Research, analysis, and implementation of security attacks in 5G and IoT

Capstone Project

Presented by,

Saurabh Jingade

University of Alberta
Master of Science in Internetworking
Edmonton, Canada

Supervisor
MSc. Sandeep Kaur

Acknowledgment

I would like to express my sincere gratitude to my mentor Ms. Sandeep Kaur for providing her invaluable guidance, comments, and suggestions throughout the project.

Also, I would like to thank Prof. Shahnawaz Mir and Dr. Mike MacGregor for providing me this opportunity.

Abstract

Mobile cellular networks have been evolving over decades now. We are already in the era of 5G which is being implemented as non-standalone (i.e., over the present 4G LTE network). Hence, it is important to research the possible upcoming developments in 5G with respect to its architecture and how it can be implemented alone (i.e., in a standalone manner).

Since this is a cutting-edge technology there is not a lot of research done on this topic yet. Its behavior, security problems, and its mitigation strategies are yet to be experimented with. Hence, analyzing and implementing some of the known security attacks on the network and transport layer of 5G can give us a clear understanding of the future problems that might occur and how to tackle them.

Starting with the evolution of cellular networks with respect to security and the technologies that enable 5G network, such as massive MIMO & Beamforming, Device-to-Device (D2D) communication, Cloud-based Radio Access Network (C-RAN), Software Defined Networking (SDN), Network Function Virtualization (NFV), Cloud Computing, and Network Slicing, this report is an attempt to explore the features of 5G, its Service Based Architecture (SBA) and its core security threats.

Our goal is to analyze the 5G network and implement some attacks like the DOS, Sink-hole, and IMSI catchers, which are common network attacks on 5G, and see how it is affecting the network. Using tools like VM-ware, we can create virtual machines like Kali as an attacking machine and Windows as the victim machine and see how the attack works. Since these attacks are network attacks (i.e., they occur at the network layer), similar issues can occur in a 5G network. Tools like Wireshark are used to analyze the packet flow between the attacker and the victim machine.

All in all, we will have a clear understanding of the evolution of cellular networks along with their security threats, 5G architecture, and the technologies that enable 5G, in-depth security analysis, and a few use case implementations.

Table of Contents

List of Tables

List of Figures

List of Graphs

Chapter 1 – Evolution of Cellular Networks

1.1 Introduction

Wireless communication has been an integral part of our lives. We live in a universe of remote systems from Wireless LAN networks at home to modern Device-to-Device communication in the advanced industries. It is not easy to envision a single day without utilizing any tiny gadgets. The gift of cellular advancements furnished us with much versatility. It hence made it conceivable to tune into the radio while going in a vehicle or on the seashore. Cellular devices are also advantageous as we no longer need to stress the links that connect us to the network. We are currently facing a daily reality such that gatherings for conferences, separation, and online courses from colleges, and clinical assistance over significant distances are considered an integral part of our daily lives. We have more prominent admittance to data than any time in recent memory, and it is all conceivable because of the innovations and advancements in cellular communication.

The number of cellular users expanded drastically throughout the last decade compared to other technologies and is still increasing. This chapter talks about the evolution of cellular communication. In that regard, we start by examining the history of cellular systems. We will subsequently go through the different generations of cellular networks and have a higher-level understanding of them. As the topic is broad, we attempt to keep ourselves to the fundamental data identified with different generations' radio interfaces and network architecture. We adjust the part to the project's focal point by discussing the advancement of security estimations during every age.

1.2 1G Cellular Systems

The 1G cellular network's primary developers were the United States, Japan, and a few parts of Europe. It used analog modulation to provide voice services. In 1979, commercial cellular systems were actualized by Nippon Telephone and Telegraph Company (NTT) in Japan. Nordic Mobile Telephone (NMT-400) is a system created in 1981 that upholds international roaming and automatic handover. [1]

Some European nations implemented this system around that time. Subscribers of NMT-400 had the option to communicate up to 15 watts of power utilizing vehicle telephones. Six countries – to be specific Finland, Sweden, Norway, Austria, Spain, and Denmark – embraced NMT-400. [1]

The Advanced Mobile Phone Service (AMPS) and its alternative Total Access Communication Systems (ETACS and NTACS) were more fruitful for 1G. From the radio angle, these above systems were indistinguishable. The fundamental contrast was the length of the channel bandwidth. [1]

1.2.1 Advanced Mobile Phone Service (AMPS) Technology

The advanced mobile phone service (AMPS) was well developed compared to the other 1G systems in the United States. It was deployed in Japan and Europe by an organization named Total Access Communication Systems (ETACS). As mentioned earlier, the systems were indistinguishable from the radio stance, just varying in channel bandwidth length. For instance, AMPS was based on a 30 kHz bandwidth, while the NTACS and ETACS used 12.5 kHz and 20 kHz for the channel bandwidth, respectively. Bell Labs and AT&T first implemented the AMPS for commercial use in the year of 1983 in Chicago and its neighboring areas, then later in Israel in 1986, in Australia in 1987, and in Pakistan in 1990. [1]

All 1G cellular systems depend on analog frequency modulation for voice and data transmission and in-band signaling to move control data among terminals and the network's remainder during a call. AMPS works on the principle of frequency division multiple access (FDMA), where every user is assigned with their frequencies. This separated the user channels within the given spectrum (see Fig 1.1). FDMA creates narrowband channels, each capable of supporting one phone circuit assigned to a particular user for the call duration. The system handles the frequency assignment, and there is continuous transmission in both downlink and uplink. This specific channel that has been assigned to a user can be used to send data, voice, or nothing at all. [2]

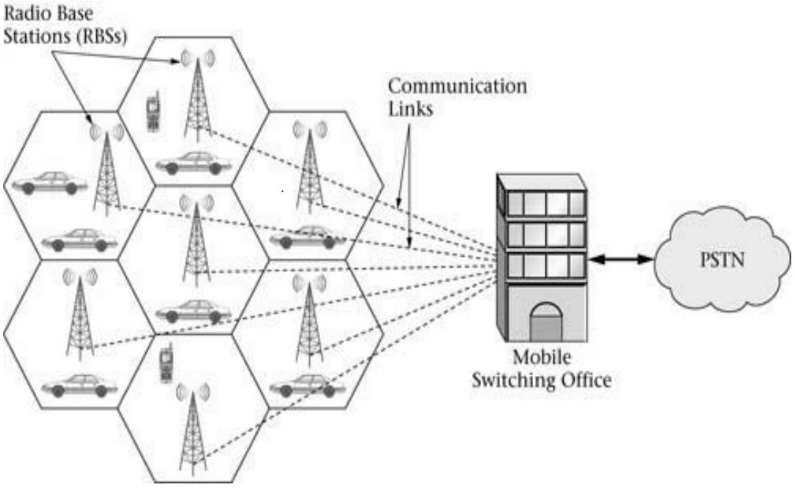


Fig 1.1: An early AMPS cellular system [2]

AMPS system components and layout include Mobile Station (MS), Radio Base Stations (BS), Communications links, and Mobile Telephone Switching Office (MTSO). Data was transmitted over Public Switched Telephone Network (PSTN) using a modem.

The MS and the BS provided Air Interface, the MTSO, now called Mobile Switching Center (MSC), is responsible for system controls like switching calls to correct cells, interacting with PSTN, monitoring traffic for billing perform diagnostic service. The MS could change its operating frequency to those diagnosed by the MSC and also its output power level if instructed. BS was mainly the interface between the MSC and MS. It received both signals and instructions from MSC.

The peak data rate for an AMPS modem call under the right conditions is usually up to 14.4 Kbps and as low as 4.8 Kbps under poor conditions. It can take anywhere up to 20 seconds or more to establish an AMPS data connection. [2]

1.2.2 Security in 1G

The first generation (1G) cellular system utilized analog communication, as stated before. Due to analog signal processing’s unstable and vulnerable nature, it was hard to provide efficient security services for 1G. For instance, eavesdropping was a pressing concern for 1G phones, as anybody could tune in to private communication between two users since all it required was a simple receiver operating at similar frequencies. There was no confidentiality in communication in 1G networks. Likewise, the cellphone’s identity could easily be duplicated, and all the call charges made from the same phone could be directed to the original owner. Since the network scale was small and a limited number of users needed servicing, the 1G cellular networks had a limited risk of mass cloning of the mobile sets. Although attempts had been made to eliminate mobile set cloning, they were proven to be unsuccessful. Even though the information about the number being dialed could be encrypted, the major problem was transmitted through the air, as signals could easily be received by using any FM receiver since the transmission used frequency modulation. [1]

1.3 2G Cellular Systems

The advancements in the processing abilities of hardware made the development of 2G wireless systems possible. A digital modulation scheme was implemented in 2G, targeting the voice market. Due to shifting from analog to digital modulation schemes, the overall performance rapidly improved. The total capacity in 2G was improved by using digital speech codecs, implementing Time Division Multiple Access (TDMA), and Code Division Multiple Access

(CDMA) techniques for multiplexing several users using a single channel. In 2G, more robust security systems were also introduced by applying encryption algorithms absent in the 1G.

Another impressive feature added to the second generation was the Short Messaging Service (SMS), which is still used today, but not for chatting purposes, but rather for security purposes like receiving One Time Password (OTP). 2G systems later evolved and supported packet data services, i.e., it started sending/receiving data in the form of packets. This was called 2.5G since this was still under transition to 3G. Unlike 2G, which utilized dial-up modems, 2.5G used Wireless Access Protocol (WAP) to provide internet access.

1.3.1 Global System for Mobile Communications

The standardized system's objective was focused on spectrum efficiency, low mobile and base station costs, international roaming, better voice quality, compatibility with other systems such as Integrated Services Digital Networks (ISDN), and the ability to support new services. The European telecommunications standards institute (ESTI) delivered the first version of the GSM standard, called the GSM Phase I, in 1990. The standard was eventually renamed the Global System for Mobile Communications (GSM).

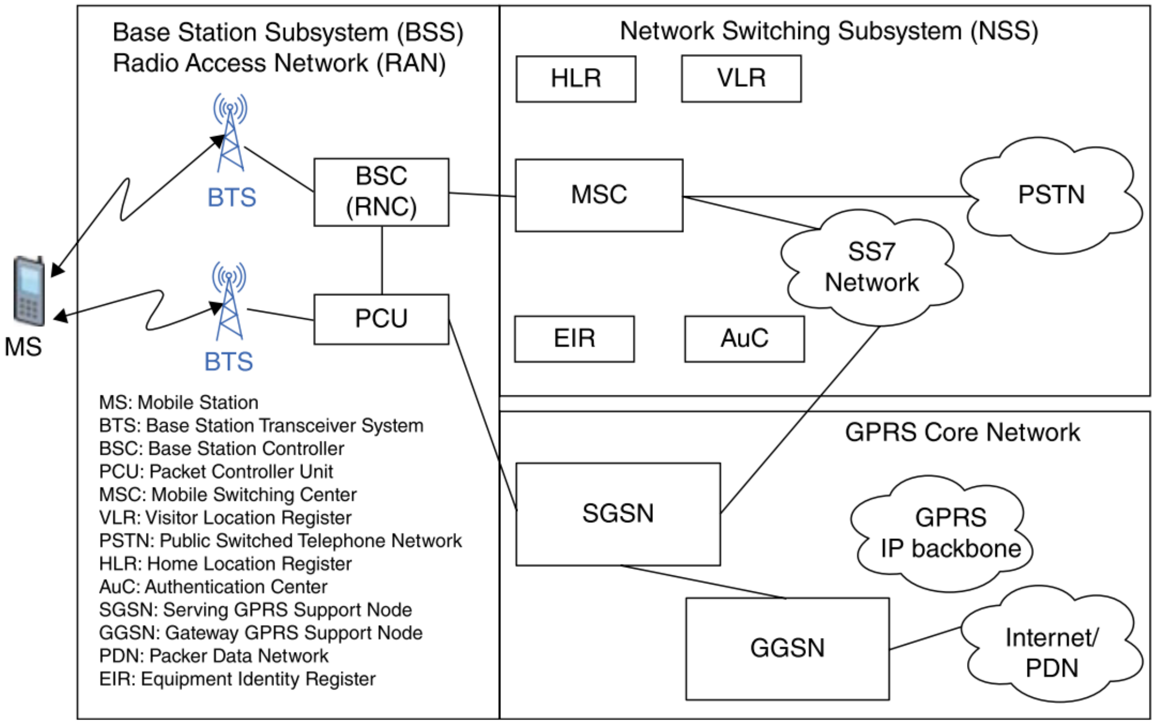
The TDMA scheme was used in the GSM air interface with the ability of multiplexing eight users in a single 200 kHz channel bandwidth, where different time slots separated the users. The GSM supported a circuit-switched data of 9.6 kbps rate, with voice and SMS service. ETSI introduced GSM Packet Radio Systems (GPRS) in the mid-1990s, the standard used in 2.5G. It was an evolutionary step of GSM systems towards higher data rates. The GPRS and GSM systems both share the same signaling links, frequency bands, and time slots. There were four different channel coding schemes to support the data, at the rates of 8 kbps to 20 kbps per slot. Theoretically, the GPRS provided a 160 kbps rate, where the 20–40 kbps rate was found in practice. [1]

1.3.2 GSM Network Architecture

In Fig 1.2, we see that the GSM Network architecture is comprised of two significant sub-components, the Base Station Sub-system (BSS), also known as Radio Access Network (RAN), and the Network Switching Subsystem (NSS). This architecture forms the basis of the next generation (3G) systems and LTE. [1]

The BSS is comprised of the Base Station Transceiver System (BTS) unit, the Base Station Controller (BSC), and the Packet Controller Unit (PCU). The BTS is responsible for effective communication between Mobile Station (MS) and the BSC via an air interface. The BSC manages traffic between BTS and Mobile Switching Center MSC. It also manages mobility across BTSs.

The NSS consists of MSC and subscriber databases. MSC carries out the switching to connect the calling party with the called party. MSC is connected with the Public Switched Telephone Network (PSTN), as shown in Fig 1.2. Home Location Register (HLR) and Visitor Location Register (VLR) are used to determine the subscriber's suggested identity for the MSC. [1]



MS: Mobile Station
BTS: Base Station Transceiver System
BSC: Base Station Controller
PCU: Packet Controller Unit
MSC: Mobile Switching Center
VLR: Visitor Location Register
PSTN: Public Switched Telephone Network
HLR: Home Location Register
AuC: Authentication Center
SGSN: Serving GPRS Support Node
GGSN: Gateway GPRS Support Node
PDN: Packer Data Network
EIR: Equipment Identity Register

Fig 1.2: GSM Network Architecture [1]

This GSM system can be upgraded to GPRS by introducing new components, such as Serving GPRS Support Node (SGSN) and Gateway GPRS Support Node (GGSN), shown in Fig 1.2. For handling data, the packet control unit (PCU) is necessary for the BTS. SGSN provides location and mobility management. The GGSN provides IP access router functionality and connects the GPRS network to the internet and other IP.

The GPRS data rate was further increased with the introduction of an Enhanced Data rate for GSM Evolution (EDGE) back in 1997. This was called 2.75G. EDGE utilized an 8PSK modulation technique that allows almost three times as much throughput compared to GPRS. An EDGE user could enjoy 80 kbps to 120 kbps of data rate. [1]

1.3.3 Security in 2G

The 2G cellular network came into the picture because of the increased demand for improving capacity, quality of transmission, and network coverage. The significant advancements in microwave devices and semiconductor technologies made digital data transmission a possibility in mobile communications. The focus was on data rather than voice as in 1G. This led to an increased demand for data confidentiality, and security became the primary concern.

1.3.4 Security in GSM

The security of 2G mainly depends on the security of GSM. GSM tries to focus on four aspects of security:

- Authentication of a user,
- Ciphering of data and signaling,
- Confidentiality of user identity, and
- Use of Subscriber Identity Module (SIM) as a security module.

Another important feature of 2G cellular networks is the Subscriber Identity Module (SIM). It is a chip containing all subscribers' information like the phone number, International Mobile Subscriber Identity (IMSI) number, the plans activated like roaming, data plans, and talk times example. All this can be used to prove the subscriber's identity with the operator and the kinds of services he is allowed to access. SIM plays a vital role in the security aspect. The authentication process requires the user to have SIM, without which the user cannot be authenticated. Ciphering takes care of the interception of all the data and signaling. GSM uses IMSI and, more particularly, uses Temporary Mobile Subscriber Identity (TMSI) to provide confidentiality for the user by ensuring that any particular user's information is in any particular area is not disclosed to anyone to avoid any intrusion of confidentiality. The SIM card uses a specific type of algorithm to develop a secure connection with the operator for further communications. Even if the SIM card is stolen, there is still a PIN code security measure in place. [1]

GSM uses symmetric algorithms like A3 and A8, which have the same key for encryption and decryption. These algorithms are implemented in the SIM card and are used for authentication between the MS and the GSM operator. These algorithms have a one-way function, meaning that output can be found if known, but the opposite is impossible. [1]

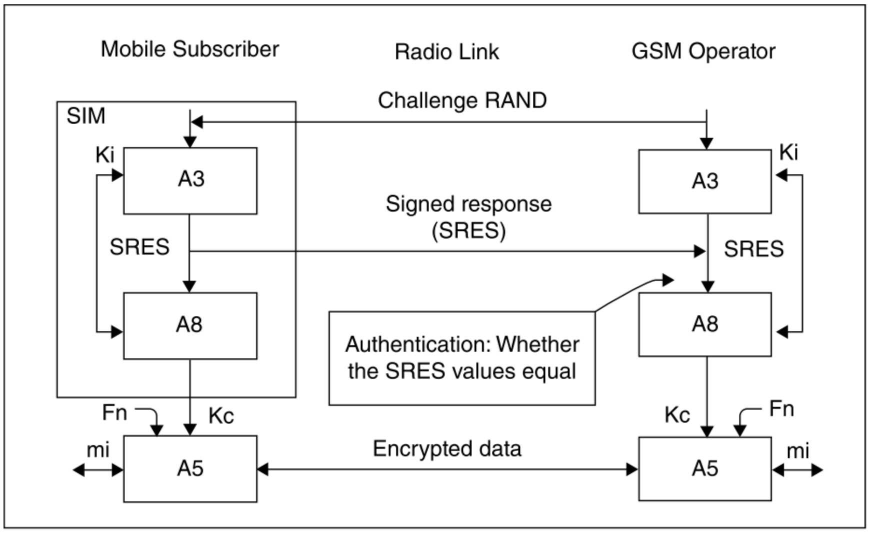


Fig 1.3: GSM authentication process [1]

1.3.4.1 IMSI

Subscriptions are identified with an International Mobile Subscriber Identity (IMSI). The IMSI is a maximum length of 15 digits. It is constructed by a Mobile Country Code (MCC), Mobile Network Code (MNC), and Mobile Subscriber Identity (MSIN). The MCC identifies the country, and the MNC identifies the network within the country. The MSIN, in turn, is a unique number for each subscriber within a particular network. The IMSI is the permanent subscription identifier, and it is used as a master key in the Home Subscriber Server (HSS). The IMSI is also stored in the User Services Identity Module (USIM - an application running on the smartcard provided by the operator). By its construction, the IMSI allows any network in the world to find the subscriber's home operator; specifically, it provides a mechanism to find the HSS in the home operator network. [3]

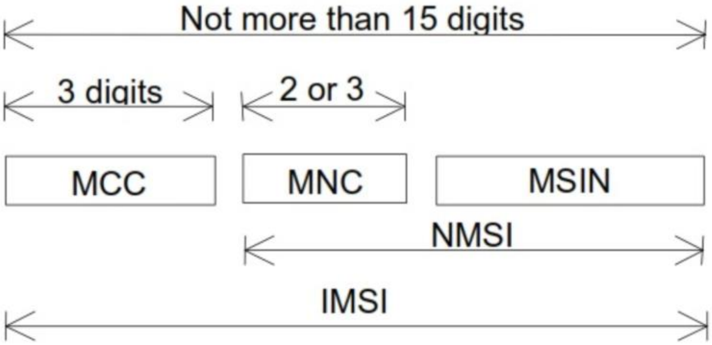


Fig 1.4 IMSI [3]

1.3.4.2 Ki

Ki is the root encryption key that is used in GSM. It is a randomly-generated 128-bit number assigned to a particular subscriber and plays a large part in generating all the GSM keys. The Ki is only known to the SIM and the Authentication Center (AuC) for protection reasons. The mobile set also has no information about the Ki, other than just feeding the information to the SIM that it needs to know to perform the authentication or generate the ciphering keys. The authentication and key generation are performed in the SIM. [1]

1.3.4.3 A3 Algorithm

The A3 algorithm provides authentication to the user so that the user can access the system. The authentication between the network and the subscriber is carried out by the so-called challenge-response method. [1]

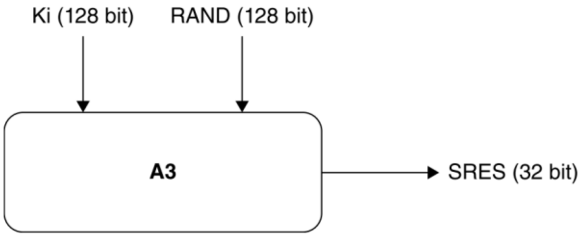


Fig 1.5: The A3 Algorithm [1]

The 128-bit number (RAND) challenge is first transmitted from the network to the subscriber through the air interface, where it is processed at the SIM card. A3 authentication algorithm and Ki are responsible for sending the RAND to the SIM card in the phone. The SIM card processes RAND and the secret 128-bit key Ki through the A3 algorithm to produce a 32-bit signed response (SRES). The A3 algorithm’s output, the SRES, is transmitted back to the subscriber’s network again through the air interface. In the network, the AuC compares SRES’s value with SRES’s value received from the subscriber. If the two values match, authentication is considered successful, and the subscriber becomes eligible to join the network. The AuC does not store the copy of SRES but takes the help of a home location register (HLR), or visitor location register (VLR) whenever required. [1]

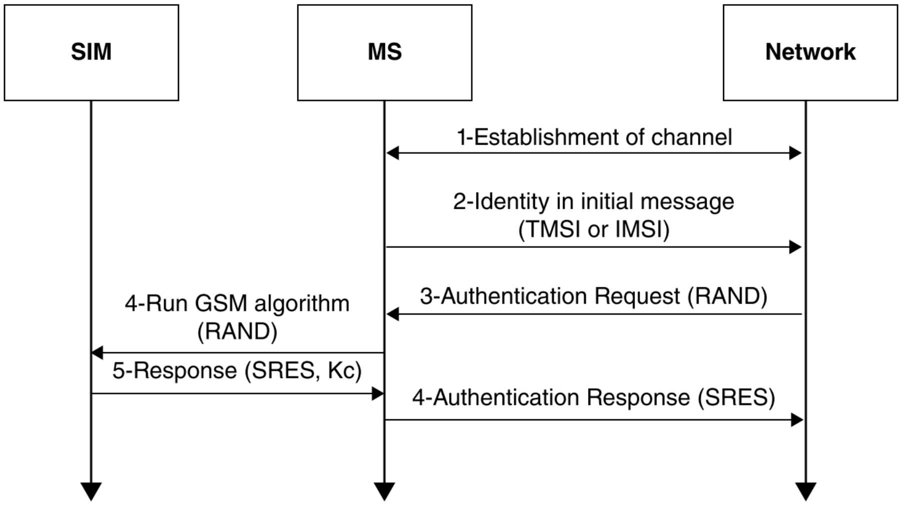

SIM
MS
Network
1-Establishment of channel
2-Identity in initial message (TMSI or IMSI)
3-Authentication Request (RAND)
4-Run GSM algorithm (RAND)
5-Response (SRES, Kc)
4-Authentication Response (SRES)

Fig 1.6: Working principle of the A3 algorithm [1]

1.3.4.4 A5 Algorithm

The A5 algorithm is a stream cipher and can be efficiently implemented on a hardware platform. Several implementations of this algorithm exist, and the most common ones are the A5/0, A5/1, and A5/2 (A5/3 is used in 3G systems). A5/1 is the most widely used, mainly in Western Europe and America, while the A5/2 is commonly used in Asia. A5/0 is used in so-called third-world countries and countries under UN sanctions, which provides no encryption. A5 works on a bit-by-bit basis, which means that error in the received cipher text will only result in the event of the corresponding bit being erroneous. [1]

GSM transmission is based on the sequence of bursts. Each burst has around 114 bits available for the information. A5/1 is used to produce for each burst a 114-bit sequence, which is XORed with the 114 bits before the modulation. A5/1 is executed using a 64-bit key together with a publicly known 22-bit frame number. [1]

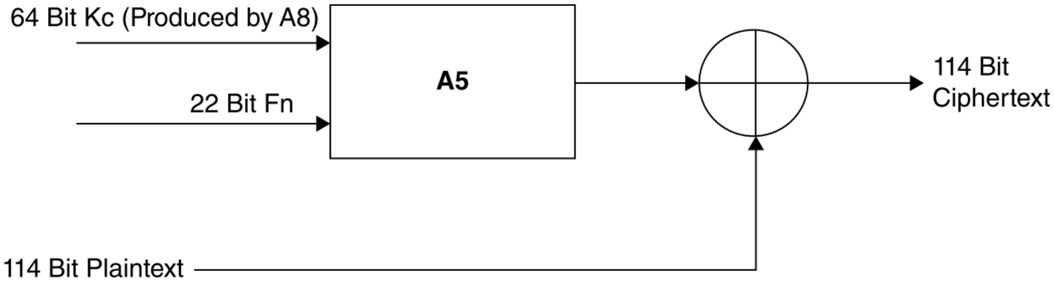


Fig 1.7: The A5 Algorithm [1]

1.3.4.5 A8 Algorithm

GSM uses ciphering to protect both user data and signaling at an air interface. Once the authentication has been successfully carried out, the RAND coming from the network together with the Ki coming from the SIM are sent through an A8 ciphering key generating algorithm to create a ciphering key (Kc). [1]

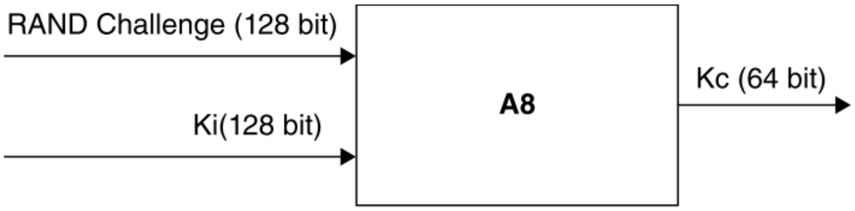


Fig 1.8: The A8 Algorithm [1]

This Kc created by the A8 algorithm is used with the A5 ciphering algorithm to cipher or decipher the data. The A5 algorithm is implemented in the mobile phone’s hardware as it encrypts and decrypts data in the air. Whenever the A3 algorithm is run to generate the SRES, the A8 algorithm also runs. Other than the A8 generating the ciphering key Kc, the network also generates the Kc and shares it with the base stations handling the connection. [1]

1.4 3G Cellular Systems

The planning for 3G was started in the early 1990s. The goal was to implement specifications for global harmony for mobile communication, which would provide global interoperability with lower costs and higher data rates and higher voice capacity, and advanced features such as multimedia applications. International Telecommunications Union (ITU), also known as IMT-200, set the requirements for the data rates as the criterion for IMT-2000:

- in building or fixed environment data rates of 2 Mbps;
- for urban environments of 384 kbps of data rates; and
- vehicular wide-area environments 144 kbps

Apart from the above requirements, 3G also intended to provide better Quality of Service (QoS) for voice telephony and interactive gaming to internet browsing, emailing, and streaming multimedia applications. [1]

1.4.1 CDMA 2000

The 3G standard for IS-95 was known as CDMA 2000 by the CDMA community. In 1999, the standard committee named the third-generation partnership project 2 (3GPP2) took responsibility for the official standardization process of CDMA 2000 from the development group Qualcomm and CDMA. CDMA 2000-1X was the first version of IS-95, where the channel bandwidth of 1.25 MHz was the same as IS-95. The data capability was enhanced by adding supplemental channels, which were separate logical channels. Each channel's capacity was 9.6 kbps, where the capacity increased to 307 kbps by using the multiple supplemental channels. As this specification of channel capacity was under 3G requirements, it was instead called 2.5G. Gradually, in the version of CDMA 2000-3X, the data rate increased to 2 Mbps by using multiple carriers. Coherent modulation was introduced to improve the uplink channel quality. The capabilities of antennas were advanced by using transmit diversity and incorporating beam steering options. The critical point of these upgrades was backward compatibility. Both A and B versions of IS-95 and CDMA could be implemented in the same carrier, convenient for migrating them. [1]

1.4.2 UMTS WCDMA

As the popularity of GSM was at its peak, the next goal was to make it universal. Hence, Universal Mobile Telecommunications Service (UTMS) came into the picture. This is a third-generation (3G) broadband, packet-based transmission of text, digitized voice, video, and multimedia at data rates up to 2 megabits per second. A collaboration was formed named 3GPP in 1998 by six regional telecommunication bodies from all over the world, and whose purpose was to continue the development of UMTS and other GSM standards.

The first UMTS standards of 3G were published in 1999, which is known as UMTS Release 99. It was a global success, with the total number of operators being 346 in over 148 countries. The number of the subscriber at that time was 450 million.

The architectural level, UMTS, was at the same level as GSM/GPRS network, but the 3G air-interface known as Wide-band CDMA (WCDMA) was a colossal modification compared to the 2G air-interface. WCDMA is a Direct Spread Spectrum CDMA system, where user data is multiplied with pseudo-random codes to provide synchronization, channelization, and scrambling. This system is designed to operate in the 5 MHz bandwidth, simultaneously supporting 100 different voice calls. The peak data rate then varies from 384 to 2048 kbps. Also, WCDMA supports using a multi-code for increasing data rate for a single user. [1]

1.4.3 UMTS Network Architecture

The UMTS network architecture consists of User Equipment (UE), UMTS Terrestrial Radio Access Network (UTRAN), and Core Network (CN). These three significant subsections are shown in Fig 1.9. The components of this architecture are derived from GSM 2G architecture, only the naming has changed, but the higher-level functionality remains the same. It is called User Equipment (UE) instead of Mobile Station (MS), as in 2G is because, in this generation, mobiles were able to perform multimedia functions like browsing data over the internet. The Node-Bs in the UTRAN is similar to that of BTS in 2G and are controlled by Radio Network Controller (RNC), which is similar to BSC, and it manages the radio resources in the data-link layer. This section has been developed for the service access point, which was absent in 2G. The CN part is also similar to the 2G, which controls different location registers. The function of SGSN and GGSN are kept the same as its ancestor, 2G. Due to this backward compatibility feature of 3G, it was easy to implement on the legacy architecture.

Even if the higher-level architecture might look the same, there are a few differences in the architecture of 2G and 3G. Here the RNCs are connected to confirm soft handover with the lowest call drop rate, whereas in 2G, BSCs were never connected. Also, the modulation scheme is a significant difference in 3G architecture.

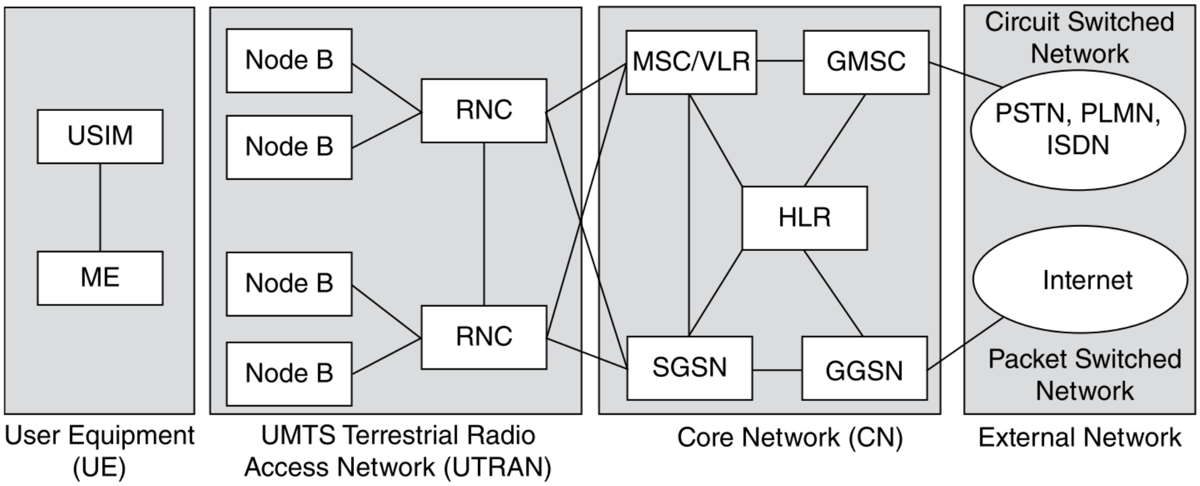


Fig 1.9: UTMS Radio Access Network [1]

1.4.4 Security in 3G

As mentioned above, the third-generation cellular networks introduced services like video, audio, and graphics applications. It additionally presented video communication and video streaming through cellular network correspondence. It was an alluring element of mobile-cellular networks. Extrapolating from the restrictions of the first era of cellular networks, it was a milestone of sorts. CDMA 2000 and UMTS CDMA went under the 3G umbrella.

3G or UMTS (Universal Mobile Telecommunications), or in particular IMT-2000, provided a single, compatible standard for cellular networks that could be used worldwide for all mobile applications. It provided support for both packet-switched and circuit-switched data communication. The security of CDMA 2000, and UMTS WCDMA, is covered below. [1]

1.4.5 Security in CDMA2000

The entities participating in the CDMA 2000 security incorporate the home network, the home location register and authentication center (HLR/AC), the serving network, the visitor location register, and the Mobile station controller/packet data serving node (VLR and MSC/PDSN), the mobile subscriber (MS), and the user identity module (UIM).

The authentication and key management (AKA) protocol used in CDMA 2000 is the UMTS AKA mechanism described in the next section. The AKA procedure involves the transfer of security credentials (Authentication Vector, AV) from the Home Environment (HE) to the Serving Network (SN).

In terms of access security, the SN network elements of interest are the PDSN, which handles packet-switched traffic, and the circuit-switched nodes VLR/MSC. An operator with a physical access infrastructure will typically have both HE and SN nodes.

1.4.6 Security in UMTS

The UMTS security architecture is gathered in five unique features, as shown in Fig 1.10. The depiction of these groups of features is given as:

1) ***Network access security***: provides the subscriber with secure access to the 3G services and gives protection against attacks to the radio interface;

2) ***Network area security***: allows all the subscribers to be able to exchange signaling data securely and protects against attacks to the wireline network;

3) ***User area security***: deals with secure access to mobile stations;

4) ***Application space security***: makes sure that applications in the user and provider domain can communicate with each other securely;

5) ***Visibility and configurability of security***: provides security information to the users as to which security features are in place and whether a specific security feature requires activation or not. [1]

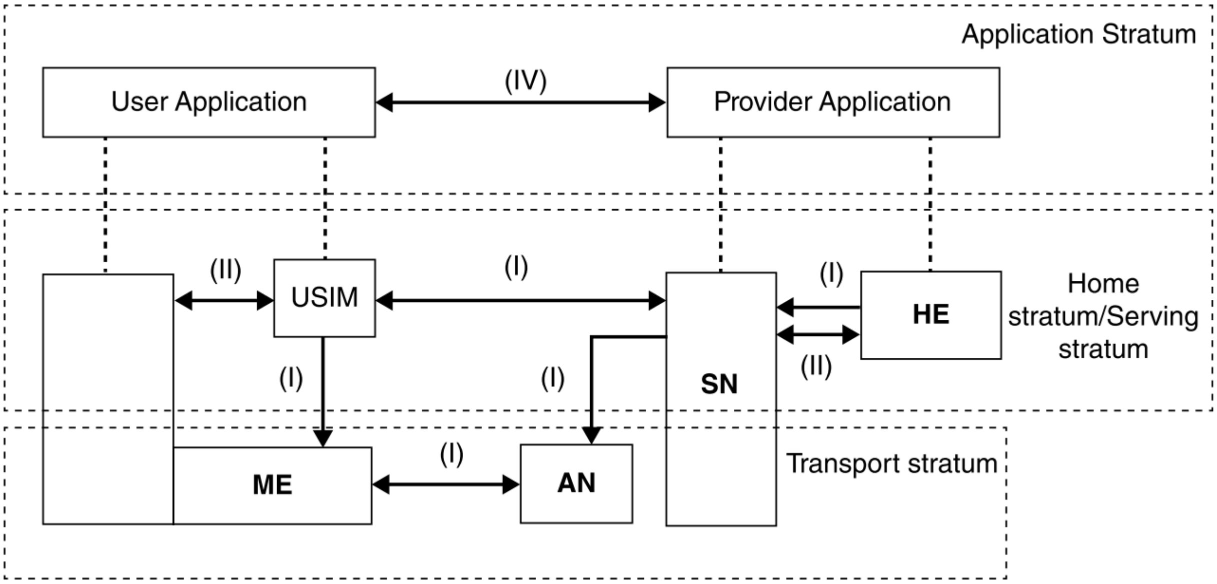


Fig 1.10: Overview of UMTS security architecture [1]

UMTS AKA is a security mechanism used to achieve the authentication features described above. This mechanism depends on the challenge/response authentication protocol executed to accomplish maximum compatibility with GSM’s subscriber authentication and key establishment protocol so that the transition from GSM to UMTS can be made.

A challenge/response protocol is a security measure utilized by one entity to verify another entity’s identity, revealing a secret password shared by the two entities involved. Each entity must prove to the other that it knows the password without revealing the information that it knows the password.

The UMTS AKA process is started by a serving network after first registration by a user, after a service request, after a location update request, after an attach request, and after a detach request or connection re-establishment request. The user’s information must be transferred from the user’s home network to the serving network to complete the process. [4]

1.5 4G Cellular Systems

The extraordinary development of the utilization of the internet was the inspiration for mobile broadband. As the cell phones were always coordinating different applications managing data, correspondence, and mode of stimulations, it was the interest of time to empower the on-demand admittance to media content from anyplace. Statics indicated that before the finish of March 2009, the number of portable broadband endorsers arrived at 225 million. To meet this massive number of services with higher performance, LTE came into the picture. Among those technologies, the key highlights of LTE are it used Multiple Access Techniques (MAT) like Single Carrier Frequency Division Multiple Access (SC-FDMA) for uplink and Orthogonal Frequency Division Multiple Access (OFDMA) for downlink. [1]

1.5.1 4G Requirements

The ITU defines the requirements of 4G systems in IMT Advanced. The requirements are:

- Having the ability to share the features worldwide will support a wide variety of services and low-cost efficiency applications.
- Internetworking capability within IMT and also with other RANs.
- Service compatibility with IMT networks.
- High-quality mobile devices.
- Worldwide roaming capability.
- 100 Mbps for high mobility and 1 Gbps for comparatively low mobility devices for supporting advanced services. [1]

1.5.2 LTE Network Architecture

The LTE network architecture is divided into two major components, Evolved-Universal Terrestrial Radio Access Network (E-UTRAN) and Evolved Packet Core (EPC). The E-UTRAN is composed of eNodeBs, which act like BTS in 2G or NodeB in 3G. It is connected to other eNodeBs via x2 interface to UE via e-Uu interface and to Mobility Management Entity (MME) via the S1-AP interface. The control interface is connected to MME, and the user interface is connected to Serving-Gateway (S-GW).

There are a few differences between UMTS architecture and the LTE systems architecture, depicted in Fig 1.12. Unlike the UMTS architecture, there are no RNC, SGSN, and GGSN blocks in the LTE. All those functions are carried by eNodeB.

The eNodeB is responsible for radio resource management. It is connected to S-GW to terminate interface towards the 3GPP radio access network and Packet Data Network Gateway (P-

GW) to control IP data services, including routing, allocating an IP address, and enforcing policy, and providing access to a non-3GPP access network.

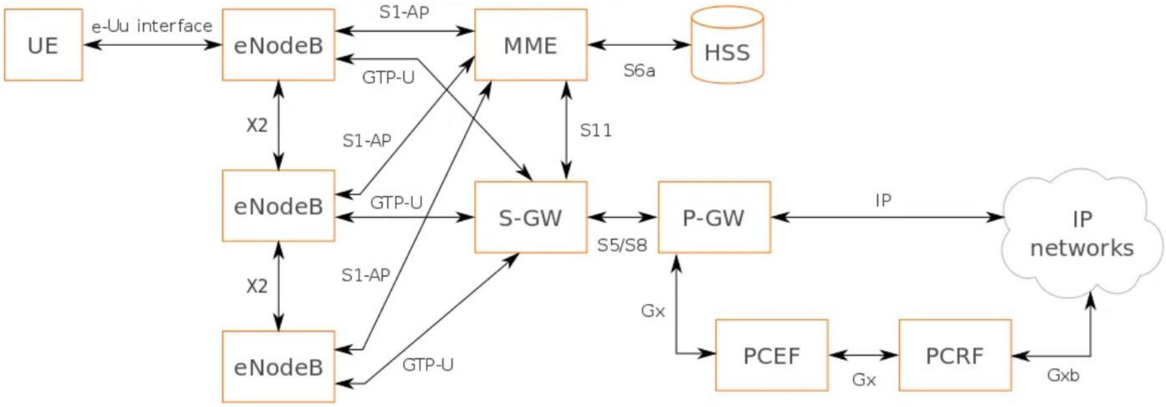


Fig 1.11: LTE Network Architecture [5]

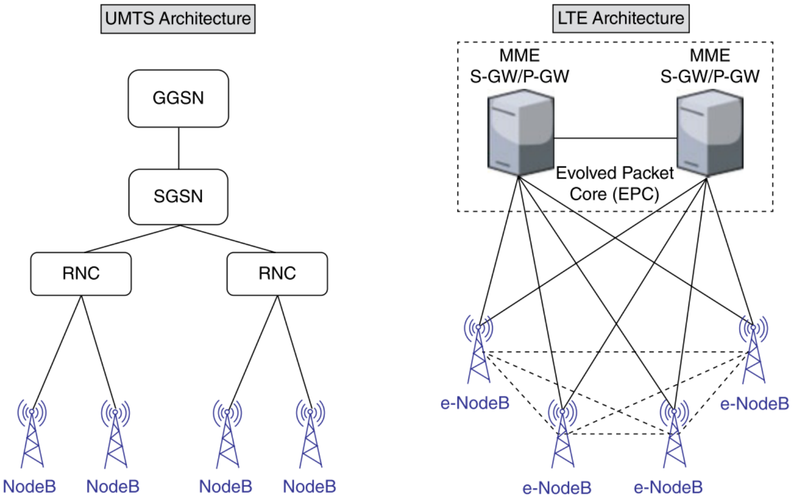


Fig 1.12: UMTS vs. LTE Architecture [1]

The function of the Mobility Management Entity (MME) is to support equipment context and identity by authenticating to the authorized users. EPC provides access control, packet routing and transfer, mobility management, radio resource management, security, and network management. [1]

1.5.3 Enhanced MIMO

Multiple-Input Multiple-Output (MIMO) is a critical technique in the cutting edge cellular system, referring to utilizing multiple antennas at both the transmitter and recipient sides. Accordingly, base stations and terminals are equipped with multiple antenna elements expected to be utilized in transmission and reception to make MIMO capabilities accessible at both the downlink and the uplink. Using MINO technology in 4G resulted in improved bandwidth and throughput, which in turn gave higher data-rates. [1]

1.5.4 Security in 4G/LTE

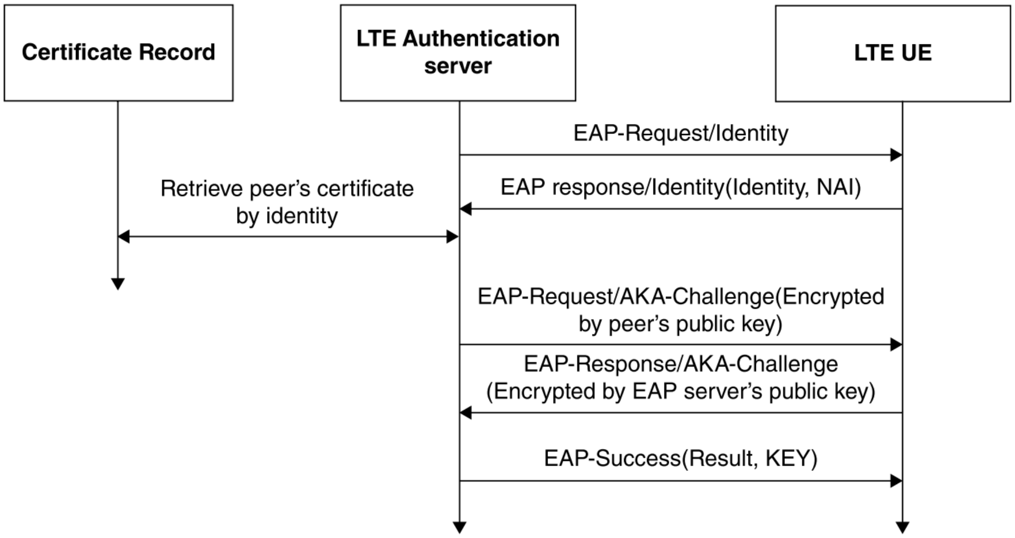

Certificate Record
LTE Authentication server
LTE UE
EAP-Request/Identity
EAP response/Identity(Identity, NAI)
Retrieve peer's certificate by identity
EAP-Request/AKA-Challenge(Encrypted by peer's public key)
EAP-Response/AKA-Challenge (Encrypted by EAP server's public key)
EAP-Success(Result, KEY)

Fig 1.13: Authentication Process in LTE [1]

Fig 1.13 shows the authentication method of LTE. The authentication server initiates the process when it sends Enhance Authentication Protocol request/Identity message (EAP) to the UE. The UE responds to this with EAP-response/Identity message containing the identity message and Network Access Identifier (NAI). Upon receipt of the EAP-response/ identity message, the authentication server tries to access the UE's certificate from its record. The authentication server then generates the EAP-Request/Authentication and Key Agreement (AKA)-Challenge message using the standard AKA process. This message is encrypted by the UE's public key and send to UE. UE then decrypts the EAP-Request/AKA- Challenge message using its private key and then sends the EAP-Response/ AKA-Challenge to the authentication server. The authentication server decrypts the information using the server's private key and verifies the EAP-response/AKA-Challenge message using the AKA algorithm. If the message is correct, the EAP server sends the EAP success message to the UE. [1]

Chapter 2 – Introduction to 5G Cellular Networks

2.1 Introduction

In the previous chapter, we have seen the history and evolution of mobile cellular networks' four generations. After over thirty years of development, mobile cellular systems have altogether changed from simple analog or circuit-based to packet-based communication systems, with likewise enormous changes in the speed and data transfer capacity improvement just as the number of connected devices. As per Ericsson's estimate, the number of connected devices is developing quickly and will reach near 28 billion by 2021, with around 16 billion Internet of Things (IoT) related devices. [1]

This massive change in users' demand and the rise of new services force a big test on the current generation of mobile cellular systems (i.e., 4G). As an outcome, it requires developing the next generation mobile systems or fifth-generation (5G), which is relied upon to be an ecosystem for each Internet-enabled gadget. In the accompanying, we will investigate more about the vision of the 5G systems and their typical use cases to perceive how it contrasts from the present 4G mobile networks.

2.1.2 Typical Use Cases

There are primarily three main categories of use cases, which have been agreed upon by most of the standardization groups, including the International Telecommunication Union (ITU) and 3GPP. These include enhanced mobile broadband (eMBB), massive machine-type communications (mMTC), and ultra-reliable and low latency communications (URLLC), as shown in Fig 2.1.

1. ***Enhanced Mobile Broadband**:* This usage scenario refers to improving the current mobile broadband services, which are typically human-centric, in terms of user-experienced data rates, traffic volume, coverage, and seamless mobility compared to services delivered in today's system. This scenario covers both broad and dense (e.g., hotspot) area coverages with different requirements. Some typical examples of 5G services in this category are high-capacity and ultrafast mobile communications for phones and infrastructure ultra-high-definition video (e.g., 4K/8K), virtual reality, augmented reality, virtual presence, and haptic feedback.

2. ***Massive Machine-Type communications**:* This usage scenario relates to deployments of a large number of connected devices, which typically transmit a relatively small amount of data such as sensors and utility meters. These devices are required to be low-cost and have

long battery life. Some typical examples of 5G services in this category are inventory control, consumer and industrial IoT, Industry 4.0 mission-critical machine-to-machine (MC-M2M), smart city, smart metering, and video surveillance.

3. ***Ultra-reliable and low latency communications**:* This usage scenario is about the capability to provide a given service with stringent requirements in terms of ultra-low latency, ultra-high reliability, and high availability, as well as high throughput. Some typical examples of 5G services in this category are vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communications, autonomous driving cars, smart grids, eHealth, Tactile Internet, remote surgery, and industrial automation.

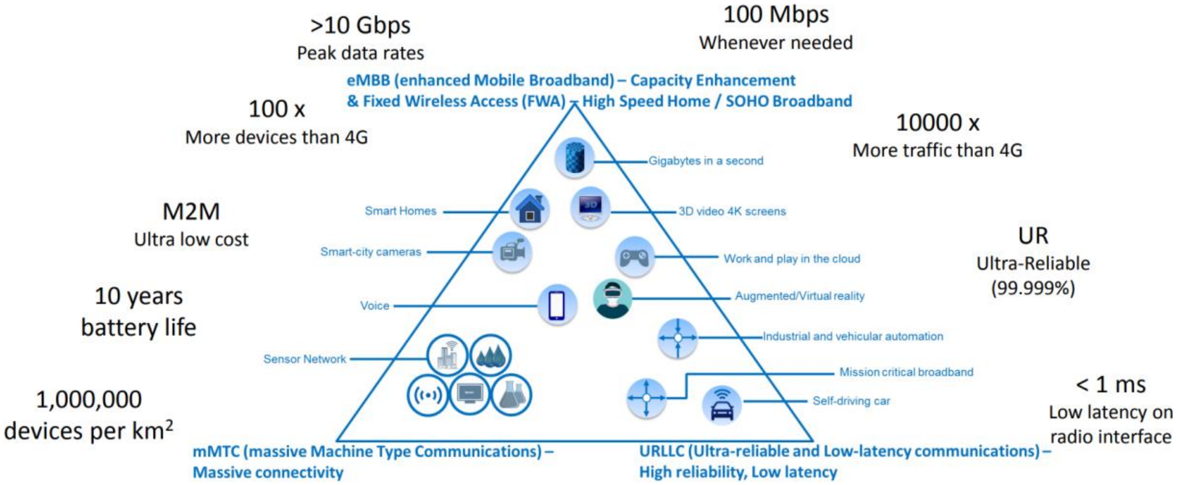

>10 Gbps
Peak data rates
100 Mbps
Whenever needed
eMBB (enhanced Mobile Broadband) – Capacity Enhancement
& Fixed Wireless Access (FWA) – High Speed Home / SOHO Broadband
100 x
More devices than 4G
10000 x
More traffic than 4G
Gigabytes in a second
Smart Homes
3D video 4K screens
M2M
Ultra low cost
Smart-city cameras
Work and play in the cloud
UR
Ultra-Reliable
(99.999%)
10 years
battery life
Voice
Augmented/Virtual reality
Industrial and vehicular automation
Sensor Network
Mission critical broadband
Self-driving car
< 1 ms
Low latency on
radio interface
1,000,000
devices per km2
mMTC (massive Machine Type Communications) –
Massive connectivity
URLLC (Ultra-reliable and Low-latency communications) –
High reliability, Low latency

Fig 2.1: Three primary use case categories as defined by ITU and 3GPP [6]

2.2 5G Requirements

ITU-R has identified some key requirements for 5G depending on the end user’s experience, system performance, services, and operation and management. Fig 2.2 shows the key requirements of 5G and some examples for each.

2.2.1 High-Data Rate and Ultra-Low Latency

To evaluate the user quality and experience in wireless communication systems, data rate and latency are two major evaluation metrics. These two matrices are key to satisfying the user's quality of experience when it comes to developing 5G.

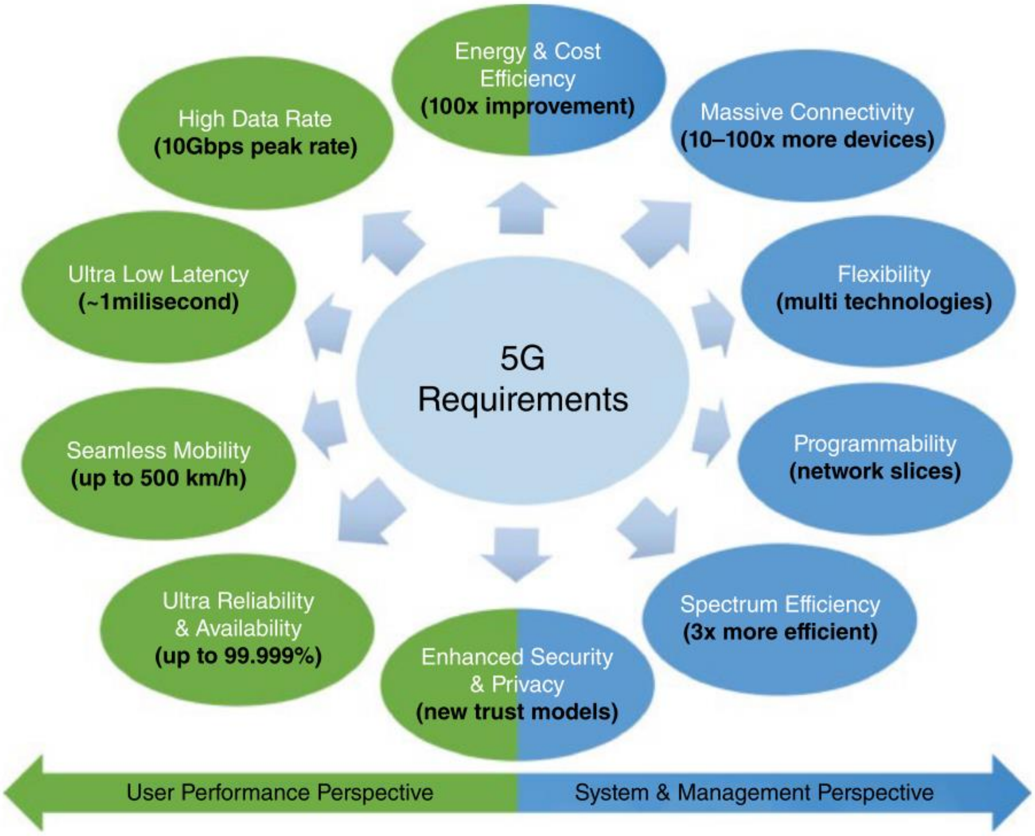


Fig 2.2: 5G key requirements and some example values [1]

The data rate requirements are expressed in terms of the peak data rate, which is the maximum achievable data rate for a user under ideal conditions, and the user experienced data rate, which is the achievable data rate for a user in the real network environment. 4G offers users a maximum peak data rate of 1 Gbps, while the maximum data rate experienced by the user is around 10 Mbps. However, for applications like virtual reality, UHD video streaming which requires more bandwidth, 5G was introduced. The peak data rate in 5G is enhanced up to 20 Gbps, while the user-experienced data rate is 1 Gbps.

Another fundamental requirement is latency, which is the end-to-end latency perceived by the end-user. With the technologies like self-driving cars, and automatic traffic control, which require real-time responses and interactions, minimizing the latency is becoming more crucial. 5G is expected to reduce the latency by ten times in the user plane, which is down to 1 ms. [1]

2.2.2 Massive Connectivity and Seamless Mobility

Massive connectivity applies to the need to endorse a significant range of connected devices and thus a large number of connections in the area unit. In the 5G age, the rise in the number of devices is not only from the introduction of different types of services and new types of devices such as sensors, smart devices, and cars, but is also from the exponential increase in the number of existing devices. Thanks to the prevalence of such smart devices, the 5G infrastructure is supposed to accommodate a link density of up to 1 billion connected devices per square kilometer or in a different context, 100 times more devices than the 4G system.

In addition to the obligation to support a wide range of wired devices, the 5G is also supposed to have a smooth service experience for smartphone devices. 5G is expected to enable an acceptable service experience for mobile devices moving at speeds of up to 500km/hr. [1]

2.2.3 Reliability and High Availability

In general, the reliability of the system relates to the ability to maintain the success rate of data transmission under defined constraints (e.g. a latency budget) for a given period. As described above, there are several third-use services and applications (i.e. ultra-reliable and low-latency communications) such as public safety, eHealth, automatic traffic control, and mission-critical services that require extremely high communication reliability. To support these types of services, the 5G system is expected to guarantee a reliability rate of up to 99,999 percent.

To provide services to end-users anywhere at any time, the 5G system must ensure its availability, which refers to the ability to withstand possible failure scenarios. Availability is usually expressed as a percentage of uptime over a given period (e.g. year and is assessed based on the number of nines in the digits (e.g. 99.99 percent). The 5G system should guarantee the availability rate with as many nines as possible e.g. five nines or 99.999 percent. [1]

2.2.4 Flexibility and Programmability

Flexibility and programmability are two network-driven requirements. As the 5G system will integrate multiple technologies to support a large number of devices and services, its network architecture should be flexible to meet a range of different properties-related requirements and attributes exposed to those devices and services. The flexibility of the network can be expressed in terms of the ability to support different types of radio access technologies, the ability to scale network resources on-demand and independently between the radio access network and the core network, as well as between the control plane and the data/user plane, the ability to install new services and applications in a very short time; and the ability to reshape the network infrastructure in real-time to adapt to changes in user or customer requirements.

Also, the 5G network infrastructure should be programmable and reconfigurable. The 5G network infrastructure will be constructed as a set of different logical virtualized networks or "slices" over the same physical infrastructure. Network programming enables on-demand and autonomous networking where mobile operators can define, program, and configure their own network 'slices' in accordance with their policies and their defined use cases. [1]

2.2.5 Energy, Cost and Spectrum Efficiency

In addition to improving network capacity and improving user experience, the design of 5G must take account of energy and cost-efficiency. In particular, the 5G system is expected to achieve a 100-fold improvement in energy efficiency compared to the current 4G system. In the meantime, the cost efficiency, which is the economic aspect of the 5G system, must be increased to guarantee the revenue of the mobile operator.

Furthermore as previously mentioned, 5G will be powered not only by human-centered devices such as smartphones but also by a wide range of “things" such as sensors, smart meters, etc. These items need much longer battery life to work in the field without any extra power source, such as at least 10 years of battery life.

Last but not least, relative to today's 4G scheme, bandwidth quality can also be greatly increased. For instance, for enhanced mobile broadband, it should be three times higher. [1]

2.2.6 Security and Privacy

Security is another significant factor that needs to be taken into account in the implementation of 5G, aside from the above criteria. Indeed at the age of 5G, the proliferation of diversified networks and devices would bring many challenges to ensuring protection. More precisely, 5G protection would need to be guaranteed at various levels, including the level of connectivity, level of technology, and level of operation. For example, 5G network architecture will be more transparent and programmable at the infrastructure level, allowed by a range of technologies such as SDN, NFV, network slicing, etc., thus pushing new security criteria such as how to securely guarantee the communication channel between the control and data planes as SDN is implemented, and how to securely separate and handle network slices.

Privacy concerns still need to be taken into account in the production of 5G. Indeed, vast amounts of user applications will be accommodated by 5G networks, meaning that a significant volume of user privacy knowledge such as user identities will be transported across the 5G network. Moreover, new forms of user identities will also be implemented in 5G, such as identifiers for IoT applications. Therefore, an effective way to handle this large volume of information is necessary, as well as to secure and avoid the leakage of personal information from users. [1]

2.3 5G Service Based Architecture (SBA)

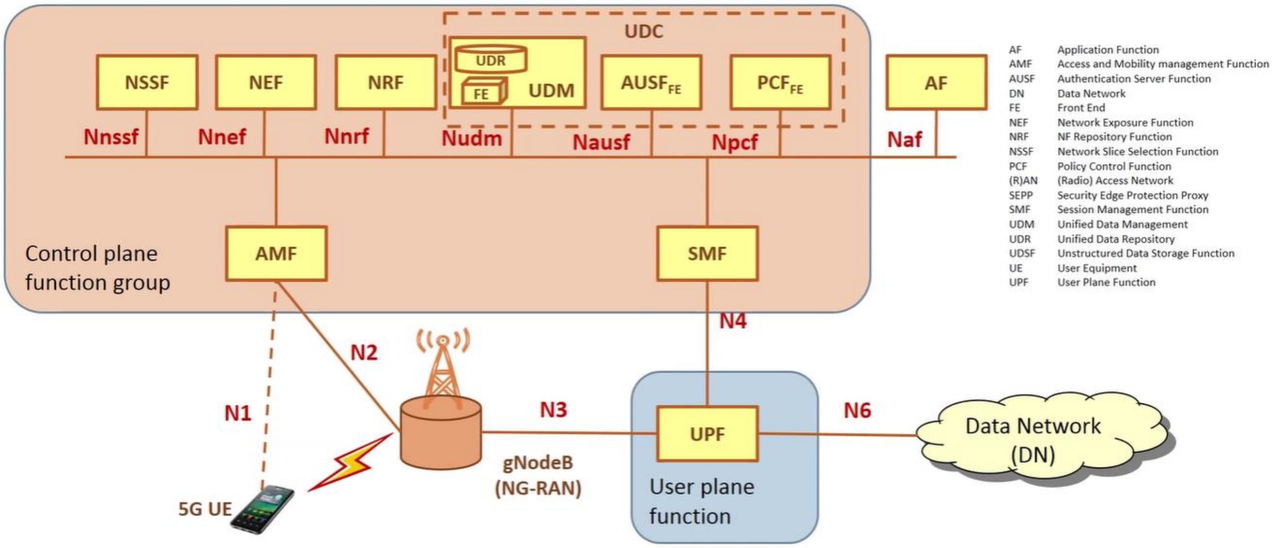

UDC
NSSF
NEF
NRF
UDR
FE
UDM
AUSF FE
PCF FE
AF
Nnssf
Nnef
Nnrf
Nudm
Nausf
Npcf
Naf
Control plane function group
AMF
SMF
N1
N2
N3
N4
N6
5G UE
gNodeB (NG-RAN)
UPF
User plane function
Data Network (DN)
AF Application Function
AMF Access and Mobility management Function
AUSF Authentication Server Function
DN Data Network
FE Front End
NEF Network Exposure Function
NRF NF Repository Function
NSSF Network Slice Selection Function
PCF Policy Control Function
(R)AN (Radio) Access Network
SEPP Security Edge Protection Proxy
SMF Session Management Function
UDM Unified Data Management
UDR Unified Data Repository
UDSF Unstructured Data Storage Function
UE User Equipment
UPF User Plane Function

Fig 2.3: 5G Service Based Architecture [7]

- **User Plane Function (UPF):**
  The U-Plane function (UPF) in the 5G core network provides functions specific to U-plane processing the same as S-GW-U and P-GW-U in CUPS. The UPF connects to the Data Network (DN) via the N6 interface.

- **Session Management Function (SMF):**
  The control plane functionality has been slightly re-organized in the 5G core network. Instead of MME, S-GW-C, and P-GW-C in EPC, the functionality has been divided between Session Management Function (SMF) and Access and Mobility Management Function (AMF). There can be multiple SMFs associated with the UE. One for each slice.

- **Access and Mobility Management Function (AMF):**
  AMF is a single node to manage all UE-related functions. The EPC functionality of MME, S-GW-C, and P-GW-C has been reallocated so that all access and mobility

functionality is done by AMF. It is connected to 5G UE via the N1 interface and to gNB via the N2 interface. There can be a single or set of AMFs for a single UE.

- **Network Slice Selection Function (NSSF):**

  It is introduced recently and supports the following functionality:

  1. Selecting the set of Network Slice instances serving the UE,
  2. Determining the allowed Network Slice Selection Assistance Information (NSSAI) and, if needed, the mapping of the Subscribed S-NSSAIs,
  3. Determining the AMF set to be used to serve UE, or, based on configuration, a list of candidate AMF(s), possibly by querying the NRF.

- **Network Exposure Function (NEF):**

  A Network Exposure Function (NEF) having a function similar to the Service Capability Exposure Function (SCEF) in EPC.

- **Network Repository Function (NRF):**

  Different Network Functions (NFs) are connected via a uniform interface called a service-based interface. Also, an individual NF consists of a smaller unit function called NF services, and an NF service in a certain NF can directly access an NF service in another NF without having to pass through another node. A Network Repository Function (NRF) provides a discovery function for NF services.

- **User Data Convergence (UDC):**

  It ensures data consistency and simplifies the creation of new services by providing easy access to the user data, as well as ensuring the consistency of storage and data models and to have minimum impact on traffic mechanisms, reference points, and protocols of network elements.

  1. **Unified Data Management (UDM):**

     It is analogous to the Home Subscriber Server (HSS) in EPC architecture and introduces the concept of User Data Convergence (UDC) that separates the User Data Repository (UDR) storing and managing subscriber information from the front end processing subscriber information.

     a. **Unified Data Repository (UDR):**

     UDR is a facility where user data can be accessed stored and managed commonly.

     b. **Front End (FE):**

     It is a core network functional entity or service layer entity or provisioning entity that can access user data stored in a unique repository. Front End identifier is defined as a name that uniquely identifies an FE within the set of all FEs accessing a UDR.

2. **Authentication Server Function (ASF):**
   The front-end section includes new specifications for an Authentication Server Function (AUSF) dedicated to authentication processing.
3. **Policy Control Function (PCF):**
   The front-end section includes new specifications for a Policy Control Function (PCF) corresponding to the Policy and Charging Rule control Function (PCRF) in EPC.

- **Application Function (AF):**
  The application function (AF) fulfills the role of an application server. It interacts with the 3GPP Core Network to provide services. [8]

2.4 5G Deployment Options and Migration Strategy

Standalone (SA) is when one Radio Access Network (RAN) is connected to one Core Network (CN). Non-Standalone (NSA) has two different RANs connected to a single CN, which means one RAN has to be connected to another RAN for signaling and transferring data, it cannot be deployed by itself.

- **Option 1 - SA LTE connected to EPC - Legacy:**
  This is the legacy standalone 4G LTE network architecture where eNB is connected to EPC.

- **Option 2 - SA NR connected to 5GC:**
  This is a standalone 5G (i.e., it is not dependent on any other network architecture like LTE). It is the only option for greenfield 5G operators. It fully supports new 5G applications and services including Enhanced Mobile Broadband (eMBB), Massive Machine-Type Communications (mMTC), and Ultra-Reliable Low Latency Communications (URLLC). It is expected to be live in the far future, maybe around the year 2025 mainly because it needs multiple spectra to provide all the above-mentioned services with ubiquitous 5G coverage.

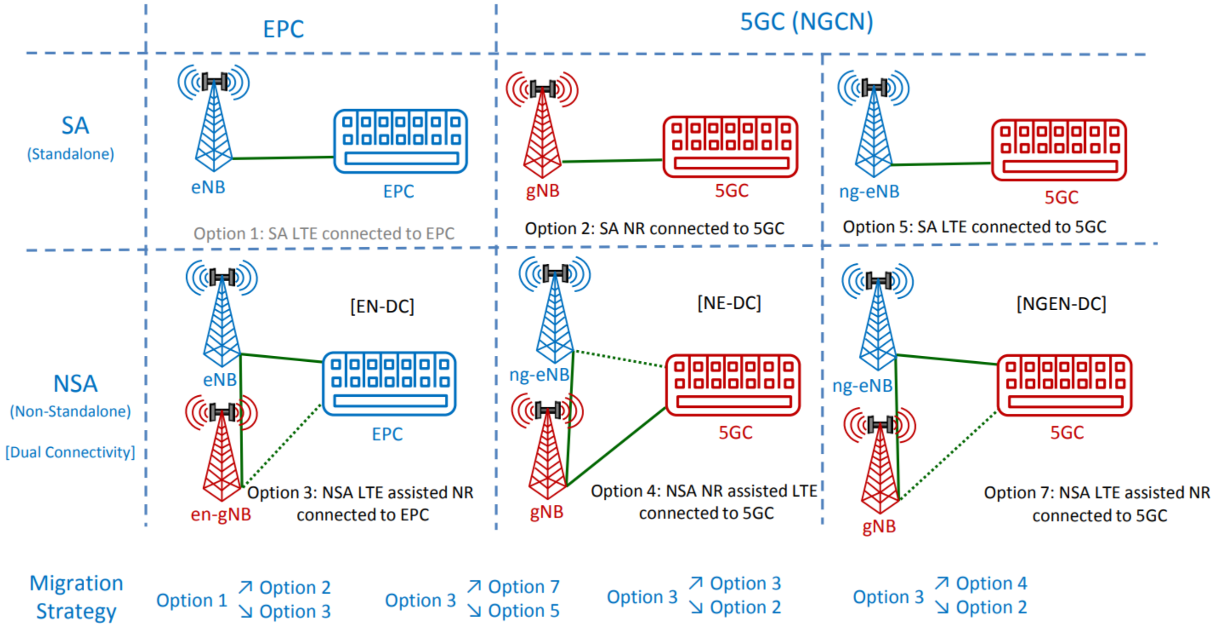

EPC
5GC (NGCN)
SA
(Standalone)
eNB
EPC
Option 1: SA LTE connected to EPC
gNB
5GC
Option 2: SA NR connected to 5GC
ng-eNB
5GC
Option 5: SA LTE connected to 5GC
NSA
(Non-Standalone)
[Dual Connectivity]
eNB
en-gNB
[EN-DC]
EPC
Option 3: NSA LTE assisted NR connected to EPC
ng-eNB
gNB
[NE-DC]
5GC
Option 4: NSA NR assisted LTE connected to 5GC
ng-eNB
gNB
[NGEN-DC]
5GC
Option 7: NSA LTE assisted NR connected to 5GC
Migration Strategy
Option 1 ↗ Option 2 ↘ Option 3
Option 3 ↗ Option 7 ↘ Option 5
Option 3 ↗ Option 3 ↘ Option 2
Option 3 ↗ Option 4 ↘ Option 2

Fig 2.4: 5G Deployment Options [9]

- **Option 5 – SA LTE connected to 5GC:**

  This is the next future technology, where the current eNB is upgraded to ng-eNB, which is then connected to the 5GC. This will allow LTE to support legacy devices and at the same time connect to the 5GC network. It is deployed in conjunction with option 2 and can provide some of the benefits that 5G NR provides in conjunction with 5GC. Allows operators to get rid of the existing 4G LTE (Option 1).

- **Option 3 – NSA LTE assisted NR, connected to EPC:**

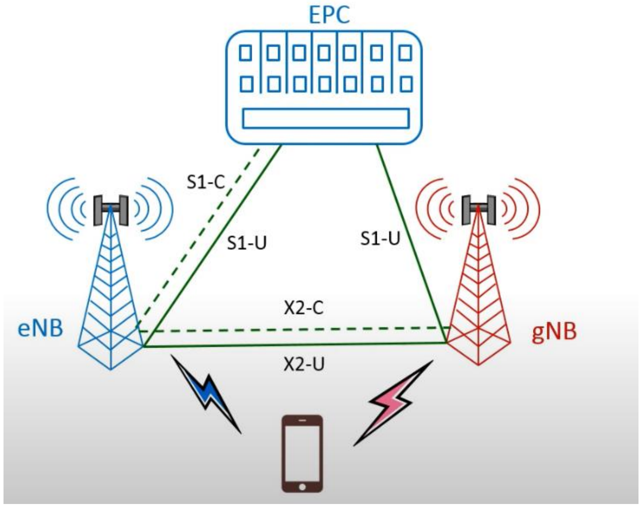


Fig 2.5: EN-DC (E-UTRA-NR Dual Connectivity) [9]

It leverages existing 4G deployments and is capable of creating 5G hotspots quickly. New 5G applications and services creation is possible with the base as LTE (4G). Option 3 is sub-divided into Option 3, Option 3A, and Option 3X.

Option 3	**Option 3A**	**Option 3X**
There is no connection from gNB to EPC. gNB is connected to EPC via eNB. So all data flows through eNB. eNB hardware upgrade is probably required because there will be a lot more additional traffic that needs to be handled.	gNB has an S1-U interface to EPC but no X2-U. The advantage of this is, traffic generated by new services can be handled by the NR and only the signaling goes through eNB.	Combination of Option 3 and 3A. Both X2 and S1 interface is available for a user plane. So, the traffic can be split based on the backhaul capacity of S1-U.

Table 2.1: Difference between Options 3/3A/3X

Option 3 is most widely talked about and is the first option to deploy 5G. The operators will be able to deploy 5G NR with their existing LTE infrastructure. This is already implemented in the year 2020. It is currently in test in major countries like the US, Canada, China, Korea, and Europe but is still not publically available.

- **Option 7 – NSA LTE assisted NR, connected to 5GC:**
  This is similar to Option 3, except, instead of using EPC it will use Next-Generation Core Network (NGCN) or 5GC, the interface will be NG instead of S1 and ng-eNB instead of eNB. The 5G is driven by capacity needs. The ng-eNB is the master node and gNB is the secondary node as shown in Fig 2.4.

- **Option 4: NSA NR assisted LTE, connected to 5GC**
  This is similar to Option 7, except, the master node is gNB and the secondary node is ng-eNB. The operator can choose this option, if they are confident with the network and technology i.e., after using Option 7 for a certain amount of time.

Initially, only Option 3 is considered in terms of 5G deployment. This is currently implemented in certain parts of the world. From Option 3, one of the migration routes would be to Option 7 or Option 5. Alternatively, we can also continue with Option 3 and also deploy Option 2 or to Option 4 and Option 2. So, it depends on the vendors on how they would like to implement this and at what stage. Issues like handover and roaming between different operators are to be considered. Samsung, one of the leading telecommunication company is suggesting Option1>Option3>Option7>Option4>Option4. The world is currently implementing Option 3. In the future, it’s better to have Option 3 and Option 1 as a backup for Option 2. [9]

2.5 5G Bands and Spectrum

There is always a debate on which frequency is the right frequency, is it low, mid-band, or high frequency. The answer to this is it depends based on what we want to achieve. Higher frequency means faster decay as shown in Fig 2.6. This means, for the same antenna height, configuration and power, the coverage area will be decreased at a higher frequency. This is independent of the technology. Similarly, low frequency means larger cell size, hence a greater number of users, which means each user will have lower throughput as there are more users in the cell.

Finally, the higher frequency gets reflected from walls and has poor penetration, while the lower frequency gets attenuated from walls but still penetrates as shown in Fig 2.6. In most cases, we want 5G to reach both indoors and outdoors.

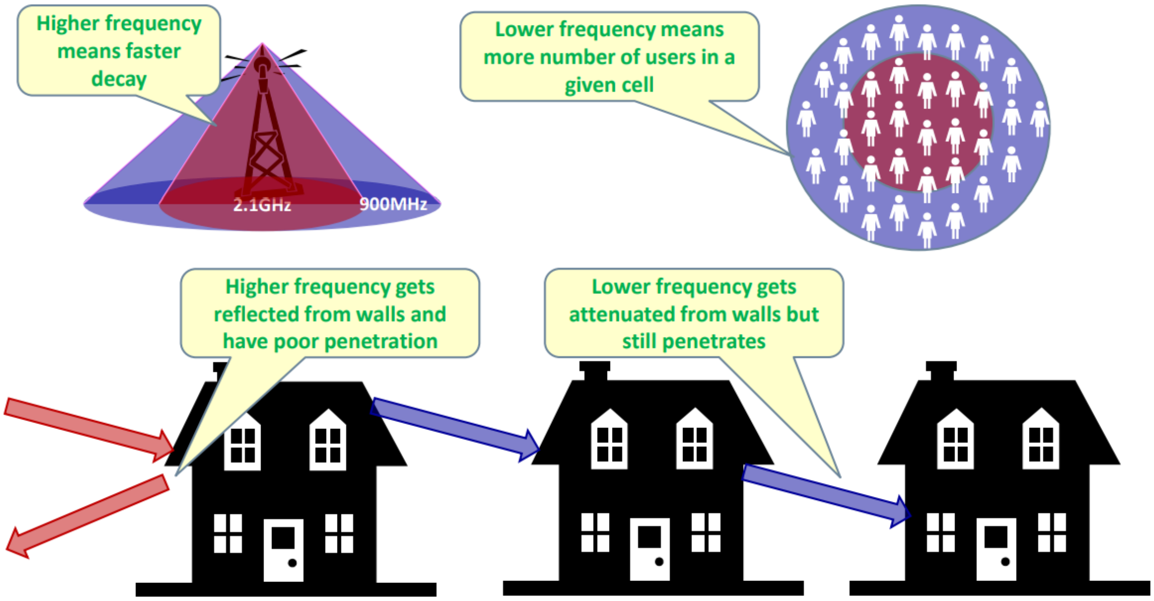

Higher frequency means faster decay
2.1GHz
900MHz
Lower frequency means more number of users in a given cell
Higher frequency gets reflected from walls and have poor penetration
Lower frequency gets attenuated from walls but still penetrates

Fig 2.6: Importance of frequency selection [10]

As a result, for 5G to be available everywhere, there needs to be three parts approach to the spectrum. The coverage layer which is below 1 GHz is required to penetrate indoors from outside. The capacity layer is between 1 GHz – 6 GHz is required to provide higher data throughput using larger bandwidths. This frequency is also compromised as it’s possible to have reasonably good indoor penetration and similarly a reasonably mid-sized cell outdoors. Finally, we have a high throughput layer that uses the frequency between 6 GHz – 100 GHz. As this is quite high frequency, the cell size is relatively smaller as shown in Fig 2.7. Beamforming can help increase the cell size in this case, but that adds to the cost and power consumption. Contrary to this, all high throughput layer solutions include massive MIMO and beamforming.

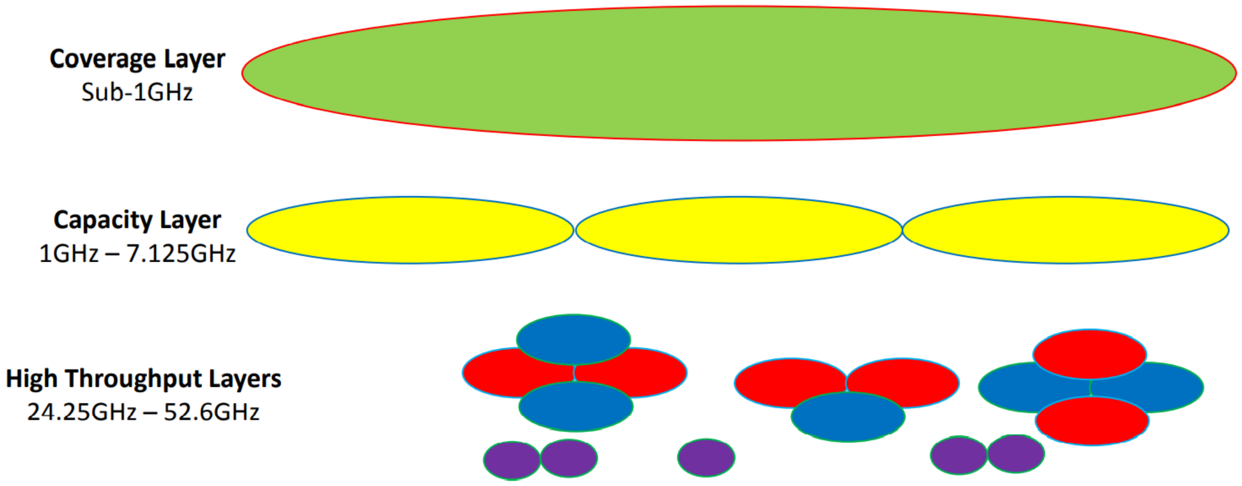


Fig 2.7: Multiple Layers for multiple needs [10]

In theory, the frequency for 5G is from 0 – 5 GHz, but in practice, 3GPP has defined two ranges of frequency called Frequency Range 1 (FR1) and FR2. FR1 covers frequency from 450 MHz – 6 GHz. For the new Wi-Fi standards, this upper range has been changed to 7.125 GHz. FR2 covers frequency from 24.25 GHz – 52.6 GHz. FR2 bands are known as mmWave bands. This is shown in Fig 2.9.

In summary, the most popular coverage layer frequency is 700 MHz, 3.5 GHz is the most popular capacity layer band and 26 – 28 GHz is the most popular high throughput layer frequency as shown in Fig 2.8.

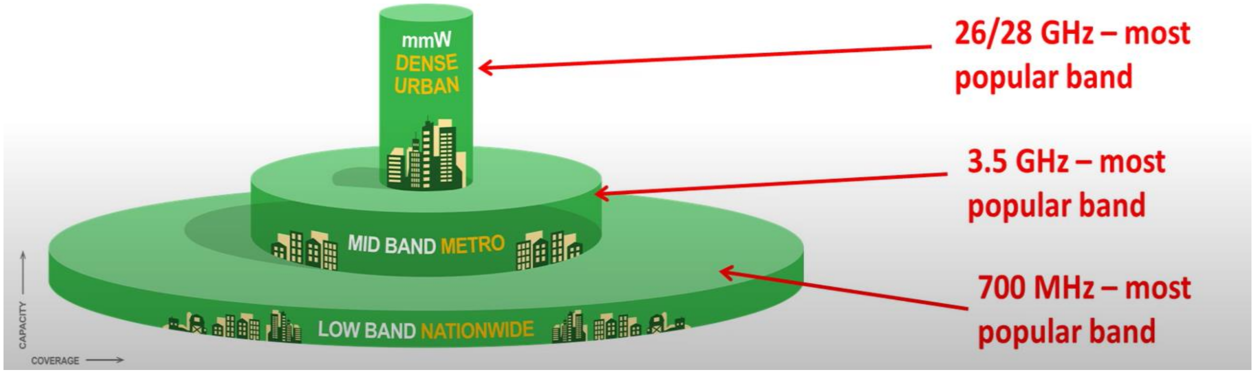

mmW
DENSE
URBAN
MID BAND METRO
LOW BAND NATIONWIDE
CAPACITY
COVERAGE
26/28 GHz – most popular band
3.5 GHz – most popular band
700 MHz – most popular band

Fig 2.8: Most popular 5G Frequency Bands [10]

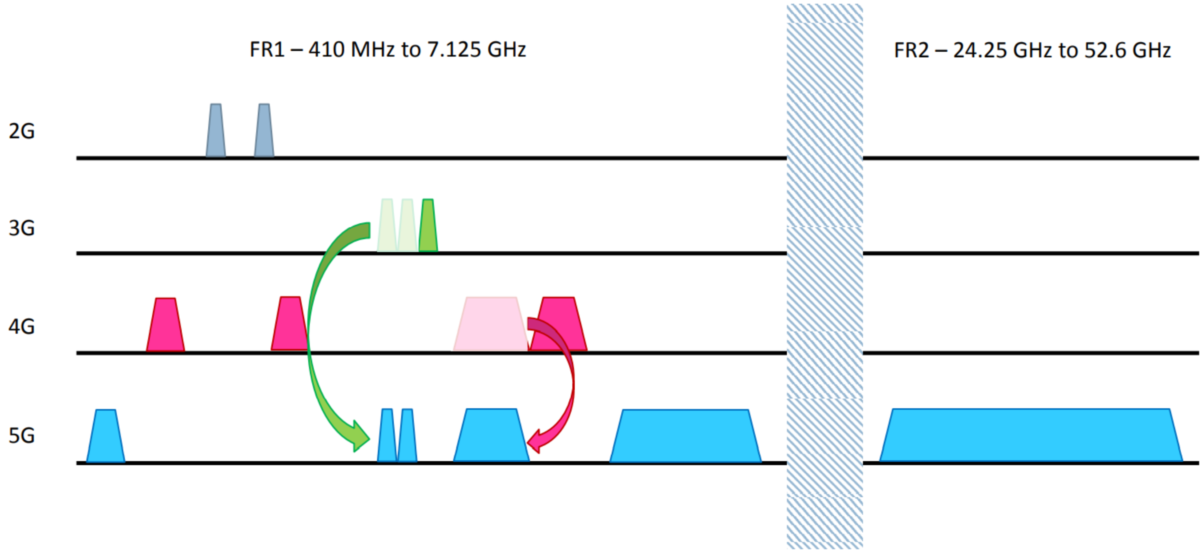


Fig 2.9: Typical Operator Frequency [10]

Looking at all the new bands that were just defined for 5G, the bandwidths available for the bands are large, but that does not mean each country will allocate large chunks of spectrum to the operators.

NR FR1 Band	***Uplink (UL) Operating Band BS Receive / UE Transmit F_{UL_low} – F_{UL_high}***	***Downlink (DL) Operating Band BS Transmit / UE Receive F_{DL_low} – F_{DL_high}***	***Bandwidth***	***Duplex Mode***
n41	2496 MHz – 2690 MHz	2496 MHz – 2690 MHz	194 MHz	TDD
n66	1710 MHz – 1780 MHz	2110 MHz – 2200 MHz	70/90 MHz	FDD
n71	663 MHz – 698 MHz	617 MHz – 652 MHz	35 MHz	FDD
n77	3300 MHz – 4200 MHz	3300 MHz – 4200 MHz	900 MHz	TDD
n78	3300 MHz – 3800 MHz	3300 MHz – 3800 MHz	500 MHz	TDD
n79	4400 MHz – 5000 MHz	4400 MHz – 5000 MHz	600 MHz	TDD

Table 2.2: NR operating bands in FR1 [11]

NR FR2 Band	***Uplink (UL) Operating Band BS Receive / UE Transmit F_{UL_low} – F_{UL_high}***	***Downlink (DL) Operating Band BS Transmit / UE Receive F_{DL_low} – F_{DL_high}***	***Bandwidth***	***Duplex Mode***
n257	26500 MHz – 29500 MHz	26500 MHz – 29500 MHz	3000 MHz	TDD
n258	24250 MHz – 27500 MHz	24250 MHz – 27500 MHz	3250 MHz	TDD
n260	37000 MHz – 40000 MHz	37000 MHz – 40000 MHz	3000 MHz	TDD

Table 2.3: NR operating bands in FR2 [11]

In several countries, 5G bands have been allocated to different operators through auctions to facilitate deployments, while in some countries the process is yet to begin and is in the research phase. The table below lists the major bands in different countries and the status of their deployment.

Country	Major NR operating bands	Status
India	n77, n78	Yet to be auctioned
Canada	n66, n71	Deployed NSA, unused
United States	n41, n261, n5, n260, n71	Deployed SA and NSA
Europe	n78, n258	Deployed NSA
China	n78, n41	Deployed NSA, unused
Japan	n78, n79, n77, n257	Deployed NSA, unused
Korea	n78, n257	Deployed NSA
Singapore	n78, n257	Unused spectrum

Table 2.4: 5G major bands and their status in different countries [12]

Chapter 3 - 5G Enabling Technologies

3.1 Introduction

In the previous chapter, we have discussed the typical used cases and the requirements of 5G and to meet those, a variety of technological candidates have been considered and extensively debated to satisfy the stringent demands discussed previously. The mapping of the 5G specifications with the relevant possible technology enablers is given in Table 3.1. In both the radio access network (RAN), the core network, and the end-to-end infrastructure, technological growth will take place. Such innovations will be listed in detail below.

Requirements	Technology Candidates
High Data Rate	mmWave, Massive MIMO, Small Cell
Ultra-Low Latency	Mobile Edge/Fog Computing, D2D
Massive Connectivity	Massive MIMO, D2D, M2M, Small Cell
Reliability and High Availability	Cloud-RAN, SDN, NFV, MANO, Cloud Computing
Flexibility and Programmability	Cloud-RAN, SDN, NFV, Network Slicing, MANO
Energy and Cost Efficiency	Cloud-RAN, SDN, NFV, Network Slicing, MANO
Spectrum Efficiency	Massive MIMO, Small Cell, D2D

Table 3.1: 5G key requirements and corresponding technology candidates [1]

3.2 5G Radio Access Network

The enabling technologies for 5G RAN include mmWave communication, massive Multiple Input Multiple Output (MIMO), ultra-dense small cell, Machine-to-Machine (M2M), and Device-to-Device (D2D) communications, and cloud-RAN. These technologies will be described in the following.

3.2.1 mmWave Communication

One of the main features of the 5G scheme, as previously stated, is to provide greater data rate capacity, for example, up to tens of Gbps at the highest data rate. More spectrum availability is needed to achieve those goals. Present wireless networks, however, usually run in a frequency range ranging from hundreds of MHz to below 3 GHz (e.g. 700 MHz) (e.g. 2.6 GHz). These applications of the spectrum are not necessary for 5G. Exploiting the very large spectrum bands that have not yet been occupied (e.g. >10 GHz) is one of the most powerful options for extending the bandwidth range. In particular, several suggested frequency bands above 10 GHz for 5G were accepted during the meeting at the WRC-15 conference hosted by ITU to be studied in advance of the next WRC conference in 2019, such as 24.25-27.5 GHz, 50.4-52.6 GHz, 81-86 GHz, etc. The best technology nominee is millimeter-wave communication (mmWave) in this context.

In 1897, the mmWave analysis was first carried out by Jagadis Chandra Bose, relating to the use of frequencies in the 30 to 300 GHz range, with corresponding wavelengths between 10 mm and 1 mm, as seen in Fig 3.1. The mmWave contact has been widely used for indoor environments or backhaul connections due to certain considerations, such as high propagation loss. However, by implementing many recent developments in propagation modeling or channel modeling, many academic projects have shown the viability of mmWave technologies for 5G cell networks to generate a greater amount of bandwidth. There is also a range of problems and transparent concerns that need to be tackled in the future, such as interference and fragmentation, in addition to the advantages of enabling wider capacity, the faster data rate that makes mmWave a promising 5G technology. [1]

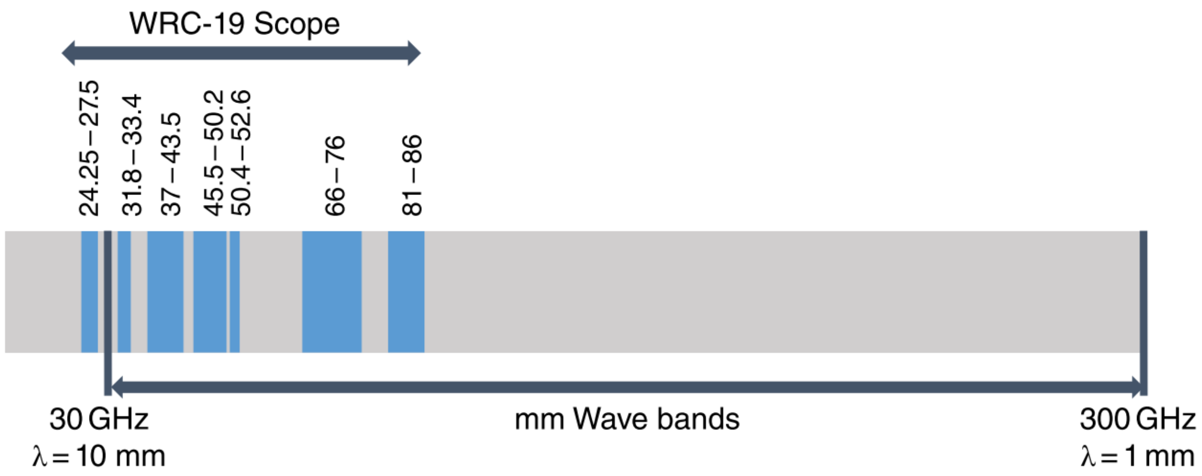


Fig 3.1: Millimeter-wave bands and potential 5G bands to be studied ahead of WRC-19 [1]

3.2.2 Massive MIMO and Beamforming

a) Massive MIMO

One of the most popular options is to densify the number of installed antennas, which applies to a technological approach called massive MIMO, to meet the 5G criteria in terms of network coverage and capability enhancement. MIMO is essentially a wireless communications antenna technology in which multiple antennas are used to send and receive data. In reality, in current 4G networks, the MIMO term has been widely used, referring to multi-user MIMO (MU-MIMO) communication, where a multiple-antenna base station is operated concurrently by many users; while massive MIMO is characterized as a multi-user MIMO system, where the number of antennas of the base station and the number of users is high. A function like providing more antennas at the base station promises to improve the power and density of the network. More specifically large MIMOs are said to dramatically increase the quality of bandwidth and electricity. These explanations make huge MIMO a relevant technology for 5G. The principle of huge MIMO is illustrated in Fig 3.2(a). Many research issues need to be answered, aside from the advantages of huge MIMO, such as plot pollution reduction, channel estimation, implementation-aware algorithmic architecture, etc. [1]

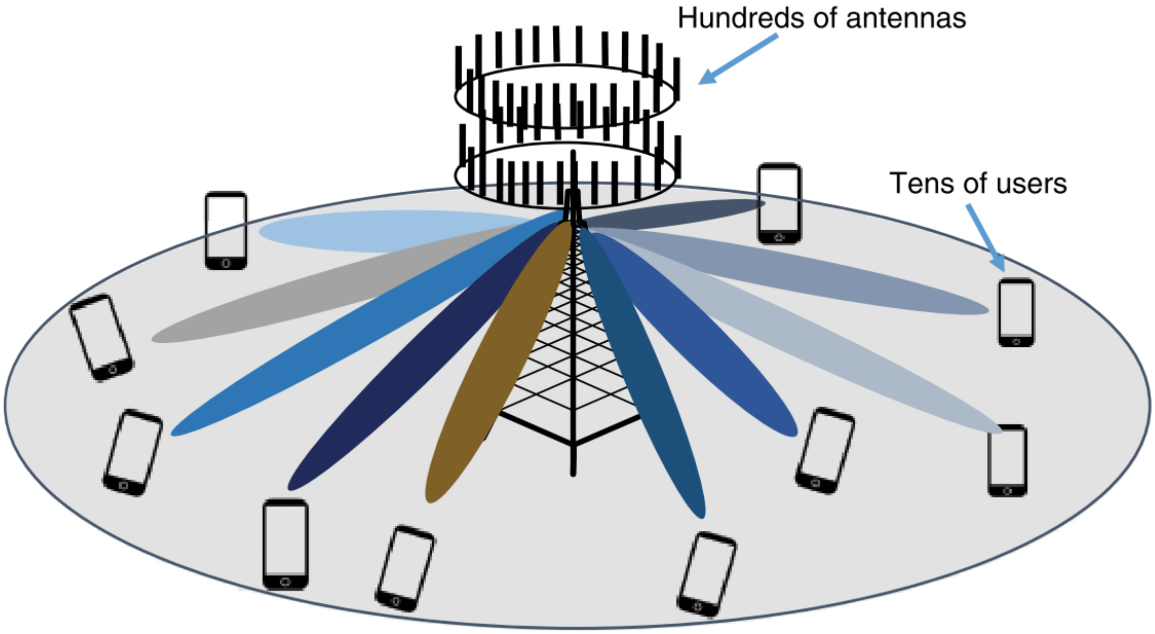


Fig 3.2(a): An illustration of massive MIMO concept [1]

b) Beamforming

Beamforming is another revolutionary technique crucial in 5G. Without regard to the location of intended users or computers, traditional base stations have transmitted signals in several directions. Signal processing algorithms can be used to determine the most appropriate propagation route to each user by the use of multiple-input multiple-output (MIMO) arrays with thousands of small antennas combined in a single formation, thus individual packets can be sent in multiple directions and then choreographed to reach the end-user in a predetermined sequence.

Free space propagation loss, relative to the smaller antenna size, and diffraction loss, inherent in higher frequencies and lack of wall penetration, are considerably greater with 5G data transmission occupying the millimeter-wave. The smaller antenna size, on the other hand, still requires the same physical space to be occupied by even larger arrays. With any of these smaller antennas likely reassigning beam path many times every millisecond, massive beamforming becomes more realistic to support the 5G bandwidth challenges. With a greater antenna density in the same physical area, with huge MIMO, narrower beams can be accomplished, offering a way of achieving high throughput with more reliable consumer detection. [13]

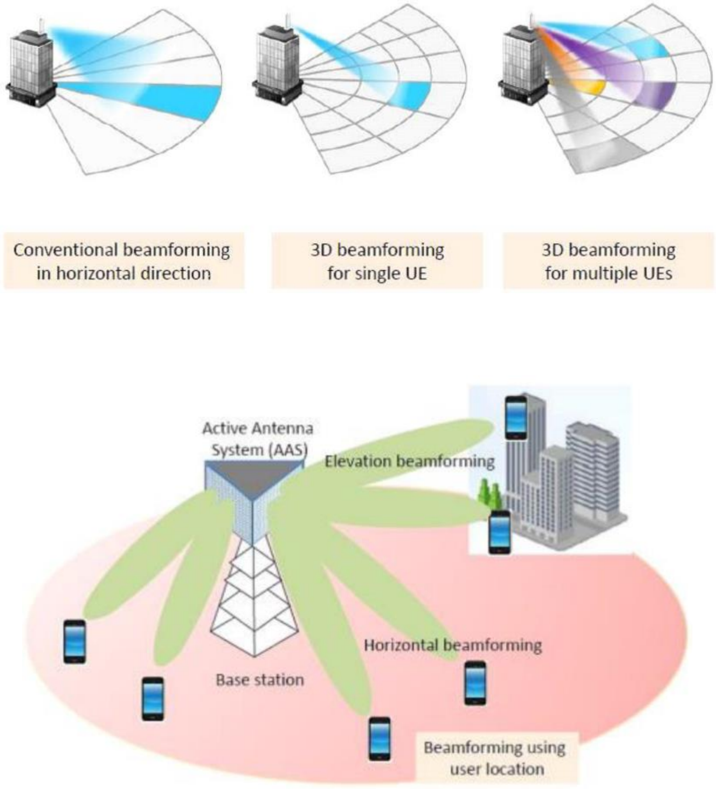


Fig 3.2(b): Beamforming [14]

3.2.3 Ultra-Dense Small Cells

The densification of the number of cellular nodes, which have a wider coverage area than the macro-cell base stations used in the legacy 3G and 4G networks, is another means of increasing network density and enhancing throughput. The scientific approach behind this theory is called the technology of small cells. The' small cells' is an umbrella word for operator-controlled, low-powered radio access nodes with a coverage range from ten to several hundred meters, including those operating in licensed spectrum and unlicensed WiFi carrier-grade, as specified by the Small Cell Forum. In Fig 3.3, an example of small cell deployment is seen.

The size of the cell is decreased for small cells, which ensures they put the network far closer to the customer, thus better covering heavy demand areas such as indoor and hotspot areas. Furthermore, the higher number of low-powered transmitting points on the small cell network makes the frequency capital available to be best used, thus improving the spectral quality.

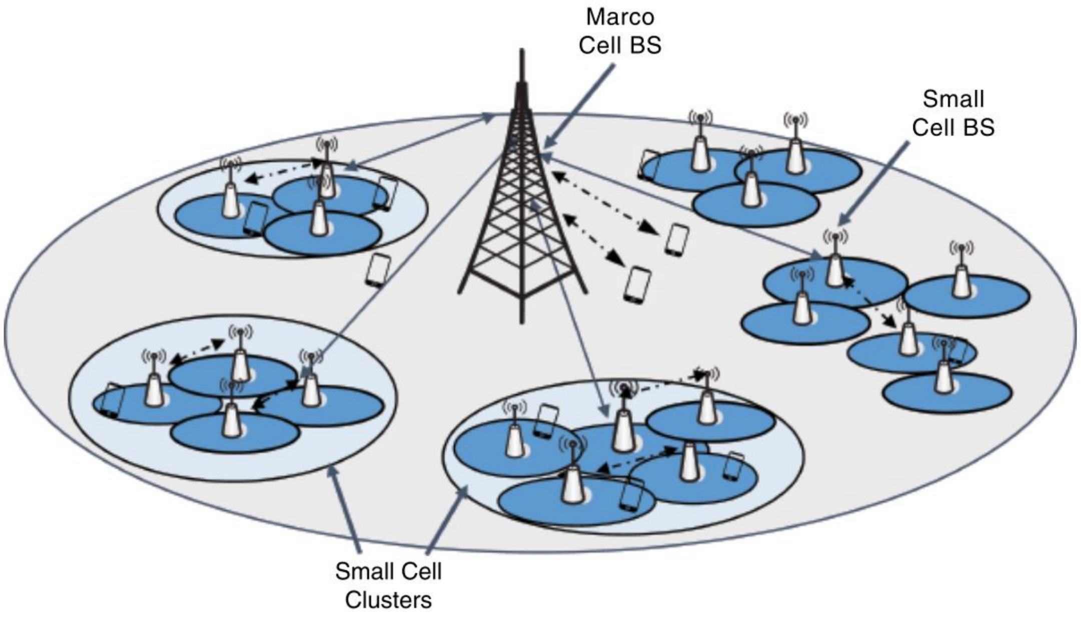


Fig 3.3: An illustration of small cells deployment [1]

Also, the 5G infrastructure would be designed in a heterogeneous manner, where macro and small cells are co-located and may be linked by wireless backhaul connections to each other, thereby offering increased amounts of network bandwidth via traffic offloading. Nevertheless, in terms of intervention and mobility control, the heterogeneity of small cells in the network can face problems, thus impacting device efficiency as a whole. [1]

3.2.4 M2M and D2D Communications

a) M2M Communication

Two-thirds of the 5G use case categories will be related to IoT and Machine Type Communication (MTC), including massive and critical communications, as previously stated. Therefore, although the idea of M2M or MTC communication was introduced by 3GPP some time ago in 4G LTE systems, it is still regarded as one of the key enablers for 5G. Fundamentally, M2M communication refers to the automated communication of data between devices and the underlying infrastructure of data transport. Data may be communicated between an MTC device and a server, or between two MTC devices directly. As shown in Fig 3.4(a), there are several M2M communication-enabled services and applications, such as monitoring and metering, home and industry automation, health care, and automotive. [1]

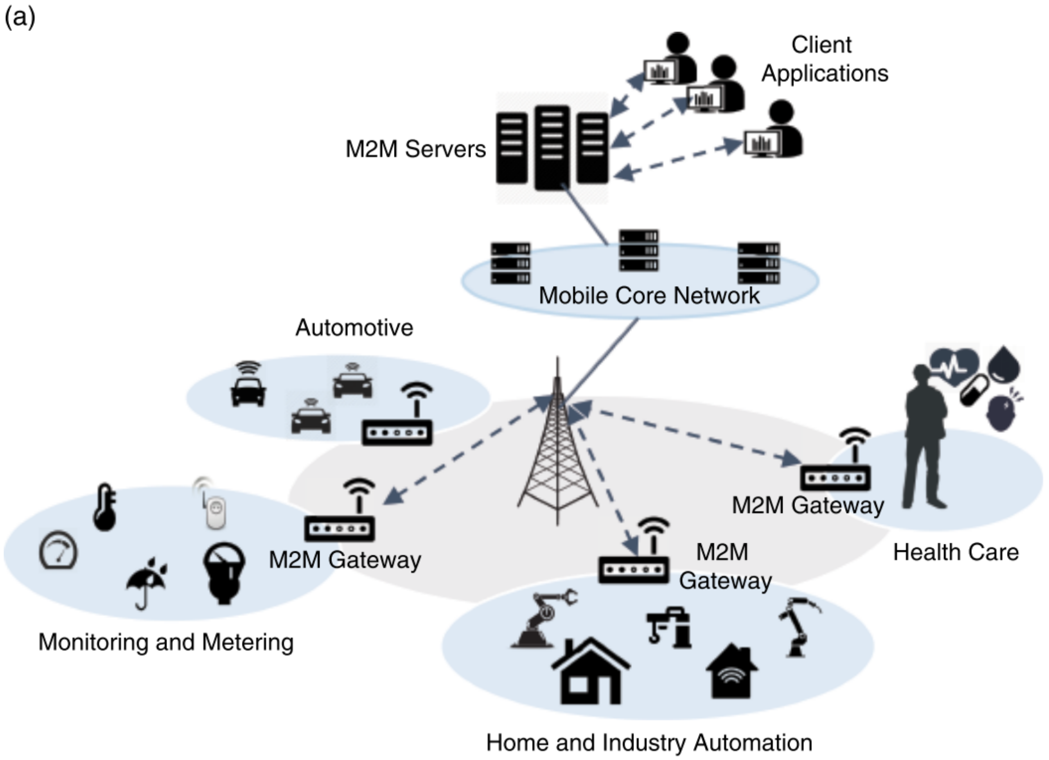


Fig 3.4(a): M2M communication and use case examples [1]

b) D2D Communication

D2D communication refers to direct communication between two mobile users/devices without a network system being traversed. It was defined in LTE Release-12 by 3GPP. D2D communication can help increase spectrum performance, consumer data rate gain, and reduce latency as well as energy consumption by leveraging direct communication between devices, thereby being considered as one of the main components of the 5G system. In general, on an approved cellular spectrum (e.g. LTE) and out-of-band D2D on unlicensed spectrum, the activity of D2D communication can be in-band D2D (e.g. WiFi). D2D use cases and implementation examples, such as proximity-based utilities, gaming, public safety, vehicular communications, and offloading, are available, as seen in Fig 3.4(b). There is however a range of open problems that need to be resolved in the future, such as interference control, services for resource management and the exploration of devices, authentication, and privacy. [1]

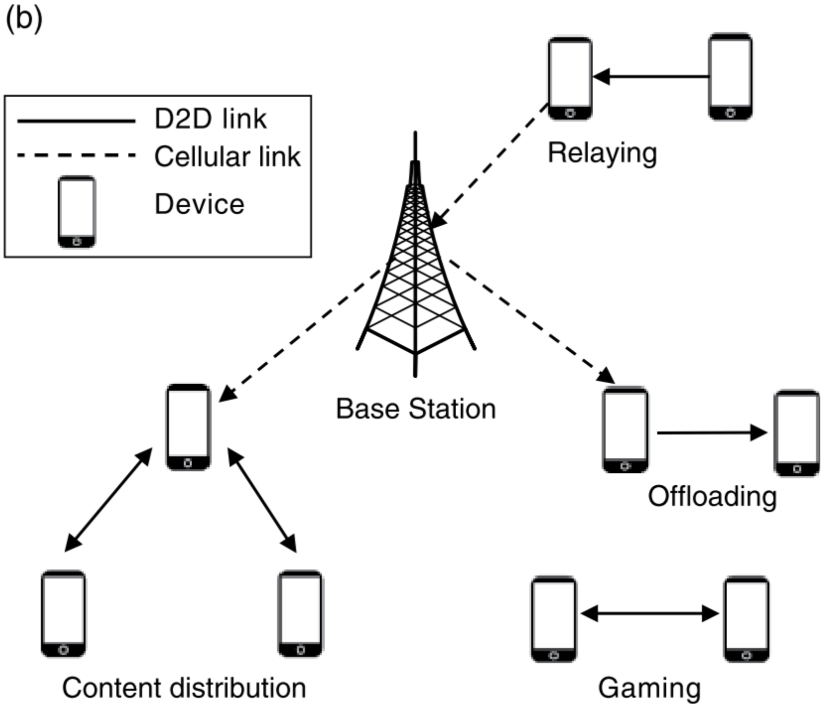


Fig 3.4(b): D2D communications and use case examples [1]

3.2.5 Cloud-based Radio Access Network (C-RAN)

Cloud-based Radio Access Network (Cloud-RAN) is an ideal solution to design the radio access part of 5G networks, since it enables energy efficiency, cost savings on baseband resources, as well as improvements in network capacity, increased throughput, etc. In essence, the Cloud-RAN is the decoupling of the Remote Radio Head (RRH) from a base station's Baseband Unit (BU) and the BU implementation in a centralized cloud computing system. By using high-speed fiber or microwave-link front-haul networks, RRHs are linked to a BBU pool.

There are multiple ways for breaking the base station feature, which applies to the RAN-as-a-service (RANaaS). These two concepts are illustrated in Fig 3.5. By making it inexpensive, scalable, and efficient, this simpler base station architecture is paving the way for dense 5G rollout.

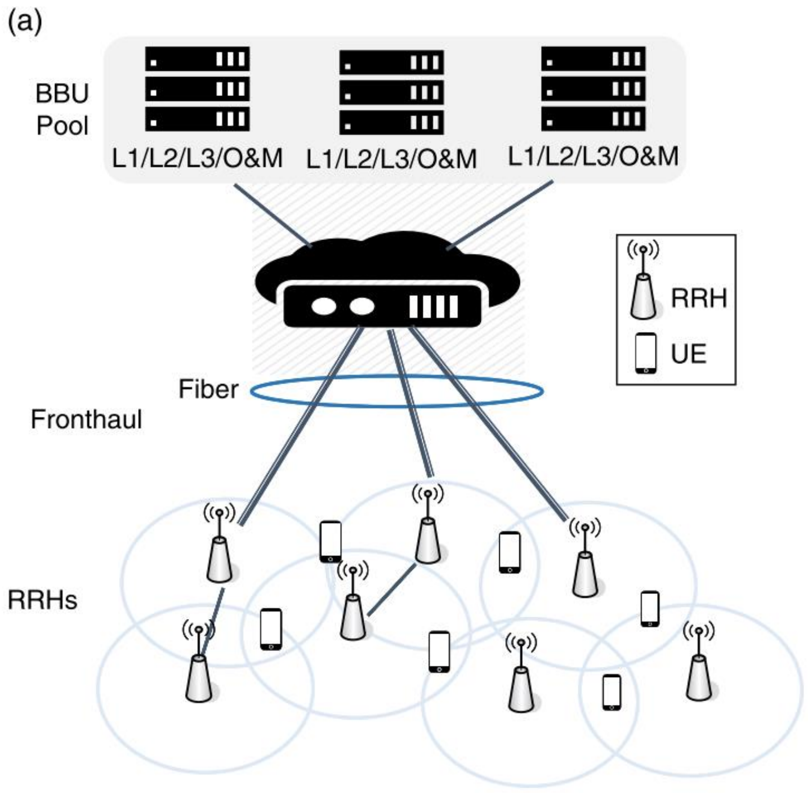


Fig 3.5: Cloud-RAN concept [1]

3.3 5G Mobile Core Network

The main technology for developing the central portion of 5G networks are SDN, NFV, and cloud computing. These technologies are described as follows:

3.3.1 Software Defined Networking (SDN)

SDN is generally regarded as the best infrastructure candidate for the development of 5G networks in terms of network flexibility and programmability. In the campus and data center network fields, the notion of SDN was first suggested. It features the isolation of the data plane from the control plane and, enabled network management by abstracting network control functionality as shown in Fig 3.6. SDN would allow a more agile and versatile core network infrastructure when implemented by 5G. Also, SDN's programmability and transparency capabilities can help mobile operators shorten the lifecycle of the business launch of their innovative offerings and creativity. The network infrastructure can be built on request and the basis of service specifications (network-as-a-service) by separating the control and data planes, thus increasing resource quality. It should be remembered that the SDN definition can also be used in the RAN domain, where radio services for base stations can be managed and planned by the SDN controller, thereby improving spectrum utilization as well as mobility management.

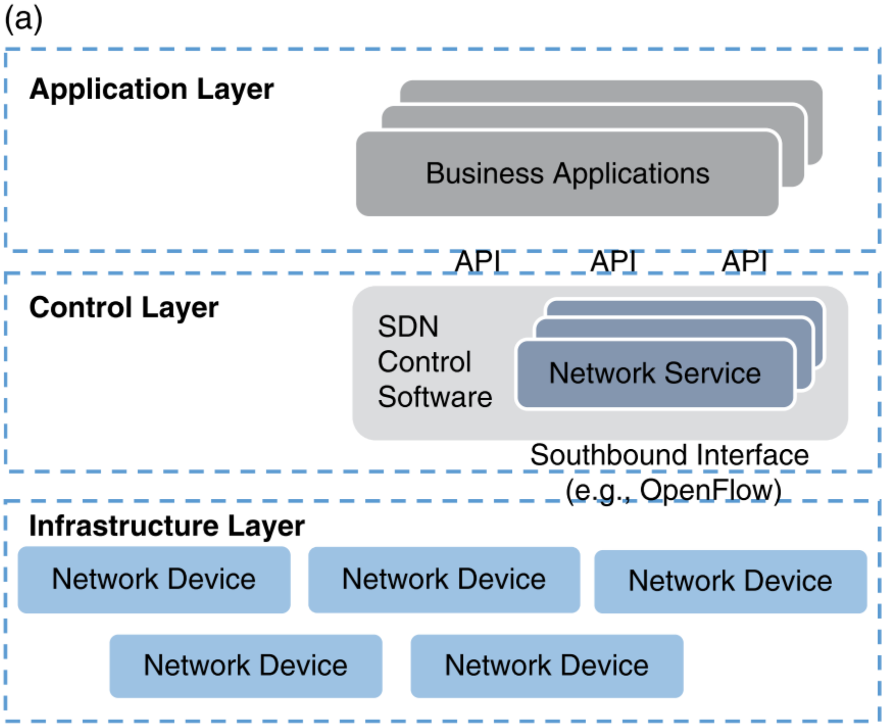


Fig 3.6: SDN architecture [1]

3.3.2 Network Function Virtualization (NFV)

The 5G infrastructure is not all about high data rate, low latency, and versatility, as mentioned earlier. It is all about cost competitiveness, which would have a bearing on mobile operators' sales. 5G network operators would expect the expense of rollout to be as low as possible, which refers to capital spending or CAPEX, and the cost of service and management, which refers to operating expenditure or OPEX. The basis for these capabilities is NFV, which is identified as the backbone of 5G core network solutions. Fig 3.7 demonstrates NFV's reference architectural structure. Essentially, NFV refers to the migration to virtual appliances operating in the cloud environment or on general-purpose commodity servers of network functions which are typically deployed on dedicated expensive hardware platforms. It is easier for mobile operators to dynamically scale capacity (computing, storage, and networking) according to shifts in traffic demands by running the network feature as software, and to speed up the time to market new networks. Also, the combination of SDN and NFV has encouraged the development of new networking paradigms, such as network slicing.

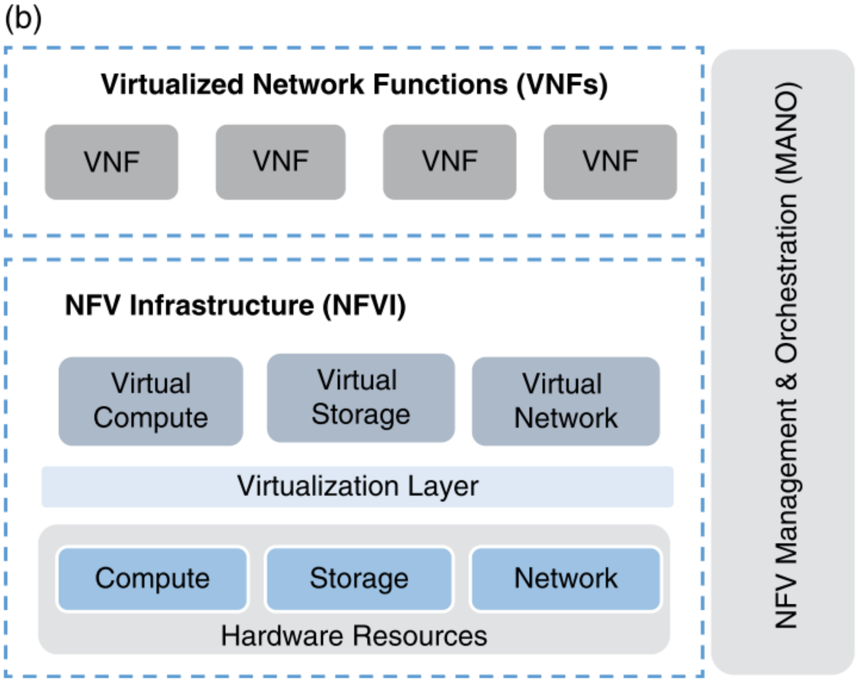


Fig 3.7: NFV architecture [1]

3.3.3 Cloud Computing

Cloud infrastructure has been seen as an optimal approach for re-designing the present RAN architecture, as mentioned in the previous section. With its intended advantages, cloud networking has been one of the main enablers for the design of 5G core networks, such as on-demand and distributed provisioning of data and infrastructure over the Internet. In this case, as virtual machines or containers are managed by the cloud manager, 5G core network functions can be realized.

The ability to deliver services in a cloud computing multi-tenant model helps telecom providers to adopt the Mobile Virtual Network Operators (MVNO) concept even more effectively than in the past. Also, a pay-as-you-use business model and the opportunity to transfer and consolidate the cloud storage services offered. Mobile operators can help to reduce their capital and optimize operating costs. Cloud computing was born to virtualize commodity IT hardware similar to NFV. NFV refers to the inspiration for virtualizing network functions in cloud computing. Several cloud technologies such as OpenStack or VMware are serving as the resource backend for virtual network functions in the recent creation of NFV. [1]

3.4 5G End-to-End System

Network slicing and control and orchestration are the main technical enablers for building a 5G end-to-end infrastructure.

3.4.1 Network Slicing

The 4G infrastructure of today has been largely optimized to serve human-to-human connectivity where cell phones are the key players. The 5G infrastructure, however, is expected to serve varied networks and apps with varying specifications and specifications in the future, where IoT devices will become prevalent. In terms of latency, data rate, accessibility, reliability, stability, etc, certain IoT-related services will need various types of features and network capabilities. To ensure these conditions and increase the efficiency of the network and the usage of energy, each type of service can also be delivered as an end-to-end, segregated, and infrastructural system in which to function. The network slicing idea would become the cornerstone of all capabilities in this context.

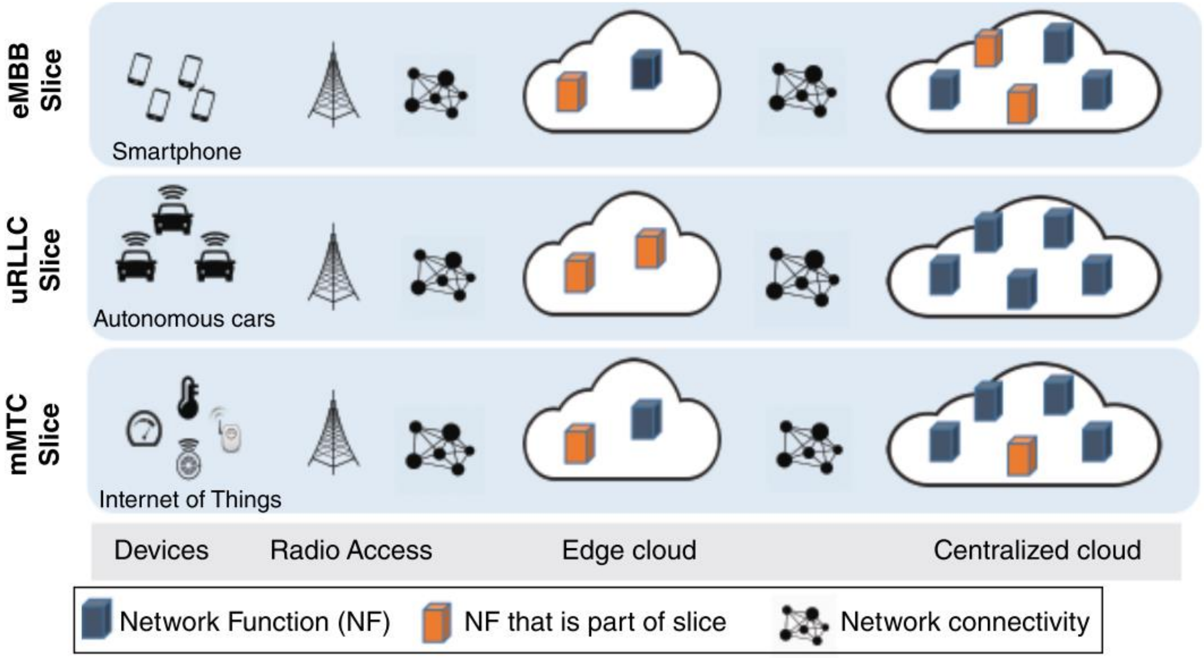


Fig 3.8: An example of Network Slicing [1]

While network slicing has been commonly recognized by many network operators and vendors as the core feature of 5G, it has not yet been standardized, so there are variations to describe what slicing is. The basic principle of Network Softwarization is slicing, according to ITU-T. Logically isolated network partitions (LINP) are enabled, with a slice assumed to be a programmable resource unit such as network, computing, and storage. A network slice, namely a "5G slice", consists of a series of 5G network functions and specific radio access infrastructure settings, as defined by NGMN in its white paper, that are combined for the specific use of a case or business model. An overview of the network slicing definition is shown in Fig 3.8, with three separate slices relating to the three major 5G usage case groups. The application of the network slicing principle is on an end-to-end basis for this reason. Multiple end-to-end slices would comprise the 5G infrastructure where dedicated resources and quality of service (QoS) are ensured.

Network slicing, however also poses several problems and gaps that need to be met in future research, such as slice specification, slice lifecycle management, slice control resilience, resource distribution, and optimization within a slice and between slices, firmly guaranteeing protection within a slice and between slices, end-to-end QoS management, convergence with other technologies (e.g. information-centric networking (ICN), D2D), etc. [1]

3.4.2 Management and Orchestration

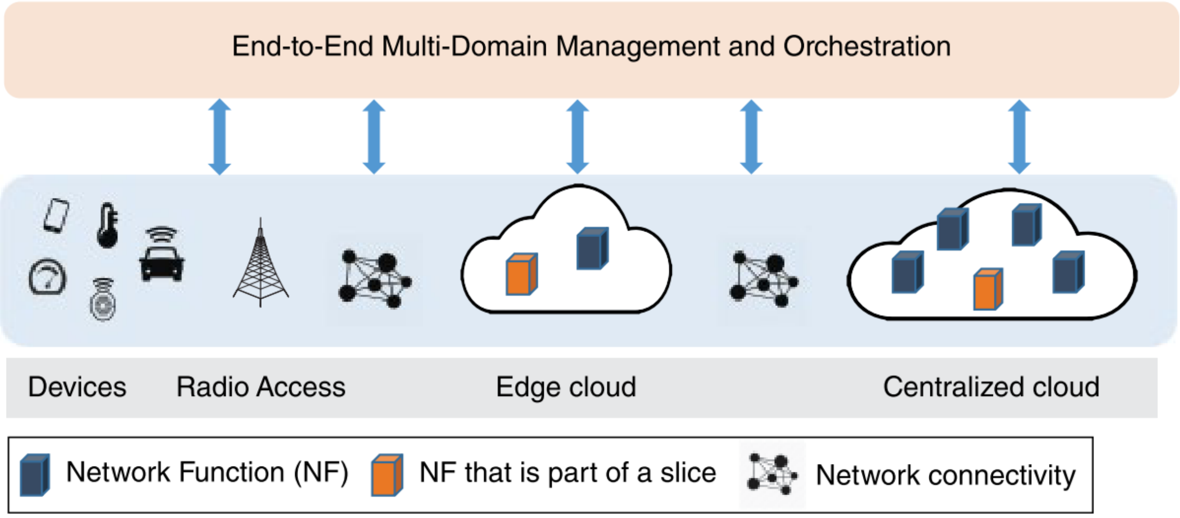


Fig 3.9: The illustration of end-to-end multi-domain management and orchestration [1]

Because of the variety of use cases, facilities, and the number of network slices generated with various resource specifications, the management and orchestration (MANO) of the network becomes more and more important as the 5G mobile networking period arrives in the next few years. In terms of fault management, setup, accounting, performance, and security, MANO's job would be to handle the entire network infrastructure. More specifically, in a complex, automatic, and effective way, MANO would be responsible for lifecycle management and provision of network services for the end-to-end connectivity of network slices.

As outlined in Fig 3.9, multi-domain, multi-operator, and multi-technology would be the role of end-to-end management and orchestration, spanning from the infrastructure layer to the application (service) layer and spanning from the RAN, to the network center. [1]

Chapter 4 – Mobile Networks Security Landscape

4.1 Introduction

The mobile network began only as a voice communication platform, yet to become the hub of our everyday life, economy, and governance, it has overshadowed every other medium in history. The future of communities and major economies, such as Europe, is highly dependent on mobile networks, where the direct financial and economic influence of the ICT (Information and Communications Technology) market is roughly 5% of GDP or hundreds of billions of euros. By the time 5G is implemented in full scale, the mobile population will reach 6 billion. This means further networking and the generation and exchange of information through mobile networks.

This sector will quickly become a prime target for anti-state and criminal actors who want to exploit this forum to undermine its development and use it to launch attacks against a larger mobile network consumer base as a result of the telecommunications boom and its vital position overall public and economic health. As occurred recently in the case of the Yahoo mega data breach that impacted 1 billion users in a single assault or the LinkedIn 2016 breach that impacted 117 million, resulting in a leak of confidential personal and technical data, the effect of breaches in this modern age of the wired world can be large and impactful.

In parallel to the evolution of the telecommunication industry, mobile network security has gradually developed. In this chapter, we will cover the overall mobile network security vulnerability environment as it has developed with the different generations of mobile networks. The history of mobile security risks, relative defensive measures, and impacts on mobile networks will also be addressed. [1]

4.2 Mobile Networks Security Landscape

The protection environment of the mobile network (as seen in Fig 4.1) can be seen in the light of the evolution of the different generations of mobile networks. In terms of infrastructure design, technological capabilities, networks delivered, and related threat vectors, there is a strong connection between the advancement of mobile network architectures and the relative evolution of security threats.

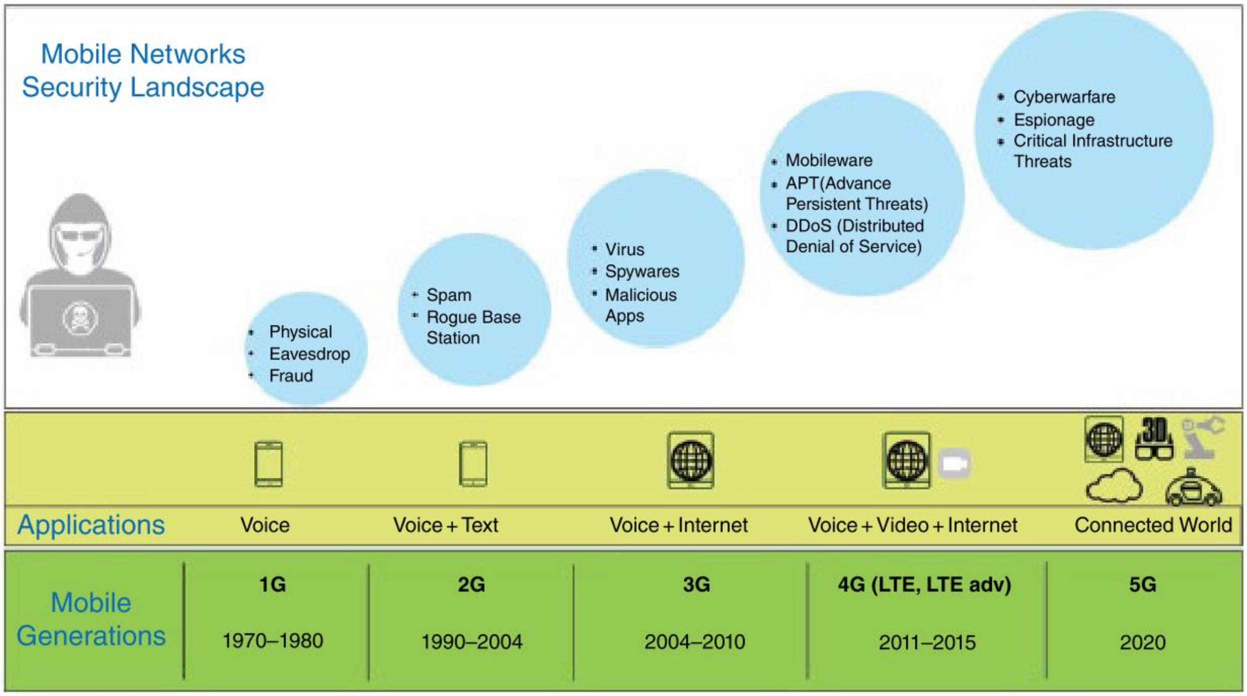


Fig 4.1: Mobile Network Security Landscape [1]

4.2.1 Security Threats and Protection for 1G

Immediately after the launch of the first generation (also called 1G) of mobile technologies, mobile networks began to experience severe risks and problems and continued to evolve as a diverse and demanding threat environment. 1G was launched specifically to give voice users versatility. Consumers began to witness the right to engage and make calls whilst they were mobile. Criminals have found an incentive and strategies to conduct telephone theft and impersonate legitimate subscribers and hack their phones and make free calls. By making and distributing illicit cloned phones, mobile phone cloning has become an enterprise. For various malicious purposes, some hackers have found new ways to hijack and eavesdrop on calls while they are being made and listen to private conversations.

As seen in Fig 4.2, to impersonate a legal network subscriber, a mobile phone cloning intruder requires some hardware and software resources. The consumer calls hitting the mobile tower are sniffed and intercepted using a radio receiver. The receiver steals the ESN (Electronic Serial Number) and mobile number detail. The intruder would burn the stolen ESN and mobile identification details using a PC with tools to replicate a replica of the original cell phone. Attackers will use off-the-shelf software such as Oki 900 to mimic and eavesdrop on AMPS communications. Between the victim and the cloned phone, it was difficult to discern.

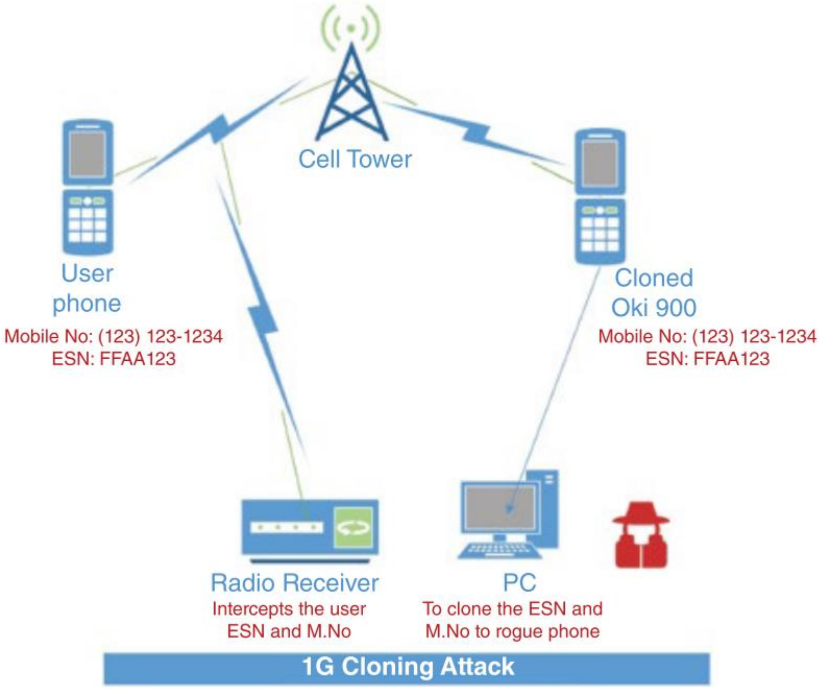

Cell Tower
User phone
Mobile No: (123) 123-1234
ESN: FFAA123
Cloned Oki 900
Mobile No: (123) 123-1234
ESN: FFAA123
Radio Receiver
Intercepts the user ESN and M.No
PC
To clone the ESN and M.No to rogue phone
1G Cloning Attack

Fig 4.2: Cell phone cloning attack in 1G network [1]

Such attacks have prompted carriers to interrupt numerous legal subscriber links and implement new authentication methods to eliminate nefariously cloned mobile devices from their networks, such as a special pin code for each user. [1]

4.2.2 Security Threats and Protection for 2G

The era of call spamming in the smartphone world arose with 2G. Spamming has been used to insert misleading information or give unwanted marketing jargon to smartphone users as a pervasive attack. Inboxes of messages were occupied by spam messages directed at a particular audience or the broader population. For their vicious ends, call spamming was used by fraudsters. By providing bogus network authentication, Rogue base stations (also called IMSI Catchers) were conceived to capture cell traffic.

When developing 2G standards, user authentication remained a key focus, as the objective was to reduce call charge fraud, channel hijack, and mobile attack surface cloning. The Subscriber Identity Module (SIM) was first used to assign unique mobile phone identification and could be securely stored inside the cell phone as a computer chip.

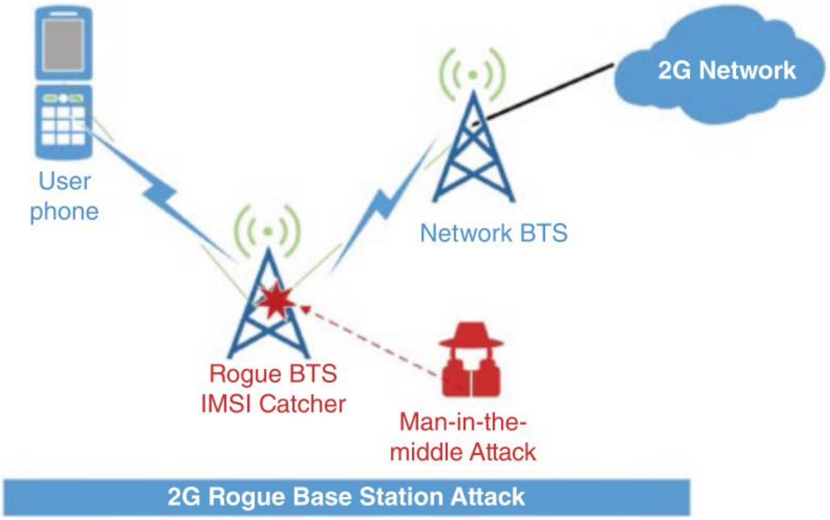


Fig 4.3: IMSI Catcher (Rogue base station) attack on 2G [1]

A new threat was introduced in 2G, masquerading as a carrier base station with a rogue base station (also called IMSI catcher) to perform a man-in-the-middle attack (MitM), as user identity protection began to be taken seriously.

Fig4.3 demonstrates the IMSI (International Mobile Subscriber Identity) catcher threat. The lack of 2-way authentication between the cell phone and the mobile network was exploited by this attack. Attackers would install a rogue base station that could impersonate a valid BTS (Base Transceiver Station) and allow the user to transmit critical identity data such as IMSI over an unsecured channel. In addition to stealing the IMSI data, a rogue base station attack was used to sniff the voice traffic over GSM, and to tap the sensitive user data transmitted over GPRS and EDGE. An attacker could easily copy the web traffic of the user and extract the information via analyzers, such as the password. Interestingly, on the market, the resources needed to conduct a GSM MitM attack are readily available, i.e. a standard BTS (Base Transceiver Station) and OSMOCOMBB open source program.

2G has implemented limited-scale encryption to secure traffic between user devices and the base station. While it was unable to defend against cryptanalytics threats, it was able to provide some simple signaling and user data encryption protection. Other recent security risks have been spam traffic in the form of smartphone short messages for fraudulent ads and marketing. [1]

4.2.3 Security Threats and Protection for 3G

As user devices become smart and resourceful, the key services offered by mobile service providers in new 3G networks have become data applications and the internet. Somewhere between 500 and 700 kbps was the average 3G data connection speed, which was sufficient to provide internet-facing applications with connectivity. The user phones, computer system, and its operating system were targeted by threat vector in 3G. Mobile OS vulnerabilities have been exploited because malicious code has been injected into mobile applications to gain unauthorized access to sensitive personal information, such as contacts, user passwords, and location data. As the data velocity increased, so did the type of malware and spyware infections.

Third Generation Mobile Network (3G) decided to base its security on the CIA (Confidentiality, Integrity, and Availability) framework. As a result, for two-way authentication between the user equipment and the network, and also to protect against attacks, such as rogue base stations, the AKA (Authentication and Key Agreement) protocol was adopted. To maintain the authenticity and integrity of signaling messages sent over the radio, AKA used strong 128-bit auth keys and hash functions.

Although two-way authentication decreased the likelihood of such attacks, in the UMTS environment, attacks such as MITM would still be possible using advanced tools such as mobile jammers and OSMOCOMBB.

As 3G was designed to provide mobile users with the next generation of data services and Internet connectivity, the system was introduced with new challenges and vulnerabilities.

Mobile networks switched to a packet switching model, IP-based RAN (Radio Access Network) and IP Core Network, and the IP-based threat vector that had not been present in previous generations of mobile networks were unleashed by these changes.

Small-sized computers called smartphones have been replaced by mobile devices, hosting typical OS vulnerabilities and weaknesses. Smartphones now require regular patching and system vulnerability updates, as any failure will expose the phone to threats from attackers who can take advantage of leaked data vulnerabilities or install viruses and spyware.

Installation of unauthorized or malicious software triggered hacking of phones which was often used to weaken the service efficiency of the mobile service providers and target the network. For a centralized app store, some handset vendors were able to enforce a stringent device protection strategy, while others find it difficult to cope with the ever-growing number of malicious codes hosted on their app store website. [1]

4.2.4 Security Threats and Protection for 4G

For the first time, LTE moved the mobile network to an all-IP-based end-to-end architecture. It helped to speed up competition for mobile service providers, offered innovative technologies and size, and even expanded the risk vector for 4G networks. The new realities for the mobile network were DDoS (Distributed Denial of Service) and APT (Advanced Persistent Threats), as the effect on the service was critical as a result of such attacks, with big financial losses. Attackers were more coordinated with their challenge of execution and began adopting a systemic strategy. It has become harder to detect their stealth presence in the mobile network, to protect and mitigate, with an average attack consisting of months' duration.

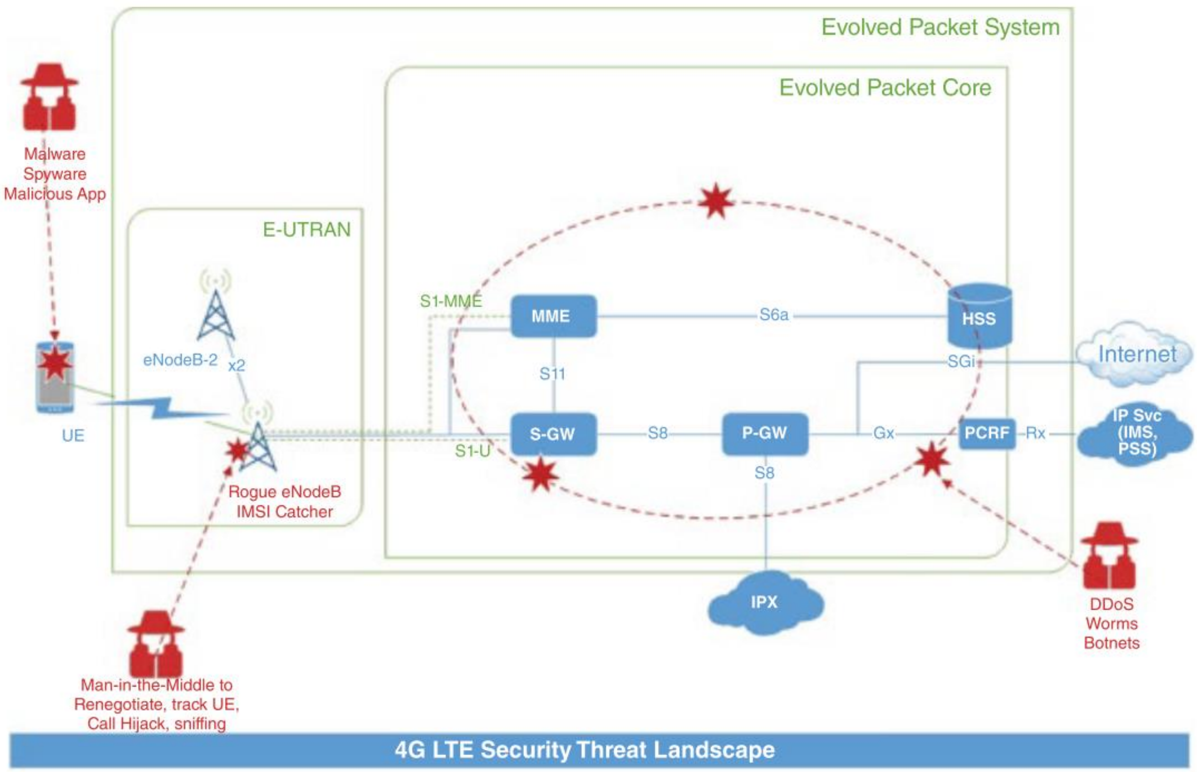


Fig 4.4: 4G end-to-end security threat landscape [1]

Overall, 4G LTE and LTE Adv. have introduced a major jump and evolution in the efficiency and throughput of the mobile network. Via a popular mobile architecture, it offers speech, data, video, and internet services.

As seen in Fig 4.4, 4G attacks have been spread across different domains of the 4G network. To steal personal identification and passwords, new forms of viruses and ransomware have been aimed at smartphones. To mimic consumer games, utilities, or fake major banking applications, millions of malicious apps have been created.

With IP core networks, 4G networks have been attacked to cause a greater effect on mobile services with well-designed DDoS (Distributed Denial of Services) attacks. 4G LTE security can be broken into several areas, such as UE, RAN, Internet Networks, and Core Network. In the following pages, individual domain security risks for 4G are covered. [1]

4.2.4.1 LTE UE (User Equipment) Domain Security

Today's UE is a versatile internet connection for lightweight, high-speed CPU (Central Processing Unit) and memory-capable portable computers. It serves not only to communicate with each other in our everyday lives, social and financial events but to make online purchases and bank transfers. Often connected via wireless LAN or cellular, smartphones may run an interdependent operating system and the software application that allows users to access their data anywhere, any time, and any place. A basic mobile operating system vulnerability may have a huge impact; a recent 'XCODEGHOST' vulnerability discovered in an iOS tool impacted 500 million users as a reference.

For about 87 percent of the time spent on mobile devices, smartphones are used today, and at least 24.7 percent of mobile apps hold one high-risk security flaw. Malicious malware can be either purposely or accidentally downloaded and installed by users or attackers. These malicious applications can be used for financial and other illegal gains to gain stealth access to user's data and passwords (stored on the phone).

Malware and worms can be mounted to initiate an attack on the local consumer network or to exploit the mobile service provider network generally since mobile service providers consider the smartphone to be a trustworthy network unit. [1]

4.2.4.2 LTE (Remote Access Network) Domain Security

Using its Cell Radio Network Temporary Identifier (C-RNTI), LTE E-UTRAN can be used to obtain access to UE positions when UE remains in a single cell or roams through several cells. This attack can be secured by the encryption of the control signal carrying traffic and the command and validation of C-RNTI messages. [1]

4.2.4.3 LTE Core Network Domain Security

LTE is built for an open end-to-end IP-based network architecture that helps to model the overall network activities but exposes the mobile network to IP-based security threats on the other side.

A distributed denial of service (DDoS) attack with an effect as high as the interruption of network services to millions of customers may be a possible target for the LTE core network. To cause a lack of operation, DDoS will attack essential EPC (Evolved Packet Core) components, such as submitting overwhelming requests for authentication and authorization to the HSS database, creating device overload conditions. By injecting bogus routing information or modifying the network database to cause service downtown, a DDoS may be run against an entire IP network backbone. The simplest type of DDoS attack is the TCP SYN attack, where millions of fake TCP SYN packets are sent to a network system to cause a denial of service status.

DDoS Protection can be achieved with a customized Anti-DDoS security system, or it is possible to use newer network technology such as Software Defined Network (SDN) to push security policies via a centralized controller.

4.2.4.4 Security Threat Analysis for 4G

The detailed threat analysis of typical 4G risks, their severity of impact, and the probability of threat occurrence are covered in Table 4.1 below:

Threat	**Threat description**	**Impact Severity**	**Threat Occurrence**
Insecure Mobile OS (Operating System)	Mobile operating systems carry vulnerabilities that are fixed, using vendor-issued patches and updates. If not fixed, can cause attackers to exploit vulnerabilities to hack into mobile systems	Moderate	4
Download unauthorized apps	Users download the app from the app store that is not verified by the vendor or checked by their IT department and can be malicious	Moderate	5
Insecure App with sensitive data	A legitimate app that leaks sensitive personal or business data and with no mechanism to encrypt or protect	Severe	5
Virus	Malicious software code with a specific purpose to damage mobile functions or files	Severe	2

Malware	An advanced virus or malicious app that can propagate and self-reproduce causing large-scale, network-wide damage	Extreme	3
Spyware	A malware type used to steal end-user data, sensitive information to transmit to remote attackers	Extreme	3
DDoS (Distrusted Denial of Service)	Launched as a coordinated attack involving hundreds of thousands of devices infected with malicious code. Targets the availability of mobile networks	Extreme	2

Table 4.1: 4G Security Threat Analysis [1]

The following steps and processes are involved in the defense mechanisms against 4G security threats:

- On your mobile devices, always install approved applications.
- Install software only from the app store of the vendor.
- With a passcode, protect smartphone entry.
- Defines a regulation of service access for each app, i.e. limiting location and communication access to certain categories of applications.
- Allow mobile device encryption.
- To encrypt and monitor confidential data, enterprise and essential software can use stable containers.
- Install Antivirus.
- Having the operating system still patched and modified. [1]

4.2.5 Security Threats and Protection for 5G

5G is aiming to link billions of smartphones, tablets, over an extremely secure, broadly dense, bandwidth-capable, fast, and fault-tolerant network system that would support several sectors and industries. Critical networks, IoT, digital cities, and the digital planet are the main usage cases for 5G. 5G would be a perfect option for attackers of these use cases that would wish to create substantial economic and social damage within a minimal timeframe. 5 G challenges can be generated around financial and politically driven benefits, carried out with comprehensive technological expertise and tools by groups of experts and offenders. The security ecosystem for 5G would be complex and dependent on advanced and diverse attacks, such as Stuxnet and fire malware.

The challenge vector for 5G can be broad, with the broad scope of 5G technologies and services and its vital position in supporting society for civil, economic development, and public safety. Motivations to challenge and attack 5G are now going to be stronger than in past iterations of networks. The 5G is more likely to be a primary focus for crime operations motivated by many diverse reasons, such as state-sponsored political motives, rivals, cartels of organized crime, spying, and cyberwarfare.

Attackers continue to evolve and discover new means of escaping identification, having learned to take advantage of the social and financial environment for their needs. With digital payment systems emerging and technology such as Bitcoin joining the mainstream, it would be much easier for offenders to hide under the blanket and continue to reap financial benefits.

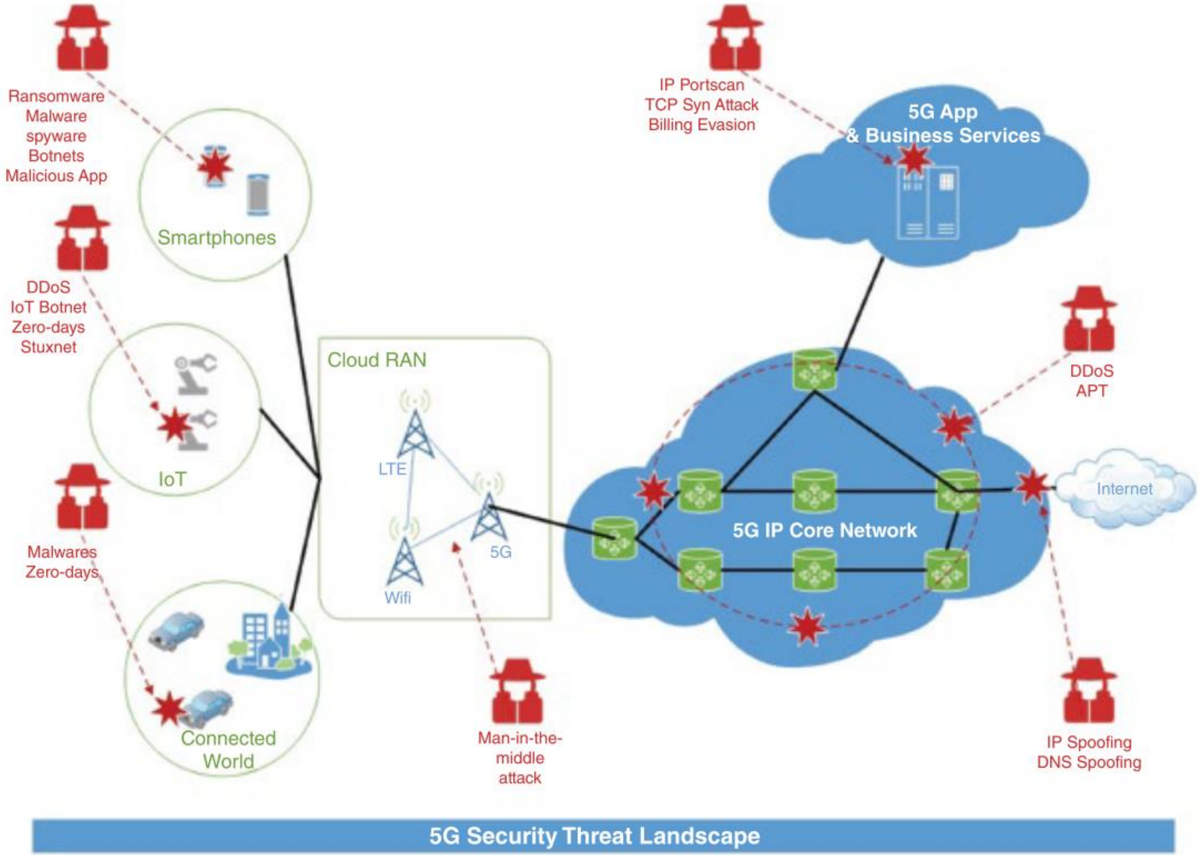


Fig 4.5: 5G security threat landscape [1]

There would be no limitation to the 5G vulnerability vector as it ranges from end-user devices such as smartphones, consumer goods, switches, home automation, autonomous vehicles, business networks, and mobile networks. For such a wider vulnerability environment, it would range from end-user machines to RAN (Radio Access Network) to the cell core network to the Internet, as seen in Fig 4.5 above. Fig 4.5 also includes the location within the network of different types of threats, such as types of smartphone threats, including ransomware, spyware, and Bots. A MITM attack on the Cloud RAN domain may be initiated, while DDoS will hit the IP core network, in such a landscape any single 5G network domain would be attacked.

For security professionals to step up and develop a defensive mechanism would be a major challenge.

With 5G networks, apps and utilities will be 10 times, as it will be the platform for the new and wired world. As with other conventional services, such as energy, 5G will act as the vital infrastructure which will be used to provide access for urban health and government, banking, commerce, and manufacturing networks.

The 5G architecture will be placed on top of the IP-based architecture which will inherit and extend the features of traditional IP-based mobile networks based on the current 5G standardization and testing work so far. [1]

4.2.5.1 Next Generation Threat Landscape for 5G

Not only will 5G be the latest generation of mobile networks, but the next generation of security vulnerabilities will also be implemented as a portal. Security risks of the next generation are projected to bear the following characteristics:

- Sophisticated: complex in nature, dynamic in design, using a multi-stage combination of different vectors and methods for the attack. The Angler exploit kit, for example, which is packaged to exploit multiple vendors into a single attack;
- Obfuscatory: threats that are blurred and very difficult to identify by several layers;
- Evasive: difficult to detect and able to cover itself, e.g. ransomware crypto wall attack; and
- Persistent: with any unsuccessful attempt, those attacks are intended to be consistent and develop themselves.

APTs (Advanced Persistent Threats) are attacks that carry out the above functionality, are difficult to counter and defend from, and consequently become severe threats that inflict great harm. Typically, APTs attack critical infrastructure installations, major service suppliers that serve the masses and create considerable disturbance and multi-dimensional effects. [1]

4.2.5.2 IoT Threat Landscape

For future industrial systems, critical infrastructure, and IoT, 5G will serve as the network platform (Internet of things). Attacks targeting such vital networks are expected to be sophisticated in nature and involve the operation of a highly complex skillset and capital. In the 5G environment, politically motivated and sponsored attacks will often be encountered. These attacks will have the potential to leverage the vulnerabilities of the undisclosed system, also referred to as zero-day flaws, and operated by a system of centralized command and control (C&C). No single tactic or method will be used for such attacks, and a combination of threat tactics such as DDoS, phishing, and specialized rootkits will often be used. The effects of such attacks are widespread and go beyond economic damage, and may include national security, public safety, and loss of life.

A centralized management system, such as SCADA (Supervisory Control and Data Acquisition), which is further linked to a network, is found in most essential infrastructures today. For normal operation, all remote devices need to transmit periodic signaling and health information to the system, so that a malware type of attack may be used to interrupt the control system and result in in-service failure. In the future, 5G will serve the underlying network and communication networks for these devices and will be vulnerable to risks to these systems. [1]

4.2.5.3 5G Evolved Security Model

We need an evolved security model (as seen in Fig 4.6) to defend 5G from advanced and diverse threat environments that provide in-depth protection not only from current threats but also from emerging and zero-day threat forms.

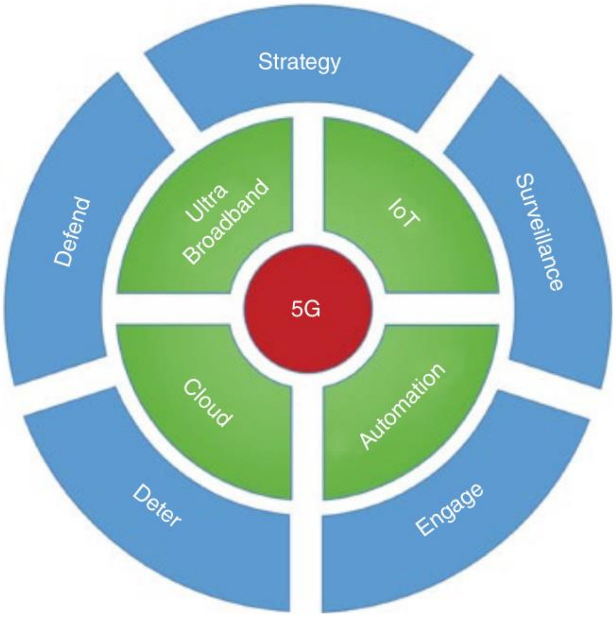


Fig 4.6: 5G Evolved Security Model [1]

A well-defined security policy and plan for securing the different components of the 5G network, including the end devices and end-users, would be needed. Based on telemetry data gained from a well-developed surveillance system, an efficient security strategy can be built.

To place the safety measures ineffective locations, a protective location strategy will be needed. To participate in the event of an attack being observed, block it, and develop the method, procedures and resources must be defined. For the continuous security of the mobile network and the related networks, a deterrent and defensive mechanism should be in operation. [1]

4.2.5.4 5G Security Threat Analysis

Mastering and studying the latest problems and risks to LTE and LTE Advanced networks is one of the best methods to plan for security challenges in 5G. As mentioned earlier, it will inherit most of the threats that currently occur in 4G (LTE and LTE Advanced) networks due to the central IP-based existence of 5G.

Threat	Threat description	Impact Severity	Threat Occurrence
Ransomware	Specialized malware use exploits, encrypt, and lock access to critical data. Access granted after paying demanded ransom money	Severe	3
Advance Malware Advance	Advance malware targeting billions of mobile and IoT devices with the capability to exploit the OS and network vulnerabilities	Extreme	3
IoT Botnets	IoT and mobile devices hosting a control agent/bot receiving remote commands and continuously leaking telemetry information to a remote bot-master running a central command and control (C&C) system. Used for both passive and active attacks	Severe	2
Critical Infrastructure Threats	Threats that are focused, damaging critical infrastructure services such as SCADA, i.e. Stuxnet, Shamoon attacks	Extreme	3
Zero-day Attacks	An advance attack exploiting the undiscovered vulnerabilities of a system. Can be a combination or package of	Extreme	1

	multiple attack types, malware, rootkits, and botnets		

Table 4.2: 5G Security Threat Analysis [1]

4.3 Mobile Security Lifecycle Functions

To secure the end-to-end security posture of mobile systems and networks, security lifecycle functions are generated for mobile devices and networks. It addresses the security at individual levels of mobile provisioning, setup, evaluation, and security control. To secure the secrecy, privacy, and compatibility of mobile devices and networks, lifecycle functions exploit protection technologies, resources, and processes.

As 5G networks, new tools such as IoT and ultra-broadband will be launched to carry out mobile e-commerce and new enterprise usage cases for mobiles. In particular, certain organizations would look to their mobile end users for well-defined security management and governance scheme to further secure their sensitive data and apps residing on personal mobile devices used by end-users as an extension of a corporate network, as in the case of BYOD (Bring Your on Device).

Main protection failures in such situations are:

- Contradictory protection policies;
- Shared media leakage;
- Limited system management;
- Data remains indisposed devices and are readable; and
- Data leakage in inter-application.

As seen in Fig 4.7, a security lifecycle will help solve the above and other typical security issues and minimize threats at different points of the mobile device's presence in the network. App provisioning, setup, maintenance, and control are the main steps of a mobile handled by the security lifecycle. We will discuss the security lifecycle roles for mobile systems and networks in depth in the sections below. [1]

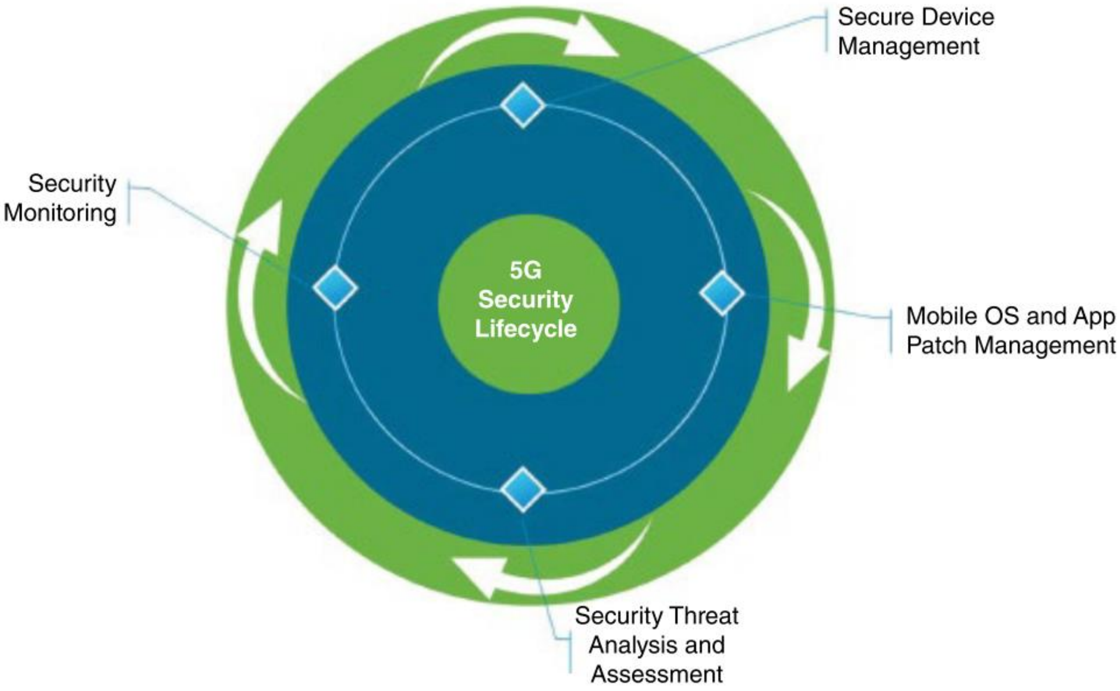

Secure Device Management
Security Monitoring
5G Security Lifecycle
Mobile OS and App Patch Management
Security Threat Analysis and Assessment

Fig 4.7: Mobile Security lifecycle functions [1]

4.3.1 Secure Device Management

Although MSP (Mobile Service Provider) has little control over mobile device security, they do provide simple mobile device authentication and authorization for network access. Companies that use BYOD to provide corporate access to personal mobile devices, however, need specialized management skills to use software such as MDM (Mobile Device Management) to centrally control and monitor their employees' mobile devices. MDM lets businesses adopt security policies, defend against cyber-attacks and limit unauthorized access to mobile devices. Before they download the security protocols, settings, and controls to secure the device, mobile users need to register or enroll with the integrated enterprise MDM. [1]

4.3.2 Mobile OS and App Patch Management

Patches are periodically published by the smartphone OS (Operating System) provider to close identified bugs and loopholes in their applications that may cause application compromise or significant security breaches in some cases. Attackers also search for ways to get hold of and hack these bugs to obtain unauthorized access to mobile devices. For consumers, it is important to periodically repair their running mobile devices. To prevent a security bug, mobile service providers also advise their customers to upgrade their devices to fix patches and make improvements.

Similarly, software creators often give their mobile applications daily patches to fill in for any dangerous functionality or code existing in their apps. Patches for apps are independent of OS updates and need to be managed by smartphone users independently. Mobile OS providers also make this task easy with their app stores, since it is impossible to keep certain applications up-to-date for custom device installs. [1]

4.3.3 Security Threat Analysis and Assessment

As we know, 5G networks will focus heavily on IP-based network connectivity and protocol and exploit the latest software-defined mobile networks (SDMN). To optimize networks and deliver innovative and better technologies utilizing new programmable networks, 5G is aiming to exploit the advantages of SDMN to detach the management plane, control plane, and data plane from cell networks. SDMN offers new protection problems with its capabilities which can trigger vulnerabilities on various network (management, control, and data) planes. Threat vectors are complex for 5G and SDMN-based networks and can dynamically implement network vulnerabilities at multiple points on mobile networks.

The stagnant and simple design of conventional mobile networks is focused on traditional vulnerability evaluations and threat detection methods and does not resolve the problems posed by complex network activities based on 5G and SDMN. Both modules and layers (planes) of the mobile network need to be protected and discussed by security evaluations for SDMN networks. New approaches to security evaluation that reflect on the complex design of SDMNs are recommended. To cover all SDMN variables and stages, such new evaluations can use attack graphs and will need to use an Analytic Hierarchy Process (AHP) to construct the structure. [1]

4.3.4 Security Monitoring

Mobile RAN and key network technology have already developed with the development of mobile networks from LTE to LTE Adv. and now 5G, and are superseded by emerging technologies such as Cloud RAN, NFV, and SDMN. Mobile operators need to provide their network operators with detailed visibility and information in real-time, not only to provide better service reliability but also to protect their vital network infrastructures from security threats. Legacy mobile security monitoring systems were not developed to secure SDN or NFV-based networks and have little or restricted capacity to interface with the mobile network's latest technical components, so security monitoring solutions that have greater efficiency, scalability, and the ability to adapt and function with these emerging mobile technologies need to be replaced.

Security management systems for LTE and 5G networks, starting from UE to RAN and LTE/5G core network modules can provide the ability to track and audit all signal and data traffic at various network locations. The solution should be able to not only inspect the IPv4 and IPv6 but also offer visibility to other protocols such as TCP, UDP, GRE, etc. Instead of traditional packet-based inspection, 5G networks could also leverage SDN control and data plane separation and perform centralized network flow traffic monitoring for deeper visibility and correlation of traffic traversing inside the network.

Mobile network compliance management needs to provide some advanced security services like:

- Tests for vulnerability;
- Daily security health check for the whole network;
- Exposure of the flow-based network;
- Security warning control system; and
- Tracking and inspection of traffic. [1]

Chapter 5 – 5G NR Use Cases

5.1 Introduction

In this chapter, we will discuss and analyze different use cases of 5G. It covers the attacks that might offer in a 5G environment in different layers. For instance, the DoS attack or SYN-FLOOD attack occurs in the transport layer, whereas the Sink-Hole attack occurs in the network layer.

5.2 5G Use Cases

5.2.1 Use Case 1: How the Distance and Throughput vary in FR1 and FR2?

The formula to calculate throughput for 5G NR is,

$$\text{data rate (in Mbps)} = 10^{-6} \cdot \sum_{j=1}^{J} \left(v_{Layers}^{(j)} \cdot Q_m^{(j)} \cdot f^{(j)} \cdot R_{\max} \cdot \frac{N_{PRB}^{BW(j),\mu} \cdot 12}{T_s^{\mu}} \cdot \left(1 - OH^{(j)}\right) \right)$$

[15]

wherein,

J is the number of aggregated component carriers in a band or band combination

$R_{\max}$ = 948/1024

For the j-th CC,

$v_{Layers}^{(j)}$ is the maximum number of layers (gNB Tx streams to UE)

$Q_m^{(j)}$ is the maximum modulation order

2 – OPSK,

4 – 16QAM,

6 – 64QAM,

8 – 256QAM

$f^{(j)}$ is the scaling factor

The scaling factor can take the values 1, 0.8, 0.75 and 0.4.

$f^{(j)}$ is signalled per band and per band per band combination

μ is the numerology

In the context of 3GPP 5G standardization contributions, the term numerology refers to the configuration of waveform parameters, and different numerologies are considered OFDM-based sub-frames having different parameters such as subcarrier spacing/symbol time, CP size, etc. [16]

Numerology can be established from the sub-carrier spacing:

0 – 15kHz SCS (Sub-Carrier Spacing)

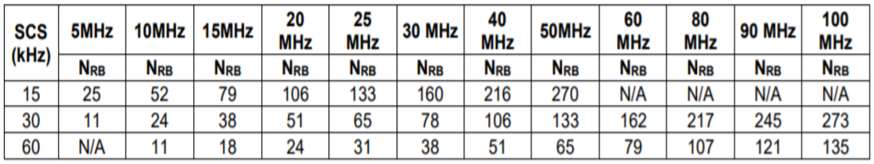
1 – 30kHz SCS

2 – 60kHz SCS

T_s^{μ} is the average OFDM symbol duration in a sub-frame for numerology μ, i.e. $T_s^{\mu} = \frac{10^{-3}}{14 \cdot 2^{\mu}}$. Note that the normal cyclic prefix is assumed.

$N_{PRB}^{BW(j),\mu}$ is the maximum Resource Blocks (RB) allocation in bandwidth $BW^{(j)}$ with numerology μ

RB allocation can be found from the table released by 3GPP



SCS (kHz)	5MHz	10MHz	15MHz	20 MHz	25 MHz	30 MHz	40 MHz	50MHz	60 MHz	80 MHz	90 MHz	100 MHz
	N_{RB}	N_{RB}	N_{RB}	N_{RB}	N_{RB}	N_{RB}	N_{RB}	N_{RB}	N_{RB}	N_{RB}	N_{RB}	N_{RB}
15	25	52	79	106	133	160	216	270	N/A	N/A	N/A	N/A
30	11	24	38	51	65	78	106	133	162	217	245	273
60	N/A	11	18	24	31	38	51	65	79	107	121	135

[17]

$OH^{(j)}$ is the overhead and takes the following values

For FR1 (450 MHz to 7.125 GHz) FR2 (24.25 GHz to 52.6 GHz)

0.14, for frequency range FR1 for DL

0.18, for frequency range FR2 for DL

0.08, for frequency range FR1 for UL

0.10, for frequency range FR2 for UL [15]

For example, with the following input:

Layer: 4, Modulation: 256QAM (Modulation order: 8), Scaling Factor: 1, Numerology: 1 (μ:30kHz), Bandwidth: 50 MHz, Resource Blocks: 133 (from Bandwidth and Numerology), Overhead: 0.14

The calculated 5G NR Throughput, in Mbps, is **976**

Layer: 4, Modulation: 64QAM (Modulation order: 6), Scaling Factor: 1, Numerology: 1 (µ:30kHz), Bandwidth: 50 MHz, Resource Blocks: 133 (from Bandwidth and Numerology), Overhead: 0.14

The calculated 5G NR Throughput, in Mbps, is **732** [18]

Now, let's try to understand how the throughput varies with distance for FR1 and FR2, we can use a tool called NetSim for this purpose.

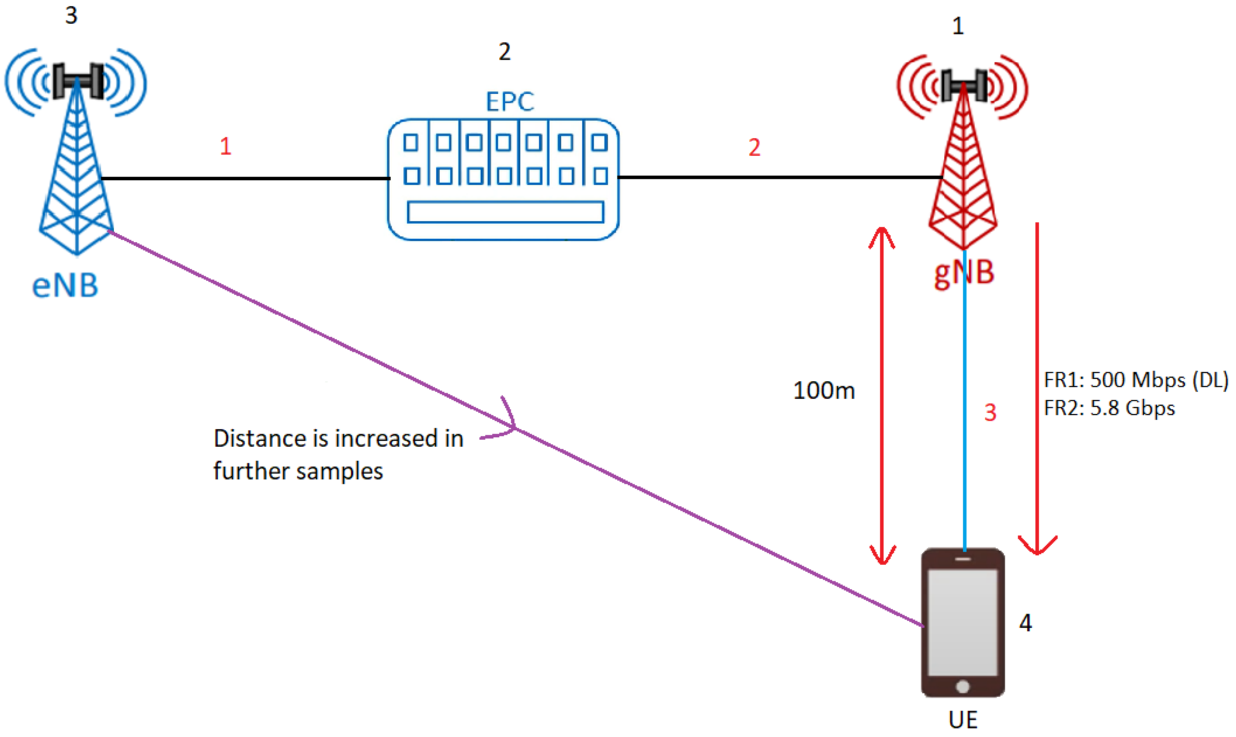


Fig 5.1: Distance and Throughput variation in FR1 and FR2

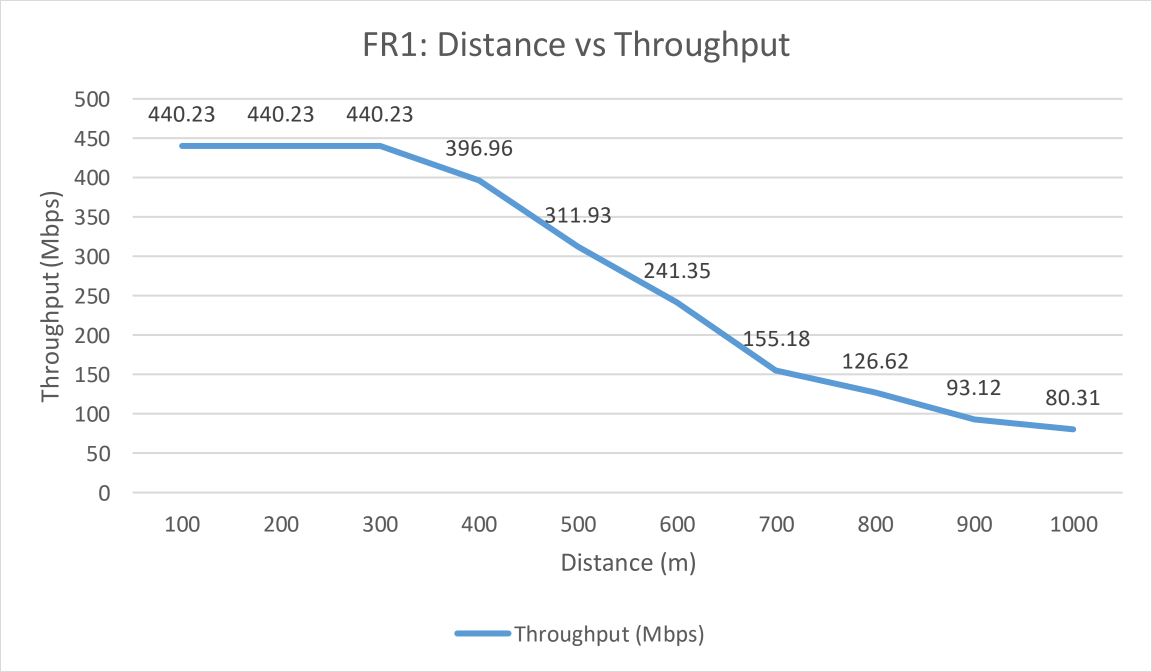
We know that the throughput drops faster in FR2. In NetSim, setup Full Buffer or Saturation download traffic and move the UE away from the gNB and study the throughput vs distance variation.

An increase in the distance leads to an increase in Pathloss, which leads to lower received signal strength and lower SNR, this, in turn, results in lower throughput. In this example, the path loss model was set to urban-macro to obtain the below results.

For FR1:

Distance (m)	Pathloss (dB)	SNR (dB)	Modulation	Throughput (Mbps)
100	97.34	40.47	64QAM	440.23
200	109.04	28.77	64QAM	440.23
300	115.92	21.89	64QAM	440.23
400	120.80	17.01	64QAM	396.96
500	124.59	13.22	64QAM	311.93
600	127.68	10.13	64QAM	241.35
700	130.30	7.51	64QAM	155.18
800	132.56	5.24	16QAM	126.62
900	134.56	3.24	16QAM	93.12
1000	136.35	1.45	16QAM	80.31

Table 5.1: Distance vs Peak Throughput in FR1 [19]



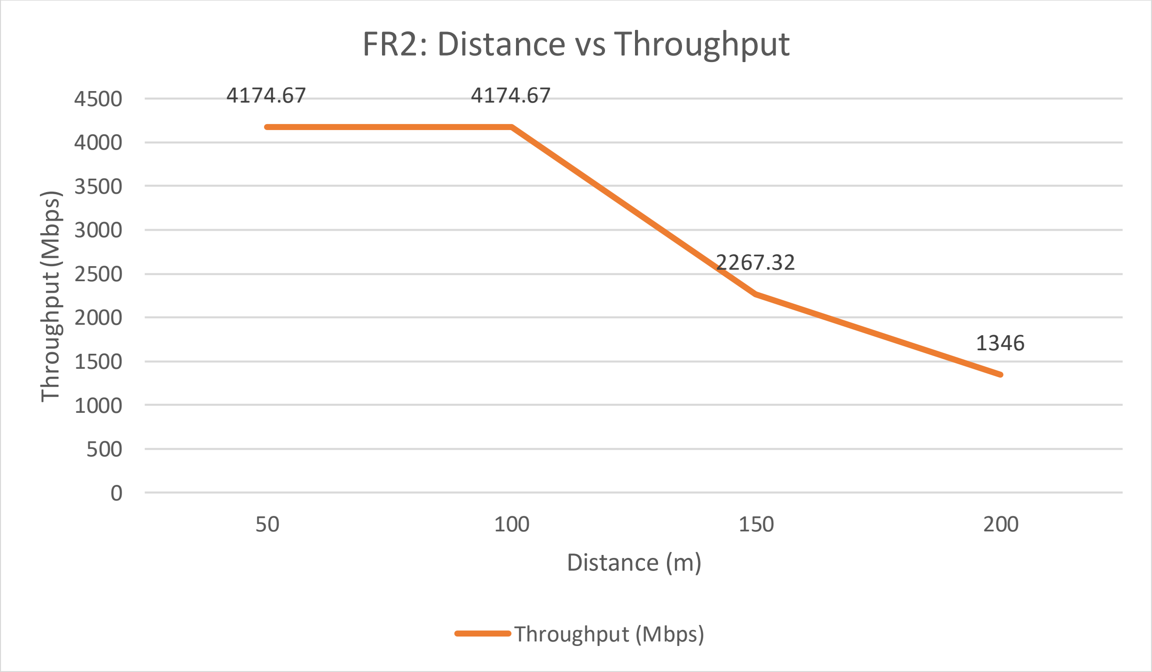
Graph 5.1: Distance vs Throughput in FR1 [19]

Throughput for FR1 was 440.23 Mbps initially and it dropped down to 80.31 Mbps over 1000 meters as shown in Graph 5.1.

For FR2:

Distance (m)	Pathloss (dB)	SNR (dB)	Modulation	Throughput (Mbps)
50	109.10	24.73	64QAM	4174.67
100	120.68	13.15	64QAM	4174.67
150	127.53	6.30	16QAM	2267.32
200	132.40	1.43	16QAM	1346.00

Table 5.2: Distance vs Peak Throughput in FR2 [19]



Graph 5.2: Distance vs Throughput for FR2 [19]

In FR2, the initial throughput was 4174.67 Mbps and it dropped almost 1/3rd of it, just in 200 meters as shown in Graph 5.2.

5.2.2 Use Case 2: DoS Attack in 5G NR

A Denial of Service (DoS) attack, such as blocking access to a website, is an attempt to make a device inaccessible to the intended user(s). A successful DoS attack uses all network or device energy, resulting in a server bottleneck or crash. If a DoS attack is orchestrated by several outlets, it becomes known as a DDoS (Distributed Denial of Service) attack. SYN flood, UDP flood, SMB Loris, ICMP flood, and HTTP GET flood are some types of DDoS attack.

SYN Flood:

DoS attacks that threaten to overload the DNS server with new TCP connection requests are known as TCP SYN floods. In most cases, a recipient initiates a TCP communication with a three-way message handshake:

- The client requests a connection by sending a SYN (synchronize) message to the server.
- The server acknowledges the request by sending SYN-ACK back to the client.
- The client answers with a responding ACK, establishing the connection.

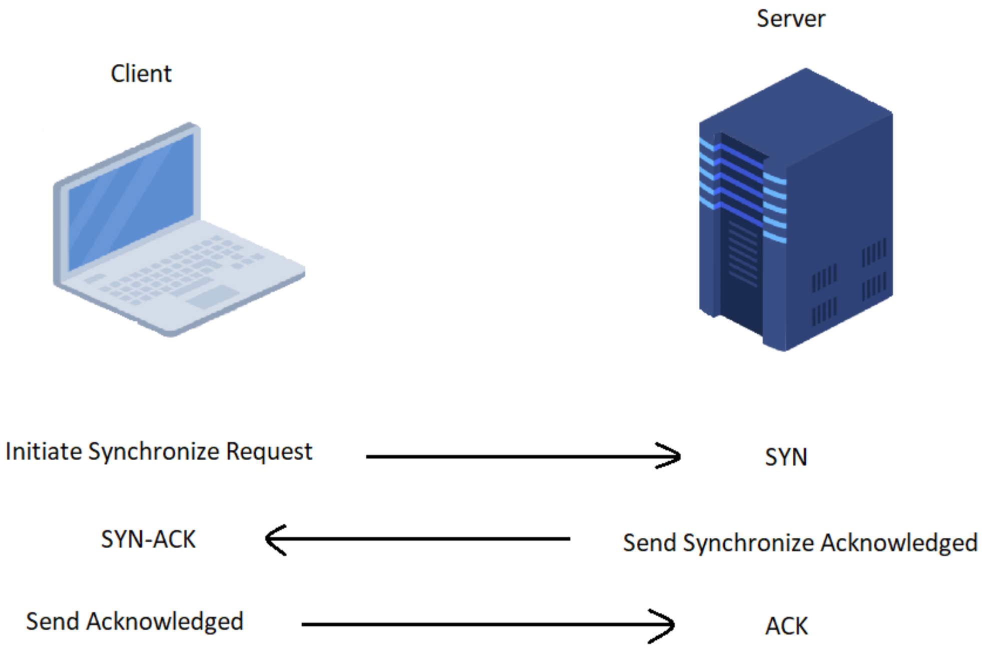


Fig 5.2: SYN Flood

The basis for any relation formed using the Transmission Control Protocol (TCP) is this triple exchange. One of the most common types of DDoS attacks is the SYN Flood. It happens when an intruder sends a series of TCP Synchronize (SYN) requests to the target to deplete the server's capacity to the point that it is inaccessible to legitimate users. Since a SYN request establishes network contact between a potential client and the target server, this happens. When a SYN request is received by the server, it responds by accepting the request and keeps the communication open while waiting for the client to recognize the open connection. However, the client acknowledgment never arrives in a successful SYN Flood, thus consuming the resources of the server until the connection times out. All available server resources are consumed by a large number of incoming SYN requests to the target server, resulting in a successful DoS attack. [20]

Create a file called “SYN_FLOOD.c” with following functions to do the above mentioned tasks:

Step 1: Before we begin, check if the node is malicious or not

Pseudo Code:

```
int malicious_node[NUMBEROFMALICIOUSNODE] = {6,8};
int is_malicious_node(DevId) {
        for (int i=0; i<NUMBEROFMALICIOUSNODE; i++)
                if (DevId == malicious_node[i]) return 1;
        return 0;
}
```

We are defining a variable called malicious_node(NUMBEROFMALICOIUSNODE) which indicates the nodes which are to be checked for maliciousness. Create a function called “is_malicious_node” with variable (DevId). DevId is the Device ID of the malicious node. This function is used to check if the node is malicious node or not.

Step 2: Create a new socket and update socket parameters

Pseudo code:

```
int socket_creation() {
        newSocket = tcp_create_socket();
        ----- Define socket parameters -----
        return newSocket;
}
```

We are defining a function called “socket_creation”, which uses a default command to create a socket. Then we define socket parameters like the Socket ID and assign it to a variable s_id.

Step 3: Create and send SYN packet to the network layer

Pseudo code:

```
static void create_syn() {
    ----- Create a SYN packet -----
    return 1;
}
static void send_syn_packet(){
    syn = create_syn();
    ----- Use pointers to send the packet in the TCP connection which was created between
    Client and Server -----
}
```

We are defining a function called "create_syn", which uses a default command to create a SYN packet and then we define a function called "send_syn_packet", which sends this SYN packet to the server on which we want to attack.

Step 4: Add timer event called SYN_FLOOD which triggers every 1000 micro-second

We are defining a function called "syn_flood", which is used to check whether the socket is present or not and also adds a timer event called SYN_FLOOD which is set to 1000 micro-seconds.

By following all the above steps we can create a malicious node in the NetSim simulation software. Now, let's create different scenarios with this aspect.

Case 1: Without malicious node in the network

In Fig 5.3, we have created a network scenario of 2UEs, 1gNB, 1EPC, 1Router, and 1eNB forming a 5G NR mmWave network. Traffic is generated from UEs towards the gNB. We will set the NUMBEROFMALICIOUSNODE value to 0 for this case and run the simulation.

The output gives the throughput value of **11.63 Mbps**.

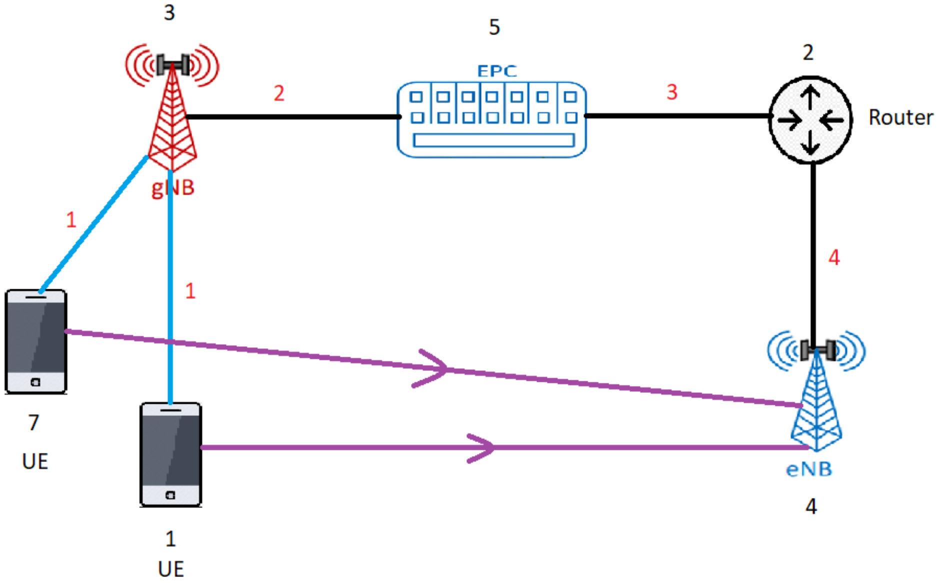


Fig 5.3: Case 1 - Without malicious node

Case 2: With one malicious node

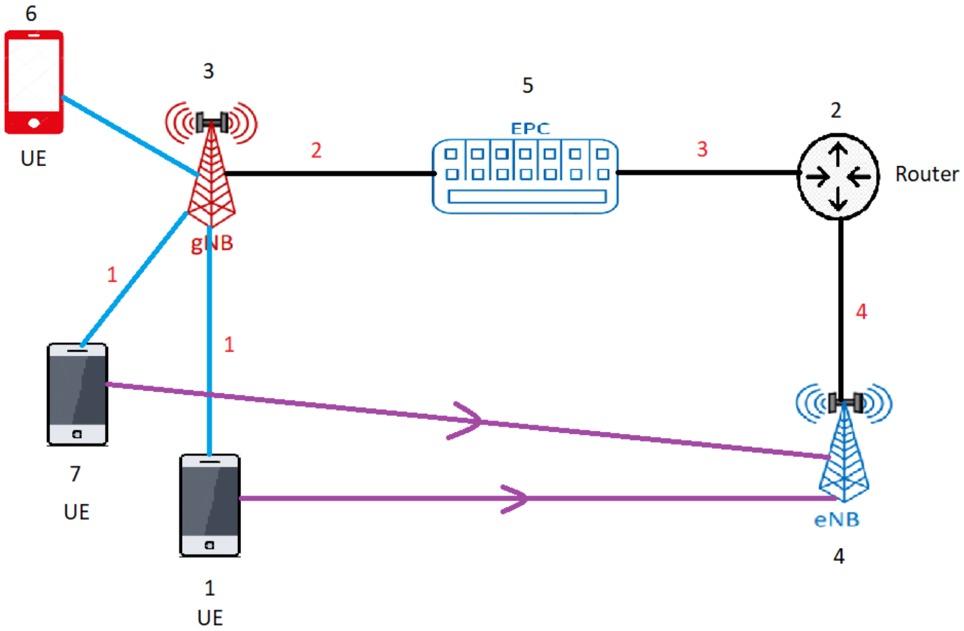


Fig 5.4: Case 2 - With one malicious node

Fig 5.4 is similar to Fig 5.3, except, it has one malicious node (node 6) in the network. In this case, we will set the NUMBEROFMALICIOUSNODE value to 1, then run the simulation.

The output gives the throughput value which is **11.47 Mbps**.

NOTE: There is a gradual drop in the throughput.

Case 3: With two malicious node

Fig 5.5 shows two malicious nodes (node 6 & 8). In this case, we will set the NUMBEROFMALICIOUSNODE value to 2, then run the simulation.

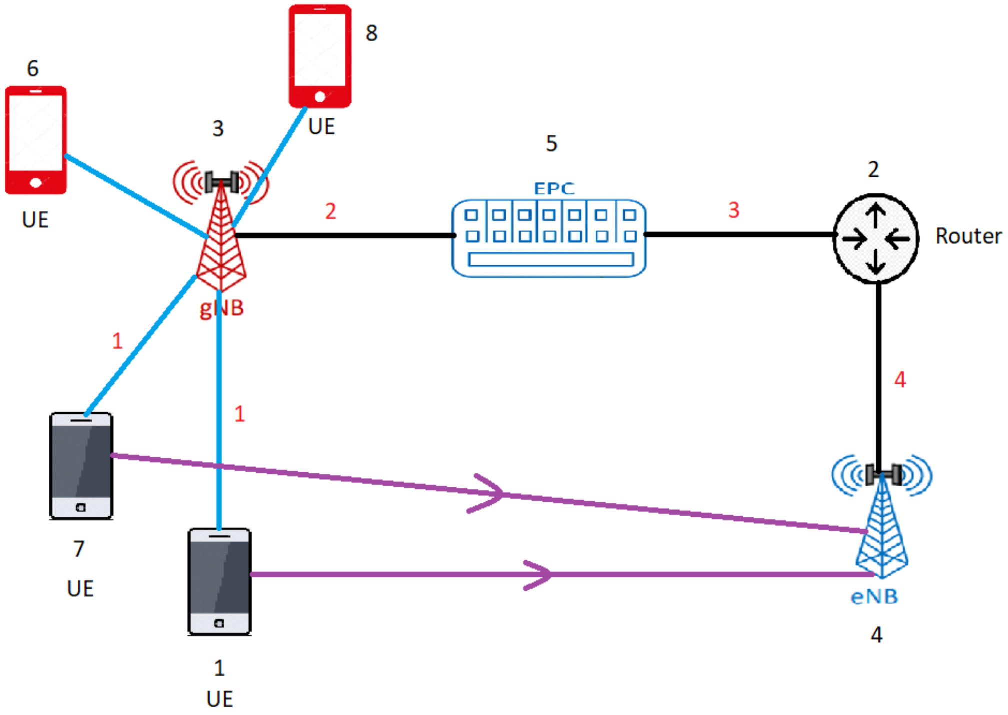


Fig 5.5: Case 3 – With two malicious node

The output gives the throughput value which is **11.28 Mbps**.

This is because of the SYN_FLOOD attack. The server resources are all allotted to the malicious nodes, hence the performance of the server decreases gradually over a while. It's important to note that, as the number of malicious nodes increases, the throughput value of the server decreases, hence the Denial of Service (DoS) to actual users (in this case node 1 & 7). [20]

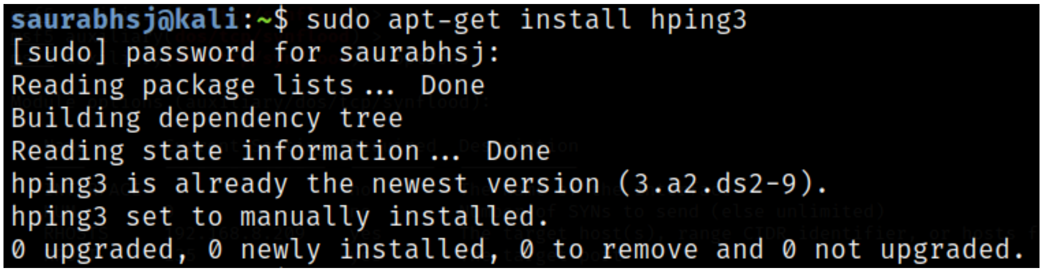
Implementation of TCP SYN FLOOD DOS attack and detect using Wireshark:

We will be using Kali VM as the attacker machine, Windows VM as the target machine, and hping3 in Kali machine to spoof IP random addresses, and Wireshark in Windows machine to detect the SYN packets.

Step 1: Install hping3 in the Kali machine

Command: **# sudo apt-get install hping3**

The above command installs the hping3 tool into the Kali (Attacker) machine.



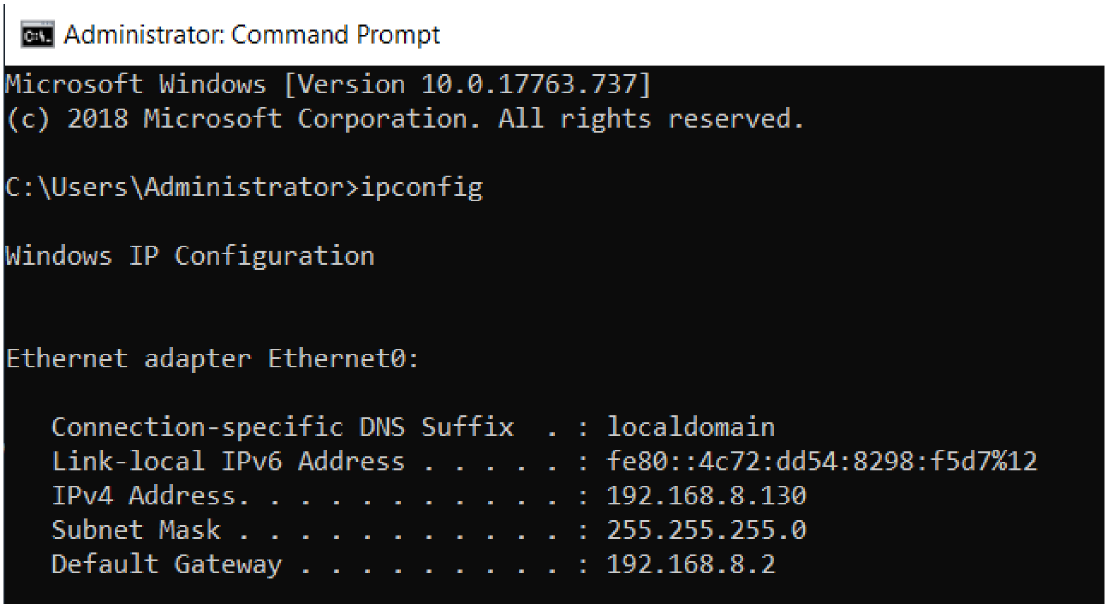
```
saurabhsj@kali:~$ sudo apt-get install hping3
[sudo] password for saurabhsj:
Reading package lists... Done
Building dependency tree
Reading state information... Done
hping3 is already the newest version (3.a2.ds2-9).
hping3 set to manually installed.
0 upgraded, 0 newly installed, 0 to remove and 0 not upgraded.
```

Fig 5.6: Install hping3

Step 2: Find the IP address of the Windows machine

Command: **# ipconfig**

The above command displays the IP address of the windows machine.



```
Microsoft Windows [Version 10.0.17763.737]
(c) 2018 Microsoft Corporation. All rights reserved.

C:\Users\Administrator>ipconfig

Windows IP Configuration


Ethernet adapter Ethernet0:

   Connection-specific DNS Suffix  . : localdomain
   Link-local IPv6 Address . . . . . : fe80::4c72:dd54:8298:f5d7%12
   IPv4 Address. . . . . . . . . . . : 192.168.8.130
   Subnet Mask . . . . . . . . . . . : 255.255.255.0
   Default Gateway . . . . . . . . . : 192.168.8.2
```

Fig 5.7: IP address of the victim machine

Note the IP address of the machine.

IP address: **192.168.8.130**

Step 3: Generate SYN packets and direct them to the target machine (Attacking the victim machine)

Command: **# hping3 -c 15000 -d 120 -S -w 64 -p 80 --flood --rand-source 192.168.8.130**

where, -c 15000: 15000 packets

-d 120: the size of 120 bytes each

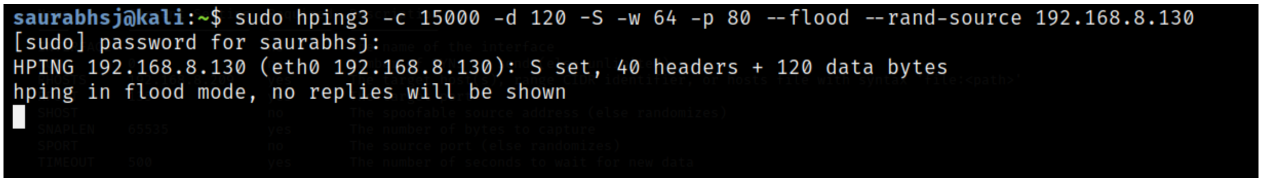
-S: Specifies SYN Flag should be enabled

-w 64: TCP window size of 64

-p 80: Direct the attack to the victim's HTTP web server

--flood: Flag to send packets as fast as possible

--rand-source: Flag that generates spoofed IP addresses to disguise the real source and avoid detection but at the same time stop the victim's SYN-ACK reply packets from reaching the attacker



```
saurabhsj@kali:~$ sudo hping3 -c 15000 -d 120 -S -w 64 -p 80 --flood --rand-source 192.168.8.130
[sudo] password for saurabhsj:
HPING 192.168.8.130 (eth0 192.168.8.130): S set, 40 headers + 120 data bytes
hping in flood mode, no replies will be shown
```

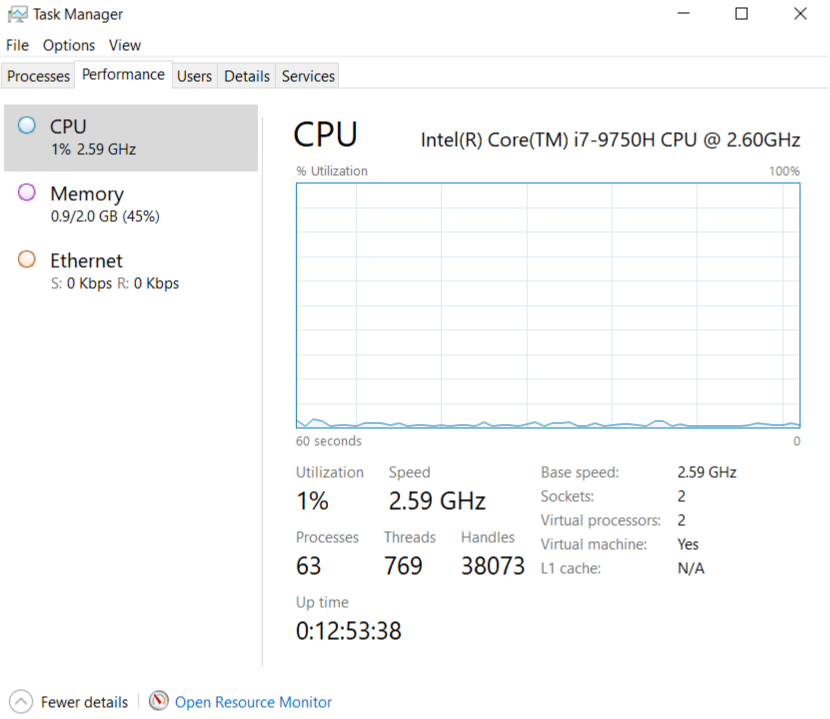
Fig 5.8: Launch the attack

Step 4: Detect a SYN FLOOD attack with Wireshark

SYN flood attacks are easy to detect as a large number of SYN packets are sent.

Before Attack:

Fig 5.9 shows the CPU utilization before the attack which is at a bare minimum of 1% and Fig 5.10 shows the Wireshark output before the attack. We can see in Fig 5.10 that there are no SYN packets received.



CPU
1% 2.59 GHz

Memory
0.9/2.0 GB (45%)

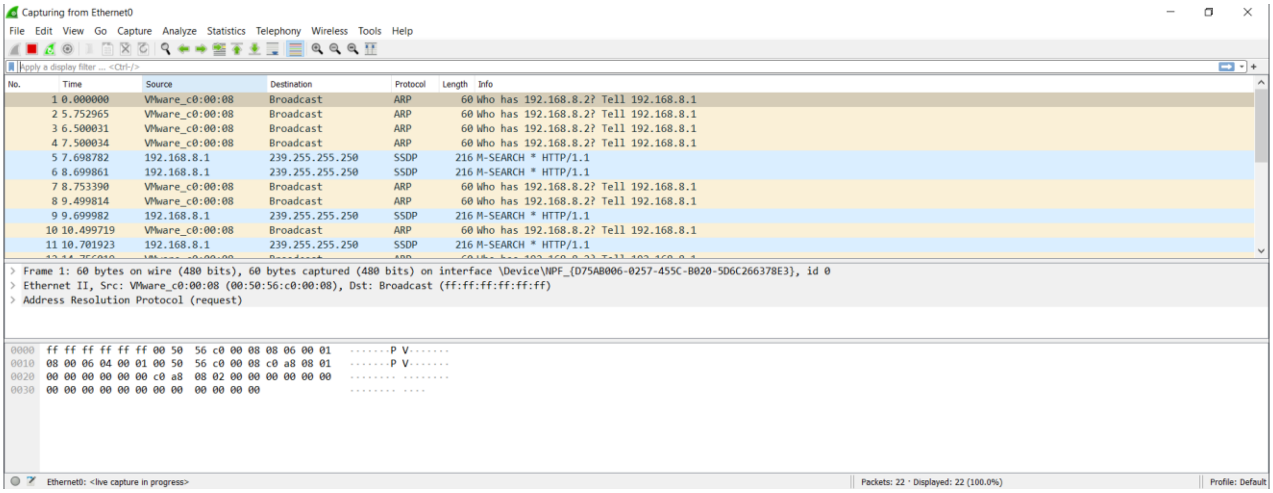
Ethernet
S: 0 Kbps R: 0 Kbps

CPU Intel(R) Core(TM) i7-9750H CPU @ 2.60GHz

Utilization 1%
Speed 2.59 GHz
Processes 63
Threads 769
Handles 38073
Up time 0:12:53:38

Base speed: 2.59 GHz
Sockets: 2
Virtual processors: 2
Virtual machine: Yes
L1 cache: N/A

Fig 5.9: CPU utilization before the attack



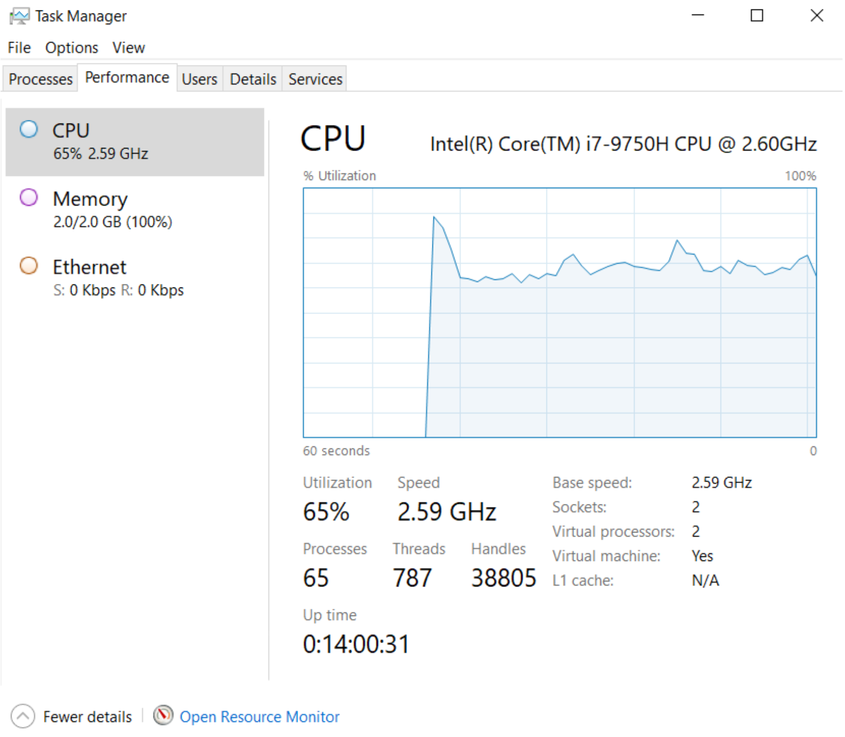
Capturing from Ethernet0
File Edit View Go Capture Analyze Statistics Telephony Wireless Tools Help
Apply a display filter ... <Ctrl-/>

```
No.   Time        Source             Destination       Protocol  Length  Info
    1 0.000000    VMware_c0:00:08    Broadcast         ARP           60  Who has 192.168.8.2? Tell 192.168.8.1
    2 5.752965    VMware_c0:00:08    Broadcast         ARP           60  Who has 192.168.8.2? Tell 192.168.8.1
    3 6.500031    VMware_c0:00:08    Broadcast         ARP           60  Who has 192.168.8.2? Tell 192.168.8.1
    4 7.500034    VMware_c0:00:08    Broadcast         ARP           60  Who has 192.168.8.2? Tell 192.168.8.1
    5 7.698782    192.168.8.1        239.255.255.250   SSDP         216  M-SEARCH * HTTP/1.1
    6 8.699861    192.168.8.1        239.255.255.250   SSDP         216  M-SEARCH * HTTP/1.1
    7 8.753390    VMware_c0:00:08    Broadcast         ARP           60  Who has 192.168.8.2? Tell 192.168.8.1
    8 9.499814    VMware_c0:00:08    Broadcast         ARP           60  Who has 192.168.8.2? Tell 192.168.8.1
    9 9.699982    192.168.8.1        239.255.255.250   SSDP         216  M-SEARCH * HTTP/1.1
   10 10.499719   VMware_c0:00:08    Broadcast         ARP           60  Who has 192.168.8.2? Tell 192.168.8.1
   11 10.701923   192.168.8.1        239.255.255.250   SSDP         216  M-SEARCH * HTTP/1.1

> Frame 1: 60 bytes on wire (480 bits), 60 bytes captured (480 bits) on interface \Device\NPF_{D75AB006-0257-455C-B020-5D6C266378E3}, id 0
> Ethernet II, Src: VMware_c0:00:08 (00:50:56:c0:00:08), Dst: Broadcast (ff:ff:ff:ff:ff:ff)
> Address Resolution Protocol (request)

0000  ff ff ff ff ff ff 00 50  56 c0 00 08 08 06 00 01   ·······P V·······
0010  08 00 06 04 00 01 00 50  56 c0 00 08 c0 a8 08 01   ·······P V·······
0020  00 00 00 00 00 00 c0 a8  08 02 00 00 00 00 00 00   ········ ········
0030  00 00 00 00 00 00 00 00  00 00 00 00               ········ ····

Ethernet0: <live capture in progress>    Packets: 22 · Displayed: 22 (100.0%)    Profile: Default
```

Fig 5.10: Wireshark output before the attack (no SYN packets)

After attack:

We can see in Fig 5.11 that after the attack, the CPU utilization has a massive increase from 1% to 65%, which will gradually increase as the attack continues limiting the CPU resource. Fig 5.12 shows SYN packets that are flooded.



CPU
65% 2.59 GHz

Memory
2.0/2.0 GB (100%)

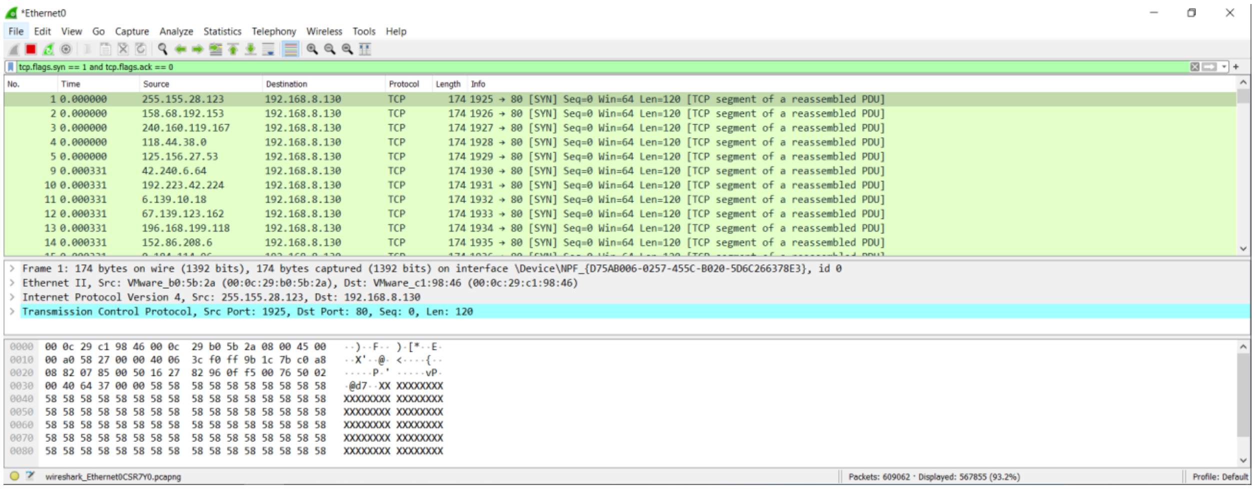
Ethernet
S: 0 Kbps R: 0 Kbps

CPU Intel(R) Core(TM) i7-9750H CPU @ 2.60GHz

Utilization 65%
Speed 2.59 GHz
Processes 65
Threads 787
Handles 38805
Up time 0:14:00:31

Base speed: 2.59 GHz
Sockets: 2
Virtual processors: 2
Virtual machine: Yes
L1 cache: N/A

Fig 5.11: CPU utilization after the attack


*Ethernet0
File Edit View Go Capture Analyze Statistics Telephony Wireless Tools Help

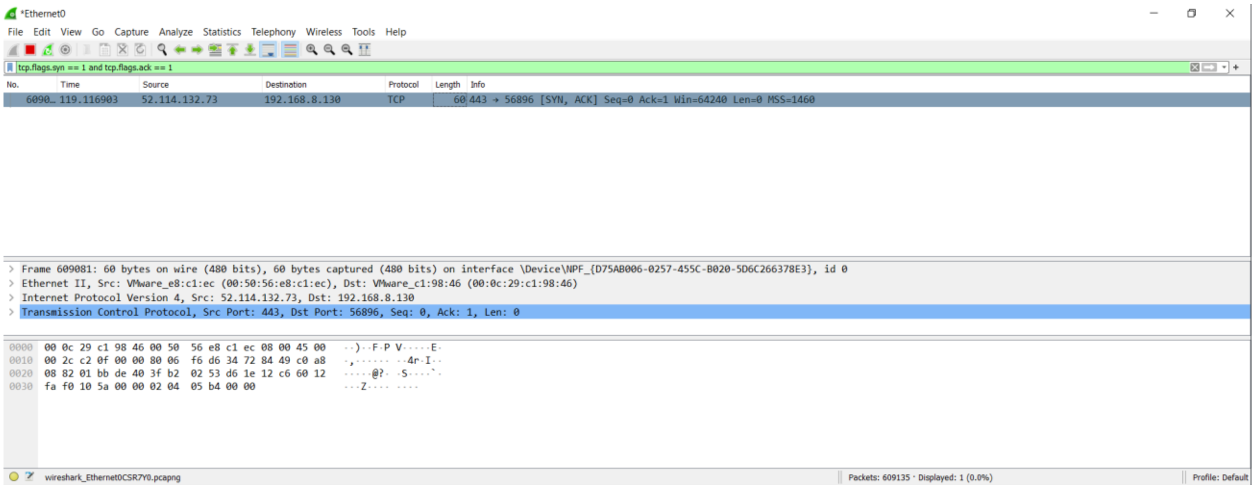
tcp.flags.syn == 1 and tcp.flags.ack == 0

```
No.  Time      Source           Destination    Protocol Length Info
   1 0.000000  255.155.28.123   192.168.8.130  TCP      174 1925 → 80 [SYN] Seq=0 Win=64 Len=120 [TCP segment of a reassembled PDU]
   2 0.000000  158.68.192.153   192.168.8.130  TCP      174 1926 → 80 [SYN] Seq=0 Win=64 Len=120 [TCP segment of a reassembled PDU]
   3 0.000000  240.160.119.167  192.168.8.130  TCP      174 1927 → 80 [SYN] Seq=0 Win=64 Len=120 [TCP segment of a reassembled PDU]
   4 0.000000  118.44.38.0      192.168.8.130  TCP      174 1928 → 80 [SYN] Seq=0 Win=64 Len=120 [TCP segment of a reassembled PDU]
   5 0.000000  125.156.27.53    192.168.8.130  TCP      174 1929 → 80 [SYN] Seq=0 Win=64 Len=120 [TCP segment of a reassembled PDU]
   9 0.000331  42.240.6.64      192.168.8.130  TCP      174 1930 → 80 [SYN] Seq=0 Win=64 Len=120 [TCP segment of a reassembled PDU]
  10 0.000331  192.223.42.224   192.168.8.130  TCP      174 1931 → 80 [SYN] Seq=0 Win=64 Len=120 [TCP segment of a reassembled PDU]
  11 0.000331  6.139.10.18      192.168.8.130  TCP      174 1932 → 80 [SYN] Seq=0 Win=64 Len=120 [TCP segment of a reassembled PDU]
  12 0.000331  67.139.123.162   192.168.8.130  TCP      174 1933 → 80 [SYN] Seq=0 Win=64 Len=120 [TCP segment of a reassembled PDU]
  13 0.000331  196.168.199.118  192.168.8.130  TCP      174 1934 → 80 [SYN] Seq=0 Win=64 Len=120 [TCP segment of a reassembled PDU]
  14 0.000331  152.86.208.6     192.168.8.130  TCP      174 1935 → 80 [SYN] Seq=0 Win=64 Len=120 [TCP segment of a reassembled PDU]

> Frame 1: 174 bytes on wire (1392 bits), 174 bytes captured (1392 bits) on interface \Device\NPF_{D75AB006-0257-455C-B020-5D6C266378E3}, id 0
> Ethernet II, Src: VMware_b0:5b:2a (00:0c:29:b0:5b:2a), Dst: VMware_c1:98:46 (00:0c:29:c1:98:46)
> Internet Protocol Version 4, Src: 255.155.28.123, Dst: 192.168.8.130
> Transmission Control Protocol, Src Port: 1925, Dst Port: 80, Seq: 0, Len: 120

0000  00 0c 29 c1 98 46 00 0c  29 b0 5b 2a 08 00 45 00   ··)··F·· )·[*··E·
0010  00 a0 58 27 00 00 40 06  3c f0 ff 9b 1c 7b c0 a8   ··X'··@· <····{··
0020  08 82 07 85 00 50 16 27  82 96 0f f5 00 76 50 02   ·····P·' ·····vP·
0030  00 40 64 37 00 00 58 58  58 58 58 58 58 58 58 58   ·@d7··XX XXXXXXXX
0040  58 58 58 58 58 58 58 58  58 58 58 58 58 58 58 58   XXXXXXXX XXXXXXXX
0050  58 58 58 58 58 58 58 58  58 58 58 58 58 58 58 58   XXXXXXXX XXXXXXXX
0060  58 58 58 58 58 58 58 58  58 58 58 58 58 58 58 58   XXXXXXXX XXXXXXXX
0070  58 58 58 58 58 58 58 58  58 58 58 58 58 58 58 58   XXXXXXXX XXXXXXXX
0080  58 58 58 58 58 58 58 58  58 58 58 58 58 58 58 58   XXXXXXXX XXXXXXXX

wireshark_Ethernet0CSR7Y0.pcapng    Packets: 609062 · Displayed: 567855 (93.2%)    Profile: Default
```

Fig 5.12: SYN packets in Wireshark without acknowledgment

We can filter for SYN packets without acknowledgement using the filter "**tcp.flags.syn == 1 and tcp.flags.ack == 0**" as shown in Fig 5.12. As we can see, there's a high volume of SYN packets with very little variance in time. Each SYN packet shows it's from a different source IP address with a destination port 80 (HTTP), identical length of 120, and window size (64).

When we filter with **"tcp.flags.syn == 1 and tcp.flags.ack == 1"** we can see that the number of SYN/ACKs is comparatively very small. A sure sign of a TCP SYN attack.



tcp.flags.syn == 1 and tcp.flags.ack == 1

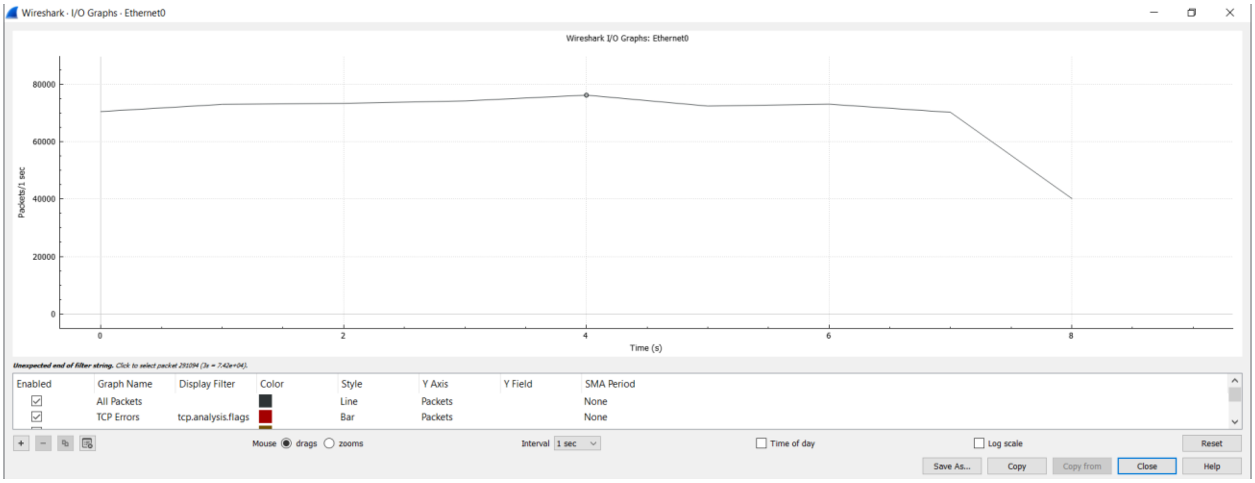
```
No.      Time         Source          Destination     Protocol  Length  Info
6090…    119.116903   52.114.132.73   192.168.8.130   TCP           60  443 → 56896 [SYN, ACK] Seq=0 Ack=1 Win=64240 Len=0 MSS=1460
```

```
> Frame 609081: 60 bytes on wire (480 bits), 60 bytes captured (480 bits) on interface \Device\NPF_{D75AB006-0257-455C-B020-5D6C266378E3}, id 0
> Ethernet II, Src: VMware_e8:c1:ec (00:50:56:e8:c1:ec), Dst: VMware_c1:98:46 (00:0c:29:c1:98:46)
> Internet Protocol Version 4, Src: 52.114.132.73, Dst: 192.168.8.130
> Transmission Control Protocol, Src Port: 443, Dst Port: 56896, Seq: 0, Ack: 1, Len: 0
```

```
0000  00 0c 29 c1 98 46 00 50  56 e8 c1 ec 08 00 45 00   ··)··F·P V·····E·
0010  00 2c c2 0f 00 00 80 06  f6 d6 34 72 84 49 c0 a8   ·,······ ··4r·I··
0020  08 82 01 bb de 40 3f b2  02 53 d6 1e 12 c6 60 12   ·····@?· ·S····`·
0030  fa f0 10 5a 00 00 02 04  05 b4 00 00               ···Z···· ····
```

wireshark_Ethernet0CSR7Y0.pcapng | Packets: 609135 · Displayed: 1 (0.0%) | Profile: Default

Fig 5.13: SYN-ACK packet



Unexpected end of filter string. Click to select packet 291094 (3s = 7.42e+04).

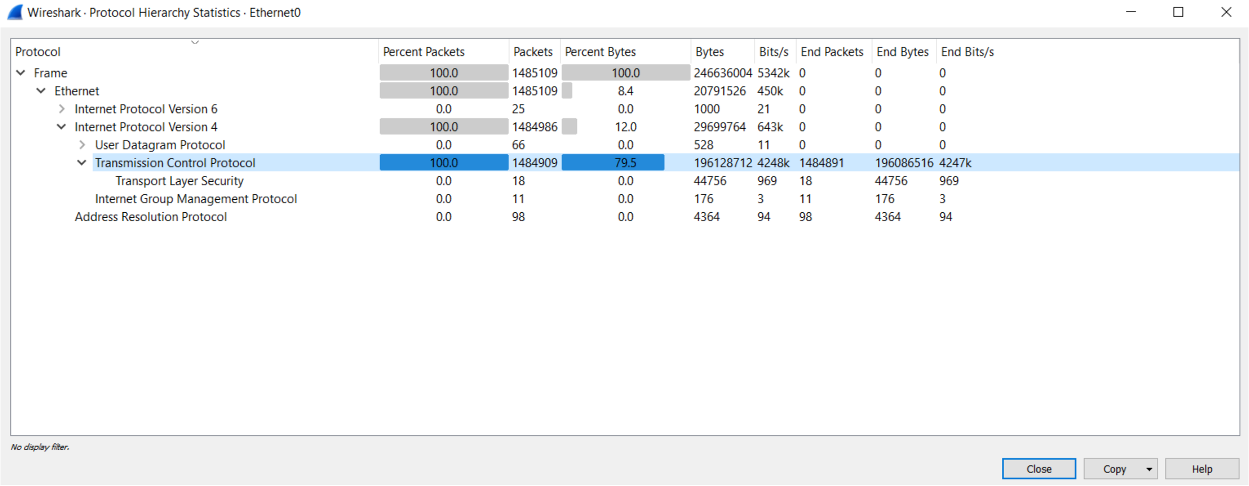
Enabled	Graph Name	Display Filter	Color	Style	Y Axis	Y Field	SMA Period
☑	All Packets		■	Line	Packets		None
☑	TCP Errors	tcp.analysis.flags	🟥	Bar	Packets		None

Mouse ● drags ○ zooms   Interval 1 sec   □ Time of day   □ Log scale

Fig 5.14: I/O Graph

We can also view Wireshark's graphs for a visual representation of the uptick in traffic. The I/O graph can be found via the Statistics>I/O Graph menu. Fig 5.14 shows a massive spike in overall packets from near **0 to up to 80000** packets a second**.**

By removing our filter and opening the protocol hierarchy statistics, we can also see that there has been an unusually high volume of TCP packets as shown in Fig 5.15. [21]



Wireshark · Protocol Hierarchy Statistics · Ethernet0

Protocol	Percent Packets	Packets	Percent Bytes	Bytes	Bits/s	End Packets	End Bytes	End Bits/s
Frame	100.0	1485109	100.0	246636004	5342k	0	0	0
Ethernet	100.0	1485109	8.4	20791526	450k	0	0	0
Internet Protocol Version 6	0.0	25	0.0	1000	21	0	0	0
Internet Protocol Version 4	100.0	1484986	12.0	29699764	643k	0	0	0
User Datagram Protocol	0.0	66	0.0	528	11	0	0	0
Transmission Control Protocol	100.0	1484909	79.5	196128712	4248k	1484891	196086516	4247k
Transport Layer Security	0.0	18	0.0	44756	969	18	44756	969
Internet Group Management Protocol	0.0	11	0.0	176	3	11	176	3
Address Resolution Protocol	0.0	98	0.0	4364	94	98	4364	94

No display filter.

Fig 5.15: Protocol Hierarchy Statistics

5.2.3 Use Case 3: Sink Hole Attack in IoT using RPL protocol in 5G NR

In a sink-hole attack, a compromised node or malicious node advertises incorrect routing information to produce itself as a specific node and receives entire network traffic. After receiving whole network traffic it can either modify the packet information or drop them to make the network complicated.

The Routing protocol (RPL) creates a topology similar to that of a tree, which is known as Directed Acyclic Graph (DAG). Each node within the network has an assigned rank, which increases as the node moves away from the root. The nodes select their parents based on their rank value. A node with a lower rank is more likely to be selected as a parent than a node with a higher rank.

In this case, the malicious node falsely advertises a low-rank value, causing the neighboring nodes to select it as the parent. Once it is selected as the parent, the malicious node ends up getting all the traffic.

Implementation in RPL (for 1 sink):

- In RPL the transmitter broadcasts the DIO (object of information of the DAG) during DODAG (tree) formation.
- The receiver on receiving the DIO from the transmitter updates its parent list, sibling list, rank and sends a DAO (object of an update to the destination) message with route information.
- Malicious node upon receiving the DIO message does not update the rank, instead, it always advertises a fake lower rank.
- The other node on listening to the malicious node DIO message updates their rank according to the fake rank.

- After the formation of DODAG, if the node that is transmitting the packet has a malicious node as the preferred parent, transmits the packet to it but the malicious node instead of transmitting the packet to its parent simply drops the packet resulting in zero throughputs. [22]

Create a file called “malicious.c” with the following functions to do the above-mentioned tasks:

Step 1: Detect if the device/node is malicious or not

Create a function called MaliciousNode(), which can be used to identify whether a current device is malicious or not, to establish malicious behavior. We can set any device and any number of devices as malicious in the network. Also, we can set Device IDs for the malicious nodes.

Step 2: Add fake rank to the malicious node

Create a function called MaliciousRank(), which can be used to assign a fake rank to the malicious node.

Step 3: Drop the packet if it’s a malicious node

Create a function called MaliciousProcessSourceRouteOption(), which is invoked whenever the data packet is received and the current device is found to be malicious. It is also used to drop the received packets if the device is malicious, instead of forwarding the packet to the next hop. This dropping of the packet upon arrival is also known as the Black Hole attack. [22]

By following all the above steps we can create a malicious node in the NetSim simulation software. Now, let’s create different scenarios with this aspect.

Fig 5.16 depicts a network scenario in a physical aspect, which comprises few sensor nodes communicating to a server via a gateway and a router. The same diagram can be depicted in terms of nodes to have a better understanding as shown in Fig 5.17 for Case1 and Fig 5.18 for Case2.

Case 1: Without Malicious Node

Fig 5.17 depicts the DODAG formation in the network without any malicious node.

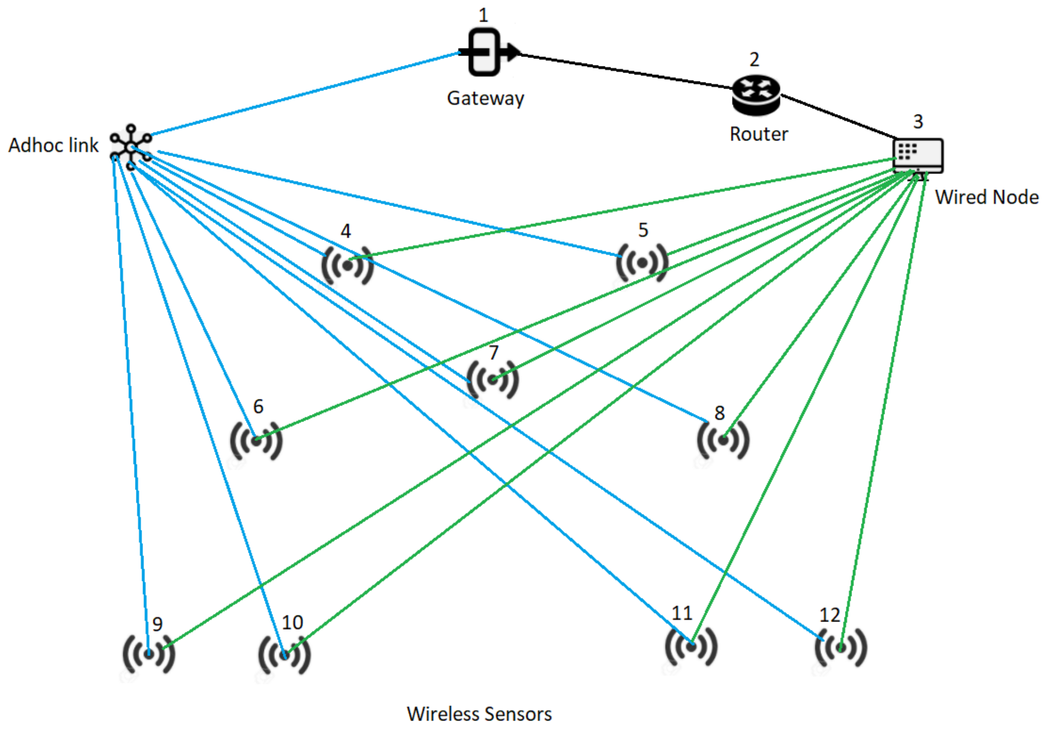


Fig 5.16: Physical representation of the IoT network

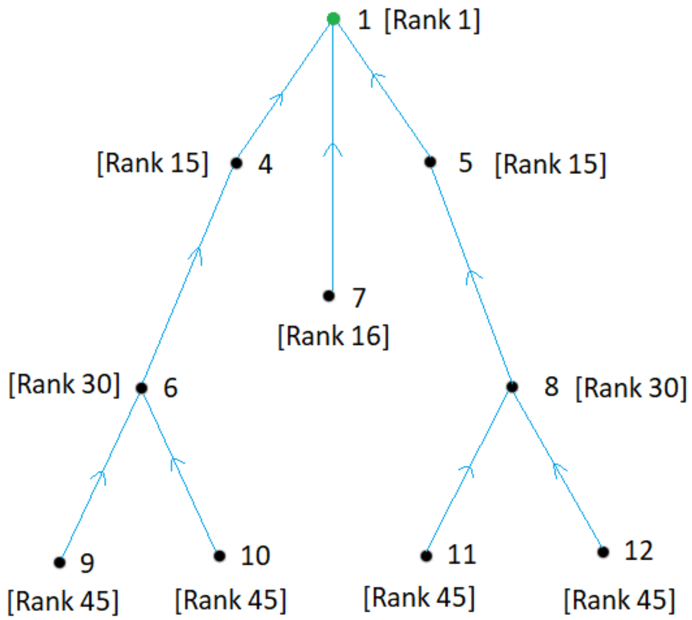


Fig 5.17: Case 1 – Without Malicious Node

Case 2: With Malicious Node

The DODAG formation in the network with malicious node 7 is depicted in Fig 5.18. It's worth noting that all traffic has been redirected to node 7. This is called a sink-hole attack. Now, node 7 can determine if this traffic has to be dropped or modified.

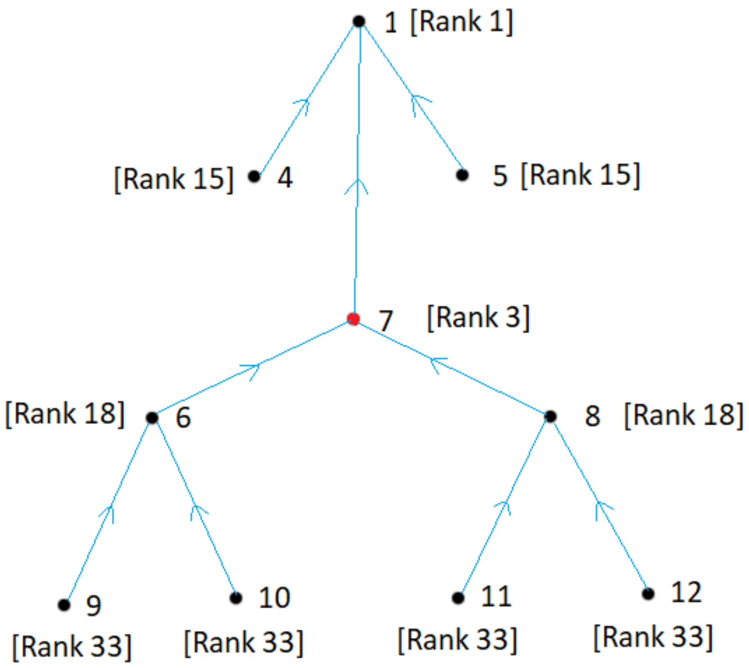


Fig 5.18: Case 2 – With Malicious Node

A similar can take place in a 5G network, this is called Rogue Base Station. In recent years, the research interest in routing-related attacks has been increasing, For example, the Internet of things as an integral component of the 5th generation (5G) communications is vulnerable to routing attacks. It needs to be protected from routing attacks. Sinkhole and selective forwarding are the most common and destructive attacks in infrastructure-based networks. With the development of SDN technology, a network needs to be able to spread quickly new code to every node in the network and a secure manner. [23]

5.2.4 Use Case 4: MITM attack - IMSI Catcher in 5G Network

An IMSI Catcher is a pesky piece of technology that can be used to detect and locate all active cell phones in a given area.

The IMSI Catcher accomplishes this by impersonating a cell phone tower, fooling your phone into linking to it, and then exposing your personal information without your consent.

IMSI Catchers are indiscriminate monitoring instruments that may be used to observe who attends a political rally or a sporting gathering such as a football game. They can also be used to listen in on your calls and edit your posts without you even noticing it.

How does IMSI Catcher work?

IMSI Catchers are machines that mimic mobile towers and trick a target's computer into connecting to them, then relay the contact to a network carrier's real cell tower. Calls, text messages, internet traffic, and other communications from the target travel via the IMSI Catcher, which then reads messages, listens to calls, and so on. Around the same moment, the victim may have no clue what is going on and it will seem to be normal. In the security sector, this is known as a Man-in-the-Middle (MITM) attack.

This is possible due to a weakness in the GSM protocol. To have the fastest commutation, cell phones are continuously searching for the mobile tower with the highest signal. This is usually the one that is closest to you. At the same time, when a device connects to a cell tower, it authenticates to it via an IMSI number. The tower, on the other hand, is not necessary to authenticate back. This is why your phone can connect to a device that functions like a cell tower if one is put near it and give away its IMSI. [24]

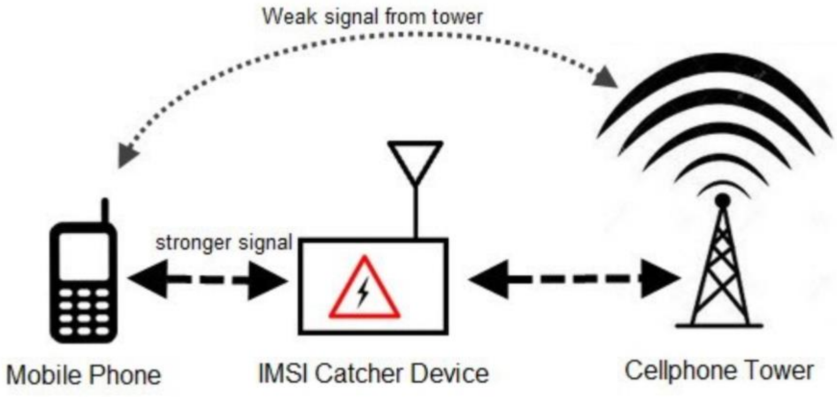


Fig 5.19: Working of IMSI Catcher [24]

How does this affect 5G?

In telecommunications systems, network operators assign a unique identifier to each SIM card, known as an International Mobile Subscriber Identity (IMSI) up to 4G and a Subscription Permanent Identifier (SUPI) up to 5G. Because the authentication between a user and their network provider is based on a shared symmetric key, it can only happen after the user has been identified. Users can be marked, found, and monitored using these permanent identifiers if the IMSI/SUPI values are transmitted in plaintext over the radio access channel.

To avoid this privacy violation, the visited network assigns temporary identifiers to the SIM card called Temporary Mobile Subscriber Identity (TMSI) before 3G networks and Globally Unique Temporary Identifier (GUTI) for 4G and 5G systems. The radio access link is then used to classify these constantly changing transient identifiers for identification purposes. [25]

The solution to IMSI Catchers in 5G:

For decades, IMSI –catching attacks have posed a threat to all generations of mobile telecommunication (2G/3G/4G). This privacy concern seems to have continued as a result of facilitating backward compatibility for legacy purposes. The 3GPP, on the other hand, has agreed to fix this issue, although at the expense of backward compatibility. Unlike previous models, 5G protection standards do not allow plain-text transmissions of the SUPI over the radio interface in the event of a 5G-GUTI malfunction. Instead, a privacy-preserving identifier based on the Elliptic Curve Integrated Encryption Scheme (ECIES) is transmitted, which includes the secret SUPI. SUCI (Subscription Concealed Identifier) is the name assigned to this anonymous SUPI.

Subscription Permanent Identifier (SUPI):

A SUPI is a globally unique Subscription Permanent Identifier (SUPI) assigned to each subscriber in 5G, as specified by the 3GPP specification TS 23.501. In 5G Core, the SUPI value is provisioned in the USIM and UDM/UDR features.

A Valid SUPI can be either of the following:

- For 3GPP RAT, an IMSI (International Mobile Subscriber Identifier) is specified by TS 23.503.
- For non-3GPP RAT, NAI (Network Access Identifier) based user identification as specified in RFC 4282 is used. [25]

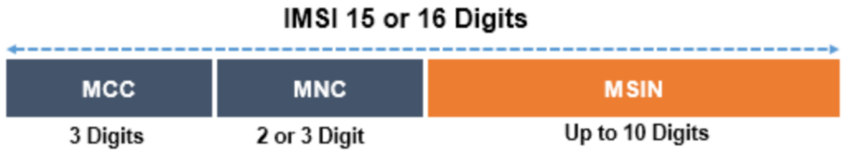


Fig 5.20: Subscription Permanent Identifier (SUPI) [25]

A SUPI is usually 15 decimal digits long. The first three digits are the Mobile Country Code (MCC), and the next two to three are the Mobile Network Code (MNC), which is used to classify the network operator. The remaining (nine or ten) digits are known as the Mobile Subscriber Identification Number (MSIN), and they identify each operator's consumer.

Subscription Concealed Identifier (SUCI):

The Subscription Concealed Identifier (SUCI) is a private identifier that contains the hidden SUPI. The UE creates a SUCI with the public key of the Home Network that was safely provisioned to the USIM during the USIM registration using an ECIES-based security scheme.

The protection scheme only hides the MSIN portion of the SUPI, while the home network identifier, i.e. MCC/MNC, is sent in plain text. The following are the data fields that make up the SUCI:

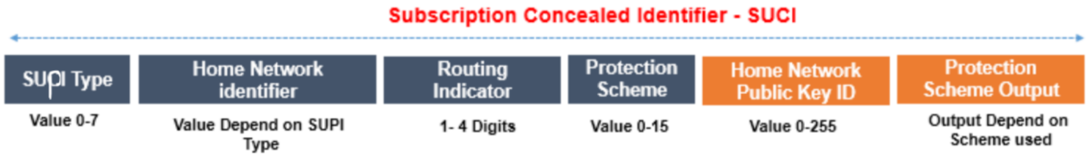


Fig 5.21: Subscription Concealed Identifier (SUCI) [25]

- **SUPI Type:** consists of a value between 0 and 7. It determines the type of SUPI hidden in the SUCI. The following values are defined:
  - 0: IMSI
  - 1: Network Access Identifier (NAI)
  - 2–7: extra values for future use

- **Home Network Identifier:** This defines the subscriber's home network. The Home Network Identifier is made up of MCC and MNC while the SUPI Type is an IMSI. When the SUPI Type is NAI, the Home Network Identifier is a string of characters with a variable length that represents a domain name. For example, user@uofa.com.

- **Routing Indicator:** It is consists of 1 to 4 decimal digits assigned by the home network operator and provisioned within the USIM.

- **Protection Scheme Identifier:** It consists of a value in the range of 0 to 15 and represented with 4 bits

null-scheme	0x0
Profile <A>	0x1
Profile <B>	0x2

- **Home Network Public Key Identifier:** It consists of a number between 0 and 255. It is used to describe the key used for SUPI security and represents a public key provided by the HPLMN. If the null-scheme is used, this data field should be set to a value of 0.

- **Protection Scheme Output:** It consists of a variable-length string of characters or hexadecimal digits, depending on the security scheme used. [25]

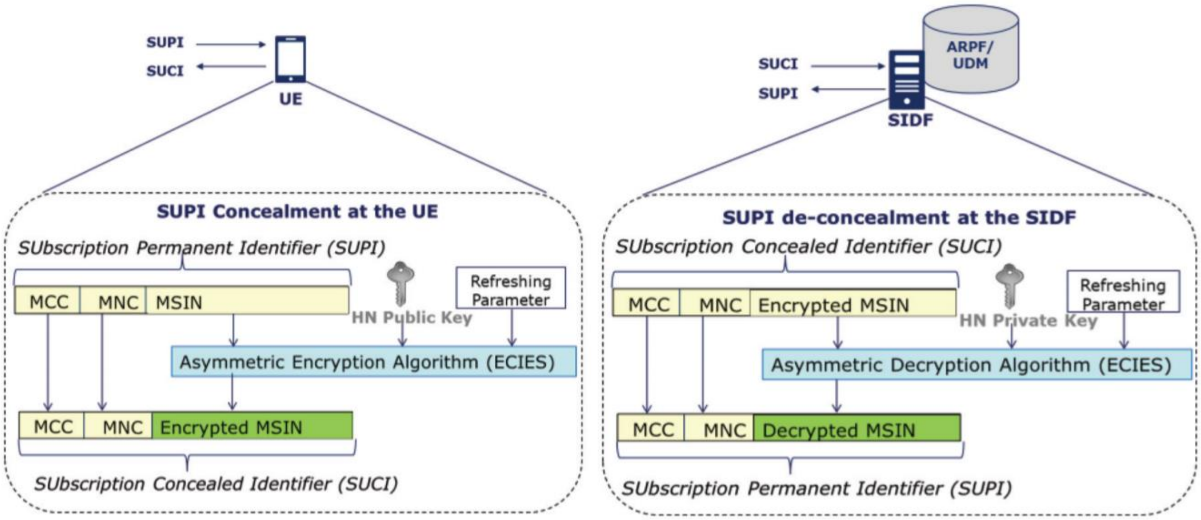

SUPI
SUCI
UE
SUPI Concealment at the UE
SUbscription Permanent Identifier (SUPI)
MCC
MNC
MSIN
HN Public Key
Refreshing Parameter
Asymmetric Encryption Algorithm (ECIES)
MCC
MNC
Encrypted MSIN
SUbscription Concealed Identifier (SUCI)
ARPF/ UDM
SUCI
SUPI
SIDF
SUPI de-concealment at the SIDF
SUbscription Concealed Identifier (SUCI)
MCC
MNC
Encrypted MSIN
HN Private Key
Refreshing Parameter
Asymmetric Decryption Algorithm (ECIES)
MCC
MNC
Decrypted MSIN
SUbscription Permanent Identifier (SUPI)

Fig 5.22: SUPI structure and concealed sensitive information [26]

5G Identity Exchange between UE and Network:

The subscriber authentication system uses the SUCI to identify user equipment on the over-the-air radio interface. Fig 5.23 depicts the UE-Network Identify exchange.

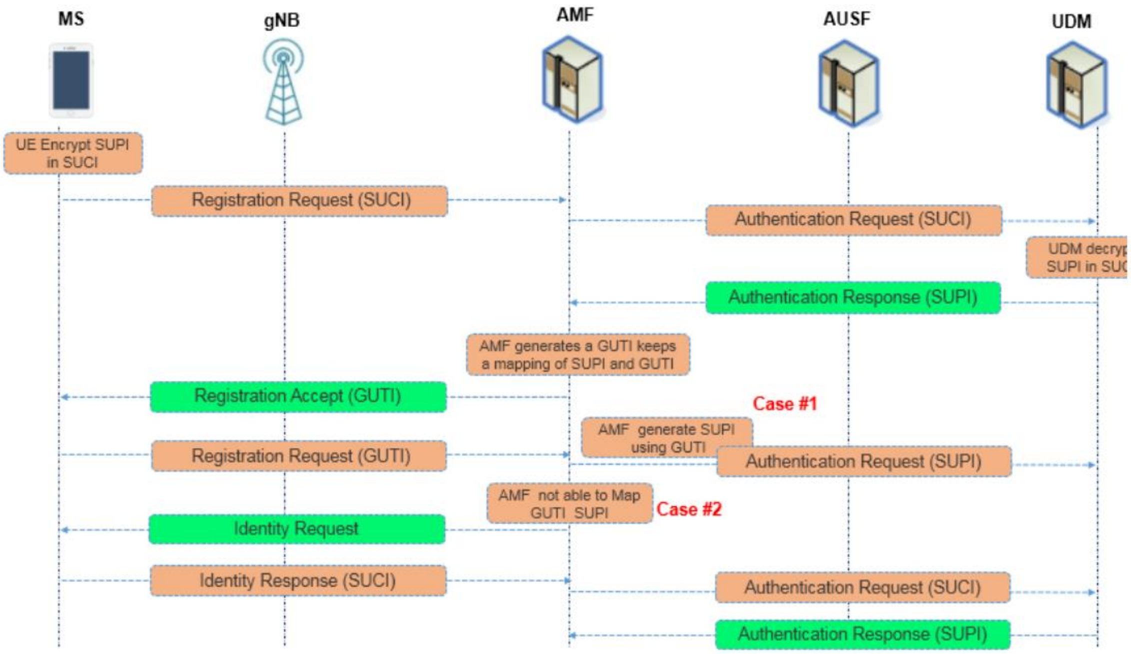

MS
gNB
AMF
AUSF
UDM
UE Encrypt SUPI in SUCI
Registration Request (SUCI)
Authentication Request (SUCI)
UDM decryp SUPI in SUC
Authentication Response (SUPI)
AMF generates a GUTI keeps a mapping of SUPI and GUTI
Registration Accept (GUTI)
Case #1
AMF generate SUPI using GUTI
Registration Request (GUTI)
Authentication Request (SUPI)
AMF not able to Map GUTI SUPI
Case #2
Identity Request
Identity Response (SUCI)
Authentication Request (SUCI)
Authentication Response (SUPI)

Fig 5.23: 5G Identity Exchange between UE and Network [25]

When a UE attempts to register for the first time, it encrypts SUPI into SUCI and sends a SUCI-encrypted Initial Registration Request. To recover the SUPI with Authentication Request, AMF sends this SUCI to AUSF and UDM. AUSF will respond with an Authentication Response containing SUPI data. AMF also creates a GUTI for this SUPI and saves the GUTI to SUPI mapping for future registrations and PDU session requests.

UE sends a validation request with GUTI in the following registration request. There are two potential outcomes now.

- SUPI can be created by AMF using GUTI and SUPI mapping.
- SUPI cannot be produced by AMF.

In the first scenario, AMF uses GUTI to produce SUPI, which can then be used to authenticate with AUSF. In the second scenario, if the UE cannot be identified using GUTI at AMF, AMF will send an identity request to the UE, and the UE will answer with an Identity Response containing the SUCI. [25]

Conclusion

The world is experiencing a dramatic increase in demand for high speed, reliable, and low latency networks and their applications, which makes it all the more important to research the latest technology of cellular networks, that is, the 5G network. It is estimated to support a 1000-fold increase in speed and seamless coverage for at least 100 billion smartphones.

Current research topics focus on the requirements of 5G, its Service Based Architecture (SBA), its enabling technologies, etc. But the security aspects of 5G, lack attention. The increase in security vulnerabilities has started to threaten mobile network assets, such as base stations. With the proliferation of data providers and mobile computer's replacement with the portable computing devices known as smartphones, security risks to mobile networks based around an IP core have increased significantly. When new threats such as espionage and DDoS attacks were added, attacks which once aimed at computers were leveraged and are now replicated for mobile devices. This report aims at covering that gap by analyzing and implementing some of the security attacks on a 5G network in an attempt to understand it better.

A security lifecycle approach to provide new mobile devices and networks, with reliable and robust defense mechanisms and network engagement is extensively discussed in this report. An end-to-end multistage lifecycle strategy will help build a security wall around emerging networks like LTE and 5G. The behavior of such networks can be understood better by analyzing and implementing the use cases. The first use case gives an overall understanding of how the throughput varies along with distance in a 5G network. The second use case depicts a DoS attack that can occur in the transport layer of the network, and the same has been implemented using Kali VM as an attacker machine and a Windows VM as a victim and by studying and analyzing the SYN packets in the victim machine. Similar behavior can be observed in mobile devices or IoT-related devices like sensors. The third use case depicts a sink-hole attack using RPL protocol, which is similar to the rogue base station attack, where all the traffic is diverted to the malicious node and that node decides whether to forward the packet or drop it.

In conclusion, it is undeniable that a mobile device is a significant source of spyware, worms, and ransomware. New flaws are detected daily in mobile OSs which may cause unthinkable harm. 5G networks, which will carry on previous network traditions and deliver new features like cloud, IoT, and ultra-broadband, will have the potential to face these obstacles and eventually fall prey to advanced security attacks like ransomware and Botnets. These should be addressed before we make 5G a standalone network.

Future Scope

Extensive research can be conducted on every aspect of 5G using simulators like Tetcos NetSim and Matlab. Topics like Peak Downlink/Uplink throughput for n78 and n258, where the values of peak throughput for upload and download in different bands can be checked, cell radius for different data rates in n78 and n258, where we can study how the data rates for specific bands varies with respect to the cell radius. We can also analyze the throughput as the User Equipment (UE) moves away from gNB in a continuous manner. We can also increase the number UE for a single gNB and see how the gNB handles the load and find the maximum number of UE a single gNB is capable to handle.

In terms of security, similar attacks like DoS and Sink-Hole can be tested using different protocols. For example, we have showcased the working of sinkhole attack using RPL protocol, instead, we can do the same using DRS protocol (which is an Ad Hoc network protocol) and see how the 5G network responds to it. Another option would be the AODV protocol which is again an Ad Hoc network protocol. Other MITM attacks and Rogue Based Station attacks can also be studied.

At present, we are working on NSA 5G (Option-3) and in the future, the plan or the goal will be to have a 5G network that is SA. So, all the topics mentioned above can be implemented and tested in different deployment options of a 5G network since the results might vary in different environments.

Glossary

A

AF - Application Function

AHP - Analytic Hierarchy Process

AKA - Authentication and Key management

AMF - Access and Mobility Management Function

AMPS - Advanced Mobile Phone Service

APT - Advanced Persistent Threats

ASF - Authentication Server Function

AuC - Authentication Center

B

BSC - Base Station Controller

BSS - Base Station Sub-system

BTS - Base Station Transceiver System

BU - Baseband Unit

BYOD - Bring Your on Device

C

C&C - Centralized Command and Control

CDMA - Code Division Multiple Access

CN - Core Network

CPU - Central Processing Unit

C-RAN - Cloud-based Radio Access Network

C-RNTI - Cell Radio Network Temporary Identifier

D

D2D – Device-to-Device

DAG - Directed Acyclic Graph

DDoS - Distributed Denial of Service

DIO - Object of information of the DAG

DODAG – Tree formation

E

EAP - Enhance Authentication Protocol request/Identity message

ECIES - Elliptic Curve Integrated Encryption Scheme

EDGE - Enhanced Data rate for GSM Evolution

eMBB - Enhanced Mobile Broadband

EPC - Evolved Packet Core

ESN - Electronic Serial Number

ESTI - European telecommunications standards institute

ETACS and NTACS - Total Access Communication Systems

E-UTRAN - Evolved-Universal Terrestrial Radio Access Network

FDMA - Frequency Division Multiple Access

F

FE - Front End

FR - Frequency Range

G

GGSN - Gateway GPRS Support Node

GGSN – Gateway GPRS Support Node

GPRS - GSM Packet Radio Systems

GSM - Global System for Mobile Communications

GUTI - Globally Unique Temporary Identifier

H

HE - Home Environment

HLR - Home Location Register

HSS – Home Subscriber server

I

ICN - information-centric networking

ICT - Information and Communications Technology

IMSI - International Mobile Subscriber Identity

IoT - Internet of Things

ISDN - Integrated Services Digital Networks

ITU - International Telecommunications Union

K

Kc - ciphering key

L

LAN - Local Area Network

LINP - Logically Isolated Network Partitions

M

M2M - Machine-to-Machine

MANO - Management and orchestration

MAT - Multiple Access Techniques

MCC - Mobile Country Code

MIMO - Multiple-Input Multiple-Output

MitM - Man-in-the-Middle

MME - Mobility Management Entity

mMTC - Massive Machine-Type Communications

MNC - Mobile Network Code

MS - Mobile Station

MSC - Mobile Switching Center

MSIN - Mobile Subscriber Identity

MTSO - Mobile Telephone Switching Office

N

NAI - Network Access Identifier

NEF - Network Exposure Function

NFV - Network Function Virtualization

NMT-400 - Nordic Mobile Telephone

NRF - Network Repository Function

NSA - Non-Standalone

NSS - Network Switching Subsystem

NSSF - Network Slice Selection Function

NTT - Nippon Telephone and Telegraph Company

O

OFDMA - Orthogonal Frequency Division Multiple Access

OS - Operating System

OTP - One Time Password

P

PCF - Policy Control Function

PCU - Packet Controller Unit

P-GW - packet data network gateway

PSTN - Public Switched Telephone Network

Q

QoS - Quality of Service

R

RAN - Radio Access Network

RANaaS - RAN-as-a-service

RNC - Radio Network Controller

RPL – Routing Protocol

RRH - Remote Radio Head

S

SA – Standalone

SBA – Service Based Architecture

SCADA - Supervisory Control and Data Acquisition

SC-FDMA - Single Carrier Frequency Division Multiple Access

SDMN - Software-Defined Mobile Networks

SDN – Software Defined Networking

SGSN - Serving GPRS Support Node

S-GW - Cisco Serving Gateway

SIM - Subscriber Identity Module

SMF - Session Management Function

SRES - Signed Response

SSGN – Serving GPRS Support Node

SUPI - Subscription Permanent Identifier

SYN - TCP Synchronize

T

TCP - Transmission Control Protocol

TDMA - Time Division Multiple Access

TMSI - Temporary Mobile Subscriber Identity

U

UDC - User Data Convergence

UDM - Unified Data Management

UDR - Unified Data Repository

UMTS - Universal Mobile Telecommunications System

UPF – User Plane Function

URLLC - ultra-reliable and low latency communications

USIM - User Services Identity Module

UTMS - Universal Mobile Telecommunications Service

UTRAN - UMTS Terrestrial Radio Access Network

V

VLR - Visitor Location Register

W

WAP - Wireless Access Protocol

WCDMA - Wide-band CDMA

Bibliography

[1] M. Liyanage, I. Ahmad, A. B. Abro, A. Gurtov, and M. Ylianttila, *A Comprehensive Guide to 5G Security*. Hoboken, New Jersey: Wiley.

[2] S. Code, E. Hours, and E. Marks, *Wireless communication*, no. 142. 2009.

[3] “www.sciencedirect.com: IMSI (International Mobile Subscriber Identity).” https://www.sciencedirect.com/topics/computer-science/subscriber-identity#:~:text=The IMSI is an E, Identity) (Figure 6.12).

[4] T. Balderas and R. a. Cumplido, “Security Architecture in UMTS Third Generation Cellular Networks,” *Coord. Ciencias Comput. Ina.*, no. Ccc, pp. 1–19, 2004, [Online]. Available: https://www.cablelabs.com/insights/a-comparative-introduction-to-4g-and-5g authentication.

[5] “LTE Network Architecture.” https://yatebts.com/documentation/concepts/lte-concepts/.

[6] “5G Typical Use Cases.” https://www.itu.int/dms_pubrec/itu-r/rec/m/R-REC-M.2083-0-201509-I!!PDF-E.pdf.

[7] “5G SBA,” [Online]. Available: https://www.google.com/search?q=5G+service+based+architecture&sxsrf=ALeKk03lwXCS39YBFhPTHrE23USQcL3V5A:1613280969788&tbm=isch&source=iu&ictx=1&fir=rZRIF31bWEfBnM%252CaZaZ07kRYJ4oyM%252C_&vet=1&usg=AI4_-kQhcTzRhcCF5PXvHlwmQ-zi9rcNqA&sa=X&ved=2ahUKEwiA04H50.

[8] “ETSI - TS 23.501 (Section 6.2),” [Online]. Available: https://www.etsi.org/deliver/etsi_ts/123500_123599/123501/16.06.00_60/ts_123501v160600p.pdf.

[9] “5G Deployment Options,” [Online]. Available: https://blog.3g4g.co.uk/2018/10/5g-network-architecture-options-updated.html.

[10] Z. Ghadialy, “5G for Absolute Beginners,” *Udemy, Inc*, no. April 2020, [Online]. Available: www.udemy.com/course/5g-for-absolute-beginners.

[11] “5G NR (New Radio) Frequency Bands,” [Online]. Available: http://www.techplayon.com/5g-nr-frequency-bands/.

[12] “5G Bands,” [Online]. Available: https://en.wikipedia.org/wiki/List_of_5G_NR_networks.

[13] “www.viavisolutions.com,” [Online]. Available: https://www.viavisolutions.com/en-us/5g-architecture#:~:text=5G Architecture,-The primary goal&text=5G is effectively a dynamic, station proximity or complex infrastructure.

[14] “Secure Beamforming in 5G-Based Cognitive Radio Network,” [Online]. Available: https://www.mdpi.com/2073-8994/11/10/1260/htm.

[15] "ETSI - TS 38.306 (Section 4)," [Online]. Available: https://www.etsi.org/deliver/etsi_ts/138300_138399/138306/15.03.00_60/ts_138306v150300p.pdf.

[16] "5G Numerology," [Online]. Available: https://www.slideshare.net/3G4GLtd/beginners-5g-numerology#:~:text=©3G4G In the context,time%2C CP size%2C etc.

[17] "RB allocation," [Online]. Available: https://www.etsi.org/deliver/etsi_ts/138100_138199/13810101/15.03.00_60/ts_13810101v150300p.pdf.

[18] "5G NR Throughput calculator," [Online]. Available: https://5g-tools.com/5g-nr-throughput-calculator/.

[19] "NetSim Tetcos Default/Example Simulation," [Online]. Available: https://tetcos.com/5g.html.

[20] S. Used, N. Standard, and V. Studio, "Dos Attack in 5G NR," vol. 2, 2019.

[21] "SYN FLOOD attack demo," [Online]. Available: http://www.firewall.cx/general-topics-reviews/network-protocol-analyzers/1224-performing-tcp-syn-flood-attack-and-detecting-it-with-wireshark.html.

[22] S. Recommended *et al.*, "Sink Hole Attack using RPL in IOT," vol. 1, 2019.

[23] "The Design of a Defense Mechanism to Mitigate Sinkhole Attack in Software Defined Wireless Sensor Cognitive Radio Networks," [Online]. Available: https://link.springer.com/article/10.1007/s11277-020-07263-9#:~:text=For example%2C Internet of things,attacks in infrastructure based networks.

[24] "Working of IMSI Catcher," [Online]. Available: https://www.paladion.net/blogs/how-to-build-an-imsi-catcher-to-intercept-gsm-traffic.

[25] "5G Identifiers SUPI and SUCI," [Online]. Available: http://www.techplayon.com/5g-identifiers-supi-and-suci/.

[26] "SUPI structure and concealed sensitive information," [Online]. Available: https://www.riverpublishers.com/journal/journal_articles/RP_Journal_2245-800X_619.pdf