alert.
        F: Flow that triggered the alert.

NumAlerts[L][T]   NumAlerts[L][T] + 1;
if NumAlerts[L][T]> = trigThreshold[L][T] then
triggResponse[F]   triggResponse[F] + 1 ;
end if
```

FIGURE 79: ATTACK RESPONSE ALGORITHM FOR SDN [57]

8.7.2 Security as a service with NFV

As shown in Fig. 78, the security orchestrator offers a RESTful API that allows the admin user to specify different management rules and policies for the instantiation and auto-scaling of the VNF instances. Based on these policies, the security orchestrator enforces the rules for a specific slice by communicating them to the VNFM of the slice, allowing it to enforce the security rules by sharing them with different security VNFs. The VNFM dynamically launches security VNF instances in other slices with pre-installed software in the cloud and monitors their performance metrics to trigger scaling actions according to the predefined policies. The scaling policies are set according to the VNF's performance requirements and behavior depending on traffic load. Moreover, the security orchestrator communicates with the SDN controller, e.g., ONOS, to connect the different security VNFs and VNF instances in the same vertical.

In the proposed architecture, an auto-scaling algorithm is executed at the VNF Manager of each slice to scale in or scale out each security VNF instance according to the predefined policies and the performance and features of that NF. Furthermore, the auto-scaling solution should consider the VM startup time, which can vary according to the cloud platform and can also be impacted by the OS image and VM type, as well as the number of requested VMs and data-center load. Lastly, a multi-slice architecture means that concurrency for resources needs to be managed at the orchestrator level by setting minimal and maximal resource limits for each slice, as well as levels of priority matching their service requirements. To set the appropriate threshold for the policies mentioned above, it should determine the maximum traffic load each security VNF can process given a certain number of resources without dropping packets or inducing latency. In that way, the scaling can be performed proactively, thus ensuring continuity of service.



```
Algorithm 2 Scale-Out Algorithm
Require:
        V ID: ID of the monitored VNF.
        VT: The type of the VNF.
        FL: Flavor of the VNF.
        SID: ID of the slice the VNF is assigned to.
        CP: Type of the cloud platform on which the VNF is deployed.

if prediction(t0 + startupTime[VT][FL][CP])> = maxThreshold[VT][FL] then
    if allocatedInstances[SID]<maxAllocate[SID] then
        requestResource(FL);
        newVID=scaleOut(FL;VT);
        loadBalance(VID;newVID);
        allocatedInstances[SID] + + ;
    end if
end if
```

FIGURE 80: SCALE-OUT ALGORITHM FOR NFV [57]

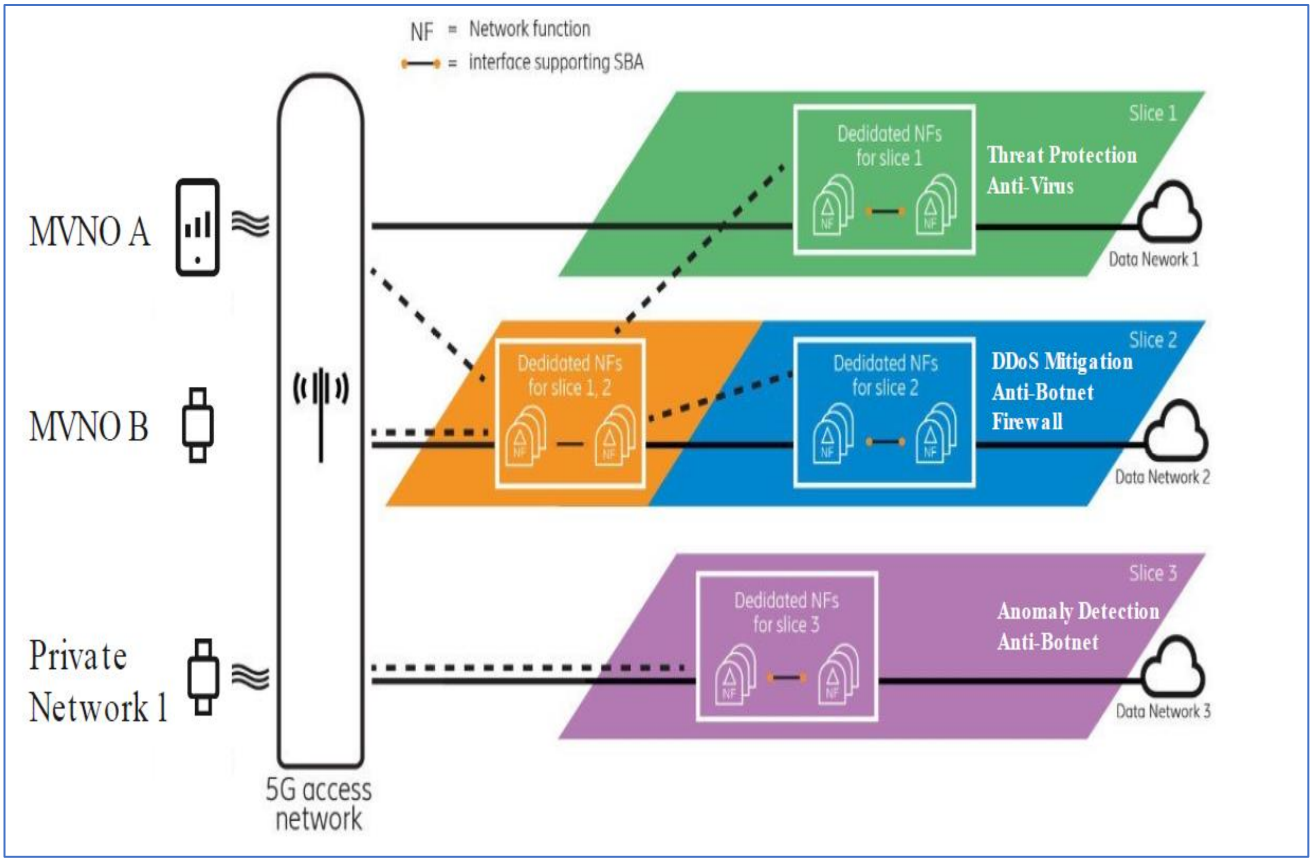
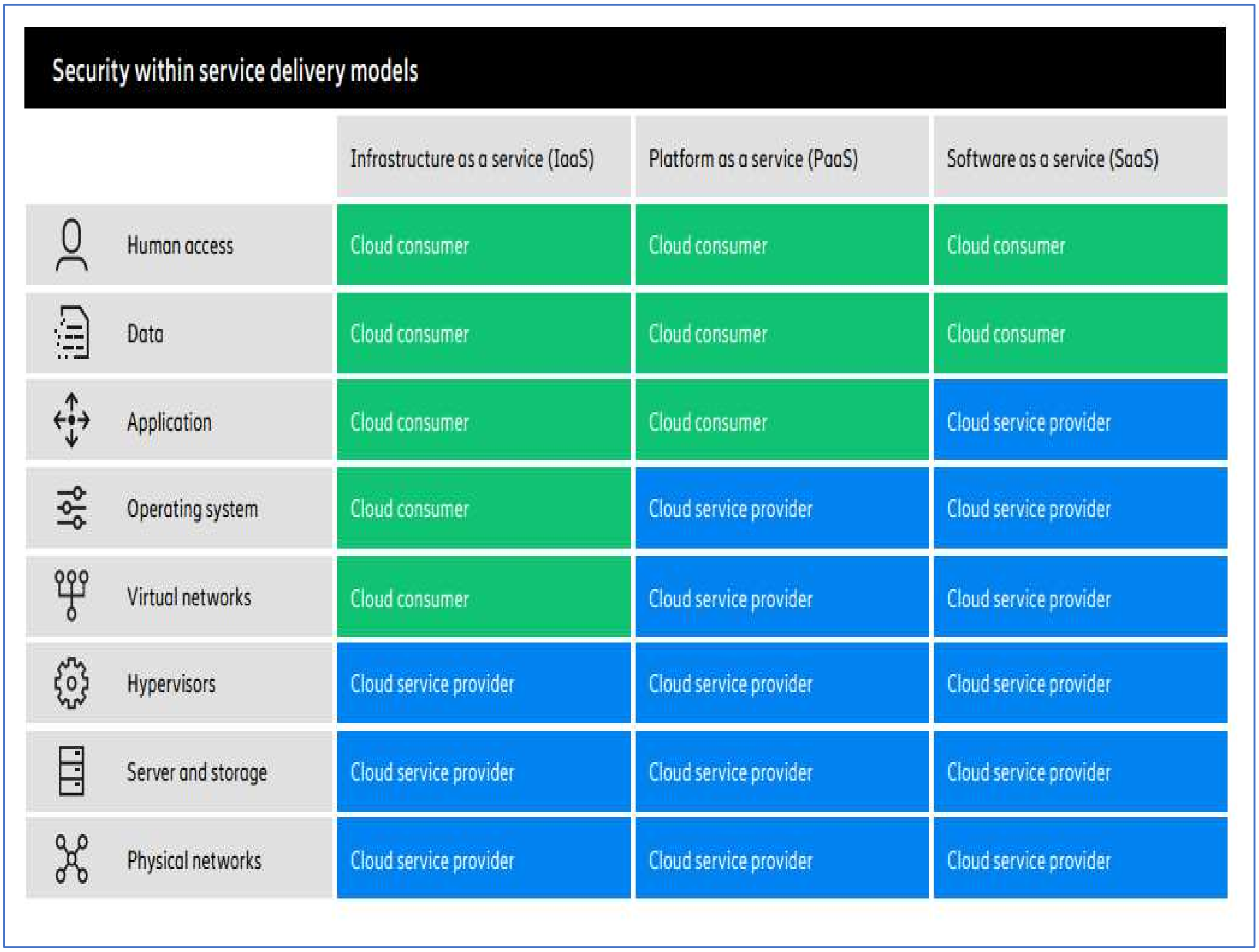


FIGURE 81: SECAAS BUILT UPON NETWORK SLICING [58]

Image Source: 5G Americas

8.8 Security in Cloud Share Responsibility [60], [61]

Considering and evaluating public cloud services, it’s critical to understand the shared responsibility model, which security tasks are maintained by the cloud provider, and which jobs are governed by the consumers. The workload responsibilities vary according to the hosted workload on Platform-as-a-Service (PaaS), Infrastructure-as-a-Service - IaaS, and Software-as-a-Service (SaaS), or in an on-premises data center. In an on-premises data center, the consumers own the whole stack. The following diagram specifies the areas of responsibility between the consumer and the cloud server provider, according to the cloud architecture stack's deployment type.



Security within service delivery models

	Infrastructure as a service (IaaS)	Platform as a service (PaaS)	Software as a service (SaaS)
Human access	Cloud consumer	Cloud consumer	Cloud consumer
Data	Cloud consumer	Cloud consumer	Cloud consumer
Application	Cloud consumer	Cloud consumer	Cloud service provider
Operating system	Cloud consumer	Cloud service provider	Cloud service provider
Virtual networks	Cloud consumer	Cloud service provider	Cloud service provider
Hypervisors	Cloud service provider	Cloud service provider	Cloud service provider
Server and storage	Cloud service provider	Cloud service provider	Cloud service provider
Physical networks	Cloud service provider	Cloud service provider	Cloud service provider

TABLE 18: CLOUD SHARED RESPONSIBILITY MODEL [61]

Source: Ericsson

As the cloud consumer, the MNO deploys 5G critical infrastructure in hybrid and public clouds. The MNO may also choose to deploy a private cloud to achieve the highest level of security gained from asset ownership, deployment control, and network visibility. The Cloud Shared Responsibility Model (CSRM) is a helpful cloud industry tool to determine which stakeholder has security responsibility at each layer of the cloud stack. The CSRM can be summarized by the following cloud industry phrase: "The cloud service provider is responsible for the cloud security, and the cloud consumer is accountable for security in the cloud." While the cloud service provider may be delegated responsibility by the MNO, the MNO is accountable for the security posture of the 5G RAN and core deployment, as reported in US DHS CISA's "Security Guidance for 5G Cloud Infrastructures" [62]. The assignment of responsibility is further complicated in hybrid cloud environments. This is because the hybrid cloud may be a mix of private and public clouds, a private cloud within a cloud service provider's public cloud infrastructure, or the cloud service provider infrastructure deployed on-premises of the MNO. The Cloud Security Alliance (CSA) has identified the additional security risks with hybrid clouds and has formed the CSA Hybrid Cloud Security Working Group [63]. Security responsibilities must be clearly defined and specified in the cloud service agreement. The MNO is always responsible for protecting sensitive data and properly configuring security tools, network functions, interfaces, and APIs.

SECTION 9. USE CASE ANALYSIS [64], [65]

9.1 UE Networks and Rouge Base Station Detection

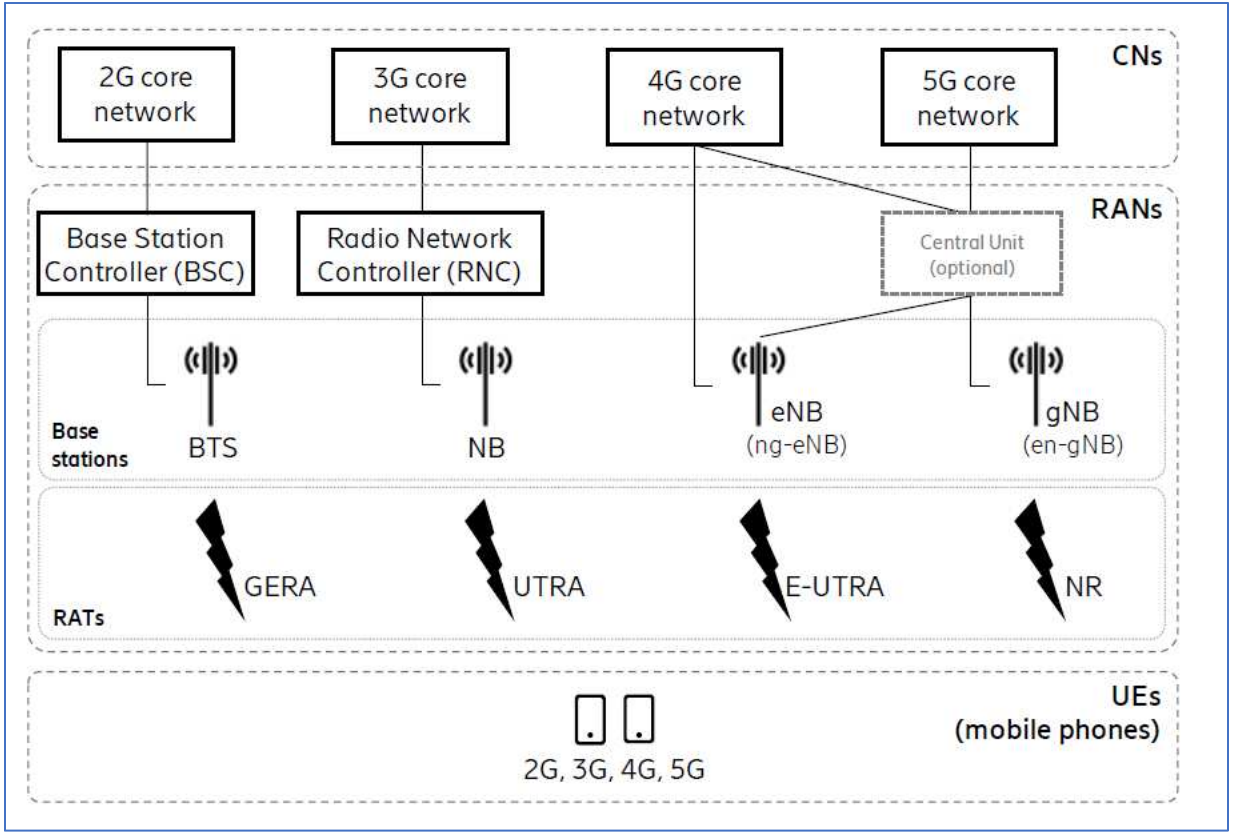


FIGURE 82: SIMPLIFIED OVERVIEW OF MOBILE NETWORKS [64]

Different generations of RAN offer radio access via a radio access technology-RAT specific to that generation, i.e.,

- 2G RAT: GSM EDGE Radio Access-GERA
- 3G RAT: Universal Terrestrial Radio Access-UTRA
- 4G RAT: Evolved-UTRA
- In 5G, however, two types of RATs co-exist the 4G RAT (E-UTRA) and a new 5G RAT, New Radio-NR.

The RAN base stations offer these RATs. Therefore, they are known as Base Transceiver Stations (BTS), NodeB (NB), Evolved NB (eNB), and next-Generation NB (gNB) for 2G, 3G, 4G, and 5G, respectively. In addition, ng-eNB (Next-Generation eNB) also connects to a 5G core network, whereas the en-gNB (EUTRA New Radio gNB) connects to the 4G core network. These base stations cover one or more cells in the smallest area where the base stations serve the UEs.

There has been a burning interest in false base station detection systems recently. The false base station is a broad name for a radio device that aims to impersonate a legitimate base station. However, the title says, “base station,” its attack capabilities have also outgrown emulating UEs towards the mobile network.

Names like IMSI catcher, Stingray, rogue base station, cell site simulator, etc., also identify it. A logical illustration of false base station attacks is shown in Fig 83.

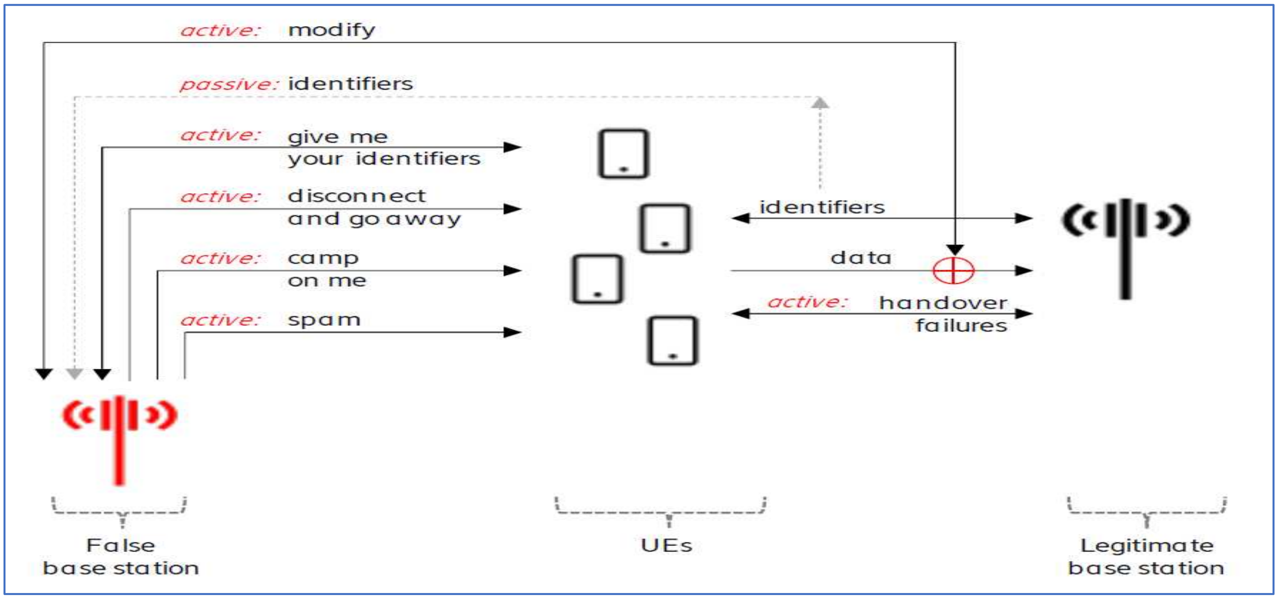


FIGURE 83: LOGICAL ILLUSTRATION OF FALSE BASE STATION ATTACKS [64]

Most of these depend on the software users download into their mobile phones. The software analyzes the mobile phone's radio environment measurements taken or reports these measurements to a server on the Internet, then analyzes the aggregated measurements collected from many mobile phones.

One of the primary attacks relates to users' privacy. An attacker either passively eavesdrops on users' identifiers from the radio interface or obtains them by communicating with the UEs. The attacker then uses those identifiers to identify or track the users. The attacker might also try to fingerprint user traffic. Under stringent assumptions, a resourceful attacker may exploit implementation flaws or vulnerabilities in application layer protocols like Domain Name System (DNS) and Internet Control Message Protocol (ICMP) by altering carefully chosen parts of the data in the radio interface. Another set of attacks relates to denial of service (DoS) on UEs and mobile networks. The attacker may use specific messages that the UEs and the network accept without authentication. The attacker may also create favorable radio conditions so the UEs keep camping on the false base station, thereby being cut off from all incoming communications from legitimate base stations.

The attacker's radio conditions may also trigger certain events in the legitimate network, like handover failures. Because of these events, some implementations in the network may take disruptive steps, such as barring even the legitimate base stations, thereby triggering service disruption. Fraud attacks with financial motives may send spam or advertising SMS messages to UEs or even try to impersonate them. There could also be a non-financial motive in which the attacker may poison UE's location or send public warning messages to create panic in public. While false base stations could be used for the good of society, such as tracking down criminals or locating lost children, they could also harm the functioning of the community with dire consequences like unauthorized surveillance, communication sabotage, unsolicited advertising, or even physical harm.

On the software side, open-source stacks are getting mature and widespread. Therefore, the increased incentive for attackers due to ever-growing connectivity and the increased feasibility of deploying a false base station are the main reasons false base station attacks are more important now than ever. Therefore, it has gotten more attention on all fronts, media, hackers conferences, academia, standardization bodies, governments, law enforcement agencies, vendors, and operators.

9.2 Detection and Mitigation

The basic principle behind detection is to collect data and perform analysis of the radio environment measurements of the network and send that data to a server for analysis. The rogue base stations (cells) can be detected and mitigated (e.g., barred) using UE-based and network-based approaches, individually or in combination. These approaches, which can be implemented in various scenarios, include base station authentication techniques, rogue base station detection using reverse mobility techniques, detection and prevention of network/RAT downgrade attacks, and signal jamming mechanism detection.

This can be categorized as below:

- A user equipment (UE) based.
- Network-based system

9.2.1 A User Equipment (UE) Based

This approach will use a rule-based strategy that works well when the measurement parameters are known. Here it determines if a rouge base station is present in the network from the view of UE. Then, the UE (mobile phone) can have a specialized application to analyze and collect the data from measurement reports. Then, this can send the data to the server on the internet for analysis.

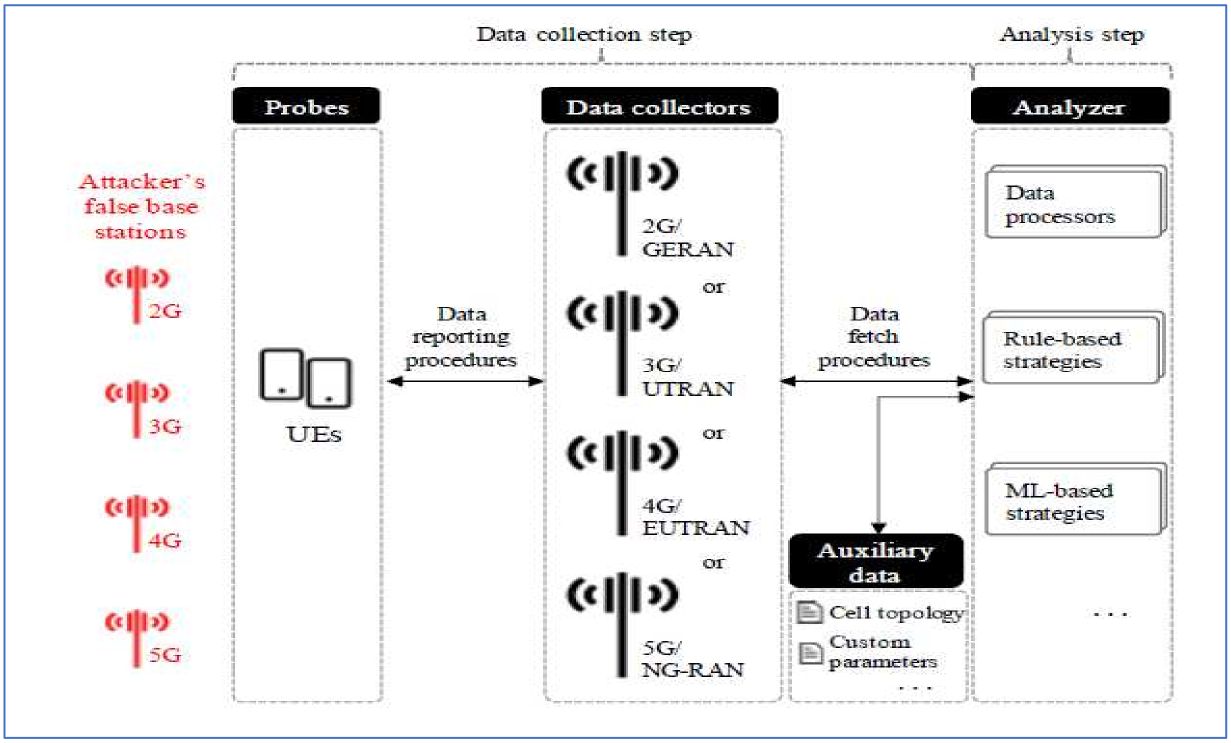


FIGURE 84: A MULTI-RAT FALSE BASE STATION DETECTION PROCESS [64]

9.2.1.1 Case Study 1

This scenario is directed to a cell selection situation where a UE powers on in a rogue base station coverage area. In this example, a UE identifies legitimate or rogue cells during an initial scan. The UE selects the cell having the strongest signal. Rogue base stations are typically configured to transmit at higher power than neighboring legitimate base stations to increase the likelihood that a UE will connect to the rogue base station rather than a legitimate base station. The UE checks the public land mobile network (PLMN) code sent by the base station to a subscriber identity module (SIM) PLMN code. If the PLMN codes match, the UE checks the S-criteria for cell selection. Assuming the rogue base station is transmitting at a higher power than the legitimate base station, the cell selection passes on the rogue cell, and the UE proceeds to camp thereon. The UE initiates an attach procedure during which the UE may send its identity before an authentication procedure is performed, thereby enabling the rogue base station to capture unencrypted UE information, such as IMSI, international mobile equipment identity (IMEI), and so on. As such, a user's private data can be stolen by the rogue base station and used to perform illegal activities.

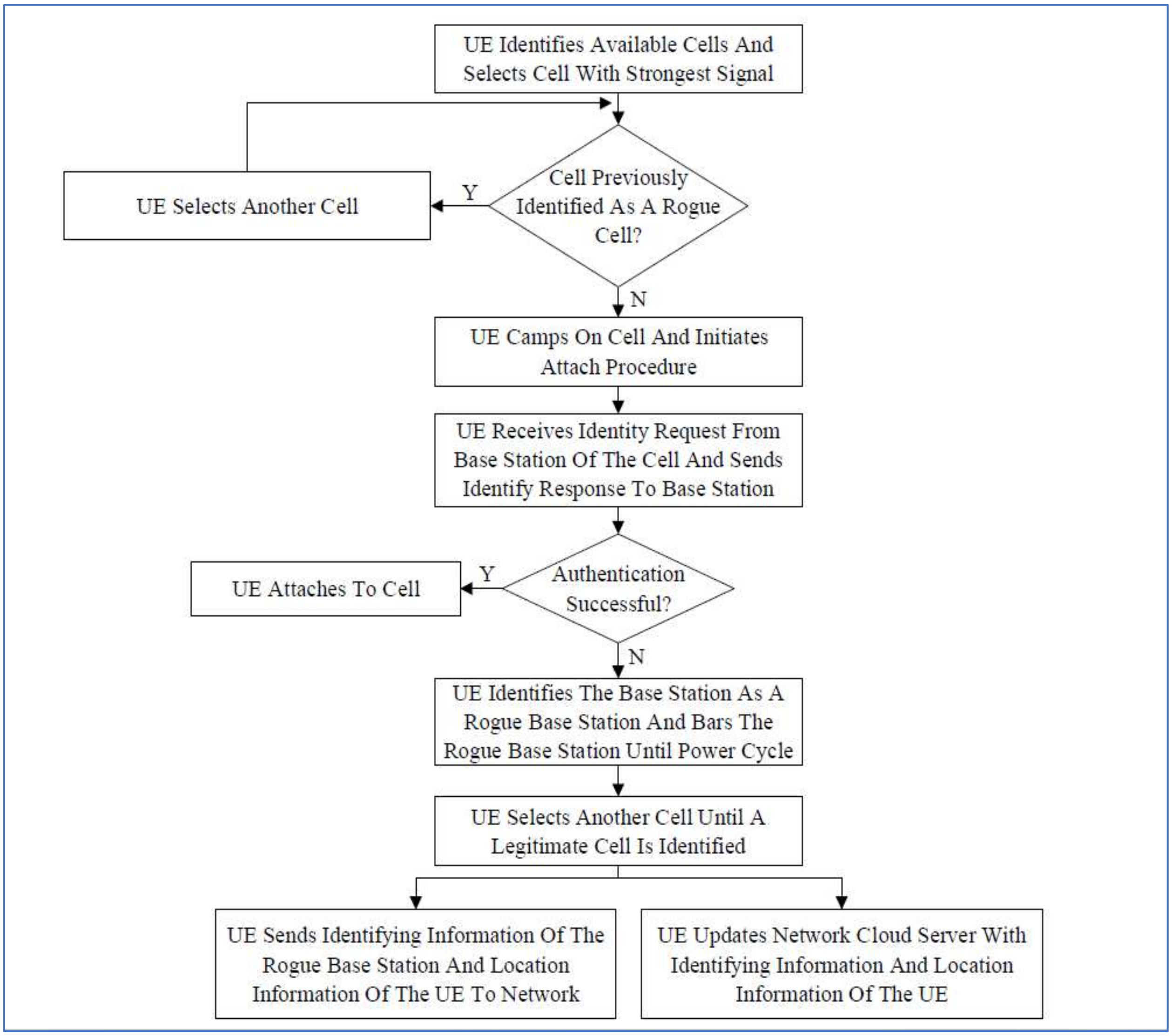

UE Identifies Available Cells And Selects Cell With Strongest Signal
Cell Previously Identified As A Rogue Cell?
Y
UE Selects Another Cell
N
UE Camps On Cell And Initiates Attach Procedure
UE Receives Identity Request From Base Station Of The Cell And Sends Identify Response To Base Station
Authentication Successful?
Y
UE Attaches To Cell
N
UE Identifies The Base Station As A Rogue Base Station And Bars The Rogue Base Station Until Power Cycle
UE Selects Another Cell Until A Legitimate Cell Is Identified
UE Sends Identifying Information Of The Rogue Base Station And Location Information Of The UE To Network
UE Updates Network Cloud Server With Identifying Information And Location Information Of The UE

FIGURE 85: CASE STUDY1 -MITIGATION PROCEDURE [65]

One approach for detecting and mitigating rogue base stations is shown below in Figure. This approach implements base station authentication techniques for detecting and mitigating rogue base stations. For example, an LTE or other standard-based network provides a robust authentication mechanism to authenticate the UE and the network. The authentication procedure occurs each time the UE powers on and attempts to attach to the network. Once the authentication procedure is completed, the UE and the network can assume each other as legitimate entities. However, the input to the authentication procedure is the ID of the UE, such as the IMSI or IMEI. This information is typically sent by the UE before the authentication and security procedures are performed. For example, an identity procedure is performed before authentication. As part of the identity procedure, the network sends an identity request to the UE, and the UE responds with an identity response, which includes the UE ID. A rogue base station mimics this process by sending the Identity request to the UE and capturing the UE's response.

The rogue base station may or may not respond with an authentication procedure. In either case, the authentication procedure fails when the base station is rogue. Therefore, once the UE sends an identity response to a base station and the authentication procedure fails, the UE determines the base station is a rogue base station (cell) and only camps on the cell during the next power cycle. The UE can store its Global Positioning System (GPS) coordinates and information, such as the cell ID, EARFCN, and PLMN ID, associated with the failed rogue base station, in a rogue base station/cell list. Therefore, the next time the UE detects a cell, the UE can compare one or more of the GPS coordinates, cell ID, EARFCN, and PLMN ID of the cell to the corresponding information in the rogue base station/cell list. If matching data is found in the rogue base station list, the UE determines the cell is a rogue base station and does not camp on the cell. Upon detecting a failed cell, the UE selects any other cell. Suppose the UE can complete a successful authentication procedure on a new (legitimate) cell and attach it to it. In that case, the UE notifies the network of the detected rogue base station(s). For example, the UE can send information associated with the rogue base station (e.g., GPS coordinates, cell ID, EARFCN, PLMN ID, etc.) to the network. The UE can send the rogue base station information to the network through a signaling message or another mechanism. Alternatively, the UE can update a cloud server of the carrier with the rogue base station information. The carrier and law enforcement can then take appropriate action on the rogue base station information based on, for example, the GPS coordinates provided by the UE. In some instances, law enforcement may deploy base stations to capture criminals. These devices can be allowed so the UE does not flag them as rogue base stations.

9.2.1.2 Case Study 2

The second scenario is directed to a cell reselection situation. In this situation, a UE device, already in idle mode and camped on a legitimate cell, comes near, and identifies a rogue cell. The UE determines that the rogue cell has a better cell strength than the legitimate cell and may reselect to the rogue cell. Cell reselection is an autonomous process the UE performs without any network intervention. The UE may initiate a tracking area update procedure in a Fourth Generation (4G) Long-Term Evolution (LTE) network. As part of this procedure, the UE may expose information specific to the UE captured by the rogue base station of the cell. Hackers can then use the captured information to extract personal information. As such, a user's private data can be stolen by the rogue base station and used to perform illegal activities.

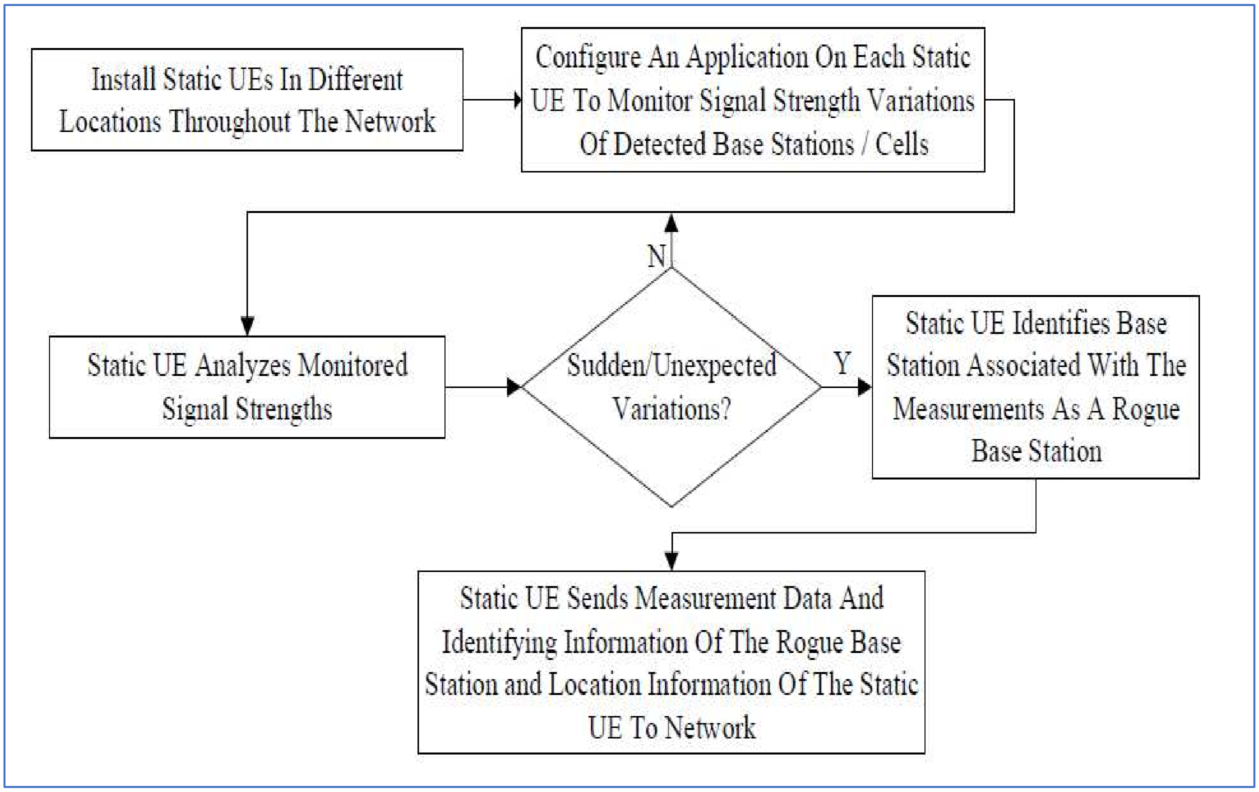

Install Static UEs In Different Locations Throughout The Network
Configure An Application On Each Static UE To Monitor Signal Strength Variations Of Detected Base Stations / Cells
N
Static UE Analyzes Monitored Signal Strengths
Sudden/Unexpected Variations?
Y
Static UE Identifies Base Station Associated With The Measurements As A Rogue Base Station
Static UE Sends Measurement Data And Identifying Information Of The Rogue Base Station and Location Information Of The Static UE To Network

FIGURE 86: CASE STUDY 2- MITIGATION PROCEDURE [65]

The second approach for detecting and mitigating rogue base stations is shown below in Figure 86. This approach implements reverse mobility techniques to detect and mitigate rogue base stations. Legitimate base stations are typically static, while UEs are generally mobile. Rogue base stations can also be portable to capture the information of as many UEs as possible. For example, rogue base stations can be installed on the top of a vehicle. The vehicle can be driven across the network to capture the details of many UEs. Mobile rogue base stations can be detected and mitigated by installing static UEs in several locations throughout the network, such as in operator stores. An application can be installed on the static UEs that monitor signal strength variations of detected cells. Wide variations in signal RSRP or Received Signal Strength Indicator (RSSI) is not expected from a well-planned legitimate base station.

However, since a mobile rogue base station moves across the network, wide signal RSRP or RSSI variations can be expected. As such, the application performs statistical analysis (e.g., mean, standard deviation, etc.) of signal characteristics, such as signal strength, using machine learning. If sudden or unexpected variations are detected, the static UE determines that a mobile rogue base station is nearby. The stationary UE determines identifying information of the rogue base station, such as the cell ID, PLMN ID, MCC, MNC, etc. static UE can send the identifying information and the static UE's GPS coordinates to the network and law enforcement authorities so that the rogue base station can be located and shut down. The static UE can also share the time plot of signal variations since the peak of the plot indicates the time when the rogue base station was close to the static UE. In another example, a network-based algorithm or technique can be implanted that configures all base stations to periodically update/send their GPS location to a network server. If one or more network components detect a significant change in a base station’s location, this base station is identified as a rogue base station. Small changes in a base station's area are sometimes filtered out to compensate for small cells.

9.2.1.3 Case Study 3

The third scenario is directed to a handover situation. In this situation, a UE is connected, identifies an intra-frequency rogue cell, and sends a measurement report to the network. The network configures intra-frequency measurements for handover/mobility. Suppose the rogue base station is transmitting in the same E-UTRA Absolute Radio Frequency Channel Number (EARFCN) as a legitimate base station. In that case, the UE may report the cell ID of the rogue base station along with the RSRP. If the cell ID is known to the legitimate base station, a handover may occur to the rogue base station. Sometimes, the authentication may already be enabled during handover, so the rogue base station may not decode the UE information. However, the UE will lose its connection to the network when the UE connects to the rogue base station. Therefore, the user will experience a service interruption.

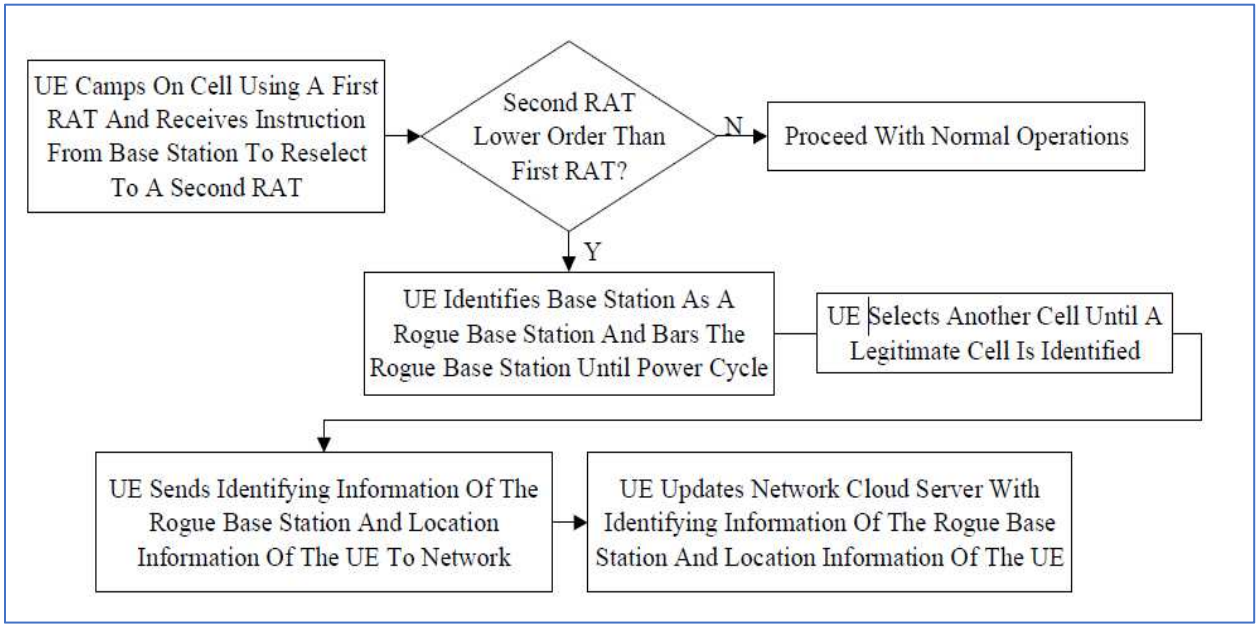

UE Camps On Cell Using A First RAT And Receives Instruction From Base Station To Reselect To A Second RAT
Second RAT Lower Order Than First RAT?
N
Proceed With Normal Operations
Y
UE Identifies Base Station As A Rogue Base Station And Bars The Rogue Base Station Until Power Cycle
UE Selects Another Cell Until A Legitimate Cell Is Identified
UE Sends Identifying Information Of The Rogue Base Station And Location Information Of The UE To Network
UE Updates Network Cloud Server With Identifying Information Of The Rogue Base Station And Location Information Of The UE

FIGURE 87: CASE STUDY 3 -MITIGATION PROCEDURE [65]

The third approach for detecting and mitigating rogue base stations is shown below in Figure 87. In this approach, network/RAT downgrade attacks are detected and prevented. Networks such as 2G (Second Generation), GSM (Global System for Mobile Communications), and 1x networks are less secure than 4G/5G (Fifth Generation) networks. As such, rogue base stations attempt to push UEs to a lower-order RAT. Due to its lesser security, rogue base stations can typically capture more UE information on a lower-order RAT. For example, a rogue base station can implement multiple RATs, such as LTE and 2G RATs. When a UE camp on the rogue base station using the LTE RAT, the rogue base station manipulates the cell reselection parameters (e.g., Cell Reselection Priority and reselection thresholds) so that the UE reselects to the 2G RAT of the rogue base station.

However, legitimate base stations typically prioritize LTE/5G over 2G. A UE can be configured to detect when the base station reverses priority and instructs the UE to reselect to a lower-order RAT. When the UE detects that the base station is instructing the UE to reselect to a lower order RAT, the UE identifies the base station as a rogue base station, and the UE does not camp on the cell until the next power cycle. The UE can store its Global Positioning System (GPS) coordinates and information associated with the failed rogue base station, such as the cell ID, EARFCN, and PLMN ID, in a rogue base station list.

Therefore, the next time the UE detects a cell, the UE can compare one or more of the GPS coordinates, cell ID, EARFCN, and PLMN ID of the cell to the corresponding information in the rogue base station list. If matching data is found in the rogue base station list, the UE determines the cell is a rogue base station and does not camp on the cell. The UE can then implement the techniques described above concerning the first approach.

9.2.1.4 Case Study 4

In this case, detection and mitigation of rouge base station that jam signals are considered. The Rouge base station may utilize a mechanism to jam signals and obstruct the signal between UE and Network by sending a signal with very high power. As a result, the rouge base station tries to connect with the UE because of a higher power, and legitimate base station signals are dropped due to less signal power.

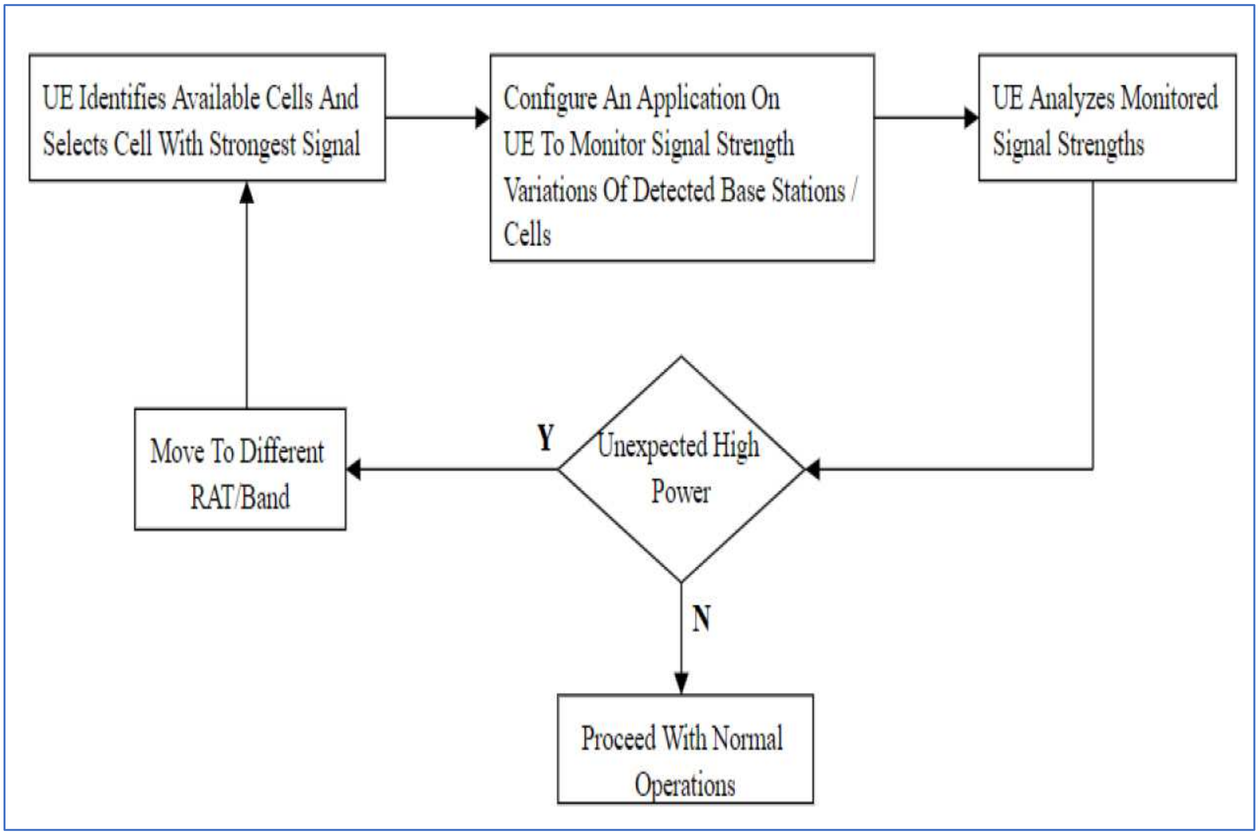

UE Identifies Available Cells And Selects Cell With Strongest Signal
Configure An Application On UE To Monitor Signal Strength Variations Of Detected Base Stations / Cells
UE Analyzes Monitored Signal Strengths
Unexpected High Power
Y
Move To Different RAT/Band
N
Proceed With Normal Operations

FIGURE 88: CASE STUDY 4 – MITIGATION PROCEDURE

The fourth approach for detecting and mitigating rogue base stations is directed to signal to jam. Some rogue base stations may utilize a jamming mechanism that sends a signal with very high power so that UEs cannot locate legitimate cells. The jamming mechanism intends to obstruct the communications between UEs and the network. Rogue base stations that utilize jamming mechanisms are typically not actual base stations but are signal generators that transmit high-power signals in the same frequency spectrum as legitimate base stations. As such, a UE can be configured to detect abnormally high RSSI, which indicates a jamming mechanism is being used on the band. When the UE detects an unusually high RSSI, the UE can attempt to find a legitimate cell on the frequency. If the UE cannot find a legitimate cell on the frequency, the UE can move to a different band or different RAT.

9.2.2 Network-Based System [64]

This approach uses a rule-based strategy that works well when the relevant measurement parameters are known. Network-based detectors can perform better in analysis because. In contrast, UEs only know their local state; a mobile network knows the global state of the system. For example, suppose there is a discrepancy between the view of the cellular network and what the network expects. In that case, this may indicate that there are unauthorized stations—keeping in view that the mobile device (UE) regularly updates the network with its local state information as a part of the regular operation of the network.

As UEs move in the area, they measure the signal strength of different base stations to connect and report the information to the network. However, it is difficult for the UE to determine whether any base station is legal or unauthorized from the collected data. On the other hand, the mobile network has information about what base station will operate in what area, what signal strength to be expected, which identifiers will be exchanged, etc. Therefore, system tuning may be required depending on the parameter and threshold involved. A limitation, however, is that the system needs an additional pre-existing network monitoring infrastructure that can collect data from various protocols or network points and supply processed data. Such network monitoring infrastructure may be absent or have widely different capabilities among mobile operators, significantly affecting detection mechanisms.

Deployment Scenario:

9.2.2.1 Step 1 - Data Collection

UEs can be used as probes for data collection without any specialized software installed on them as they exchange measurement reports with the network. Although depending on severe factors such as the transmission power, distance from UEs, the signal strength of legitimate base stations, and radio conditions, some UEs may fall victim, while others will not. Those UEs, which are sufficiently near the false base station to receive its signals but have not yet fallen victim to it, observe false cells and report them along with other legitimate cells to the operator's legitimate cells. Generally, a base station is the data collector that engages directly with UEs to receive reports using standard procedures using standard 3GPP methods. Other data collectors can be servers that manage RAN.

Data can be further classified into types:
Main Data: Measurement report. Identify the views of the network from the UEs' perspective, e.g., how many and what types of cells they observed in a particular area.

Auxiliary Data: Other than measurement reports, these data types can be used to detect false base stations. They enrich or augment the measurement reports. One example of Auxiliary data is cell topology data containing information about base stations in the operator's mobile network, like their location, cells, and RAT types. The cell topology enables the mobile network to compare the UE's view of the mobile network with its expected view.

UEs and RAN engage with each other on-air interface through standard Radio Resource Control (RRC) procedure enabling measurement report mechanism. The reporting mechanism is fundamental to all generations of mobile networks. It allows the network to decide what conditions suit an event, such as UE switching to a different base station.

First, the RAN configures UEs in a connected state for measurements using RRC messages, so-called reconfiguration and resumes. These messages contain measurement configurations with many parameters, including measurement objects and reporting configurations. Measurement objects (such as carrier frequency and cell identifiers) are the radio resources on which UE is asked to perform measurements. For example, a 4G network can configure UEs to measure 4G/intra-frequency, 4G/inter-frequency, inter-RAN, 5G/NR, 3G, and 2G.

The network can configure mobile phones to measure 5G or inter-RAT 4G and 3G frequencies. Measurement objects in 4G cover more RATs than 5G because mobility from 5G is restricted to 4G and 3G. Reporting configurations consist of reporting criteria and format. In 5G, they also contain a reference signal type that indicates the reference signal UE uses for beam and cell measurements. Reporting requirements trigger UEs to send a measurement report that can be periodic or event-based (for example, a neighbor cell's signal getting better than some threshold). Reporting format specifies quantities, e.g., numbers of cells, that UE includes in the measurement report. The reporting configuration could also indicate reporting of Cell Global Identifier (CGI) to get a complete cell identifier. Finally, a measurement object is linked to a reporting configuration by a measurement identifier that identifies a measurement configuration. Multiple measurement identifiers can link several measurement objects to several reporting configurations.

Next, the UEs measure and report according to the measurement configuration in an RRC message called measurement report. The standards mandate that UE sends this message only after successful security activation. It means the message is encrypted and integrity protected so that unauthorized parties or sniffers cannot read or modify the reports sent by the UEs. The report consists of measurement results identified with measurement identifiers to be linked to corresponding measurement configurations. As per measurement configuration, the report may further consist of measurements on serving and neighbor cells such as physical cell identifiers (PCI), received signal received power (RSRP), received signal received quality (RSRQ), signal to interference plus noise ratio (SINR).

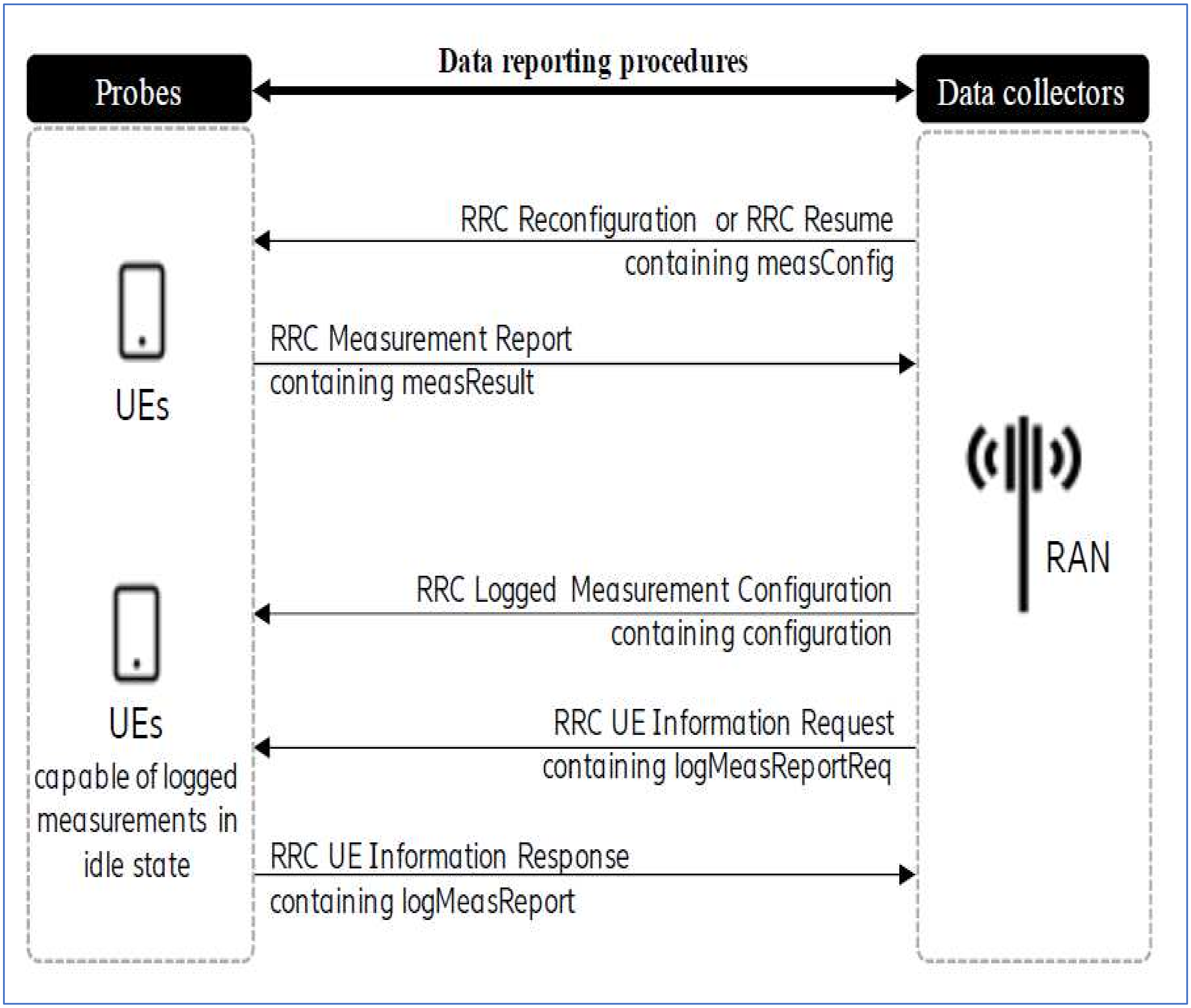
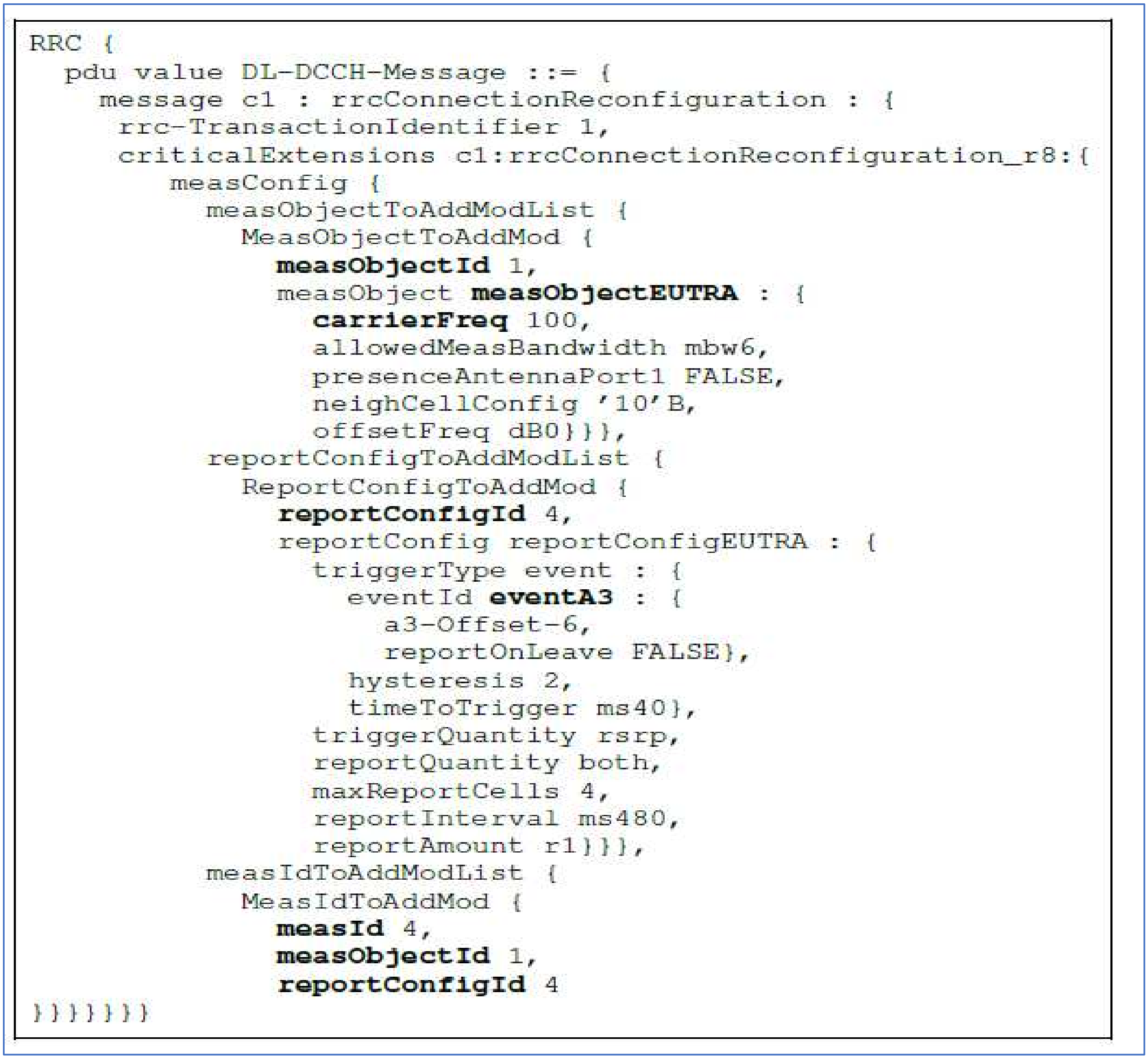

Data reporting procedures
Probes
Data collectors
RRC Reconfiguration or RRC Resume
containing measConfig
RRC Measurement Report
containing measResult
UEs
RAN
RRC Logged Measurement Configuration
containing configuration
UEs
capable of logged measurements in idle state
RRC UE Information Request
containing logMeasReportReq
RRC UE Information Response
containing logMeasReport

FIGURE 89: RRC PROCEDURES CARRYING MEASUREMENT CONFIGURATION AND REPORTS [64]

As shown in Fig. 89, there is yet another procedure that the RAN can also use to configure the capable UEs to collect logs even during idle or inactive states. It is called logged measurement configuration. For the UEs configured to perform logged measurements, the RAN can obtain the reports when the UEs come back to the connected state by sending an RRC message called information request, and UEs will return pieces in an information response message.



```
RRC {
  pdu value DL-DCCH-Message ::= {
    message c1 : rrcConnectionReconfiguration : {
     rrc-TransactionIdentifier 1,
     criticalExtensions c1:rrcConnectionReconfiguration_r8:{
        measConfig {
          measObjectToAddModList {
            MeasObjectToAddMod {
              measObjectId 1,
              measObject measObjectEUTRA : {
                carrierFreq 100,
                allowedMeasBandwidth mbw6,
                presenceAntennaPort1 FALSE,
                neighCellConfig '10'B,
                offsetFreq dB0}}},
          reportConfigToAddModList {
            ReportConfigToAddMod {
              reportConfigId 4,
              reportConfig reportConfigEUTRA : {
                triggerType event : {
                  eventId eventA3 : {
                    a3-Offset-6,
                    reportOnLeave FALSE},
                  hysteresis 2,
                  timeToTrigger ms40},
                triggerQuantity rsrp,
                reportQuantity both,
                maxReportCells 4,
                reportInterval ms480,
                reportAmount r1}}},
          measIdToAddModList {
            MeasIdToAddMod {
              measId 4,
              measObjectId 1,
              reportConfigId 4
}}}}}}}
```

FIGURE 90: EXAMPLE 4G MEASUREMENT CONFIGURATION [64]

9.2.2.1 Step 2 – Analysis

The Analyzer component performs the Analysis step. The overall goal of this step is to identify the presence of false base stations in a mobile operator's network. A high-level overview of the Analysis step is illustrated in Fig. 91.

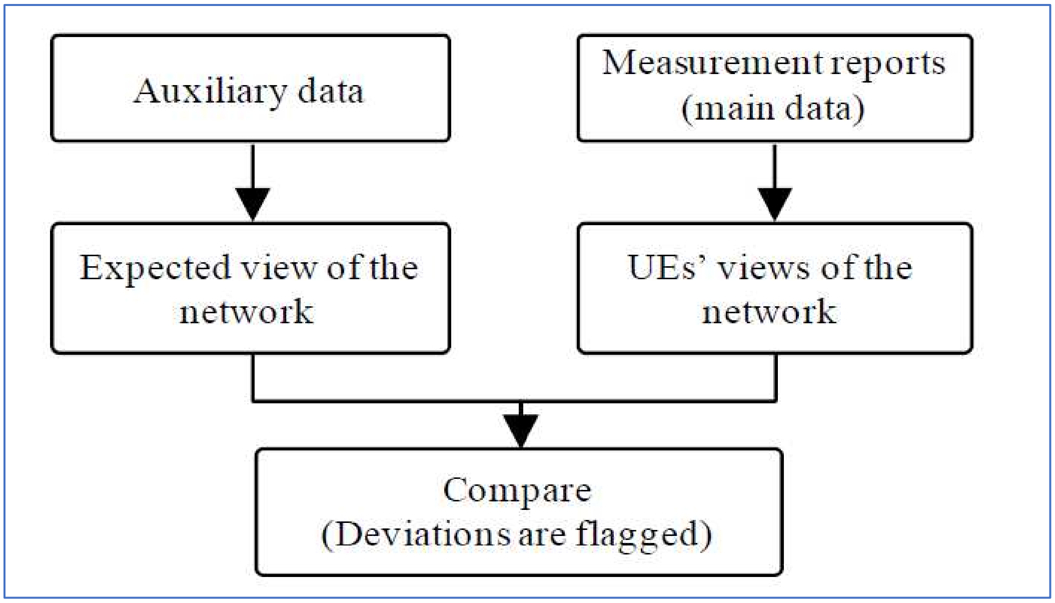

Auxiliary data
Measurement reports (main data)
Expected view of the network
UEs' views of the network
Compare (Deviations are flagged)

FIGURE 91: OVERVIEW OF ANALYSIS STEP [64]

One or more functions of the Analyzer component are executed in the Analysis step. First, the data processor function in Analyzer prepares the data for analysis by parsing the measurement reports obtained from Data collectors and Auxiliary data. It can then apply different strategies, e.g., rules-based or machine learning, which uses the processed data and detects if any information indicates the presence of a rogue base station. If there are any deviations, they are flagged and termed as rouge.

Example of Rules:

- Rule: PCIs in range 0-450! legitimate cell; otherwise! false cell
- Rule: PCIs in the range of 400-410! false cell; otherwise! legitimate cell
- Rule: PCIs were reported together with 312, 313, and 314! false cell
- Rule: PCIs other than 263 were reported between 18:00-8:00! false cell
- Rule: RSRP < -60 dBm! legitimate cell; otherwise! false cell
- Rule: RSRQ > -9 dB! false cell; otherwise! legitimate cell
- Rule: 2G/GERA cells! false cell; otherwise! check other rules!
- Rule: 3G/UTRA cells! false cell; otherwise ! check other rules
- Rule: mobile network code among (11,12,13)! Legitimate cell; otherwise! check other rules

TABLE 19: EXAMPLE OF RULES

The ranges, thresholds, and other parameters mentioned above can be either hard-coded in rules or taken as input from customization parameters as part of Auxiliary data.

9.2.3 Effectiveness and Limitations

The approaches mentioned above that have been used to detect and mitigate rogue base stations are rule-based and categorized into User Equipment (UE)- and Network-based systems. The user equipment-based technique examines the network from the perspective of the user equipment to determine whether a rogue base station is present. However, it only has local information on the system. Additionally, they require specialized applications or servers and privileged root access to collect and analyze the measurement report, reducing the measurement for analysis. Therefore, this approach may give false positives as it does not know the view of the network, e.g., if a new base station is installed, all the UEs may determine it as false as never seen before.

On the other hand, a network-based system has the view perspective of the entire network and knows the global states of the system. The approach relies on measurement reports from UEs, so it functions if at least one or a few UEs are connected to the legitimate network. If a false base station actively connects to all UEs so that no UE can send measurement reports to the legitimate network, this approach will not function.

Rule-based strategies work well if the relevant measurement parameters are known. So, it can be the first step in detecting rogue base stations before another approach. However, there could also be cases when rules are impractical, and the values of one parameter fluctuate from minimum to maximum. In such cases, a more intelligent and robust strategy is required, like using Machine Learning (ML) algorithms and cloud-based services to quickly implement a complete proof detection mechanism on UE and Network. Also, static rules are a disadvantage when the cell topology frequently changes, and manual maintenance is not practical in a large-scale deployment. Therefore, rules should automatically be updated according to cell topology for a real-time database.

Section 10. OPEN RESEARCH ISSUES AND FUTURE RESEARCH TRENDS

10.1 Open Research Issues [49]

This research paper compares the current work on the C-RAN security requirements. The result is also shown in section 7, Table 15. The table is classified according to the C-RAN logic layers that face different security threats or attacks. It observes several open issues in C-RAN security.

i) The literature needs a comprehensive and universal C-RAN security framework to fulfill all security requirements. Most existing work only concerned specific security issues regarding different planes of the C-RAN logic architecture. All existing solutions can only defend against some security threats and satisfy all security requirements.

ii) A more efficient radio resource allocation and management scheme should be studied to improve the security of the C-RAN system. Secure spectrum resource management (e.g., spectrum sensing, spectrum sharing, spectrum allocation, etc.) is considered the most critical challenge among the security requirements. However, original spectrum sensing techniques generally use energy detection methods, which only resist some radio spectrum resource threats in the complex C-RAN communication environment. For example, the centralized and virtual BBUs pool can resist the SSDF attack. But security weakness exists in that adversaries can attack the pool with massive attacks in a centralized way. Unfortunately, the literature still needs more relevant research to solve this problem.

iii) Privacy preservation has been a hot topic discussed widely. But based on different surveys, there is no related work on privacy preservation in the field of C-RAN. In many C-RAN application scenarios, the service providers must obtain user personal information, such as user locations, individual identities, and behaviors, due to business requirements. So, it is necessary to propose a C-RAN privacy preservation method to avoid the leakage of user personal information. From a user's point of view, they expect high QoS without worrying about sacrificing privacy. How to solve this problem is still an open research issue.

iv) Trust management in C-RAN is expected in practice but has yet to be seriously explored. As discussed in Section 4, trust is essential for virtualization security. In the current literature, only some schemes exist for trustworthy environment establishment in C-RAN. Most existing schemes request further investigation to show their applicability. For example, when a node overly allocates shared spectrum resources or hinders other nodes' communication, its trust rating will be judged worst. However, the availability of the model was not rigorously proven.

v) Achieving physical layer security is incredibly challenging due to the open nature of C-RAN. Therefore, physical layer security has always been a hot spot of research. However, it tried to analyze various methods to prevent physical layer attacks in the literature. Still, effective solutions for C-RAN physical layer security should be included.

vi) Finally, other open issues need to be discussed and researched, such as cloud computing security issues, virtualization security, etc. Therefore, the open issues about cloud computing security are worth exploring for achieving C-RAN security.

10.2 Future Research Trends

Based on the open research problems discussed above, several promising research directions can be suggested to motivate future research.

1) **Investigation of a Universal and Comprehensive C-RAN Security Framework**

This framework should integrate the current advance of C-RAN security technologies, which can resist various security threats and attacks in different logic layers. It should also consider all security requirements for supporting different C-RAN deployment scenarios.

2) **Investigation of a Uniform, Efficient, and Secure Authentication Mechanism**

When users access or switch a radio network node in C-RAN, this mechanism can uniformly authenticate a user and verify data security in all scenarios of the C-RAN system. The traditional core network can authenticate user identities with Evolved Packet System (EPS) and Key Agreement Protocol. However, it needs to meet the practical security requirement of C-RAN, especially for roaming and inter-operator cases.

3) **Investigation of a Security Technology that Allows Different Operators to Share the Maximum Amount of Resources in the Virtualized BBU Pool in a Trustworthy Way**

Concretely, it needs a trust mechanism to let an operator audit and monitor how many resources have been consumed by another operator, especially the ones borrowed from another operator.

4) **Investigation of New Security Solutions that Enhance the Security of the C-RAN System Based on Trust Relationships among Users and Operators**

For example, the C-RAN system can inspect users' historical trust relationships to decide whether to issue access or provide services accordingly.

5) **Investigation of a Privacy Preservation Mechanism for C-RAN**

When a service provider needs to obtain user personal information, this mechanism can prevent the leakage of user personal information.

6) **Investigation of Secure Virtualization Mechanisms in the Virtualized BBU Pool**

How to ensure the security of the virtualized BBUs pool has yet to be explored seriously in the literature, which is a promising research topic.

SECTION 11. CONCLUSION

Telecommunication networks are evolving every day. 5G is a significant step forward for communications networks. However, advanced technologies such as 5G, IoT, and virtualization services may also affect the network's security. Furthermore, the new architectures that allow 5G to progress can also expose new vulnerabilities. Therefore, securing 5G must be designed and not an afterthought. Hence, a careful approach to these unique aspects of cloud-native services, open-source software, APIs, SDN, and NFV can improve their security. Additionally, taking a Zero-Trust approach, combined with the advanced techniques of cyber threat intelligence, and Network Slicing that 5G offers, will further enhance 5G's security. The transformation to a secure 5G will occur; however, its level of success will depend on making it deployable and operational.

In addition, C-RAN has become an essential component of 5G infrastructure. 3GPP 5G standards allow physical and virtual overlap between RAN and core networks in deployed networks. It also needs to consider the technological evolution around us, which goes beyond 5G standards. This paper analyzed the C-RAN architecture and its deployment scenarios to illustrate its differences from the traditional RAN. It highlighted its specific characteristics by comparing the C-RAN with the traditional RAN. Existing security solutions of C-RAN were reviewed based on its logic layers. By applying the security requirements of C-RAN as a measure, this paper compared the existing solutions to figure out open issues and direct future research. This research explored that C-RAN security is a new research area in its infancy. A comprehensive C-RAN security framework still needs to be included in the literature. Trust management and privacy preservation are highly requested in such a framework to support advanced networking services and gain user adoption. Network operators adopt security practices that can mitigate threats like those described in this paper, DoS, MitM attacks, and configuration attacks. Network operators should consider techniques, as referenced in this paper, for more robust security, such as zero trust, multi-layer security, cross-domain solutions, post-quantum cryptography, and isolation.

In conclusion, some important insights must be considered when starting future network deployment with security in mind. First, new services are evolving rapidly and becoming omnipresent. The ability of 5G RAN to support massive bandwidth, massive interconnectivity of machines, and reliable, low-latency communication will enable various innovative applications and one not yet envisaged. Not only do these applications consume large amounts of bandwidth, but they also have stringent delay and control requirements. Therefore, 5G networks must be adaptable, resilient, and flexible to support these applications. In addition, those technologies will enable network functionality and most security controls, highlighting the importance of a comprehensive security architecture. Similarly, as more user devices such as IoT are connected, and a new diverse set of services are provided, communication security and privacy issues will become more prominent.

SECTION 12. REFERENCES

[1] “Wireless Revolution,” Wikipedia, [Online]. Available: https://en.wikipedia.org/wiki/Wireless#Wireless_revolution

[2] “The Evolution of telecom networks,” LTEMagazine, [Online]. Available: https://lte.ma/the-evolution-of-telecom-networks/

[3] V. K. Garg, "WIRELESS COMMUNICATIONS AND NETWORKING," [Online]. Available: https://www.academia.edu/19503445/Wireless_Communications_and_Networking

[4] E. Sabir, "Evolution of Telecommunication systems and rise of network heterogeneity,"[Online]. Available: https://www.researchgate.net/figure/Evolution-of-Telecommunication-systems-and-rise-of-network-heterogeneity_fig1_48908510

[5] "The Evolution of Mobile Technologies: 1G → 2G→ 3G→ 4G LTE," Qualcomm, [Online]. Available: https://www.qualcomm.com/content/dam/qcomm-martech/dm-assets/documents/the_evolution_of_mobile_technologies-wireless-networks.pdf

[6] B. I. Bakare and E. E. Bassey, "A Comparative Study of the Evolution of Wireless Communication Technologies from the First Generation (1G) to the Fourth Generation (4G)," [Online]. Available: https://www.researchgate.net/publication/355397163_A_Comparative_Study_of_the_Evolution_of_Wireless_Communication_Technologies_from_the_First_Generation_1G_to_the_Fourth_Generation_4G

[7] "1G,” Wikipedia, [Online]. Avalable: https://en.wikipedia.org/wiki/1G

[8] A. Sutton and N. Linge, "Mobile Network Architecture Evolution - 1G to 4G," [Online]. Available: https://www.academia.edu/13885065/Mobile_Network_Architecture_Evolution_1G_to_4G

[9] "2G,” Wikipedia, [Online]. Avalable: https://en.wikipedia.org/wiki/2G

[10] I. Eid, "Wireless & Mobile Networks - 7. Cellular Wireless Networks," [Online]. Available: https://www.academia.edu/11824099/Wireless_and_Mobile_Networks_7_Cellular_Wireless_Networks

[11] "3G,” Wikipedia, [Online]. Avalable: https://en.wikipedia.org/wiki/3G

[12] "LTE Telecommunication,” Wikipedia, [Online]. Avalable: https://en.wikipedia.org/wiki/LTE_(telecommunication)

[13] L. Fattouh, H. A. Salman, Z. K. Taha, N. Akkari, G. Aldabbagh, and O. Bamasak, "A survey on heterogeneous mobile networks planning intense indoor areas," [Online]. Available: https://www.researchgate.net/publication/333886291_A_survey_on_heterogeneous_mobile_networks_planning_in_indoor_dense_areas/figures?lo=1

[14] Z. Ghadialy, "R&S Webinar on LTE-A Pro and evolution to 5G," [Online]. Available: https://blog.3g4g.co.uk/search/label/LTE-Advanced%20Pro

[15] A. Komninos, "Personal Predictive Internet Content Pre-caching for Mobile Devices," [Online]. Available:https://www.researchgate.net/publication/228790601_Personal_Predictive_Internet_Content_Pre-caching_for_Mobile_Devices/figures?lo=1)

[16] "Enhanced Data rates for GSM Evolution,” Wikipedia, [Online]. Avalable: https://en.wikipedia.org/wiki/Enhanced_Data_rates_for_GSM_Evolution

[17] "5G,” Wikipedia, [Online]. Avalable: https://en.wikipedia.org/wiki/5G#References

[18] N. GUL, "5G! GOOD OR BAD?" [Online]. Available: https://www.deepcurious.com/5g-good-or-bad

[19] Y. Yu, H. Lee, and H. Jeon, "What is 5G? Emerging 5G Mobile Services and Network Requirements," [Online]. Available: https://www.mdpi.com/2071-1050/9/10/1848

[20] "Introducing 3GPP," 3GPP, [Online]. Available: https://www.3gpp.org/about-us/introducing-3gpp

[21] "IMT Vision – Framework and overall objectives of the future development of IMT for 2020 and beyond," ITU-R, [Online]. Available: https://www.itu.int/dms_pubrec/itu-r/rec/m/R-REC-M.2083-0-201509-I!!PDF-E.pdf

[22] "Release 17," 3GPP, [Online]. Available: https://www.3gpp.org/specifications-technologies/releases/release-17

[23] "ENISA Threat Landscape for 5G Networks Report," ENISA, [Online]. Available: https://www.enisa.europa.eu/publications/enisa-threat-landscape-report-for-5g-networks

[24] "Managing the Future of Cellular: What 5G Means for the Radio Access Network (RAN)," arm, [Online]. Available: https://www.arm.com/-/media/global/solutions/infrastructure/managing-the-future-of-cellular.pdf

[25] J. Saqlain, "IoT and 5G history evolution and its architecture their compatibility and future," 31 March 2018

[26] "5G System Overview," 3GPP, [Online]. Available: https://www.3gpp.org/technologies/5g-system-overview

[27] R. Goodwins, "5G New Radio: The technical background," ZDNET, [Online]. Available: https://www.zdnet.com/article/5g-new-radio-the-technical-background/

[28] "5G NR Terminologies – Subcarrier Spacing, Fram-Subframe, Slot and Symbol," Techplayon, [Online]. Available: https://www.techplayon.com/understanding-basic-5g-nr-terminologies-subcarrier-spacing-frame-and-subframe-slot-and-ofdm-symbols/

[29] W. Sitch, "MIMO in 5G Networks: Engineering & Test Challenges," LightReading, [Online]. Available: https://www.lightreading.com/mobile/5g/mimo-in-5g-networks-engineering-and-test-challenges/a/d-id/737055

[30] "5G; System Architecture for the 5G System," ETSI, [Online]. Available: https://www.etsi.org/deliver/etsi_ts/123500_123599/123501/15.02.00_60/ts_123501v150200p.pdf

[31] "An Introduction to Network Slicing," GDMA, [Online]. Available: https://www.gsma.com/futurenetworks/wp-content/uploads/2017/11/GSMA-An-Introduction-to-Network-Slicing.pdf

[32 "5G network slicing," Wikipedia, [Online]. Available: https://en.wikipedia.org/wiki/5G_network_slicing

[33] "Network slicing," Ericsson, [Online]. Available: https://www.ericsson.com/en/network-slicing

[34] "Network slicing," Ericsson, [Online]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3144

[35] "Network Functions Virtualisation (NFV) Release 2; Management and Orchestration; Architectural Framework Specification," ETSI, [Online]. Available: https://www.etsi.org/deliver/etsi_gs/NFV/001_099/006/02.01.01_60/gs_nfv006v020101p.pdf

[36] "Network function virtualization," Wikipedia, [Online]. Available: https://en.wikipedia.org/wiki/Network_function_virtualization

[37] "SDN Architecture 1.0 Overview," ONF, [Online]. Available: https://opennetworking.org/wp-content/uploads/2014/11/TR_SDN-ARCH-1.0-Overview-12012016.04.pdf

[38] "Software-Defined Networking (SDN) Definition," ONF, [Online]. Available: https://opennetworking.org/sdn-definition/

[39] "View on 5G Architecture," 5GPPP, [Online]. Available: https://5g-ppp.eu/wp-content/uploads/2019/07/5G-PPP-5G-Architecture-White-Paper_v3.0_PublicConsultation.pdf

[40] "Software-defined networking," Wikipedia, [Online]. Available: https://en.wikipedia.org/wiki/Software-defined_networking

[41] J. Silver, "SDN 101: What It Is, Why It Matters, and How to Do It," Cisco Blogs, [Online]. Available: https://blogs.cisco.com/ciscoit/sdn-101-what-it-is-why-it-matters-and-how-to-do-it

[42] P. Teppo and K. Norrman, "Security in 5G RAN and core deployments," Ericsson, [Online]. Available: https://www.ericsson.com/en/reports-and-papers/white-papers/security-in-5g-ran-and-core-deployments

[43] BLOG, "What is Cybersecurity?," CISA, [Online]. Available: https://www.cisa.gov/news-events/news/what-cybersecurity

[44] "Potential Threat Vectors to 5G Infrastructure?," CISA and NSA, [Online]. Available: https://www.cisa.gov/sites/default/files/publications/potential-threat-vectors-5G-infrastructure_508_v2_0%20%281%29.pdf

[45] "5G Security Evaluation Process Investigation," CISA and RE, [Online]. Available: https://www.cisa.gov/sites/default/files/publications/5G_Security_Evaluation_Process_Investigation_508c.pdf

[46] C. Rizzo and C. Brookson, "Security for ICT - the work of ETSI," ETSI, [Online]. Available: https://www.etsi.org/images/files/ETSIWhitePapers/etsi_wp1_security.pdf

[47] "ESF Potential Threats to 5G Network Slicing," NSA and CISA, [Online]. Available: https://media.defense.gov/2022/Dec/13/2003132073/-1/-1/0/POTENTIAL%20THREATS%20TO%205G%20NETWORK%20SLICING_508C_FINAL.PDF

[48] "The Evolution of Security in 5G- A "Slice" Mobile Threats," 5G Americas, [Online]. Available: https://www.5gamericas.org/wp-content/uploads/2019/08/5G-Security-White-Paper_8.15.pdf

[49] P. Zhang, Z. Yan, and F. Tian, "A Survey on C-RAN Security," IEEE Xplore, [Online]. Available: [https://ieeexplore.ieee.org/document/7954591]

[50] "Evolving to a strong Cloud RAN security posture," Ericsson, [Online]. Available: https://www.ericsson.com/4ae608/assets/local/ran/doc/evolving-strong-cloud-ran-security-posture-report.pdf

[51] "Open RAN Risk Analysis 5GRANR," Secunet and Federal Office of Information Security, [Online]. Available:https://www.bsi.bund.de/SharedDocs/Downloads/EN/BSI/Publications/Studies/5G/5GRAN-Risk-Analysis.pdf?__blob=publicationFile&v=5.

[52] "IDENTIFYING AND ADDRESSING THE VULNERABILITIES AND SECURITY ISSUES OF SDN," Ericsson, [Online]. Available: https://www.ericsson.com/4ac60f/assets/local/reports-papers/ericsson-technology-review/docs/2015/etr-sdn-security.pdf

[53] Y. Jarraya, "NFV builds 5G trustworthiness through security compliance," Ericsson, [Online]. Available: https://www.ericsson.com/en/blog/2020/11/nfv-security-improves-5g-trustworthiness

[54] Zhang, Shunliang, Y. Wang, and W. Zhou, "Towards secure 5G networks: A Survey," Cryptography and Information Security Lab, [Online]. Available: http://www.parkjonghyuk.net/lecture/2019-2nd-lecture/IoTSecurity/8.pdf

[55] S. Lal, T. Taleb, and A. Dutta, "NFV: Security Threats and Best Practices," [Online]. Available: http://anastacia-h2020.eu/publications/NFV_Security_Threats_and_Best_Practices.pdf

[56] I. Ahmad, T . Kumar, M. Liyanage, J. Okwuibe, M. Ylianttila, and A. Gurtov, "Overview of 5G Security Challenges and Solutions," [Online]. Available: http://jultika.oulu.fi/files/nbnfi-fe201902124647.pdf

[57] Y. Khettab, M. Bagaa, D. L. C. Dutra, T. Taleb, and N. Toumi, "Virtual Security as a Service for 5G Verticals," [Online]. Available: http://www.anastacia-h2020.eu/publications/Virtual_Security_as_a_Service_for_5G_Verticals.pdf

[58] "SECURITY CONSIDERATION FOR THE 5G ERA," 5G Americas, [Online]. Available: https://www.5gamericas.org/wp-content/uploads/2020/07/Security-Considerations-for-the-5G-Era-2020-WP-Lossless.pdf

[59] "Zero Trust security," Microsoft, [Online]. Available: https://learn.microsoft.com/en-us/azure/security/fundamentals/zero-trust

[60] "Evolving to a strong Cloud RAN security posture," Ericsson, [Online]. Available: https://www.ericsson.com/4ae608/assets/local/ran/doc/evolving-strong-cloud-ran-security-posture-report.pdf

[61] "Shared responsibility in the cloud," Microsoft, [Online]. Available: https://learn.microsoft.com/en-us/azure/security/fundamentals/shared-responsibility

[62] "SECURITY GUIDANCE FOR 5G CLOUD INFRASTRUCTURES," CISA and NSA, [Online]. Available: https://www.cisa.gov/sites/default/files/publications/Security_Guidance_For_5G_Cloud_Infrastructures_Part_I_508_Compliant.pdf

[63] "Secure Connection Requirements of Hybrid Cloud," CSA, [Online]. Available: https://cloudsecurityalliance.org/artifacts/secure-connection-requirements-of-hybrid-cloud/

[64] P. K. Nakarmi, M. A. Ersoy, E. U. Soykan, and K. Norrman, "Murat: Multi-RAT False Base Station Detector," Ericsson Research Security, [Online]. Available: https://arxiv.org/pdf/2102.08780.pdf

[65] M. Venkata, "Rogue Base Station Detection Techniques," Technical Disclosure Commons, [Online]. Available: https://www.tdcommons.org/cgi/viewcontent.cgi?article=5081&context=dpubs_series

SECTION 13. ACROYNMS AND ABBREVIATION

Term	Description
1G	First Generation
2G	Second Generation
3G	Third Generation
3GPP	3rd Generation Partnership Project
4G	Fourth Generation
5G	Fifth Generation
AF	Application Function
AI	Artificial Inteligence
AKA	Authentication and Key Agreement
AMF	Access and Mobility Function
AMPS	Advanced Mobile Phone System
APCO	the Association of Public-Safety Communications Officials
API	Application Programming Interface
ARLC	Air Radio Link Control
ARP	Address Resolution Protocol
ATM	Asynchronous Transfer Mode
AUC	Authentication Center
AUSF	Authentication Server function
BBU	Baseband Unit
BCCH	Broadcast Control Channel
BS	Base Station
BSS	Business Support System
BTS	Base Transceiver System
CAPEX	Capital Expenditure
CDMA	Code Division Multiple Access
CDS	Cross-Domain Solutions
CEPT	the Conference of European Posts and Telegraphs
CGI	Cell Global Identifier
CISA	Cybersecurity and Infrastructure Security Agency
COTS	Commercial off-the-Shelf
CPDI	Control to Data-Plane Interface
C-RAN	Cloud Radio Access Network
CSA	Cloud Security Alliance
CSRIC	Communications Security, Reliability, and Interoperability Council
CSRM	Cloud Shared Responsibility Model
Cus	Centralized Units
DC	Data Center
DDoS	Distributed Denial of Service
DMZ	Demilitarized Zone

DN	Data Network
DNS	Domain Name System
DoS	Denial of Service
DPI	Deep Packet Inspection
Dus	Distributed Units
EAP-AKA	Extensible Authentication Protocol – Authentication and Key Agreement
EARFCN	E-UTRA Absolute Radio Frequency Channel Number
EDGE	Enhanced Data Rates for GSM Evolution
eMBB	Enhanced Mobile Broadband
EMS	Element Management System
eNB	Evolved NB
ENISA	The European Union Agency for Network and Information Security
ENSURE	Enablers for Network and System Security and Resilience
ETSI	European Telecommunications Standards Institute
E-UTRAN	Evolved UMTC Terrestrial Radio Access Network
FCC	Federal Communications Commission
FDD	Frequency Division Duplexing
FDMA	Frequency Division Multiplexing
FFT	Fast Fourier Transform
FTP	File Transfer Protocol
GERA	GSM EDGE Radio Access
GPRS	General Packet Radio Services
GPS	Global Positioning System
GRE	Generic Routing Encapsulation
GSM	Global System for Mobile Communication
GSMA	GSM Association
IaaS	Infrastructure as a Service
ICT	Information and Communications Technologies
IEEE	the **Institute of Electrical and Electronics Engineers**
IMEI	International Mobile Equipment Identity
IMSI	International Mobile Subscriber Identity
IMT	International Mobile Telecommunications
IOT	Internet of Things
IP	Internet Protocol
IPSec	Internet Protocol Security
ISDN	Integrated Services Digital Networks
ISM	NFVI Security Manager
ITU-R	International Telecommunication Union-Radiocommunication Sector
LTE	Long-Term Evolution Technology
MAC	Medium Access Control

MANO	Management and Orchestration
ME	Mobile Equipment
MEC	Multi-Access Edge Computing
MFA	Multi-Factor Authentication
MIMO	Multiple Input Multiple Output
MitM	Man in the Middle
ML	Machine Learning
MLS	Multi-Layered Security
mMTC	Massive Machine Type Communication
MNO	Mobile Network Operators
MPLS	Multiprotocol Label Switching
MR	Measurement Report
MTC	Machine Type Communication
NAS	Non Access Stratum
NB	NodeB
NBI	North Bound Interface
NE	Network Elements
NFV	Networks Functions Virtualization
NFV MANO	NFV Management and Network Orchestration
NFVI	NFV Infrastructure
ng-eNB	Next-Generation eNB
NG-RAN	Next Generation RAN
NMT	Nordic Mobile Telephone
NR	New Radio
NSA	U.S. National Security Agency
NSA	Non-Stand Alone
NSD	Network Service Descriptor
NSM	NFV Security Manager
NSSF	Network Slice Selection Function
OFDMA	Orthogonal Frequency Division Multiple Access
ONF	Open Networking Foundation
ONOS	Open Network Operating System
openCIT	Open Cloud Integrity Tool
OPEX	Operational Expenditures
O-RAN	Open Radio Access Network
OS	Operating System
OSS	Operation Support Subsystem
OVS	Open Virtual Switches
PaaS	Platform as a Service
PCCH	Paging Control Channel
PCF	Policy Control Function
PCI	Physical Cell Identifiers
PCRF	Policy and Charging Rules Function

PDCP	Packet Data Convergence Protocol
PDP	Policy Decision Point
PDU	Packet Data Unit
PLMN	Public Land Mobile Networks
PQC	Post-quantum cryptography
PSK	Phase Shift Key
PUEA	Primary User Emulation Attacks
QAM	Quadrature Amplitude Modulation
QoE	Quality of Experience
QoS	Quality of Service
QPSK	Quadrature Phase Shift Keying
RAN	Radio Access Network
RAT	Radio Access Technology
RB	Resource Block
RLC	Radio Link Control
RRC	Radio Resource Control
RRH	Remote Radio Unit
RSRP	Received Signal Received Power
RSRQ	Received Signal Received Quality
RSSI	Received Signal Strength Indicator
SA	Stand Alone
SaaS	Software as a Service
SCAS	Security Assurance Specifications
SDAP	Service Data Adaptation Protocol
SDN	Software Defined Network
SDNC	SDN Controller
SDU	Service Data Unit
SECaaS	Security as a Service
SIM	Subscriber Identity Module
SINR	Signal to Interference plus Noise Ratio
SLA	Service Level Agreement
SLR	Service Level Requirements
SMF	Session Management Function
SMTP	Simple Mail Transfer Protocol
SNR	Signal-Noise Ratio
SOCs	Service Operation Centers
SS7	Signalling System 7
SSDF	Spectrum Sensing Data Falsification
SUCI	Subscription Concealed Identifier
SUPI	Subscription Permanent Identifier
sVirt	Secure Virtualization
TACS	Total Access Communication System
TCP	Transmission Control Protocol

TDMA	Time Division Multiple Access
TETRA	Trans European Trunked Radio
TPM	Trusted Platform Module
TS	Technical Specification
UDP	User Datagram Protocol
UE	User Equipment
UPF	User Plane Function
uRLLC	Ultra-Reliable and Low-Latency communication
USIM	Universal Subscriber Identity Module
UTRAN	UMTC Terrestrial Radio Access Network
vEPC	Virtual Evolved Packet Core
VIM	Virtualized Infrastructure Manager
VM	Virtual Machine
VMM	Virtual Machine Manager
VNF	Virtualized Network Functions
VSF	Virtual Security Functions
WAN	Wireless Area Network
WCDMA	Wideband Code Division Multiple Access
WSPRT	Weighted Sequential Probability Ratio Test
ZTA	Zero Trust Architecture