Study of 5G Roaming Security

Shubhra Kulshrestha

Study of 5G Roaming Security

Author: Shubhra Kulshrestha (5025516)

Supervisor: Ir. Rogier Noldus (TU Delft, Ericsson)

Thesis Committee Members: Dr. Ir. Eric Smeitink (KPN, TU Delft)

Dr. Qing Wang (TU Delft)

in partial fulfilment of the requirements for the degree of

Master of Science

in Electrical Engineering
Track: Wireless Communications and Sensing

Delft University of Technology

Delft, The Netherlands
To be publicly defended on Friday, December 01, 2023

Contents

List of Figures

List of Tables

Abstract

This thesis titled ´Study of 5G Roaming Security´ investigates the potential network vulnerabilities of 5G roaming reference points. The 5G Non-Standalone (NSA) is already being deployed in different countries across the world. With 3G becoming obsolete, mobile communication will primarily depend on 4G and 5G-NSA. In the later stages of the 5G rollout, 5G Standalone (SA) deployment will also take place gradually. Since 5G will support various use cases such as connected vehicles, smart farming, and smart healthcare, the number of connected devices will eventually increase. This would mean more data traffic generation compared to the LTE. In such a scenario, the privacy protection of the User Plane becomes highly significant. This thesis work provides a randomization-based security solution that would make the Standalone 5G reference points more robust to interception attacks.

The goal is to implement a negotiation-based randomization solution over N9, N3, and N32 as these reference points cross the HPLMN boundary. This solution will be a part of the GTP-U header and can be easily implemented by modifying the existing signaling procedures. Random bytes will be added to the GTP-U header before the start of the payload. The idea of adding randomization bytes has been extended to include TCP-based randomization and IMS-based randomization. The TCP-based randomisation also includes two different algorithms for the addition of random bytes. After analysing the User Plane security, the vulnerability analysis of SEPP was undertaken, to understand the vulnerabilities that can be a threat to the network infrastructure. A vulnerability assessment matrix was made and high-risk vulnerabilities were highlighted along with a few precautionary steps. The implementation details and architectural changes for the implementation of GTP and TCP-based randomization are provided. The randomization is useful in masking the signature distribution of an application's packet length and can be a powerful protection mechanism against data traffic analysis attacks.

Acknowledgements

I would like to express my gratitude towards my supervisor Ir. Rogier Noldus for his continuous guidance, support, and generosity with his time. I would not have been able to accomplish my thesis without his valuable feedback and guidance.

I would like to thank master thesis committee members Dr. Ir. Eric Smeitink and Dr. Qing Wang for providing me the opportunity to present my master thesis. I could not have embarked on this journey without the support of my family, Sanjay Kulshrestha, Ritu Kulshrestha and Shikhar Kulshrestha. Special thanks to my husband Abhikalp Kulshrestha for always motivating me during my master's journey and beyond.

Finally, I would like to express my gratitude to everyone who helped, whether directly or indirectly, in the completion of this thesis, including the entire team of Electrical Engineering at Delft University of Technology.

Shubhra Kulshrestha
Delft, November 2023

Chapter 1

Introduction

The telecommunication industry plays a major role in connecting several other industries such as banking, manufacturing, healthcare, military, etc. A country´s economy is heavily dependent on reliable and effective communication services. The Mobile Network Operators (MNOs), provide communication services such as telephony, wireless service, and internet to subscribers and regularly process the sensitive data of millions of subscribers. Thus, a security threat has the potential to damage multiple industries relying on these services.

The **Telecom Vendors**, **Mobile Network Operators (MNOs)** and **Subscribers** are a part of this dynamic sector. Network outages or data compromise caused due to security vulnerabilities can impact the following in different ways:

- **Telecom Vendors**: These companies manufacture Information and Communication Technology infrastructure and devices, and provide services for managing their products. Examples are Cisco, Nokia, Ericsson, and Huawei. These vendors sell their products to the MNOs and are not directly involved with the users of these services. Network outages caused by any security vulnerability such as misconfigurations, unprotected sensitive information, and non-compliance to security standards can lead to revenue loss, damage to reputation and a decline in sales.

- **Mobile Network Operators (MNOs)**: The MNOs acquire the technology infrastructure from the Vendors and use it to provide communication services to their customers or users. The MNOs are responsible for controlling the acquired infrastructure to sell their services. This includes spectrum purchasing, drafting the billing policies, and handling customer care. Examples are KPN, T-Mobile, Vodafone, etc. Network outages caused due to an open security flaw can lead to fines from regulators, damaged reputation, and lawsuits from companies and people using their services as well as revenue loss.

- **Customers/Users**: The customers or users purchase voice and data services from the MNOs. Network outages or data compromise can result in hampering critical day-to-day tasks and eventually loss of trust in the Operator. Many telecom security attacks can cause the disclosure of the personal data of millions of subscribers putting their privacy at risk.

Figure 1.1 combines real-life security incidents due to open security vulnerabilities which led to huge financial losses for the MNOs.

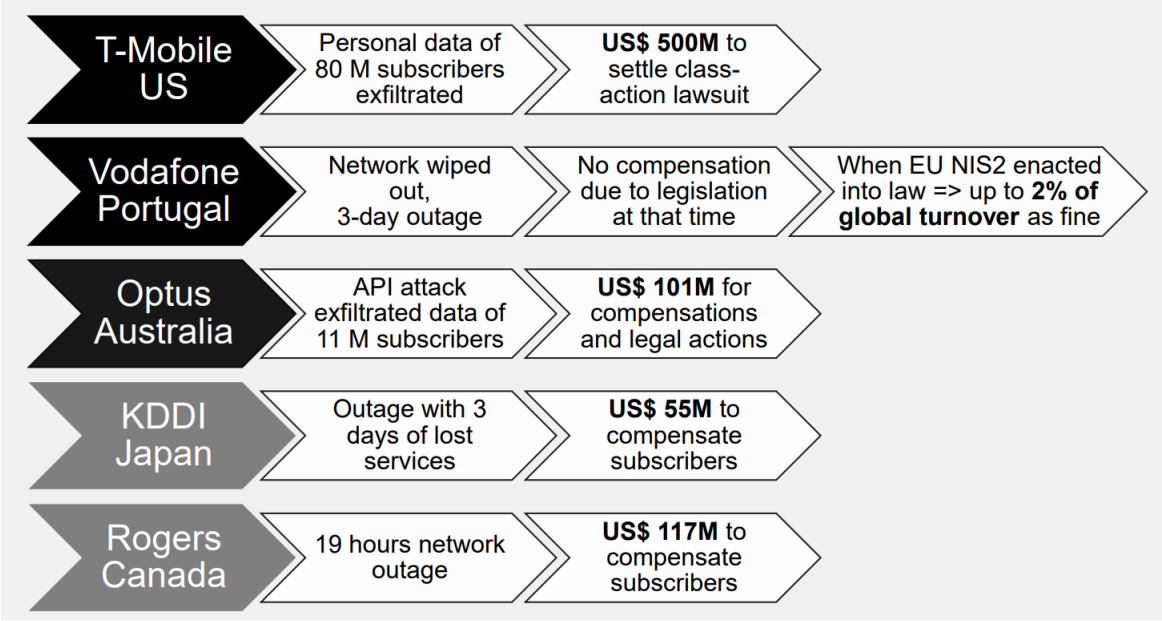


Figure 1.1: MNO revenue loss as a result of exposed security flaws [34],[28],[60],[51],[29]

1.1 Roaming Security in Mobile Networks

The Internet Protocol (IP) Packet eXchange (IPX) Network facilitates the efficient connection between different mobile network operators (MNO), fixed-line operators, or any other internet service providers based on their service level agreements (SLAs) and commercial arrangements [38]. IPX networks are often used for services like international roaming, ensuring that users can connect and use services across different telecommunications networks..

As described in [30], within the realm of roaming agreements, the vulnerabilities associated with interconnection are arguably the weakest security point in the legacy 3GPP standards. These vulnerabilities can serve as an entry point for cybercriminals and authoritarian governments from various parts of the world to carry out cyber and physical attacks on an operator's clientele [30].

For instance, in July 2021, the Pegasus Project [30] uncovered instances of governments abusing smartphone spyware to target political opponents, activists, and journalists going against the terms of their original licenses. The report highlighted the extensive network of businesses involved in selling and supporting such spyware products [30].

Moreover, nation-states frequently exploit signaling vulnerabilities present in 2G, 3G and 4G networks. This exploitation allows them to engage in activities like gathering intelligence on other telecom networks or tracking the whereabouts of specific individuals both domestically and when they are using foreign networks. Cybercriminals and individuals without the technical know-how in SS7 and Diameter often enlist the services of grey market actors to design customized attacks for them [30].

This chapter provides an introduction to 5G network architecture, its general security architecture, and roaming security provisions. Section 1.1 covers 5G features followed by security architecture in Section 1.3 and vulnerabilities from the literature are described in Section 1.4. The final Sections 1.4 and 1.5 discuss the research motivation and research objectives.

1.2 5G Network

The Fifth Generation of Mobile Networks is expected to achieve high data rates with 10 Gbps upload speed and 20 Gbps download speed [19]. To achieve such data rates, multiple changes have been made to the existing LTE network architecture, thus, 5G Standalone architecture evolved and came into existence. The Standalone deployment means that the core network will be a 5G-Next Generation only core consisting of virtualized network functions, connected to the user devices via a 5G Radio Access Network. A few network architecture-related concepts that have been introduced for 5G networks are [9]:

- Slicing of Core network and RAN network
- Strict separation of User Plane and Control Plane with the aid of Software-Defined Networks
- Migrating from reference-point-based architecture to service-based architecture in the core network.
- Enforcement of Quality-of-Service flows as opposed to EPS bearer level for quality of service.

In contrast to the Standalone deployment (SA) which will be compatible with existing technologies and provide full functionality for 5G use cases, there exists another deployment called Non-Standalone (NSA) deployment. The NSA deployment option will be widely used in the beginning by the Mobile Network Operators (MNO) as it entails only upgrading the existing eNodeB to ng-enodeB in a master-slave configuration with the EPC core remaining intact. In the next few deployment stages, the EPC core will be replaced with the NGC core.

1.3 5G Security Architecture features

The upgraded 5G security architecture offers enhanced protection mechanisms [22] against fake base station attacks, IMSI attacks, Man-in-the-Middle (MITM) attacks, etc. Some of these mechanisms are described below:

- **Increased Home control**: Increased home control allows for verification of device location while roaming if a request is received from the visited network. This was done to mitigate the vulnerabilities in 3G and 4G where fake signaling messages were sent to the home network to get hold of the IMSI.

- **Introducing Security Anchor Function (SEAF) in the AMF**: The SEAF acts as an anchor key and allows re-authentication when moving between different access networks. This reduces the signaling load on the HSS. The SEAF and the AMF can be co-located or separated.

- **Use of SUCI before authentication**: The 5G Subscription Permanent Identifier (SUPI) is a unique identifier associated with a subscriber in a 5G network. It plays a crucial role in identifying and authenticating subscribers, ensuring secure communication and access to services. It is never disclosed over the air when a mobile is establishing a connection. Unlike 3G and 4G, where IMSI is disclosed during the UE attachment. Another identifier, Subscription Concealed Identifier (SUCI), is used until device authentication. This has been done to prevent IMSI catchers, who can impersonate base stations and use this information for voice and SMS interception. The SUCI is a part of the SUPI.

- **Continuous evaluation of signal strength**: The UE which is in RRCCONNECTED mode transmits assessment reports to the network. These measurement reports have security values that can prevent IMSI catchers. Eg. the signal strength of a fake BS can be unusually high.

- **Secure Steering of Roaming (SoR) information**: The 3GPP Release-15 standard for 5G added native support for a secure Steering of Roaming (SoR) solution. The 5G SoR solution enables the home network operator to steer its roaming customers to its preferred VPLMN networks to enhance roaming customers' experience and reduce roaming charges. The home network can steer roaming information to the preferred VPLMN for a more comfortable roaming experience.

1.4 5G Security Vulnerabilities

As described in Section 1.1, the 5G Standalone core supports Service-Based Architecture (SBA), and there is a shift from Telecom proprietary protocols to more generic Internet protocols. The 5G SA core is built on well-known Internet protocols such as HTTP and TLS instead of Diameter in 4G. These are internet protocols that are well known and open to attackers [37]. Since these protocols have been a part of the Internet for a long period, many of their vulnerabilities have been exposed before, which is a threat to the current security scenario. Thus, a lot of tools and techniques are already available for exploiting these vulnerabilities. The 5G SA core does not use Diameter for routing "roaming messages". Instead, these roaming communications will be performed HTTP/2 signaling in a Producer-Consumer model, which will be elaborated in Chapter 3. The roaming communication in 5G SA will be either TLS-protected or PRINS-protected [17]. The implementation of these protocols in 5G architecture will bring in new challenges and risks. Hence, a risk analysis of PRINS protected reference point can help prevent or mitigate known vulnerabilities.

Moreover, GPRS Tunnelling Protocol (GTP), which carries the roaming traffic, is prone to certain vulnerabilities, that can impact the core network along with the user plane present in the radio access network. There can be attacks on core signaling protocols such as HTTP/2

and TCP. The 5G Non-standalone (NSA) architecture will still be using Enhanced Packet Core (EPC) from LTE and the vulnerabilities that are a part of Diameter will also be present in 5G. For protecting the network against such risks, it is recommended that mobile operators use Diameter, GTP, and SS7 firewalls which can filter invalid signaling traffic.

1.5 Research motivation

The NG-Core uses internet protocols such as HTTP/2 for communication between servers and clients. It also uses JSON which is a data interchange format used for exchanging information between different network components and entities. Also, EPC and 5G will be used together for Non-Standalone deployment. In such cases, the end-to-end roaming connection will fall back to SS7 or Diameter signaling. It is important to secure both the core network and the RAN network during roaming for standalone and non-standalone deployments. During roaming, attackers can take advantage of vulnerabilities present in 5G networks. These attackers can exploit such vulnerabilities to gain access to the network. To protect the network, it is necessary to understand these vulnerabilities and use protective measures to prevent such network attacks.

1.6 Thesis Objectives

- The objective of this thesis is to perform a thorough vulnerability analysis of 5G roaming reference points and identify the reference points that are highly prone to roaming attacks for further analysis.
- To propose a protection mechanism for the identified vulnerable reference points. This protection mechanism will take the form of an upgraded GTP Randomization protocol for securing the roaming interconnection.
- To implement Randomization on TCP level to safeguard mobile applications against fingerprinting.
- To perform a risk analysis of N-32 reference point and analyze the risks associated with the newly introduced interconnection protocol PRINS which runs over the N32 reference point.

1.7 Thesis Outline

The thesis outline that shall be followed is provided below :

- Chapter 1 introduces roaming security for 5G and describes the research motivation and the thesis objectives.
- Chapter 2 provides details of existing literature on GTP and application fingerprinting.
- Chapter 3 provides background information on 5G architecture and system processes.

- Chapter 4 identifies the vulnerable reference points based on the literature survey and protocol vulnerability analysis.
- Chapter 5 proposes an updated GTP Randomization for additional security and provides implementation procedure details.
- Chapter 6 proposes TCP Randomization and provides implementation procedure details.
- Chapter 7 provides a conclusion for different Randomization methods implemented in the thesis. It also describes the future work.

Chapter 2

Literature Review

This chapter provides details of existing research work that was undertaken in the past to evaluate the 5G interconnection vulnerabilities. Description of GTP vulnerabilities and N-32 risks have been covered.

2.1 Interconnection Security

The roaming security of a network depends largely upon its security architecture. Commonly, GSMA-recommended firewalls and filtering mechanisms are used to filter out illegitimate traffic. Some vulnerabilities still exist in 2G, 3G and 4G networks. The legacy networks (2G, 3G, and 4G) standardized by 3GPP are built on trusting relationships with their roaming partners. The S6a is the signaling reference point between the MME and the HSS in 4G, which allows unencrypted signaling traffic [30]. The interconnection protocols that are used in the network are SS7 (2G and 3G), Diameter (4G) and GTP. The SS7 protocol used in 2G and 3G, was considered vulnerable to IMSI attacks, Denial-of-Service, and Eavesdropping on traffic or signaling [35]. These legacy networks will continue to be a part of future network generations before being completely decommissioned. The vulnerabilities associated with SS7, Diameter, and GTP persist in 5G, as part of the network uses 2G for selected applications. The GTP protocol, considered vulnerable, has been a part of 2G, 3G, and 4G, and it will be a part of 5G-Standalone and 5G-Non Standalone. GTP is an important interconnection protocol for 4G and 5G which tunnels the user plane data across the roaming reference point. It tunnels the data from the RAN to the core network and vice-versa.

2.2 Previous work related to GTP vulnerabilities in 5G

The report by Positive Technologies [52] suggests that GTP in LTE networks is vulnerable to attacks because the GTP does not verify subscriber location and subscriber credentials are checked only at the S-Gw, which increases the chances of attacker impersonating the S-Gw [52]. Other vulnerabilities include disclosure of subscriber information (TEID and location data), spoofing and DoS attacks on network equipment. These GTP vulnerabilities will still exist after migration to 5G Standalone. The user plane will use GTP-U for encapsulating PDUs. In the past mobile generations, there have been attacks in which management protocol

packets were encapsulated in the user session. Park et al. [56] further explain how a GTP-in-GTP attack is carried out to force information leak and this method can be used to deplete IP pools allocated to terminals in the core network [56]. The GTP-C echo request is used to acquire IP for core network equipment, by sending a create session request to the core network. The attacker can increase the terminal number in the create session request so that PGw allocates multiple IPs. This way, creating session requests from the actual terminals will be rejected. An attacker can also send an attached request message to access the 5G NSA by configuring multiple terminals as botnets and repeating airplane mode on and off. This will bring excessive traffic to the core network. The scanning attack was introduced by ERNW at a 2011 conference [55]. It used GTP echo messages, that were sent to the system and the system responded. Using these messages IP address of the system can be checked with the response. Thus, this information is enough to prepare for an advanced attack. The Create Session Request message is a GTP-C message sent when a UE registers to the 5G NSA core network for the first time after it is switched on. The P-GW allocates an IP to the UE and transmits a reply. If a malicious attacker constantly transmits the create session message to P-GW, all IP resources of P-GW will be assigned, and the normal UEs will not be allocated an IP. When a malicious attacker transmits a create session Request message with a fraud number of a normal user to a 5G NSA network, the normal user will be unable to use the Vo5G voice services and data communication. The service is disturbed through the altering of the GTP message. Not all anomalies in the data in the messages can be detected by systems like P-GW that use the GTP protocol in the 5G network. Therefore, an attack can be launched using the erroneous Create Session Request message [55].

The 5G NSA will use the EPC core with an upgraded 5G RAN. In 5G SA, GTP-C for the control plane will be replaced by HTTP/2, but GTP-U will be used over N3 and N9 reference points. GTP was not designed with security mechanisms. These vulnerabilities were documented in GSMA FS.20 GTP-Security. These interfaces become vulnerable as attackers try to exploit GTP vulnerabilities. Another report claims that another way of attacking can be controlling IPX/GRX roaming links through cyber criminals and malicious peers [18]. Common GTP issues are confidential data disclosure, denial of service, network overloads, and other fraudulent activities. In LTE networks the roaming traffic through S8 has increased drastically after the launch of EU Roam like Home legislation in 2017. With 5G there will be the same roaming traffic in addition to roaming IoT devices. Since roaming traffic was relatively less in the past, the roaming operators had minimal to no security over the roaming interface. According to [25] half of the successful attacks were due to the major flaw that users' location is not verified. Testing confirmed that this vulnerability can be exploited via an inter-operator IPX network and in some cases from a mobile device [25]. The extensive use of GTP to shift networks makes it attractive for attackers.

The report published by SecurityGen [31] provides a detailed analysis of GTP test results obtained from testing different telecom networks. Figure 2.1 shows the protection levels of 2G, 3G, 4G and 5G networks. From the test results, it is clear that almost 54 percent of networks have medium protection levels, which means that the security mechanisms were unable to block all the illegitimate packets. Only 8 percent had low protection mechanisms.

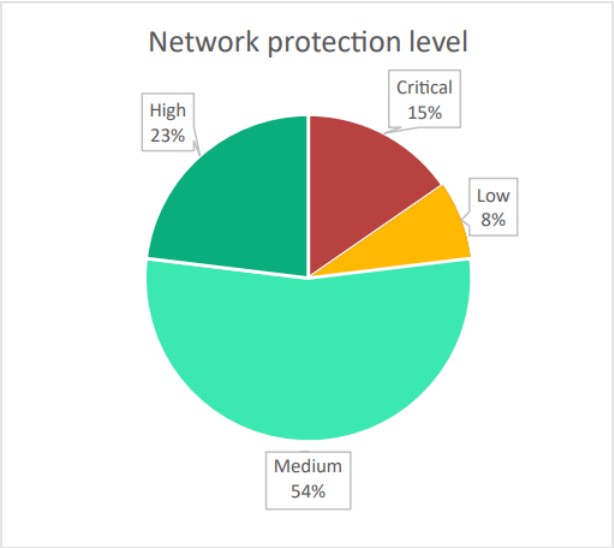


Figure 2.1: Protection levels of tested networks with exposed GTP [31]

2.2.1 Mitigation of GTP vulnerabilities

Some mitigation strategies suggested by [52] are the implementation of GSMA security recommendations such as monitoring and analyzing signaling traffic using a threat detection system. These systems also detect suspicious activities and block them. A regular security assessment is needed especially after reconfiguration and the addition of network equipment.

An Intrusion Detection System (IDS) with a traffic-based security threat detection system can be introduced as suggested in [56]. The information generated by the detection system is continuously monitored by the operators. The exploitation of a 5G NSA core network resource exhaustion attack involves sending a Create Session Request message from the device to deplete the IP address resources of the P-GW. To effectively counteract this traffic attack, it is crucial to capture and identify traffic from S1-U or S5 interfaces, which are channels for encapsulating and transmitting the attack through GTP-U. Implementing a packet-based algorithm becomes essential for discerning the transmission of Create Session Request messages from the device. Similar to the algorithm employed for scanning attacks, the process involves capturing outgoing packets destined for the 5G network and verifying whether the port number aligns with 2152 (GTP-U port) or 2123 (GTP-C port). Further scrutiny checks if the flag value corresponds to 0x4820, and the 12th 1-byte value, serving as a spare field, aligns with 0x00, indicating the payload as a Create Session Request message. Confirming the legitimacy of an attack packet entails comparing the length of the Create Session Request message, leveraging the fact that UDP length typically falls within the range of 200–280. This determination involves subtracting the header length (12) from the UDP length in the Create Session Request message [55].

To enhance the efficacy of post-attack detection measures, extracting the MSISDN is proposed. This involves scrutinizing the IP or TEID of the attacker's device and promptly reporting the findings to the EPC network manager.

Apart from these, GSMA FS.37 GTP-U Security has provided measures for the protection of the GTP-U protocol which are summarised in Table 2.1.

Table 2.1: Measures for protection of GTP-U [36]

Prevention Methods	Description	Effective against
Rate Limiting	Limiting the amount of traffic per subscriber to ensure network service availability.	DoS, DDoS, overloading
Anti-spoofing	It finds out the mismatch of subscriber identity between different layers. By monitoring GTP-C traffic with TEID, IP spoofing can be detected. GTP-U traffic is then monitored and if the IP address in the GTP-U packet does not match the one assigned to the TEID, spoofing is detected.	Spoofing attack
Barring GTP-in-GTP	GTP-U traffic embedded into GTP is not allowed and might be an indication of fraudulent activity. GTP-U traffic inside GTP will be barred.	GTP-in-GTP attack
N9 operator-to-operator security	N9 user plane traffic will be confidentiality, integrity and replay protected over IPX.	Roaming attacks
Flow policing	Inspection of GTP-U payload with application recognition.	Malicious applications, congestion prevention, overload
Traffic encryption for protecting user data	Using IPsec encryption for transfer of data over public internet and for dedicated leased lines.	False packet injection
GTP-U/C correlation	Inter-PLMNUP Security (IPUPS) at the perimeter of the PLMN in a home routing scenario protects user plane messages traversing through the N9 interface between a visitor network and a home network.	Roaming attacks

The GSMA has published security guidelines [36] for the implementation of secure GTP which are covered in Table 2.1. However, most of these guidelines are not properly implemented in the network. Another interesting observation from [31] is that none of the networks that were tested could not implement all the prevention methods recommended by the GSMA. In fact, only 8 percent of the networks had IP filtering of roaming partners. Figure 2.2 describes the protection measures implemented for GTP, 77 percent of the tested networks had no security measures in place.

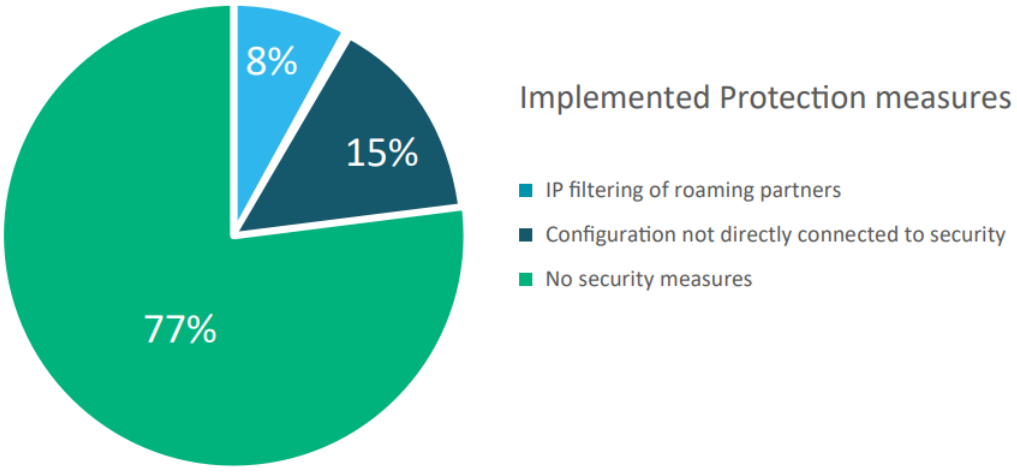


Figure 2.2: Protection levels of tested networks with exposed GTP [31]

2.3 Existing research on N-32 vulnerabilities

Few research studies have critically analyzed the problems with the N-32 protection mechanism called PRINS and vulnerabilities of the 5G core network protocol HTTP/2. The author Geir M Køien of [45] concluded that over N32, the associated roaming partner needs to be trusted to a certain degree. Full trust is necessary for IPX operators. Additionally, verifying consistency and completeness becomes challenging in the absence of formally defined JSON requirements. JSON is a data interchange format used for exchanging information between different network components and entities. In a multi-operator environment, such as the N32 interface, processing might reveal JSON inconsistencies and inherent library incompatibilities. These include whitespace handling (JSON does not specify whitespace handling), Library-specific features (Some libraries support additional features that might be incompatible with other libraries) and Data Types (JSON supports a limited set of data types, including strings, numbers, booleans, arrays, objects, and null.) The JOSE standard, known for its complexity, offers numerous opportunities for inappropriate usage. The standard may also have additional problems. Also suggested by [62] there are two deployment options for security, one is TLS and the other is PRINS. TLS is the most secure variant with all signaling encrypted end-to-end in TLS tunnels between mobile roaming partners. However, this solution is complex to operate with nonroaming connections which more principally, completely excludes the role of roaming VAS operators, and roaming HUBs and degrades the role of IPX carriers to an IP routing service. So, this option only works for the top roaming relations but not for all roaming relations in the global ecosystem of around 800 mobile network operators. The other option, PRINS is an application layer security solution whereby part of the signaling information is sent in the clear, to enable roaming VAS operators, roaming HUBs, and for IPX carriers to inspect and/or modify signaling traffic in transit. This is technically rather complex and comes with operational hassle as the terminating mobile operators need to verify the modifications from the intermediate carriers. More problematic is that the sending operator is not in control of what is modified and by whom, so again an

open door to fraud/abuse and not resolving all the present vulnerabilities with 2G/3G/LTE roaming.

2.4 Application Fingerprinting for Mobile Networks

The term Application Fingerprinting is a more generic term used for identifying web applications based on their specific and unique features. This technique has been used for many years to extract relevant information from mobile data traffic patterns. According to Eurostat [50], mobile phone call data is considered to be a vital source for the compilation and enhancement of official statistics on tourism, economy, environment, crowd flows, and mobility. This is also what has spurred Statistics Netherlands (CBS) to research the use of aggregated anonymized mobile call data for statistics since 2009. The CBS is no exception in this regard. Around the world, not only organizations such as universities, and national statistical offices are using traffic analysis but also commercial organizations are developing methods and software for the analysis and processing of telephony data for statistics, science, and policy development.

Mobile traffic identification plays an important role in network management, marketing research, and user characteristic analysis. For example, based on this technology, a network administrator can obtain the popular apps in the network and optimize the resource allocation accordingly to improve the user experience. A company can monitor whether employees use prohibited apps during working hours, such as games and shopping. For advertisers, understanding a certain app is popular with users in which area and period can help them create a better advertising strategy. For market researchers, understanding the use of apps of concerned users can help them analyze the interests and needs of users; then further business activities can be carried out. For example, if a person uses a flight booking app frequently, then the user may be a potential customer of travel services.

First, mobile traffic is always carried over HTTP/HTTPS, making the port-based approach identify mobile traffic as Web only. Second, lots of apps use encryption protocols for data transmission to protect user privacy. Indeed, some encryption protocols may expose useful information during its negotiation process, such as the TLS SNI (Server Name Indication), so that part of the encrypted traffic can be identified by the DP (Deep Packet Inspection) approach [67]. However, the SNI field sometimes is blank and not every SSL/TLS connection has a negotiation phase, which decreases the effectiveness of this method. For MOMO, which is a social app in China, random 50 HTTPS connections were checked, and out of those 50, 9 connections did not have a negotiation process and 2 Security and Communication Networks connections had a negotiation process but the SNI field was blank. Third, mobile apps often access third-party libraries, resulting in the fact that different apps would generate similar traffic. It is difficult to discriminate such traffic via DPI technology or IP address. This problem can be circumvented if such traffic is considered an individual category. Fourth, CDN (Content Delivery Network) is used by many apps to improve the user experience. As a result, a server's IP address can be shared by multiple apps.

To preserve user integrity and confidentiality, an additional protection mechanism that uses randomization has been proposed in Chapters 5 and 6 of the thesis.

2.5 Randomisation for 5G Network

The work done in [26] presents a privacy-preserving technique effective against application fingerprinting. The authors have presented two ways for anonymizing application traffic, either (1) the distribution of one app. can be made similar to the distribution of all other apps implying similar distribution for all the apps or (2) the length of data packets could be mutated in such a way that the app length distribution of one app resembles the other app length distribution e.g. modifying Facebook´distribution so that it becomes similar to the distribution of Youtube. This method successfully masks the signature distribution of any application at the cost of increased processing time and data length. The distribution is plotted between inter-arrival time and packet lengths. With such a distribution each application can be classified accurately, which can be dangerous if the traffic is intercepted by a potential attacker. The Figure 2.3 shows different distributions belonging to different apps based on their traffic. Thus, randomization can be used to provide additional security to the network. In this thesis, the Randomisation concept is considered for 5G networks as an additional security mechanism for roaming reference points.

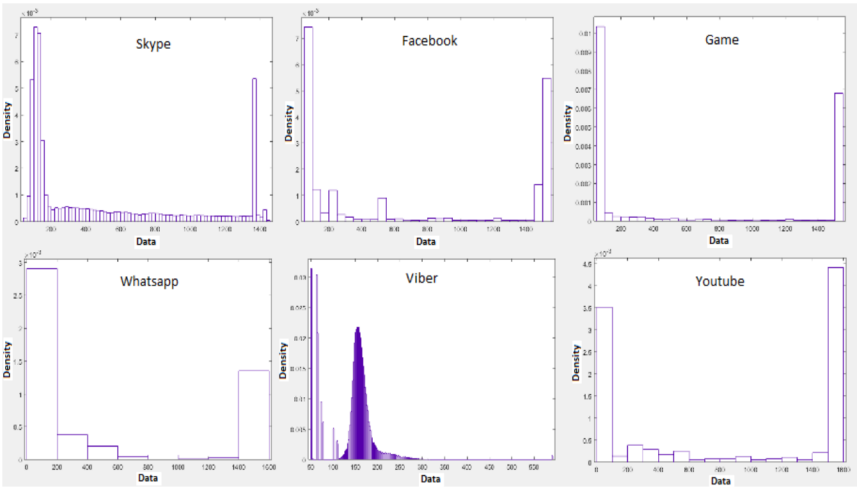


Figure 2.3: Different applications having unique packet length distribution. the Y axis shows the density and the x axis shows the application data. [26]

Chapter 3

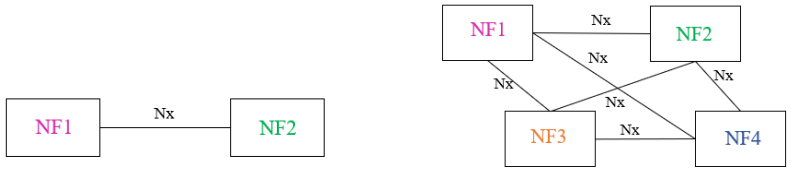
5G Network Architecture

This chapter consists of 5G Network Architecture details including a description of Core Network and Radio Access Network, related protocols as well as 5G system procedures.

3.1 5G Network Architecture

The Third Generation Partnership Project (3GPP) released 5G Standalone specification in 2018, after the release of the Non-Standalone specification in 2017 which involved EPC with an upgraded eNodeB. Later, Phase-2 of 5G specification was released which covered the entire Next Generation Core, NG-RAN along with Vehicle-to-Vehicle communication, and Network-Slicing aspects. The fifth generation of mobile networks promises high-speed data connectivity, low latency, increased bandwidth coverage, reduced power consumption, and an increased number of end device connections as described in Chapter 1 (Section 1.1). It uses a higher frequency spectrum to provide increased data rates and virtualized network functions for flexibility. The 5G architecture can be represented through reference point-based architecture and service-based architecture.

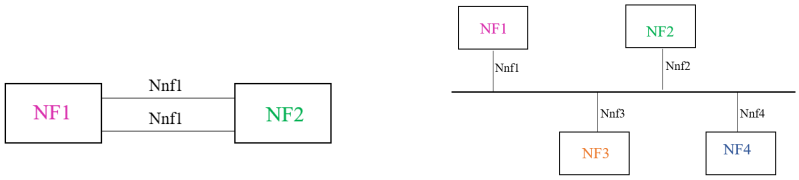
- Reference point architecture - A reference point forms a logical connection that joins two non-overlapping functional groups. In this representation, shown in Figure 3.1, Network Functions (NF) have a one-to-one connection with each other and communicate with the help of the Nx reference point. The entire 5G network can be modeled as a reference point-based architecture.



(a) Nx connecting NF1 and NF2 **(b)** Nx connecting multiple NFs

Figure 3.1: Connection of NFs using reference points Nx where x can be any integer [9].

- Service-based architecture - In this representation, virtualized network functions (VNFs) offer services to each other or other authorized functions using the restful APIs. The connection between NF1 and NF2 is called the service-based interface denoted by Nnf-name. Nnf1 and Nnf2 are equivalent to Nx. The 5G core network follows a service-based architecture. In Figure 3.2, multiple NFs will interact using a message bus which forms a common interface to support communication between every NF. It may include reference points where necessary.



(a) Service-based interface Nnf connecting NF1 and NF2

(b) Service based interface Nnf connecting multiple NFs

Figure 3.2: Connection of NFs using Service-based interface Nnf [9].

The service-based architecture and reference point-based architecture are two different models for representing network function interactions within a network. Till now, only reference point-based architecture was in use as 2G, 3G, and 4G were modeled in this manner. However, with the upcoming 5G network, the service-based architecture was introduced for the 5G Core Network. The inclusion of this architecture in the core network saved a lot of bandwidth as a common message bus is used instead of point-to-point links, and adding and deleting new NFs became easier, making the network more efficient and flexible. The entire 5G System architecture can be decomposed into a 5G Core Network and Next Generation-Radio Access Network (NG-RAN) with the UE directly connected to the RAN as shown in Figure 3.3:

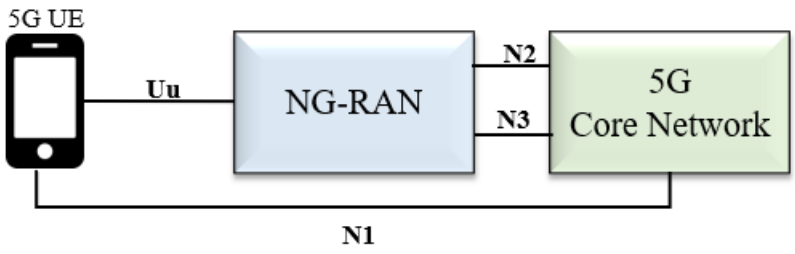


Figure 3.3: 5G Network Schematic [9]

3.2 User Equipment (UE)

The User Equipment (UE) is a mobile wireless device that allows the user to access the 5G network services. There are two frequency ranges in which 5G will operate, namely FR1, which extends from 450 Mhz to 6000 Mhz and FR2, which extends from 24250 MHz to 52600 MHz [4]. The UE is compatible with both the frequency bands, FR1 and FR2 supported by the 5G NR. The UE can connect to the core network directly over N1 as well as Access Network over the N2 interface. The UE in 5GS contains Universal Subscriber Identity Module (USIM) which is similar to 3G and 4G. In previous generations, a unique identifier known as IMSI was allocated to every SIM. In 5G NR, the Subscription Permanent Identifier (SUPI) is allocated to the UE or SIM by the network operator for identification. The authentication between a user and the network is done with a shared key after the identification. The UE detects signals that the gNodeB broadcasts over the air interface and attaches itself to the network after the exchange of NAS signaling and network identifiers. Some of these identifiers are listed below [9] :

- **Subscription Permanent Identifier (SUPI)** is provisioned in the SIM and UDM/UDR. It contains IMSI and it is capable of interworking with EPC for non-standalone deployments. The SUPI is concealed in SUCI to prevent the plaintext transmission of SUPI over the air interface.
- **5G Global Unique Temporary Identifier (5G-GUTI)** is allocated by the AMF. It is a temporary identifier used for identification over the radio link. It is a combination of PLMN ID, AMF ID, and TMSI.
- **Permanent Equipment Identifier (PEI)** is assigned to the UE for accessing the network.

3.3 5G Core Network Elements

The 5G core network follows a service-based architecture model, which means that each network element or virtualized network function (VNF) will provide its services to other authorized VNFs or network elements. The access network allows the user to access the services of the core network which includes authentication, charging, connection establishment, handover support, etc. The 5G core is an access-independent or unified core network. The components of the core network are explicitly defined in 3GPP TS 23.501 [9]. Figure 3.4 shows major core network functions in a reference point representation. The actual interaction between core network functions is service-based interaction which is explained in Section 3.3.2.
The major NFs forming the 5G core are described below:

- Access and Mobility Management Function (AMF)
  This function can be compared to the MME in a 4G network. However, AMF is responsible for UE registration, connection management, and mobility procedures, while session establishment is handled by SMF. All the session management information is transferred over reference point N11 by the AMF. The AMF also supports overload prevention mechanisms: N2 overload control and NAS congestion control. The AMF is identified by Globally Unique AMF Identifier (GUAMI) which acts as an AMF ID.

This ID is specified by the UE during the first NAS signaling and ensures the proper selection of the AMF. It acts as a termination point of the N2, and N1 interface along with NAS ciphering and integrity protection. In addition to the above, AMF also supports non-3GPP access over N3IWF/TNGF. It provides mechanisms for NAS signalling, UE authentication, and mobility management for Non-3GPP access. The AMF is responsible for SMF selection.

- Session Management Function (SMF)
  In 5G, SMF is responsible for handling all the session-related activities, similar to the MME of LTE. In the case of roaming, the home PLMN´s SMF can be called h-SMF while the visited PLMN´s SMF is called v-SMF, further details follow in Chapter 5, Section 5.1.1, which covers roaming scenarios thoroughly. It handles PDU session establishment, modification, and release by maintaining the tunnel between the access network and UPF. The SMF is responsible for assigning the IP address to the UE and routes the UPF traffic to specific destinations. It acts as a DHCP server while allocating the IP address and client while requesting IPs from the external servers. It supports charging data collection, controls the UPF, and transports the session management information via AMF to AN over the N2 interface. It is also responsible for determining the SSC mode of a session. It also provides a lawful interception feature, which provides the operator the authority to track and trace any PDU session.

- Policy Control Function (PCF)
  The PCF can be compared to the PCRF in the 4G core network. It fetches the subscriber details from the UDM for making policy-related decisions and provides the control plane function with rules. It governs network behavior by supporting a unified policy framework.

- Authentication server Function (AUSF)
  Its role is similar to AuC in 4G. It provides authentication to users wanting to access the network for 3GPP access and non-3GPP access.

- Unified Data management (UDM)
  The UDM functionality is similar to HSS. It stores subscriber information. UDM is usually stateful, as it stores the data in local memory, but it can be stateless wherein the information is stored in UDR. It provides service to AMF, SMF and NEF. It is also responsible for 5G-AKA credentials for UEs serving registration management. The UDM contains the subscriber data, including both the static subscription data and the dynamic location information. The Nudm is a service-based interface (SBI) for UDM, which accomplishes all the functions of UDM. The UDM can interact with the following NFs with the help of certain reference points [9].

- User Plane Function (UPF)
  UPF is comparable to PDN Gateway in 4G. UPF establishes a connection with the data network for the PDU session, which is a data connection. The PDU session gets established between UPF and DN and data transfer takes place through the PDU session. Moreover, it also allocates UE IP address/prefix. It forms an anchor point for Inter/Intra-RAT mobility. It supports IPUPS functionality to safeguard the network

from invalid inter PLMN N9 traffic. The SMF is responsible for activating IPUPS. Lawful intercept, policy enforcement and QOS charging, traffic reporting.

- Data Network (DN)
  The data network constitutes the applications in the operator internet and public internet.

- Network Exposure Function (NEF)
  The Network Exposure Function (NEF) exposes the overall 5G core network services and capabilities.

- Network Slice Selection Function (NSSF)
  The Network Slice Selection Function (NSSF) selects the Network Slice Instance (NSI) based on information provided during UE attach. A set of Access and Mobility Management Functions (AMF) are provided to the UE based on the slices accessible to the UE.

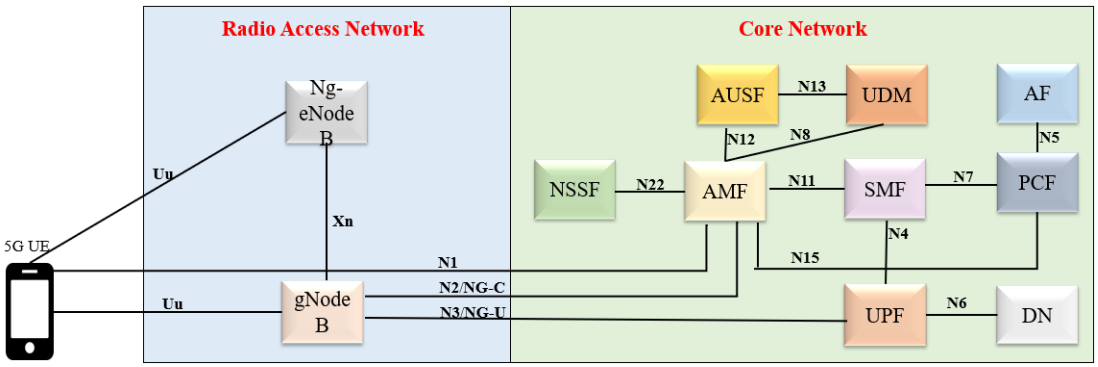


Figure 3.4: 5G network having 4G connectivity in Reference point representation [9]

3.3.1 Overview of 5G Reference Points

The major reference points connecting various NFs are present in Figure 3.4. The description of each reference point is given in Table 3.1. The Reference point-based representation is used in 5G SA Radio Access Network. To define interactions between these NFs, a protocol is also defined per reference point. Thus, for every reference point, there is a separate protocol that runs over that reference point and governs the interaction between respective NFs. The reference points for the RAN network will be taken up in Section 3.6.

Table 3.1: Description of major 5G-SA Reference Points [10]

Reference Point	Description
N1	It is the interface between the AMF and the UE for non-3GPP access. It supports NAS signaling for security control, PDU session establishment, modification, UE registration, and mobility. Multiple N1 instances are supported.
N2	The NAS control plane signalling across the AMF and the gNodeB runs over N2.The N-GAP protocol runs over N2. The SCTP ensures the delivery of messages belonging to the application layer.
N3	It comprises the user plane for data transfer between the gNodeB and the UPF. The GTP-U tunnel runs over it during PDU session establishment.
N4	The AMF communicates with the other SMF over N4 reference point with the help of PRFCP protocol.
N5	The PCF and AF communicate over N5 for policy and charging control for roaming and LBO scenarios.
N6	It connects UPF to the DN, thus providing connectivity between UPF and the external network or internet.
N7	It is present between SMF and PCF. The usage reporting combines SMF and PCF data over N7.
N8	During registration procedure, the subscriber data is retrieved from UDM over N8, as it links AMF and UDM.
N9	It is present between UPF of home PLMN and UPF of visited PLMN.
N11	It is the reference point between AMF and SMF. The signaling messages for adding, modifying, or deleting a PDU session are exchanged over N11. The AMF transfers information to the SMF relevant to PDU session establishment.

3.3.2 5G Core Network Architecture (Service-based representation)

The core network functions were introduced in Section 3.3. In this section, the actual interaction amongst these core network functions will be discussed along with design principles and benefits of incorporating service-based, software, and cloudified core networks. A core network function NRF provides service discovery between NFs and maintains a database of various NF instances along with their supported service (function type, function ID, network slice identifiers, capacity information, IP addresses). The core NFs which are a part of the control plane interact with each other via NRF over SBI using HTTP/2 transport [3]. Each NF acts as an independent module which can be reused similarly to a microservice and each NF can act both as a producer or consumer for a particular service. The consumer NF can request a response from the producer NF. In this model, the communication between NFs uses HTTP-based APIs, analogous to Diameter in previous generations. This shift towards a service-based core has improved extensibility, efficiency, and QoS.
The service-based architecture was introduced in Section 3.3. The 5G offers versatile features

such as network slicing, multiple connected devices, edge computing, and vendor-independent architecture. Hence, the 5G SA core components are implemented inside containers, as opposed to virtual machines in EPC, and communicate via HTTP/2. This design helps in reducing power consumption, enabling more device connections, flexibility in adding new functionalities, and bringing more network designing options for MNOs. Also, enabling distributed, cloudified RANs can provide low latency for enhanced streaming and gaming experience. The service-based architecture is shown in Figure 3.5.

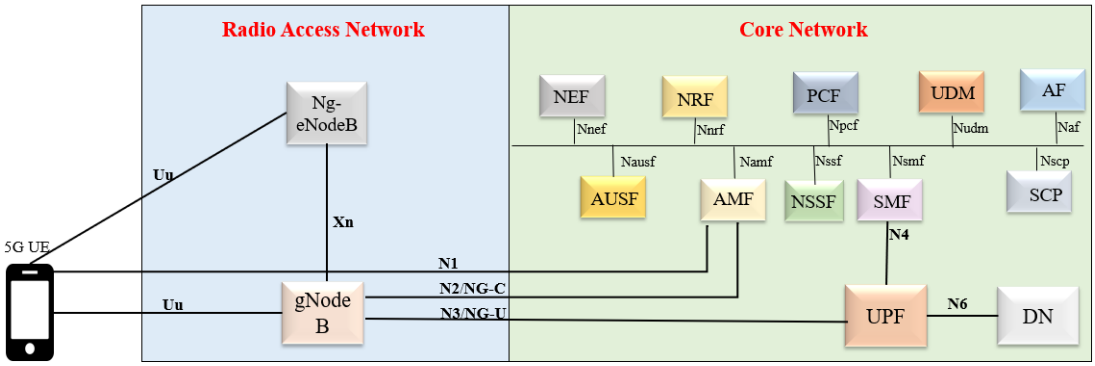


Figure 3.5: Service-based architecture of 5G core network elements [3]

3.3.3 Protocols over Service Based Interface

The SBI architecture protocol stack is described in Figure 3.6. The HTTP/2 is an application for transferring messages. The control-related messages are written in JSON format which is a serialization protocol. The HTTP/2 is used to transport messages written in JSON format [3]. The HTTP/2 performs operations such as GET, DELETE, PUT, etc. while communicating with other NFs. Any NF can be a producer or consumer or both.

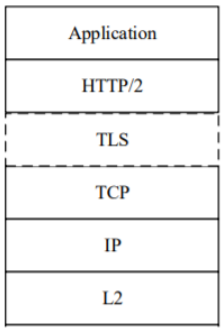


Figure 3.6: 5G Protocol Stack for the Core Network [3]

3.4 Radio Access Network Architecture

The radio access network utilizes radio frequencies to provide core network access to the user equipment, acting as a communication medium between the user equipment (UE) and the core network (CN). In mobile communication, a site is divided into multiple cells and each cell can be served by one or more radio transceivers. Thus, RAN components provide cell coverage, error detection, and correction, power management, bandwidth management, etc. The following RAN components are crucial for realizing these functions.

- Baseband Unit (BBU) - It manages signal processing for error detection and correction, resource utilization, and securing the wireless links. Earlier, these were present just below the antenna tower but in 5G NR, these can be present anywhere as a virtualized function.
- Remote Radio Head (RRH) - It converts digital information into signals required for wireless communication and serving at different frequencies.
- Antenna - It emits electrical signals into radio waves. It continuously emits radio signals that are detected by the UE during the attachment procedure and receives radio signals from the UE.

3.5 5G RAN Architecture

The Next Generation-RAN (NG-RAN) or 5G RAN is divided into the Transport Network Layer (TNL) and Radio Network Layer (RNL). The RAN nodes and the interfaces between them are a part of the RNL while the TNL is responsible for the transport of signaling and user plane data. The protocols that run over the Uu and NG interface can be subdivided into control plane and user plane protocols. The gNode-B has a layered architecture and the separation of the user plane and control plane has been followed which is explained below [5]:

3.5.1 Integrated gNode-B

The RAN architecture is relatively flat as there is no aggregation which is evident from Figure 3.7. The RAN architecture is made up of an integrated g-NodeB. In 5G RAN, multiple g-NodeBs will be connected to the 5G core network. Each of these g-NodeBs consists of an antenna, a Baseband Unit (BBU) and Remote Radio head (RRH). The RRH consists of analog-to-digital and digital-to-analog converters along with RF circuits. In the transmission segment of the Remote Radio Head (RRH), the digital signal undergoes conversion to radio frequency (RF), followed by amplification to reach the desired power level. Subsequently, the antenna linked to the RRH broadcasts the RF signal into the surrounding environment. On the receiving end, within the RRH's receiver segment, the antenna captures the targeted signal band. The received RF signal is then both amplified and transformed back into a digital signal within the receiver chain. It's noteworthy that the RRH is strategically positioned near the antenna [59].

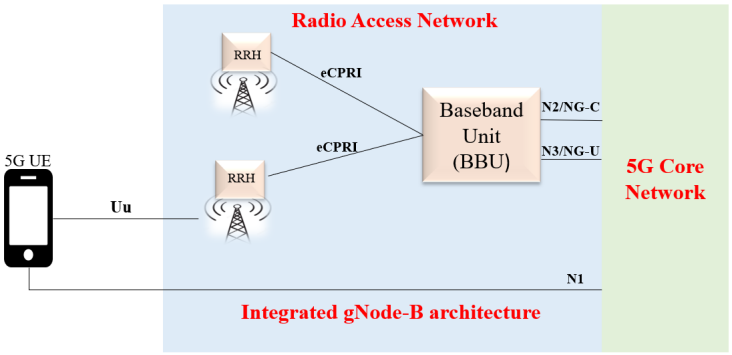


Figure 3.7: Components of Integrated gNodeB [59]

3.5.2 Integrated gNodeB with Centralised and Distributed Baseband unit

The baseband unit is further split up into Centralised-BBU (BBU-C) and Distributed-BBU (BBU-D) as shown in Figure 3.8. This is done to reduce the load on the transport network and provide more flexibility as distributed units can be deployed close to the RRH. The distributed unit (BBU-D) is under centralised-BBU and provides functionalities such as real-time L1, and L2 scheduling while BBU-C handles non-real-time L3 functions. The lower layers namely RLC, MAC, and PHY can be processed by the Distributed Baseband Unit (BBU-D) while higher layers (SDAP and PDCP) can be processed in the Centralised Baseband (BBU-C) Unit. The processing of the radio protocol stack is variable and depends on the type of split option which is discussed in Section 3.5.4.

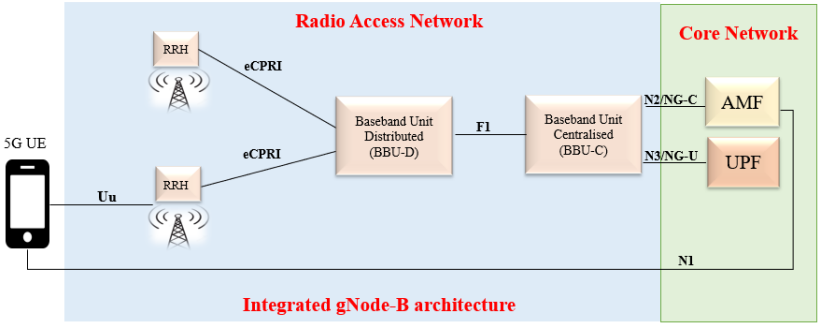


Figure 3.8: Integrated gNodeB with split Baseband Unit [59]

3.5.3 Integrated gNodeB with split centralized Baseband Unit

The BBU-C is further split into the User Plane and Control Plane as shown in Figure 3.9. In the 5G SA, the NG interface connects the gNB to 5G Core, analogous to the S1 interface in 4G LTE that connects eNB to EPC. The NG-C permits signaling between a gNB and an AMF. NG-U permits the transfer of application data between a gNB and a UPF [59].

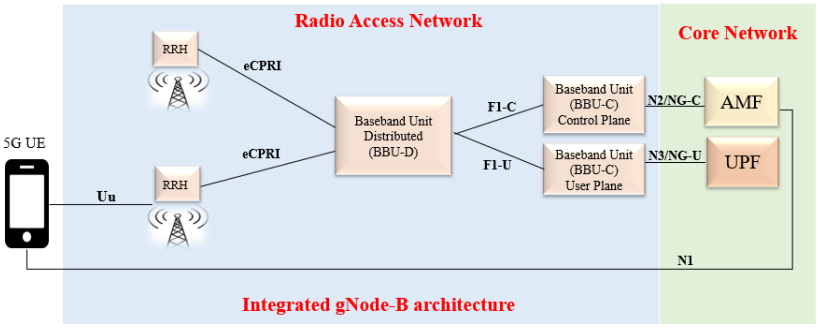


Figure 3.9: Control and User plane separation inside Baseband Unit [59]

3.5.4 Radio Network Layer Split Options

The radio protocol stack for the user plane and control plane consists of the three protocol layers mainly L1 (physical sub-layer), L2 (MAC and RLC sub-layer), and L3 (PDCP, SDAP, and RRC sub-layer). These sub-layers can be mapped to different gNodeB units (BBU-C and BBU-D). There can be eight possible options for the functional split of BBU [41]. In option 1, RRC will be processed in the central unit while PDCP, RLC, MAC, physical layer, and RF will be processed in the distributed unit. Similarly, in option 6, the MAC and upper layers are in the central unit (CU) while the PHY layer and RF are mapped to the DU. The selection of a split option is completely dependent on the use case, network deployment, and transport network requirement.

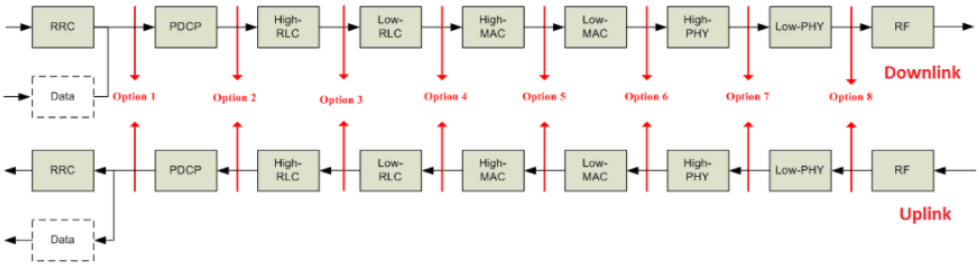


Figure 3.10: RAN split options [41]

3.6 Major RAN Reference Points and Protocols

The major RAN Reference Points that connect different gNodeB components and the main protocol that runs over those Reference Points are described below:

3.6.1 Xn Reference Points

The Xn Reference Point is a logical interface that connects two gNode-Bs or an eNode-B to a gNode-B. It supports mobility functions, dual connectivity, and data forwarding. The

Xn-AP is an application layer signaling protocol that runs over the Xn. It helps in the exchange of signaling messages between two NG-RAN nodes and then sends the PDUs to the tunnel endpoints [14]. The Xn control plane and user plane stacks are in Figure 3.11. The Xn-Application protocol (Xn-AP) runs on top of the Stream Control Transmission Protocol (SCTP), which is a transport layer protocol. SCTP guarantees the delivery of signaling messages between two endpoints. The Xn user plane supports the GPRS Tunneling Protocol-User (GTP-U) for PDU forwarding. It transports PDUs and signaling messages between the tunnel endpoints. The Tunnel Endpoint ID (TEID) which is part of the GTP header, helps in identifying, to which tunnel a particular PDU belongs to. It runs over the UDP transport protocol which does not guarantee the user data delivery. The GTP-U is responsible for creating, releasing, and changing the tunnel for the transfer of IP packets [14].

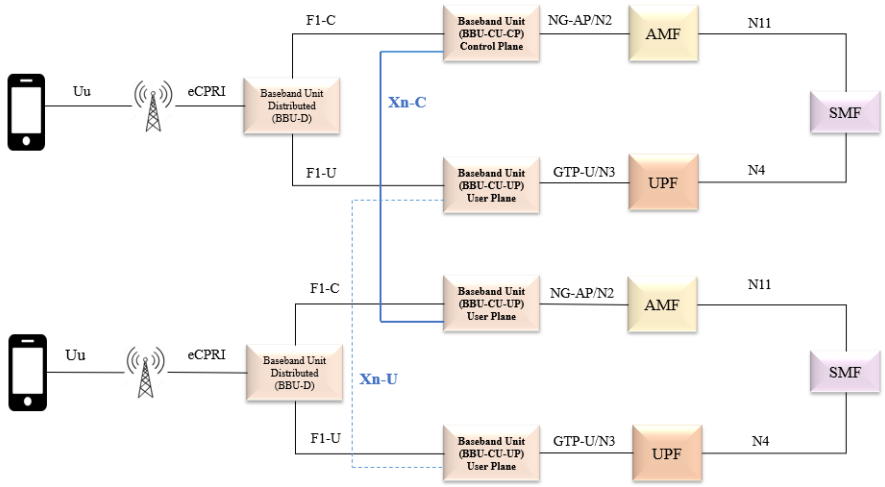
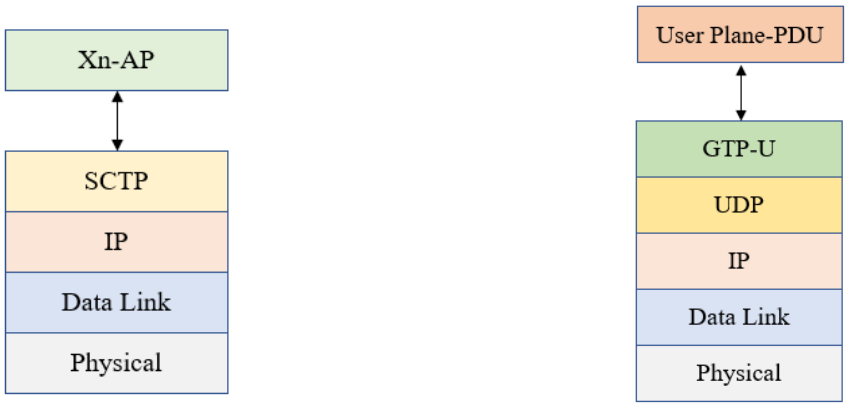


Figure 3.11: Xn reference point between BBU-CU-CP and BBU-CU-UP of two different gNodeBs [14]



(a) Xn-C **(b)** Xn-U

Figure 3.12: Xn control plane and user plane protocols running between two gNodeBs [14]

3.6.2 NG Interface

The NG-C Reference Point connects BBU-CU-CP with AMF and BBU-CU-UP with UPF. The NG is also divided into the NG-User plane (NG-U) and NG-Control plane (NG-C) similar to the Xn interface. The NG-C is also referred to as N2 when using the reference point representation of core network architecture. The NG-Application protocol (NG-AP) runs on top of SCTP/IP between BBU-CU-CP and the AMF. It is responsible for procedures to perform inter-RAT handover and intra-RAT handover, the segregation of UE on the protocol level for user-related signaling management, and the transfer of NAS signaling messages between AMF and UE. Similarly, NG-U, also known as N3, provides nonguaranteed delivery of PDUs using the GTP-U protocol. The user data or the payload is transported from BBU-CU-UP to UPF over the NG-U interface. The protocol stack [13] of NG-U and NG-C is depicted in Figure 3.14 [13].

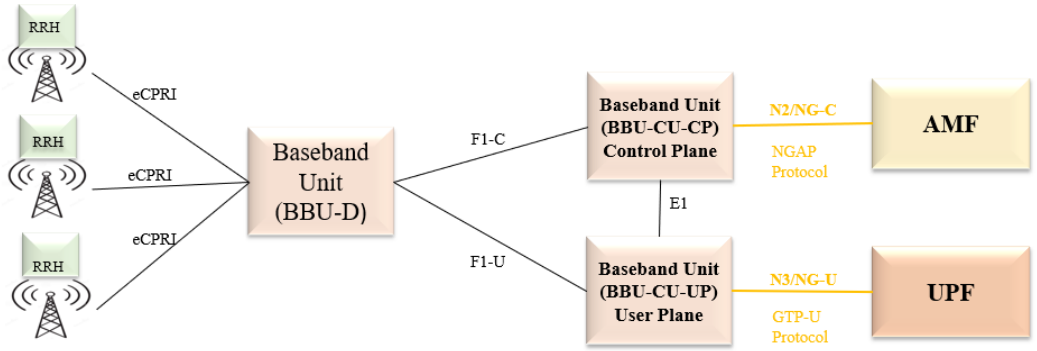


Figure 3.13: NG interface running between Access Network and Core Network [13]

(a) gNodeB and Core network control plane **(b)** gNodeB and Core network user plane

Figure 3.14: NG control plane and user plane between BBU-C and AMF, UPF [13]

3.6.3 E1 Interface

The E1 interface links the BBU-CU-UP and the BBU-CU-CP as shown in Figure 3.15. It separates the radio network layer from the transport network layer. It is responsible for error indication, interface management, and bearer management for the UE [15]. The application level protocol is E1-Application Protocol (E1-AP) which runs over SCTP/IP as shown in Figure 3.16. E1-AP controls the traversal of the payload tunnel through BBU-CU-UP.

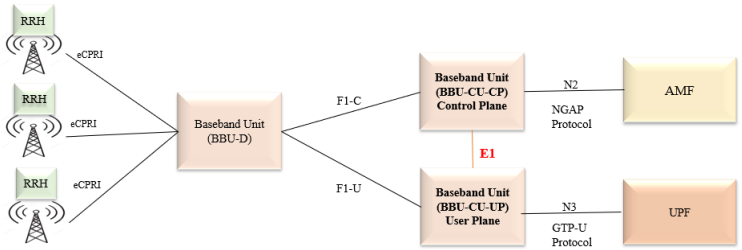


Figure 3.15: RAN Deployment scenarios [15]

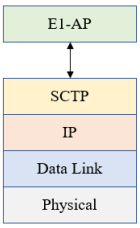


Figure 3.16: E1 Protocol stack [15]

3.6.4 F1 Interface

The F1 interface connects centralized BBU and distributed BBU visible in Figure 3.17. It also supports the interconnection of a gNB-DU and a gNB-CU belonging to the different manufacturers. The protocol stack of F1-U and F1-C is similar to Xn and NG stack, with the only difference being the F1-AP protocol which runs over the F1-C interface between distributed Baseband unit and centralized baseband unit control plane as shown in Figure 3.18a. The F1-C provides functionalities such as error indication, reset function, exchange application-level data, paging, UE context management, trace function, etc. The F1-U supports the transfer of user data and flow control functions with the help of GTP-U protocol [16]. F1 is made up of Elementary Procedures (EP), which are defined separately and used for making complete sequences. There are two kinds of EPs: one with response and the other without response as specified in [16].

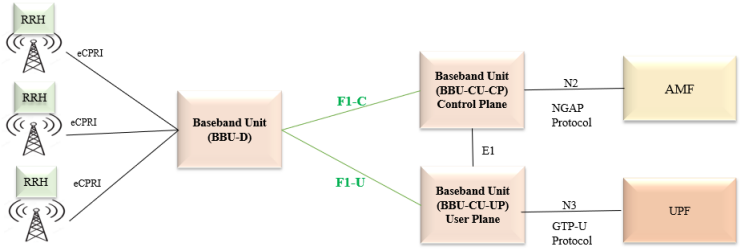


Figure 3.17: F1 runs between BBU-D and BBU-C [16]

(a) F1 control plane between BBU-D and BBU-CU-CP **(b)** F1 control plane between BBU-D and BBU-CU-UP

Figure 3.18: F1 User and Control Plane Protocol stacks [16]

3.6.5 Radio Protocol Architecture

The combination of the three protocol layers, which are also the lower layers mainly L1 (physical layer), L2 (data link layer), and L3 (network layer) forms the radio protocol stack. These are responsible for header compression, segmentation, and error correction. The radio protocol architecture for the user plane and control plane is in Figures 3.21, 3.20. It runs over the Uu interface between the UE and the gNodeB. The protocol stack for 5G is similar to the protocol stack of LTE except for a few layers. The L2 layers are the same for both the user plane and the control plane. In the control plane, control-related information is transmitted between the gNodeB and the UE while the user plane sends the data packets for payload. Each of these sublayers is explained below:

Non-Access Stratum (NAS) layer - For network registration and deregistration, PDU session establishment, PDU session release, mobility management, and security control, Non-Access Stratum (NAS) signaling transfer takes place over the control plane [12], [5]. It is the highest stratum between the AMF and the UE. It supports both 3GPP and non-3GPP access.

Radio Resource Control (RRC) layer - It is responsible for broadcasting the system information related to NAS and AS, NAS message transfer, mobility, and maintaining RRC connection between UE and NG-RAN. RRC protocol is used over the air interface or the Uu interface. There are three RRC states defined for 5G NR: RRC idle, RRC inactive, and RRC connected [5]. These different states are associated with different amounts of radio resources. When the UE is switched ON it is in the RRC idle state and moves to the RRC connected state after attachment. If there is a period of inactivity it goes into RRC inactive to save power.

Packet Data Convergence protocol (PDCP) layer - The PDCP layer is a part of both the user plane as well as control plane protocol stack. It compresses only the user data packets and not the control plane signaling messages, using the Robust Header Compression (ROHC) algorithm. It provides services only for radio bearers which are mapped on DCCH and DTCH type logical channels. A single radio bearer is associated with only one PDCP entity. Each PDCP entity can be associated with 1, 2, or 4 RLC entities and it carries data of one radio bearer. The Packet Data Convergence protocol takes the packets from RRC and transfers them to RLC after adding the PDCP header. These data packets are now referred to as PDCP PDUs and the outgoing packets will be called RLC SDUs for RLC sublayer. The PDCP layer adds PDCP sequence numbers to the data packets coming from the RRC. It also provides integrity protection and ciphering for control and user plane messages. The PDCP sublayer is configured by higher layers. The data is first stored in the transmission buffer and undergoes PDCP sequence numbering. Then the user data goes through header compression and the integrity protection is done to the control plane messages. Then ciphering is performed to both the control plane and user plane data. The PDCP protects the data from UE to the BBU-Centralised as it runs between UE and the Access network.

Radio Link Control (RLC)layer - The Radio Link Control layer supports three transmission modes: Unacknowledged Mode (UM mode), Transparent Mode (TM mode), and Acknowledged Mode (AM) Transfer of upper layer PDUs. It performs sequence numbering, duplication detection, segmentation, and error correction through Automatic Repeat Request (ARQ). The ARQ retransmits RLC SDU segments based on the status report. An SDU is

an incoming data packet that needs processing.
Medium Access Control (MAC) layer - It performs mapping between logical channels and transport channels [1]. It performs error correction HARQ whenever necessary and prioritization between UE through dynamic scheduling.
Service Data Adaption (SDAP) layer - The Service Data Adaption layer in the user plane stack is responsible for mapping QoS Flow within a PDU session to a Data Radio Bearer (DRB) and marks the transmitted data packets with QoS Flow ID, necessary for correct forwarding.
Physical (PHY) layer - Finally, the physical layer communicates directly with the user equipment through the physical channels. Physical channel characteristics include timing, access protocols, and data rates.

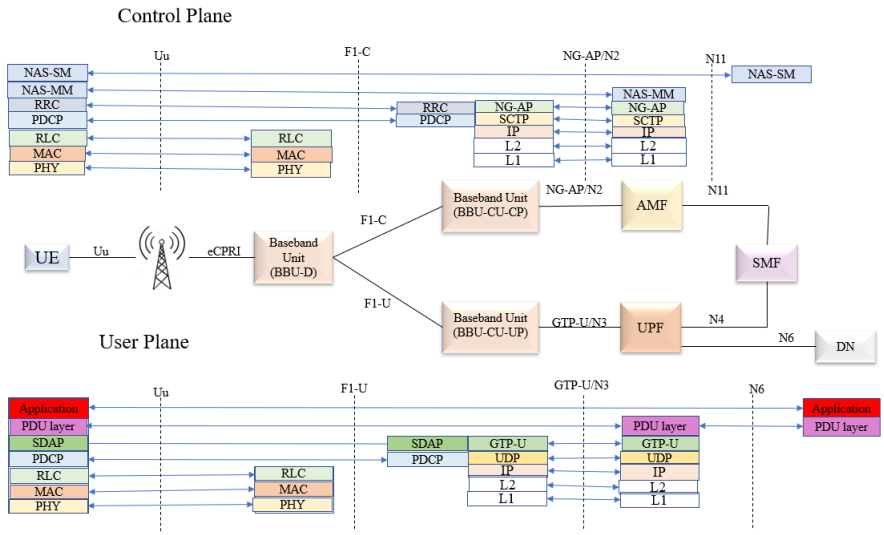


Figure 3.19: Representation of termination points of various NAS and AS protocols [9]

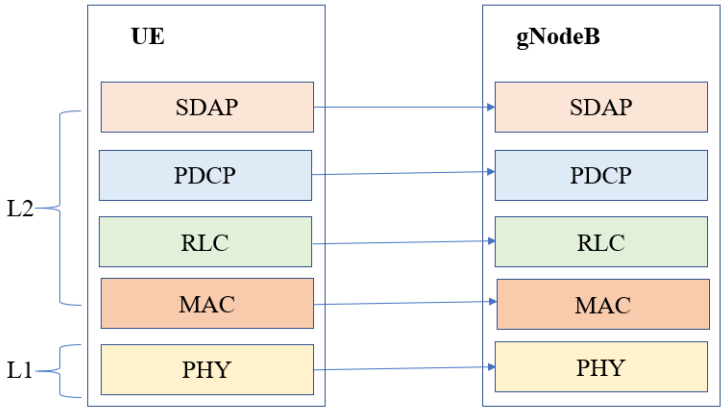


Figure 3.20: 5G User Plane Protocol stack [9]

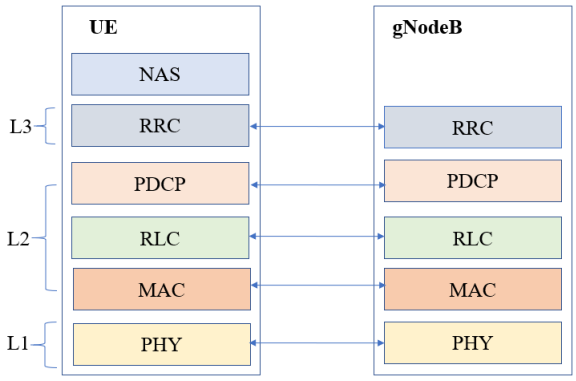


Figure 3.21: 5G Control Plane Protocol stack [9]

3.6.6 Transport Network

The transport network manages control and user plane data transfer between access and core network, and also within the access network. The **fronthaul** network is between RRU (Remote Radio Unit) and DU (CPRI and eCPRI interfaces), **midhaul** is the network between DU and CU, and **backhaul** network is between CU and core network as shown in Figure 3.22. 5G supports low latency and higher throughputs, which calls for a high-speed connection at the front end. The LTE network uses CPRI for fronthaul. Following are the parts of the transport network as shown in Table 3.3 [41], [44]:

- Fronthaul Transport - For fronthaul networks, CPRI has been used in previous generations. The Common Public Radio Interface (CPRI) provides standardizes the connection between a Radio Equipment Controller (REC) and a Radio Equipment (RE) for wireless medium. In LTE RAN, the CPRI is used for the transport of signals over the fronthaul. The CPRI was not enough to fulfill the latency and bandwidth requirement of the 5G system, thus, e-CPRI was introduced. The eCPRI consists of an enhanced Radio Equipment (eRE) and enhanced Radio Equipment Controller (eREC) segregated physically, but connected via a transport network [32]. The Service Application Protocol (SAP) runs over the transport network between the eRE and eREC. The SAP transmits a variety of data packets including control, and user data packets.

- Midhaul Transport - The Midhaul will occur when the DU and CU are in different sites. The mid-haul is the transport network infrastructure between the DU and CU. The latency on this link should be around 1 ms for some 5G use cases.

- Backhaul Transport - Optical fiber is always preferred for backhaul transport because of its low latency. However, connecting various cells to the core using fiber is not possible due to high deployment costs and unavailability. Since the massive deployment of small cells will be the key technique for 5G networks and the backhaul requirements of the small cells can significantly vary with the small cell location, the fiber cannot be the optimal approach for a 5G backhaul solution. On the other hand wireless backhauling (e.g. microwave and millimeter wave) has become popular due to its

availability, deployment time, and cost-effective approach but the weather condition and multipath propagation lowers its efficiency.

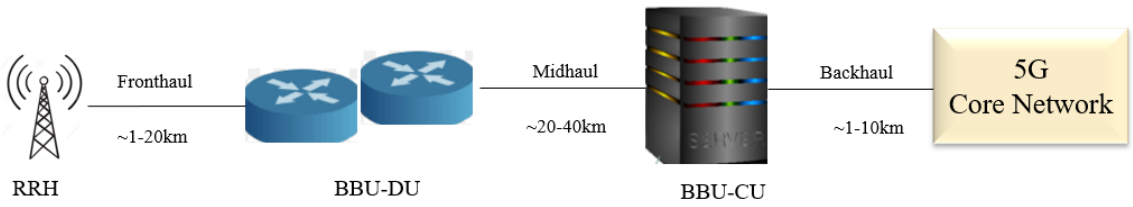


Figure 3.22: Transport Network schematic [41]

Table 3.2: Distance ranges of fronthaul, backhaul and midhaul [41]

Fronthaul	1∼20km
Midhaul	20∼40km
Backhaul	1∼10km Aggregation: 5-80km Core: 20∼300km

Table 3.3: latency requirements [41]

Service type		Latency requirement
eMBB	User plane (UE-CU/MEC)	4 ms
	Control plane (UE-CN)	10 ms
URLLC	User plan (UE-CU/MEC)	0.5 ms ∼1 ms
	Control plane (UE-CN)	10 ms

The RAN requires very low latency for some of the 5G use cases as shown in Table 3.3. This means an increased load on the transport network.

3.7 5G System Procedures

This section elaborates on how the UE registers with the network for the first time or periodically from time to time. It also elaborates on PDU session establishment for different scenarios. A PLMN is made up of a core network and an access network provided by the same network operator. When the subscriber moves to another geographic area, that is not covered by its Home-PLMN, it has to register itself with the Visited-PLMN to get the mobile services. There are two arrangements for providing the services to the subscriber when he roams outside the Home PLMN: Home Routing (HR) and Local Breakout (LBO). In Home-Routing, the subscriber attaches itself to the visited network, and all the traffic is routed through the home network. In the LBO, the UE attaches itself to the visited network and

uses the VPLMN's resources after fetching the subscriber data from the UDM. In this case, the user traffic does not go through the HPLMN.

3.7.1 UE Scanning Procedure

Scanning is the first process that a UE undergoes after the power is on. The UE scans all the possible frequencies for synchronization. The detailed steps are described below [10]:

1. When the UE is turned ON, it scans all the neighboring frequencies and tunes to a specific frequency.
2. It detects Primary Synchronisation Signal (PSS), and Secondary Synchronisation Signal (SSS) for frequency and time synchronization. It then learns about cell ID and decodes other channels such as PBCH, PDCCH, and PDSCH for RMSI.
3. The SSS, PSS and NR-PBCH are sent in a synchronization signal block (SS Block).
4. SSBs are sent in a batch which is used during beam sweeping by altering beam direction for each SSB transmission.
5. UE chooses the best beam and decodes the PBCH MIB contents data.
6. UE reads the data from SIB and decodes the PLMN ID, RACH parameters and cell selection parameters.
7. If the PLMN ID sent by transmitter is present within the UE´s PLMN ID list then UE runs the cell selection procedure, and the UE registers with the new cell and the entire registration process takes place again.

3.7.2 UE Registration procedure

Registration is the next step once scanning is done and synchronization is achieved. There are different types of registration. The procedure for initial registration is described below [10]:

1. The initial registration is the first procedure that the UE performs once it's switched on. The UE registers itself with the 5G network using any one of the following registration types:
    - Initial Registration - This is done when the UE registers itself with the network for the first time.
    - Periodic Registration - It is performed periodically after every predefined period of inactivity.
    - Mobility Registration - This is done when UE shifts to a new tracking area which is outside UE's registration area.
    - Emergency Registration - It is used by the UE when it wants to use an emergency service when the UE is already registered to the 5G network.

2. UE sends registration requests to the RAN or gNodeB. The registration request contains parameters such as registration type, SUPI, 5G-GUTI, tracking area update, requested NSSAI, and PEI. The PEI is taken from UE during the initial registration.

3. The RAN selects the AMF based on the requested NSSAI. If it is unavailable, it selects the default AMF based on 5G-GUTI.

4. If the UE was previously connected to an old AMF, then the new AMF will communicate with the old AMF to get the UE context over the N3 interface.

5. The new AMF registers with the UDM over the N8 interface. The UDM transfers all the subscriber data to the new AMF and AMF subscribes to all the changes.

6. UDM deregisters the old AMF. UE deregisters with the old AMF. The new AMF fetches policy data from the PCF over the N15 reference point.

7. AMF also selects the SMF and uses SM context for PDU sessions.

8. AMF sends registration accept to the UE over N3 and the UE responds by sending registration complete message.

3.8 PDU Session

A Protocol Data Unit (PDU) session is a logical connection between the UE and the Data Network, similar to a Packet Data Network (PDN) session in the EPS, which provides internet connectivity to the UE as shown in Figure 3.23. The UPF acts as the PDU session anchor. The data network is identified by the Data Network Name (DNN). Each PDU is initially called a Service Data Unit (SDU). To transport the SDU, the current radio network layer adds encapsulation to the SDU by adding a protocol header (PCI). The combined PCI and SDU are known as PDUs. A PDU session transports IP packets through a logical connection that runs on top of UDP. Each PDU session contains bearers that carry the user payload between the UE and the Data Network. These bearers are encapsulated in a tunnel with the help of the GTP protocol. Each tunnel endpoint is assigned a value called Tunnel Endpoint Identifier (TEID) for identifying the tunnel, which is carrying a particular data traffic. There are two kinds of bearers: Signalling Radio Bearer (SRB) and Data Radio Bearer (DRB). The SRBs carry all the signaling messages (NAS and RRC) while DRBs form a tunnel that carries the payload.

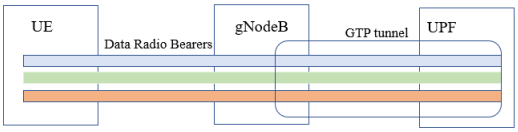


Figure 3.23: PDU Session establishment [10]

3.8.1 UE requested PDU session establishment

This is the scenario where the UE is in the Home-PLMN. The home PLMN of a mobile subscriber is called HPLMN, which has subscription profiles of subscribers in the HLR. For Local-Breakout (LBO) and non-roaming, the PDU session establishment procedure is explained below [10]:

1. The UE sends PDU session establishment requests to the AN. This request consists of the following:
   - PDU Session ID
   - UE requested DNN
   - S-NSSAI
   - Request type
   - N1 SM container
   - SM PDU DN request container
2. The AMF determines the message based on the request type. If the NAS message does not contain S-NSSAI, AMF determines the S-NSSAI of the serving PLMN from the current allowed NSSAI for the UE.
3. If the AMF does not have an association with the SMF, AMF transmits NSMF-PDUSession-CreateSMContextRequest, else Nsmf-PDUSession-UpdateSMContextRequest is transmitted to the SMF.
4. According to the data given by the UE, SMF connects with UDM over N10 and PCF over N7 to get data for PDU session creation.
5. If the request type is an initial request, SMF begins an N4 Session establishment request with selected UPF, otherwise, it transmits an N4 change request.
6. UPF accepts the request and sends N4 session establishment/change response.
7. SMF gets the GTP tunnel info from the UPF and the tunnel is established.
8. After the successful establishment of the tunnel endpoint, SMF transmits Namf-Communication-N1 N2 MessageTransfer with tunnel data for the N2 message and PDU session data in the N1 Container.
9. Upon receiving the aforementioned message, the AMF initiates the process by sending an NGAP PDU session Setup Request. This request includes the N2 parameter obtained from the SMF in the initial message. The parameters accompanying the PDU Session Setup Request comprise PDU Session ID, QoS Profile, CN Tunnel Information, PDU Session type, and Session AMBR.
10. The NG RAN creates the AN tunnel based on N2 data. The gNode-B forwards the N1 message to the UE for setting of PDU session.
11. The AMF updates SMF about tunnel setup by transmitting Nsmf-PDUSession-UpdateSMContext-Request and receives feedback from the SMF.

3.8.2 UE requested PDU session establishment for Home-Routing

The PDU session establishment procedure for the Home-Routing case is explained below [10]:

1. The UE sends PDU session establishment requests to the AN of the visited PLMN. This request consists of the following: PDU Session ID UE requested DNN S-NSSAI Request type N1 SM container SM PDU DN requests a container

2. The AMF determines the message based on the request type. If S-NSSAI is not included in the request, then serving PLMN s-NSSAI and home-PLMN s-NSSAI are chosen by the AMF. Using these values, SMF is selected in both home-PLMN and serving-PLMN.

3. The AMF provides the identity of HSMF and alternate H-SMF. The V-SMF does not use DNN selection mode but it transfers this information to H-SMF. If the AMF does not have any link with the SMF, AMF sends NSMF-PDUSession-Create SM Context Request otherwise, Nsmf-PDUSession-UpdateSMContextRequest is sent to SMF.

4. The V-SMF selects a UPF in VPLMN and initiates the N4 session establishment procedure by sending an N4 session establishment request if the request type is an initial request, otherwise it sends a Session Modification request.

5. The V-UPF responds with N4 Session Establishment Response/ N4 session modification Response.

6. The V-SMF forwards the session management-related information to the H-SMF. The H-SMF might use the DNN selection mode when contemplating to either accept or reject the UE request.

7. The H-SMF registers for the PDU session with the UDM using NudmUECM registration.

8. The H-SMF transfers QoS requirements associated with the PDU session to the V-SMF. This takes place during or after the PDU session establishment. The V-SMF may check these QoS requirements concerning the roaming agreements.

9. The V-SMF starts an N4 Session Modification process with the V-UPF. The V-SMF shall provide N4 rules to the V-UPF for the PDU Session, the rules relating to forward UL traffic to the H-UPF.

10. The V-SMF gets the GTP tunnel info from the V-UPF and the Tunnel is created.

11. After the successful creation of the tunnel endpoint, V-SMF sends Namf-Communication-N1N2MessageTransfer with tunnel details for the N2 message and PDU session details in the N1 Container to the AMF.

12. Upon Reception of the above message AMF Sends anThe NGAP PDU session Setup Request along with an N2 parameter from SMF in the above message with parameters, PDU Session ID, QoS Profile, CN tunnel Info, PDU Session type, Session AMBR.

13. The NG RAN sets up the AN tunnel based on N2 information.

14. The gNodeB forwards the N1 message to the UE for setting of PDU session.

15. The AMF updates SMF about tunnel setup by sending Nsmf-PDUSession-Update SM Context-Request and receives a response from the SMF.

3.9 Quality of Service Flows

Quality of Service (QoS) means overall performance of a service experienced by the users of the network. To quantitatively measure the QoS packet loss, throughput, bit rate, availability, transmission delay and jitter, etc. related aspects of service can be measured.

5G Quality of Service (QoS) model is based on QoS Flows. Each QoS flow has a unique identifier called QoS Flow Identifier (QFI). There are two classes of flows: guaranteed bit rate (GBR) QoS flows and non-gBR QoS flows. "The QoS Flow is the finest granularity of QoS differentiation in the PDU Session" [9]. User Plane (UP) traffic with the same QFI receives the same forwarding treatment.

At the Non-Access Stratum (NAS), packet filters in 5GC and UE map UL and DL packets respectively to QoS flows. At the Access Stratum (AS), rules in UE and Access Network (AN), map QoS flows to DRBs. The 5G-RAN forms at least one Data Radio Bearer (DRB) together with the PDU Session and additional DRB(s) for QoS flow(s) of that PDU session can be subsequently configured for each UE [5].

Chapter 4

Analysis of 5G Reference Points

This chapter identifies the vulnerable roaming reference points in 5G, based on Chapter 2, and provides a detailed analysis of how these vulnerabilities, if overlooked, in the future, can be exploited in the 5G network.

4.1 Analysis of 5G Roaming Reference points

The reference points that cross the network boundary are shown in Figure 4.1 and Figure 4.2. Figure 4.1 explains the relation between Network Functions in Home PLMN and Network Functions present in the Visited PLMN. These reference points are carried over the N-32 which connects the SEPPs. The N-32 is discussed in Section 4.4. These reference points are different for Local Breakout roaming and Home-Routed (HR) roaming. The purpose of Home Routing is to give the network operator in the home PLMN more control over charging and policy-related activities. As a result, the calling services become expensive. The Local Breakout (LBO) is relatively less expensive but relies on foreign network operators' services and infrastructure. Most of the network operators use Home-Routing widely for 2G and 3G, instead of Local Breakout. For LTE, HR is recommended in the early phase of deployment [2].

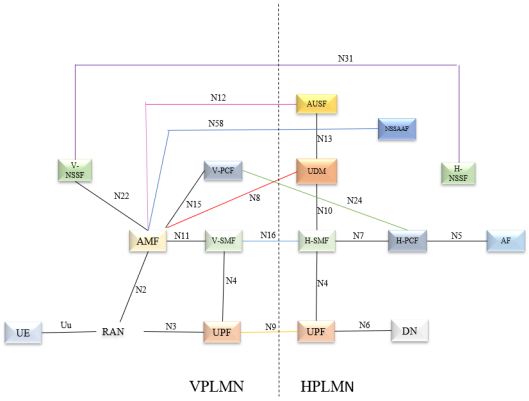


Figure 4.1: Reference points extending from HPLMN to VPLMN for Home-Routing [9]

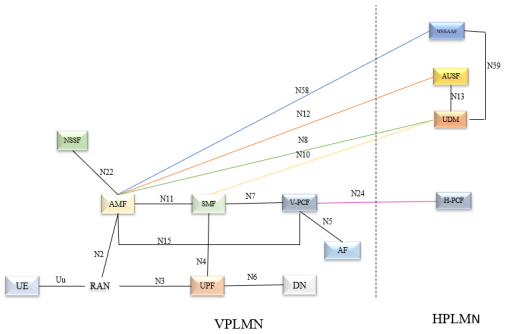


Figure 4.2: Reference points extending from HPLMN to VPLMN for Local Breakout [9]

The detailed description of roaming reference points, along with their protection mechanisms can be found below [9],[12]:

- N8 reference point
  This is the reference point between the AMF of V-PLMN and UDM of H-PLMN, part of LBO and HR, similar to S6a present between HSS and MME in LTE. It is a part of the SBI Core of 5G, which uses HTTP/2 protocol with JSON as the application layer serialization protocol.

  Protection mechanisms

  - The SBI interfaces are protected via TLS and OAuth 2 [40], [42] security protocols.
  - It is protected by SEPP which hides network topology.

  From above, it can be summarized, that it is a sensitive interface but well-protected by 5G security by design architecture.

- N9 reference point
  The N9 reference point can be present within the same PLMN or between different PLMNs. It uses the GTP-U protocol for transferring user data.

  Protection mechanisms

  - It is protected by IPsec [43] and DTLS [53].
  - The N9 uses GTP-U as an interconnection protocol. The N9 is also protected by IPUPS which sits at the perimeter of user plane for filtering illegitimate traffic.

  From Section 2.2, it can be summarized, that the GTP protocol has known security flaws, which makes N9 vulnerable to roaming attacks. The F1-U has an optional IPsec protection which is directly connected to N3.

- N10 reference point
  The N10 is present between UDM of H-PLMN and SMF of V-PLMN in LBO. It uses HTTP/2 protocol with JSON, as it is a part of the core network. The N10 offers services to Unified Data Management (UDM) and Session Management Function (SMF) via the Nudm interface.

  Protection mechanisms

- The SBI interfaces are protected via TLS and OAuth 2 security protocols.
- It is protected by SEPP which hides network topology.

From above, it can be summarized, that N10 is a sensitive interface but well-protected by 5G security by design architecture.

- N12 reference point
  The AUSF within the 5G Core offers services to the AMF via the Nausf service-based interface or N12 Reference Point.

  Protection Mechanism

  - The SBI interfaces are protected via TLS and OAuth 2 security protocols.
  - It is protected by SEPP which hides network topology.
  - It is a sensitive interface but well-protected by 5G security by design architecture.

  From above, it can be summarized, that N12 is a sensitive interface but well-protected by 5G security by design architecture.

- N24 reference point
  It is a service-based interface or reference point between the H-PCF and the V-PCF in LBO. Both the V-PCF and H-PCF need to support charging.

  Protection Mechanism

  - The SBI interfaces are protected via TLS and OAuth 2 security protocols.
  - It is protected by SEPP which hides network topology.

  From above, it can be summarized, that N24 is a sensitive interface but well-protected by 5G security by design architecture.

- N31 reference point
  The visited NSSF interacts with the home NSSF over the N31 Reference Point in Home-Routing. This capability allows Network Slice selection in both the H-PLMN and the V-PLMN.

  Protection Mechanism

  - The SBI interfaces are protected via TLS and OAuth 2 security protocols.
  - It is protected by SEPP which hides network topology.

  From above, it can be summarized, that N31 is a sensitive interface but well-protected by 5G security by design architecture.

- N32 reference point
  The N-32 connects SEPPs in H-PLMN and V-PLMN. The is similar to Diameter Edge Agent (DEA) in LTE, as these hides the network topology and supports interconnection signaling between H-PLMN and V-PLMN.

  Protection Mechanism

- It is protected by TLS encryption and PRINS protocol. The PRINS is considered vulnerable due to its complexity.

From above, it can be summarized, that N32 is vulnerable to roaming attacks.

- N58 reference point
  The Network Slice Specific Authentication and Authorization Function (NSSAAF) acts as the service producer while AMF acts as the service consumer. The NSSAAF interacts with AMF over the N58 service-based interface shown in Figure 4.1. It performs slice-specific authentication and re-authentication for a UE.

  Protection Mechanism

  - The SBI interfaces are protected via TLS and OAuth 2 security protocols.
  - It is protected by SEPP which hides network topology.

  From above it can be summarized that it is a sensitive interface but well-protected by 5G security by design. architecture.

4.2 Vulnerable 5G Roaming Reference Points

From the analysis of reference points provided in Section 4.1, N3, N9, and N32 are identified as vulnerable reference Points as these are prone to roaming attacks. Since GTP runs over N3 as well, it has been selected for further analysis. The numbered interfaces in Figure 4.3 will be considered for further analysis as these were found more vulnerable to attacks than the others because of the reasons listed in Section 4.1.

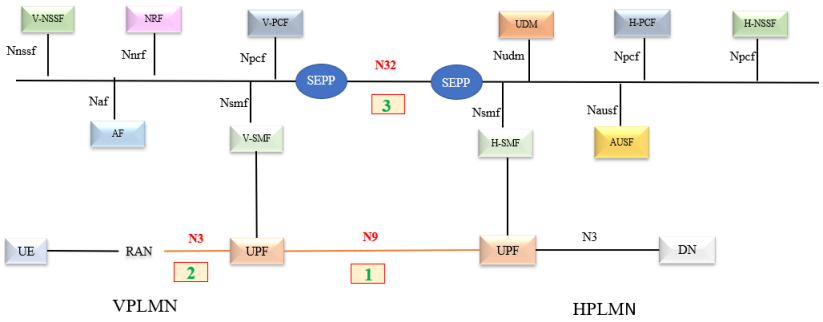


Figure 4.3: Selected interfaces for Vulnerability Analysis

4.3 N9 Reference Point

The N9 reference point connects two different UPFs. These UPFs might be located in the same PLMN or different PLMNs. In the case of roaming, the subscriber roams to a different PLMN, called the visited PLMN (V-PLMN) and to get access to the voice and data services, a PDU session is established from the visited network to the home country, which is called

Home-Routing (HR). In Home-Routing, the user plane is terminated in the home network, the DN is located in the home network and HPLMN is in charge of the PDU session anchor and its connectivity with the DN. The PDU session forms a continuous tunnel that follows the subscriber. The HPLMN controls policy and charging. The VPLMN need not know the specific data services offered to the subscriber, as it has limited responsibility for ensuring proper QoS, handover support, and handling the transition from the user plane active to the user plane inactive state. The HR mode is ideal for cases where the HPLMN is providing special data services to the user. In the case of Local Breakout (LBO) roaming, the VPLMN is in charge of PDU session anchor and policy control. When a subscriber roams to a foreign network, then normally N9 Home-Routing will be used for voice as well as data usage. The GTP tunneling protocol is used over N9 for tunneling user data packets. Figure 4.4 shows the N9 interface for the home routing case.

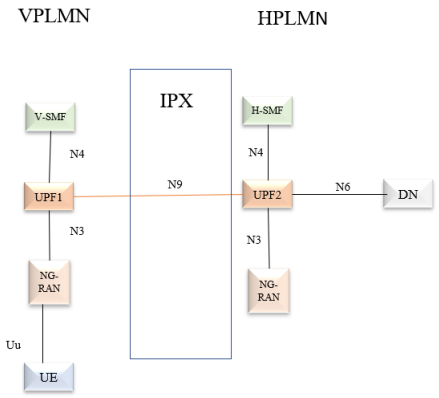


Figure 4.4: N9 interface [58]

If a subscriber is not within the reach of the UPF service area, then an intermediate SMF (I-SMF) is added between AMF and SMF. The I-SMF selects and controls the I-UPF connected over N3 to RAN. The SMF insertion takes place when a UE, which has already created a PDU session, moves out of the SMF service area. To maintain the PDU session, I-SMF is inserted. In 3GPP Release-16, I-SMF insertion, I-SMF removal, and I-SMF change are supported. The architecture with I-SMF is similar to the home-routed roaming architecture but with V-SMF replaced by I-SMF and H-SMF replaced by SMF [61]. To maintain the PDU, the proposed GTP-based randomization protocol can be used for both Inter and Intra-PLMN N9 configurations depicted in Figure 4.5.

4.3.1 GTP Protocol used over N9

The GPRS tunneling protocol for the User Plane (GTP-U) is used over the N3, N4, and N9 interfaces. The GTP-U runs over UDP and the data belonging to a specific PDU session is identified by a Tunnel Endpoint Identifier (TEID) along with the IP address and UDP port number. The signaling required for tunnel establishment is established by HTTP/2 in 5G-SA between SMF and AMF, instead of GTP-C in 4G and 5G NSA. The protocol stack in Figure 4.6 for the user plane describes these layers. The GTP path is generally defined by an IP address and UDP port number.

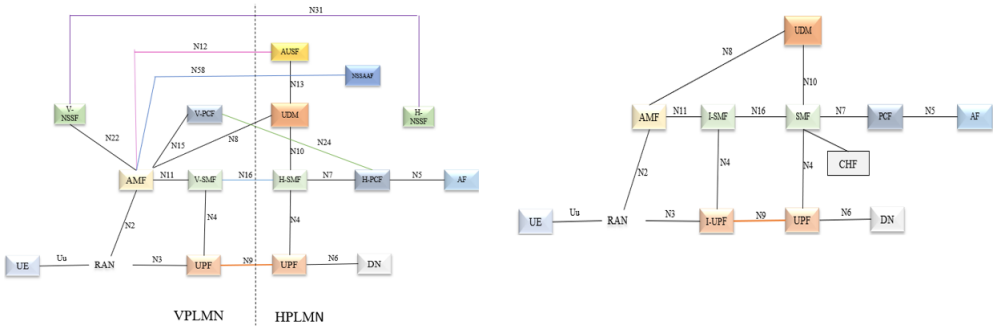


Figure 4.5: UPF topology for Inter-PLMN and Intra-PLMN data transfer [58]

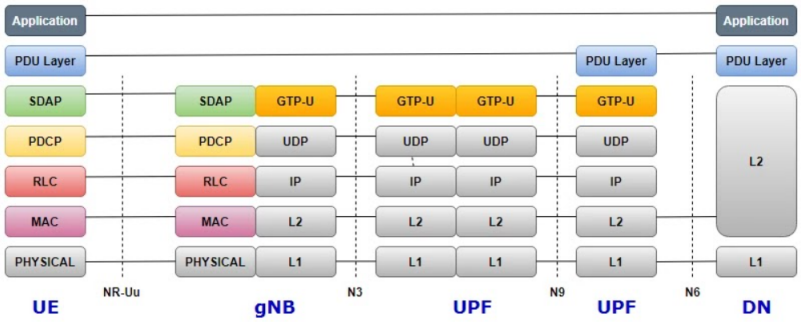


Figure 4.6: User Plane Protocol stack[61]

The GTP protocol was a part of the legacy network and is a part of the 5G user plane. The GTP protocol has certain flaws, which are described in Section 2.2. The GSMA-recommended mitigation measures are described in Section 2.3. Despite the availability of security guidelines and mechanisms, the GTP protocol remains vulnerable to security flaws [31].

From my experience as a Security Engineer, having performed multiple GTP Penetration tests, I think that the GTP is vulnerable and susceptible to interception attacks due to the following reasons:

- Lack of additional protection mechanisms recommended by GSMA such as GTP firewall. Many operators do not secure their networks with GTP firewalls at present.

- The GTP firewalls should be configured correctly to filter out all types of illegitimate GTP packets. However, this is not the case, in reality, some firewalls still allow malicious packets due to the configuration of incorrect filtering rules.

- The current mobile networks use GTP-Cv0 which has significant security weaknesses such as non-random TEID.The latest GTPP-C standard is GTP-Cv2.

- The GTP-Cv2 used in 4G and 5G-NSA also has some vulnerabilities that can be exploited to eavesdrop or launch an interception attack.

4.3.2 GTP Roaming risks in 5G

The GTP protocol is vulnerable to interception because of the different causes and known exploits described in Subsection 4.3.1. In 5G-SA, the user plane protocol stack in Figure 4.6 shows the GTP-U running from BBU-D, part of gnodeB, to the UPF. If N3 and N9 are intercepted then the outcome might reveal relevant user information that could be used to launch attacks. A piece of exposed information may be able to determine which application is being used by the subscriber by analyzing the IP packet flow between BBU-D and the UPF. This kind of analysis can be done using the methods discussed in Section 2.2, by gaining access to the network and then intercepting the user traffic. From [26], it is known that every application has a unique fingerprint which could be based on different parameters such as Packet Length, Inter-arrival packet times, Communication patterns, and Volumes of Exchanged traffic. This can serve as a basis for disclosing sensitive user information such as their entertainment preferences, and religious preferences, even if the packets are encrypted. This can help in uniquely identifying the subscriber through his application preferences. If in a GTP flow, interception happens, then TLS encryption will be insufficient to protect the subscriber´s identity. These kinds of interception attacks can take place in the subscriber´s Home network or any foreign network.

In the case of the Home network, the subscriber roams within the same country but across different areas with a limited distance range using the network provider´s trusted infrastructure and service providers. The security measures for keeping the network safe from threats can be applied uniformly across a country. During roaming, when a subscriber crosses the network boundary and roams to a distant location, ensuring secure communication between two different PLMNs becomes complicated. The UE gets attached or registered to a foreign network provider´s gNodeB. The network provider could be trusted or untrusted and the exchange of data flow across the GTP tunnel will take place between the Home network and the foreign network. Hence, this could pose a security risk to GTP running over N3 and N9, which have already known vulnerabilities. Thus, interception attacks are more likely to occur over roaming Reference Points, N3 and N9. For determining interconnection security, the GTP-U protocol in 5G should be made more secure and robust against existing vulnerabilities.

4.3.3 Methods to exploit GTP in 5G

The different ways in which an attacker can gain access to N3 and N9 or any other roaming reference points are as follows:

- The malicious attacker can get access to the network if unknown domains used for data and C2 (Command and Control) are not blocked. The hackers use malware domain generation algorithm which bypasses the security controls.
- If the MNOs are not detecting and blocking malicious traffic inside the GTP-U tunnel, which originates through bots or clicks fraud, from connected devices like the UE or any IOT device, then the attacker can gain access to the network.
- If the MNOs are not carrying out statistical measurement of inbound, and outbound roaming traffic linked to specific IMSI/MSISDN/IMEI, the anomalies in the subscriber´s

profile will not be detected. For IoT devices, timing, data size, and frequency are useful features for determining traffic characteristics, an attacker can use these to know subscriber preferences and personal information.

- If the MNO is not using IPsec tunnel encryption for transferring packets, then any malicious attacker can alter the traffic. "IR.77 requires IPsec encryption only for the data transferred over the public internet, but even in the case of a dedicated/leased line, there is a risk that the link security might be compromised leading to data theft, fraud, DDoS attack, etc." [38].

- The MNOs might have GTP firewalls in place to protect their roaming traffic. However, firewall filtering is based on preconfigured rules, which may not be sufficient to filter out all types of illegitimate traffic or any misconfiguration can let the illegitimate packet through the network. Mostly Type Length Value / Information Element (TLV/IE) structure and presence of mandatory IEs are checked.

- If the GRX or IPX node gets compromised, the attacker can use various GTP messages to intercept the user traffic across N3 and N9 roaming interfaces. These GRX or IPX nodes are a part of the IPX infrastructure, which provides roaming capabilities to different MNOs. If the message is coming from an untrusted or spoofed source, the firewall should immediately block those packets. If the firewall is not adequately configured, then these packets can pass through and reach the core network nodes. Figure 4.7 shows traces from a real network´s GTP test, the message Create SGSN context was successful despite coming from an untrusted source. These messages include Create Session Request, Context Request, Relocation Cancel Request, Delete Session Request, and Modify Bearer Request.

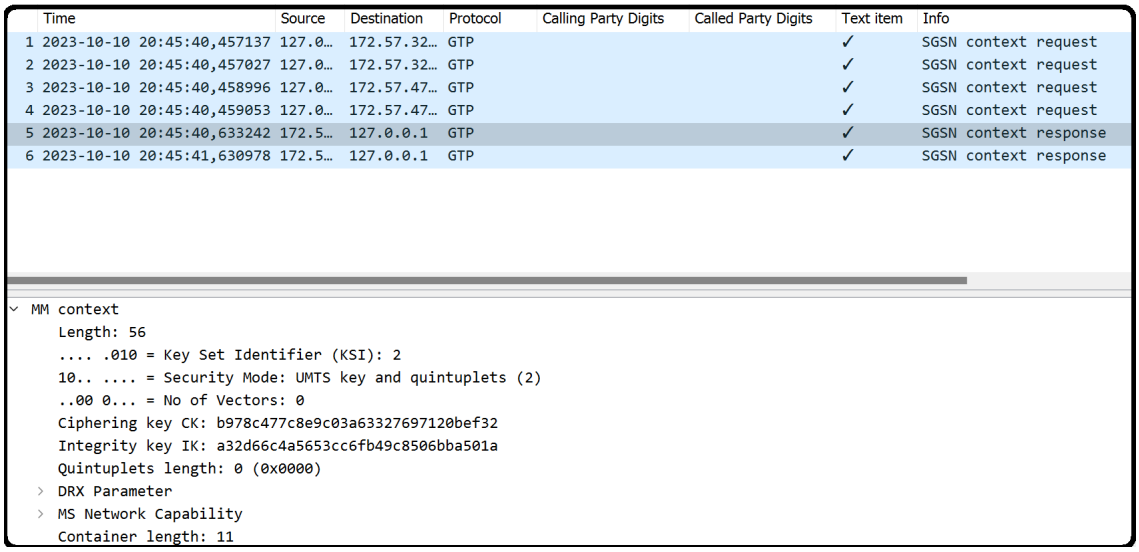


Figure 4.7: Trace of GTPv1 header showing key information

4.3.4 Reason for additional protection of N9 reference point

The reasons for additional protection of N9 reference point are listed below:

- The N9 Inter-PLMN user traffic carries home-routed user traffic from the visited network to the home network. This traffic consists of sensitive information that needs to be protected from external visibility and manipulation [6].
- Inadequate protection of Inter-PLMN user traffic on N9 might lead to attacks that cause unauthorized modification of information and leakage of sensitive information [6].
- The GTP protocol has vulnerabilities and some of these vulnerabilities were exploited in the past to launch roaming attacks on UMTS and LTE networks [57], [65]. These attacks include overbilling attacks, session hijacking attacks, and DoS attacks. The major flaw is that GTP does not provide encryption and does not verify the location of the user [54]. In 5G SA, GTP runs over N3 and N9 interfaces for data transfer. Hence, its secure implementation is necessary for a safe intra and inter-PLMN data transfer.

4.4 N32 interface

The 5G core architecture has gone through a complete change when compared to the previous generations, because of which 3GPP has migrated to a new interconnection or roaming security protocol, PRINS. Since, this protocol, namely PRINS, will be implemented for the first time, there will be new implementation challenges.

The N-32 interface forms the link between two Security Edge Proxy Protection (SEPP) which can be seen in Figure 4.8 belonging to two different networks. The SEPP provides security to the Service-based core network against roaming attacks. The SEPPs are similar to the Diameter Edge Agent of LTE networks, however, the 5G core network uses HTTP/2 signaling for interaction with other NFs and JSON for encoding information. These SEPPs are located at the edge of the network and protect by hiding the network topology and separating the core network of one PLMN from the core network of another PLMN. The N-32 interface is made up of N-32-c (control plane) and N-32-f (forwarding plane) which is shown in Figure 4.9 [12]. After the completion of the HTTP/2 handshake, the N32-c connection is torn down, as it was established to exchange control signaling messages for forming the N-32-f, in the first place. The connection established is end-to-end which may or may not involve IPX in between. The N32-f is the forwarding interface that is involved in exchanging data between the NF consumer and NF producer. All the core network reference points that cross the PLMN boundary will be carried over N-32. The SEPP protects all outgoing messages before sending them to a second PLMN over N32 and receives and verifies all incoming messages on the N32 interface before sending them to the correct NF.

The IPX may modify the content which the receiver SEPP may apply after verifying the integrity of the modifications [17]. The SEPP will act as a non-transparent Proxy for the NFs when service-based interfaces are used across PLMNs, however inside IPX service providers, an HTTP proxy may also be used to modify information elements (IE's) inside the HTTP/2 request and response messages [12]. Acting similarly to the IPX Diameter Proxy used in

EPC roaming, the HTTP/2 Proxy can be used for inspection of messages, and modification of parameters.

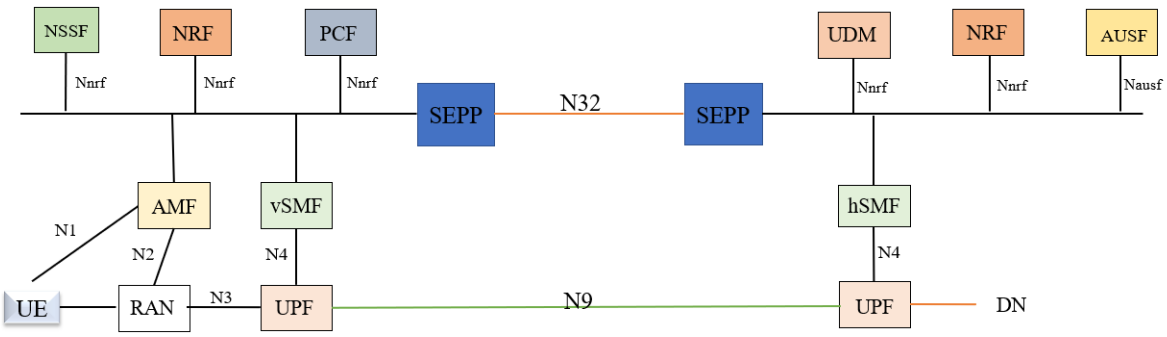


Figure 4.8: N32 interface [12]

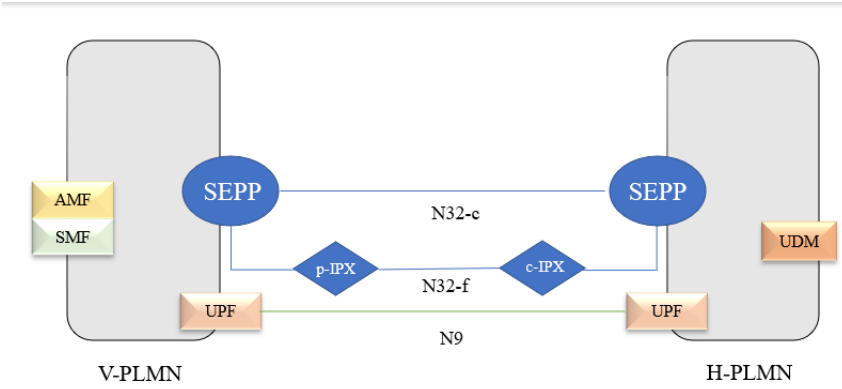


Figure 4.9: N32 interface with N32-c and N32-f [12]

4.4.1 Reason for choosing N32

The Protocol for Interconnect Security (PRINS) for securing the N-32 is highly complex as suggested in [45], [62]. Summarising N-32 vulnerabilities, it has been observed that the roaming relation is based on trust between the roaming partners and IPX and there are few JSON incompatibilities amongst different SEPP vendors. These call for additional security mechanisms over N-32 to mitigate future attacks that can exploit these vulnerabilities. It can be concluded from above that N-32 protection mechanisms are highly complex with certain ambiguities concerning JSON specifications.

The reasons for choosing N-32 for this thesis as a vulnerable roaming interface are described below in detail.

- The HTTP/2 messages will be exchanged over this interface which will carry a lot of signaling load related to roaming. This interface traverses through different IPX depending on the location of the VPLMN. The IPX is considered as the weakest link when it comes to roaming security.
- The application layer traffic consists of all the IEs present in the HTTP message payload. These IEs get different security treatment from SEPP. Few IEs require end-to-end (e2e) encryption, while others require only e2e integrity protection, and still, others may require e2e integrity protection but modifiable by intermediate IPX providers while in transit [12].
- There are inconsistencies with JSON which was mentioned in the previous section and the signaling protocol HTTP/2 also comes with its vulnerabilities such as dependency cycle attacks, resource-intensive attacks and stream prioritization abuse [24]. Thus, it becomes necessary to analyze other risks that these vulnerabilities could pose for Mobile Network Operators.

4.5 SEPP Overwiew

The Security Edge Protection Proxy (SEPP) is the last node/NF of a PLMN, which provides inter-PLMN security during roaming scenarios. The SEPP belonging to consumer NF is called c-SEPP and the SEPP belonging to producer NF is called p-SEPP. The concept of SEPP is analogous to the Diameter routing Agent (DRA) specifically Diameter edge router of LTE which sits outside the LTE network, providing routing support, load balancing, and roaming security. The logical connection between p-SEPP and c-SEPP is the N32 interface. The N32 handshake process takes place between the SEPPs in two different PLMNs in order to authenticate each other and negotiate regarding the security mechanism to be selected and security configuration parameters for N32-f [17]. The specific N32-c and N32-f procedures are described below: The N32-c interface includes the following handshake procedures [17]:

- Security capability negotiation procedure: An end-to-end TLS connection is required before security capability negotiation. The initiating SEPP will use HTTP Post along with " SecNegotiateREQData" IE towards the responding SEPP. The responding SEPP responds with the selected security capability.
- Parameter Exchange Procedure: If the selected security capability is PRINS then the parameter exchange procedure is performed. The first step is cipher suite negotiation for protection of N32-f signalling and then an exchange of protection policies takes place.
- N32-f Context termination: The initiating SEPP creates a request for the responding SEPP for N32 ContextID termination. The responding SEPP stops message transfer over N32-f and deletes the "N32-f ContextID".
- Error Reporting Procedure: When the SEPP encounters errors in the messages, it sends the error information to the sending SEPP which will log the messages that were not processed.

The N32-f interface includes the following procedures [17]:

- Message protection to the information exchanged between the NF service producer and the NF service consumer across PLMNs by applying application layer security procedures.
- Forwarding of the application layer protected message from a SEPP in one PLMN to a SEPP in another PLMN. This kind of forwarding can include IPX providers in between the SEPPs.
- If IPX providers are present between SEPP in PLMN A and SEPP in PLMN B, the forwarding on the N32-f interface may include the insertion of content modification policy which the receiving SEPP applies after verifying the integrity of such modification policy.

4.5.1 SEPP Security Mechanisms

The different protection mechanisms supported by SEPP are explained below:

Protection between NF and SEPP: There are two ways to protect NF and SEPP which include custom HTTP header [12] as shown in Figure 4.10 and TLS protection based on telescopic FQDN and wildcard certificate. The dotted yellow line shows the protection mechanism between SEPPs over N-32-c and the orange dotted line represents protection between NF and SEPP as well as N-32-f. TLS will be used for N32-f connections between the SEPPS if no IPX is present. In case, there is IPX for IP routing and modification of information, the Protocol for N-32 Interconnect Security (PRINS) is used for N32-f connections. For PRINS, TLS VPN or NDS/IP will be used to give additional protection.
Protection between SEPPs: As stated above, TLS is used for N32-f in the absence of IPX infrastructure 4.10. If there are IPX that offer IP routing, offers services that require modification and addition of information, then the Application Layer Security (ALS) protocol called PRINS is used over N32-f. The SEPPs implement protection policies regarding application layer security providing confidentiality and integrity protection for those elements that are to be protected [12]. The JSON Web Encryption (JWE) is used for the protection of N32-f messages by . JSON Web Encryption (JWE) is a standard for encrypting JSON (JavaScript Object Notation) data in a compact and URL-safe way.

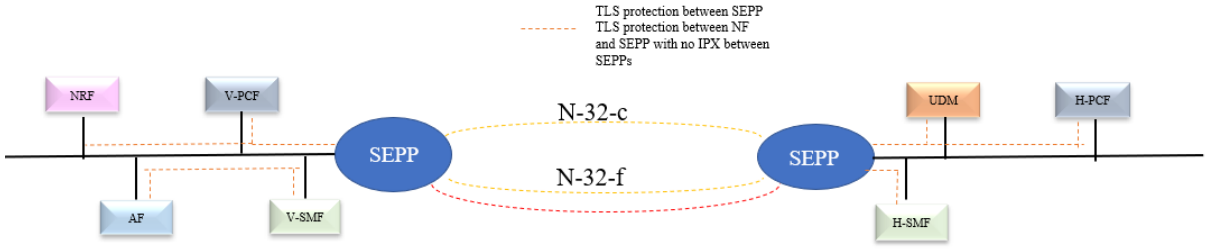


Figure 4.10: Protection using TLS between SEPPs [12]

4.6 SEPP Vulnerability Analysis

The list of vulnerabilities of SEPP that might be exploited by attackers to gain access to the network is listed below:

1. **Eavesdropping on sensitive SBI-interface messages:** The sensitive SBI interfaces are a part of core network architecture. If eavesdropping takes place on these interfaces then sensitive signalling-related data could be leaked, which can be a threat to other parts of the network such as the RAN [8]. The likelihood of this event is extremely low since the core network cannot be directly accessed; there is a separate access network that uses core network services with proper standardized procedures.

2. **Crashing NF by sending malformed messages:** These kinds of attacks were common in previous generations of mobile networks. There are different ways to generate and introduce these malformed packets into the network, by injecting GTP packets that have incorrect headers, TCP packets with incorrect source destination address, TCP packets with invalid flag bits, etc. These protocols will also be a part of the 5G network, which means the likelihood of such incidents remains high.

3. **Overwhelming NF by flooding:** This means that the number of signaling messages for each NF will exceed the pre-configured buffer size or the allocated bandwidth will be less for the amount of traffic data that will be generated. As 5G supports machine-to-machine communication, there will be large volumes of data, transferred continuously over the N-interfaces. The likelihood is medium if apt changes (relating to bandwidth and buffer size) are made before deployment.

4. **Obtaining unauthorized access to core network services:** The likelihood of this event remains low as the core network is highly secure with TLS and OAuth2 protection in place. Only network operators can have direct access to the core network functions. Obtaining authorized access will be extremely difficult.

5. **Modifying N-interface messages on the fly:** The likelihood of this event also remains low. Due to reasons similar to the previous vulnerability, modification of N-interface messages on the fly remains highly unlikely.

6. **Malicious changes to NF configuration:** The likelihood remains extremely low as the probability of accessing core network elements is unlikely for the reasons listed in vulnerability 4.

7. **Misusing cryptographic material of peer SEPPs and IPX providers:** This can happen if one of the IPX nodes has been compromised. The likelihood of such an event remains low as explained in 4.

8. **Incorrect handling for PLMN ID mismatch:** This can occur if there are configuration errors in the network product. The likelihood remains low and the SEPP should be able to send an error message if there is a PLMN ID mismatch.

9. **Incorrect handling for protection policies mismatch:** This can occur if there are configuration errors in the network product. The likelihood remains low and the SEPP should be able to send an error message if there is a protection policies mismatch.

10. **Exposure of confidential IEs in N32-f message:** This can occur if there are configuration errors in the network product. The likelihood remains low and the SEPP should be able to send an error message if there is a PLMN ID mismatch.

11. **Compromise of IPX node:** The likelihood of this event is Medium since the IPX is a roaming interconnection infrastructure that provides services to network operators operating from different PLMNs. These incidents have been reported in the past.

12. **JSON parser version inconsistencies:** According to [45], determining consistency and completeness becomes challenging in the absence of formally defined requirements for JSON. The processing of JSON in a multi-operator environment, typical for the N32 interface, may reveal inconsistencies and inherent incompatibilities in a JSON library. Additionally, the standard might encounter additional issues. Given that JSON is slated for use in a core network, the probability of such events remains high, and their impact could be significant.

13. **Inter-vendor incompatibility:** The likelihood is high and the impact can be catastrophic. Inside a network, there will be inter-working of products from different vendors. The N-32-related procedures are specified in [12], however, the procedures when SEPPs are from different vendors have not been described.

The vulnerabilities listed above have been taken from different sources [45], [62]. Based on the nature of these vulnerabilities, a vulnerability assessment matrix has been created. These vulnerabilities have been classified based on the impact they could have on the network and how frequently these might occur (likelihood). The aim was to focus only on the SEPP vulnerabilities and evaluate those with likelihood and impact factors. The risk assessment matrix focuses on the vulnerabilities as well as other external factors that might pose a threat. Hence, vulnerability assessment was performed instead of risk assessment. In Table 4.1, if there is a high likelihood of that event happening then it is colored red, yellow is assigned for medium likelihood, and green is for low likelihood.

Table 4.1: Vulnerability Assessment matrix for SEPP

Risk ID	Risk Description	Likelihood	Impact
1	Eavesdropping on sensitive N-interface messages	**Low**	**High**
2	Crashing NF by sending malformed messages	**Low**	**High**
3	Overwhelming NF by flooding	**High**	**High**
4	Obtaining unauthorized access to core network services	**Low**	**High**
5	Modifying N-interface messages on the fly	**Low**	**Medium**
6	Malicious changes to NF configuration	**Low**	**High**
7	Misusing cryptographic material of peer SEPPs and IPX providers	**Low**	**Low**
8	Incorrect handling for PLMN ID mismatch	**Low**	**Low**
9	Incorrect handling for protection policies mismatch.	**High**	**High**
10	Exposure of confidential IEs in N32-f message.	**Low**	**Medium**
11	Compromise of IPX node	**Medium**	**High**
12	JSON parser version inconsistencies	**Medium**	**High**
13	Inter-vendor incompatibility	**High**	**High**

4.7 Conclusion from Vulnerability Analysis Matrix

It can be concluded from above that vulnerabilities with IDs: 2, 3 and, and 10 are the ones that should be given priority by MNOs. Since the entire core network implementation will be new for operators and service providers, incorrect handling of protection policy mismatch becomes highly probable. In "3GPP TS 33.501", the steps that are to be taken in case of a mismatch are clearly defined but still, the interpretation of those procedures may vary. The vulnerability with ID 3 is also highly possible as there will be excessive signaling for different 5G-connected devices which will include not just mobile phones but also diverse IoT devices, connected vehicles, etc. The operators will have to be careful about the bandwidth allotted for control signaling because if overload happens then other core network functions will also fail. The ID 2 vulnerability can take place if an IPX node gets compromised.

The vulnerabilities with ID 1, 2, and 5 can be mitigated using the Randomisation method introduced in Chapter 5.

4.8 Survey Results

To strengthen the inferences obtained from the literature study as well as the analysis of vulnerable 5G Reference points, a brief survey was conducted. There were ten participants, belonging to German and Swedish telecom companies. The results obtained from this secondary data source are summarised below:

- **Vulnerability level of 5G SA Roaming reference points**.
  The participants were asked to rate the vulnerability level of 5G SA roaming reference

points from the following choices: High, Medium, and Low. The pie chart in Figure 4.11 shows that the majority of participants believe that 5G Interconnection reference points are prone to attacks. Therefore, an additional security mechanism would be ideal for these reference points.

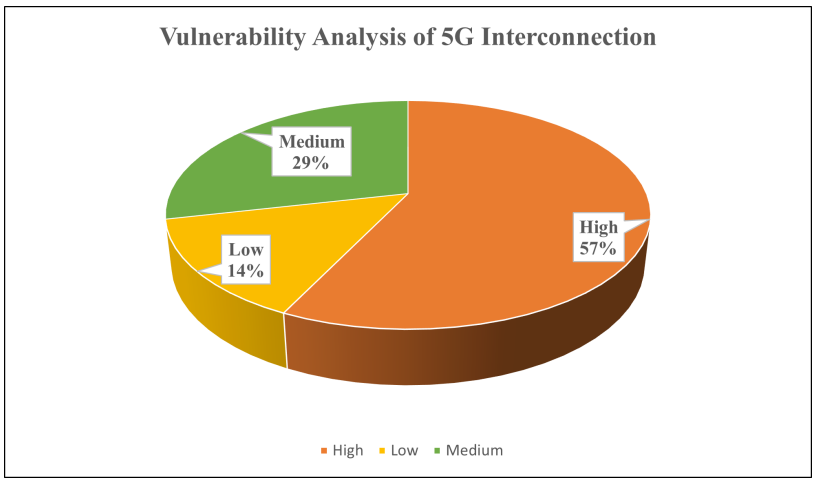


Figure 4.11: Results of 5G SA interconnection vulnerability analysis

- **Vulnerability level of GTP-U Protocol**.
  The participants rated the vulnerability level of the GTP-U protocol from the same choices as before. The pie chart in Figure 4.12 shows all the findings. It is evident that GTP-U had some shortcomings that will continue to exist in 5G SA in the user plane, Additional security measures have been recommended by GSMA, but not all operators will be able to implement those immediately. This makes GTP-U vulnerable making N3 and N9 also vulnerable to attacks.

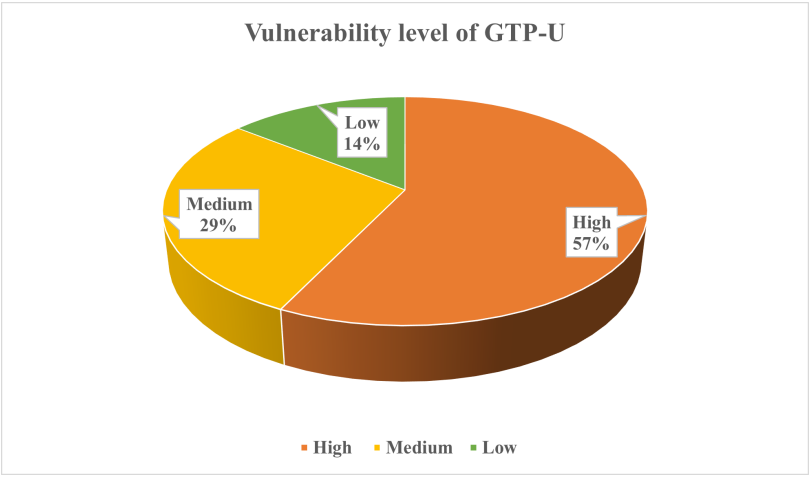


Figure 4.12: Results of GTP-U vulnerability analysis

- **Enhanced User Plane protocol for 5G SA**.
  The participants were asked if they would prefer to have a more secure User Plane Protocol for 5G SA at additional costs and the majority of them selected Yes.

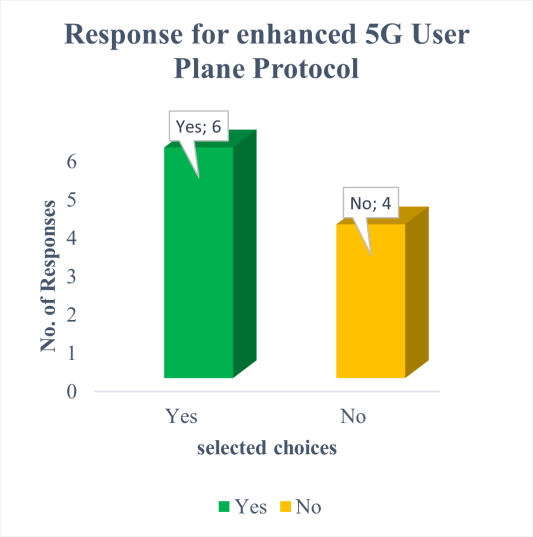


Figure 4.13: Response for enhanced User Plane Protocol

- **Complexity level of PRINS Protocol**. Figure 4.14 indicates that most of the participants also think that the new proposed protocol PRINS is highly complicated, this will make it difficult to implement and configure for the MNOs for the first time. Moreover, this added complexity, in conjunction with new core network architecture and protocols, can render N32 vulnerable to roaming attacks.

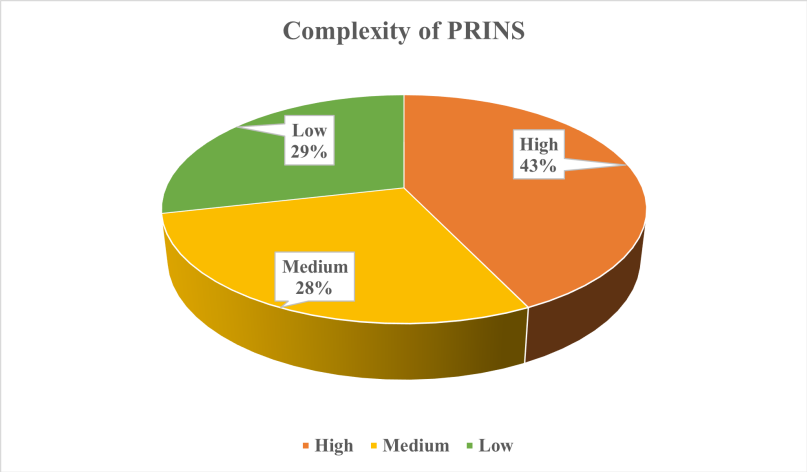


Figure 4.14: Results of PRINS Complexity level analysis

.

Chapter 5

GTP-based Randomisation

This chapter introduces GTP-based Randomisation, negotiation process, and implementation details. The new updated GTP protocol can make the N9 interface resistant to the existing GTP-based interception attacks.

5.1 GTP for 5G network

The GTP protocol, as described in Section 4.3.1 is a tunneling protocol that encapsulates the data packet and transports them from one end to the other end. The GTP-Control plane protocol and GTP-User plane protocol are two variants of the GTP. The GTP-C was a part of the core network in 2G, 3G, 4G and 5G-NSA. The latest standard is GTP-Cv2 which is a part of LTE and 5G-NSA, The GTP-U was a part of the legacy network and will continue to function in 5G-SA networks. The GTP-U will be carrying data from the UE to the external network (Data network) and the other way around. The GTP runs over the UDP layer to ensure fast and connection-less delivery of the IP datagrams.

The GTP protocol version 0, as defined in GSM TS 09.60, was used in the system based on the GSM Base Station Subsystem (BSS) [21] and version 1 is described in 3GPP TS 29.060 and applies to GSM and UMTS systems. The 3GPP created a new version of the GTP protocol, which was incompatible with version 0. The GTP-v1 came with new features which were not supported by GTPv0. The group members wanted to decompose GTP into separate user plane protocols GTP-U and control plane protocols GTP-C. The reason behind this separation was to support GTP-U-based tunneling of user data over the Iu interface (between UMTS core and RAN) without GTP-C for Iu tunnel setup. The new GTP version supported multiple QoS levels per IP address as well as multiple bearers. The GTPv0 supported both TCP and UDP for transport. However, GTPv1 was transported over UDP only. The GTP-Uv1 will be used in 5G over the N9 interface. The user plane protocol remained the same for setting up user plane tunnels, but GTP-C protocol version 2 came with enhancements for bearer handling in the EPS network.

The user plane security is part of the 5G security network architecture. The use of a GTP firewall prevents security risks. The N9 connects two different UPFs and runs through various IPX during HR roaming, depicted in Figure 5.1. The security of the entire user plane is ensured by Inter-PLMN User Plane Security (IPUPS) [39] which acts as a firewall

for filtering illegitimate packets. IPUPS can be a part of UPF or can exist as an independent function outside of the UPF. It provides the filtering functionality at the UPF by the Packet Detection Rules (PDR); these rules are configured at the UPFs by the SMFs in both the Uplink and Downlink directions. Thus, any packet that does not align with a valid TEID gets dropped immediately [58].

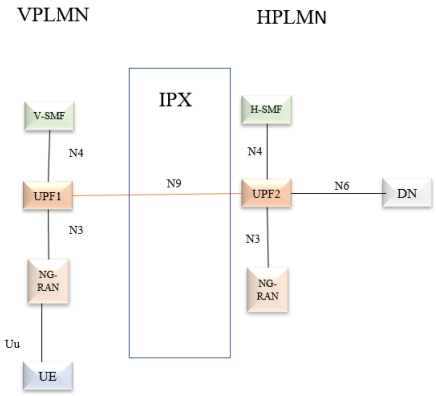


Figure 5.1: N9 interface [58]

5.2 Randomisation for 5G Network

Randomization means creating random data lengths to eliminate the effect of bias or chance. In the context of this thesis, it involves adding random bits or bytes to the user data according to pre-defined rules. From [26], it can be concluded that every application has a signature distribution. This distribution is plotted with inter-arrival time and packet lengths. With such a distribution each application can be classified accurately, which can be dangerous if the traffic is intercepted by a potential attacker. Thus, randomization can be used to provide additional security to the network. In this thesis, the Randomisation concept is considered for 5G-Standalone networks as an additional security mechanism for roaming reference points. The two variants of randomization at different network levels are as follows:

- GTP-based Randomisation - described in Chapter 5
- RandomTCP-TCP based Randomisation - described in Chapter 6

It has been described in previous sections that N3 and N9 are vulnerable reference points due to the use of the GTP-U protocol which is considered vulnerable to various attacks. The N9 reference point spans over IPXs and connects the UPFs between two different PLMNs traversing large distances. It has been observed that despite encryption, the distribution of mobile traffic can reveal necessary information about the user [26].
Randomization can help in preserving user integrity and confidentiality for such cases. We have considered three variants of this approach which are implemented on different network levels:

- GTP-based randomization

- TCP-based randomization

5.3 GTP-based randomisation

The proposed GTP-level randomization will be applied between the distributed BBU and the UPF to ensure additional user plane data confidentiality and integrity. Before applying the randomization to the data packets, a negotiation between the UPF and BBU-D has to take place. This negotiation will be signaled through the core network signaling (between AMF and SMF) through HTTP/2. These two endpoints will negotiate about the randomization parameters such as the factor by which the packet length has to be appended, the type of randomization to be applied based on the subscriber preference, etc. For the packets going from the UE towards the DN, in the uplink direction, the BBU-D will be responsible for encapsulating the payload with added randomized bytes inside the GTP-U header. For the downlink direction, when the flow of data is from DN towards the UE, this step will be performed by the UPF with the rules configured by the SMF. For this solution to work properly, a few amendments will be required in the existing PDU session establishment procedure which are given below:

- Provisioning of randomization policy in the PCF of V-PLMN and H-PLMN, for roaming scenarios.
- Inclusion of randomization header inside the GTP-U header, which is explained in the next section.
- Signalling related to randomization negotiation between BBU-D and the UPF during PDU session establishment.
- A prior agreement is needed between the roaming partners to facilitate the GTP-based Randomisation between BBU-D in VPLMN and UPF in HPLMN.

5.3.1 GTP-U Header for 5G

The different header fields of GTP-U are presented in Figure 5.2. The description of various header fields of the GTP-U header is given in Table 5.1. It shows the GTP header, which is specified in 3GPP TS 29.281, can be described as the reference header. Section 5.4.2 explains each of the extension header types that will be used in the proposed Randomisation header.

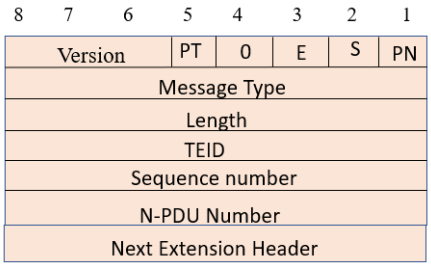


Figure 5.2: GTP-U Header fields [57]

Table 5.1: GTP-U Header fields [11]

Header Field	Description for randomisation mechanism
Version	Its value will be 1 for 5G NSA and SA variants.
Protocol Type (PT)	This demarcates GTP prime and GTP . For our case it will be set to 1 as GTP will be used.
GTP Sequence Number (S) (optional)	When it is set to '0', the Sequence Number field is absent. When it is set to '1', the Sequence Number field is available and used during inter-RAT handover for data forwarding. For the Echo Response, Echo Request, Supported Extension and Error Indication Headers Notification messages, the S flag shall be set to '1'.
Presence of NPDU (PN)	This flag indicates the presence of a viable value of the N-PDU Number field. When it is set to '0', the N-PDU Number field is absent. Useful for the inter SGSN Routing Area Update and handover procedure. It coordinates the data transmission between SGSN and MS.
Message Type	This field refers to the kind of GTP-U message.
Length	It refers to the length in octets of the payload. Minimum length- 0 bytes Maximum length- 65,535 bytes
TEID	This field is useful in finding a tunnel endpoint in the receiving GTP-U protocol entity. The other end side of a GTP tunnel locally allocates the TEID value the transmitting side should use.
Sequence Number (optional)	Sequence Number field is useful for G-PDUs (T-PDUs+headers), an increasing sequence number for T-PDUs is sent through the GTP-U tunnels, for preserving transmission.
N-PDU Number	This field is helpful for the Update procedure and Inter SGSN Routing Area Update procedures (e.g. between 2G and 3G radio access networks). This will be 0 for our case.
Next Extension header	This field defines the type of Extension Header that follows this field in the GTP-PDU.
E bit	This bit is set if there is an extension header present. For randomisation, it will be set to 1.

5.3.2 Extension header types

The randomization parameter is added to the GTP header as an extension header type and other header types which are described below [11]:

- UDP Port -This extension header is sent in Error Indication messages to indicate the UDP Source Port of the G-PDU that triggered the Error Indication. This has a length of 4 octets so the length of field is 1.
- PDCP PDU Number - This extension header is transferred to provide the PDCP sequence number of not yet acknowledged N-PDUs. Its Length field has a value of 1 and is 4 octets long.
- Long PDCP PDU Number - This extension header is used for explicit X2 or indirect S1 DL data forwarding during a Handover procedure between two eNBs. The Long PDCP PDU header is also applicable for explicit Xn or indirect N3 DL data forwarding during a Handover procedure between two NG-RANs. The Long PDCP PDU number extension header is 8 octets; therefore, the length of the field is 2.
- RAN Container - This extension header may be sent inside a G-PDU over the X2 user plane interface between the eNBs. It has a variable length and its content is specified in 3GPP TS 36.425 [25]. A G-PDU message with this extension header may be transmitted without a T-PDU.
- NR RAN Container - This extension header may be sent inside a G-PDU over the X2-U, Xn-U, and F1-U user plane interfaces, within NG-RAN and, for EN-DC, within E-UTRAN. The NR RAN Container has a fluctuating length and constitutes the information generated by the user plane protocol.
- PDU session container- It is transmitted over N3 and N9. The expansion of QoS in 5GC means the UPF should set the QoS Flow Identifier on a per-packet basis, including delay measurements or signal that Reflective QoS is being used per packet, for this, GTP has been extended to include PDU session container.
- Randomisation header- This header will be added as the next extension header to the GTP-U header. The presence of randomization will be indicated by the spare bit. If it is set to 1, it will mean randomization has to be applied.

5.3.3 Negotiation process for GTP-based randomization

The GTP-based randomization negotiation will take place between the UE and the mobile network. Figure 5.3 shows the control plane signaling that takes place during the session establishment procedure indicated by the red arrows. During session establishment the UE interacts with BBU-D through RRC message exchange and sends session establishment requests to BBU-D which further forwards the request to BBU-CU-CP and then to the AMF. The session establishment request will also contain a request for **GTP-based Randomization**. The BBU-D and BBU-CU-CP act as transparent routers to route the request to the AMF. The AMF selects the SMF based on the NSSAI received from the UE. The SMF then

interacts with UDM and the PCF to obtain the randomization policy, if it has been added as a part of the subscriber profile. The SMF takes care of the user plane packet forwarding for traffic detected by a PDR by providing FAR (Forwarding Action Rule) which consists of Forwarding target information and Forwarding operation information [10]. The UPF will add randomization during the downlink. The PDU session between UE and the DN in a 5G network is implemented as continuous flows, from Figure 5.4, the data packets are classified into different flows based on the QoS Flow Identifier (QFI).

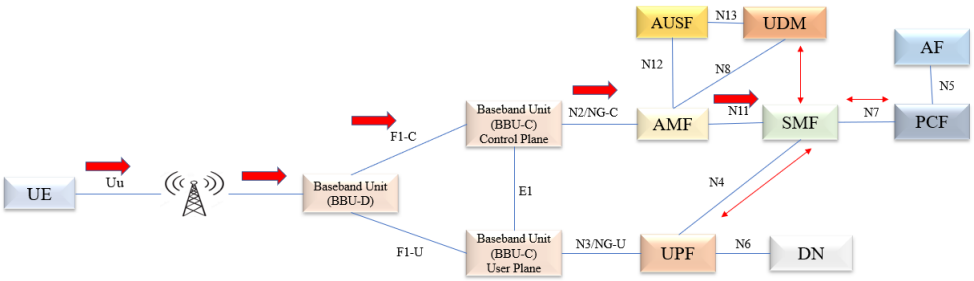


Figure 5.3: Core Network Signalling for PDU session establishment- [10]

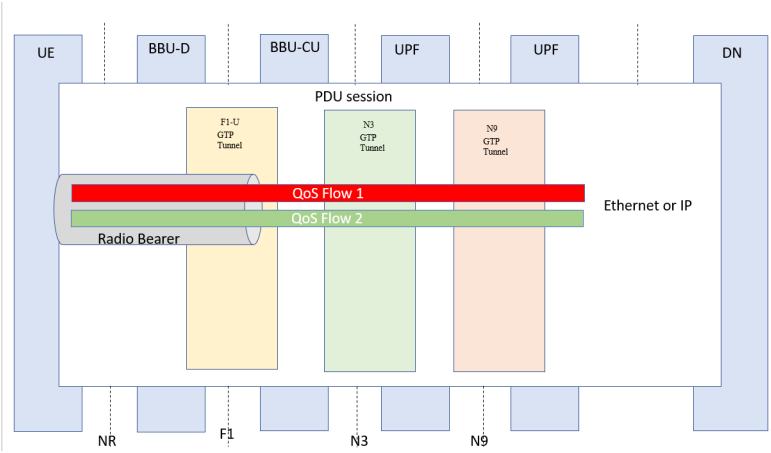


Figure 5.4: PDU session between UE and DN [10]

5.3.4 Session establishment for randomization

The Session establishment process followed for randomization will remain similar to the usual PDU Session establishment process [10] except for a few minor changes. Since GTP-based Randomisation will be offered as an add-on service, the UE will indicate this to the network. The Randomisation Session Establishment will take place as follows:

1. The UE sends a PDU session establishment request to the AN indicated by the first arrow in Figure 5.5. This request consists of the following:
    - PDU Session ID
    - UE requested DNN
    - S-NSSAI
    - Request type
    - N1 SM container
    - SM PDU DN request container
    - **Request for GTP Randomisation**
2. The AMF determines the message based on the request type. If the NAS message does not have S-NSSAI, AMF determines the S-NSSAI of the serving PLMN from the current allowed NSSAI for the UE.
3. If the AMF does not have a link with the SMF, AMF sends NSMF-PDUSession-CreateSMContextRequest else Nsmf-PDUSession-UpdateSMContextRequest is sent to SMF.
4. Based on the data given by the UE, SMF interacts with UDM over N10 and PCF over N7 to get information such as the PCC rule which has been modified to include randomization at the UPF, based on location policy for PDU session creation.
5. If the request type is an initial request, SMF starts an N4 Session establishment request with selected UPF otherwise, it sends an N4 modification request. It performs binding of SDF to QoS Flows and generates QoS rules for the UE.
6. UPF acknowledges the request and sends N4 session establishment/modification response.
7. SMF gets the GTP tunnel info from the UPF and the Tunnel is created.
8. After the Successful creation of the Tunnel endpoint, SMF sends Namf-Communication, N1N2 MessageTransfer with Tunnel Details for N2 message and PDU session details in N1 Container along with **Randomisation parameter**.
9. Upon Reception of the above message AMF Sends an NGAP PDU session Setup Request along with the N2 parameter from SMF in the above message with parameters, PDU Session ID, QoS Profile, CN tunnel Info, PDU Session type, Session AMBR. The AMF should send the randomization values to BBU-D for UL packages.
10. The NG RAN sets up the AN tunnel based on N2 information. The gNode-B forwards the N1 message to the UE for setting of PDU session.
11. The AMF updates SMF about tunnel setup by sending Nsmf-PDUSession-Update SMContext-Request and receives a response from the SMF.

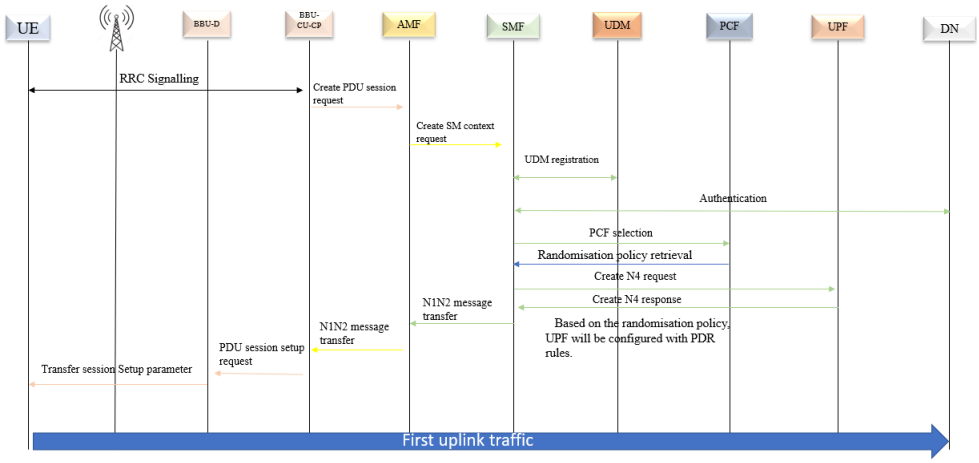


Figure 5.5: PDU Session Establishment Procedure with Randomisation

5.3.5 Randomization implementation in the User Plane

The control plane signaling for GTP-based Randomisation was explained in the previous section with relevant diagrams. In this sub-section, the flow of data from the Data Network to the UE will be explained along with the functioning of new GTP-based randomization. Figure 5.6 explains data traffic flow in the Downlink direction (DL).

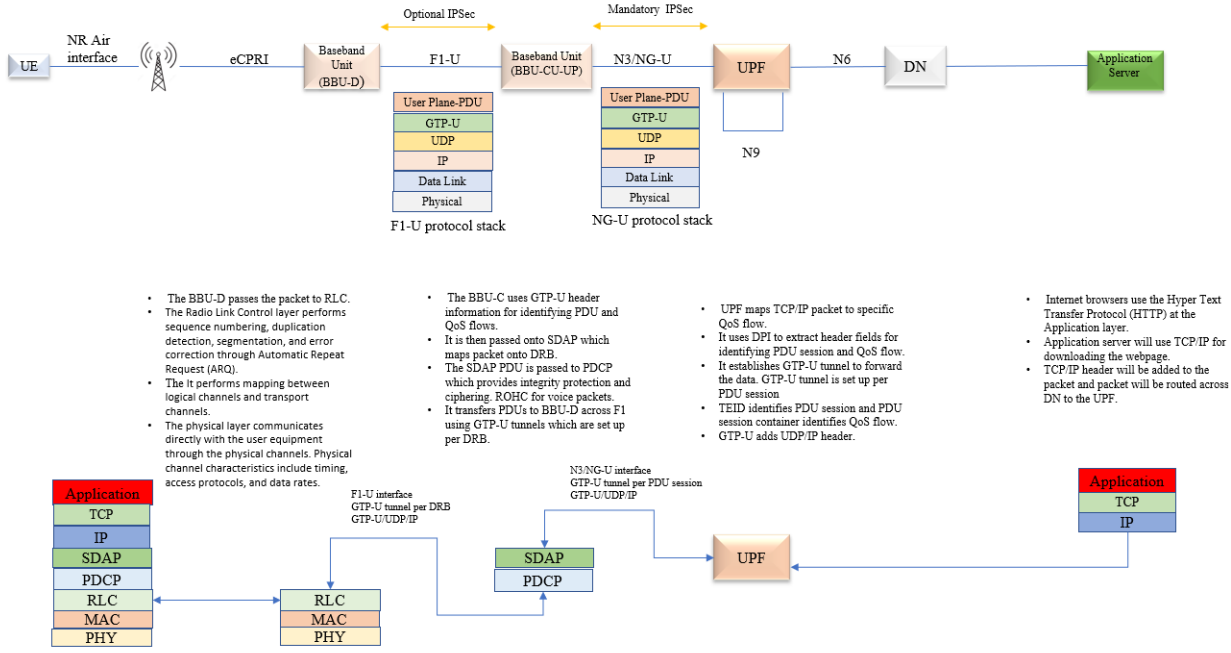


Figure 5.6: Data flow across RAN interfaces.

1. The user wants to browse the internet or use an application and download a webpage. The end user sends the HTTP GET command, as browsers use HTTP(over TCP) at the

application level, to the server hosting that webpage. The server will start downloading the web page for the end-user using the HTTP. Figure 5.7 shows the TCP header which is added by the TCP layer.

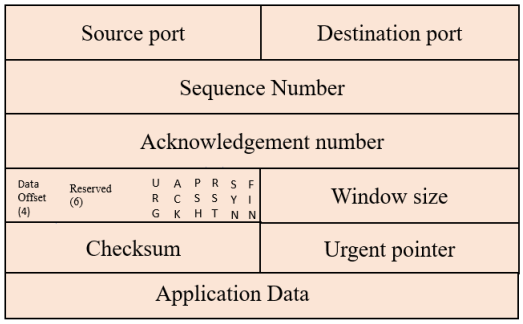


Figure 5.7: TCP header [47]

2. The header added by the IP layer is shown in Figure 5.8 below. The IP header is 20 bytes long. The header length indicates the length of the header while the total length indicates the packet size. The Differentiated Service Code Point (DSCP) is used to prioritize the packet. The Explicit Congestion Notification (ECN) conveys information about network congestion. The protocol indicates the content of the payload (application).

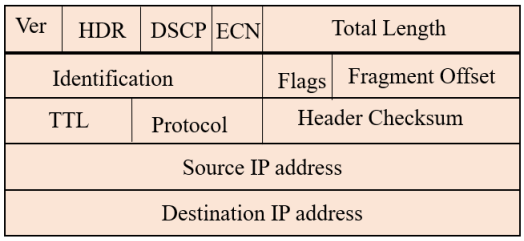


Figure 5.8: IP header

3. After the addition of the IP header, the packet is routed toward the UPF via the IP network. The IP network uses layer 2 protocols to transport the packet. The UPF then maps the packet to a specific QoS Flow. The UPF extracts various header fields using Deep Packet Inspection (DPI). These header fields are then compared against SDF rules, which the SMF sets amidst the session establishment procedure. Once the PDU session and QoS flow are identified, the UPF initiates a GTP-U tunnel for every PDU session. The TEID identifies the PDU session and the PDU session container, 5.9 is added to the header for identifying the QoS flow. The randomization header will also be added as an extension header by the UPF. The reserved fourth spare bit will be set to 1 to indicate the presence of a randomization header.

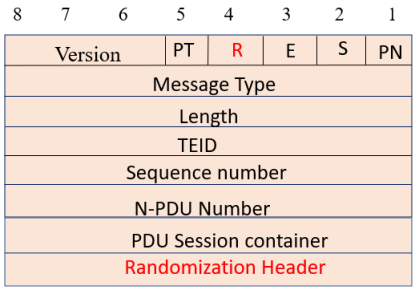


Figure 5.9: GTP-U header containing PDU session container

4. The PDU session container header is shown in Figure 5.10. The different elements of the header are as follows: Paging Policy Presence (PPP) - specifies the presence of Paging Policy Indicator (PPI). Reflective QoS Indicator (RQI) specifies whether the reflective QoS should be applied to a particular QoS

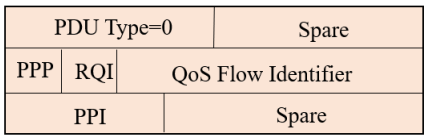


Figure 5.10: PDU session container elements [11]

5. The GTP-U tunnel runs on top of the UDP/IP protocol stack. Hence, these headers are added to the GTP-U header before routing it towards the the transport network. Figure 5.11 shows the UDP header.

Figure 5.11: UDP header [11]

6. The data packet now consists of two IP headers which is also called the IP in IP encapsulation in Figure 5.12. The packet prioritization will be done at the UPF using DSCP in the outside IP header. The GTP-U headers are removed at the distributed BBU.

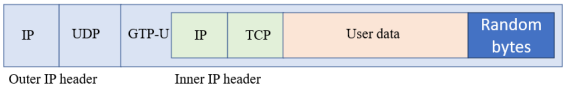


Figure 5.12: Protocol headers for Downlink dataflow.

7. The BBU-CU-UP identifies the PDU session and QoS flow from the GTP-U header. Then, the data packet is processed by the SDAP layer which maps the packet to a specific DRB. It is then passed on to the PDCP which provides ciphering and integrity protection to the payload. The PDCP PDU is then sent to the BBU-D using GTP-U over F1-C. The GTP-U header is added to the NR RAN container, described in Figure 5.13.

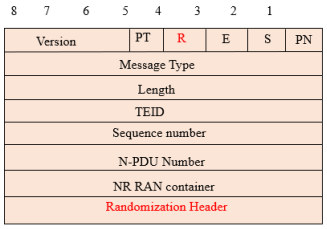


Figure 5.13: GTP-U containing NR-RAN header [11]

5.3.6 Architecture Diagram for GTP-based randomization

The proposed implementation of GTP-based randomization for 5G-SA for downlink data transfer is shown in Figure 5.14. The updated GTP header can also be included in 5G-NSA and LTE network architecture. During the downlink, the GTP header, containing the randomization bytes, is added by the UPF after mapping each packet to the appropriate QoS flow. The UPF will be provisioned with the randomization functionality, which will start adding randomization bytes after the core network signaling initiation.

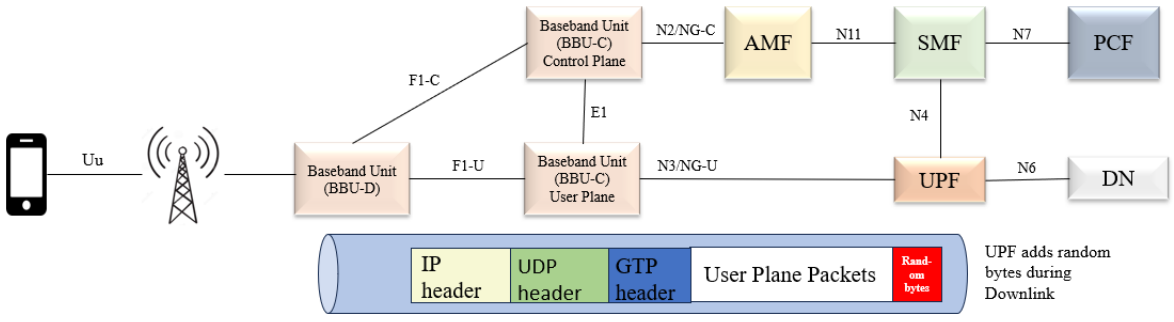


Figure 5.14: GTP-based randomisation implementation in 5G-SA for Downlink

The proposed implementation of GTP-based randomization for 5G-SA for uplink data transfer is shown in Figure 5.15. During the uplink, the BBU-D will be responsible for inserting the random bytes in the GTP-U header. The data packets are tunneled from the BBU-D to the UPF. The UPF strips off all the headers that encapsulate the data and routes the data towards the DN. In this way, the randomized bytes will be removed from the GTP

header without having any separate functionality for de-randomising the data. However, the addition of a randomization header will make the protocol bulkier.

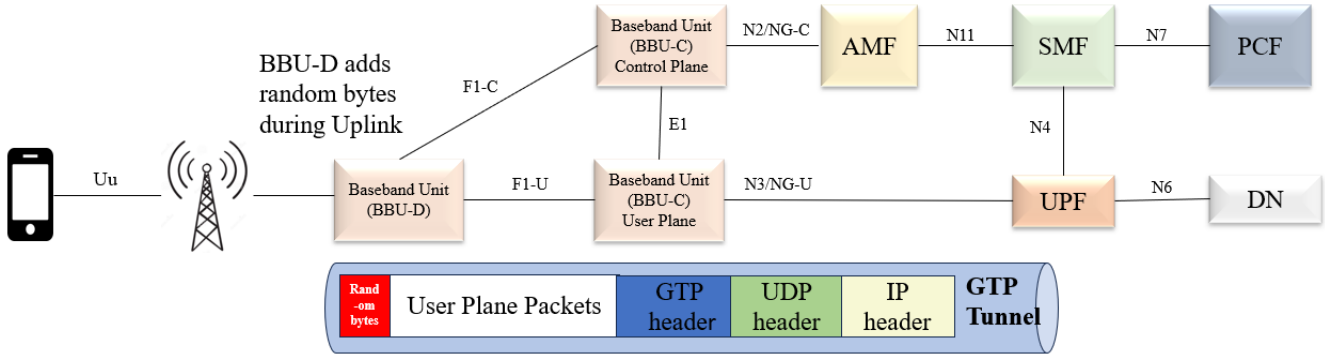


Figure 5.15: GTP-based randomisation implementation in 5G-SA for Uplink

Figure 5.16 shows network connections during roaming and GTP-based randomization implementation across different networks. In this case, for uplink, the randomization will be done by the BBU-D, and for downlink, randomization will be done by the UPF of the Home network. The network operators should negotiate earlier to include this capability in the BBU-D. Figure 5.17 shows the different network layers forming the part of the User Plane across which the randomization is being implemented.

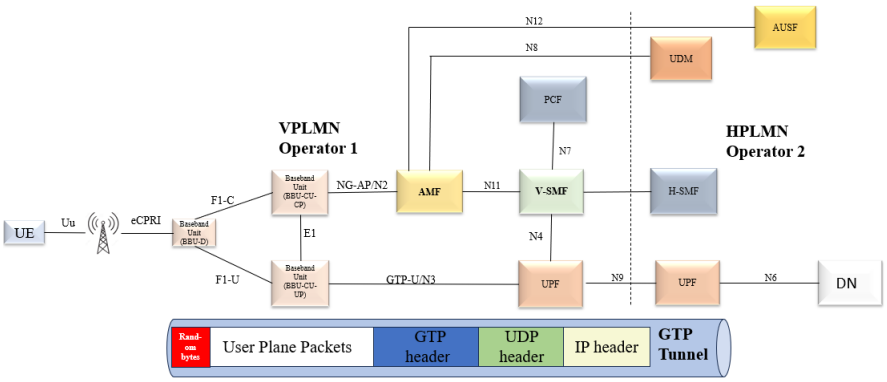


Figure 5.16: GTP-based randomization implementation for roaming scenario

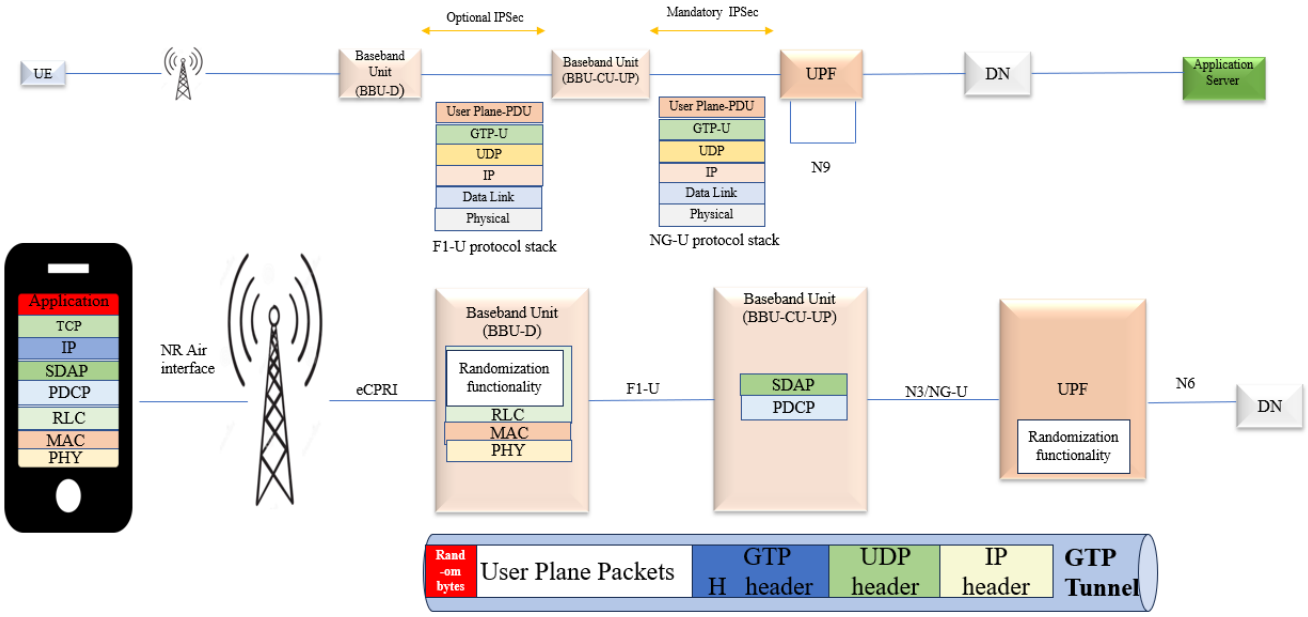


Figure 5.17: 5G randomization across different network layers

5.4 Existing Traffic Randomisation techniques

The different Randomisation techniques that have been used in computer communication are listed below:

- By shaping an application's smallest packet length in a flow to the smallest packet length of the other application's flow. The other app will be called the target application [26].
- By comparing the distribution of packet length to any other distribution and generating packet lengths from that distribution and modifying the packet lengths accordingly [26].
- By morphing one class of traffic to look like the other using convex optimization methods [63].
- Using the PCAPAnon which provides different anonymization functions such as Block Black Marker for MAC address, Prefix-Preserving and Length-PrefixPreserving (LPP) for IP address, Length-Semantics-Preserving (LSP) for pattern substitution with Reg-Exp matching and checksum adjustment [49].
- Using Anonymisation Application Programming Interface (AAPI), which consists of different functions that are applied to a traffic stream. First, the anonymization function changes fields of the packets by randomizing them, replacing them, or by performing prefix-preserving anonymization on IP addresses, etc. Filtering functions are used to differentiate traffic stream structures and then complex policies are applied such as "leave all the UDP packets unchanged but randomize the payload of all TCP

packets". Finally, there are application-level stream functions, which are called cooking and uncooking, that provide this framework the ability to compose and decompose application-level streams [46].

- Divide the data into cluster size and compute the mean of each column within each cluster, then add Laplace noise to the mean and replace every value with a perturbed mean.

5.5 Packet length Calculation for GTP-Randomisation

The exact number of random bytes to be generated will be influenced by various network conditions and parameters such as link bandwidth, cell load, throughput, and peak bitrate. The random byte generation will be dynamically controlled according to the criteria mentioned. The addition of randomized bytes will increase the packet length, the increased packet length will have impact on buffer size, Throughput and Packet loss.

The impact of increased packet length on Buffer size, Throughput, and Packet loss is explained below:

- **Impact on UPF Buffer Size**: The increase in length after randomization will increase the packet length, which will further increase the RTT. Also, the required data rates are application use case specific since video streaming, gaming, and VR will require higher data rates compared to file transfer or voice. The buffer size of UPF and RLC will be affected during Downlink and Uplink data transfer. The buffer capacity is implementation-specific. It depends a lot on 5G Key Performance Indicators (KPI) such as peak bitrate, experienced bitrate, cell bitrate, etc. The buffer size can be dynamically controlled based on the traffic load conditions.

  The empirical formula for calculation of buffer size [64] is given as;

  $$B = RTT.C/sqrt(n)$$

  where RTT is the Round-Trip Time, C is the network capacity and n is the number of concurrent flows. For DL we have taken n=6 as an assumption. The RTT values and Packet Size are taken from [66]. In Table 5.2, C is the sum of F1-U and N3 capacities. The Table 5.2 gives the approximate buffer size values for increasing packet lengths after the addition of random bytes. The buffer capacities can be planned according to the chosen values of the packet lengths.

Table 5.2: Buffer Size Calculations for various Packet Lengths

Packet Size (B)	RTT (ms)	C (GBps)	n	Buffer Size (MB)
60	10.44	28	6	1.193
512	11.82	28	6	1.351
1024	11.03	28	6	1.26

- **Impact on Throughput**: The increased packet length can significantly lower the network throughput. The maximum throughput calculated for the UPF for the downlink packet is given by [27],

$$Throughput = \frac{(L_{L1} + L_p) * B * T_x}{BW} * 100$$

where L1 is 20 bytes, Lp takes values 60 B, 512 B, and 1024 B, BW is bandwidth and is taken as 28 Gbps [44], B is the size of the UDP header, 8B and Tx is the maximum throughput. The addition of random bytes can severely degrade the throughput. The Throughput calculations in percentage are given in the Table 5.3:

Table 5.3: Throughput Calculation for various Packet Sizes

Packet Size (B)	L1(B)	B	Tx (Mbps)	Throughput (Mbps)
64	20	8	$1.12*10^9$	26.88
512	20	8	$4.41*10^7$	6.7
1024	20	8	$1.14*10^7$	3.4

- **Impact on Packet Loss at F1-U**: The packet loss at the F1-U interface will increase with the increase in packet size lengths. The packet loss is calculated as [7],

$$PacketLoss = \frac{10^6 * (MissingDLpackets)}{TotalULPackets}$$

where Missing DL packets are calculated by counting the missing GTP-U sequence numbers and UL packets are measured as the total number of packets that were generated by the UE and sent towards the BBU-D or F1 interface.

5.6 Advantages and Disadvantages of GTP-based Randomisation

The potential advantages of this approach are listed below:

- The GTP-based Randomization will be applied through the existing GTP protocol, which runs over N3, N9, and F1-U. This method comes as an in-built solution for protection during roaming as well as non-roaming scenarios.
- It can be easily integrated into the existing network procedures and protocols. It requires only one-time configurational changes and then it is easy to operate. This reduces configurational errors.
- It will be a means of making the network more trustworthy so that customers will be able to trust the technology. The stress and uncertainty around roaming security will be reduced.

- It can be extended to IoT device data for enhancing subscriber privacy.
- It can reduce the likelihood of other data privacy-related interception attacks.

The potential disadvantages of this approach are listed below:

- This method introduces additional data overhead since extra bytes are added to the user payload.
- It protects user-plane data only against a specific type of attack category (data analysis), not suitable for protection against other attack categories.
- The processing time will increase which will lead to transmission delays, increased buffer capacity and lower throughputs.
- This protocol has been designed to provide additional protection to the GTP-based reference points only. It is not suitable for other reference points.

5.7 Use Cases

The GTP-based Randomization can be helpful to different actors in different ways, who are directly or indirectly, involved with the telecom industry. The following actors will be involved:

- Subscribers: The GTP-based randomization is suitable for social networking applications that generate a lot of personal user data. The subscribers can use this method to mask their application data fingerprints.
- Roaming subscribers: It is ideal for roaming scenarios where additional protection can make the subscriber connection more secure as roaming involves attaching to lots of trusted and untrusted networks.
- Wifi subscribers: This method can also be extended to incorporate VoWifi applications since GTP runs over the S2b interface that connects PGw and ePDG.
- MNOs: The MNOs can include GTP-based randomization for offering enhanced services.
- Vendors: The vendors or any other third-party services can develop the software for adding randomization functionality to their network product.

5.8 Vulnerabilities addressed by GTP Randomisation

The 3GPP defined GTP protocol was found vulnerable to interception, impersonation, and DoS as explained in Chapter 2. Most of the GSMA recommendations, summarised in Chapter 2, are not always followed by MNOs in real-time scenarios. Since the proposed solution is an in-built alteration to the existing GTP, it may be possible to implement it easily in the network. The different ways in which the addition of GTP-based Randomisation in the network can lead to a more secure network are:

- Reduce threat to user privacy.
- Reduce threat from Roaming partners.
- Bring down User plane communication sniffing.
- Reduce the risks of subscriber information leakage.

5.9 Conclusion

The GTP-based randomization is designed for implementation on N3 and N9 reference points for the 5G-SA. This idea can be extended to 5G-NSA and 4G. It is an effort to safeguard a user's privacy when he/she roams abroad by randomizing the application data (packet lengths) which could become a means to breach individual privacy. By improvising on the existing GTP header we were able to provide additional security mechanisms over N3 and N9. Most networks do not have proper filtering rules in place, making GTP firewalls ineffective-against GTP attacks. The operators can benefit from this protection mechanism of the GTP protocol against interception, even if the firewall fails to do so. It reduces the risk of a subscriber's privacy breach in a foreign network. It can reduce the burden of configuring additional firewalls and their monitoring for MNOs. The randomization was applied on the application layer which does not have any security mechanism therefore it is possible to intercept GTP messages in the absence of the GTP firewall. This idea can be applied to 4G, 5G SA, VoWifi as well as 5G NSA. There are potential drawbacks to this method as this method introduces additional data overhead, and it will require major architecture and protocol changes for proper implementation.

Chapter 6

TCP-based Randomisation

This chapter introduces the RandomTCP, which is the proposed enhanced version of TCP, that will be more robust to application fingerprinting and that can be a viable solution for ensuring security during roaming scenarios. It also gives an overview of the generic TCP protocol and its various functions.

6.1 Overview of TCP protocol

TCP is a transport layer protocol that falls just below the application layer of the OSI stack. TCP is used in LTE on many reference points and now it will continue to be a part of the 5G system. TCP, in 5GS, is primarily responsible for the transport of application layer messages between the application in the Data Network and the UE. It provides reliable delivery of data along with mechanisms for error control, flow control, and congestion control. Since TCP provides reliable data delivery, it is preferred over UDP for non-real-time applications in which packet loss is critical. This protocol works over the client-server model, which means that the client which can be a 5G user device will initiate a connection request to a server that is listening at a specific port. The user acting as a client, will use HTTP to browse webpages on a web server using any web browser. The HTTP packet is delivered to the server's port 80 over TCP protocol. The application layer protocol will rely on TCP for the reliable delivery of messages. The application data that needs to be transferred is encapsulated inside the TCP header and then it becomes the TCP segment. The TCP header, in Figure 6.1, is a variable length header with a minimum length of 20 bytes and a maximum length, that can extend up to 60 bytes. The TCP header fields are explained in Table 6.1 The first 10 fields are mandatory which account for the fixed 20 bytes. The entire TCP connection can be easily identified by specifying 5 tuple fields: destination port, Protocol, source IP address, destination IP address, and source port. A TCP connection between two entities is defined by a TCP socket which is a combination of IP address and port number. The details related to TCP socket establishment are given in Section 6.1.1.

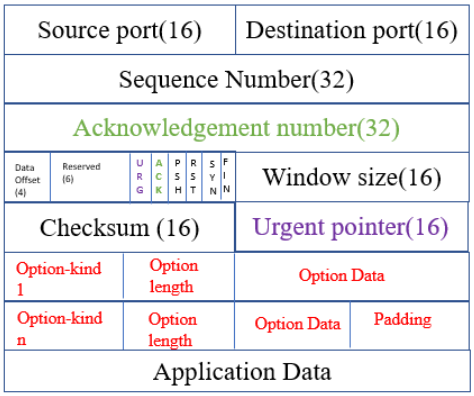


Figure 6.1: TCP header fields [33]

Table 6.1: TCP header fields description [33]

Header Field	Description
Source Port	Address of the sender port.
Destination Port	Address of the receiver port.
Sequence Number	It is a 32 bit field indicating the amount of data sent.
Ack Number	It is a 32-bit field used by the receiver to ask for the next frame.
Data Offset	It indicates the beginning of the data.
Reserved/ Randomization field	It has been added for using Randomization.
Control bits	ACK, URG, PSH, RST, FIN and SYN.
Window	It is used for flow control.
Checksum	checks the TCP header for errors
Urgent Pointer	Indicates the end of urgent bit when URG is set.
Options	randomization option along with MSS, Window Scaling, Selective acknowledgement, Nop.

6.1.1 Connection establishment procedure/Three-way Handshake

A TCP connection is always established with a three-way handshake between the client and the server, which is the first step towards connection establishment. In Figure 6.4, the first three arrows show the handshake process and fourth arrow shows the start of the data transfer. Before the three-way handshake, the server should bind to and listen at a specific port which means that the server is in the passive open state. The client first creates a socket object. During socket object creation it specifies the IP address of the server and port number of the process. Once the socket is created, the client then initiates active open by sending a SYN message and a sequence number to the server for connection establishment. The SYN

message is a flag inside the TCP header which is set to 1 when initiating connection. The rest of the TCP header is empty as there is no data, only the source port, destination port, SYN, and sequence number field are filled with values. Only with the help of these flags, do servers decide how to process the packet. After a successful handshake, the server creates a new socket for the client by dedicating a specific port for the client. Thus, a TCP connection exists between the client socket and the server socket. The Table 6.2 describes the additional header fields.

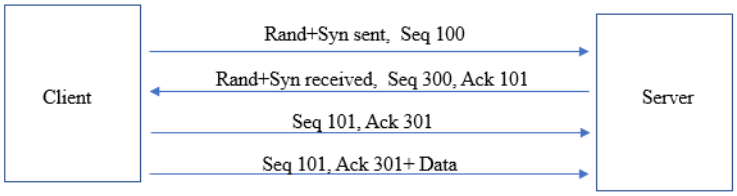


Figure 6.2: Three-way TCP handshake process

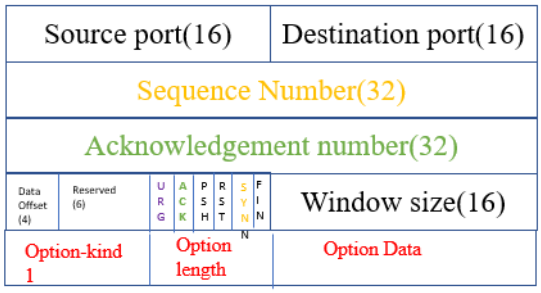


Figure 6.3: Special TCP segment used for connection establishment with no payload

Table 6.2: TCP Options [47]

Option Kind	Option Length	Option-Data	Description
0			End of Option List: It marks the end of all options present in the segment. It is included only when the end of the options doesn't match with the end of the TCP header.
1			No-Operation: A "spacer" that can be used between the options to align a subsequent option on a 32-bit boundary if needed.
2	4	Maximum Segment Size value	Maximum Segment Size: Describes the size of the largest segment, chosen by the receiver. Used only in connection request (SYN) messages
3	3	Window size shift bits	Window Scale: Implements the optional window scale feature,which permits devices to specify much larger window sizes than would be available with the conventional Window field.
4	2		Selective Acknowledgment Permitted: Describes that the device supports the selective acknowledgment (SACK) feature.
5	variable	Data selectively acknowledged	Selective Acknowledgment: Helps devices supporting the optional selective acknowledgment feature to specify non-contiguous blocks of information that have been received so they are not re-sent if intervening segments do not show up and need to be re-sent.
14	3	Alternate checksum algorithm	Alternate Checksum Request: Allows a device to request for a checksum generation algorithm other than the standard TCP algorithm to be used for this connection. Both devices must agree to the algorithm to be used.
15	variable	Alternate checksum	Alternate Checksum: If the checksum value required to implement an alternate checksum is too bulky to fit in the standard 16-bit Checksum field.
r	variable	Randomised bytes	These are added to the application data to randomize packet lengths.

6.1.2 Data transfer mechanism

The UE, which consists of TCP protocol stack, acts as a client and will use HTTP for browsing through the web pages. An HTTP packet is delivered to the server's port 80 using TCP while the source port will be a random number. TCP will divide the data into TCP segments and number each one of them. TCP will also ensure that none of the segments is delivered out of order. Only the sequence number of the first data byte in a block (TCP segment) is sent to the destination host since data is transmitted in blocks (TCP segments). Sequence numbers are used by TCP servers to rearrange segments when they arrive out of order and to eliminate duplicate segments [47]. In an ACK, the receiving TCP server informs the sender how many bytes it will be able to receive (beyond the last TCP segment) without overflowing or overrunning its internal buffers. This information is sent in ACK. The TCP also waits for acknowledgment from the receiver in a defined time slot. Once the handshake is done, the server TCP allocates a dedicated port for that application. In this manner, the client and server will be able to run multiple applications using different dedicated ports.

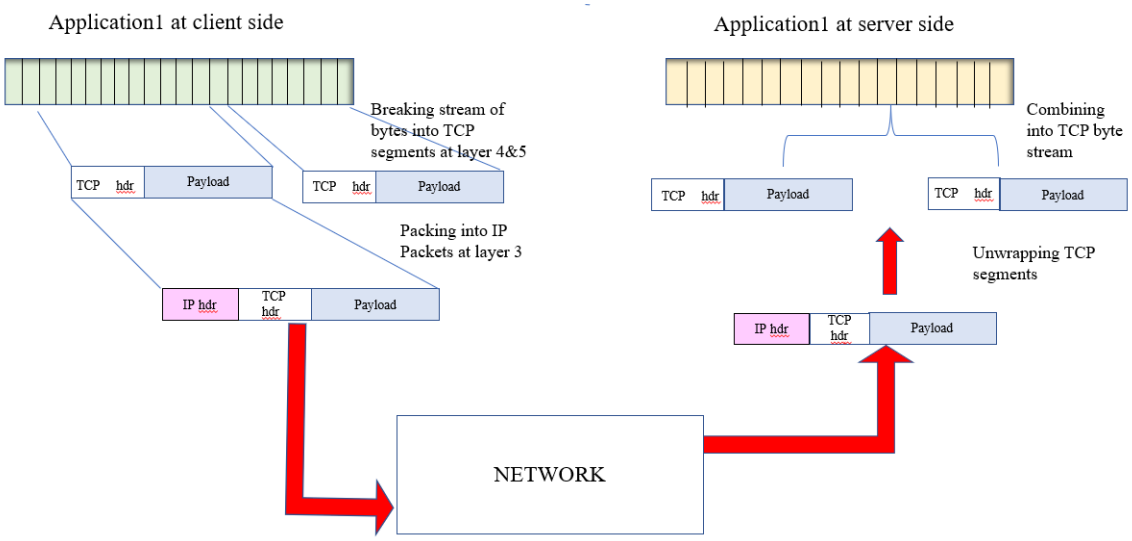


Figure 6.4: TCP data transfer process [47]

6.1.3 Connection maintenance

With an active TCP connection, keepalive timers (special TCP messages) are associated with connection maintenance. When the keepalive timer reaches zero, an empty probe packet is sent to the peer. This probe packet does not contain any data, only the acknowledgment flag is set. It is like a duplicate acknowledgment that is sent to verify if the connection is still up and running. If there is a reply from the other peer with another set acknowledgment flag and no data, it means that the connection is still valid. This method of checking the connection generates extra overhead which can impact routers.

6.2 RandomTCP - A TCP-based Randomization protocol

The TCP-based randomization will be implemented as a two-step process. In the first step, negotiation for randomization will be carried out between the UE and application server through an extended handshake process. In the second step, the actual randomization will take place. The TCP header currently does not have extension capabilities. However, the reserved 6 bits, in Figure 6.1 can be used in the future to indicate the presence of a randomization parameter list along with options for randomization. The proposed enhanced header is presented in Figure 6.5. The Reserved field of the TCP header will used to create an additional flag called RAND which is 1 bit long and the new RESERVED field will be 5 bit long. If the RAND flag is set then the server will assume that the client has opted for randomisation. The randomization parameters will be included in the OPTIONS field. If the server is capable of implementing randomisation then it will reply with a set RAND flag along with the chosen randomisation parameter from the list. If it is incapable of applying randomization, then it will proceed with the classic TCP without randomization.

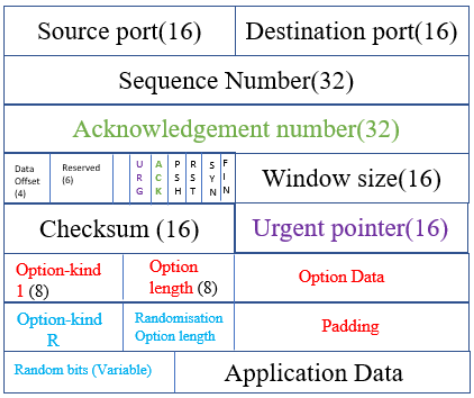


Figure 6.5: Extended TCP header fields for TCP randomization

6.2.1 RandomTCP negotiation mechanism

The negotiation process for the implementation of RandomTCP is given below:

1. The client will create a TCP socket by specifying its source port, destination port, source IP address, destination IP address, and Protocol. The UE will establish a connection, i.e. TCP socket, by initiating the extended three-way handshake.

2. The first message will be SYN, RAND directed at port 80 along with the list of randomization parameters as explained in Figure 6.6. The server acknowledges the SYN and RAND by selecting the randomization parameter.

3. Then UE acknowledges the received messages and sends ACK. This completes the three-way handshake process and now the randomized data transfer will take place between UE and the server. In the next step, data randomization will take place at a different server port and not on port 80 as this process will require a dedicated port for the addition of random bytes.

The client starts the process by sending the TCP SYN message which contains the sequence number and randomization flag, set to 1. After receiving the message, the server will send its sequence number and the client's incremented sequence number as a positive acknowledgment. The client will respond by sending acknowledgment (server's sequence number incremented by 1). Figure 6.6 below explains this process more clearly. The client sends a SYN segment and RAND segment with a sequence number 100, along with the window size, denoted by Win and MSS. The server receives the SYN+RAND, acknowledges it and sends back ACK 101 and gives its sequence number 300. The client acknowledges the SYN sent by the server and starts sending data.

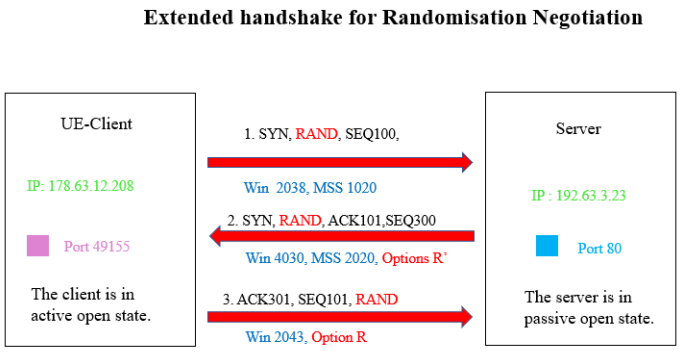


Figure 6.6: Extended Three-way handshake for TCP connection establishment

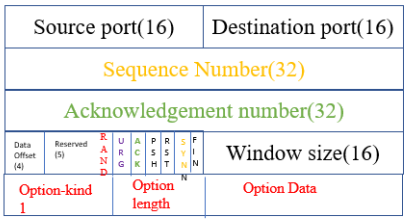


Figure 6.7: RandomTCP header

6.3 RandomTCP Architecture Diagram

The UEs or the mobile device is configured with the TCP protocol stack which establishes an end-to-end connection with the TCP server during the PDU session establishment explained through Figure 6.8. The TCP protocol has been a part of 3G, 4G, and now 5G in the core network and for transport of application-related data. The data flows in both the uplink as well as the downlink direction. During the downlink, the application data packets, which will be carried by the Random-TCP, are routed to the UPF which maps the data packets to the appropriate QoS flow. These are then tunneled via the GTP and over the air interface to the UE, the GTP tunnel terminates at the BBU-D. During the uplink, all the application packets are encapsulated in the Random-TCP header and routed across the internet to the specific application server. The Random-TCP will help in masking the application data distribution. The masking pattern will differ for different applications since this masking will be based on Congestion Window size and segment size. For the roaming scenario in Figure 6.9 the data packets are routed from the home network to the visited network across the N9. The random-TCP in the roaming scenario will behave in a similar manner to the non-roaming scenario. The only difference is the increased distance from the UE.

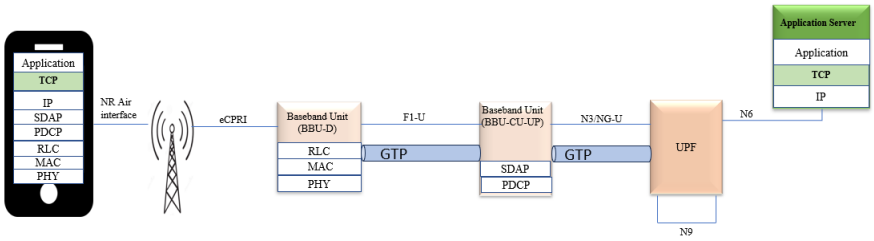


Figure 6.8: RandomTCP implementation for non-roaming scenario

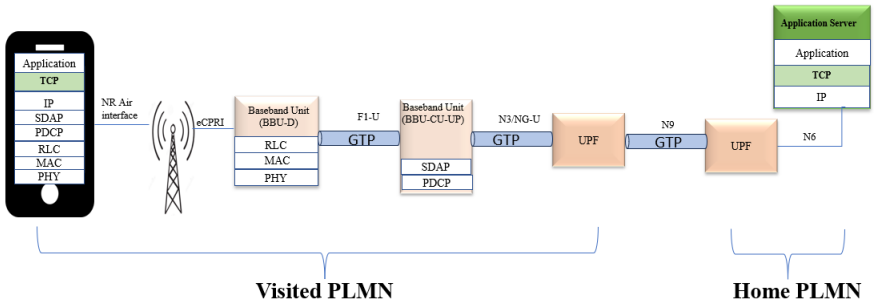


Figure 6.9: RandomTCP implementation for roaming scenario

6.4 Overwiew of RandomTCP Operation

At this stage, the TCP connection exists between the given pair of sockets as evident from Figure 6.10. It is also evident from Figure 6.10 that the server has created a new port number 40510, which is an ephemeral port number, for running any specific application. We have assumed that the subscriber has chosen to use the RandomTCP and Randomization-related signaling was successfully performed during the handshake along with the Randomization option negotiation between the client and the server. The RandomTCP protocol works by randomizing the payload length of each TCP packet before transmitting the packet out of the buffer. This randomization is useful in hiding the signature distribution of each application. As seen earlier, each application can be classified by observing the distribution between packet length and inter-arrival times of these packets. This variant of TCP will consist of a Random Byte calculator. The working operation is described in detail in Sections 6.4.2 and 6.4.3. Two approaches have been worked out for the randomization of payload length. The first approach as described in Algorithm 1 calculates Trand as the difference between Wmax and CWND. This difference will calculate the unique number of bytes that will be added to each packet. The second approach described in Algorithm 2 calculates the payload length based on the CWND size and distribution parameters of each application. The next few subsections provide more details of the exact working of the RandomTCP protocol.

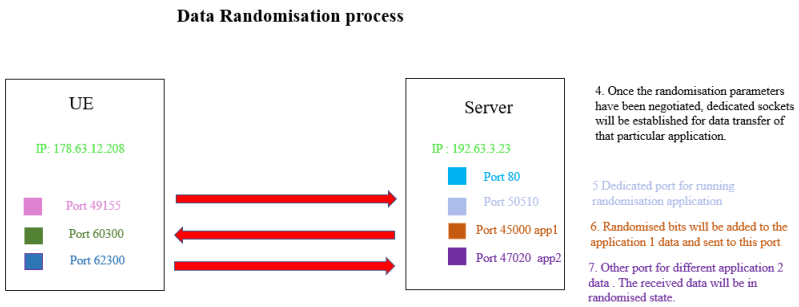


Figure 6.10: Randomisation process

6.4.1 Revision of Randomization parameters

The TCP-based Randomization protocol also provides a provision to update randomization parameters throughout the lifetime of the connection. This parameter update can be done with the help of the existing TCP keepalive mechanism. The acknowledgment flag in the header can be accompanied by the set RAND flag along with the randomization parameter list. When there is an acknowledgment reply, the chosen randomization parameter can be accompanied by the ACK. In this manner, at the cost of extra overhead, randomization parameter updates can be done by both parties.

6.4.2 RandomTCP Principle

The transport layer protocol, TCP is used in different ways for classification of application. One such study takes into account the first five TCP packets to classify different applications. It further concluded that only the first five packet lengths are enough for precise app classification. Another study takes into account the length of the first n flows for grouping the applications. The RandomTCP could be a potential solution for randomizing data lengths. It has been applied for TCP BBR [23], which is the most suitable TCP variant for long-distance congestion control and ideal for our 5G roaming scenarios. The RandomTCP variable Trand will calculate a new payload length L' that will lie between 1 and the MSS. The two algorithms for generating two different L' are provided in Section 6.4.3. This can be applied to all the applications running at different ports.

6.4.3 Random Byte calculator

For the generation of Random bytes, two approaches are proposed. The first algorithm which calculates the number of padding bytes based on the difference is given below. This algorithm has been designed to work with TCP BBR, but with slight changes, it can work with other TCP variants too. The input parameters are Wmax, CWND, MSS, and L. The Trand is the variable that stores the difference between CWND and Wmax. The new length L' is defined with Trand and original packet length L. Based on L', the number of bytes can be added to the payload. The input variables are defined below :

- Wmax (Maximum Window size): The maximum Window size is an indication of how much data the receiver can take without dropping the packets. It is also known as the

maximum Receiver Window size, in most operating systems it is fixed at 65,536 bytes or 12 MSS. It is possible to scale up this window size depending on network traffic conditions. The Wmax is involved in flow control as it puts an upper bound on how much data is to be sent. The data to be sent should be less than or equal to Wmax.

- CWND (Congestion Window size): The TCP has its mechanism for congestion control which is different for its different variants. The congestion window is known only to the sender. The congestion window size or CWND is initially kept as 1 MSS and later increases exponentially depending upon the number of acknowledgments. The data to be sent has to be less than or equal to the congestion window. The sender window size is the minimum of RCWND and CWND.

- L (original Payload Length): It is the length of each segment. The maximum length is equal to equal to 1460 bytes.

- MSS (Maximum Segment size): The payload length is divided into TCP segments for transmission. The maximum segment size that can be transmitted is 1460 bytes.

The first algorithm changes the payload length based on the normal distribution of each application. The work done in [26] explains the different normal distributions that were plotted between inter-arrival packet times and payload lengths using Normal distributions and Poisson distributions. The Figure 6.11 shows how the payload length will change on applying randomization. To plot this kind of distribution in MATLAB, the standard deviation and mean values should be known. Then it alters the length according to the L' calculations as shown below:

Algorithm 1: For randomisation of TCP flow

Input: Wmax,CWND,MSS,L,μ, θ
Output: L'
Data: Testing set x

$X = \frac{cwnd}{Wmax}$
if $CWND < Wmax$ **then**
  $\Phi = \frac{1}{2\sqrt{(\pi)}} \int_{\frac{-x-\mu}{\sigma\sqrt{(2\pi)}}}^{\frac{x-\mu}{\sigma\sqrt{(2\pi)}}} \exp t^2 dt$
  $Trand = \phi * \theta$
  $L' = L * Trand$
end
else
  Trand = MSS
end

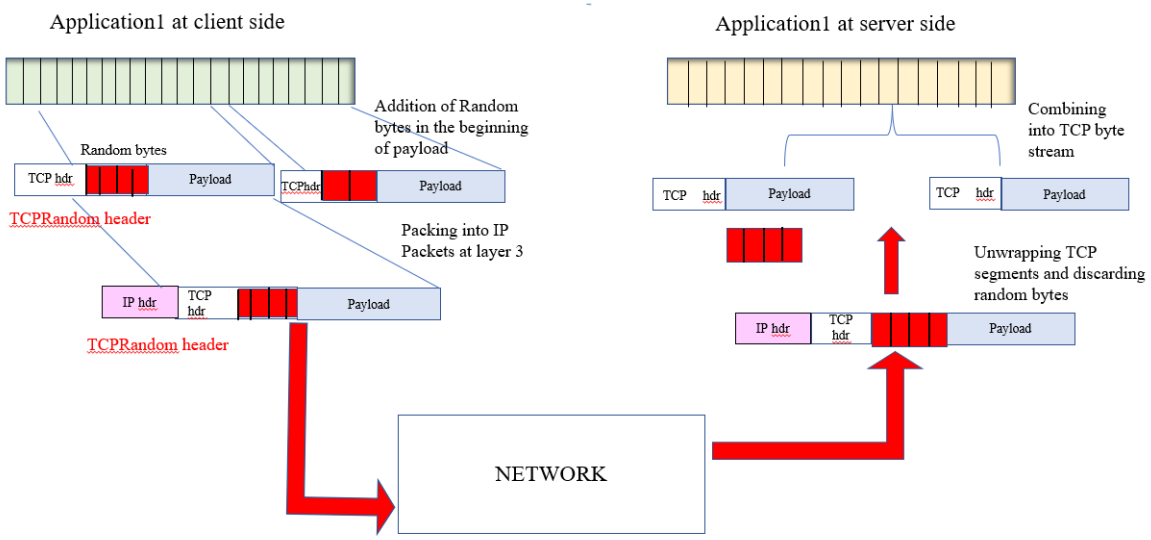


Figure 6.11: Diagram showing the addition of random bytes to payload using Algorithm1 and Algorithm 2

Algorithm 2: For adding Random bytes to TCP flow

Input: Wmax, CWND, MSS, L

```
if Wmax < CWND then
    Trand = CWND − Wmax
else
    Wmax > CWND
    Trand = Wmax − CWND
L'= L + Trand
else
    Trand = MSS
```

1

In Figure 6.13, X-axis represents congestion window size (cwnd) and Y-axis represents random bytes (trand) to be generated. The value of 'cwnd' is inversely proportional to 'trand' that means when the value of 'cwnd' is low then corresponding value of 'trand' is high. Relationship between 'cwnd' and 'trand' shows number of random bytes to be added for optimized functioning of the TCP protocol.

```
wmax = 65;      % It specifies the maximum receiver window size.
l = 1400;       % It is the segment length.
mss = 536;      % It is the maximum segment size.
trand=zeros(1,10);
cwnd = 1:10;    % It depends on RTT and BDP

for x = 1:length(cwnd)
    if wmax < cwnd(x)
        trand(x) = cwnd(x) - wmax;
    elseif wmax > cwnd(x)
        trand(x) = wmax - cwnd(x);
        l1 = l + trand(x);

    else
        trand(x) = mss;
    end
end
figure(11)
plot(cwnd, trand)
xlim([0 10])
ylim([55 65])
xlabel('CWND')
ylabel('trand')
```

Figure 6.12: MATLAB code addition of random bytes to payload using Algorithm 1

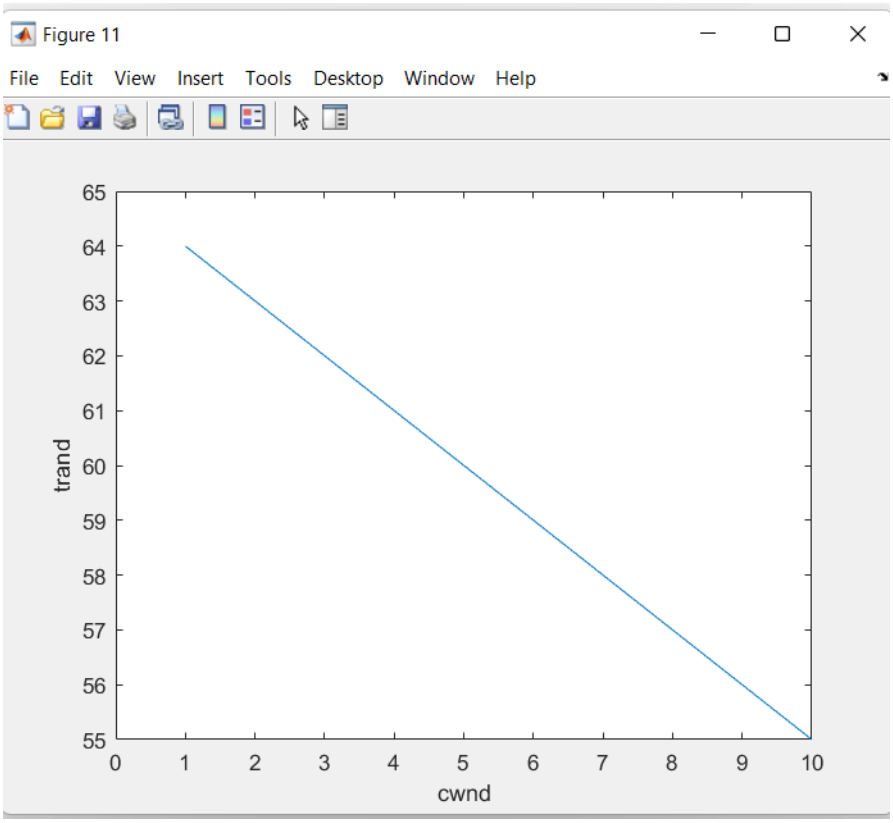


Figure 6.13: Diagram showing the addition of random bytes to payload using Algorithm 2

6.5 Advantages and Disadvantages of TCP-based Randomisation

The following are the advantages of this method:

- The Random-TCP protocol can act as an additional security mechanism in combination with the TLS.
- It makes the TCP robust to different kinds of interception attacks.
- All applications using the TCP protocol can widely adopt the TCP-based randomization method.
- This method may come as an in-built solution for protection during roaming as well as non-roaming scenarios.
- It might be easily integrated into the existing network procedures and protocols. It may require only one-time configurational changes which can result in reduced configuration errors.
- It will be a means of making the network more trustworthy and customers will be able to trust the technology and the stress and uncertainty around it will be reduced.
- It can be extended to IoT device data for enhancing subscriber privacy.
- It can reduce the likelihood of other data privacy-related attacks.

The following are the disadvantages of TCP-based Randomisation:

- This method introduces extra data overhead, similar to GTP-based randomization and requires extra processing time.
- It may add extra costs for the MNOs.
- New updates must be made to the current TCP protocol header which requires additional efforts.
- Due to a change in the TCP protocol stack, terminal OS update will be required.

6.6 Use Cases

The following actors will be involved:

- Subscribers - The TCP-based randomization can be used for all mobile applications like Facebook, YouTube, and Viber that generate a lot of personal user data.
- Roaming subscribers - It is ideal for roaming scenarios where additional protection for masking application data can make the subscriber´s TCP connection more secure.
- MNOs - The MNOs can provide enhanced TCP services for various applications like home IoT, IoT and roaming security.

6.7 Conclusion

The transport layer messages are already protected with the TLS protocol. However, the distribution of packet length with time can reveal the type of application used by the customer, as explained in Chapter 2, Section 2.3. Thus, in order to make the existing TCP robust to this vulnerability in 5G-SA, the TCP-based Randomisation has been suggested here. The TCP-based Randomisation is applied on the transport layer between the UE and the application server. The TCP protocol header has provisions such as empty unused bits that make it ideal for the enhancements proposed in this chapter. The TCP randomization will be useful for the 5G user plane. It can benefit core network signaling if the correlation between signaling packets are found in future communications. For implementing TCP-based Randomisation, significant changes have to be made to update the existing Android OS, since, the protocol stack of our TCP will have extra randomization bytes and additional headers compared to the traditional TCP. The MNOs can also use this method for Home IoT, since TCP is considered a better choice for IoT, as specified in the standards [48].

Chapter 7

Conclusion and Future Research

This chapter provides a conclusion to the thesis work and provides ideas for Future Research that can be undertaken later based on the current findings.

7.1 Conclusion

In this thesis, a vulnerability analysis of 5G reference points was done and highly prone reference points were identified. Randomization was introduced as a protection mechanism for TCP and GTP and risk analysis of N-32 was performed.

The following conclusions can be derived from the research work done on GTP-based randomization and TCP-based randomization for enhancing 5G roaming security:

- It was concluded that the operators should be careful with JSON versions and mismatching of configuration policies.
- From the SEPP vulnerability analysis matrix, it can be concluded that crashing Network Functions with unwanted messages is highly likely since more devices will be connected to the 5G core. The increased number of 5G devices also makes the risk of attacks by signaling overloads highly likely since a lot of overhead will be generated if PRINS is implemented.
- Both the TCP-based and GTP-based randomization methods require some form of header modification but if the subscriber is unwilling to use randomization, he/she can also proceed with the classic TCP and GTP without randomization.
- The TCP-based randomization approach is more generic compared to the GTP-based randomization as it can be implemented in the 5G core network, IMS network, and mobile network as well. The GTP-based randomization can only be implemented between UPF and BBU-D in both directions.
- Algorithm 2, for adding the random bytes to the TCP header takes into account the currently available space in each segment for the addition of random bytes. Algorithm 1 modifies the length by using the packet length distribution for other applications. Amongst these two algorithms, algorithm 1 masks the signature distribution better than Algorithm 2.

- The TCP segments are not uniform in size, their lengths vary despite a fixed MSS value. This variation in segment lengths depends on network conditions such as the number of devices and rate control algorithm.

- The randomization methods were applied on the TCP, which is a transport-level protocol and GTP which is an application-level protocol. The transport layer implementation calls for alterations in the TCP header.

- For making the Random-TCP ready for use in the telecommunication industry, the IETF will need to impart the reserved bit field space to indicate the presence of the Randomisation header. This change is a mandatory requirement.

- The Random-TCP is difficult to implement since it does not comprise an extension capability through a header like GTP-U.

- The updated GTP-U will occupy more buffer space but will provide improved protection for roaming abroad, which will minimize the risk of mobile traffic data analysis-related threats.

- For implementing GTP-based randomization, an additional functionality that would provide extra bytes will be added to BBU-D and the UPF as part of the RLC layer and the GTP extension header capability should be sanctioned by 3GPP for using it as a means to add random bytes.

- The GTP-based Randomisation solution can be implemented by making available the updated GTP protocol as a special service to subscribers. These subscribers can choose this subscription when roaming abroad. The TCP-based Randomization solution can be implemented for all kinds of internet-based applications.

- The GTP-based Randomization is effective in masking the application fingerprint over the N3 and N9 roaming reference points in 5G-SA as it is implemented over the application layer protocol, which may or may not be secured according to the GSMA-recommended measures inside the operator´s network. On the other hand, TCP-based Randomization effectively masks the entire application fingerprint between the mobile phone and the application server. The TCP is considered secure when there is TLS protection in place, which encrypts the TCP data. However, there are application classification methods available, which segregate the traffic based on TCP flow property.

7.2 Future Research

Certain topics were partially explored in this thesis. Hence, further research work can be undertaken on these topics in the future. A few future research ideas are summarised below:

- **Analysis of Randomisation Function placement in the BBU-D.** For this work, placement of the Randomisation function inside the RLC was considered but the Randomisation Function can be a part of other BBU-D layers.

- **Analysis of different Randomization variants.** In Chapter 5 of the thesis, message length-based Randomisation has been worked upon. There are other types of randomization algorithms such as those based on Pseudorandom Bit Generation (PRB), Warner´s randomized response, and output perturbation [20].

- **Cost analysis for N-32.** Since there will be increased signaling overhead due to PRINS, it will lead to an increase in infrastructure costs. Further work can be done to estimate the percentage increase in the overhead, in the future, when PRINS will be implemented.

- **Application-based Randomisation.** In Chapters 5 and 6, GTP-based Randomization and TCP-based Randomisation were explored. Another means of protection mechanism can be Application-based Randomization, which will include an option for applications to implement Randomization themselves. There can be an initial negotiation phase wherein the mobile application can negotiate with the application server whether to implement or not implement Randomization. In this manner, no modifications will be required to the TCP protocol stack in the mobile phone

- **IMS-based Randomization.** The IMS-based randomization for voice services was found infeasible, as the addition of random bytes can severely degrade the voice quality. More work can be done to find out the impact of Randomization for IMS-based text services.

List of Abbreviations

AF	Application Function
AMF	Access and Mobility Management Function
AN	Access Network
API	Application Programming Interface
BBU	Baseband Unit
BBU-CU-CP	Baseband Unit-Centralised Unit-Control Plane
BBU-CU-UP	Baseband Unit-Centralised Unit-User Plane
BBU-D	Baseband Unit-Distributed
BSF	Binding Support Function
CDN	Content Delivery Network
C-RAN	Centralized Radio Access Network
CUPS	Control and User Plane Separation
DHCP	Dynamic Host Configuration Protocol
DN	Data Network
DPI	Deep Packet Inspection
DRA	Diameter Routing Agent
EPC	Enhanced Packet Core
GTP	GPRS Tunnelling Protocol
GTP-C	GPRRS Tunnelling Protocol-Control Plane
GTP-U	GPRS Tunnelling Protocol-User Plane
GUAMI	Globally Unique AMF Identifier

GUTI	Globally Unique Temporary Identifier
HPLMN	Home-Public Land Mobile Network
HR	Home Routing
IMS	IP Multimedia Services
IMSI	International Mobile Subscriber Identity
IP	Internet Protocol
JSON	JavaScript Object Notation
LBO	Local Breakout
LTE	Long-Term Evolution
MAC	Media Access Conrol
MCG	Master Cell Group
MITM	Man-in-the-Middle
MME	Mobility Management Identity
MN	Master Node
MNO	Mobile Network Operator
NAS	Non-Access Stratum
NEF	Network Exposure Function
NG-AP	Next-Generation Application Protocol
NGC	Next Generation Core
NSA	Non-Standalone
NSI	Network Slice Identifier
NSSF	Network Slice Selection Function
PCF	Policy Control Function
PCRF	Policy and Charging Rules Function
PDCP	Packet Data Convergence Protocol
PDU	Protocol Data Unit
PHY	Physical Layer

PNF	Physical Network Function
QoS	Quality of Service
RAN	Radio Access Network
RF	Radio Frequency
RLC	Radio Link Control
RNL	Radio Network Layer
RRC	Radio Resource Control
RRH	Remote Radio Head
SCG	Secondary Cell Group
SCP	Service Communication Proxy
SCTP	Stream Control Transmission Protocol
SDAP	Service Data Adaptation Protocol
SEAF	Security Anchor Function
SEAF	Security Anchor Function
SEPP	Security Edge protection proxy
SMF	Session Management Function
SS7	Signalling System 7
SSL	Secure Socket Layer
SUPI	Subscription Permanent Identifier
TCP	Transmission Control Protocol
TLS	Transport Layer Security
TNL	Transport Network Layer
UDM	Unified Data Management
UDP	User Datagram Protocol
UE	User Equipment
UE	User Equipment
VNF	Virtualised Network Function

V-RAN	Virtualized-Radio Access Network
3G	Third Generation
3GPP	3rd Generation Partnership Project
4G	Fourth Generation
5G	Fifth Generation

Bibliography

[1] GTP vulnerabilities: A cause for concern in 5G and LTE networks . Technical report, Motorola.

[2] LTE International Roaming Whitepaper. Technical report, Huawei.

[3] 3GPP. *3GPP TS 29.500 version 15.4.0 Release 15,5G System; Technical Realization of Service-Based Architecture; Stage 3*.

[4] 3GPP. *3GPP TS 38.101-1 V15.3.0 (2018-10),5G; NR; User Equipment (UE) radio transmission and reception; Part 1: Range 1 Standalone ; Stage 2*, 2018. Rel. 15.0.

[5] 3GPP. *3GPP TS 38.300 V15.3.1 (2018-10),5G; NR; Overall description; Stage-2 (3GPP TS 38.300 version 15.3.1 Release 15) ; Stage 2*, 2018. Rel. 15.0.

[6] 3GPP. *Security Aspects; Study on security aspects of the 5G Service Based Architecture (SBA (3GPP TR 33.855 V1.4.0 (2019-03))*, 02 2019. Rel. 16.

[7] 3GPP. *3GPP TS 28.552 version 16.6.0 (2020-08), 5G; Management and orchestration; 5G performance measurements*, 2020. Rel. 16.

[8] 3GPP. *LTE; Security Assurance Specification (SCAS) threats and critical assets in 3GPP network product classes (3GPP TR 33.926 version 16.3.0 Release 16)*, 10 2020. Rel. 16.3.0.

[9] 3GPP. *3GPP TS 23.501 V17.0.0 (2021-03),System architecture for the 5G System (5GS); Stage 2*, 2021. Rel. 17.0.

[10] 3GPP. *3GPP TS 23.502 V17.0.0 (2021-03),Technical Specification Group Services and System Aspects; Procedures for the 5G System (5GS) ; Stage 2*, 2021. Rel. 17.0.

[11] "3GPP". *"3GPP TS 29.281 V17.1.0;General Packet Radio System (GPRS) Tunnelling Protocol User Plane (GTPv1-U)*, September 2021.

[12] 3GPP. *3GPP TS 33.501 version 17.2.1 Release 16,Security architecture and procedures for 5G system (Release 17)*, 2021.

[13] 3GPP. *3GPP TS 38.410 V16.3.0 (2020-09),Technical Specification GroupRadio Access Network; NG-RAN; NG general aspects and principles (Release 16) ; NG-RAN; Architecture description (Release 16); NG-RAN;F1 general aspects and principles ; Stage 2*, 2021. Rel. 16.0.

[14] 3GPP. *3GPP TS 38.420 V16.0.0 (2020-07),Technical Specification Group Radio Access Network; NG-RAN; Xn general aspects and principles (Release 16)*, 2021. Rel. 16.0.

[15] 3GPP. *3GPP TS 38.460 V16.2.0 (2021-01),Technical Specification Group Radio Access Network; NG-RAN; E1 general aspects and principles (Release 16)*, 2021. Rel. 16.0.

[16] 3GPP. *3GPP TS 38.470 V16.5.0 (2021-07),Technical Specification Group Radio Access Network; NG-RAN;F1 general aspects and principles (Release 16) ; Stage 2*, 2021. Rel. 16.0.

[17] 3GPP. *55G System; Public Land Mobile Network (PLMN) Interconnection; Stage 3 (3GPP TS 29.573 version 16.7.0 Release 16); Procedures for the 5G System (5GS) ; Stage 2*, 2021. Rel. 16.7.0.

[18] a10networks. GTP firewall in 4G and 5G mobile networks, 2020.

[19] Patrik Persson Ali Zaidi Afif Osseiran, Stefan Parkvall. 5G wireless access: An overview. Technical report, Ericsson.

[20] Charu C Aggarwal and Philip S Yu. A survey of randomization methods for privacy-preserving data mining. *Privacy-preserving data mining: models and algorithms*, pages 137–156, 2008.

[21] Alessio Casati Alex Shneyderman. *Mobile VPN: Delivering Advanced Services in Next Generation Wireless Systems*. John Wiley Sons, 2003.

[22] G Americas. The evolution of security in 5G. *5G Americas Whitepaper, Tech. Rep., Oct*, 2018.

[23] Eneko Atxutegi, Fidel Liberal, Habtegebreil Kassaye Haile, Karl-Johan Grinnemo, Anna Brunstrom, and Ake Arvidsson. On the use of tcp bbr in cellular networks. *IEEE Communications Magazine*, 56(3):172–179, 2018.

[24] David Beckett and Sakir Sezer. Http/2 cannon: Experimental analysis on http/1 and http/2 request flood ddos attacks. In *2017 Seventh International Conference on Emerging Security Technologies (EST)*, pages 108–113. IEEE, 2017.

[25] Bruno Carvalho. Why only have the GI-FW and GTP inspection isn't enough for 5G security?, 2021.

[26] Louma Chaddad, Ali Chehab, Imad H. Elhajj, and Ayman Kayssi. Mobile traffic anonymization through probabilistic distribution. In *2019 22nd Conference on Innovation in Clouds, Internet and Networks and Workshops (ICIN)*, pages 242–248, 2019.

[27] Whai-En Chen and Chia Hung Liu. High-performance user plane function (upf) for the next generation core networks. *IET Networks*, 9(6):284–289, 2020.

[28] Catalin Cimpanu. Cyberattack brings down vodafone portugal mobile, voice, and tv services, 2022.

[29] Ismail Shakil Divya Rajagopal. Rogers network resuming after major outage hits millions of canadians, 2022.

[30] Patrick Donegan. 5G Roaming Drives Security by Redesign. Technical report, Hardenstance, 02 2022.

[31] Patrick Donegan. GTP vulnerabilities: A cause for concern in 5G and LTE networks . Technical report, SecurityGen, 09 2023.

[32] eCPRI. *Common Public Radio Interface: eCPRI Interface Specification*, 2019-10.

[33] W Eddy. Rfc 9293: Transmission control protocol (tcp), 2022.

[34] Nic Fildes George Steer. Millions of t-mobile customers' data leaked in 'sophisticated cyber attack', 2021.

[35] Vasileios Gkioulos, Stephen D Wolthusen, and AJNJ Iossifides. A survey on the security vulnerabilities of cellular communication systems (gsm-umts-lte). In *Proc. Norweigian Inf. Security Conf.(NISK)*, pages 1–12, 2016.

[36] GSMA. *FS.37 GTP-U Security version 3.1.*

[37] GSMA. *5G Security Issues*, 1 2019.

[38] GSMA. *Guidelines for IPX Provider networks (Previously Inter-Service Provider IP Backbone Guidelines)*, 05 2021.

[39] GSMA. *5GS Roaming Guidelines, Version 6.0*, 05 2022.

[40] Dick Hardt. Rfc 6749: The oauth 2.0 authorization framework, 2012.

[41] ITU-T. *Transport network support of IMT-2020/5G*, 2 2018.

[42] Michael Jones and Dick Hardt. Rfc 6750: The oauth 2.0 authorization framework: Bearer token usage, 2012.

[43] Charlie Kaufman, Paul Hoffman, Yoav Nir, Parsi Eronen, and T Kivinen. Rfc 7296: Internet key exchange protocol version 2 (ikev2), 2014.

[44] Abdul Kayum. 5G TRANSPORT REQUIREMENT; (A Guiding Tool for Planning 5G Transport Network). Technical report, Telecommunications Engineering Centre, Ministry of Communications and Information Technology, 2020.

[45] Geir M Køien. On threats to the 5G service-based architecture. *Wireless Personal Communications*, 119(1):97–116, 2021.

[46] D. Koukis, Spiros Antonatos, Demetris Antoniades, Evangelos Markatos, and P. Trimintzios. A generic anonymization framework for network traffic. volume 5, pages 2302 – 2309, 07 2006.

[47] Charles M Kozierok. *The TCP/IP guide: a comprehensive, illustrated Internet protocols reference*. No Starch Press, 2005.

[48] Chansook Lim. Improving congestion control of tcp for constrained iot networks. *Sensors*, 20(17):4774, 2020.

[49] Ying-Dar Lin, Po-Ching Lin, Sheng-Hao Wang, I. Wei Chen, and Yuan-Cheng Lai. PCAPLib: A System of Extracting, Classifying, and Anonymizing Real Packet Traces. *IEEE Systems Journal*, 10(2):520–531, June 2016.

[50] Yvonne Gootzen May Offermans. Pilot study: Mobile phone meta data records – introduction to the research method, Feb 2021.

[51] KAZUAKI NAGATA. Kddi to spend ¥7.3 billion to compensate users for major network outage, 2022.

[52] Positive Technologies. *Threat vector: GTP*, 2020.

[53] Eric Rescorla and Nagendra Modadugu. Rfc 4347: Datagram transport layer security, 2006.

[54] RoamsysNext. 11 ways that mobile network operators pay for telecom fraud, 2020.

[55] Youngkwon Park Dowon Kim Eunsoen Jeong Seongmin Park, Bomin Choi. Vestiges of past generation: Threats to 5G core network. In *Innovative Mobile and Internet Services in Ubiquitous Computing*, 2021.

[56] Youngkwon Park Hyungjin Cho Dowon Kim Sungmoon Kwon Seongmin Park, Daeun Kim. 5G security threat assessment in real networks. In *Sensors 2021*, 2021.

[57] Mostafa Taha Member Silvere Mavoungou, Georges Kaddoum and Georges Matar. Survey on threats and attacks on mobile networks. In *IEEE Access Volume-4, Pages 4543 - 4572*, 2016.

[58] Magnus Olsson Lars Frid Shabnam Sultana Catherine Mulligan Stefan Rommer, Peter Hedman. *5G Core Network*. Elsevier, 2019.

[59] Techplayon. 5G core, https://www.techplayon.com/.

[60] Tiffanie Turnbull. Optus: How a massive data breach has exposed australia, 2022.

[61] Joni Tyagi. 5G protocol stack, 2021.

[62] Pieter Veenstra. Reasons for the gsma to reconsider the solutions for 5G roaming, 2020.

[63] Charles V Wright, Scott E Coull, and Fabian Monrose. Traffic morphing: An efficient defense against statistical traffic analysis. In *NDSS*, volume 9. Citeseer, 2009.

[64] Dongzhu Xu, Anfu Zhou, Xinyu Zhang, Guixian Wang, Xi Liu, Congkai An, Yiming Shi, Liang Liu, and Huadong Ma. Understanding operational 5G: A first measurement study on its coverage, performance and energy consumption. In *Proceedings of the Annual conference of the ACM Special Interest Group on Data Communication on the applications, technologies, architectures, and protocols for computer communication*, pages 479–494, 2020.

[65] Yingyou Wen Xuena Peng and Hong Zhao. Security issues and solutions in 3G core network. In *JOURNAL OF NETWORKS, VOL. 6, NO. 5*, 2011.

[66] Kaiyue Zeng, Wei Deng, Rui Wang, Long Zhang, Jinxia Cheng, Tao Chen, and Na Yi. 5G network performance evaluation and deployment recommendation under factory environment. In *2021 IEEE 32nd Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, pages 1370–1375. IEEE, 2021.

[67] Shuang Zhao, Shuhui Chen, Yipin Sun, Zhiping Cai, and Jinshu Su. Identifying known and unknown mobile application traffic using a multilevel classifier. *Security and Communication*, 2019.

Appendix A

Appendix Section

A.1 Evolution of Mobile Network Generations

A new generation of mobile networks emerges every ten years roughly. Below are the evolution stages of mobile network generations starting from 2G:

- 2G Architecture The Global System for Mobile Communication was the widely deployed 2G standard, developed by ETSI. The Mobile stations contain a subscriber identity module (SIM) and a wireless mobile device. The mobile device communicates with the Base Transceiver Station (BTS) over the air interface with Um as the reference point as shown in Figure A.1. Each of these BTS is controlled by the Base Station Controller (BSC) which is further controlled by the Mobile Switching Centre (MSC). The MSC is a part of the core network and it is the main switching entity, that deals with roaming, handovers and authentication. The Home Location Register (HLR) contains all the necessary subscription data and user profiles. The entire network works on circuit switching. It provides circuit switching-oriented services. The architecture was simple, and limited to speech telephony, SMS, and supplementary services such as call forwarding, call hold, and call waiting.
  The General Packet Radio Service (GPRS) standard was defined as the 2.5G. The GPRS provides internet connectivity, which can be seen as a shift from circuit-switched towards packet-switched networks. The Serving GPRS support node (SGSN) and Gateway GPRS support node (GGSN) provide internet connectivity to the user.

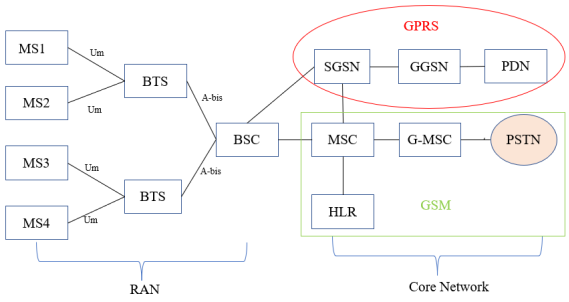


Figure A.1: 2G and 2.5G Network Architecture

- 3G Architecture The third generation of mobile communication is also known as the Universal Mobile Telecommunication System (UMTS). It switched faster data throughput, higher peak rates, lower latency, and better spectrum efficiency when compared to GPRS. Both circuit and packet-switched services can be used. The MS was renamed to User Equipment (UE). The UEs are connected to the NodeB. The NodeBs are further controlled by RNC. The RNCs manage radio resources and are connected to the core network. The NodeB handles power management.

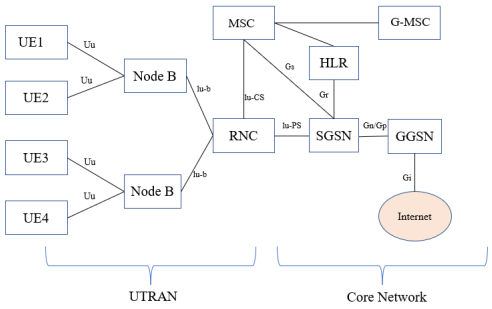


Figure A.2: 3G Network Architecture

- 4G Architecture The UE is connected to the eNodeB and the eNodeB is connected to the MME (which is a part of the core network). The 4G RAN has a flat architecture as BSC/RNC nodes are not present. As a result, the latency gets reduced. The eNodeB is capable of handling radio network control functions and it is split up in BBU and multiple RRU. The fronthaul network, which was introduced in LTE, connects the RRU and BBU. The different network entities are explained below:
  - User Equipment (UE): The UE is a device that the user carries. The Uu interface connecting the UE to the e-NodeB is the LTE version of the air interface.
  - Enhanced-NodeB (e-NodeB/eNB): This is the 4G base station and provides the user with radio interfaces and performs Radio Resource Management, header compression, security modulation, interleaving, handover, and re-transmission control. eNBs are connected through an interface called the X2 interface. There are two S1 reference points: S1-MME and S1-U. S1-MME carries all the control plane signaling information. S1-U carries the payload (data plane) information.
  - Mobility Management Entity (MME): E-NodeBs connect to the MME. MME is the mobility server and acts as a registrar for access to the core network. It operates in the control plane and is responsible for sending data to the eNBs. MME also helps with bearer establishment procedures. It also performs user profile authentication in association with the Home Subscriber System (HSS).
  - Serving-Gateway (S-Gw): This acts as the terminating interface towards the E-UTRAN. The control plane signaling traverses through the S1-MME reference point from the eNB to the MME. The user plane payload traverses from the eNB to the S-Gw over the S1-U reference point by passing the MME. The S-Gw serves

as a local mobility anchor point of data connections for inter-eNB handover and handles user data tunnels between the eNBs and PDN-Gw which is explained next.

– Packet Data Network-Gateway (PDN-Gw/P-Gw): A PDN refers to the network that is established and operated to provide data transmission services to the public (like the Internet). P-Gw provides UE with access to a PDN by assigning an IP address from the address space of the PDN. It serves as the mobility anchor point for handover between 3GPP and non-3GPP and also performs policy enforcement, packet filtering, and charging based on the rules set by the policy and charging functions. The PDN-Gw acts as the gateway router to the internet.

– Home Subscriber System (HSS): It is the central database that stores user profiles, and provides user authentication information and user profiles to the MME. It is similar to the HLR in the GSM architecture. HSS also keeps track of the location information of the subscribers allowing them to connect to the nearest eNB.

– Policy Charging Control Function (PCRF): This node determines the policy rules for the data bearers. It supports service data flow detection, policy enforcement and flow-based charging in a centralized manner

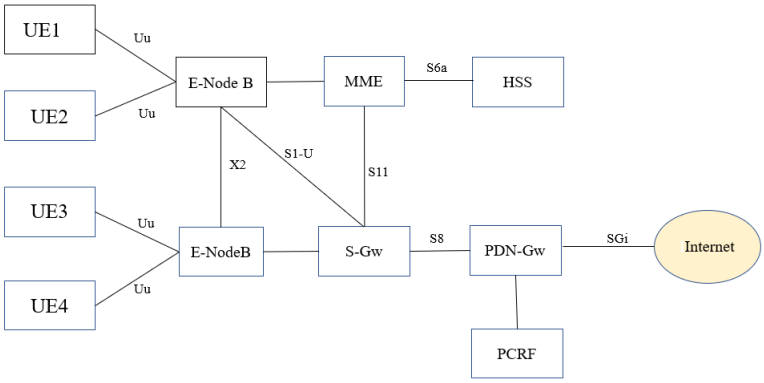


Figure A.3: 4G Network Architecture

A.2 RAN Deployment Options

The RAN can be deployed as a standalone, which means only one kind of radio access technology will be used to communicate with the EPC core/5G core network. Hence, the radio access network can "stand-alone (SA)" on its own and does not need any other radio access technology to function. However, RAN can also be deployed as a non-standalone (NSA). The NSA deployment is ideal for operators who wish to upgrade their current LTE network to benefit from 5G data speeds. The core can be either EPC or 5G core together with eNodeB and gNodeB in the radio access network forming the NSA architecture. The various standalone and non-standalone options are described below in detail:

A.2.1 Standalone Options

- Option 1: EPC network
  This option is the current 4G network that operators are using. The LTE RAN is connected to a 4G core (EPC) as shown in Figure A.4. There are no 5G components. These networks will be slowly upgraded to 5G networks.

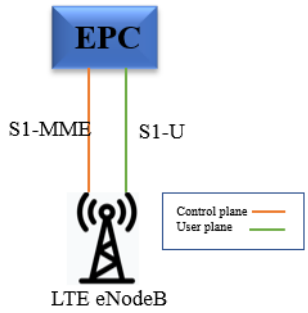


Figure A.4: Standalone option 1: LTE network

- Option 2: 5G NR with 5G Core
  This option shows a 5G core and 5G radio access network as in Figure A.5. There is no existing radio access technology in the network. It is considered the holy grail as it can support all the 5G use cases. Initially, the 5G will be deployed as a non-standalone. After the initial deployment, migrating to this option will be every operator's preference.

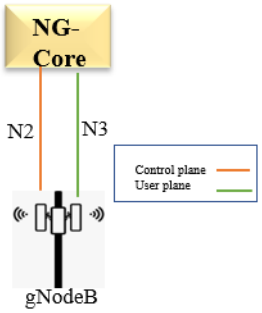


Figure A.5: Standalone option 2: 5G Network

- Option 5: 5G core with LTE RAN
  The 5G core is deployed with LTE RAN as shown in Figure A.6. The eNodeB has to be upgraded to ng-eNodeB for compatibility with the core. These upgraded eNodeBs will not have all the functionalities of gNodeB. Thus, this deployment option does not serve all the 5G RAN-related advantages.

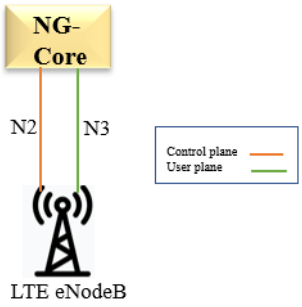


Figure A.6: Standalone option 5: 5G core with LTE network

A.2.2 Non-Standalone Options

- Option 3: eNodeB as Master Node (MN) - It is ideal for those cases in which operators want RAN benefits without deploying the 5G core. The eNodeB serves as the master node (MN) which means that all the signalling-related data from the EPC must go through it. The Xn interface connects gNodeB and eNodeB while the S interface connects EPC with eNodeB. The gNodeB is not connected to EPC. The user data can either go through Path 1 or directly go through Path 2. Path 1 provides 5G speed as the data passes through gNodeB. The traffic is split at the eNodeB by the PDCP sublayer using the Master Cell group (MCG) split bearer. The S1-U will require higher bandwidth compared to the Xn and Uu interface as the traffic will be split at the eNodeB.

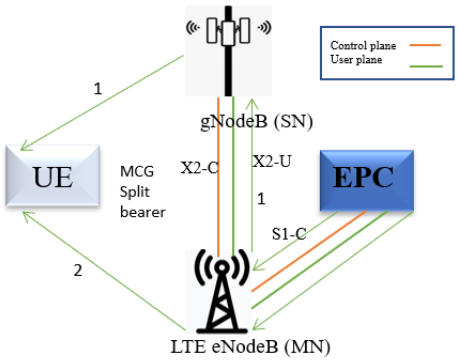


Figure A.7: Non-Standalone option 3: EPC core with LTE and gNodeB

- Option 3a: eNodeB as MN with split data - In option 3a, EPC is directly connected to gNodeB through the S1-U interface. The control plane traffic runs between eNodeb and gNodeB through the X2 interface. In this option, there are no split bearers, only SCG and MCG bearers. The user data can go through Path 1 and Path 2 after the EPC splits it.

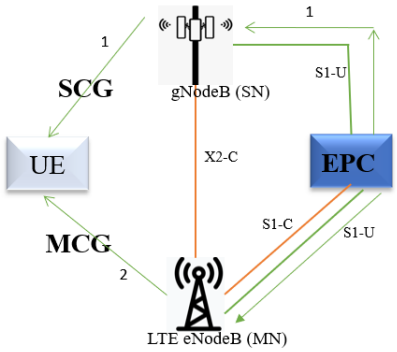


Figure A.8: NON-Standalone option 3a: EPC core with LTE and gNodeB

- Option 3x: eNodeB as MN with separate user and control plane paths This option is heavily favoured by the operators as it can be implemented faster as very little modification needs to be done to the existing network. The EPC splits the control and user plane data. The user plane data runs through the gNodeB and control-related signalling goes through eNodeB. When the traffic is on a higher side, the path between the gNodeB and eNode B can be used to lower the load as a fraction of the data can be sent through the X interface. EPC splits the traffic for eNB and gNB. The user data goes through gNodeB while control-related signalling goes through the eNodeB. The X2 interface which is present between eNodeB and gNodeB can be used for carrying the data traffic.

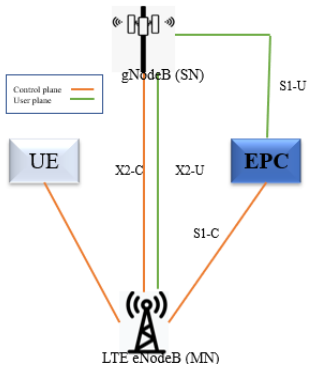


Figure A.9: Non-Standalone option 3x: EPC core with LTE and gNodeB

- Option 4: 5G core with LTE network
  This option provides dual connectivity since eNodeB and gNodeB are both connected to the 5G core. The gNodeB serves as the master node (MN) while eNodeB is the secondary node (SN). In this case, the control plane signalling has to go through the gNodeB which acts as a signalling anchor. The user data can be split by the core. Both technologies are used in the access network along with the 5G core. This will give better coverage and data rates to the users. The eNodeB in this case has to be upgraded for compatibility with the 5G core. The N2 interface is used between the

eNodeB and the 5G core. The user traffic goes through the N3 interface while control signalling goes through N2.

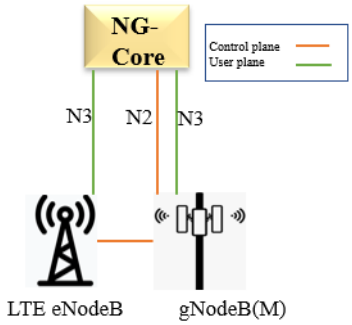


Figure A.10: Non-Standalone option 4: 5G core with LTE network

- Option 7: 5G core with eNodeB and gNodeB
  This option is similar to option 3 with the only difference being the deployment of 5G core. The upgraded eNodeB will be used and it will serve as the master node while gnodeb will act as the secondary node. The control plane signaling goes through eNodeB. The user data will go to gNodeB through eNodeB. The eNodeB splits the user traffic and part of it can follow path 1 while the remaining user data can go through path 2. The dual connectivity along with next next-generation core will provide good coverage and bandwidth.

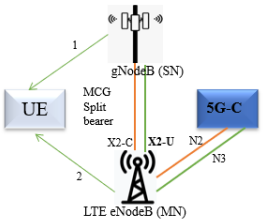


Figure A.11: Non-Standalone option 7: 5G core with LTE network

- Option 7a: eNodeB as MN with split data path
  This option is similar to option 3a. There are no split bearers in place. The gnodeB and eNodeB are directly connected over the X interface. The core splits the user data and controls messages over N3 and N2 interfaces. Part of the data can follow path 1 while the rest of the user data can follow path 2. This reduces the load on both interfaces.

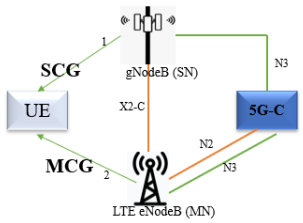


Figure A.12: Standalone option 7a: 5G core with LTE network

- Option 7x: eNodeB as MN with separate user and control plane paths
  This option is similar to option 3x. The SCG split bearer will be implemented. The eNodeB is the master node while gNodeB is the secondary node. The fraction of user plane data goes through gNodeB and the rest of it can go through the eNodeB. The user plane data can also take only one path depending on the network load.

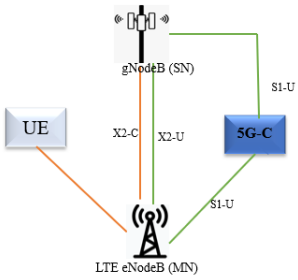


Figure A.13: Standalone option 7x: 5G core with LTE network