Research and implement solutions to tackle security threats in Mobile Cloud Gaming

Capstone Project MINT 709

by

Gurdev Singh

University of Alberta

Master of Science in Internetworking

Edmonton, Canada

Supervisor

Sandeep Kaur

ABSTRACT

Cloud gaming allows gamers to enjoy high-quality gaming experiences at any time and from any location. Complex game software is hosted on powerful servers in data centres, rendered game scenes are broadcast to players in real time over the Internet, and players interact with the games using lightweight software that runs on a variety of heterogeneous devices.

Cloud gaming enables people to play video games on their device remotely by connecting to the client's server. It is no longer necessary to transport a separate gaming device, making gaming more convenient. Computer games can now be played on devices that are significantly less powerful than traditional gaming consoles. This means that everybody, including those without a strong computer, may enjoy playing games. Cloud gaming is a method of playing video games online from anywhere in the world. You can use your computer, phone, or other device to play them.

The distributed denial of service (DDoS) assault is the most visible and often used danger to the cloud-based gaming industry. The user account and any saved games are kept in the cloud. As a result, without service, it would be impossible to play any game. These assaults are launched directly against the game servers. They can, however, be the target, along with the entire gaming platform, and do significant damage. They may potentially take the entire system offline, for example, denying access to authentications on cloud-based accounts and saved games.

ACKNOWLEDGMENT

I would like to express my gratitude to my primary mentor **Mrs. Sandeep Kaur**, who guided me throughout this project. She offered me the freedom to work on my project while keeping ensuring that I stayed on track and did not stray from my subject's core. Without her wise instruction, my thesis would not have been possible.

I would also like to thank **Mr. Shahnawaz Mir**, my program coordinator, for providing me with such a wonderful opportunity and for allowing me to choose a project of my own choosing.

I wish to extend my special thanks to **Dr. Mike McGregor**, who allowed me to start this project.

I would also like to express my heartfelt gratitude to all the instructors, professors, seniors, classmates, colleagues, and the entire University of Alberta who has assisted me in this study, either directly or indirectly, and have been supportive and cooperative at all times in helping me achieve my goal.

TABLE OF CONTENTS

TABLE OF FIGURES

LIST OF TABLES:

1 Introduction

1.1 Evolution Of Wireless Technologies

It is undeniable that the development of mobile technology has come a long way. Initially, cell phones could scarcely maintain a call, however, current technology now allows us to keep a call connected, stream material, and do much more at the same time.

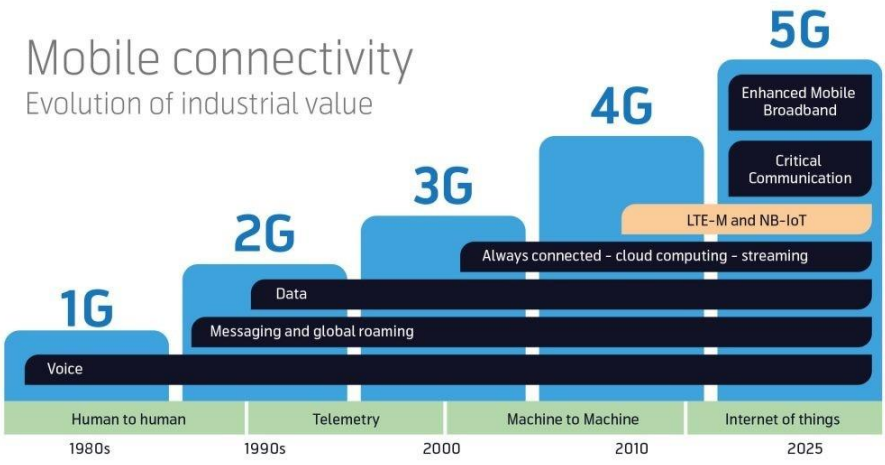

Figure 1: Network Evolution[47]

As depicted in Figure 1, the evolution of cellular networks has progressed through the first generation (1G), which provided public voice service with speeds up to 2.4kbps, to the second generation, which was based on a digital system and featured text messaging, to the third generation (3G), which had a packet-switched advancement and offered data transfer rates of at least 200kbit/s. This led to the fourth generation (4G), which is an IP-based voice communication technology comprising LTE (long term evolution), UMB (ultra-mobile broadband), and IEEE 802.16 (wimax). The fifth generation (5G) is a more advanced version of the current 4G/IMTAdvanced standards, with greater capacity that would enable increased mobile user density, ultra-reliability, and massive communication.

Features	1G	2G	3G	4G	5G
Start/development	1970/1984	1980/1999	1990/2002	2000/2010	2010/2015
Technology	AMPS,NMT,TACS	GSM	WCDMA	LTE, WiMax	MIMO, mm Waves
Frequency	30 KHz	1.8 GHz	1.6-2 GHz	2-8 GHz	3-30 GHz
Bandwidth	2 Kbps	14.4-64 Kbps	2 Mbps	200 Mbps - 1 Gbps	1 Gbps and higher
Access System	FDMA	TDMA/CDMA	CDMA	CDMA	OFDM/BDMA
Core Network	PSTN	PSTN	Packet network	Internet	Internet

Table 1: Comparison of wireless networks

1.2 FIRST GENERATION(1G)–ANALOG SYSTEM

1G refers to the first generation, which was introduced in the 1980s. It was used indefinite until it was replaced by 2G. The primary distinction between 1G and 2G is that 1G is analogue in nature, whereas 2G is digital in nature. The user must enable a transmission button while disabling reception, resulting in a 'push to transmit' system. Because this method does not allow for simultaneous listening and talking, the IMTS (Improved Mobile Telephone System) was introduced in the 1960s. This method consists of two channels, one for sending and the other for receiving; consequently, the 'push to transmit' mechanism was eliminated. IMTS utilises 23 channels ranging from 150Mhz to 450Mhz. Bell Labs invented AMPS (Advanced Mobile Phone Service) and, as a result, first generation (1G) cellular networks were released in 1980. In Japan, AMPS was known as MCS-LI, while in England, it was known as TACS. The main idea behind first-generation cellular networks is that geographical areas are divided into cells. The cell ranges in length from 10 to 25 kilometers, and each cell has its own base station. Because AMPS cells are shorter in length than IMTS cells, they can sustain other neighbouring cells. Also, cells in AMPS demand less power and are less expensive.

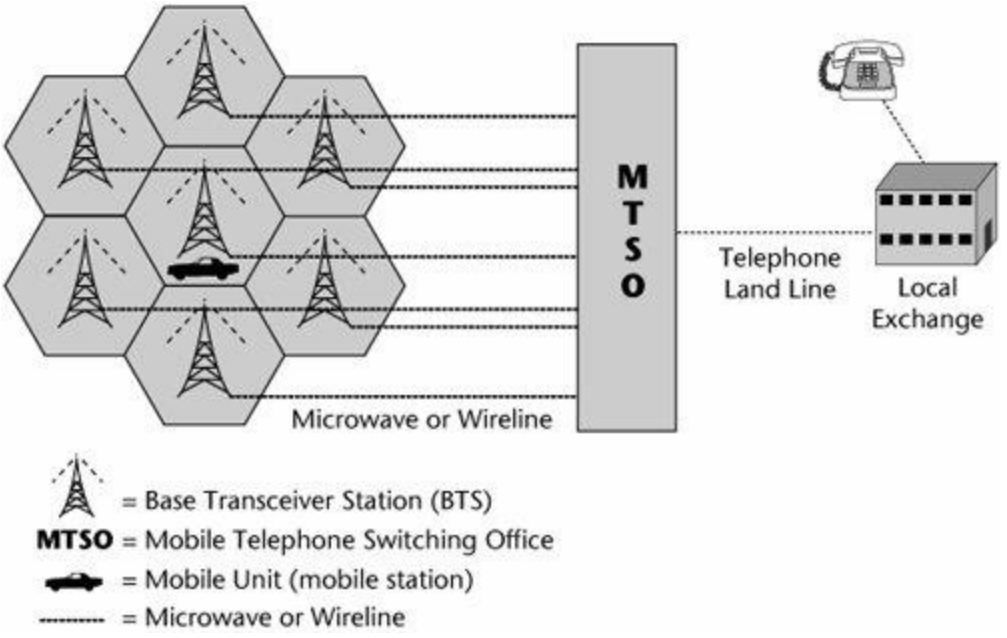

Figure 2: 1st Generation- analog systems[48]

The system had several significant flaws. First, in order to operate efficiently, the analogue signal required a broad transmission spectrum. As a result, a considerable frequency separation between users was required to reduce interference. Second, it only allowed one user per channel, drastically limiting its application. Third, it was prohibitively expensive and inefficient. [8] Lastly, it had various security issues, including the fact that the data was not encrypted and could be easily intercepted. Notwithstanding these limitations, 1G introduced a number of critical characteristics of network systems. For example, it relied on base stations to provide coverage, with neighbouring cells using separate frequencies to avoid interference, and it used automated techniques of coordination to maintain a flawless connection.

1.3 SECOND GENERATION (2G –DIGITAL SYSTEM)

The GSM standard was used to launch the second generation of mobile networks in Finland in 1991. In the end, the calls might be encrypted, increasing safety and providing numerous businesses with new channels of communication. In the meantime, voices could be heard much better, and there was less static and other background noise. [9] It's important to note that 2G was more than just a new technology. It set the stage for the next era of interactive media. In addition to calling, people could now send text, photo, and multimedia messages by clicking buttons on their phones. Digital exchanges like these seemed to offer a seemingly endless number of possibilities. In response to the new options, businesses quickly changed their marketing strategies and increased sales in many different industries.

2G transfer speeds were initially limited to 9.6 kbit/s. 40 kbit/s was a reality at the end of the 2G era. [10] EDGE technology offered speeds of up to one megabit per second. In the 1990s, 2G changed telecommunications, even though the speeds were not as fast as they are now.

1.3.1 Key Benefits of 2G over 1G

The previous generation of the network, which debuted in the early 1990s, was named 1G. [8] Among the advantages of the second-generation network over its predecessor were:

- Data and speech signal digital encryption
- Clearer voice with reduced static
- More concurrent network users
- More efficient spectrum use
- Cheaper and smaller phones
- Roaming enabled
- Texting supported.
- MMS supported
- Digital encryption of data and voice communications
- Lower power consumption (mobile phones become more energy efficient)

1.3.2 GSM

The Global System for Mobile (GSM) was initially developed by the European Telecommunications Standards Institute (ETSI). GSM operates on three frequency band ranges: GSM-900, GSM-1800, and GSM-1900, which have 124, 374, and 299 radio channels, respectively. Time-division multiple access (TDMA) is used in the 900 and 1800 MHz bands of GSM to multiplex up to 8 calls per channel. The frequency range covered by GSM is from 30 to 200 kHz, with one channel dedicated to uplink transmission and another to

downlink transmission. Thus, GSM provides both voice and circuit-switched data speeds of up to 14.4kbps.

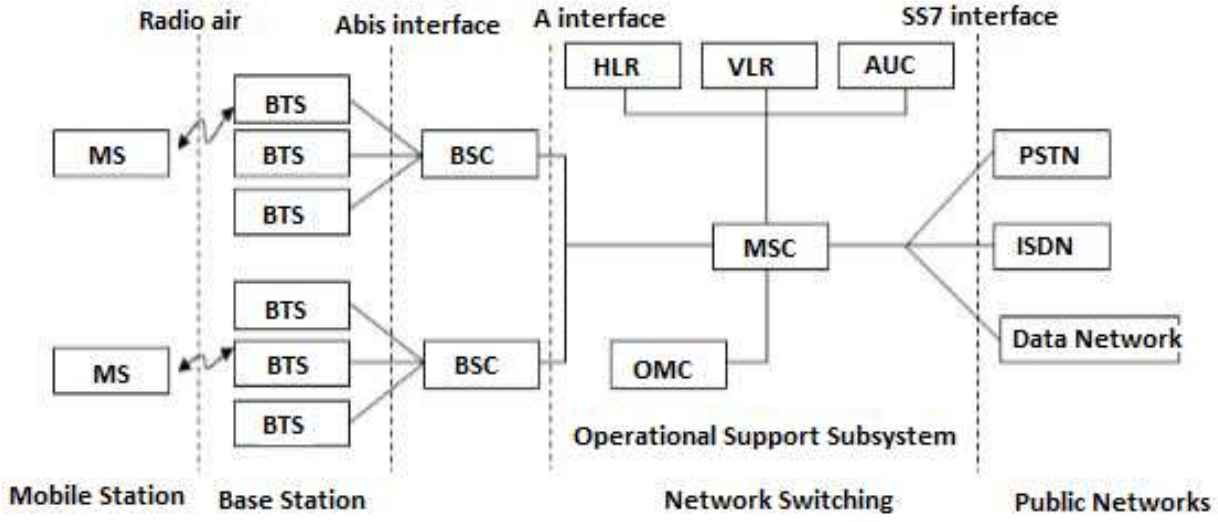

Figure 3: GSM Architecture [49]

1.3.3 2.5G

GPRS, commonly known as 2.5G, is a proposed extension of the existing 2G GSM network that will allow the launch of packet-based services and enhanced data rates. As demand for voice and data expanded, the key concept was to supply IP services without upgrading network infrastructure. It used various coding systems CS1, CS2, CS3, and CS4 to enhance data speeds. The data transmission rate ranged between 50 and 384 kbps. At the core network, serving GPRS Support Node (SGSN) and gateway GPRS Support Node (GGSN) are also introduced.

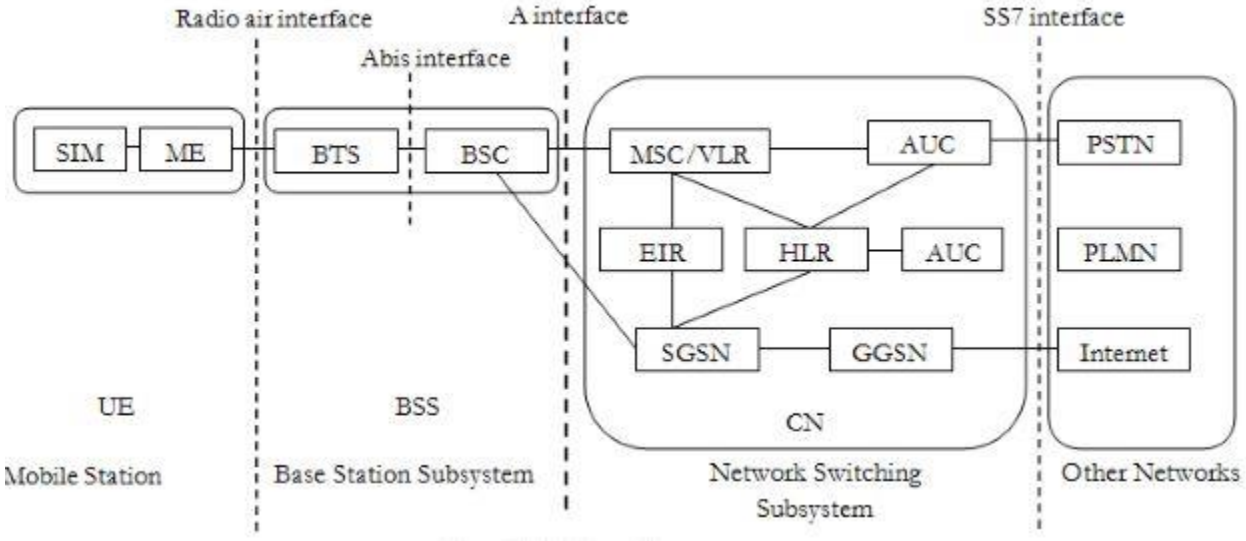

Figure 4: GPRS Architecture [50]

1.3.4 2.75 G

With the launch of Enhanced Data Rates for GSM Evolution (EDGE) networks in 2003, considerable improvements were made to existing GSM systems. IMT Single Carrier is another name for a backward-compatible digital mobile phone technology (IMT-SC). To improve data speeds, the system utilised nine separate modulation and coding methods (MCS- 8PSK). As a result, the downlink data rate is increased to 384 kbps and the uplink data rate is increased to 60 kbps.

Features

- Digital data services introduced (SMS, email)
- Higher transmission rates
- Backward compatible mobile technology 1.3.5

Disadvantages

- Weak signals in less populated areas
- Unable to handle complex data (Video)

1.4 Third Generation (3G)

Because the data rate is high in the third generation, video calling is feasible, which was not possible with Edge [10]. High volume data transfer was achievable in Edge, but when the packet travelled over the air, it behaved like a circuit switch, reducing connection efficiency. The highest possible transmission speeds have now been raised to 8 Mbps. Quality of Service (QoS) needs differ based on the type of service being used; when sending speech, slower data rates are allocated, but faster data rates are allocated for video chats. The 3rd Generation Partnership Project has established a mobile system that is compatible with the IMT-2000 Standard. [3] [4]

UMTS is Europe's 3G wireless Standards which was ETSI (European Telecommunication Standard Institute) driven.

Universal mobile telecommunication system includes Wideband CDMA and a combination of CDMA & TDMA. CDMA is a mechanism in which each user is given a unique code and the entire available bandwidth is used after that. WCDMA uses a wide band of frequencies due to which more users can be accommodated as compared to CDMA.[4]

The only downside of CDMA-2000 is that it is incompatible with W-CDMA since it use different chip speeds and a different multicarrier method. In the third generation, packet switching is used to transmit data while voice communications are handled through circuit switching. The channel's bandwidth is limited to 5MHz. The 3G technology follows a layered architecture strategy, which ensures that voice and data services are utilized efficiently. The layered method assists network operators in rolling out new features because it is standardised with open interfaces, and improved spectral efficiency can be obtained.

1.4.1 3G UMTS Core Network Architecture

The core network design of 3G UMTS was essentially an evolution of the core network architecture of GSM, but with additional features incorporated to enable the enhanced capabilities needed for UMTS. To account for the different ways data could be transported, the UMTS core network was divided into two distinct sections:

- Circuit switched elements: These entities typically belong to the GSM network and carry data in a circuit-switched manner, which means that a dedicated channel is established for the entire duration of the call.
- Packet switching elements: The main purpose of these network entities is to facilitate the transmission of packet data. This approach significantly increases network utilization since capacity can be shared, and data is transported in packets that are directed to their respective destinations.

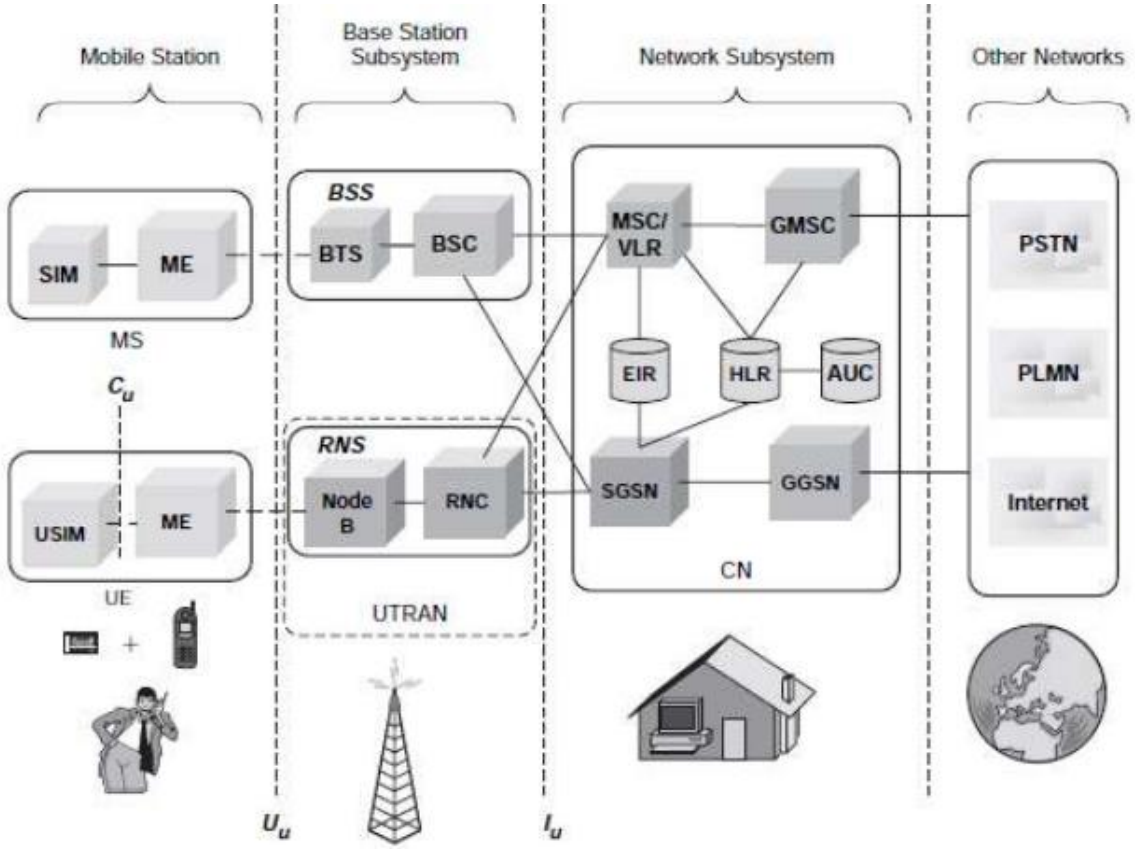

Figure 5: UMTS Architecture[50]

Certain network elements, particularly those related to registration, were shared by both domains and worked in the same manner as GSM. [3]

Elements with circuit switching

The UMTS core network architecture's circuit switching elements featured the following network entities:

- Mobile switching centre (MSC): The functionality of this entity was similar to that of GSM, and it oversaw the circuit-switched calls that were currently active.
- Gateway MSC (GMSC): This served as the interface to the outside networks.

Packet switched elements

The 3G UMTS core network architecture included the following entities for packet switching: SGSN (Serving GPRS Support Node):

SGSN (Serving GPRS Support Node): As the name suggests, this entity was initially introduced with the introduction of GPRS and its usage was subsequently retained in the UMTS network design. [9] In the UMTS network design, the SGSN played a number of roles.

1. One of the key functions of the SGSN was mobility administration. When a user equipment (UE) connected to the UMTS Core Network's Packet Switched domain, the SGSN generated Mobility Management (MM) information based on the mobile device's current location.
2. Session management: Another important task carried out by the SGSN was session management. It was responsible for managing data sessions and ensuring the proper quality of service. It also managed the PDP (Packet Data Protocol) contexts, which were the channels through which data was transmitted.
3. The SGSN was able to manage only the elements within its network by establishing connections with other network areas, such as the MSC and other circuit-switched areas. It was responsible for coordinating communication and data transfer between these different areas.
4. In addition to its other responsibilities, the SGSN also played a role in billing. It achieved this by monitoring the flow of user data across the GPRS network and generating Call Detail Records (CDRs). These records were then forwarded to the charging entities, such as the Charging Gateway Function (CGF).

Gateway GPRS Support Node (GGSN): The Gateway GPRS Support Node (GGSN) was also introduced in the GPRS network and continued to be used in the UMTS packet switching network. It acted as the central hub for the network and was responsible for managing the inter-working between the UMTS packet switched network and external packet switched networks. As a sophisticated router, the GGSN received data targeted to a specific user and checked whether the user was active before forwarding the data to the SGSN responsible for that user's service.

Shared elements

The following network entities were shared aspects of the 3G UMTS core network architecture:

HLR (Home Location Register): The subscriber information, along with their last known location, was stored in this database. This allowed the UMTS network to direct calls to the correct RNC/Node B. When a user powered on their UE, it connected to the network, which allowed the network to determine the Node B it was communicating with, making it possible to route incoming calls to the appropriate location. [9] [10]

Even when the UE was not active (but switched on), it re-registered on a regular basis to ensure that the network (HLR) was aware of its current or last known location on the network.

Equipment identity register (EIR): Just to clarify, the statement "As previously stated, this number was placed in the equipment and was checked by the network during registration" was not previously stated in this conversation. However, I can confirm that the International Mobile Equipment Identity (IMEI) is a unique identifier assigned to each mobile device, and the EIR (Equipment Identity Register) is responsible for verifying whether a device is allowed to access the network based on its IMEI. If the IMEI is found to be invalid or blacklisted (due to being reported as stolen, for example), the EIR will deny network access to the device. [9][10]

AuC (Authentication Center): The AuC was a password-protected database that held the secret key, which was also stored in the user's USIM card. [9][10]

The 3G UMTS wireless communications technology was the first phase in the transition from a mobile speech network to a data network, which meant that significantly greater data capability was required. The network architecture reflected this. [9][10]

1.4.2 3G NETWORK SECURITY TECHNOLOGIES AND MEASURES

Data Integrity

The features include three areas: integrity algorithm consultations (users and service networks consult the integrity of the used algorithm), integrity key consultations (users and service networks consult the key the integrity of the algorithm used by both sides), data integrity and signalling data certification (the recipient can verify signalling data in the event of an unreliable link). Figure 6 depicts the signalling message's data integrity verification process. [13]

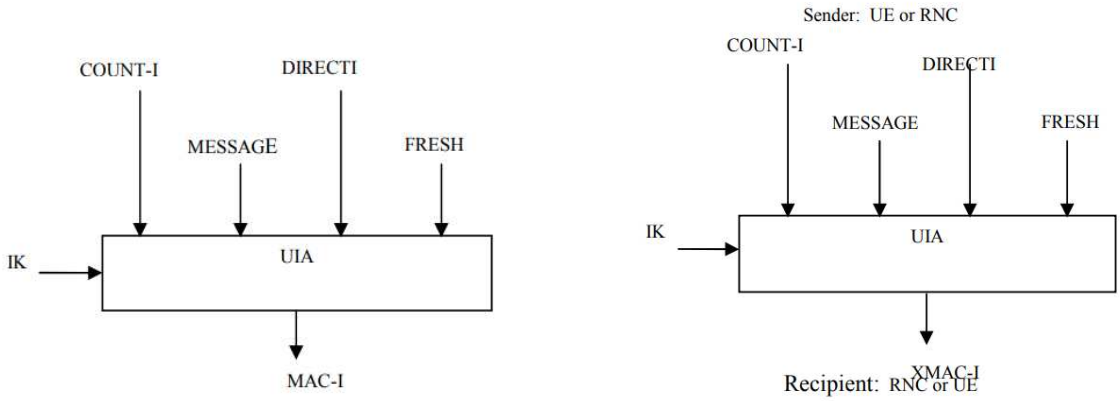

Figure 6: Exporting signaling messages[13]

Whole Network Key Management

Figure 7 shows the entire network-wide key management structure.

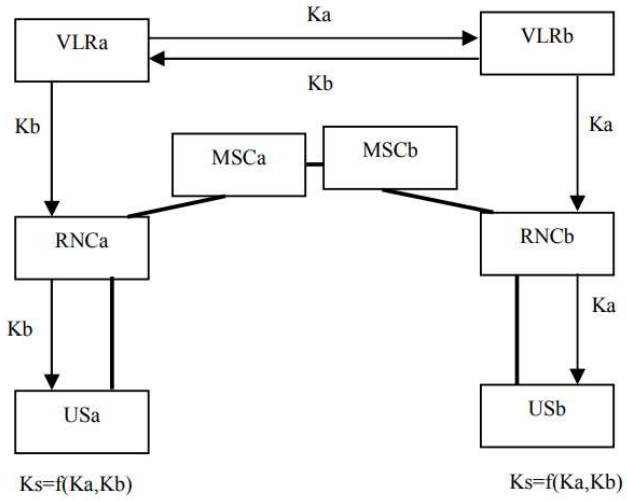

Figure 7: Key management structure[13]

IV. 3G SECURITY LOGIC STRUCTURE

The security structure of a 3G system is a mix of security features and security techniques. Security features are utilised to meet one or more security requirements, while security mechanisms are used to put security features into action. Figure 8 depicts the entire 3G system security architecture. [13]

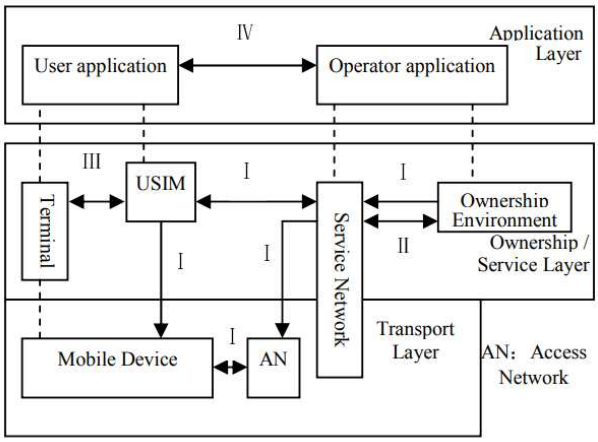

Figure 8: 3G system's security architecture diagram[13]

We identify five security feature categories in the 3G system's security architecture. They involve the transport layer, the ownership/service layer, and the application layer, but they also involve mobile users (including mobile devices and MS), the service network, and the ownership environment. Each security feature group is used to combat specific threats and assaults in order to achieve certain security objectives.

1. Network Access Security: provides security methods for 3G service network access and wireless link protection.
2. Network Domain Security: ensures the security of signaling in the core network and resists attacks on wired networks.
3. User Domain Security: primarily ensures the safety of mobile stations.
4. Application Domain Security: ensures the safe exchange of information between user domain and service provider apps.
5. Visibility of Security Features and Configured Capacity: indicates to users whether the use of services and service providers need to secure a service base.

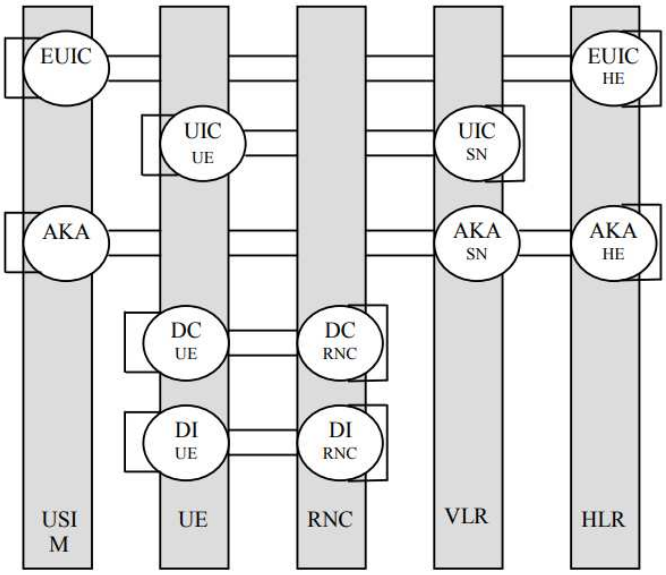

Figure 9: 3G Security architecture[13]

The grey bars in Figure 9 depict network elements that are part of the 3GPP security architecture. Based on their interest, the entities in the network are divided into three groups: [11]

1) The user equipment and smart cards (User Equipment, UE), as well as the user service identification module (USIM), are the user components.
2) components of the service network, such as visitor location registers (VLR) and radio network controllers (RNC).
3) The user identity decryption node (UIDN) and the home location registry (HLR) are components of the ownership network.

The five categories of security measures depicted by the ellipse in Figure 9 are as follows:

1) Enhanced User Identity Confidentiality (EUIC) for smart card authentication using the UIDN within the network identity information ownership.
2) Client Personality Security (UIC).
3) Visit the Home Location Register (VLR), bidirectional authentication and key distribution between the HLR and Home Location Register for USIM cards, and Authentication and Key Agreement (AKA).
4) Data Cryptogram (DC), which encrypts data between radio network controller RNC and user equipment (UE).
5) DI, which is utilized for source and destination authentication, interactive message integrity, and efficacy.

1.5 4G LTE networks

LTE (long term evolution)

LTE, which stands for Long-Term Evolution, is a 4G wireless data transmission global standard that was introduced in 2008. It succeeded the 3rd generation Universal Mobile Telecommunications System (UMTS) or 3G network technology. The high-speed wireless technology of LTE was developed and is

maintained by the 3rd Generation Partnership Project (3GPP) standards organisation, which ensures uniformity in the use of the technology by telecom standard development organisations.

As per the January 2022 report by the GSA, LTE has been a global success, with 6.6 billion subscriptions linking two-thirds of the world's mobile customers.[14]

- 791 telecom operators operate LTE networks in 240 nations and territories worldwide.
- 336 have deployed LTE-Advanced networks.
- 227 have launched VoLTE networks.

Is 4G LTE equivalent to 3G?

No, LTE stands for 4G, and the only observable difference is speed [12]. To put it in perspective, third generation (3G) mobile communications technology was first introduced in 2001 and theoretically promised speeds of 7.2Mbps, but in reality, it could only deliver up to 3Mbps. However, with the advancement of technology, 3G HSPA+ or 3G++ can now deliver up to 42Mbps, and a maximum speed of 6Mbps. On the other hand, 4G LTE (fourth-generation long-term evolution) offers data rates of up to 100Mbps, which is 2.5 times faster than 3G HSPA+ and 15 times faster than 3G.

1.5.1 Difference between LTE and 4G

They're not competing, after all. LTE is the technology that powers 4G. (the fourth generation of mobile communications - an architecture).

In 2022, all 4G phones will use LTE technology. It boosts mobile and broadband data speeds.

The theoretical maximum speed of LTE is 100Mbps. In actuality, it is limited to 15Mbps. Of course, it all depends on your location. [12]

what is LTE-A?

LTE-A stands for Long-Term Evolution-Advanced, which is a standard for mobile communication networks that offers faster data transfer speeds and improved network performance compared to earlier versions of LTE.

LTE-A is an enhancement of the original LTE standard, and it offers a number of advanced features, including:

1) Carrier Aggregation: This allows multiple LTE carriers to be combined to increase the available bandwidth, thereby increasing the data transfer speeds.
2) Enhanced Multiple Input Multiple Output (MIMO): This allows the use of multiple antennas to improve the network's signal quality, resulting in faster data transfer speeds and better coverage.
3) Coordinated Multi-Point (CoMP): This technology improves network performance in areas with high user density by allowing the coordination of signals from multiple base stations to improve the signal quality and reduce interference.
4) HetNet (Heterogeneous Networks): This allows for the integration of different types of wireless networks, such as LTE, Wi-Fi, and small cells, to provide better coverage and network performance.

LTE-A is used by many mobile network operators worldwide, and it is widely used for mobile broadband services, including video streaming, online gaming, and cloud-based applications.

1.5.2 Features of 4G LTE

Some of the key features of 4G LTE (Long-Term Evolution) include:

1) High data transfer speeds: 4G LTE is designed to provide faster data transfer speeds than previous generations of wireless networks. It can deliver download speeds of up to 1 Gbps and upload speeds of up to 100 Mbps, which makes it suitable for high-bandwidth applications such as video streaming and online gaming.
2) Low latency: 4G LTE has low latency, which means that there is minimal delay in data transmission between devices. This makes it suitable for real-time applications such as video conferencing and online gaming.
3) Improved spectral efficiency: 4G LTE uses advanced modulation techniques such as Orthogonal Frequency Division Multiplexing (OFDM) to improve spectral efficiency, which means that it can support more users and devices on the network without sacrificing performance.
4) Quality of Service (QoS): 4G LTE has built-in QoS features that allow network operators to prioritize different types of traffic, such as voice or video, to ensure that they are delivered with the appropriate level of performance.
5) Security: 4G LTE uses advanced encryption algorithms to protect data as it is transmitted over the network, which makes it more secure than previous generations of wireless networks.
6) Backward compatibility: 4G LTE is designed to be backward compatible with previous generations of wireless networks, which means that it can coexist with 3G and 2G networks. This allows users to stay connected even when they are outside the coverage area of a 4G LTE network.

1.5.3 What is VoLTE?

VoLTE stands for Voice over LTE, which is a technology that allows voice calls to be transmitted over 4G LTE networks. In traditional cellular networks, voice calls are transmitted over 2G or 3G networks, while data is transmitted over 4G LTE networks. However, with VoLTE, voice calls are transmitted as data over the 4G LTE network, which provides several benefits, including: [14][15]

But 4G LTE is equally important for IoT devices (connected things).

4G LTE for the IoT (Internet of Things)

Within 4G, which is already widely used to connect industrial-grade IoT devices, there are three primary types. [15]

- The LPWAN, or low-power wide-area network, comes in two flavours: category M (Machine to Machine) (Cat-M or LTE-M) and category NB-IoT. (Cat NB-IoT).
- LTE-1 is the category for mid-range bandwidth (LTE Cat 1).
- LTE Advanced (LTE-A) or LTE Advanced Pro networks are commonly used for high bandwidth applications.

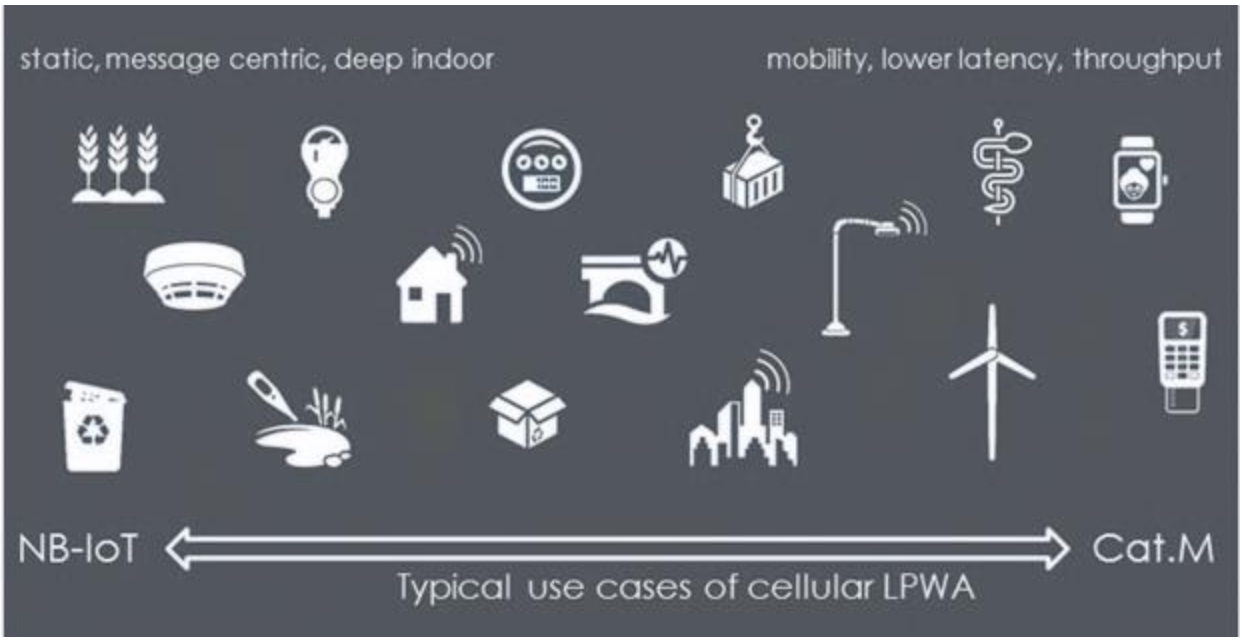

Figure 10: Use cases of cellular LPWA

For linked items, 4G comes in a variety of flavours. Below are some instances of IoT devices and their bandwidth requirements. The appropriate 4G LTE module can then be chosen based on the requirements.

1.5.4 LTE Architecture

LTE's high-level network architecture comprises three significant components: the User Device (UE), UMTS Terrestrial Radio Access Network Evolution (E-UTRAN), and the Evolved Packet Core (EPC). The evolved packet core facilitates communication with packet data networks, such as the internet, a private company network, or the IP multimedia subsystem. The interfaces between the different parts of the system are labelled as Uu, S1, and SGi. [15]

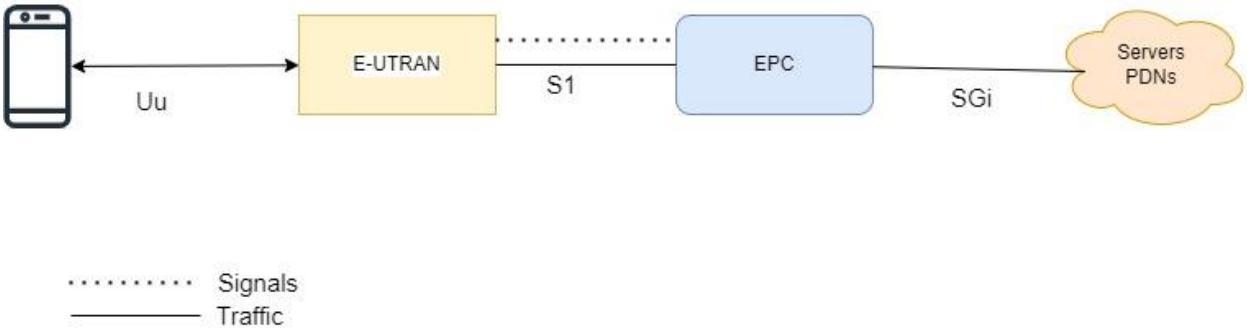

Figure 11: LTE architecture

User Equipment (UE)

The user equipment (UE) in LTE has an internal architecture that is similar to that of GSM and UMTS, also known as mobile equipment (ME). The fundamental modules present in the mobile equipment include the Mobile Termination (MT), Terminal Equipment (TE), and the Universal Integrated Circuit Card (UICC), which is the SIM card used in LTE equipment. The program stored in the UICC is called the Universal

Subscriber Identity Module (USIM), and it stores information such as the user's phone number, home network identity, and security keys, which is similar to the information stored in a 3G SIM card.[15]

The E-UTRAN (The Access network)

Here is an example of the evolution of a UMTS Terrestrial Radio Access Network (E-UTRAN) design.

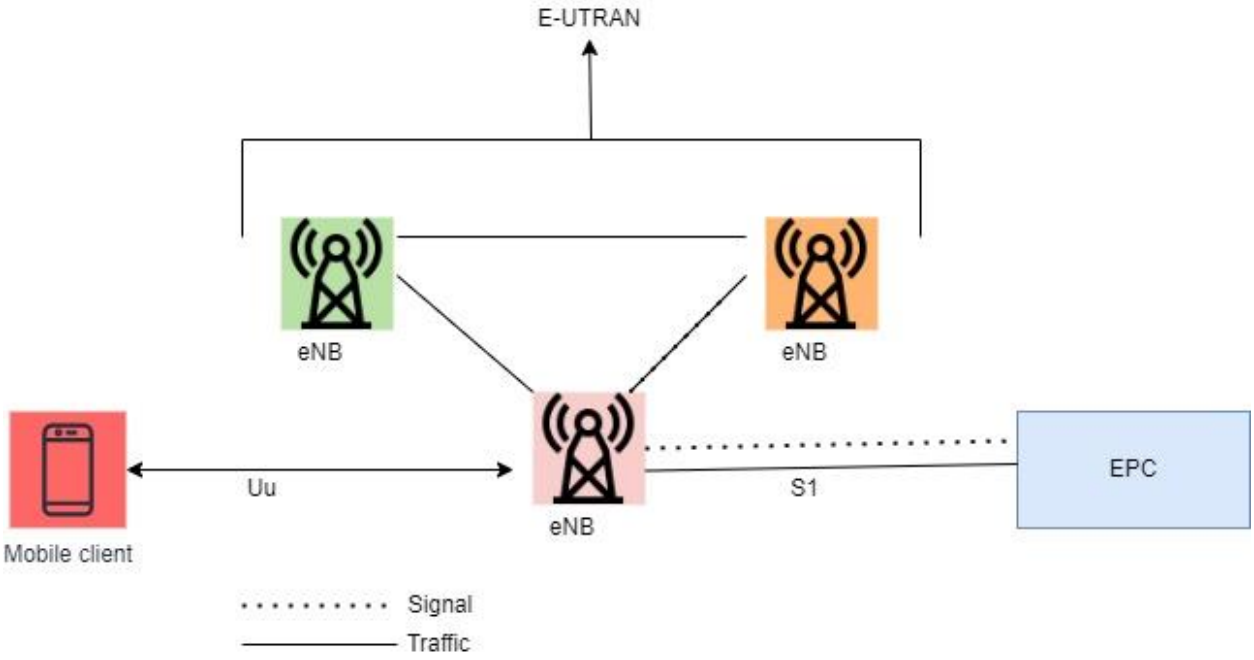

Figure 12: E-UTRAN architecture

To clarify, an Evolved Packet Core (EPC) is a network component in LTE that handles the routing of data packets between the mobile device and the core network. It is responsible for functions such as mobility management, session management, and authentication.

An eNodeB (Evolved Node B) is a base station in the LTE network that communicates with mobile devices over the air interface. It is responsible for tasks such as radio resource management, scheduling, and handover. The eNodeB connects to the EPC via the S1 interface, which is used for control and user plane traffic between the eNodeB and the EPC.

An eNodeB, when connected to an LTE mobile device, can perform two main functions: sending and receiving radio signals to and from all mobile devices using the LTE air interface's analogue and digital signal processing functions, and transmitting low-level handover commands to all of its mobiles, controlling their functioning.

The EPC enables each eNodeB to connect to neighbouring base stations via the S1 and X2 interfaces for signalling and packet forwarding during handover. It can also connect to them via the S1 interface. Additionally, a home eNodeB (HeNB) is a user-owned base station that provides home femtocell coverage. It is part of a closed subscriber group (CSG) and can only be accessed by mobile phones equipped with a USIM that is also part of the CSG.

EPC is an abbreviation for Evolved Packet Core (The core network)

The diagram shows the Evolved Packet Core (EPC) architecture of LTE network, which consists of several key elements, including the Mobility Management Entity (MME), the Serving Gateway (S-GW), and the Packet Data Network Gateway (P-GW). These elements work together to ensure that data packets are delivered securely and efficiently between the LTE user equipment (UE) and external packet data networks such as the internet.

The MME manages the UE's mobility, security, and signalling connections. The Serving Gateway (S-GW) routes user data packets between the UE and the P-GW, while also managing the UE's IP address allocation and mobility. The Packet Data Network Gateway (P-GW) connects the LTE network to external packet data networks, such as the internet or private corporate networks.

Other important elements of the EPC that are not shown in the diagram include the Equipment Identity Registry (EIR), which maintains a database of mobile equipment identities to prevent fraud and unauthorised access, and the Policy Control and Charging Rules Function (PCRF), which enforces quality of service (QoS) policies and manages charging and billing for data services.

The Earthquake and Tsunami Warning System (ETWS) is another element not shown in the diagram. It is a feature that enables mobile operators to send emergency messages to mobile devices during natural disasters or other emergencies, providing critical information to users to keep them safe.

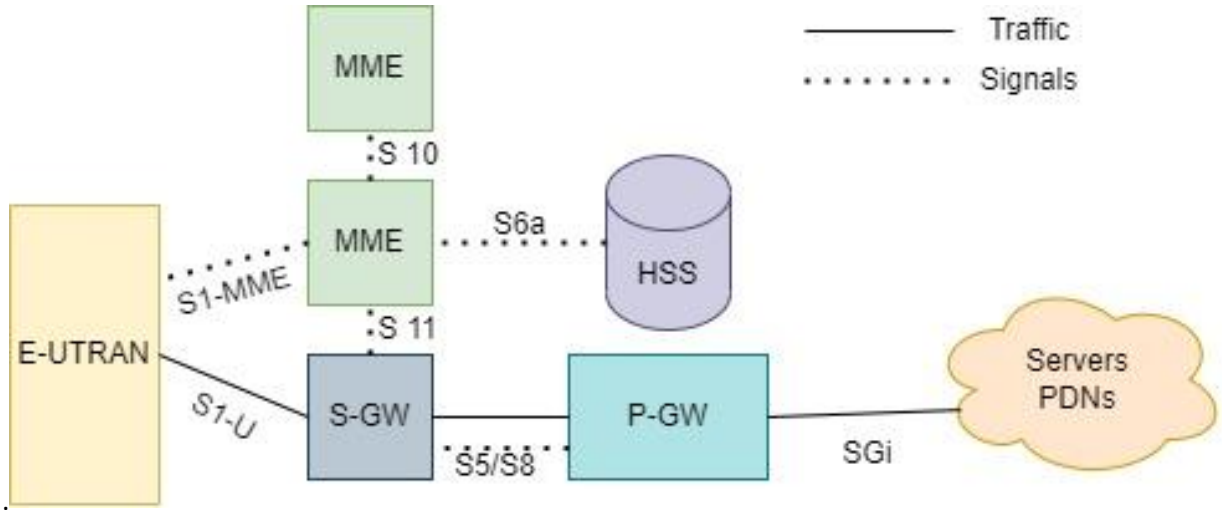

Figure 13: EPC architecture

The architecture of the Evolved Packet Core (EPC) includes several key components, including:

1) Home Subscriber Server (HSS): This is a core database that contains information about all of the network operator's subscribers, including their phone numbers, services subscribed to, and security keys.

2) SGi: This interface is used to connect with external packet data networks, such as the internet. Each packet data network is identified by an Access Point Name (APN). The P-GW is responsible for routing data packets between the mobile device and the appropriate PDN.

3) P-GW (Packet Data Network Gateway): The P-GW is responsible for routing data packets between the mobile device and the appropriate PDN. It also serves as a GPRS Support Node (GGSN) and a Serving GPRS Support Node (SGSN) for UMTS and GSM networks.

4) S-GW (Serving Gateway): The S-GW handles data forwarding between the base station and the P-GW.

5) Policy Control and Charging Rules Function (PCRF): This subsystem manages policy control and charging capabilities in the EPC. The Policy Control Enforcement Function (PCEF) controls policy decision-making and charging functionality in the P-GW.

Other components that may be present in the EPC but are not depicted in the diagram for clarity include the Earthquake and Tsunami Warning System (ETWS) and the Equipment Identity Registry (EIR).

1.6 LTE Security Architecture

The diagram depicts the whole security architecture for LTE. The identified stratums are application, home, serving, and transit, each addressing a suitably isolated category of security issues. [17]

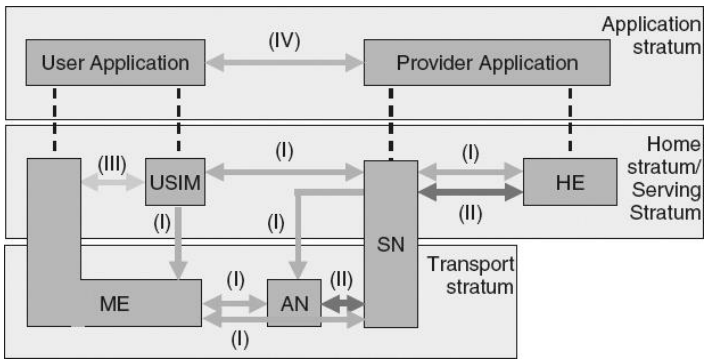

Figure 14: LTE Security architecture [51]

A summary of the LTE Security Architecture. Reprinted with permission from the 2010 3GPP. Any further use is strictly forbidden.

As shown in the diagram, the 3GPP defines five sets of security features:

1) Secure access to the network: a collection of security measures that protect users from attacks on the radio access link and give them safe access to services.
2) Security of the network domain: a set of security features that guard against wireline network attacks and allow nodes to safely transmit user data and signaling data within the Access Network (AN) as well as between the AN and the Serving Network (SN).
3) Security for user domains: the collection of security measures that restrict access to mobile stations.
4) Securing the application domain: the set of security features that make it safe for user and provider domain apps to communicate with each other.
5) Visibility and configuration of security: a set of options that let the user decide if a security feature works or not, as well as whether services should be used and dependent on the security feature.

In what follows, we will go over several of these feature sets in further detail.

Network access security includes several specific features such as confidentiality of user identity, entity authentication, confidentiality of particular agreements and data exchanges, and data integrity. To ensure user identity and location confidentiality, as well as untraceable, short-lived temporary identities are usually assigned. Entity authentication is implemented through authentication at each connection established between the user and the network. General confidentiality pertains to cypher algorithm and key agreements, as well as user and signalling data. Many systems have achieved the characteristics of integrity algorithm and key agreements, along with data integrity and origin authentication of signalling data.

Table 2 shows the security termination points. Reprinted with permission from the 2010 3GPP. Any further use is strictly forbidden.

	Ciphering	Integrity Protection
NAS Signalling	Required and terminated in MME	Required and terminated in MME
U-Plane Data	Required and terminated in eNB	Not Required (NOTE 1)
RRC Signalling (AS)	Required and terminated in eNB	Required and terminated in eNB
MAC Signalling (AS)	Not required	Not required
NOTE 1: Because U-Plane integrity protection is not required, it is not provided between UE and Serving Gateway or for the transit of user plane data between eNB and Serving Gateway through the S1 interface.		

Table 2: Security termination Points

RRC-signaling may be encrypted to prevent UE monitoring during over-the-air RRC interactions, such as measurements or handover. NAS signalling can also be encrypted for privacy. The PDCP layer should ensure the confidentiality of user plane exchanges. This measure, on the other hand, is optional. Meanwhile, integrity must be provided (that is, it is required) for both NAS and RRC signalling. These measures will be discussed further below. Table 14.1 provides the NAS signalling, U-plane, and AS termination points (RRC and MAC signaling) [16] [17]

The security measures related to IP networks fall under the umbrella of "network domain security" and use various methods endorsed by IETF. The details of these measures are explained in 33.210 and 33.10. User domain security involves authentication of users and permission for the USIM-Terminal link, which are essential for user and terminal authentication. Additionally, the USIM Application Toolkit provides security features that allow authentication of applications residing on the USIM. In the case of non-3GPP accesses, a comparable architecture is used, where the access and serving networks are non-3GPP access networks.

1.6.1 Cryptographic Overview

LTE utilizes advanced cryptographic techniques and a distinct key structure that differ from those used in GSM and UMTS. The EPS Encryption Algorithms (EEA) and EPS Integrity Algorithms (EIA) are two cryptographic algorithms used for confidentiality and integrity respectively. EEA1 and EIA1 are based on SNOW 3G algorithms, which are similar to UMTS algorithms. EEA2 and EIA2 are based on the Advanced Encryption Standard (AES), with EEA2 using AES in CTR mode (as a stream cipher) and EIA2 using AES-

CMAC (Cipher-based MAC). EEA3 and EIA3, on the other hand, are based on the Chinese cipher ZUC. While these new algorithms are used in LTE, older algorithms are also implemented in network deployments to ensure compatibility with outdated devices and cellular networks. [17]

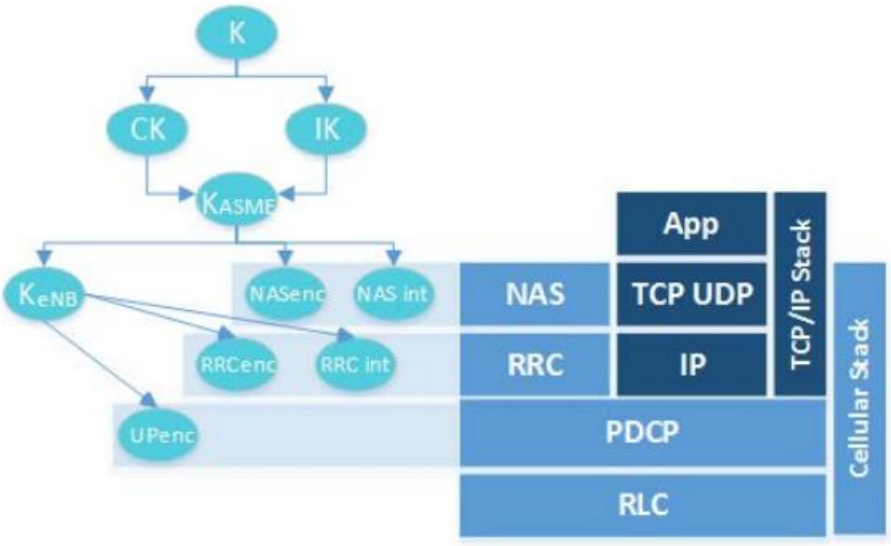

Figure 15: Keys protecting the network stack [52]

LTE keys typically have a length of 256 bits, but some implementations use only the 128 least significant bits. The LTE specification allows for a system-wide upgrade from 128-bit to 256-bit keys. The control and user planes of LTE may use different cryptographic algorithms and key sizes. Figure 6 illustrates the various keys and their application for a particular protocol. [15] [17]

1.6.2 Hardware Security

The UICC, which stands for Universal Integrated Circuit Card, is a newer version of the Subscriber Identity Module (SIM) card and is a critical component of the LTE security framework. It contains the Universal Subscriber Identity Module (USIM) application, which is responsible for carrying out all of the security-related operations required by LTE networks, including authentication and cryptographic functions. The UICC is designed to be a secure and portable storage device that allows users to transfer their cellular service to different devices while retaining their contacts and other data. It consists of a processor, ROM, and RAM, and can run small Java applications for various purposes such as updates and games. Besides, the UICC has network awareness and can potentially support identity services and Near Field Communication (NFC).

1.6.3 UE Authentication

The Authentication and Key Agreement (AKA) protocol is used as the primary authentication mechanism in LTE networks for mobile devices to connect. According to the 3GPP TS 33.401 standard, the usage of AKA is mandatory in LTE. The AKA protocol confirms the knowledge of a secret key (K) by both the UICC and MNO, thereby authenticating the UICC to the network, but not the user or the mobile device itself. The specific steps involved in this procedure are outlined and explained further below:

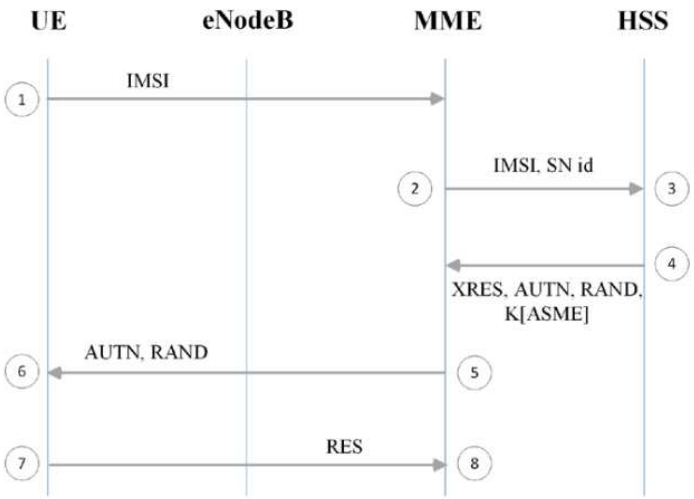

Figure 16: Authentication and key agreement protocol [52]

1.6.4 Air Interface Security

The UE and the eNodeB communicate with each other using the Uu interface, which is a Radio Frequency (RF) connection. Both endpoints convert IP packets into an RF signal, which is transmitted over the air interface. The UE and EPC then demodulate the RF signal into IP packets that can be understood by both parties. The eNodeB routes these packets through the EPC, while the UE uses the IP packets to perform specific functions. However, since these radio waves are transmitted over the air, anyone or anything in the wave path can intercept them, making over-the-air communication not always secure. A diagram below illustrates the network's interception points where this can occur.

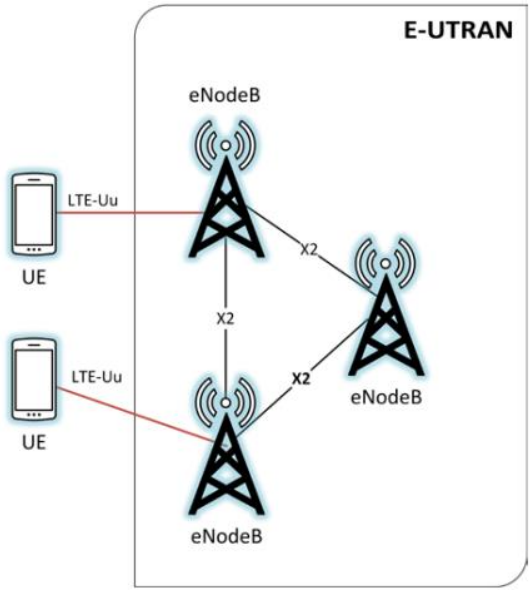

Figure 17: Highlighting the air interface [52]

1.6.5 E-UTRAN Security

The E-UTRAN portion of the LTE network connects a phone to an MNO's core network via the radio access network and associated interfaces. Handover is one of a cellular network's most important tasks. This enables the user to remain connected to their phone while remaining mobile, such as while driving on a highway. Base stations will frequently need to connect with one another through the X2 interface in order to achieve this "mobility." To guarantee the safe transmission of call-related information, 3GPP offers a variety of security measures.

Handovers can take one of two forms: Handovers for X2 and S1 The MME is aware that a handover is about to take place during an S1 handover. The transfer only takes place between eNodeBs via the X2 interface, and the MME is unaware of any X2 handovers. Different kinds of handover have different effects on security. With an S1 handover, the MME can renew the cryptographic parameters that protect the air interface before the connection is cut off. Only after an X2 handover can new keying material be delivered for use in the next handover.

New keys are generated when a session is handed over, partially isolating the new session from the previous one; however, no new master session key—also known as KASME—is established. It should be noted that the source base station and MME control key derivation, in addition to the new eNodeB, are not intended to have knowledge of the keys used in the original eNodeB session. The KeNB is used to create the KeNB*, which is used to safeguard the new session after handover, along with additional cryptographic parameters and the cell ID of the new eNodeB.

1.6.6 Backhaul Security

To keep certain LTE network interfaces secret, 3GPP has defined optional features. According to LTE technical specifications 33.401, confidentiality protection is optional between eNodeBs and the Evolved Packet Core S1 interface. The use of IPsec in accordance with 3GPP TS 33.2104 NDS/IP should be implemented to provide confidentiality on the S1 interface, according to the 3GPP specification. However, the specification also states that confidentiality protection is an operator option if the S1 interface is trusted or physically protected. The 3GPP specification does not provide any additional definitions for the terms "physically protected" and "trusted."

Endpoints that are frequently thousands of kilometers apart are connected using the S1 interface. From the cell tower to the EPC site, any data sent over the LTE network may travel a significant distance. This backhaul connection can be established in a variety of ways for instance, Ethernet, microwave, satellite, underground fiber, and so on. The MNO must implement security controls at each connection route point in order to physically protect the S1 interface. It will be difficult for the MNO to ensure that the S1 interface is physically protected because the cellular MNO is unlikely to own or operate the physical link needed to backhaul LTE network traffic. The operator of the network may employ additional network security mechanisms, such as MPLS VPN and layer 2 VPN, to safeguard and verify the trustworthiness of traffic crossing the S1 interface.

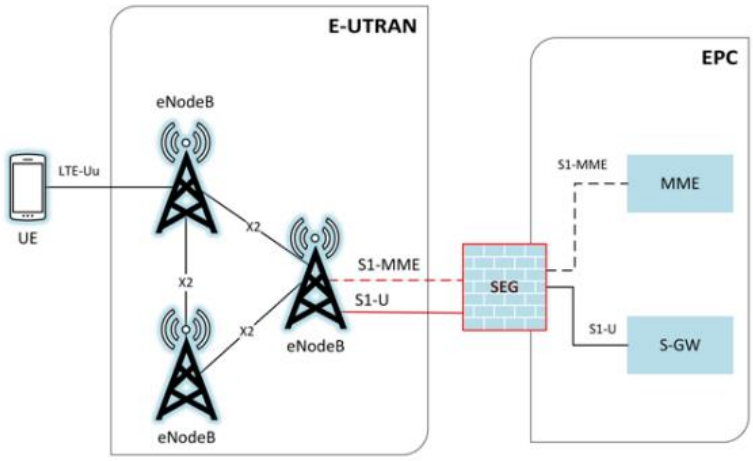

Figure 18: Protecting the S1 interface [52]

Some security risks are introduced by an all-IP-based system that are not present in prior cellular networks. Prior to LTE, an attacker needed specialised hardware to intercept traffic on a cellular network. The transport mechanism used by LTE between the eNodeB and the EPC is all IP-based communications. To intercept communications, you only need basic networking knowledge, a computer, a network cable, and access to a switch port. If secrecy is not enabled on the S1 interface, all intercepted traffic is relayed in clear text.

1.6.7 Core Network Security

In addition to logical and physical segregation, other security measures are also employed in the LTE network. For example, user equipment and eNodeBs authenticate each other using the AKA protocol, as discussed earlier. The LTE network also employs encryption to safeguard user traffic confidentiality, and integrity protection to ensure that traffic is not modified while in transit. The encryption and integrity protection functions are provided by the LTE security algorithms, which are used to generate keys and to encrypt and authenticate data in transit. These algorithms, such as the Advanced Encryption Standard (AES) and Hash-based Message Authentication Code (HMAC), provide strong security for the LTE network.

Moreover, the LTE network supports several security mechanisms, including mutual authentication, confidentiality, integrity protection, and replay protection. These mechanisms provide a high level of security for both user and control plane traffic. The user data is also protected by the use of bearer-level encryption, where a separate encryption key is used for each bearer between the UE and the EPC. The network also supports message authentication, which ensures that the messages are not modified during transmission.

Finally, the LTE network employs several security protocols, such as IPSec, Internet Key Exchange (IKE), and Diameter Security Protocol (Diameter), to provide end-to-end security for different network elements. IPSec is used to encrypt and authenticate traffic between different security domains in the core network. IKE is used to establish and manage IPSec security associations between network elements. Diameter is used for secure communication between network elements, such as the UE and the EPC. These protocols provide a layered approach to security and help to ensure that the LTE network is highly secure.

The 3GPP standards do not include any stringent security criteria or requirements for core network security. SCAS's future development may necessitate the implementation of specialised security safeguards within the different main components.

1.7 Threats to LTE Networks

Many of the risks described below were discovered through academic study, while others may be verified and reported real-world attacks against deployed cellular systems. The higher-level danger categories given in this section can be used as a starting point for organisations' own thorough threat models. [18]

While some of these risks may have an effect on network availability and resilience, others only have an impact on user data integrity and confidentiality. Therefore, the majority of the risks outlined above would only affect a small fraction of the network. Many of the threats outlined below can be implemented with a minimal level of sophistication due to the rising availability of cheap cost LTE hardware and software [21].

1.7.1 General Cybersecurity Threats

LTE infrastructure components (e.g., eNodeB, MME, S-GW) may be built on commodity hardware, firmware, and software, rendering them vulnerable to publicly known software defects in general-purpose operating systems (e.g., FreeBSD and other *nix versions) or other software applications. Although systems may be heavily customised, commodity hardware and well-known operating systems should be identified and understood. This means that these systems must be properly setup and patched on a regular basis to address known vulnerabilities such as those listed in the National Vulnerability Database [28]. The subsections that follow will cover malware risks to specific network components as well as LTE network management. [18] [19]

1.7.2 Malware Attacks on UE's

A mobile device's ability to connect to a cellular network may be prevented by harmful software that infects its operating system, firmware, and installed applications. Malware can directly target the baseband OS and its firmware, potentially modifying crucial network access configuration files or disrupting important processes like parsing base station signalling, ultimately leading to service interruptions. [18] [19]

If certain components are infected with malware, it could allow a carrier's core network infrastructure to monitor network activity, modify crucial communication gateways, and intercept user traffic, such as call or SMS/MMS traffic. While these types of attacks have already been discovered in GSM networks, there have been no reported incidents of such attacks on the backend LTE infrastructure to date. However, a series of attack requests could potentially be used to initiate a Distributed Denial of Service (DDoS) attack against an MME. [18]

1.7.3 Device and Identity Tracking

The IMSI (UICC) and IMEI (handset) are unique identifiers that can be utilized to identify the owner of a mobile device and determine its physical location. Given that individuals generally keep their mobile devices in close proximity, an eavesdropper could potentially determine if a particular individual is present in a specific location by intercepting traffic with a rogue base station in that area. This intrusion into privacy could reveal whether or not a subscriber is located in a specific place. Geolocation data is accessible via signalling channels, which are transmitted over the air interface during handset hookup and authentication. [19]

1.7.4 Downgrade Attacks

By using a rogue base station that transmits at a high-power level, an attacker can coerce a user into downgrading to GSM or UMTS. Currently, there are no significant publicly known vulnerabilities in the cryptographic methods used to protect the secrecy and integrity of the UMTS air interface. However, the 2G GSM cryptographic techniques that ensure the confidentiality and integrity of the air interface have significant flaws, such as A5/1 and A5/2 algorithms. If the air interface cryptographic methods used to safeguard the air interface are compromised, depending on the algorithm negotiated while attaching to the rogue base station, call and data secrecy may be lost. Although GSM is not the main focus of this text, it is included because real-world deployments use GSM networks to link with LTE networks. [19]

1.7.5 Air Interface Eavesdropping

If the operator fails to encrypt user plane LTE traffic on the Uu interface, there is a possibility of a sophisticated eavesdropping attack. The attacker would require specialized equipment to record and store the radio traffic exchanged between the UE and the eNodeB. In addition, the attacker would need software that can identify the exact LTE frequencies and timeslots employed by the UE to demodulate the intercepted traffic into IP packets. [19]

1.7.6 Radio Jamming Attacks

Jamming attacks are a method of interrupting cellular network access by exploiting the radio frequency channel used to transmit and receive data. This type of attack diminishes the signal-to-noise ratio by transmitting static and/or noise at high power levels across a defined frequency band. Such an attack can be carried out in various ways, each requiring a different level of expertise and access to specialized equipment. Smart jamming is a form of jamming that specifically targets certain channels in the LTE spectrum and is deliberately timed to avoid detection. In contrast, dumb jamming is the act of broadcasting noise across a wide range of RF frequencies.

1.7.7 Physical Attacks on Network Infrastructure

A cell site is the actual physical location that contains all the necessary equipment to operate and maintain an eNodeB. Even though these sites are sometimes enclosed by a fence and secured by a physical security system, they can still be breached. If the equipment used to operate the eNodeB is taken offline or destroyed, a denial-of-service attack can be carried out. For instance, the theft of copper is a common occurrence that could lead to a DoS. In addition, if an attacker gains control over the systems operating the eNodeB, more advanced and hard-to-detect attacks are also possible. [18] [19]

1.7.8 Attacks Against K

The security of LTE system is highly dependent on cryptographic keys, which are used to protect different layers of communication. These keys are derived from a pre-shared key known as 'K', which is stored in two places: the USIM on the UICC and the HSS/AuC of the carrier. However, if the USIM manufacturer is responsible for embedding the key into the USIM, they may also have access to 'K'. If a malicious actor gains access to 'K', they would be able to impersonate a subscriber and decrypt their communication. Thus, the security of LTE is highly dependent on the protection of this pre-shared key.

1.7.9 Stealing Service

UICC cards are tiny cards that are designed to be removed from mobile devices. An MNO's service is linked to a user's UICC. This means that a UICC can be stolen from one mobile device and installed in another with the intention of stealing service, including voice and data. Another method of stealing service is when

an insider with access to the HSS or PCRF gives unauthorised network access. This might be an employee who, unbeknownst to the MNO, activates UICCs and sells them for personal gain. [18]

1.8 Disadvantages of 4G Technology

1) Illegally obtaining information from people becomes easier, 4G technology involves the possibility of some interference, though not much, it is vulnerable to assault (jamming frequencies), and the invasion of privacy increases. [20]
2) The consumer is required to purchase a new device to enable 4G, as new frequencies require new components in cell towers. Consumers will pay more for data. Your current equipment may not be 4G network compatible. It has many network bands for various phones. It is costly and difficult to deploy.
3) The 4G LTE network has higher data prices for consumers (expensive), consumers are forced to purchase a new device to support 4G LTE, it consumes a lot of battery when in use, it consumes data very quickly, and your battery becomes hot when used for an extended period of time (like a microwave).
4) A 4G LTE network requires complex hardware; 4G technology is still limited to specific carriers and locations; nevertheless, the number of cities with 4G coverage is growing by the day; it will take time for this network to be available in all major cities around the world.
5) Mobiles compatible with 4G networks are less expensive than previously, but new equipment must be installed to provide these services, which is a time-consuming process for most mobile carriers planning to launch these services. 4G mobile technology is still relatively new, but it will almost certainly have initial glitches and bugs, which could be quite annoying for the user.
6) Because 4G technology employs several antennae and transmitters, you will have considerably lower battery life on your mobile device while using this network. As a result, you will need to use larger mobile devices with more battery capacity to stay online for longer periods of time.
7) Users would be forced to use 3G or Wi-Fi connectivity in areas where 4G mobile network coverage does not yet exist. While this is a problem in and of itself, the worse issue is that they would still have to pay the same amount specified by the 4G network plan. This situation can only be resolved once mobile carriers expand their 4G network coverage to include more regions.
8) 4G technology necessitates costly infrastructure for operation, which is embodied in eNodeB's (Access Points) and primarily EPC's (Gateways or Routers). 4G is optimal for data rates, but not always for voice services. Some of these services are offloaded (delegated) to Wi-Fi or 3G/GSM cellular technologies on your phone.

2 5G Cellular Networks

2.1 5G Cellular Networks

The cause of rapid growth in cellular communication networks is the continual expansion in cellular devices, greater data demand, and the need for improved service quality (QoS). By 2020, it is predicted that 50 billion cellular devices will be using cellular network services, resulting in a significant rise in data traffic. At the moment, available solutions/technologies are incapable of meeting this challenge. In a nutshell, the significant increase in 3D (Device, Data, and Data Rate) necessitates the creation of 5G cellular networks. [21]

The vision of 5G is more than just user requirements. 5G wireless networks can be considered in three major perspectives:

User Centric Network: 5G cellular networks are considered user centric networks since they are designed to provide continuous connectivity and a positive user experience.

Service Provider Centric Network: 5G cellular networks are designed to link to the internet of apps and other services.

5G wireless networks are expected to be energy efficient, less expensive, highly scalable, and secure.

The three broad perspectives of 5G cellular networks defined above define the three major features of future cellular networks:

Connection that is ubiquitous: In the future, many types of gadgets will link ubiquitously to provide an uninterrupted user experience. In fact, ubiquitous connectivity will make the user-centric perspective a reality. [21]

Zero latency: Life-critical systems, real-time applications, and services will be supported by 5G networks with zero delay tolerance. As a result, it is expected that 5G networks would have zero latency, or extremely low latency on the order of 1 millisecond. The zero latency will, in fact, actualize the service-provider-centric view.

High-speed Gigabit connection: The zero-latency property could be achieved by using a high-speed connection for fast data transmission and reception, with users and machines receiving data at rates of Gigabits per second.

5G cellular networks are not simply upgrades to 4G cellular networks; they are made up of new system architectures and concepts that rethink each communication layer. DOCOMO, Alcatel-Lucent, Huawei, GSMA Intelligence Network, Qualcomm, Nokia Siemens, Samsung, 5GPPP, Vodafone, and many other companies are collaborating to improve 5G cellular networks.

2.2 Different modes in 5G:

Another component of 5g that few of us are aware of until they become vendors or telecom service providers (MNOs) is the implementation architecture of 5g Access Network, which is separated into two categories: [22]

- Non-StandAlone

- StandAlone

2.2.1 Non-StandAlone (NSA) mode:

With the existing 4G Core or EPC, NSA includes a brand-new RAN that is deployed alongside the 4G or LTE radio. On the other hand, 5G SA includes a brand-new radio as well as the 5G Core (5GC), which has a cloud-native architecture (CNA) that is completely virtualized and allows for new ways to design, deploy, and manage services. Faster performance than is required by a 5G network is possible thanks to 5GC's high throughput. Its virtualized service-based architecture (SBA) makes it possible to use edge computing to deploy all 5G software network operations. [22] [23]

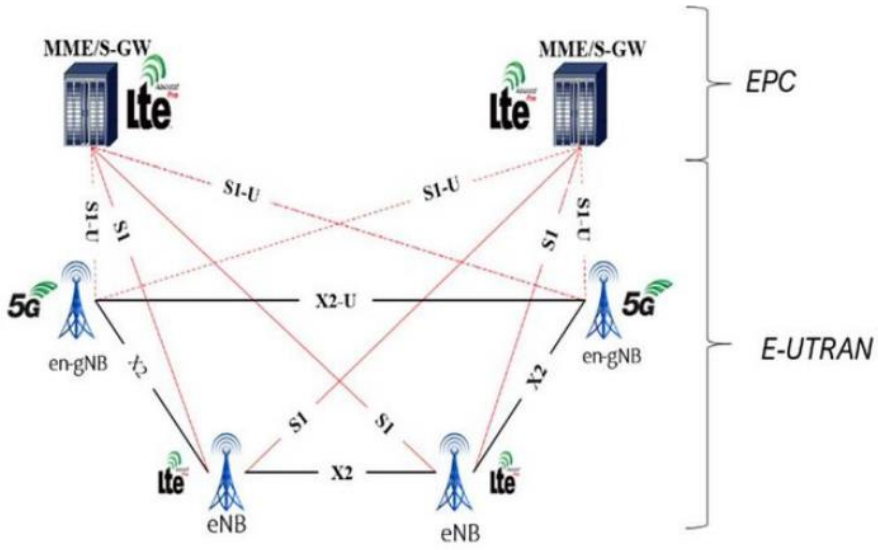

Figure 19: Non-Stand alone mode [23]

The NSA 5G NR mode is an early form of the SA 5G NR mode, in which 5G networks are supported by current LTE infrastructure. It primarily focuses on eMBB, in which 5G-enabled handsets and gadgets will employ mmWave frequencies for greater data capacity while continuing to use existing 4G infrastructure for voice communications.

In order for eMBB to gain a competitive advantage in the telecom industry, the NSA assists MNOs in rapidly launching 5G. The National Security Agency (NSA) also lends a hand in making use of the LTE/VoLTE footprint it already has in order to boost capacity, expand the LTE installed base, and boost delivery efficiency. It will not be able to use URLLC, mMTC, or network slicing, but its faster internet speeds will make it possible to stream videos, use augmented and virtual reality (AR), and have an immersive media experience.

By making use of the two brand-new radio frequency ranges, Non-Standalone 5G NR will increase data bandwidth:

- The frequency range 1 (450 MHz to 6000 MHz) is referred to as sub-6 GHz because it overlaps with 4G LTE frequencies. The groups are numbered 1 through 255.
- The frequency range 2 (24 GHz to 52 GHz) is the primary mmWave frequency band. The sizes

of the bands range from 257 to 511.

2.2.2 StandAlone (NSA) mode:

The Service-Based Architecture (SBA) and the functional separation of distinct network operations are two key features of the brand-new 5G SA core architecture, which is described by the 3GPP. MNOs planning to launch new enterprise 5G services like smart cities, smart factories, or other vertically integrated market solutions will benefit most from its design's unique advantage of end-to-end high speed and service assurance. With the deployment strategy, new services can be launched quickly with a short time to market. However, it comes with additional costs and difficulties associated with operating multiple cores in the network.

Standalone 5G NR is a brand-new end-to-end architecture that doesn't use the 4G LTE infrastructure and uses mm-waves and sub-GHz frequencies instead. Utilizing enhanced mobile broadband (eMBB), ultra-reliable and low latency communications (URLLC), and massive machine-type communications (mMTC), the SA 5G NR will achieve multi-gigabit data rates at lower costs and with improved efficiency.

In addition, 5G SA provides more advanced network slicing capabilities, facilitating operators' rapid transition to 5G New Radio (NR) and 5G as the core network. Ultra-low latency is provided by URLLC, mMTC, and network slicing for a wide range of next-generation applications like remote control of critical infrastructure, self-driving cars, improved healthcare, and more. On the other hand, advanced NR scenarios are not compatible with the EPC, which is the framework for providing integrated voice and data on a 4G LTE network. The dependability and latency of 5G will be crucial for controlling and coordinating drones, industrial automation, robotics, and smart grid control machines.

2.3 Technical Differences between 5G SA and 5G NSA

The primary difference between NSA and SA is that NSA sends 5G control signals to the 4G base station; in SA, on the other hand, the 5G base station is directly connected to the 5G core network, so control signals are sent without regard to the 4G network. In layman's terms, NSA is equivalent to adding a solid-state drive to an outdated computer to boost system performance, whereas SA is equivalent to replacing it with a new computer that uses the most recent technology and performs optimally. [23]

Some of the benefits include:

- The cost of NSA is significantly lower than that of SA.
- By reusing 4G facilities, the National Security Agency (NSA) makes it easier to deploy 5G networks, accelerating the time to market for 5G mobile broadband.
- Because 4G spots can be used to install 5G radio, NSA speeds up deployment and reduces time to market. SA needs to build 5G base stations and the back-end 5G core network in order to fully utilize 5G's features and characteristics.
- For advanced 5G use cases, SA combines a 5G core with SBA for scalability and flexibility, resulting in an ultrafast network with extremely low latency.

We have numerous deployment configuration options in each of these categories (SA and NSA).

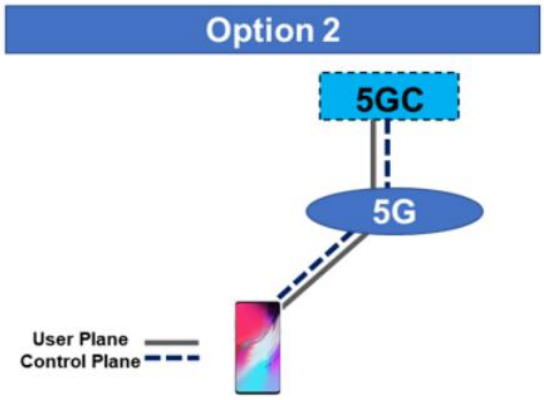

Figure 20: Stand-alone option 2

As per 3gpp , all the 5g features related to 5g for example network slicing , lower latency, high throughput and high density of device connectivity can only be obtain by using 5g Core. Here both control plane and user planes are initiated from 5g-NR for UE. Also, Lower call setup time in 5g because if 5g service is needed here once UE need to start data it will be moved from idle to connected and call will be setup , while in **option 3x** idle mode UE is connected with LTE and to start make data UE has to make the measurement of the 5g cell to make sure if the UE is in good coverage of 5g then after making measurement report, the 4g site will try to establish a connection over the X2 interface with 5g site and 5g can be accessed which is a longer time to access a 5g compare to a SA solution. Option 2 will provide more and more new business opportunities to Mobile Network Operators , more services will be introduce along with new feature or network slicing. Since 5g core is using Service Based Architecture therefore introducing any new function within the network is much easier.

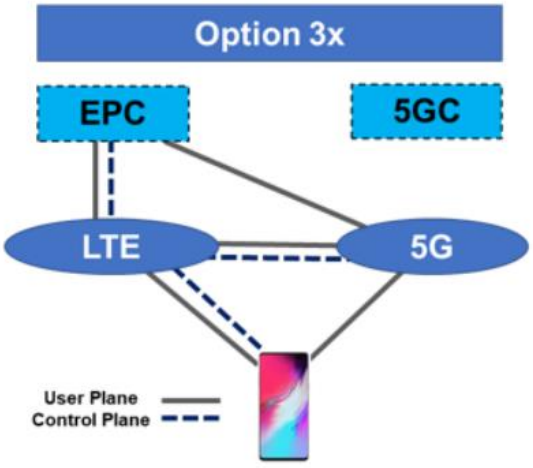

Figure 21: Non Stand-alone option 3x

According to the 3gpp, this method has a faster deployment because EPC is employed here, and existing LTE idle mode is used to fill coverage gaps caused by a low number of 5g stations at an early stage. Dual connection with LTE will also improve throughput over 5g alone because there are two data sources. In this situation, VoLTE is already in use; however, VoNR is not yet ready or has not been deployed in the network. The point to be noted here is that in IDLE mode UE will always be connected via 4g site because control plane is coming from LTE site and not from the 5g Node , however once the UE need to access data it will be connected with both LTE and 5g node. And once the data activity is finished the UE will be in IDLE mode connected back to LTE node , Now the question raised why SA is required once we have high throughput and low budget deployment already here , for that need to check the StandAlone architecture.

Above defined Architecture can be implemented by operator choices , for example if they already a nation wide network of 4g up and running they can simple upgrade the core and implement some of the 5g node anchoring with 4g , or if they are late in deploying 4g and area is small so its good if they can simply implement the 5g network etc..

Coming back to the further Deployment options.

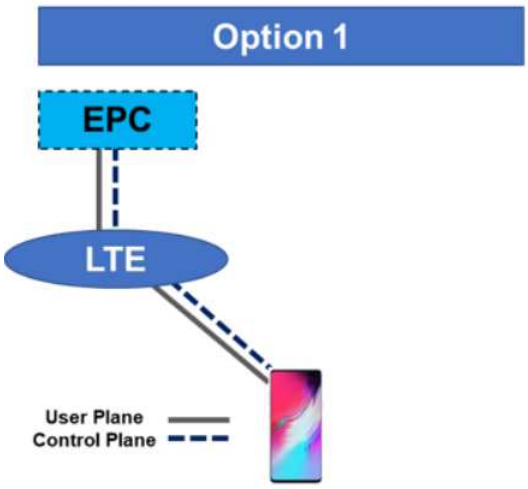

Figure 22: Non Stand-alone option 1

As per the 3gpp, Option 1 is an StandAlone deployment that represents the current 4g Deployments in many countries by different operators, in other words it can be said that legacy deployment of LTE radio connected to the Evolved Packet Core , here EPC has no any relation with 5gC or 5gNR. Here UE is only connected with LTE node to both user and control plane.

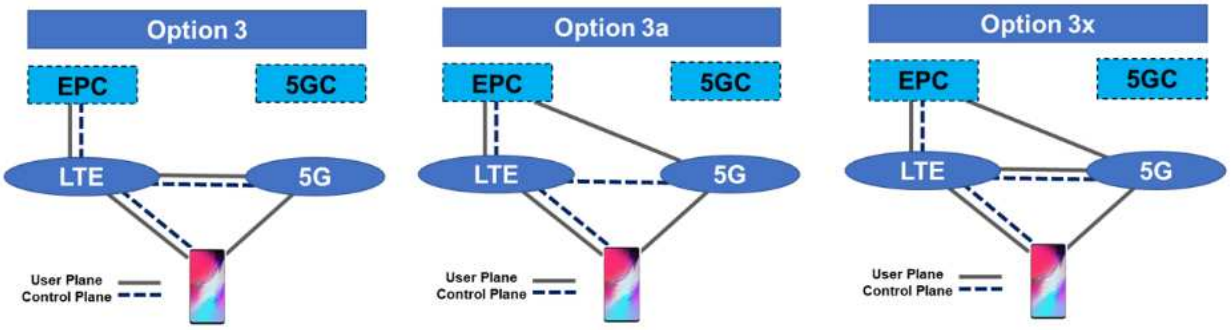

Figure 23: Non Stand-alone option 3,3a,3x

Option 3 is a Non-Standalone deployment, as mentioned in 3gpp, in which LTE and 5g NR radio access networks are available but managed by only the EPC core connected to the LTE access. In this case, LTE access is used as a control plane signalling anchor for 5g NR, and user data traffic (user plane) to the UE can be transmitted in both LTE and 5gNR modes.

Option 3x is a hybrid of options 3 and 3a, in which some user plane data traffic is routed directly from the EPC to the 5gNR and then to the UE. Alternatively, it is feasible to transfer a portion of the data from EPC to 5gNR through LTE RAN before it reaches the UE.

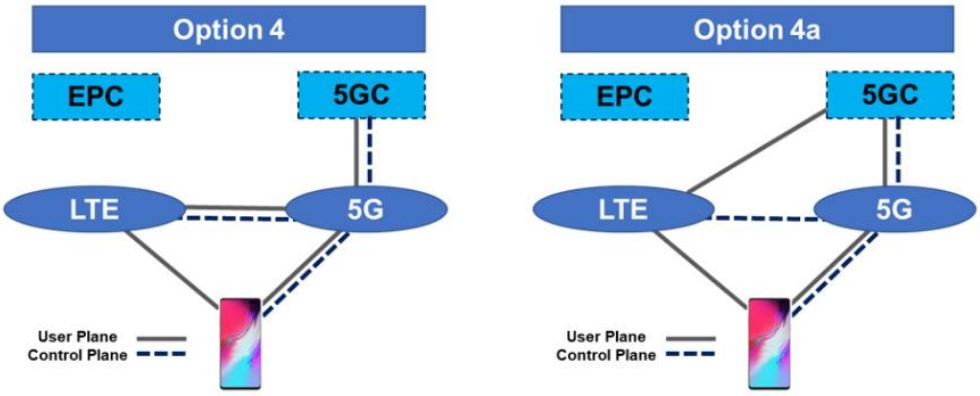

Figure 24: Non Stand-alone option 4 and 4a

Option four According to 3GPP specifications, there is also a Non Standalone (NSA) deployment option, in which both LTE and 5gNR radio access technologies are deployed and controlled using only 5gC. With this deployment option, user plane data from the LTE RAN is delivered directly or via 5gNR to the 5gc. In addition, the interface between LTE and 5gNR will be Xn.

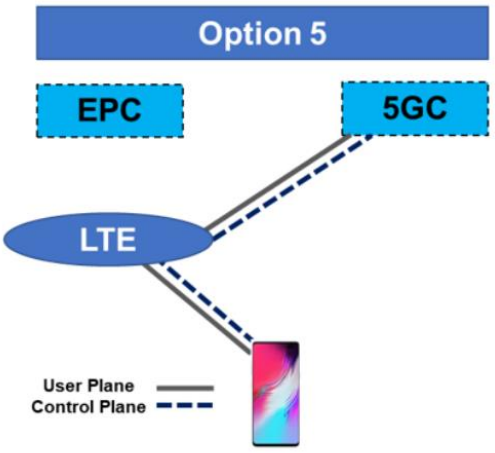

Figure 25: Non Stand-alone option 5

Option 5 is a StandAlone (SA) configuration in which the LTE RAN is linked to the 5gc. Other than for research purposes, this option appears unlikely to be implemented by any other MNOs, as the majority of the benefits of 5g come from shifting to a 5gNR access to network.

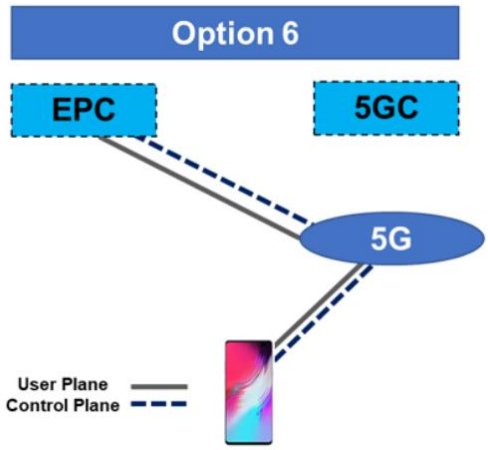

Figure 26: Non Stand-alone option 6

Option 6 as per the 3gpp is also an Standalone (SA) deployment and it relevant in scenarios where the radio network is completely migrated to 5gNR, but it keep the EPC. *Option 6 is later removed by 3gpp so the 5g devices can not support deployment option 6.*

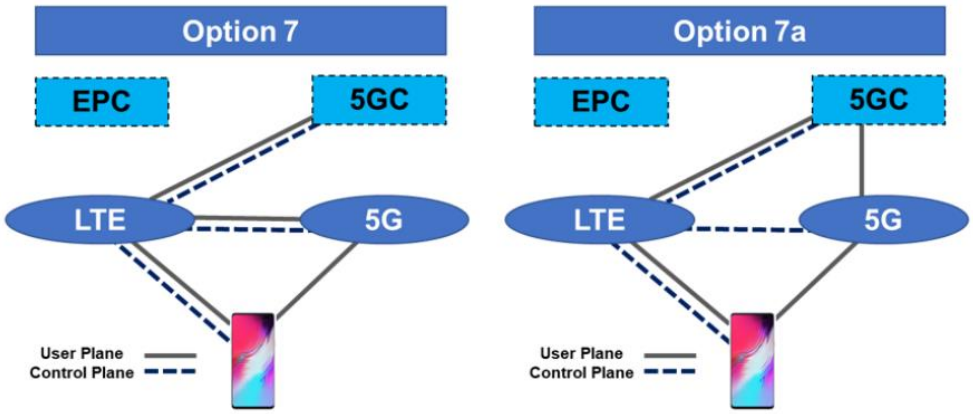

Figure 27: Non Stand-alone option 7 and 7a

Option 7 in the 3gpp depicts Non Standalone Deployments (NSA), in which the next generation core or the 5gc will be deployed with a mix of LTE and 5gNR. LTE is the master node in this case since it carries the 5gC control plane.

Option 3x seems to be the most suitable as it relies on the existing LTE and EPC which lead to faster deployment of 5g Network.

Option 3x can improve performance in several aspects for example , aggregating throughput using both 5gNR and LTE or optimized transmission of data using 5gNR downlink when the 5gNR coverage is good. It will also provide seamless mobility across LTE and 5gNR as the mobility is anchored by LTE.

2.4 5G Usage Scenarios in NSA and SA Operation

The 5G NR standards for the SA give a comprehensive set of criteria for the 5G core network that go beyond the NSA. The three key 5G usage scenarios outlined by the 3GPP and GSMA are as follows: [22]

1. Improved mobile broadband (eMBB)
2. Communication that is ultra-reliable and has a short latency (URLLC)
3. Large machine-to-machine communication (mMTC)

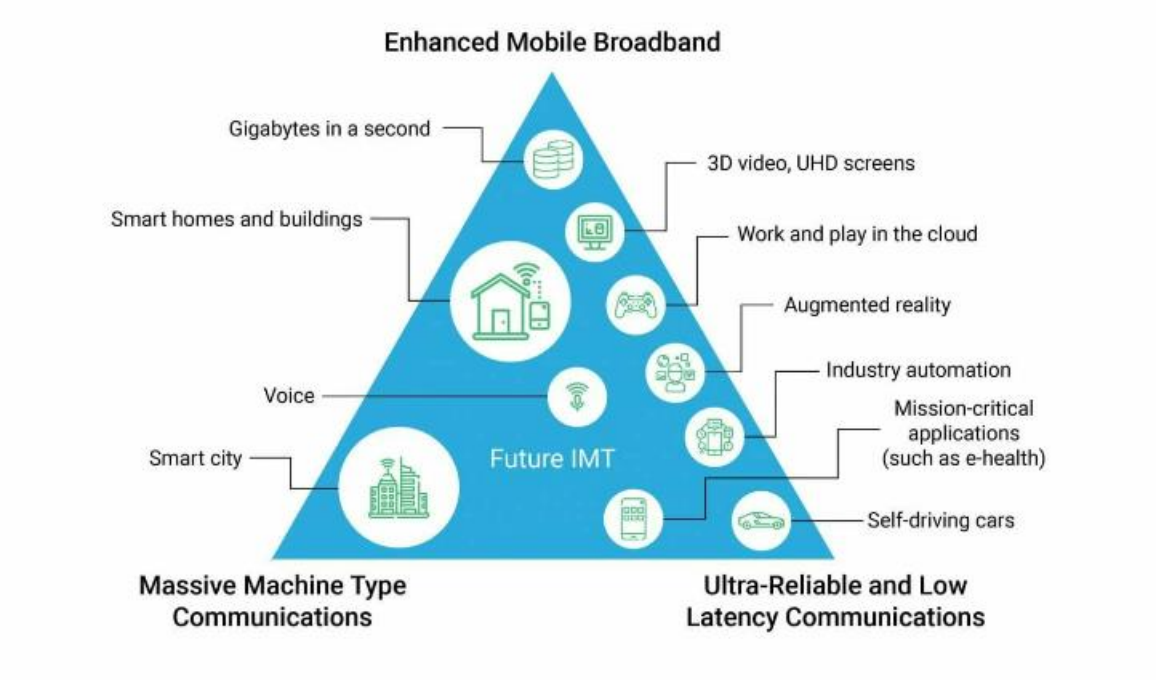

Figure 28: Major 5G usage Scenarios [22]

Advanced systems like Vehicle-to-Everything (V2X) contain components of all three use cases because more features and functionality are being linked every day.

The variety of IoT applications will drive a surge of diversity into 5G network slicing, despite the common perception that the Internet of Things is a single use case with similar requirements. Flexible latency requirements will exist for some Internet of Things devices, but remotely operated surgical or medical monitoring equipment may jeopardize latency reduction. Cameras and smoke detectors, for instance, will require higher levels of security that can be set in the network slice and prioritized traffic routing.

The numerous use cases emphasize the significance of E2E testing methods. Some examples of essential practices are as follows:

TeraVM is a software-based solution that is excellent for 5G security assessment and application emulation. It also provides network slice functionality verification and node selection, device variety, and volume simulation. The TeraVM is completely virtual and can simulate the activity of a city's worth of 5G network subscribers, making it easy to spot issues with flow and bottlenecks.

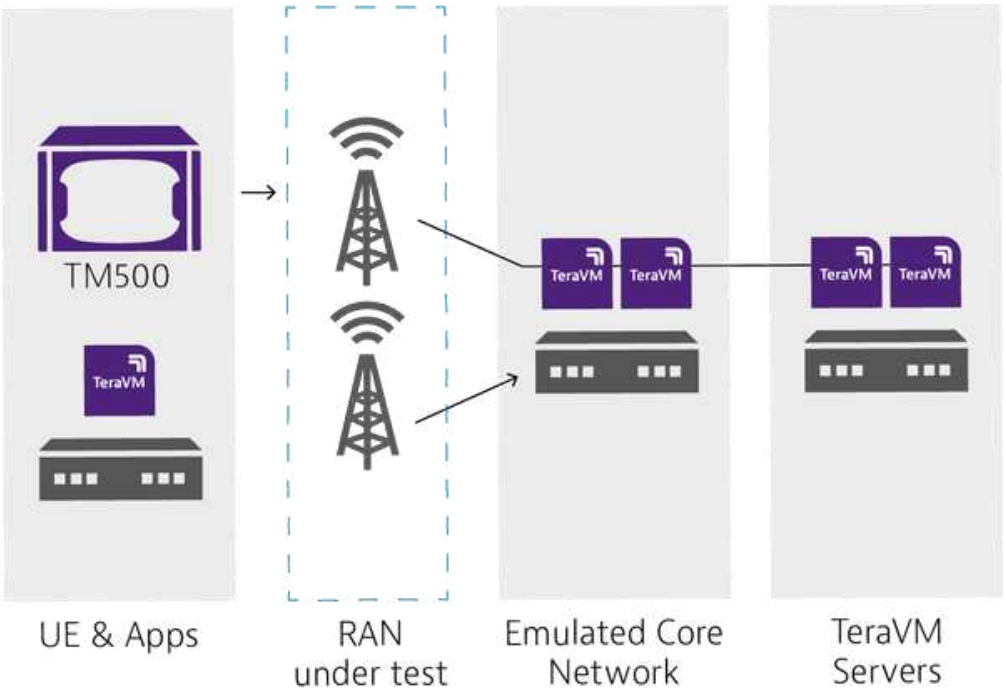

Figure 29: 5G core network emulation[23]

The TM500 test system is a powerful alternative for confirming end-user network performance through testing issues. The TM500 is able to support millimeter wave frequencies and a large number of 5G UEs in both standalone and non-standalone modes. The first-ever wrap-around testing of 5G standalone (SA) base station equipment was carried out using the TeraVM and TM500, both of which are renowned as market leaders in 5G tools. [23]

As a crucial architectural lynchpin, network slicing takes center stage when commercial 5G deployment becomes a reality. In addition to the revolutionary increase in speed and capacity, the diverse services that will drive fundamental improvements in numerous industries, including manufacturing, transportation, and medical, set 5G apart from previous wireless generations.

Network slicing can make it possible to create the ideal conditions for each use case as part of the network's overall picture. Utilizing test tools to proactively validate functionality is a prudent course of action because the new 5G highway will soon be clogged with traffic of all sizes and types.

2.4.1 EMBB – Enhanced Mobile Broadband

Enhanced Mobile Broadband (eMBB) is a 5G use case class that aims for peak download speeds of more than 10 Gbps. This is significantly faster than the 3 Gbps that 4G LTE Advanced Pro can achieve.

For the majority of potential 5G users, we naturally anticipate the extremely fast mobile broadband speeds. Consequently, this is a clear use case to which all of us can still relate today. We are all aware of how crucial it is to have fast internet connections, and many of us have looked into the potential of 4G LTE networks as home broadband in the past. However, with peak speeds exceeding 10 Gbps, 5G raises the bar for mobile broadband speeds to an entirely new level. However, because they represent the absolute maximum that a network can provide under ideal conditions, peak speeds are essentially theoretical. We have to deal with the route loss that a mobile signal must endure

(signal fading) and the fact that we share the network with many other people, which limits a person's reception speed. We also live in an environment that is not ideal. Consequently, the actual 5G data rates in the real world are significantly lower than the 10 Gbps, which we discuss in greater detail in our dedicated piece on average 5G speeds. High information speeds are expected for eMBB, which 5G organizations might furnish either alone or related to 4G LTE organizations.

2.4.2 URLLC – Ultra-Reliable Low Latency Communication

Ultra-Reliable Low Latency Communication (uRLLC) is a use case class that sets the fundamental requirements for 5G networks to support low data rates (bps or kbps) and extremely low latencies (below one millisecond).

uRLLC, or ultra-reliable low latency communication, is utilized when extremely low latencies of one millisecond or less necessitate highly dependable connectivity (99.99% reliable). uRLLC does not require the massive data rates that 5G networks are renowned for; Instead, the connection's dependability and speed at low data rates are the primary concerns. 5G networks can operate at high, medium, and low frequencies, among others. In terms of overall performance, higher frequencies offer faster data rates and shorter latencies. For instance, think about oneself driving vehicle. Latencies must be extremely low because real-time communication is required for the vehicle-to-network connection. By allowing latencies of one millisecond or less, 5G NR can be used for a wide range of applications, including industrial automation, mission-critical applications, and self-driving cars.

2.4.3 MMTC – Massive Machine-Type Communication

In 5G networks, Massive Machine Type Communication (mMTC) is a use case class that outlines the minimum requirements for enabling 1 million low-powered, low-cost, and low-complexity devices per square kilometer with a battery life of up to ten years.

A use case called "Mega Machine Type Communication" requires the network to be able to handle the massive deployment of billions of low-cost, low-power devices. This is one of the main use-case classifications important to lay the preparation for the digitalization of various areas through cell IoT advances. The deployment of billions of devices necessitates low-powered devices with a battery life of up to ten (10) years and a very low complexity level in order to keep costs low. Low-data-rate applications like home automation with sensors and actuators and machine monitoring systems can be accommodated by the mMTC requirement for 5G. Smart meters, for instance, transmit text-based data at modest data rates and do not require a real-time connection, so a delay of a few seconds is not a problem.

2.5 Basic Network Architecture

Mobile devices can communicate wirelessly thanks to the cellular network. Smartphones and tablets have typically been considered User Equipment (UE) devices; however, UE devices will increasingly include automobiles, drones, industrial and agricultural machinery, robotics, home appliances, medical devices, and so on. [24]

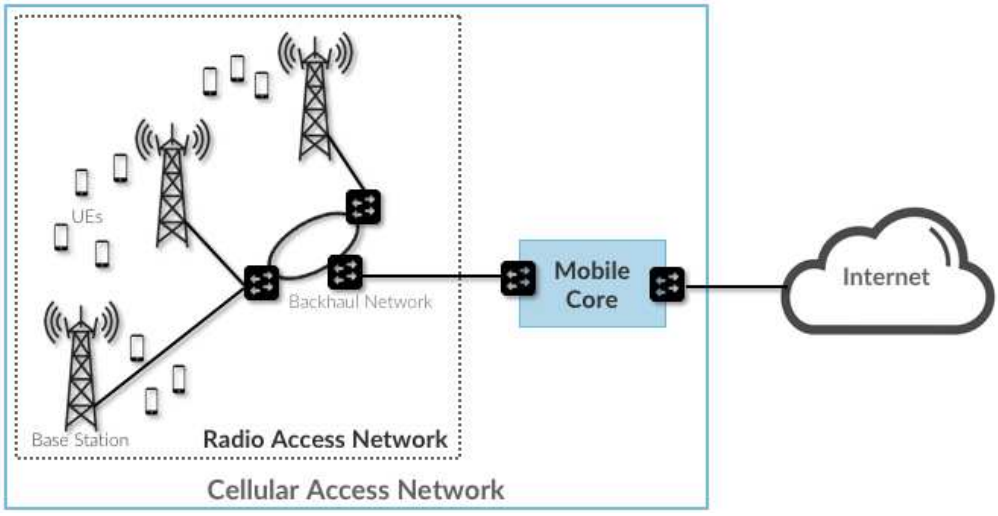

Figure 30: Cellular networks consists of a Radio Access Network (RAN) and a Mobile Core [52]

The two primary subsystems of the cellular network are shown in Figure 30: the Mobile Core as well as the Radio Access Network (RAN). The radio spectrum is maintained by the RAN, which ensures that it is utilized effectively and that all users' requirements for quality of service are met. A dispersed network of base stations is what it means. These are referred to as gNB in 5G, where the letter g stands for "next Generation." In 4G, these are referred to somewhat cryptically as eNodeB (or eNB), which stands for evolved Node B.

Instead of a device, the Mobile Core is a set of capabilities that perform various tasks.

• Provides connectivity via Internet Protocol (IP) for voice and data services.

• Ensures that the link satisfies the promised QoS requirements.

• Keeps an eye on user mobility to guarantee uninterrupted service.

• Monitors subscriber usage for the purpose of billing and charging.

Mobile Core is yet another example of a generic term. In 4G, this is referred to as the Evolved Packet Core (EPC), and in 5G, it is referred to as the Next Generation Core (NG-Core).

Even though its name includes the word "Core," the Mobile Core is still a part of the access network from the perspective of the Internet. It acts as a link between a RAN in a specific area and the larger IP-based Internet. Albeit 3GPP considers broad adaptability in how the Portable Center is topographically conveyed, assuming that every launch of the Versatile Center serves a metropolitan locale is a decent turning out model for our motivations. The associated RAN would then cover a number of cell towers, if not hundreds.

As shown in Figure, a Backhaul Network links the Mobile Core to the base stations that implement the RAN. This network is frequently wired, may or may not have the ring topology shown in Figure 1, and is frequently constructed using common Internet components. For establishing RAN backhaul, for instance, the Passive Optical Network (PON) that provides Fiber-to-the-Home is an excellent

option. The backhaul network is unquestionably an essential component of the RAN, but the 3GPP standard does not mandate its implementation.

Network operators have traditionally purchased proprietary implementations of each subsystem from a single vendor, despite the fact that 3GPP specifies all of the parts that implement the RAN and Mobile Core in an open standard, including sub-layers that we have yet to introduce. The perception of "opaqueness" associated with the cellular network in general and the RAN in particular is exacerbated by the absence of an open-source implementation. There is a great deal of potential to open and disaggregate both the RAN and the Mobile Core, even though an eNodeB implementation does include sophisticated algorithms for scheduling transmission on the radio spectrum—algorithms that equipment manufacturers consider to be important intellectual property. In the subsequent two sections, each is discussed individually.

Figure 31 redraws parts from Figure 30 to show two important differences before getting into the specifics. The first is that a base station has both an analog and a digital component, which is shown by an antenna and a processor pair. The Mobile Core has a Control Plane and a User Plane, which is similar to the control/data plane split that Internet users are accustomed to. This is the second difference. Control and User Plane Separation (CUPS) was also coined by 3GPP to represent this idea.) In the following discussion, the significance of these two distinctions will be emphasized.

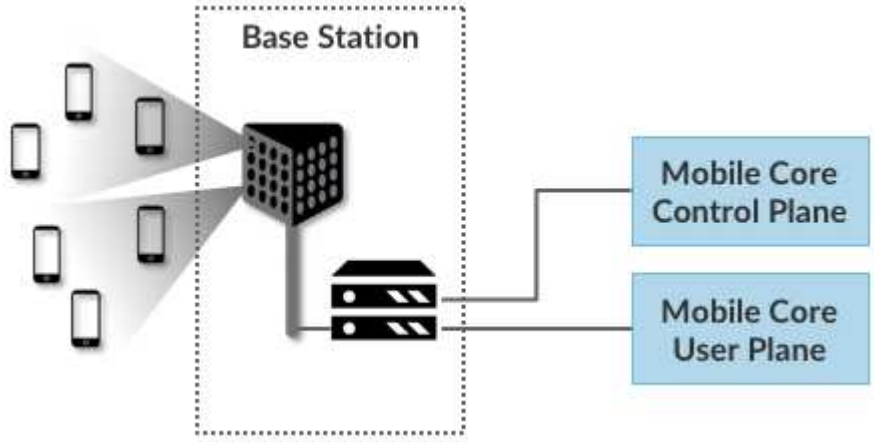

Figure 31: Mobile Core divided into a Control Plan and a User Plane, an architectural feature known as CUPS: Control and User Plane Separation

2.6 Radio Access Network

By describing the functions of each base station, let's break down the RAN. Keep in mind that this is similar to talking about the Internet by explaining how a router works. While this is a good place to start, it doesn't fully explain the story.

Second, each base station establishes the wireless channel for a subscriber's UE at power-up or handover while the UE is active. After the UE has been idle for a predetermined amount of time, this channel is released. In 3GPP jargon, this wireless channel is referred to as providing a bearer service. In the past, telecommunications, particularly early wireline technologies like ISDN, have used the term "carrier" to refer to a data channel rather than a signaling channel.

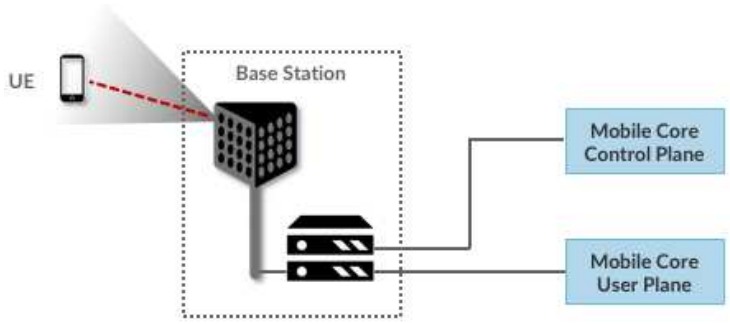

Figure 32: Base Station detects (and connects to) active UEs.

Second, the "3GPP Control Plane" connects each base station to the matching Mobile Core Control Plane component and facilitates signaling communication between the two. UE authentication, registration, and mobility tracking are made possible by this signaling flow.

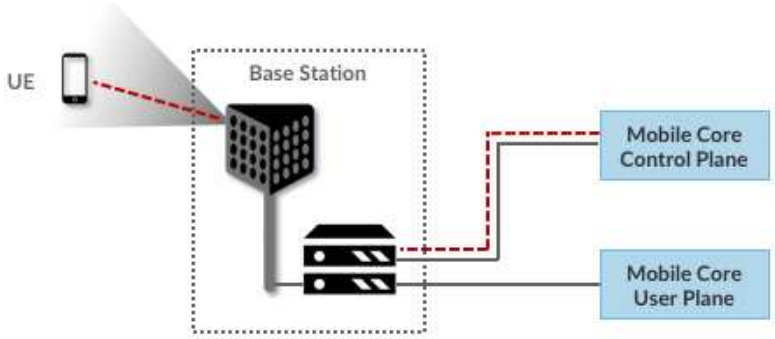

Figure 33. Base Station establishes control plane connectivity between each UE and the Mobile Core.

Thirdly, the base station creates one or more tunnels between the Mobile Core User Plane component that corresponds to each active UE.

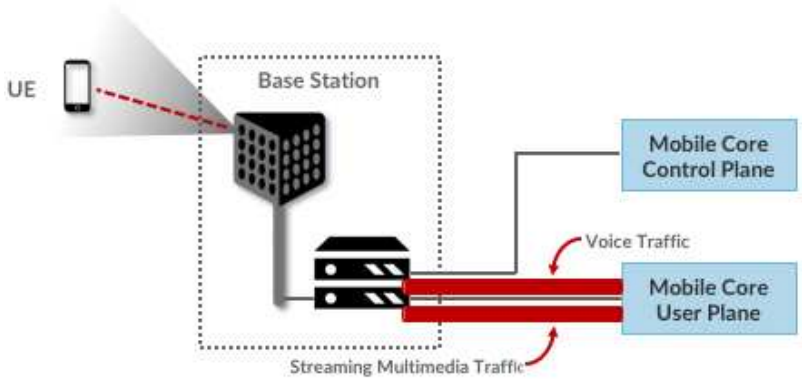

Figure 34: Base station establishes one or more tunnels between each UE and the Mobile Core's User Plane.

Fourth, control and user plane packets are sent by the base station between the Mobile Core and the UE. The SCTP/IP and GTP/UDP/IP tunnels are utilized for the routing of these packets, respectively. A reliable alternative to TCP, SCTP (Stream Control Transport Protocol) is intended to transmit control (signal) information for telecommunications services. GTP, which stands for "General Packet Radio Service Tunneling Protocol," is a UDP-based 3GPP-specific tunneling protocol.

It is important to note that the RAN-to-Mobile Core connection is IP-based. Prior to 4G, the cellular network's internals were circuit-based, which is understandable considering the network's origins as a speech network. This was introduced as one of the most significant differences between 3G and 4G.

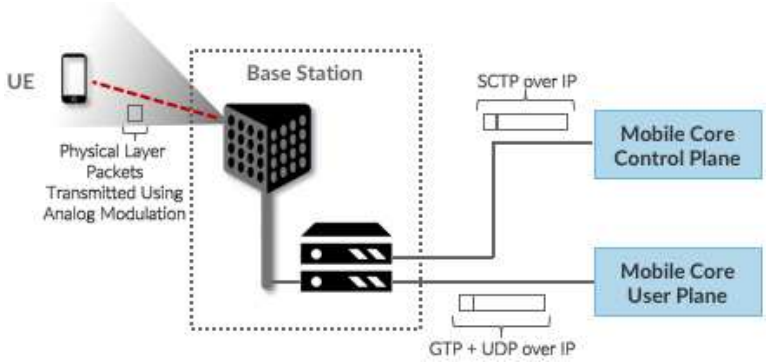

Figure 35: Base Station to Mobile Core (and Base Station to Base Station) control plane tunneled over SCTP/IP and user plane tunneled over GTP/UDP/IP [24]

Fifth, each base station coordinates UE handovers with nearby base stations through direct station-to-station communications. Both control plane (SCTP over IP) and user plane (GTP over UDP/IP) packets are transferred over these lines, just like the station-to-core connectivity shown in the previous figure.

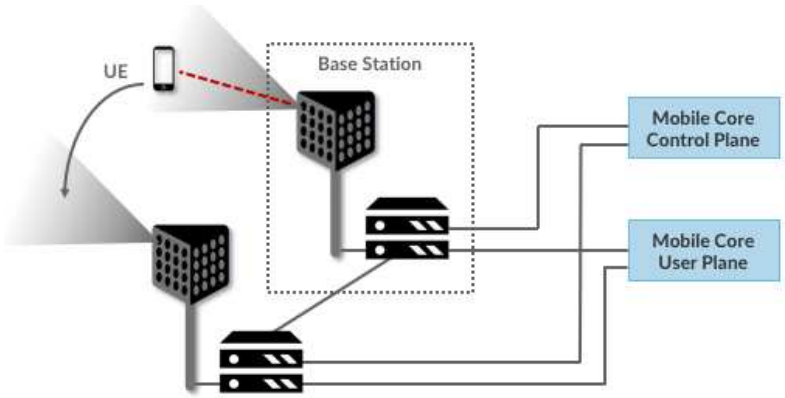

Figure 36: Base Stations cooperate to implement UE hand over [24]

Sixth, multi-base station wireless multi-point transmission to a UE is coordinated by the base stations. This transmission may or may not be part of a UE handover from one base station to another.

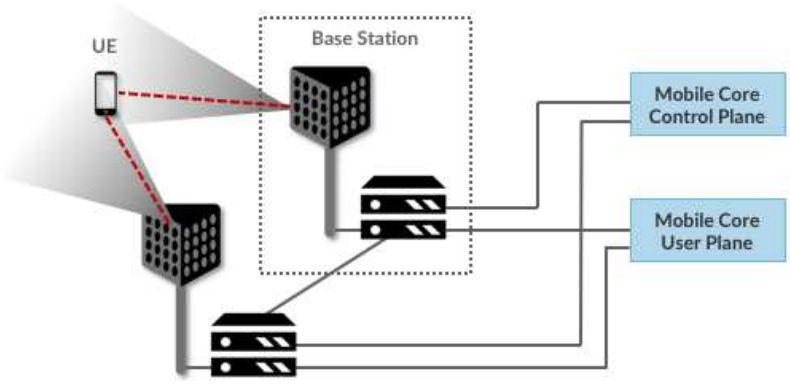

Figure 37: Base Stations cooperate to implement multipath transmission (link aggregation) to UE [24]

The fact that the base station functions as a specialized forwarder is crucial. It assembles physical layer segments into IP packets and forwards them to the Mobile Core's upstream user plane via a GTP/UDP/IP tunnel after fragmenting outgoing IP packets into physical layer segments and scheduling their transmission for the Internet-to-UE direction over the available radio spectrum. Depending on observations of wireless channel quality and per-subscriber regulations, it also decides whether to (a) forward outgoing packets directly to the UE, (b) indirectly send packets to the UE via a neighboring base station, or (c) use multiple pathways to reach the UE. In the third scenario, the physical payloads might be spread out among a lot of base stations or among different carrier frequencies of a single base station (such as Wi-Fi).

Scheduling is challenging and multifaceted, even when taken into consideration as a localized option at a single base station. Now that we know more, we can see that there is also a global component. In order to make better use of the radio spectrum over a larger area, traffic can be routed to a different base station or multiple base stations.

To put it another way, the RAN as a whole, as opposed to just a single base station, enables not only handovers—an obvious requirement for mobility—but also link aggregation and load balancing—techniques that are familiar to anyone who has studied the Internet. We will examine how SDN approaches can be used to make such RAN-wide (global) decisions in a subsequent chapter.

2.7 Mobile Core

The Mobile Core's primary responsibility is to provide mobile customers with connectivity to the external packet data network (Internet) while also ensuring that they are authenticated and that the observed service quality satisfies their subscription service level agreements (SLAs). The Mobile Core is responsible for managing the mobility of all subscribers by tracking their last location down to the serving base station's granularity. Because the Internet's core does not keep track of individual subscribers, the Mobile Core's architecture is very complicated, especially when those subscribers move around. [25]

As we move from 4G to 5G, the way that functionality is virtualized and integrated into individual components varies, but overall functionality stays the same. The 5G Mobile Core is largely influenced by the cloud's move toward a microservice-based (cloud native) architecture. This shift toward cloud native is more widespread than first appears, in part due to the increased flexibility and specialization

it provides. The 5G Mobile Core could evolve to serve large IoT, which has a fundamentally different latency requirement and usage pattern (i.e., many more devices connecting intermittently) rather than just phone and broadband connectivity. A one-size-fits-all approach to session management is undermined, if not destroyed, by this.

5G Mobile Core

A microservice-like architecture is used in the 5G Mobile Core, which is known as the NG-Core by 3GPP. Although the 3GPP specification specifies this level of disaggregation, it is essentially just a set of functional blocks rather than an implementation, which is why we refer to it as "microservice-like." A microservice-based system's set of technical decisions are not the same thing as a set of functional blocks. Nevertheless, a reasonable working paradigm is to consider the components in Figure 38 as a collection of microservices.

The functional block collection is broken up into three groups, as shown below. The first group is mirrored in the EPC and operates in the Control Plane (CP).

Core Access and Mobility Management Function, or AMF for short: manages mobility, access authentication and authorization, connectivity and reachability, and location services. manages the mobility-related components of the EPC's MME.

Session Management Function, or SMF: each UE session is managed, with IP address allocation, UP function selection, QoS control, and UP routing control all included. approximately corresponds to a portion of the EPC's MME and the PGW's control-related components.

Policy Control Function, or PCF: manages the policy rules that other CP functions use to enforce. corresponds roughly to the PCRF of the EPC.

Unified Data Management, or UDM: manages user identification and the generation of authentication credentials. Included is a portion of the EPC's HSS's functionality.

Authentication Server Function, or AUSF: In essence, this is an authentication server. Included is a portion of the EPC's HSS's functionality.

The second group operates similarly in the Control Plane (CP), but the EPC does not directly mirror it:

Structured Data Storage Network Function, also known as SDSF: a service for "assistant" storage of structured data. A "SQL Database" could be utilized in a system that is based on microservices.

Network Function for Unstructured Data Storage (UDSF): a service that acts as a "helper" to store unstructured data. In a microservices-based system, this could be done with a "Key/Value Store."

Network Exposure Function, or NEF: a way to make particular capabilities available to third-party services, like the ability to convert data between internal and external representations. This could be done by an "API Server" in a microservices-based system.

NF Repository Function, or NRF: An instrument for finding accessible administrations. This could be done by a "Discovery Service" in a microservices-based system.

The Network Slicing Selector Function, or NSSF, is: a method for selecting a Network Slice that will serve a particular UE. Network slices are basically a way to divide up network resources so that different users can get different services. That is a significant aspect of 5G that we will discuss in greater detail in a subsequent chapter.

One component operates in the User Plane (UP) of the third group:

UPF, or User Plane Function,: The S/PGW blend in EPC is utilized to advance correspondences between the RAN and the Web. In addition to packet forwarding, it is in charge of policy enforcement, lawful intercept, traffic consumption reporting, and QoS policing.

While the second group is 3GPP's method of pointing to a cloud native solution as the preferred end-state for the Mobile Core, despite the unnecessary inclusion of new terminology, the first and third groups are best understood as a fundamental restructuring of 4G's EPC. It is important to note that by introducing distinct storage services, all other services can be stateless, making scaling easier. However, in microservice-based systems, a message bus that links all of the components is depicted in Figure 38 rather than a complete set of pairwise connections. Additionally, this requires a clearly defined plan for implementation.

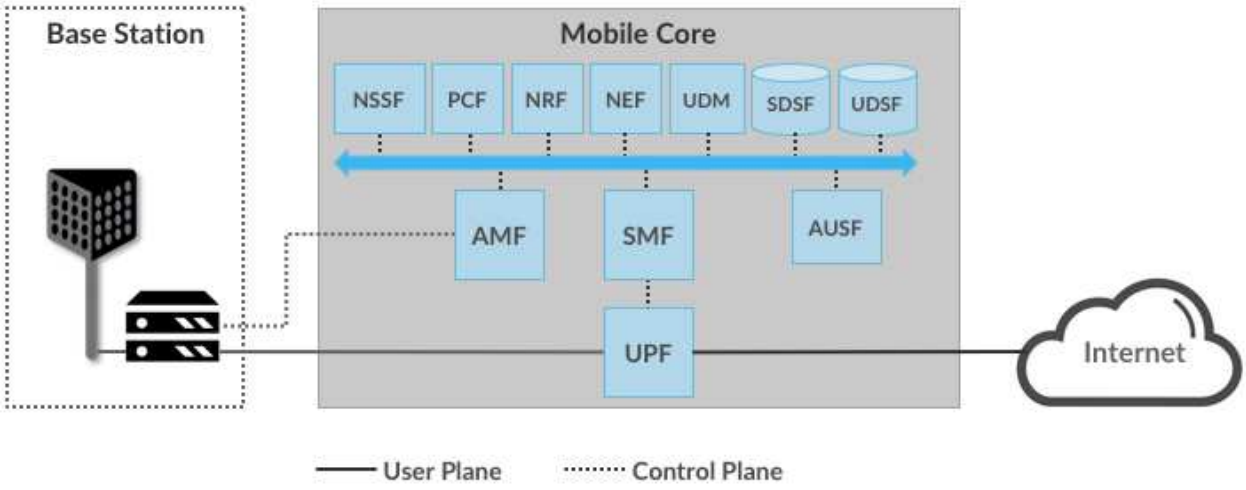

Figure 38: 5G Mobile Core (NG-Core) [25]

Taking a step back from these particulars and assuming implementation, the important conclusion is that the Mobile Core can be viewed as a service graph. This may also be referred to as a service chain or a service graph, with the latter term appearing more frequently in NFV-related writings. "Service mesh," on the other hand, "has taken on a pretty specific meaning in cloud native terminology; In order to avoid simplifying things too much, we won't use it here. 3GPP is silent on the specific wording because it is considered an implementation decision rather than a part of the specification.

2.8 5G Network Slicing

On top of a shared physical infrastructure, it is a network configuration that enables the creation of multiple virtualized and independent networks. As it has developed, this configuration has become an essential component of the broader 5G architectural environment. 26] The requirements of the application, use case, or customer can be used to assign each "slice," or section, of the network.

While some services, like smart parking meters, place a premium on high dependability and security and are less concerned about latency, others, like driverless cars, may require ultra-low latency (URLLC) and rapid communication speeds. Support for these various services and efficient resource reassignment between virtual network slices are made possible by 5G network slicing.

Applications that are 5G-enabled or better require a lower latency, more connections, and more bandwidth than previous generations could provide. A one-size-fits-all approach to service delivery is out of date because each use case will have unique performance requirements.

The design of organization cutting in 5G is like that of a perplexing public transportation framework. In contrast to rows of identical lanes and automobiles, some aspects of transportation are universal, such as roads and bridges. On the other hand, other modes of transportation and automobiles are tailored to the speed, cost, and volume requirements of the user. The fundamental tenets of the architecture are E2E network slicing (end-to-end) and logical separation from other slices, even though each unique slice traverses numerous common network elements.

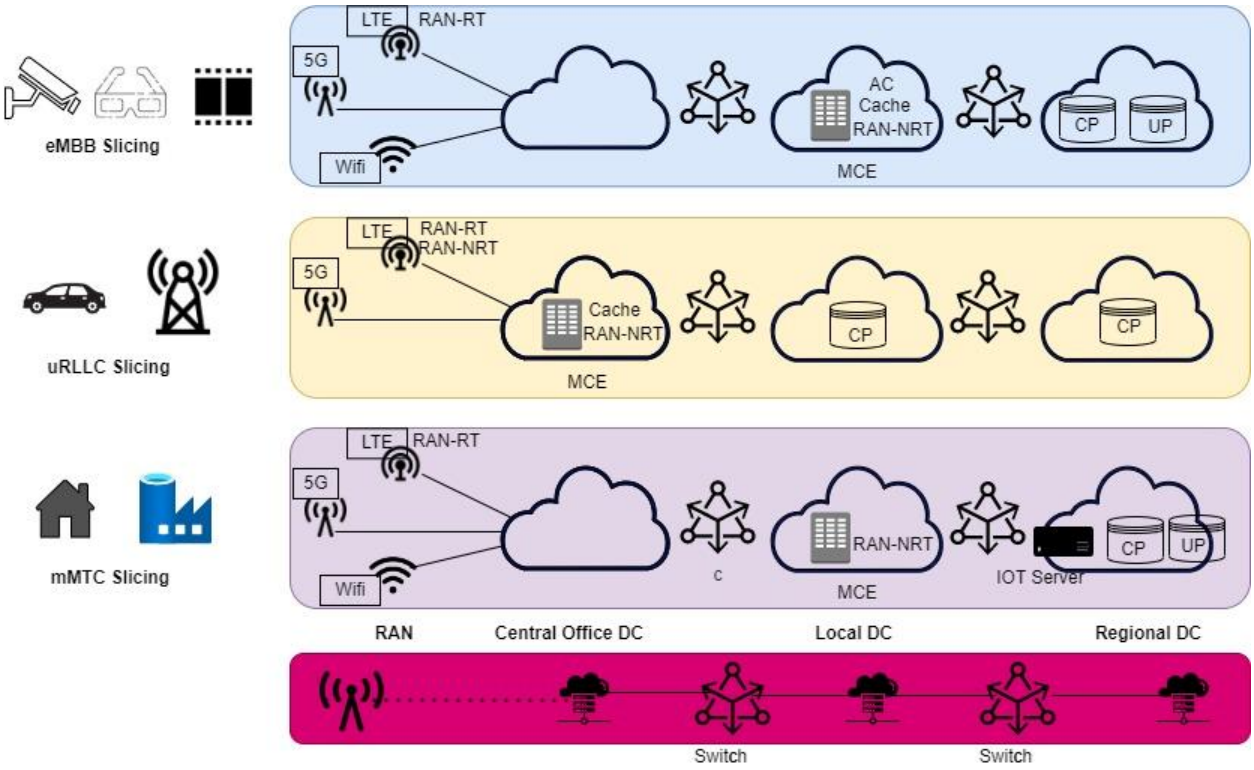

Figure 39: 5G Network Slicing

A crucial part of the architecture used to control traffic flows through the application program interfaces (APIs) of a central control plane is a network segmentation SDN (software defined network). In order for the application layer to be able to offer the client individualized services, the control plane configures resources. Additionally, SDN includes an infrastructure layer that is in charge of data forwarding and control plane rule processing in addition to providing fundamental network services. Services are mapped and inter-layer functionality is monitored by the network slice controller (or orchestrator).

SDN virtualization makes it possible for each client instance to unlock and arrange the specific resources that make up a slice along with the services that are required. A dynamic function that must constantly monitor performance and isolate slices in order to meet these requirements. Recursion, which enables the control plane to construct numerous sub controllers to support slice composition, is yet another important component of network slicing SDN.

Network function virtualization (NFV) is an additional requirement for slicing. The goal of the NFV concept is to provide services that were previously only available on proprietary hardware by installing network functionality on virtual machines (VMs) on a virtualized server.

NFV can also be used to manage the lifecycle of network slices and the infrastructure resources associated with them. In either the core or edge clouds, the SDN is used to manage the provisioning of virtual machines. SDN and NFV network slicing and exploitation of virtual and physical resources, including RANs, can be supported by these technologies as a whole.

Network Slicing by the 3GPP Working groups that are developing the 5G core architecture with network slicing as a fundamental element have continued to focus on slicing because the 3GPP has identified network slicing as a crucial overall component. Stage 2 with slicing is described in 3GPP Technical Specification (TS) 23.501, while TS 22.261 provides network slice provisioning, device association to slices, and performance isolation during regular and elastic slice operation. [26]

Low latency industrial IoT and autonomous driving are two examples of 5G potential found in the Version 16 of the 3GPP 5G specification. The report discusses 5G core solutions for the cellular Internet of Things, as well as the bandwidth and cost implications of unlicensed NR spectrum.

Challenges and Opportunities

Because a single network can be segmented to fit a variety of use cases based on client demand and segmentation, the benefits are obvious. The ability to provide a network slice as a service reduces both operating and capital expenses (CAPEX), and operators can then allocate resources to each slice based on the required speed, throughput, and latency to cover the breadth of network slicing in 5G. In terms of coverage, capacity, and connectivity, it has the ability to give priority to essential public entities like medical emergency teams and first responders.

Although 5G necessitates network slicing, this is not the case. In contrast to other essential features of 5G, it can be implemented on existing 4G/LTE networks, providing immediate advantages while preparing for the upcoming transition.

Using AI-powered orchestration, it also offers a safe and effective alternative for testing and deploying new services. In order to evaluate new services, it is no longer necessary to make changes that break existing services because the network has been separated. There are fewer functions to deploy for each new slice.

2.9 5G Identifiers SUPI and SUCI

Each SIM card in a telecom network is given a unique identification number by network operators. This number is called an IMSI (International Mobile Subscriber Identity) up to 4G and a SUPI (Subscription Permanent Identifier) for 5G. A user cannot be authenticated until the user has been

identified because authentication between the user and the network provider is based on a shared symmetric key. However, these persistent identifiers can be utilized to identify, locate, and track individuals if the IMSI/SUPI values are broadcast in plaintext over the radio access link. [27]

The visiting network gives the SIM card temporary IDs, which are called Temporary Mobile Subscriber Identity (TMSI) until 3G systems and GUTI for 4G and 5G systems, to prevent this invasion of privacy. The radio access link is then used to use these dynamic temporary identifiers for identification. However, temporary identifier authentication is not possible in some cases, such as when a user registers with a network for the first time and is not yet assigned a temporary identifier or when the visiting network is unable to resolve the IMSI/SUPI. from the TMSI/GUTI provided.

This can be deliberately imitated by an active man-in-the-middle opponent to get an unsuspecting user to reveal its long-term identity. Today's mobile networks, including 4G LTE/LTE-Adv, are still vulnerable to these attacks, which are referred to as "IMSI catching".

Subscription Permanent Identifier (SUPI)

According to 3GPP specification TS 23.501, a SUPI is a globally unique Subscription Permanent Identifier (SUPI) assigned to each subscriber in 5G. In USIM, SUPI is enabled, and in 5G Core, the UDM/UDR function is enabled. [28]

A legitimate SUPI can be one of the accompanying:

an NAI (Network Access Identifier) as described in RFC 4282 for non-3GPP RAT, and an IMSI (International Mobile Subscriber Identification) as defined in TS 23.503.

There are typically 15 decimal digits in a SUPI. The Mobile Network Code (MNC), which identifies the network operator, is made up of two or three numbers, the first three of which are the Mobile Country Code (MCC). The Mobile Subscriber Identification Number (MSIN) is the remaining nine or ten digits, and it identifies the particular operator's unique user. Similar to IMSI, SUPI is a 15-character string that uniquely identifies the ME.

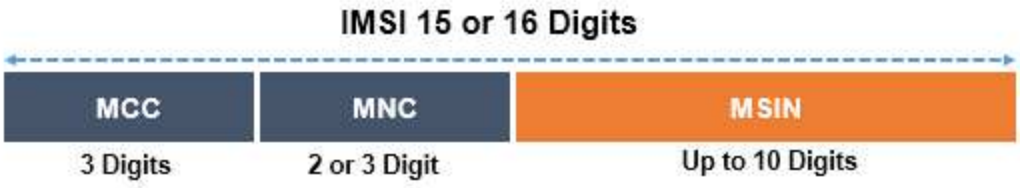

Figure 40: Subscription permanent identifier [27]

Solution to IMSI Catchers in 5G

IMSI-catching attacks have posed a threat to all mobile telecommunications generations (2G, 3G, and 4G) for decades. Supporting backwards compatibility for legacy reasons appears to have maintained this privacy issue. In contrast, this issue has finally been resolved by the 3GPP, albeit at the expense of backward compatibility. In contrast to previous generations, the security specifications for 5G do not permit plain-text transfers of the SUPI over the radio interface in the event that a 5G-GUTI fails

to identify the device. Instead, the disguised SUPI is sent along with an ECIES-based privacy-preserving identifier. Subscription Concealed Identifier, or SUCI, is the name given to this secret SUPI.

Subscription Concealed Identifier (SUCI)

A privacy-preserving identifier that includes the hidden SUPI is the Subscription Concealed Identier (SUCI). An SUCI is created by the UE using the Home Network's public key, which was securely provisioned to the USIM during registration using an ECIES-based protection scheme.

Only the MSIN portion of the SUPI is hidden by the security system, while the home network identity, MCC/MNC, is sent in plain text. The SUCI is made up of the following data fields:

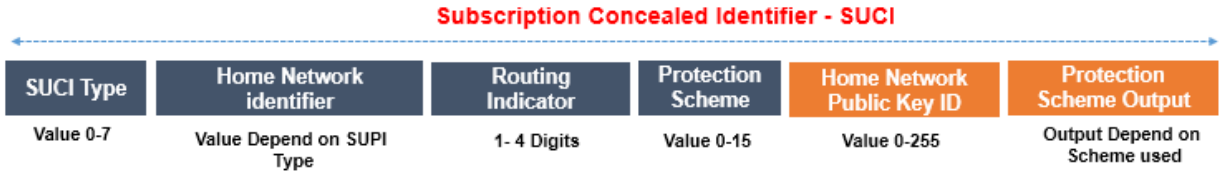

Figure 41: Subscription concealed identifier [27]

SUPI Type: consisting in a value in the range 0 to 7. It identifies the type of the SUPI concealed in the SUCI. The following values are defined

- 0: IMSI
- 1: Network Access Identifier (NAI)
- 2 to 7: spare values for future use.

Identifier for the Home Network: identifies the home network of the subscriber. When the SUPI Type is an IMSI, MCC and MNC make up the Home Network Identifier. The Home Network Identifier is a variable-length string of characters that represents a domain name when the SUPI type is Network Access Identifier. User@techno.com, for instance [27]

Indicators of Routing: It consists of one to four digits that are stored in the USIM and provided by the home network operator.

The Protection Scheme's Identifier: It is represented by four bits and has a value between 0 and 15.

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

Public Key Identifier for the Home Organization: It has a number between 0 and 255 in it. It is used to identify the SUPI protection key and is an HPLMN-provisioned public key. This data field must be set to 0 if the null scheme is used.

The Protection Scheme's Output: It is comprised of a variable-length series of characters or hexadecimal digits, contingent upon the insurance conspire used.

5G Identity Exchange between UE and Network

On the over-the-air radio interface, a UE can be identified thanks to the subscriber identification mechanism (SUCI). The UE-Network Identify exchange is depicted in the following diagram. [28] A UE encrypts SUPI and sends an Initial Registration Requested with SUCI when it attempts to register for the first time. In order for AUSF and UDM to obtain the SUPI with Authentication Request, AMF sends this SUCI to them. An authentication response containing SUPI data from AUSF is required. After that, AMF creates a GUTI for this SUPI and stores the GUTI to SUPI mapping for use in subsequent registrations or requests for PDU sessions.

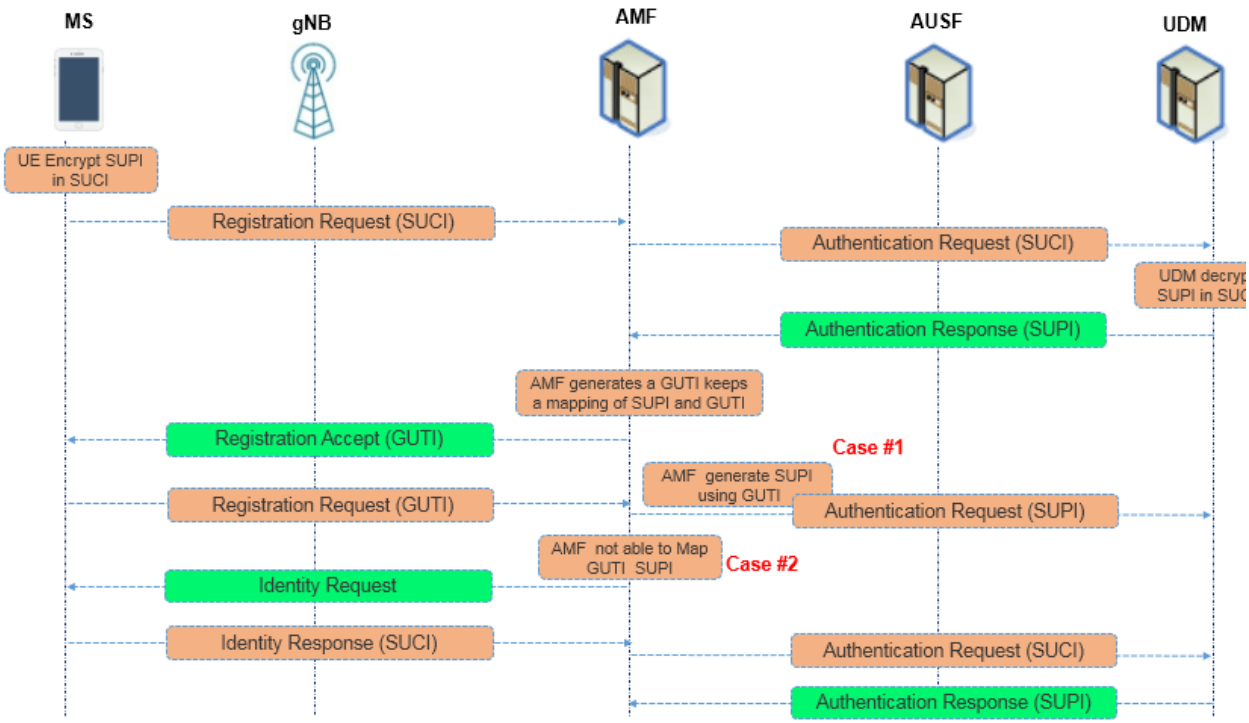

Figure 42: 5G identity exchange between UE and network [27]

UE sends a registration request using GUTI in a future registration request. There are now two conceivable outcomes.

1. AMF is capable of generating SUPI via GUTI and SUPI mapping.
2. AMF is unable to generate SUPI.

In the first scenario, AMF generates SUPI with GUTI and then uses SUPI to authenticate with AUSF. In the second situation, if the UE cannot be identified via GUTI at AMF, AMF will send an identification request to the UE, and the UE will respond with an Identity Response including the SUCI.

The Challenges

Numerous obstacles confront operators and developers despite the enormous potential. Radio access networks (RANs) must be modified in order to enable complete E2E network slicing, which includes RAN deployment. Even though standard progress is being made, there is still no industry consensus on how to implement network slicing on 5G networks with other architectural features.

Operators are put under even more stress as a result of the addition of additional networks to the same physical infrastructure. Managing spectrum slicing and allocation for extremely dynamic

circumstances, as well as maintaining SLA, QoS, and security assurance for each individual slice, are common issues.

Security

Security aspects of the network are also impacted by the increased complexity. In addition to its own device authentication to verify users, each slice will have its own security requirements that are proportional to the use case it is designed to support. Like the Internet of Things, network slicing's scalability factor provides billions of new attack vectors. A successful attack from a single point of 5G network management could simultaneously penetrate multiple network slices and/or domains. [29]

The roles and responsibilities of slice operators and organizations must be explicitly defined in order to effectively address security issues associated with network slicing. In order to meet this demand, new security solutions like micro-segmentation are developing. Security will be a constant focus for 5G operators, who intend to invest a significant amount in security solutions prior to and following commercial deployment. As a result, security-as-a-service (SECaaS) providers can now offer comprehensive solutions that benefit everyone.

2.10 5G NR Radio Protocol Stack

5G-NR The user plane has the same Phy, MAC, RLC, and PDCP as LTE and has added a new layer called SDAP (Service Data Adaptation Protocol).

On the other hand, the 5G-NR control plane is identical to LTE, with AMF serving as the MME equivalent node (Access and Management Mobility Function).[30]

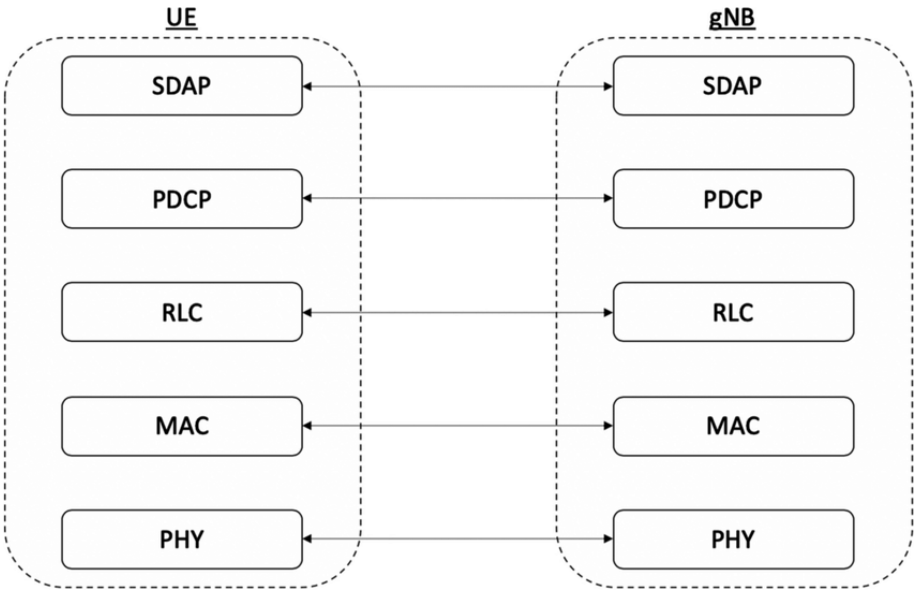

Figure 43: User plane protocol stack

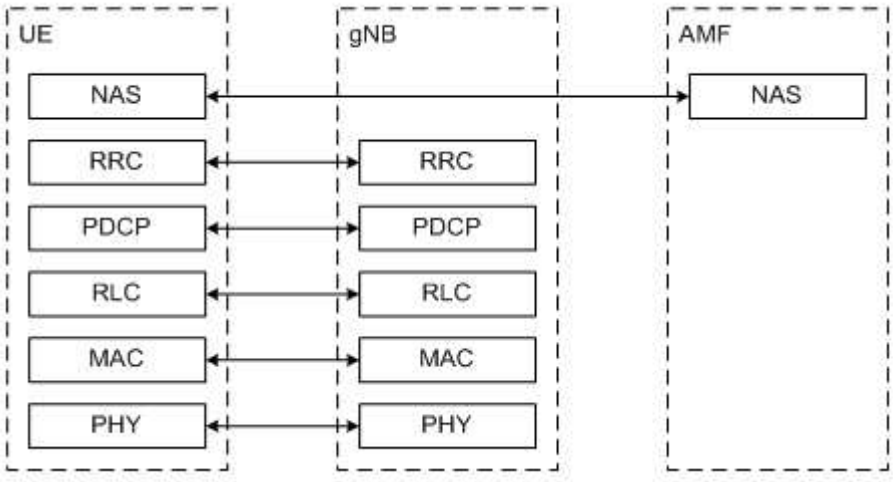

Figure 44: Control plane protocol stack

5G-NR Layer 3 (RRC) Functions:

The RRC sub layer's primary services and functions include:

- UE measurement reporting and control
- QoS management functions
- Radio link failure detection and recovery
- NAS message transfer between UE and NAS
- Paging that is initiated by 5GC or NG-RAN
- Establishment, maintenance, and release of an RRC connection between the UE and NG-RAN, including the addition, modification, and release of carrier aggregation, Dual Connectivity in NR, or between E-UTRA and NR, and the addition, modification, and release of Dual Connectivity.
- functions related to security, like managing keys
- establishing, configuring, securing, and disabling Data Radio Bearers (DRBs) and Signaling Radio Bearers (SRBs)
- Handover and context transfer are mobility functions; UE cell selection and reselection, in addition to control of cell selection and reselection; and mobility between RATs.

5G-NR Layer 2 Functions:

The layer 2 of NR is split into the following sub layers:

- Radio Link Control (RLC)
- Medium Access Control (MAC)
- Service Data Adaptation Protocol (SDAP)
- Packet Data Convergence Protocol (PDCP)

The two figures below depict the Layer 2 architecture for downlink and uplink, where:

- The physical layer provides transport channels for the MAC sublayer.
- Logicial channels are provided to the RLC sublayer by the MAC sublayer.

- RLC channels are provided to the PDCP sublayer by the RLC sublayer.
- The PDCP sublayer offers to the SDAP sublayer radio conveyors.
- The SDAP layer provides 5GC QoS flows.
- Comp. refers to segment and header compression. to splitting up
- Control channels (for clarity, BCCH and PCCH are not depicted).

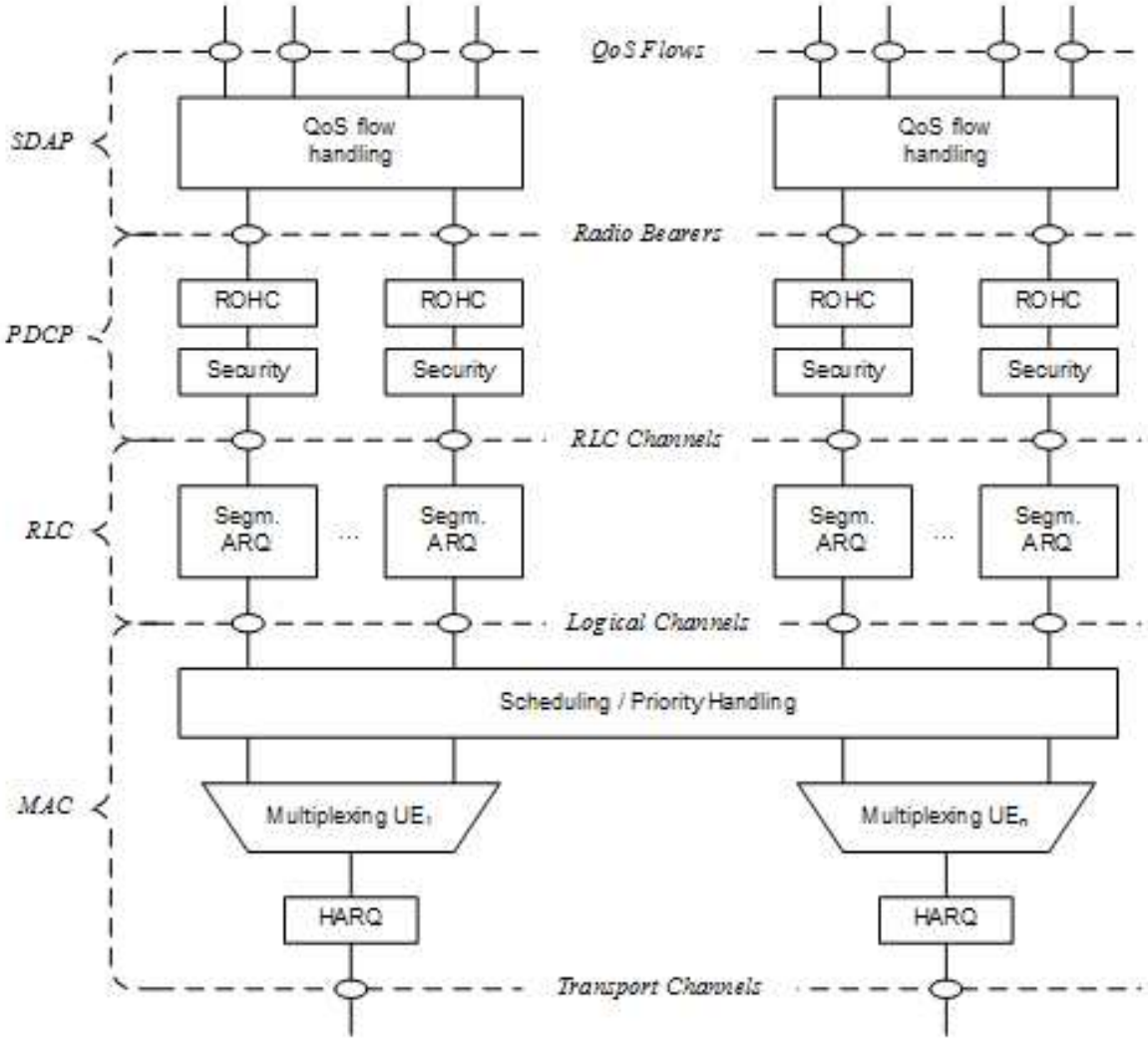

Figure 45: Downlink Layer 2 Structure

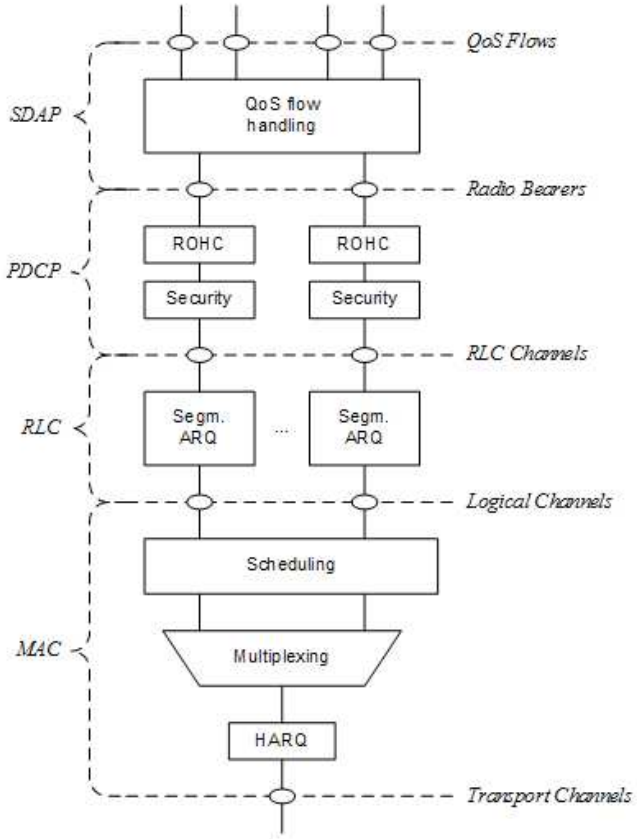

Figure 46: Uplink Layer 2 Structure

SDAP (Service Data Adaptation Protocol) Protocol Functions :

The main services and functions of SDAP include:

- Mapping between a QoS flow and a data radio bearer (Due to new QoS framework)
- Marking QoS flow ID (QFI) in both DL and UL packets (DL: due to reflective QoS and UL: due to new QoS framework)

Except for DC, where two SDAP protocol entities can be specified, each individual PDU session is configured with a single SDAP protocol entity.

PDCP (Packet Data Convergence Protocol) Layer Functions:

The following are the primary services and functions of the PDCP sublayer for the user plane:

- Sequence Numbering
- Header compression and decompression: ROHC only
- Transfer of user data
- Reordering and Duplicate detection (if in order delivery to layers above PDCP is required)
- PDCP PDU routing (in case of split bearers)

- Retransmission of PDCP SDUs
- Ciphering and Deciphering
- PDCP SDU discard
- PDCP re-establishment and data recovery for RLC AM
- Duplication of PDCP PDUs

The following are the primary services and functions of the PDCP sublayer for the control plane:

- Sequence Numbering
- Ciphering, deciphering and integrity protection
- Transfer of control plane data
- Duplicate detection
- Duplication of PDCP PDUs.

RLC (Radio Link Control) Layer Functions:

The RLC sublayer's principal services and functions are dependent on the transmission mode and include:

- Transfer of upper layer PDUs
- Sequence numbering independent of the one in PDCP
- Error Correction through ARQ
- Segmentation and re-segmentation
- Reassembly of SDU
- RLC SDU discard
- RLC re-establishment

MAC (Media Access Control) Layer Functions

The main services and functions of the MAC sub layer include:

- Mapping between logical channels and transport channels
- Multiplexing and demultiplexing MAC SDUs belonging to one or more logical channels into or out of transport blocks (TB) delivered to the physical layer on transport channels
- Scheduling information reporting
- Error correction via HARQ
- Priority handling between UEs via dynamic scheduling
- Priority handling between logical channels of a single UE via logical channel prioritization
- Padding

A single MAC entity can support one or more numerologies and/or TTI durations, while logical channel prioritisation mapping limits control which numerology and/or TTI duration a logical channel can employ.

3 Edge computing

3.1 Edge Computing

Edge computing is a theory of networking that focuses on reducing latency and bandwidth consumption by placing processing as close to the data source as possible. In layman's terms, edge computing is the process of moving fewer processes from the cloud to local locations like a user's computer, an Internet of Things device, or an edge server. 31] The amount of long-distance transmission between a client and server is reduced by bringing processing to the network's edge.

What is the network edge?

An Internet device or the local network that houses the device communicates with the Internet at the network edge. The term "edge" is a bit hazy; The user's router, internet service provider, or local edge server can also be referred to as the network edge, as can a user's computer or the processor in an IoT camera. The most important takeaway is that, in contrast to origin servers and cloud servers, which may be extremely far from the devices with which they connect, the network's edge is geographically close to the device.

What differentiates edge computing from other computing models?

The first computers were big, bulky machines that could only be used directly or through terminals, which were basically extensions of the computer. The use of computers may spread significantly as personal computers become more widely available. In the past, personal computing was the most popular computing model. On the user's device or, in some instances, in an on-premises data center, programs were run and data was saved locally.

The relatively recent development of cloud computing offered some advantages over on-premise computing. Cloud administrations are united in a merchant made due "cloud" (or set of server farms) and are open through the Web from any gadget.

However, cloud computing may cause delays because of the distance between customers and the data centers that host cloud services. Edge computing brings computers closer to end users, preserving the centralized nature of cloud computing while reducing the distance that data must travel.

In conclusion:

- Early calculations: applications that were centralized and ran on a single isolated computer
- Personal computing: local, decentralized applications
- Cloud computing: applications that were centralized and ran in data centers.
- Cloud computing: programs that are centralized and run close to users, either on the device or at the edge of the network.

Example of edge computing

Imagine a building that is guarded by dozens of high-definition Internet of Things video cameras. These "dumb" cameras simply send a raw video signal to a cloud server on a regular basis. A motion-detection application on the cloud server processes the video output from all of the cameras to ensure that the server's database only contains clips that contain activity. Because of the large

amount of video footage that is being transferred, the building's Internet infrastructure is under constant and significant strain. In addition, processing video from all of the cameras simultaneously puts a lot of strain on the cloud server.[31] Consider moving the motion sensor calculation to the edge of the network. What if each camera uploaded footage to the cloud server as needed while running the motion-detecting software on its own internal computer? The amount of bandwidth used will be significantly reduced because a large portion of the camera footage will never need to be transferred to the cloud server.

Additionally, since the cloud server would now be solely responsible for storing the crucial footage, it would be able to communicate with more cameras without becoming overloaded. Edge computing appears in this manner.

3.2 Other possible use cases for edge computing

It is possible to incorporate edge computing into a wide range of products, services, and applications. The possibilities include:

- System monitoring for security: As stated previously.
- IoT gadgets: Smart devices that connect to the Internet may benefit from running code on the device itself rather than in the cloud for better user interactions.
- Autonomous cars: Instead of waiting for commands from a server, autonomous vehicles must respond immediately.
- Better caching: An application can modify how content is cached in order to provide it to users more efficiently by running code on a CDN edge network.
- Medical equipment for monitoring: Medical equipment must respond immediately rather than waiting for a cloud server to respond. Conferencing via video: Moving backend processes closer to the source of the video can reduce lag and latency because interactive live video uses a lot of bandwidth.
- Cloud gaming is a new type of gaming in which the game itself is processed and hosted in data centers and delivered live to devices.

3.3 Benefits of edge computing

- **Cost efficient.**

As the preceding illustration demonstrates, edge computing contributes to the reduction of bandwidth and server resources. Cloud resources and bandwidth are limited and costly. By 2025, over 75 billion Internet of Things (IoT) devices will be in use worldwide, with smart cameras, printers, thermostats, and even toasters in every home and workplace. To support all of those devices, significant amounts of processing will need to be moved to the edge.

- **Performance**

Reduced latency is another important benefit of pushing operations to the edge. A delay occurs whenever a device needs to communicate with another remote server. Because each message must be routed out of the building, communicate with a server located someplace else in the world, and be returned before it appears on the recipient's screen, two employees in the same workplace using an instant messaging platform may experience a significant delay. There will be no noticeable delay if that procedure is moved to the edge and the company's internal router is in charge of transferring intra-office communications. [31]

In a similar vein, users of a variety of web applications will experience delays whenever they encounter operations that necessitate communicating with an external server. These delays can be completely eliminated by moving more operations to the network edge, but their duration will vary depending on the server's location and available bandwidth.

- **New functionality**

Edge computing can also provide capabilities that were previously unavailable. Edge computing, for instance, can be used by a business to process and analyze data at the edge in real time.

In conclusion, the primary benefits of edge computing are as follows:

- Reduced latency
- Reduced bandwidth utilisation and associated cost
- Reduced server resources and associated cost
- Increased functionality

3.4 Drawbacks of edge computing

The fact that attack vectors are expanded by edge computing is one of its drawbacks. Bad attackers now have new methods for infiltrating "smart" devices like edge servers and Internet of Things devices with powerful built-in computers.

Edge computing's requirement for additional local hardware is yet another drawback. An IoT camera, for instance, requires a built-in computer to transfer raw video data to a web server; however, running its own motion-detection algorithms would require a significantly more powerful computer with more processing power. By the by, as equipment costs fall, it turns out to be more reasonable to develop more astute contraptions.

One way to completely eliminate the need for additional hardware is to use edge servers. Thanks to Cloudflare's network of 275 edge facilities spread across the globe, Cloudflare users, for instance, can have edge code running globally with Cloudflare Workers.

4 Cloud gaming:

Introduction:

Computationally complex games are run on powerful cloud servers, rendered game scenes are streamed to gamers using thin clients on heterogeneous devices, and control events from input devices are sent back to cloud servers for interaction in a novel way to deliver computer games to users. The operation of cloud gaming services is shown in Figure 1. A cloud gaming platform is hosted in one or more data centers on cloud servers. The computer game programs that run on the cloud gaming platform are broken down into two main parts:

(i) game logic, which is in charge of converting player commands into interactions in the game
(ii) a scene renderer, which is in charge of creating game scenes in real time. The command interpreter issues the gamer commands, and the video capture converts the game scenes into videos, which are then compressed by the video encoder [32].

The command interpreter, video capturer, and video encoder are all part of the cloud gaming platform. The cloud gaming platform, as depicted in this diagram, receives user inputs and provides video frames to gamers' thin clients. It only requires two low-complexity components, making it a thin client: I an order collector that connects to game regulators, for example, gampads, joysticks, consoles, and mice, and (ii) a video decoder that can be accomplished utilizing efficiently manufactured (minimal expense) decoder chips. Real-time computer games are difficult to maintain because interactions between the cloud game platform and thin clients use the best Internet available.

Cloud gaming services were first offered by startups like OnLive, Gaikai, G-cluster, and Ubitus in the latter part of the 2000s. Additionally, we witnessed the prominent game console developer SONY acquire Gaikai. Following this, competition between Nvidia's Grid Game Streaming Service and Sony's PlayStation Now (PS Now) heated up the cloud gaming market even more. The number of people using cloud gaming increased from 30 million in 2014 to 150 million in 2015, as estimated by Strategy Analytics in 2014. The same report asserts that additional significant game console manufacturers will soon enter the cloud gaming industry.

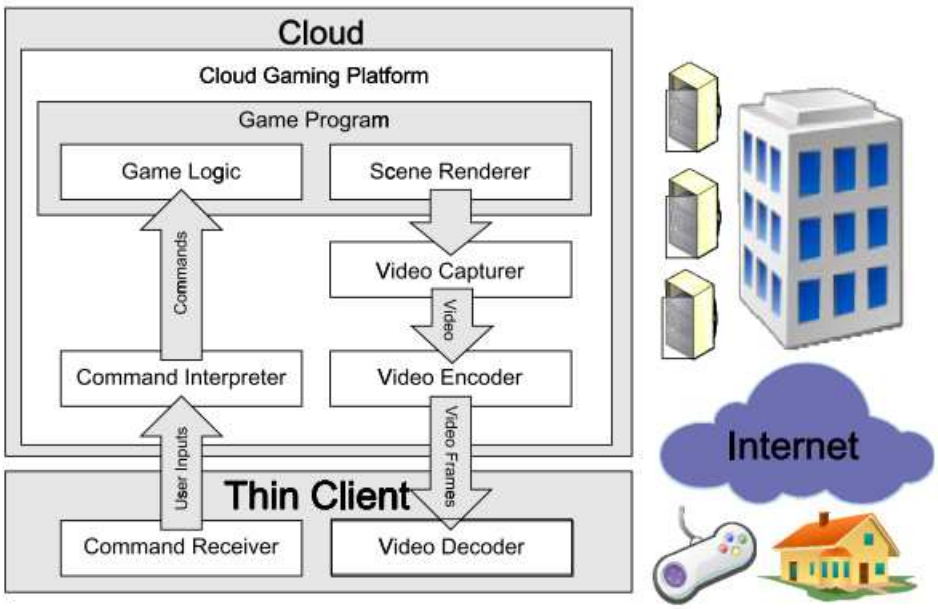

Figure 47: Cloud gaming architecture [34]

There are a number of reasons why cloud gaming is so popular, including the potential advantages it offers gamers, game developers, and service providers. Gamers can use cloud gaming to:

- Enjoy unique features like migrating across client computers during game sessions, observing Ongoing tournaments, and sharing game replays with friends.
- Purchase or rent games on-demand.
- avoid regularly upgrading their hardware.

Game designers now have the ability to:

- Concentrate on a single platform to save money on testing and porting.
- Avoid retailers for higher profit margins.
- Reach out to more gamers and avoid piracy because client computers never receive the game software.

Cloud gaming: for service providers:

- Demonstrates the potential of other/new remote execution applications.
- Increases demand on already-deployed cloud resources.
- Leads to new business models.

because cloud gaming places the greatest demands on a variety of networking and computational resources. Before the research community can fully realize its potential to attract more gamers, game developers, and service providers, the numerous benefits of cloud gaming must be overcome. 32] The following is a summary of the most significant aspect. To start, cloud gaming stages and testbeds should be laid out to direct broad execution assessments. Quality of Experience (QoE) data, such as player perceptions of the experience, and Quality of Service (QoS) indicators like energy consumption and network metrics are included in the evaluations. Assessing the intricate interplay between QoS and QoE metrics is much more difficult than creating platforms and testbeds, designing test scenarios, and

conducting evaluations. Second, the emerging platforms and assessment methods make it possible for the research community to optimize a variety of components, such as communication routes and cloud servers. More specifically, optimization techniques can help cloud servers improve resource allocation, distributed design, efficient content coding, and adaptive transmissions through communication channels. Third, there are various game classifications in PC games. There are two categories of these genres: perspective and focus. A gamer's point of view is how they see the game's action. It is in charge of how the produced video looks on the screen. The majority of perspectives are first-person, second-person, third-person, and omnipresent. Graphical viewpoints are used in first-person games like Counter-Strike, which are told from the point of view of the characters in the game. Grand Theft Auto and other second-person games allow players to see the characters on the screen by showing them from behind. The players' perspectives in third-person games are fixed on 3D scenes that are projected onto 2D areas. In current third-person video games, the sky perspective, also known as the "God view," is frequently employed. Examples include Diablo, Command & Conquer, and FreeStyle, among other classic third-person games. Last but not least, omnipresent gives players full command of viewpoints of the region of interest (RoI) from a variety of vantage points and distances. Age of Empires 3, Stronghold 2, and Warcraft III are just a few of the contemporary war games that fall under this category. How players react to the game's content is influenced by the game's theme. Common themes include sports, shooting, fighting, turn-based role-playing games (RPGs), action role-playing games (ARPGs), turn-based strategy, real-time strategy (RTS), and management simulation. Albeit the perspective might be restricted by the game point, a game kind can be characterized by a blend of perspective and topic, like first-individual shooter, third-individual ARPG, pervasive RTS, etc. The most challenging games for cloud gaming service providers are fast-paced first-person shooters with a lot of scene complexity. On the other hand, turn-based third-person RPG games are better suited for cloud gaming because they are less susceptible to delays.

4.1 Cloud Gaming Based on 5G and Edge Computing

A technology called "cloud gaming" lets you play games that are hosted on distant servers. The video and audio for the game are streamed to the player's device by a powerful server. The player uses a controller, keyboard, mouse, or other similar input device to interact with the game.

The fact that no software or hardware is installed locally gives cloud gaming its name. Instead, gamers connect to a centralized service that runs on remote servers in global data centers. This could mean that your home computer isn't powerful enough to play games like Fortnite that require more graphics processing power. However, you are able to access cloud services that will enable you to play them regardless.[33] The most significant advantage of cloud gaming is that it enables you to play games without having to purchase costly hardware. As a result, you can play on desktop computers with low specs as well as mobile devices like phones and tablets. Creators benefit from cloud gaming as well because they are able to sell their games as a service rather than dealing with physical copies and downloads.

4.2 Features of cloud gaming

With cloud gaming, you can play games on any computer or mobile device with an internet connection. It's a great option for people who don't want to spend a lot of money on a big computer but still want to play their favorite games in high quality. The adaptability of cloud gaming is its primary benefit. If you are traveling and do not have access to your home computer, all you need to do is download the file once. Giving up your favorite gaming collection is not necessary.

Customers can also get remote access to new content upgrades through cloud gaming before the company makes them available locally offline via hard drives and storage devices. Hard drives for Sony Corporation's PlayStation and Microsoft Corporation's Xbox One are examples of offline devices.

Internet games with multiple players will greatly benefit from this. For instance, Fortnite necessitates frequent updates to keep players informed of the most recent game environment updates and modifications. [33]

Cloud gaming enables gamers to simultaneously play their favorite games on multiple devices. It's especially helpful for people who want to play online games on their smartphones while they wait in line but don't want to miss important updates. In addition, players can access the same game collection from any location. They only require a device that can run Linux and an internet connection, whether they are at home or.

4.3 How does cloud gaming work?

A method called cloud gaming lets you play video games without having to install them on your computer or console. Instead, the game is broadcast to you while you play on a remote server. This means that you can play games of high quality on any device, regardless of how old or slow it is.

Cloud gaming streams games to your smartphone or computer from a large, centralized server. The process is similar to streaming movies or music, but you need a fast connection.

It uses the same technology as other streaming services like Netflix. The game sends commands to a cloud server over the internet in response to constant input from your device. After that, the server returns an image of the current situation, which your device displays as quickly as it can on screen.

This means that every time a player enters data into a game, it is recorded. Whether they are pressing buttons with their mouse or controller, navigating their character in first-person, or conversing in chat. Milliseconds are used to describe everything. Something is wrong with the connection if there is any delay between when a player inputs and when the image appears on the screen.

The cloud gaming server delivers the game directly to your computer or device. Long download times and large files are no longer necessary thanks to this. You won't have to worry about downloading and installing games to your smartphone because of this.

4.4 General Architecture of a Cloud Gaming System (CGS)

There are three main categories of real-time systems for remote rendering. Video Streaming, 3D Graphics Streaming, and Video Streaming with Post-Rendering Operations In a video streaming system, the server renders the 3D commands, converts them to 2D, and then transmits the video stream to the client. In a 3D graphics system, the client interprets the graphics-related orders and draws the scene accordingly. Between the first two, the third system uses Thin Client to handle low-processor-intensive tasks while the server does the heavy lifting of rendering 3D images.

For distributed gaming systems, a number of thin client strategies have been proposed. There are two categories of them: image-based systems and systems based on instructions. The primary difference between the two is that in instruction-based systems, only the instructions for creating graphics that correspond to a control event are sent over the network. In image-based systems, however, the entire

computationally demanding rendering of the game scene is done on the server side and sent over the network in the form of a video stream. Because Cloud Gaming does not require the client to have computer capabilities, all CGSs utilize image-based thin client architectures.

A CGS's general image-based architecture is the subject of this article.

Client Thin: A Client Cooperation module and a video decoder are incorporated. All control movements made by the end user with a mouse, keyboard, or other input device are collected by the User Interaction module. The video decoder plays the video that is being streamed by the server in response to the player's movements. [34]

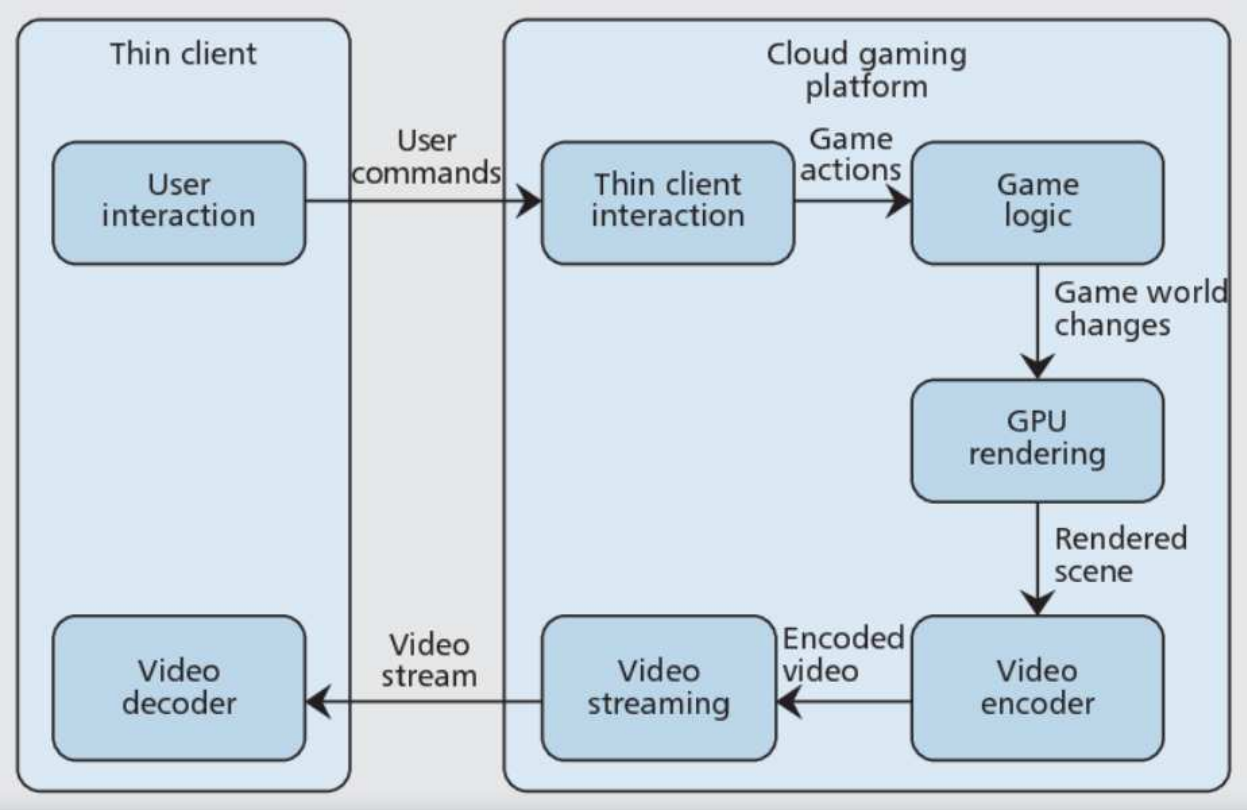

Figure 48: Thin client architecture [34]

Cloud Gaming Platform: There are four modules to it: Game Logic, the GPU Renderer, the Thin Client Interface, and the Video Encoder. All client commands are handled by the Thin Client Interaction module, which also converts network messages into game actions. In the game world, the game logic module interprets them before the GPU renders the scene. The scene is then compressed by the encoder and sent to the thin client by the Video Streaming service, which then decodes the stream and plays the video frames to the client. To provide various CGS implementations, a variety of design options can be investigated:

- The method by which the encoded game screen is delivered to the client.
- The manner in which the existing game software is altered and executed on the server.
- The manner in which the game screen is encoded and decoded (on the server)
- How to deal with short-term network instability to keep the game responsive and good looking.

4.5 Cloud Gaming Architectures

Cloud gaming systems often employ powerful servers hosted in data centres to operate and transmit video games to players via the internet. The following are the fundamental components of a cloud gaming architecture: [34]

- Game servers: These are the servers that actually run the games and are in charge of processing game logic and rendering game images.
- Network infrastructure: This refers to the network gear and software used to transfer data between game servers and user devices.
- User devices: These are the computers, smartphones, and gaming consoles that players use to access and play the games.
- User interface: The software that enables users to access and engage with games, such as a web browser or a specific app.

Users' devices do not need to have strong CPUs or graphics cards in a cloud gaming architecture because the game processing and rendering is done on the game servers. Users can now play high-quality games on a variety of devices. Users' devices, however, must have a reliable internet connection and enough capacity to stream the games.

Cloud gaming architectures are multiple techniques to distributing gaming application components across a number of geographically distant devices. Table 3 shows a high-level classification of cloud gaming systems based on game application component distribution. These categories are also covered in depth in the next section.

Cloud gaming model	**Components of game at Client**	**Components of game at Cloud**
Remote Rendering	Input Controller	Game Logic, Networking, Database, Video Renderer
Local Rendering	Input Controller, Video Renderer	Game Logic, Networking, Database
Cognitive	Dynamically decided	Dynamically decided

Table 3: Abstract architectures for a cloud gaming application

4.5.1 Remote Rendering Model:

In this model, the client gets client input, makes an interpretation of it to orders, and conveys them to the game server. The architecture allows users to interact with the gaming application and game logic with input controllers. The video rendering unit renders scenes' frames as the game scenes are updated. The frames are then delivered to the client via the internet [21][5]. Customers who lack the hardware necessary to produce high-definition game graphics can benefit from the model's intelligent solutions. Due to the constant transfer of game video frames or scenes over the Internet, this tactic may, regrettably, cause a bottleneck. For instance, a basic first-person shooter game needs a lot of consistent data because it has a frame rate of about 35 frames per second.

4.5.2 Local Rendering Model:

The video renderer and input controller can both sit at the client because this approach supports component divisions. The cloud is where additional game components are kept. This strategy reduces the amount of bandwidth required by rendering the scenes on the client side. This architecture is comparable to the remote rendering model up until the point where input commands alter game scenes on the cloud

server. However, instead of rendering scenes and sending packets, output instructions are formed and sent to the client. Before rendering the scenes and presenting them to the end user, the client receives and interprets these instructions [2]. Rendering instructions must be provided for each frame using a standard tool like OpenGL, even though this paradigm reduces traffic.

4.5.3 Cognitive Approach:

In this way, the game's client and cloud divisions are determined dynamically on behalf of user resources (such as computation, rendering power, or bandwidth). Cai et al. This strategy, which divided a cloud gaming application into multiple parts, was suggested by [22][23][20]. Game components could be transferred from the cloud to the client by the application. However, the program had to be divided into components in such a way that dependency issues could be resolved in order to achieve this collaboration (on-loading and off-loading components). In addition, neither the dynamic decision-making approach for on-loading nor off-loading actions nor the control mechanism for the components were discussed in depth in the study.

The architecture that is suggested in this work does not make use of remote rendering. In contrast to the cognitive method, this work does not dynamically adapt to user resources. In contrast to standard local rendering solutions, this one does not send any data at all and the instructions it sends have a longer time horizon.

4.6 Proposed Architecture

Our architecture is designed to execute game components independently on client machines without requiring remote rendering. We suggest a three-part architecture to accomplish this. Data transfer between gaming clients and the cloud server (s), as well as client-end and cloud-end Figure 1 portrays the high-level design. At the beginning of a game on the Client End, the Cloud End provides initial information that is sufficient to operate the game for some time without consulting the Cloud End. An Instruction Set serves as the delivery medium for this data. To get information about the next steps, completing a game stage requires a connection with Cloud-End. 35] The following sections provide an explanation of the architecture's various components:

Client-End: The client module consists of two subcomponents in addition to the game's fundamental logic.

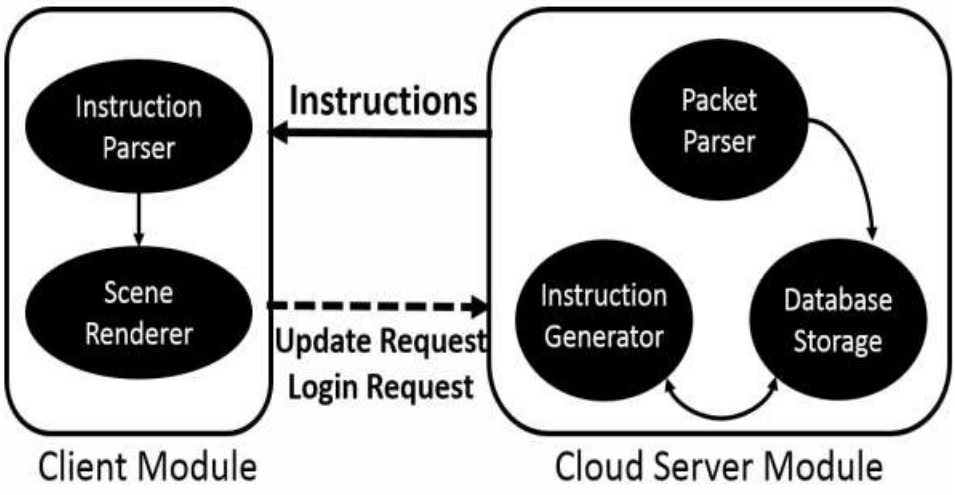

Figure 49: Generic Proposed Architecture

Instruction Parser: This is in charge of parsing the server's data (instructions) to find the information needed to render the game scenes.

Renderer of Scenes: Scenes are rendered using data from the first module in this module. Additionally, it runs a portion of the game without having to communicate with servers and records user actions. When necessary, the Scene-Renderer contacts the Cloud-End to obtain game statistics.

Server-End: The game's supported instruction set, user data, and gaming statistics are all stored in the Server-End. Additionally, there is a module in the Server-End that sends game instructions to the Client-End as needed while keeping an eye on the client's current game state.

Client-server data in video games: Since we want to run different parts of the game independently on the client, we use an instruction set rather than remotely rendered game scenes. In contrast to the local renderer, the instructions do not include OpenGL commands or any other similar tool commands that represent an entire scene. Instead, the instruction set is designed so that rendering a large number of scenes requires only a small amount of data transmission and does not necessitate constant communication with the Server-End. A proof-of-concept game and an example instruction set are discussed in the following section.

4.7 Various QOS Parameters

A CGS's success can be evaluated from a number of different angles. Resource allocation is important from the service provider's perspective, but metrics that affect the gaming experience are important from the end user's perspective. The time scale can also be used to measure QOS. It can be smaller during individual game sessions and larger over multiple gaming sessions. Because the majority of CGSs use a single Virtual Machine without offloading to serve each client, we will only examine systems with a limited timeframe here. [40] These are the primary metrics:

1. Characteristics of traffic: a single gaming session's bandwidth consumption. The payload size and packet rate (uplink and downstream) are also specified.

2. Latency: When evaluating CGS performance, this is probably the most significant parameter. It encompasses all of the distinct latencies experienced by various components and is referred to as the system's response time.

3. Image Quality: The user's Quality of Experience (QOE) is directly impacted by the quality of the images and videos sent across the network. Additionally, quality changes over a variety of network conditions are measured. Typically, the statistic Frame rate (or FPS) is used to evaluate streaming quality. Graphic quality can be evaluated using the SSIM metric or the Peak Signal to Noise Ratio (PSNR) method.

Latency is a measure of Response Delay, which is the time between when a user issues a command and when the associated game frame is displayed to the user. Latency quantifies the responsiveness of the game.

There are four distinct delays in the Response Delay (RD): Delays in the network (ND), processor (PD), game (GD), and playout (OD).

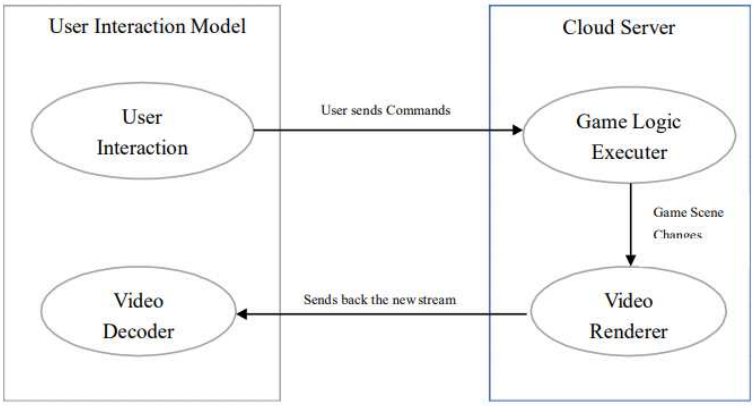

Figure 50: General framework of CGS

Network Delay (ND): The network's round trip time, which measures the time it takes for a client's instruction to reach the server and the time it takes for the game screen to appear on the client.

Playout Delay (OD): The amount of time it takes for the client to receive, decode, and play the current frame.

Processing Delay (PD): The amount of time it takes the server to receive and process the user's command. It also contains the time required to encode and packetize the client's current frame.

Game Delay (GD): The amount of time it takes the game software to execute the user's command and generate the accompanying game frame. This is commonly assumed to be the same in the cloud context, therefore GD in standalone gaming = GD in cloud gaming.

RD = ND + PD + GD + OD

ND can be measured using ICMP pings, Wireshark, or other network tools, whereas GD is game dependent. Assessing PD and OD is difficult because they happen internally at the server and client, respectively.

As previously said, RD is a crucial aspect in determining the quality of the CGS, and several studies on delay tolerance for various types of games have been conducted. The findings can be summarised as follows:

Example Game Type	Perspective	Delay Threshold
First Person Shooter	First Person	100 ms
Role Playing Game	Third Person	500 ms
Real Time Strategy	Omnipresent	1000 ms

Table 4: Delay tolerance in traditional gaming

5 Software Defined Networking

SDN is an architecture that separates the control plane and the data/infrastructure plane from one another, allowing us to program the network to meet our requirements [1]. This makes networks more programmable and flexible. Because it enables users to address various changes as they occur, SDN is important in Cloud Computing. It simplifies network design and enhances monitoring and performance, two of its primary benefits. SDN makes it possible for networks to be adaptable and responsive to shifting network demands, which has resulted in a number of advantages for cloud computing and cloud storage services when combined with automation and virtualization. However, the only obstacle they face is the limitations imposed by networks. [36].

5.1 Software Defined Network (SDN) Architecture

The SDN Architecture consists of three main layers:

SDN Applications make up the application layer. These applications include all of the enterprise's network applications, like load balancers, firewalls, and intrusion detection systems. In a traditional network, all of this might be controlled by a physical device; however, SDN lets us use an application to monitor and manage their operation.

An SDN controller application makes up the control layer. It oversees the flow of network traffic and all network policies.

All of the network's physical components, like switches, are in the infrastructure layer, as the name suggests. [2].

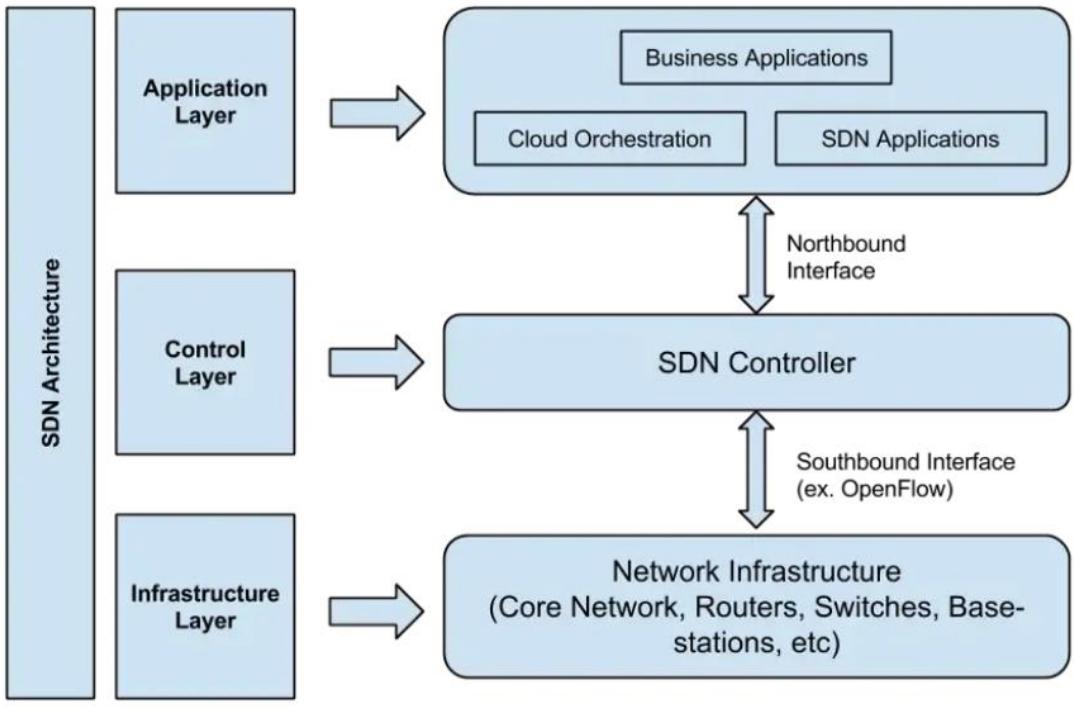

Figure 51: SDN Architecture

5.2 Components of Software Defined Network (SDN)

There are 4 parts of Programming Characterized Organization:

The SDN application acts as a link between devices, network resources, and the SDN controller via a Northbound API.

Controller SDN: The SDN controller sends the requirements of the SDN Application to the SDN Datapaths. Any network policies that the SDN Applications might require are also included in this. Additionally, it provides them with an overview of the network's traffic. The controller interfaces with the SDN Datapath/Data plane via the Southbound API [37].

Datapath in SDN: It gives data packets a place to go on the network by establishing switches.

API SDN: The SDN Application Programming Interface (API) gives the SDN Controller and network routers a way to talk to each other freely.

Software Defined Network Implementation: The Software Defined Network can only be implemented in the following ways:

Find a problem: Before anything else, we have to figure out a problem that might need to be solved using SDN. When we fully transition to SDN, we will be able to measure and implement the results if we do this.

Create a team with a variety of skills: We need a team with a variety of skills because SDN has a lot of moving parts and people with only one area of expertise won't be able to do it well.

Examine a subset: To ensure that we do not suffer significant losses in the event of a fault, we should first test SDN in a low-risk area of the network prior to fully implementing it.

Evaluate: We must examine the data following the SDN implementation to determine whether it produced the anticipated results; We should only implement for the entire network after that.

5.3 SDN Infrastructure w.r.t Cloud Computing

There are three main SDN Infrastructures:[38]

1. SDN without Cloud Computing

Even though it is extremely difficult to find SDN deployment without Cloud Computing today, we may find it in fairly large businesses with a lot of networking devices that just need a centralized location, probably the IT department, to control the changes that are happening to the network in real time. This layout might be good for managing the company's network, but the client might have to ask the IT department for resources, which could cause a server bottleneck.

2. **Cloud Computing without SDN**

This layout is used by the majority of recent cloud computing models. Because the server manages the new servers and virtual machines on its own, users can easily manage rising demand without worrying about scalability. However, in order to implement SDN, we will need to wait for a number of networking configurations to be met, which may not be feasible for small businesses seeking a quick solution. This can also be used in private clouds, where security isn't as important, so SDN isn't necessary in conjunction with cloud computing because it would be too much.

3. **SDN along with Cloud Computing**

As a result, the company might decide to combine SDN and Cloud Computing because they might need to coordinate cloud resources from a single location. These services aren't as common as they used to be, but big companies are using them because they have cost savings, better performance, and more security, among other benefits that will be discussed below. [5].

5.4 SDN Based Cloud Network

In an SDN-based cloud environment, the OpenFlow Protocol is used to communicate between the controller and the data plane. It allows network administrators to remotely manage routing tables and centralized packet-switching decisions, which allows switches and data centers to be programmed independently and decouples the network from individual switches.

Because the controller can be either a physical or cloud-based entity, this protocol also has a virtualization component. An OpenFlow switch, which can be software or hardware, is what runs the OpenFlow protocol. OpenFlow switches can be connected to network devices to provide access to SDN capabilities. OpenFlow switches manage the Data and Control paths independently, whereas conventional switches manage both in the same device. The Controller layer is hosted on a separate server, while the Datapath layer is present inside the switch. The OpenFlow protocol is then used to communicate with the controller. This protocol stores information like forwarding table modifications and sent and received packets. [9]

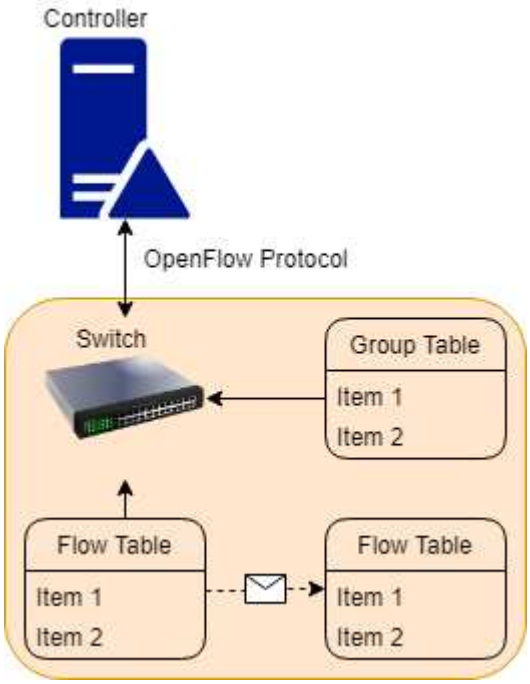

Figure 52: Communication between switch and controller [38]

In the figure above we can see the communication happening between the controller and the switch, the Flow Tables communicate with one another and share information such as set of instructions, matching fields and matching packets etc.

5.5 Functionality of Software Defined Network w.r.t. Cloud Computing

The data plane determines how packets are transferred, i.e. which route must be taken through the switches, while the control plane, as previously stated, determines how traffic flows within the network. The control plane layer's (centralised controller) functionality kicks in when a packet reaches a network switch. The policies and rules it provides to the firmware of the switch determine where the packet will be sent.

The switch relies on the controller for policies because it is unable to make decisions on its own. It also sends the controller information about the traffic it has handled. The same protocols are utilized because all packets destined for the same destination are treated in the same manner.

A virtual cloud network runs tiered on top of the physical network when SDN is combined with a cloud network and a virtualization component is included. In a virtualized environment, this could be done to separate the physical network from network traffic. Cloud Vendors operating multi-tenant cloud environments greatly benefit from this type of network segmentation because it enables them to design a distinct set of policies and regulations for each of their tenants. [2].

5.6 Implementing SDN on the Cloud Implementing

It is impossible to implement SDN in the cloud without taking into account every minute detail; Because it is not a static system and updates must be carried out while the system is running without interruption, we need to carefully plan out the entire implementation. The following factors must be taken into account when planning the Cloud SDN implementation:

1. Look for cloud service providers that have a framework that is strong enough to support SDN on their platform.

We must first identify Cloud Vendors whose architecture supports SDNs before we can implement SDN in the cloud. We must move the SDNs to a third-party cloud service once they have been identified. This third-party cloud service can be any public cloud service like Google Cloud Platform (GCP) or Amazon Web Services (AWS). We will be able to offer load balancers, firewalls, and other Cloud-based services that tenants can scale to fit their needs through this type of service.

2. In order to create a dynamic network, you should be aware of the application requirements.

Migrating SDNs to the Cloud is preferred because of the network's ability to respond to application requests and recognize shifting demands. It will be extremely difficult to make the network dynamic without a thorough understanding of the applications' requirements. All programmable aspects that have an impact on network scalability, throughput, reaction time, and security must be specified. The network will be most effective when it is made to be the most dynamic.

3. Determine the method of deployment We can deploy our SDN in one of three ways, which are outlined in greater detail below:

3.1. SDN private: As a result, the SDN cannot take advantage of cloud advantages like cost reduction and scalability because it will reside on the user's infrastructure.

3.2. SDN public: We will be able to take advantage of every advantage of the cloud because this type of SDN will be entirely hosted in the cloud.

3.3. SDN hybrid: With half and half SDN, a part of the framework is facilitated in the cloud, while the rest of facilitated on the organization's premises. When we need to frequently push and pull data from the database, this strategy is best suited to the use case.

4. Plan and test the current network's migration.

Because we will need to define how the apps will use the network services that have been migrated to the cloud, this is a crucial step in the process of migrating an SDN to the cloud. We will also need to evaluate these services, and we can do so by evaluating, among other things, scalability, security, and performance. We will also need to describe the SDN's future maintenance and the roles that each actor will play [8].

The Benefits of Using Software Defined Networks in Cloud Computing The following are a few of the benefits of using Software Defined Networks in cloud computing:

SDN gives network administrators the ability to control the protocols and rules that apply to the switches in the network, allowing them to make even minute adjustments like preventing particular packets from reaching a switch. This kind of control is especially useful when the network is running a cloud multi-tenant architecture because traffic loads can be managed very well.

The network administrator can manage all of the network's switches from a single location thanks to the centralized system. Because it is inefficient to configure each device individually and a shift in focus could send the wrong policies, this makes it much simpler to send security policies by analyzing network activity. This has a security benefit as well.

Additionally, SDN makes it possible to virtualize hardware equipment, allowing businesses to avoid managing actual hardware and incurring hardware costs.

Software Defined Wide Area Network (SD-WAN), a new technology, has emerged as a result of SDN. Similar to SDN, SD-WAN lets an organization manage its WAN without having to worry about all of the connection details between the switches of different departments. Instead, the controller decides where traffic needs to go [4].

By virtualizing the majority of physical networking devices, SDN cuts down on network downtime. We are able to quickly recover from errors and upgrade subsystems without having to shut down the network as a whole thanks to this.

We are able to control and monitor network traffic down to the packet level thanks to the SDN controller; We are able to quickly identify irregularities and send out the necessary security policies thanks to this in-depth data observation from a centralized location [6].

5.7 Challenges with Software Defined Network on the Cloud

Implementing SDN may present unique challenges, as outlined below, given that it is a relatively new technology:

Too soon turning on: Because there are numerous moving parts that need to be accommodated appropriately, moving to SDN without first looking at all of the minute details poses a significant risk of failure [8].

There is only one failure: There is a single point of failure because everything is handled by a centralized controller, and gaining access to the SDN cloud will be extremely destructive.

Cost: It has not received the anticipated level of acceptance since its introduction. This is because implementing SDN and effectively connecting it to the cloud may require additional resources, which may not be feasible for mid- to small-sized businesses.

Uncertainty regarding the best model: There is uncertainty regarding the best model—a combination of the two—because there is no obvious alternative for SDN deployment and manufacturers offer a variety of options, including private (hardware-oriented), public (totally cloud-based), and hybrid models. This makes it hard to determine which model is best because it costs a lot to set up one model and then switch to another. [4]

5.8 Cloud Gaming: Issues and Challenges

For cloud gaming to work, a lot of cutting-edge technology is needed, like live video streaming with low latency and high-performance 3D rendering. The essential design considerations that cloud gaming companies have already addressed serve as the basis for our investigation. A cloud gaming system needs to collect a player's actions, send them to a cloud server for processing, render the results, encode or compress changes to the game environment, and stream video (game scenes) back to the player. Interactivity can only be guaranteed if each of these serial processes occurs within milliseconds. [39]

Example Game Type	Perspective	Delay Threshold
First Person Shooter	First Person	100 ms
Role Playing Game	Third Person	500 ms
Real Time Strategy	Omnipresent	1000 ms

Table 5: Delay tolerance in traditional gaming.

This amount of time, which is referred to as interaction latency, needs to be kept as low as is practical in order to provide cloud gamers with an excellent gaming experience. However, there are consequences: The player's tolerance for interaction delay decreases the less time the system has to complete crucial tasks like scene rendering and video compression. In addition, the likelihood that greater network latency will have a significant impact on a player's interaction experience increases with the time threshold. With delay tolerance in mind, we begin the design conversation.

5.8.1 Interaction Delay Tolerance

This amount of time, which is referred to as interaction latency, needs to be kept as low as is practical in order to provide cloud gamers with an excellent gaming experience. However, there are consequences: The player's tolerance for interaction delay decreases the less time the system has to complete crucial tasks like scene rendering and video compression. In addition, the likelihood that greater network latency will have a significant impact on a player's interaction experience increases with the time threshold. With delay tolerance in mind, we begin the design conversation. [6].

5.8.2 Video Streaming and Encoding

The requirements of a cloud gaming system for video streaming and encoding are then examined. The requirements for video streaming for cloud gaming are very similar to those for live media streaming, another traditional application. Both live media streaming and cloud gaming require that incoming video be encoded and compressed before being distributed to end users as quickly as possible. In both cases, we are only concerned with a small number of the most recent video frames, and we do not have access to future frames before they are made, so encoding must be done with only a small number of frames.

However, live video streaming and cloud gaming have significant differences. To begin, unlike live media streaming, cloud gaming has virtually no client-side video frame buffering capacity. This is because a player's command must travel to the cloud via the Internet when it is sent to the local thin client. There, it is processed by the game logic, rendered by the processing unit, compressed by the video encoder, and returned to the player. There isn't much room for a buffer because everything needs to be done in less than 100 to 200 milliseconds. On the other hand, live video streaming can support a buffer of hundreds of milliseconds or even a few seconds without affecting the QoE of the end user. Because cloud gaming requires delicate real-time encoding, choosing the right video encoder is important for every cloud gaming service. At this time, the H.264/MPEG-4 AVC encoder versions are used by Gaikai and Onlive, the largest cloud gaming companies. Onlive compresses their cloud gaming video feeds with specialized hardware, whereas Gaikai encodes using software. The H.264 encoder was chosen for both scenarios because it has a very high compression ratio and can be set to meet strict real-time requirements.

6 Security

6.1 Security in 5G

Recent advances in mobile communications have created security challenges that have significant privacy concerns for commercial and industrial 5G applications. Many of the security concerns associated with the introduction of 5G are addressed through security measures.

For a long time, business and academic researchers have been striving to improve 5G security. Since 2017, a 3GPP working group dedicated to service and system aspects has been researching and establishing security specifications for 5G systems [41][42].

The 5G network, according to 3GPP, is divided into two parts:

- Standalone Network
- Non-standalone Network

6.2 Attacks in 5G NSA

In this scenario, the evolved packet core is operational, which is the core of 4G networks, and prefers E-UTRAN, the access network of the 4G LTE. As a result, it is crucial to remember that threats and vulnerabilities in 4G LTE networks can equally affect 5G networks.

The following are the critical threats to the security of the 5G NSA network [27]:

6.2.1 Downgrade Attack

It forces the UE LTE connection to 2G or 3G, but the end-user can connect using more advanced technology. Ultimately, the attacker could perform a man-in-the-middle or eavesdropping attack to collect information. For example, a customer can identify whether they have an LTE connection at a location, and it suddenly drops to "E," "G," or another symbol by looking at the indication on their devices.

6.2.2 Data modification Attack

Secure methods of intercepting traffic do not protect the integrity of UMTS and LTE communications. It can result in active men-in-the-middle such as data injection or modification. Mobile devices and base stations can prevent man-in-the-middle type attacks by authenticating and verifying each other. The 5G protocols such as AKA and EAP-AKA initiate the authentication process on 5G networks and are new solutions for recording connection requests

6.2.3 IMSI Tracking

The International Mobile Subscriber Identity (IMSI) is transmitted wirelessly, unencrypted, allowing an attacker to find the SIM card used by the connected user when IMSI (International Mobile Subscriber Identity) requests are created. Additionally, base station spoofing is a fake base station that can unknowingly track and collect personal data.

6.2.4 LTE Roaming

Vulnerabilities in 2G, 3G, and 4G users may be exposed to attacks such as eavesdropping on calls, reading and forwarding data, and tracking due to the usage of obsolete signaling protocols such as SS7 for PSTN networks and diameter for authentication and authorization.[43]

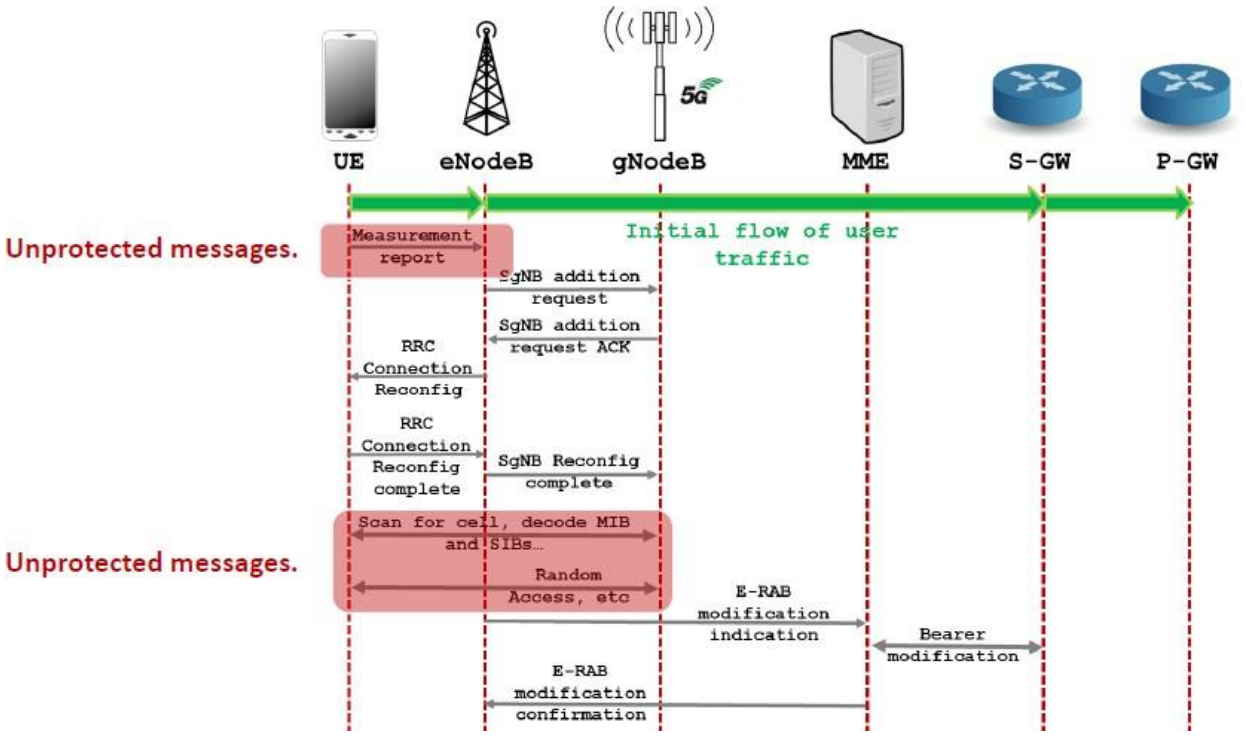

Figure 53: 5G NSA Attach Procedure [42]

6.3 attacks in 5G SA

The service-based architecture of 5G SA, which uses technologies like NFV and SDN to boost security solutions, is a significant feature that distinguishes it from 5G NSA. The service-based architecture runs on general-purpose hardware with software applications built on it. The open- source nature creates vulnerabilities. To address security concerns with 4G LTE, the new ITU standards are included in 5G stand-alone services, protocols, and implementations. The information transfer is software-based in 5G, which is sensitive and confidential, poses a risk to the network applications if not targeted with proper authentication and encryption.[44]

Threats/ Vulnerabilities In 5G SA

6.3.1 SUPI/SUCI Privacy

Subscription Permanent Identifier (SUPI) is encrypted to Subscription Concealed Identifier (SUCI) and transmitted over the network in 5G, preventing attackers from targeting and tracking subscribers' privacy.

SUPI can be targeted and monitored over the air under specific circumstances [29] :

- An unauthenticated device is trying to make an emergency call.
- UE with outdated SIM not provisioned with 5G public key.
- Customers bring their device with an un-updated SIM card to the operator's network.

6.3.2 Man-in-the-middle (MitM)

In 5G, there is a significant concern related to the unprotected user plane integrity protection that an attacker can exploit with Man-in-the-middle during its registration on the network. IP is enabled on

control plane messages, but the data plane and control plane are separated, leaving the data plane vulnerable.

The three classes of attacks that can occur are:

- Identification attacks, detection of devices on the network, knowledge of their properties, and applications.
- Reduce attacks and data rates with capturing device capabilities.
- Battery drain attacks.

6.3.3 Roaming

Roaming from 5G SA to NSA is a network security issue. Even now, the NSA, like 4G EPC, uses SS7 and protocols that are vulnerable to attacks such as text message decryption, location monitoring, and eavesdropping. Therefore, protocols such as HTTP/2 and JavaScript Object Notation (JSON) should be used to mitigate these vulnerabilities with roaming flows.[42][43]

6.4 Threats Related To gNB

6.4.1 Spoofing

Rouge gNBs might have a significant impact depending on the network design of the operator, e.g., the interface between gNB and AMF is protected and secured by IPSec. However, installing a fake gNB is easy and cheap using a software-defined radio-based solution. Suppose the gNB cannot access and penetrate the core network of the compromised network operator. In that case, the effect is limited to misconfiguration of the UE attempting to register or re- register with the core network via this gNB. In these rare cases at the protocol level, previously reported DoS attacks on UEs in 4G networks are common. [17]

The fake gNB can still access the KgNB associated with the UE and integrate itself into the core network, and the UE will not be able to identify the gNB as a fake. User plane encryption is done directly between the gNB and the UE, so knowing the KUPenc key is also required. Therefore, If end-to-end encryption is not provided at the application layer between the UE and the final data sink, the user plane data is mapped directly to the gNB level. However, knowledge of KgNB is still insufficient and does not imply knowledge of KNASint or KNASenc. Therefore, encryption should be enabled by AMF at the NAS level so that gNB cannot generate or intercept NAS messages that request or include User PII such as IMEI or PEI [17]. Remember that identification request messages at the NAS level (response may contain PEI) are sent only after successful authentication if the security context has been established at the AMF level.

6.4.2 Tampering

A software update method is to be used to update a gNB. For example, if the gNB firmware contains a backdoor and an attacker intentionally or unintentionally inserts the backdoor. At the same time, the debug feature is enabled, the modified node can violate the security features of security like IPSec and can lead to leakage of user secrets. IPsec implementation must meet specific requirements compared to long-term 5G device keys. However, the standard does not specify the minimum level of security. After removing the IPSec protection layer, it is also possible to use the decrypted data to access the storage directly.[44]

6.4.3 Jamming by rogue gNB

A rogue gNB with a higher link budget will try to connect to UE, causing the legitimate gNB's connection to be disrupted. As a result, registration may fail. A rogue gNB may indicate that the user is unauthorized while transmitting a denial of connection with a reject cause, possibly resulting in roaming not permitted update status. Generally, the UE will not register until the device is powered off or the SIM card is removed and reinserted. The communication interface can be disconnected entirely for machine-to-machine communication for both fixed and mobile use if the device is altogether standard compliant. The registration with the new gNB is likely for mobile devices if the link budget with the existing serving cell has changed recently and for stationary if the link budget has not changed. [44]

6.5 Security risks in mobile cloud gaming

The following are the most serious risks of online gaming: [45]

6.5.1 Viruses and Malware

You run the risk of downloading spyware and viruses without your knowledge if you look for cheaper or free versions of your favorite games. This is also true when purchasing products from third-party dealers or accessing cheat codes. A security flaw could put you in danger even if you download a game legitimately. If malware is installed on your laptop, hackers can steal personal information.

6.5.2 Identity theft

In order to create profiles of potential victims, cybercriminals gather information that can be used to identify an individual. One of the potential dangers of playing online games with strangers is the chat tool, which lets you talk to other players. The chat function could be used by criminals to obtain private information like your name, phone number, and home address. As a result, it's critical to exercise caution when sharing information while gaming.

6.5.3 Account takeover

Hackers may gain access to all of your accounts and potentially take control of them if you use the same username and password for all of your preferred gaming platforms, which is not recommended. Hackers occasionally use brute force attacks, in which automated programs attempt to gain access to your account using credentials obtained from other sources.

6.5.4 Swatting and doxing

Doxing is a method by which hackers may post your home address or phone number online once they have access to your personal information. The goal of doxing is to punish, scare, or humiliate the target. Doxers act in a variety of ways for a variety of reasons, such as having fun online without realizing the harm they do, pursuing justice (often in the wrong way), retaliation, enmity, harassment, and even profit. Doxing is a one-time occurrence that can cause irreparable harm to a person's life without them realizing it. Worse still, there have been instances of game-related swatting, in which criminals send law enforcement to your home by falsely reporting an emergency to scare you.

6.5.5 Spyware

Spyware can sometimes target gamers, especially if they are working with a bad online gaming company. Without an individual's knowledge, spyware monitors their online activities. It is possible that this information will be sold to third parties if it is gathered, which would be an invasion of privacy.

6.5.6 Data breaches

Game publishers can be directly targeted by hackers. If they gain access to the systems of a publisher, they can steal a lot of information, including the source code for games and personal information stored in users' accounts. The Zynga data breach, in which hackers took usernames, passwords, and email addresses of players of Draw Something and Words With Friends, was a significant example. It was one of the largest data breaches in history, affecting approximately 172 million accounts.

6.5.7 Cross-site scripting

Older gaming systems may use insecure methods to process your login credentials. As a result, they are exposed to cross-site scripting (XSS), a hacking technique that enables thieves to intercept and steal such information.

6.5.8 DDoS attacks

When hackers attempt to overwhelm gaming servers, DDoS attacks cause the service to crash and go offline. Even though this does not include the theft of user data, it still annoys users and can be costly for the game's provider in terms of downtime and recovery time.

6.5.9 Phishing emails

Hackers can also use phishing URLs or phishing emails spread through online gaming chat to trick people into installing game malware on their devices. Emails or chats, for instance, may appear to come from reputable sources and may ask you to download additional content or go to a login page. In reality, the emails are fakes with malicious intent.

6.5.10 Cyberbullying

Other gamers may occasionally be abusive to gamers. In addition to humiliating their victims, cyberbullies may attempt to persuade them to reveal personal information that they can use against them.

7 Security As A Service (SECaaS)

5G offers a wide variety of service types associated with technical requirements. It also allows for various alternative use cases, each with security concerns. It thus requires the use of tools like NFV and SDN to provide differentiated services to customers. The operator's Security (SECaaS - Security-as-a-Service) offering is based on the network slicing, providing "security" for the use case and service exposure, providing service (as-a-Service).

Service exposure covers aspects such as:

- Additional services such as security, geofencing, etc., for differentiation.
- Integration of slice properties such as latency, bandwidth, proximity, QoS, assessment, etc., is standardized.
- Insights and monitoring via instrumentation.

One of the enabling technologies that makes it possible for 5G services is network slicing. It makes it possible for each vertical service to have its own slice of the network that provides the resources it needs. SDN, NFV, and cloud computing are the primary foundations of network slicing. Putting up multiple VNFs in different slices might make some VNFs more vulnerable than the static network.[53]

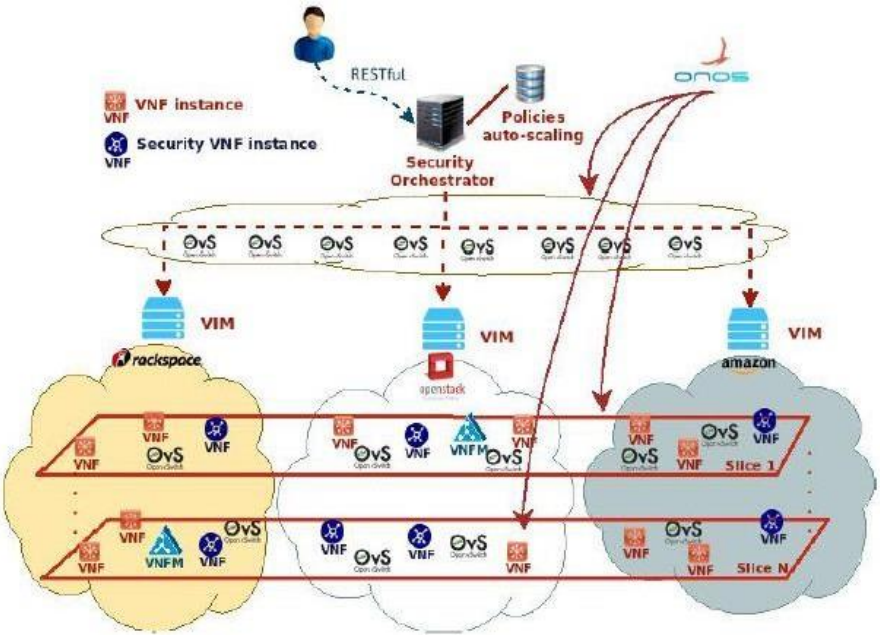

Figure 54: Security as a Service architecture consideration [53]

Fig. shows an overview of the proposed architecture that enables SECaaS on cross-domain platforms. Various security VNFs, such as intrusion detection and prevention systems and deep packet inspection, will be deployed and managed using this architecture. The proposed architectural framework aims to dynamically deploy secure VNF instances maintain elasticity, monitor performance, and implement predictive auto-scaling based on pre-defined policies and metrics. [46]

7.1 Examples of security breaches in cloud gaming

7.1.1 OnLive, in August of 2010

One example of a security breach in the cloud gaming industry is the incident that occurred with the gaming platform, OnLive, in August of 2010.

OnLive was a cloud gaming platform that allowed users to stream and play video games on a variety of devices, including PCs, Macs, and tablets, without the need for expensive hardware or downloads. The company stored all of its users' saved games and other data in the cloud, making the service very convenient for users.

However, on August 17, 2010, OnLive announced that its systems had been hacked, and that user accounts and credit card information had been compromised. The company stated that it had taken immediate steps to secure its systems and that user data was no longer at risk.

The attackers were able to gain access to the OnLive servers through a SQL injection attack, a type of attack that allows an attacker to execute malicious SQL code on a database server. The attackers were able to obtain a large amount of sensitive user data, including usernames, email addresses, and encrypted password. The company promptly notify their user to update their password for the safety measure.

The incident was a major blow to OnLive's reputation and trust in the security of their service, which already was a new concept for many people. The company was also forced to spend a significant amount of time and money on forensic analysis, incident response, and user notification. They also had to struggle to regain the trust of their customers.

It is important to note that the incident not only affected OnLive, but also sent a negative message to the gaming industry, which was still relatively new to cloud gaming and the security risks that come with it. This incident also highlighted the importance of protecting sensitive user data in the cloud, as well as the need for companies to have incident response plans in place to quickly and effectively respond to security breaches.

7.1.2 Xbox Live, in 2011

Another example of a security breach in the cloud gaming industry is the incident that occurred with the gaming platform, Xbox Live, in 2011.

In April of 2011, Microsoft, the company behind Xbox Live, announced that a number of user accounts on the service had been compromised. The attackers had gained access to user email addresses, passwords, and other personal information. Microsoft stated that credit card information was not affected, but they also recommended the users to be vigilant and monitor their credit card statement.

The attackers were able to gain access to the Xbox Live accounts by using a technique called phishing, in which the attackers send out emails that appear to be from legitimate sources, such as Xbox Live, in an attempt to trick users into revealing their login credentials.

As a result of the incident, Microsoft took a number of steps to secure the service, including the use of two-factor authentication for all Xbox Live accounts and increased security for email verification. They also help user to understand how to spot phishing email and how to prevent it.

7.1.3 Ubisoft in 2013

Another example occurred on the game developer, Ubisoft in 2013, the company, which is behind the popular Assassin's Creed and Far Cry series, suffered a major data breach, In which attacker were able to access the company's servers and steal a large amount of sensitive user data, including names, email addresses, and encrypted passwords.

The company stated that it had taken immediate steps to secure its systems and that user data was no longer at risk, but like the above examples, it had a damage on the reputation and trust of their service.

These examples highlight the importance of companies in the cloud gaming industry to take steps to protect sensitive user data, and to have incident response plans in place to quickly and effectively respond to security breaches. Additionally, it is important for users to be aware of the risks associated with cloud gaming and to take steps to protect their own personal information, such as by using strong, unique passwords and being vigilant against phishing emails.

7.1.4 Steam in November of 2018

A recent example of a security breach in the cloud gaming industry occurred with the platform, Steam in November of 2018. Steam is a popular digital distribution platform for PC and mobile games, developed and run by Valve corporation.

On November 6, 2018, Valve announced that a number of Steam accounts had been compromised by a hacking incident. The attackers had been able to gain access to user email addresses, encrypted passwords, and purchase histories. Valve stated that no financial information had been compromised, but recommended users to change their password and be vigilant for suspicious activity on their accounts.

Valve's investigation showed that the attack was caused by a phishing campaign, in which the attackers sent out emails that appeared to be from Steam, in an attempt to trick users into revealing their login credentials. Steam also implemented additional security measures to prevent similar attacks in the future.

7.1.5 Epic Games in 2020

Another example happened in 2020, with the gaming platform, Epic Games, where a data breach had exposed the personal data of millions of Fortnite players and other users of the Epic Games Store. The data included usernames, email addresses, and salted and hashed passwords. The company quickly responded by disabling the ability for users to sign in with the affected account and forcing password reset for all users. Epic Games also hired a top cyber security firm to investigate the incident and help protect the users.

These examples show that, despite the companies best efforts, breaches can still happen, that's why it's important for companies to be transparent about the incident and communicate with their users, take responsibility and provide solutions for the affected user, and to learn from the experience and improve their security measures. It is also important for users to be aware of the risks associated with cloud gaming and to take steps to protect their own personal information, such as by using strong, unique passwords, being vigilant against phishing emails and monitoring their accounts for suspicious activity.

7.1.6 Blizzard’s Battle in 2016

One example of a security breach in cloud gaming through a DDoS (Distributed Denial of Service) attack occurred in 2016 with the platform, Blizzard's Battle.net service. The Blizzard's Battle.net service is a gaming platform that includes popular titles such as World of Warcraft, Hearthstone, and Diablo III.

In April 2016, the Blizzard's Battle.net service was hit by a massive DDoS attack, which caused significant disruption to the service. The attack made it difficult or impossible for users to log in and play their games, or access other features of the platform, such as account management and customer support.

DDoS attacks work by overwhelming a server or network with a flood of fake traffic, making it difficult or impossible for legitimate users to access the service. This can cause significant service disruption and financial losses for the company, as well as inconvenience and frustration for users.

Blizzard reacted quickly to the attack and work to mitigate the issue, and after the attack, they were forced to improve their infrastructure, and their security measure by implementing DDoS protection. They also communicated the status of the service with their user frequently, keeping them informed about the progress.

The attack on Blizzard's Battle.net service serves as a reminder of the potential dangers of DDoS attacks, which can cause significant disruptions to online services, and the importance of having adequate DDoS protection in place. Additionally, it showed how a company should handle the situation, by being transparent and proactive, as well as learning from the experience to improve the security of their platform.

8 Case study on DDoS attacks

Type of DDOS attack impacting mobile cloud gaming and their mitigation techniques.

DDoS (Distributed Denial of Service) attacks are a type of cyber attack in which a large number of devices are used to flood a network or server with traffic, causing it to become overwhelmed and unavailable to legitimate users. In the context of mobile cloud gaming, DDoS attacks can disrupt gameplay, cause lag, and even prevent players from accessing the game server.

These types of attacks are performed using a different type of botnet for example Mirai-based botnet, which is a type of malware that can infect Internet of Things (IoT) devices such as routers, cameras, and other connected devices. The botnet can then be controlled remotely to launch coordinated attacks on a target. The attackers likely used a variety of methods to infect the IoT devices, including exploiting known vulnerabilities or using default passwords that had not been changed. These botnets are able to control a large number of these devices and use them to generate a massive amount of traffic aimed at the game's servers.

The attackers use multiple techniques to launch the DDoS attacks on game servers. These techniques included UDP flood, SYN flood, and DNS amplification attacks.

UDP flood

A UDP flood attack is a type of DDoS (Distributed Denial of Service) attack that targets the User Datagram Protocol (UDP) network protocol. UDP is a connectionless, unreliable protocol that does not require a handshake between the sender and receiver before transmitting data. This makes UDP an attractive target for attackers who want to flood a network with a large volume of traffic.

In a UDP flood attack, the attacker sends a large number of UDP packets to the target network or system. These packets are often sent with a spoofed source IP address, which makes it difficult for the target system to identify the source of the attack. The target system will receive a large number of UDP packets, but because UDP is a connectionless protocol, it will not be able to distinguish between legitimate and malicious traffic.

The effect of a UDP flood attack is that the target system becomes overwhelmed with traffic and is unable to process legitimate traffic. This can cause the system to slow down or even crash, resulting in a denial of service.

There are several variations of UDP flood attacks, including:

1. UDP amplification attack: In this attack, the attacker sends a small number of UDP packets to a large number of servers that have a vulnerability that allows them to respond with a larger volume of traffic. The attacker spoofs the source IP address of the packets so that the response traffic is directed at the target system.
2. NTP amplification attack: This is another type of UDP amplification attack that targets Network Time Protocol (NTP) servers. The attacker sends a small number of NTP packets with a spoofed source IP address to a large number of NTP servers that have a vulnerability that allows them to respond with a larger volume of traffic. The response traffic is directed at the target system.

Preventing UDP flood attacks can be challenging, as UDP is designed to be a fast and efficient protocol that does not include many of the safeguards found in other protocols like TCP. However, there are some mitigation strategies that can be used, such as:

1. Implementing firewalls and intrusion prevention systems (IPS) that can identify and block UDP flood traffic.
2. Configuring network devices to limit the rate of UDP traffic that can be sent to a system.
3. Using traffic analysis tools to monitor network traffic and identify anomalous traffic patterns that could be indicative of an attack.
4. Configuring servers to limit the rate of incoming UDP traffic and to drop packets with spoofed source IP addresses.

SYN Flood

A SYN flood attack is a type of DDoS (Distributed Denial of Service) attack that targets the Transmission Control Protocol (TCP) network protocol. TCP is a connection-oriented protocol that requires a three-way handshake between the sender and receiver before transmitting data. This makes TCP more reliable than UDP, but it also makes it vulnerable to SYN flood attacks.

In a SYN flood attack, the attacker sends a large number of SYN (synchronization) packets to the target system but does not complete the three-way handshake by sending the final ACK (acknowledgement) packet. This leaves the target system waiting for the final ACK packet, and the half-open connections can consume system resources like memory and CPU cycles.

The effect of a SYN flood attack is that the target system becomes overwhelmed with half-open connections and is unable to process legitimate traffic. This can cause the system to slow down or even crash, resulting in a denial of service.

There are several variations of SYN flood attacks, including:

1. Direct SYN flood attack: In this attack, the attacker sends a large number of SYN packets directly to the target system.
2. Spoofed SYN flood attack: In this attack, the attacker spoofs the source IP address of the SYN packets to make it appear that they are coming from legitimate sources. This makes it more difficult for the target system to identify the source of the attack.
3. Distributed SYN flood attack: In this attack, the attacker uses a botnet or a network of compromised computers to send the SYN packets to the target system from multiple sources. This makes it even more difficult for the target system to identify the source of the attack.

Preventing SYN flood attacks can be challenging, but there are some mitigation strategies that can be used, such as:

1. Implementing firewalls and intrusion prevention systems (IPS) that can identify and block SYN flood traffic.
2. Configuring network devices to limit the rate of SYN traffic that can be sent to a system.
3. Implementing SYN cookies, a technique that allows the target system to continue processing legitimate traffic while discarding half-open connections from SYN flood attacks.

4. Using traffic analysis tools to monitor network traffic and identify anomalous traffic patterns that could be indicative of an attack.
5. Implementing rate limiting, such as limiting the rate of SYN packets from a single IP address, to prevent a single source from overwhelming the system.

DNS amplification

A DNS amplification attack is a type of DDoS (Distributed Denial of Service) attack that targets Domain Name System (DNS) servers. DNS is a critical component of the Internet that translates domain names into IP addresses. In a DNS amplification attack, the attacker exploits the vulnerability of DNS servers that respond with a larger volume of traffic than the initial request.

The attacker sends a large number of DNS queries to a vulnerable DNS server with a spoofed source IP address that appears to be the target system. The DNS server, which is designed to respond to DNS queries, returns a much larger response to the target system than the original query. This amplifies the volume of traffic directed at the target system, overwhelming it and resulting in a denial of service.

The effect of a DNS amplification attack is that the target system becomes overwhelmed with traffic and is unable to process legitimate traffic. This can cause the system to slow down or even crash, resulting in a denial of service.

There are several variations of DNS amplification attacks, including:

1. Direct DNS amplification attack: In this attack, the attacker sends a large number of DNS queries directly to the vulnerable DNS server.
2. Reflection DNS amplification attack: In this attack, the attacker sends DNS queries to a large number of third-party DNS servers that have a vulnerability that allows them to respond with a larger volume of traffic. The response traffic is directed at the target system, amplifying the volume of traffic.

Preventing DNS amplification attacks can be challenging, but there are some mitigation strategies that can be used, such as:

1. Implementing firewalls and intrusion prevention systems (IPS) that can identify and block DNS amplification traffic.
2. Configuring DNS servers to restrict access to recursive queries, which can prevent the DNS server from being used in an amplification attack.
3. Implementing rate limiting, such as limiting the number of queries that can be sent to a DNS server from a single IP address.
4. Monitoring DNS traffic for anomalous patterns that could indicate an attack.
5. Updating DNS server software to patch known vulnerabilities that can be exploited in amplification attacks.

9 Conclusion

Mobile cloud gaming is becoming increasingly popular and has the potential to revolutionize the gaming industry. However, it also introduces security threats that need to be addressed to ensure the safety and privacy of users.

Security threats in mobile cloud gaming are a major concern for both players and game developers. These threats include unauthorized access, data breaches, and identity theft. The current security measures used in mobile cloud gaming are insufficient to protect against the growing number and sophistication of security threats. Game developers need to implement more robust security measures to protect their players' personal and financial data.

To address these security threats, game developers can use a combination of encryption techniques, multi-factor authentication, and secure cloud storage solutions.

Education and awareness are essential to ensure players understand the potential security risks associated with mobile cloud gaming. Players should be informed about best practices for protecting their personal information, such as using strong passwords and avoiding public Wi-Fi networks.

Continued research and development of security solutions is necessary to stay ahead of evolving security threats. As the popularity of mobile cloud gaming continues to grow, game developers must remain vigilant in protecting their players' data and ensuring a safe and enjoyable gaming experience.

In summary, security threats in mobile cloud gaming are a significant concern that requires a proactive approach by game developers, players, and the broader gaming industry. Through the implementation of robust security measures, education and awareness campaigns, and ongoing research and development, it is possible to mitigate these threats and ensure a secure and enjoyable gaming experience for all.

10 References:

[1] N. U. M. O. C. O. a. A. J. A. Opeoluwa Tosin Eluwole, “From 1G to 5G, What Next?,” IAENG International Journal of Computer Science, 45:3, IJCS_45_3_06, 28 August 2018.

[2] P. Sharma, “Evolution Of Mobile Wireless Communication; Networks-1g To 5g,” International Journal Of Computer Science And Mobile Computing, vol. 2, no. 8, p. 47 – 53, 2013.

[3] G. 01.02, “Digital cellular telecommunications system (Phase 2+), General description of a GSM Public Land Mobile Network,” ETSI, 1996.

[4] M. Y. Rhee, Mobile Communication Systems and Security, John Wiley & Sons, 2009.

[5] H. K., S. P. H. J. Mamta Agiwal, “A Survey on 4G-5G Dual Connectivity: Road to 5G Implementation,” IEEE Access, vol. 9, pp. 16193-16210, 2021.

[6] A.K.Pachauri, “5G technology-Redefining Wireless Communication in upcoming Years,” 2012.

[7] C. K. Toh, “Ad Hoc Mobile Wireless Networks: Protocols and Systems,” 2002

[8] [Online]. Available: https://www.ecstuff4u.com/2018/05/1g-technology-advantagesand.html.

[9] W. Lee, “Cellular and Mobile Communication,” 2010.

[10] W. Stalling, “Data and Computer Communication,” 2011.

[11] G. A, “Wireless Communications,” Cambridge University Press, Cambridge,, 2005

[12] “network-evolution-3g-vs-4g-vs-5g,” [Online]. Available: https://medium.com/@sarpkoksal/core-network-evolution-3g-vs-4g-vs-5g7738267503c7.

[13] Wenqiong Yu, “The Network Security Issue of 3G Mobile Communication System Research”, 2010 International Conference on Machine Vision and Human-machine Interface.

[14] K. J, Introduction to 4G Mobile, Artech House, London,, 2014.

[15] “techdifferences.com/difference-between-3g-and-4g-technology,” [Online]. Available: https://techdifferences.com/difference-between-3g-and-4g-technology.html.

[16] Christopher Cox, An Introduction To Lte LTE, LTE-Advanced, Sae And 4g Mobile Communications, A John Wiley & Sons, Publication, 2012.

[17] Guide to LTE Security. [online] Available : https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-187.pdf

[18] L. He, Z. Yan and M. Atiquzzaman, "LTE/LTE-A Network Security Data Collection and Analysis for Security Measurement: A Survey," in IEEE Access, vol. 6, pp. 4220-4242, 2018, doi: 10.1109/ACCESS.2018.2792534.

[19] R. Piqueras Jover, "Security attacks against the availability of LTE mobility networks: Overview and research directions," 2013 16th International Symposium on Wireless Personal Multimedia Communications (WPMC), Atlantic City, NJ, USA, 2013, pp. 1-9.

[20] 4G Technology uses , features , advantages and disadvantages [Online] Available:

https://www.online-sciences.com/technology/4g-technology-uses-features-advantages-and-disadvantages/

[21] H. K., S. P. H. J. Mamta Agiwal, “A Survey on 4G-5G Dual Connectivity: Road to 5G Implementation,” IEEE Access, vol. 9, pp. 16193-16210, 2021

[22] 5G Standalone (SA) vs 5G Non-Standalone (NSA). [Online] Available:

https://www.alepo.com/5g-sa-vs-5g-nsa-what-are-the-differences/

[23] O. O. Erunkulu, A. M. Zungeru, C. K. Lebekwe, M. Mosalaosi and J. M. Chuma, "5G Mobile Communication Applications: A Survey and Comparison of Use Cases," in IEEE Access, vol. 9, pp. 97251-97295, 2021, doi: 10.1109/ACCESS.2021.3093213.

[24] 3. T. 2. v. 15.3.0, “System Architecture for the 5G System,” ETSI TS 123 501 V15.3.0, 2018.

[25] A. G. A. R. K. Jha, “A Survey Of 5g Network Architecture And Emerging Technologies,” IEEE Access, vol. 3, pp. 1206-1232, 2015.

[26] H. Baba et al., "End-to-end 5G network slice resource management and orchestration architecture," 2022 IEEE 8th International Conference on Network Softwarization (NetSoft), Milan, Italy, 2022, pp. 269-271, doi: 10.1109/NetSoft54395.2022.9844088.

[27] Fuwen Liu, Li Su, Bo Yang, Haitao Du, Minpeng Qi, Shen He “Security Enhancements to Subscriber Privacy Protection Scheme in 5G Systems” 2021 International Wireless Communications and Mobile Computing (IWCMC) | 978-1-7281-8616-0/21/$31.00 ©2021 IEEE | DOI: 10.1109/IWCMC51323.2021.9498591

[28] 5G Identifiers SUPI and SUCI, [Online] Available: https://www.techplayon.com/5g-identifiers-supi-and-suci/

[29] 3. T. 3. v. 1. Release, “5G Security architecture and procedures,” ETSI TS 133 501 V15.2.0, 2018.

[30] W. L. D. P. D. A. M. G. B. V. L. Gerrit Holtrup, “5G System Security Analysis,” arXivLabs, 2021.

[31] M. Caprolu, R. Di Pietro, F. Lombardi and S. Raponi, "Edge Computing Perspectives: Architectures, Technologies, and Open Security Issues," 2019 IEEE International Conference on Edge Computing (EDGE), Milan, Italy, 2019, pp. 116-123, doi: 10.1109/EDGE.2019.00035.

[32] P. Shrivastava and M. Damle, "Investment decision in cloud gaming-based businesses opportunities: An analysis of the cloud gaming industry," 2022 International Conference on Decision Aid Sciences and Applications (DASA), Chiangrai, Thailand, 2022, pp. 1224-1228, doi: 10.1109/DASA54658.2022.9765298.

[33] Y. Zhang and Y. Zhang, "Discussion on Key Technologies of Cloud Game Based on 5G and Edge Computing," 2020 IEEE 20th International Conference on Communication Technology (ICCT), Nanning, China, 2020, pp. 524-527, doi: 10.1109/ICCT50939.2020.9295741.

[34] R. Shea, J. Liu, E. C. . -H. Ngai and Y. Cui, "Cloud gaming: architecture and performance," in IEEE Network, vol. 27, no. 4, pp. 16-21, July-August 2013, doi: 10.1109/MNET.2013.6574660.

[35] W. Cai, V. C. M. Leung and M. Chen, "Next Generation Mobile Cloud Gaming," 2013 IEEE Seventh International Symposium on Service-Oriented System Engineering, San Francisco, CA, USA, 2013, pp. 551-560, doi: 10.1109/SOSE.2013.30.

[36] D. King, C. Rotsos, A. Aguado, N. Georgalas and V. Lopez, "The Software Defined Transport Network: Fundamentals, findings and futures," 2016 18th International Conference on Transparent Optical Networks (ICTON), Trento, Italy, 2016, pp. 1-4, doi: 10.1109/ICTON.2016.7550669.

[37] A. S. Prasad, D. Koll and X. Fu, "On the Security of Software-Defined Networks," 2015 Fourth European Workshop on Software Defined Networks, Bilbao, Spain, 2015, pp. 105-106, doi: 10.1109/EWSDN.2015.70.

[38] [Online], "Software Defined Networking (SDN) and Cloud Computing". Available=

https://medium.com/@danish_raza/software-defined-networks-sdn-7b5e3c25ba97

[39] M. Claypool and D. Finkel, "The effects of latency on player performance in cloud-based games," 2014 13th Annual Workshop on Network and Systems Support for Games, Nagoya, Japan, 2014, pp. 1-6, doi: 10.1109/NetGames.2014.7008964.

[40] Anmol Gupta1, Kamlesh Dutta, "parameters used for measuring the End User ExperiCloud Gaming: Architecture and Quality of Service" CPUH-Research Journal: 2015, 1(2), 19-22

[41] E. J. S.-C. D. H.-M. F. A.-S. Juan Aranda, "5G networks: A review from theperspectives of architecture, business models, cybersecurity, and research developments," Novasinergia, 2021.

[42] R. P. Jover, 5g Protocol Vulnerabilities And Exploits, Bloomberg, 2020.

[43] Security Considerations for the 5G Era," 5G Americas, 2020.

[44] Y. W. W. Z. Shunliang Zhang, "Towards secure 5G networks: A Survey," Computer Networks, vol. 162, 2019

[45] [Online], "The 10 biggest online gaming risks and how to avoid them. Available= https://www.kaspersky.com/resource-center/threats/top-10-online-gaming-risks

[46] A. Furfaro, A. Garro and A. Tundis, "Towards Security as a Service (SecaaS): On the modeling of Security Services for Cloud Computing," 2014 International Carnahan Conference on Security Technology (ICCST), Rome, Italy, 2014, pp. 1-6, doi: 10.1109/CCST.2014.6986995.

[47] Evolution of mobile technology, [online] Available: https://iot.telenor.com/technologies/evolution-mobile-technology/

[48] Wireless WANs: Cellular Radio and PCS Networks, [online] Available: https://flylib.com/books/en/2.567.1.77/1/

[49] Explain GSM Network architecture in detail, [online] Available: https://www.ques10.com/p/11989/explain-gsm-network-architecture-in-detail-1/?

[50] GPRS (General Packet Radio Services) Architecture, [online] Available: https://dattashingate.wordpress.com/2018/09/18/gprs-general-packet-radio-services-architecture/

[51] LTE, LTE-Advanced and WiMAX: Towards IMT-Advanced Networks, [online] Available: https://www.oreilly.com/library/view/lte-lte-advanced-and/9781119970453/ch14-sec002.html

[52] LTE Key Hierarchy [online], Available: https://ethz.ch/content/dam/ethz/special-interest/infk/inst-infsec/system-security-group-dam/education/SOWN_AS19/cellular-security-2.pdf

[53] M. B. D. L. C. D. T. T. a. N. T. Yacine Khettab, "Virtual Security as a Service for 5G Verticals," IEEE Wireless Communications and Networking Conference, 2018