Mobile Network Protocols from 2G to 5G – Evolution and Lessons Learned.

1st Anita Francis Archibong
27729790
a_rchibo@live.concordia.ca

2nd Valentine Ozonyia
40202470
v_ozonyi@live.concordia.ca

3rd Josephine Famiyeh
40262544
j_famiye@live.concordia.ca

4th Elvis Okoye
40274904
em_okoye@live.concordia.ca

5th Tweneboath Kodua
40227652
k_anyimadu@live.concordia.ca

6th Ugochukwu Kizito Ugwu
40244315
u_ugwu@live.concordia.ca

7th Ihekweazu Samuel
40265794
s_ihekwe@live.concordia.ca

***Abstract*— Mobile network protocols are the backbone of wireless communication, enabling seamless connectivity and delivering services to users. From the foundational GSM technology to the cutting-edge 5G networks, these protocols govern how mobile devices connect, transmit data, and interact with network infrastructure, all while ensuring security and reliability. This paper provides a comprehensive review of the evolution of mobile generations, highlighting differences in types, data transmission rates, challenges, techniques, features, and applications. Furthermore, it examines the SS7 protocol, exploring its significance, associated security vulnerabilities, and implementation strategies within modern mobile networks.**

Keywords— 1G, 2G, 3G, 4G, 5G, 5G architecture, SS7

I. INTRODUCTION

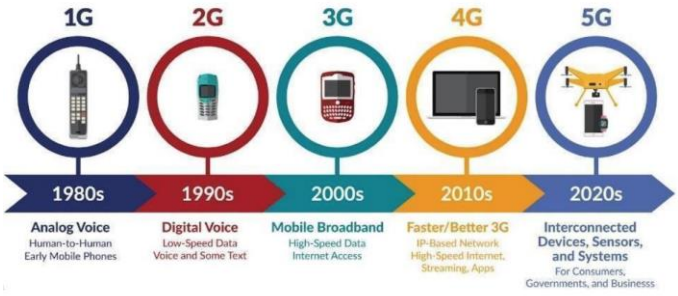
WHAT distinguishes humans from animals is our unique ability to communicate. The history of communication traces back to early humans who gathered around fires, akin to today's social networking on platforms like Facebook. As civilization progressed, cave inscriptions evolved into modern-day blogging. However, this primitive form of communication was confined to localized areas, prompting the need for long-distance communication as people ventured beyond caves.

Smoke signals marked the advent of long-distance communication, employed by Native Americans in the early 19th century, each tribe with its own signalling system [1]. Similarly, ancient China utilized smoke signals to convey messages across towers, covering distances of up to 750 km in a matter of hours. Even today, smoke signals persist in certain contexts, such as the selection of a new pope by Catholic activists. Pigeons emerged as another means of long-distance communication due to their natural roaming ability. In the 19th century, pigeons transmitted star quotations between cities. The Pony Express, a mid-19th-century courier service, relied on human messengers on horseback to relay messages, mail, and newspapers between the east and west coasts of America. Semaphore flags, a telegraphy system, were later employed for visual communication using handheld flags, rods, and disks, still finding application in emergency communication and underway replenishment at sea.

The last two centuries hold particular significance in the evolution of communication technology. While ancient cultures, such as the Greeks, Romans, and Chinese, conducted random experiments that hinted at the relationship between electricity and magnetism, the intentional exploration of this connection occurred in the 19th century. Danish physicist Hans Christian Orsted's discovery in 1820 laid the foundation, demonstrating that a current-carrying wire influenced a compass needle.
Further developments ensued with Dominique Francois Jean Arago's discovery of rotary magnetism (Arago's rotation) and Andre-Marie Ampere's contributions to electromagnetism. Michael Faraday's experiments on electromagnetic induction paved the way for wireless signal transmission. He predicted the existence of electromagnetic waves, and his work inspired Samuel Finley Breese Morse to create the electric telegraph, marking the first use of electromagnetism for communication [1].

James Maxwell's electromagnetic theory and the discovery of radio waves in 1895 played a crucial role [1]. Marconi's demonstrations in 1920 marked the advent of wireless communication, leading to the first commercial radio broadcast in 1920. Subsequent years witnessed significant milestones, including the use of satellites for communication proposed by John R. Pierce in 1955 and the launch of Sputnik I by the Soviet Union in 1957 [1]. World War II spurred advancements in radio technology, and the post-war era saw innovations like the Carter Fone in 1968, connecting two-way radios to the telephone system. The continuous evolution of wireless communication has shaped our modern world, with each discovery and invention building upon the achievements of the past [1].

II. Evolution of Cellular Networks



A. *First Generation Systems (1G)*

The advent of frequency modulation (F.M.) around 1930 was a pivotal development in communications, significantly aiding military operations during World War II. F.M.s combine an audio signal with a carrier frequency, which improves signal robustness and range. Despite its advantages, the early mobile phone systems in metropolitan areas were capacity-constrained, prompting AT&T's creation of the Advanced Mobile Phone Service (AMPS) by AT&T in 1983. As the inaugural analog cellular network in the U.S., AMPS utilized the 800-900 MHz frequency range and segmented this into 30 kHz channels via frequency division multiple access (FDMA).

In parallel, Europe made strides with Nordic Mobile Telephony (NMT) and NMT900, operating in 450 MHz and 900 MHz bands. NMT, a pioneer in automated mobile networks, enabled long-distance calling and incorporated F.M. with Fast Frequency Shift Keying (FFSK) modulation, transmitting at 600-1200 bits per second and allowing for channel multiplexing. The U.K.'s contribution was the Total Access Communications System (TACS), a 900 MHz band system and an adaptation of AMPS. AMPS, NMT, and TACS, laid the groundwork for the first generation of communication technologies, introducing foundational principles such as frequency reuse, mobile subscriber management, and seamless hand-offs.

General Architecture of 1G Systems

In discussions about cellular communication, the focus often turns to either AMPS or TACS [11], representing prominent elements of the first-generation (1G) system architecture.

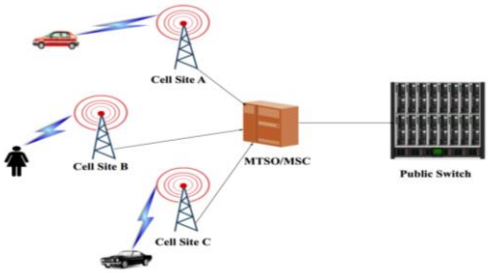


Figure 1 General 1G System Architecture [11]

As illustrated in Figure 1, the architecture encompasses various high-level system blocks within the cellular network. This framework involves radio transmissions occurring between the mobile device and the base station within each cell. The radio transmissions operate in a full-duplex configuration, where distinct frequency bands are allocated for transmitting and receiving signals. A cell site serves as the intermediary between the mobile device and the Mobile Telephone System Office (MTSO), conveying signals from the mobile device to the MTSO through either T1/E1 lines or a microwave system. On the end-user side, the Mobile Subscriber Unit (MSU) comprises a control unit and transceiver responsible for transmitting to and receiving from the cell site [11].

Additionally, according to [1], the MTSO functions as the central processing unit of the network, managing calls and establishing connections between cell site radio links and the Public Service Telephone Network (PSTN). The MTSO maintains comprehensive call records and subscriber statuses, including call routing and billing information. All traffic between the cellular network and PSTN or other networks traverses through the MTSO via landline cable connections, with the MTSO converting the received energy from base stations into another medium. The architecture of the MTSO includes components such as the Mobile Switching Centre (MSC), field monitoring stations, and relay stations. The MSC's primary role is to route mobile phone calls, while the MTSO facilitates the integration of mobile telephones with the land telephone network, offering services like direct dialed mobile-to-mobile, mobile-to-land, and land-to-mobile calling.

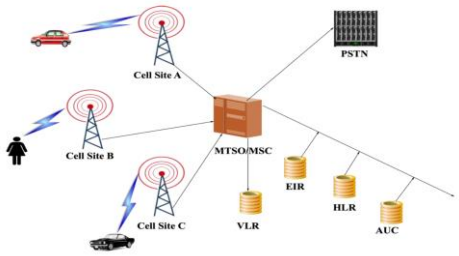


Figure 2 Components in MTSO/MSC [11]

The MTSO in Figure 2 is central to cellular network operations, incorporating critical databases such as the HLR for subscriber information and the VLR for tracking roaming users. It also hosts the EIR to record and block phones and the AUC for user authentication and communication encryption.

Parallelly, the PSTN, which has evolved since Bell's invention, comprises transmission networks using multiplexing and optical fibres or coaxial cables and switching networks that facilitate circuit-switched calls. Signalling networks manage analog signal transitions and resource allocation, with the PSTN acting as a gateway routing calls locally or across area codes [11].

Figure 3 Cell Site Configuration Overview [11]

As outlined in [11], cellular networks consist of hexagonal cells, each with a cell site equipped with antennas and electronics for two-way communication, distributed to cover areas without overlap. Operators use 25MHz of spectrum, divided equally for transmission and reception.

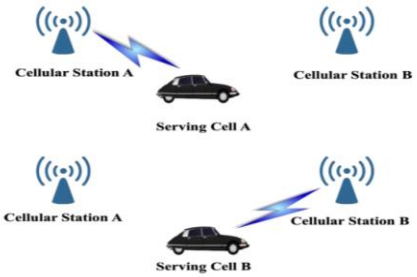


Figure 4 Analog Handoff [2]

A key attribute of 1G technology is the implementation of Handoff; the handoff process in 1G systems ensures continuous calls when moving between cells by transferring the connection to adjacent base stations at specific RF signal levels. An additional crucial feature is frequency reuse, which aims to increase capacity by managing the Carrier-to-Interference (C/I) ratio, optimizing frequency use, and minimizing interference within the network's geographic area.

B. Second Generation Systems (2G)

Following the inception of the first generation, a pivotal transition occurred with the introduction of digital radio technology in 2G systems. This evolution involved utilizing various modulation formats to enhance the quality and capacity of existing cellular networks. During the 2G era, one of the primary services offered was mobile fax, leveraging a transmission speed of 9.6 kbps for efficient information content transport.

In 2G systems, digital radio technology was used in Cellular, Personal Communication Services (PCS), and Specialized Mobile Radio (SMR), all designed to optimize voice traffic throughput. This technological shift marked a significant departure from the analog methods employed in 1G, introducing digital advancements that facilitated more transparent communication and increased capacity for data transfer.

For a comprehensive understanding, Figure 5 [11] visually highlights the critical differences between the first-generation and second-generation cellular systems. This visual aid helps discern the technological advancements that paved the way for improved communication capabilities during the transition from 1G to 2G.

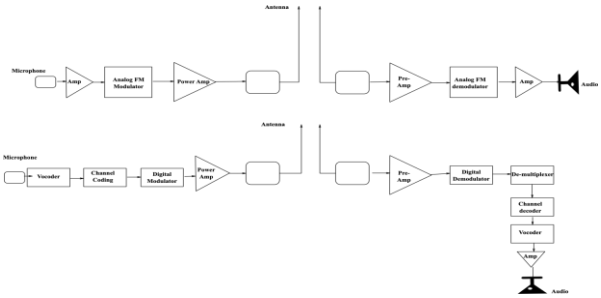


Figure 5 Analog and Digital Radio [11]

Voice signals are digitized by a vocoder using Pulse Code Modulation, which involves sampling, quantizing, compression, and encryption. This stream is then channel-encoded to add security and efficiency. Digital modulation transfers this data onto an RF wave using ASK, FSK, or PSK, which is then filtered and emitted by an antenna [11].

Received signals are filtered, amplified, and demodulated back into a digital stream, decoded, and converted into an analog signal for the listener. Digital technology is favoured for its security, efficiency in bandwidth usage and long-distance transmission quality. Advantages of 2G include higher capacity, lower costs, reduced fraud, and enhanced features with encryption for security, despite integration challenges with existing systems [11]. Noteworthy advantages of the 2G architecture encompass increased capacity compared to analog, reduced capital infrastructure costs, lower per capita subscriber costs, minimized cellular fraud, improved features, and the implementation of encryption measures for heightened security. Nevertheless, while implementing the 2G architecture, individuals encountered challenges seamlessly integrating this new framework with existing legacy systems.

C. Third Generation Systems (3G)

According to the ITU specification, 3G, also known as IMT-2000, has been recognized as a pivotal advancement in high-speed wireless data for mobility. IMT-2000 serves as a comprehensive radio and network access standard, addressing the challenges of mobile and high-speed data services across previous generations [1].

IMT-2000 sets specific data speed requirements, including 144 Kbps for driving speeds, 384 Kbps for stationary or outdoor walking speeds, and 2 Mbps for indoor scenarios. Despite the predominance of voice in the 3G landscape, Short Message Service (SMS) emerged as a significant packet data service. Wireless operators faced critical decisions during the transition from 1G/2G, choosing transition methods to support the diverse platforms within IMT-2000. Understanding the concept of 2.5G is crucial in this transitional phase, serving as a vital link between established 2G and envisioned 3G platforms.

The decision-making process regarding platform selection involves speculation and choices based on the belief that specific platforms will facilitate future services. 2.5G,

encompassing technologies like GPRS/HSCSD, EDGE, and CDMA 2000 (phase 1), operates as a platform-independent, data-centric technology, offering high-speed data services (144.4 k) over 2G. The primary goal of 2.5G was to bridge existing 1G or 2G platforms with those of 3G. Operators were tasked with selecting migration paths based on resources, capital, spectrum availability, and workforce. However, a commonality persisted between the two generations: deploying a packet-based data network, regardless of the chosen platform (GPRS/EDGE/CDMA).

Figure 6 below presents a table outlining the comparative advantages of each 2.5G platform about its underlying foundational technology platform.

2G Technology	2.5G Technology	Enhancements	Migration-to-3G Platform
GSM	GPRS	• High speed packet data services (144.4K)	WCDMA
		• Uses existing radio spectrum	
IS-136	EDGE	• High speed packet data services (144.4K)	WCDMA
		• Uses existing radio spectrum	
CDMA	CDMA2000 (phase1)	• High speed packet data services (144.4K)	CDMA2000 –MC multi carrier
		• Uses existing radio spectrum	
		• 1XRTT used	

Figure 6 2G and 2.5G [11]

*D. **Fourth Generation Systems (4G)***

The proliferation of mobile device applications prompted a rise in internet use, enhancing mobile broadband infrastructure. The IEEE introduced the 802.16 standard to cater to Wireless Metropolitan Area Networks, which was later adapted for mobile use. WiMAX, supporting IEEE 802.16 standards, emerged to certify and ensure the interoperability of wireless systems, focusing on IP protocols and VoIP [1].

Simultaneously, LTE was developed, utilizing OFDM and OFDMA technologies and known as evolved UTRA in 3GPP standards. This marked a move beyond 3G systems, with 3GPP2's counterpart being UMB [1]. LTE supports packet-switched traffic with services like mobility and QoS, facilitated by a notable structural change: the removal of the RNC and integration of its functions into the eNB. This change reduces latency, as the eNB manages nodes directly, and LTE's all-IP interface architecture enhances connectivity. eNBs connect via the X2 interface and the MME/GW through the S1 interface. The MME/GW comprises the S-GW, which handles UE mobility, and the P-GW, which acts as the point of access to external networks, managing IP assignments, policy, and routing within the LTE network.

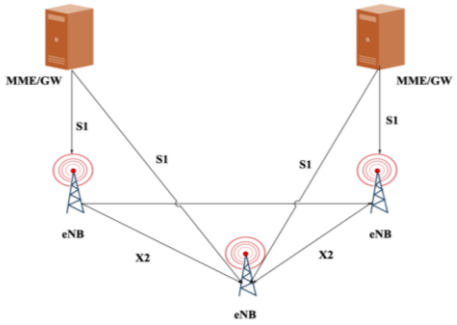


Figure 7 4G Network Architecture [12]

The Evolved Node-B (eNB) in LTE networks merges the functionality of 3G's Node B with the RNC's protocols. It manages header compression, ciphering, data transmission reliability, radio resources, and admission control [12]. The Mobility Management Entity (MME) oversees signalling processes, including tracking user equipment (UE), paging, roaming, authentication, security, and selecting and managing gateways and bearers.

In LTE's user plane, the PDCP and RLC, previously ending at the RNC, now terminate at the eNB, enhancing efficiency [12]. The RRC, formerly part of the RNC, is embedded within the eNB, handling system information, connection management, mobility, and measurement reporting. This integration simplifies the control plane and bolsters network responsiveness.

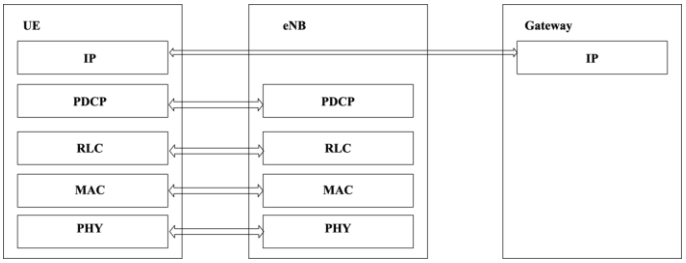


Figure 8 User Plane

In LTE architecture, the non-access stratum (NAS) protocol is crucial for network operations. It ends at the MME on the network side and at the UE itself on the terminal side. The NAS handles tasks like EPS bearer management, authentication, and security control. As mentioned in [12], LTE's Quality of Service (QoS) is vital for supporting diverse services like VoIP and video streaming. LTE's QoS differentiates packet flows to meet varied service needs. QoS flows, or EPS bearers, are set up between the UE and the P-GW, and each IP flow is assigned a unique EPS bearer for traffic prioritization.

The P-GW classifies and directs incoming IP packets to the correct EPS bearer, which the eNB maps to the associated radio QoS bearer. This ensures consistent one-to-one mapping between EPS and radio bearers, which is critical for maintaining service quality.

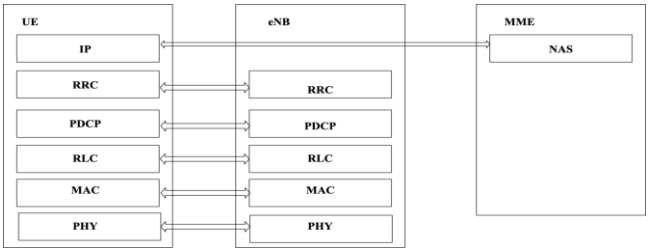


Figure 9 Control Plane

E. Fifth Generations Systems (5G)

Mobile cellular systems have evolved over three decades from analog to sophisticated packet-based 4G systems, leading to speed, bandwidth, and connectivity gains. With increasing demands from sectors like agriculture, health, and transport, 4G faced connectivity and service diversity challenges, paving the way for 5G. As outlined in [1], 5G is envisioned as a comprehensive ecosystem, enhancing value creation across industries with seamless and consistent experiences supported by robust business models.

5G exceeds the capabilities of previous systems, offering new services from various perspectives. Architecturally, it combines non-standalone (leveraging existing LTE) and standalone systems, including New Radio (NR) and WLAN technologies. The spectrum needs of 5G are addressed using bands up to 86GHz, approved by WRC-15, with additional sub-6GHz frequencies allocated for mobile use. From the user perspective, 5G ensures ubiquitous connectivity, high data rates, low latency, and superior QoS.

Incorporating cloud computing, SDN, and NFV, 5G's programmable, software-driven architecture supports emerging applications such as 3D video and autonomous driving, showcasing its broad potential across various industries and use cases [1].

III. Security Protocols of Mobile Networks in Different Stages

A. 1G Security Protocols & Architecture

Frequency Division Multiple Access (FDMA)

In [18], FDMA is a prevalent multiple-access technique employed in 1G, particularly for Analog communication. This method entails the division of frequency bandwidth or a channel into multiple individual channels, each with a specific bandwidth, allowing each conversation to occur on a distinct frequency. Guard bands are maintained between each frequency band segment to prevent crosstalk. In FDMA, a central controller allocates frequency bands to users based on their specific requirements. Once a band is allocated to a user, it remains dedicated until the entire flow of information is completed. Techniques like frequency dividers, such as flip-flop clock frequency dividers, can be employed to divide frequencies in FDMA.

B. 2G Security Protocols & Architecture

Digital techniques for cellular communication can be broadly categorized into two primary groups: the AMPS and TACS spectrum. The Global System for Mobile Communications (GSM) is the chosen digital modulation technique for markets utilizing TACS spectrum allocation. However, for the AMPS markets, the choice lies between Time Division Multiple Access (TDMA) and Code Division Multiple Access (CDMA) radio access platforms. The Integrated Digital Enhanced Network (IDEN) radio access platform is also available as an alternative to the AMPS/TACS spectrum decision, operating within the SMR band [17].

The radio channel is a communication medium shared by numerous subscribers within a single cell. As mobile stations contend for frequency resources to transmit their information streams, the potential for the multiple access problem arises. Consequently, specific multiple access procedures must be implemented to partition the available frequency band effectively. The ensuing methods delineate the multiple access approaches employed in 2G technology.

Time Division Multiple Access (TDMA)

Time Division Multiple Access (TDMA) operates on the premise of allocating the entire bandwidth of a channel to all users, with each user having limited time for communication. User time slots are organized into frames, as illustrated in the figure, where each slot is designated for an individual user within a frame comprising six slots. These frames repeat at regular intervals of Tf. Guard bands are allocated to mitigate user interference stemming from synchronization time variations. [18].

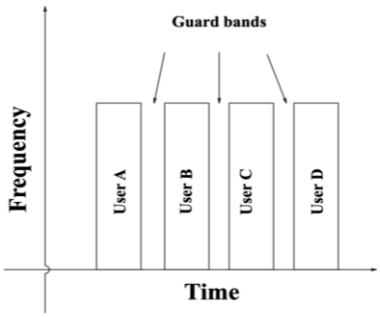


Figure 10 Time Division Multiple Access (TDMA) [18]

Code Division Multiple Access (CDMA)

Code Division Multiple Access (CDMA) enables a single channel to accommodate simultaneous transmissions without collisions. In a scenario with four stations—Station 1, 2, 3, and 4—each station possesses a unique code (C1, C2, C3, and C4). All stations transmit data (D1, D2, D3, and D4), and the channel carries the algebraic sum of the product of codes and data (C1D1, C2D2, C3D3, and C4D4). In [4], on the receiver side, if a user desires to receive data from a specific user, they multiply the algebraic sum with that user's code, then sum it up and divide it by the number of users. This technique effectively delivers the sender's data to the intended receiver.

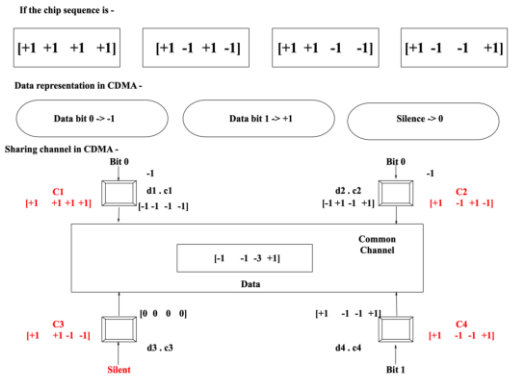


Figure 11 CDMA Technology [18]

Global System for Mobile Communication (GSM)

The GSM network facilitates mobile communications through a well-structured architecture that links mobile stations to the network via Base Transceiver Stations (BTS) and Base Station Controllers (BSC). The Mobile Switching Center (MSC) oversees switching functions and critical operational tasks, including user location registration and handovers. Externally originated calls access the network through the Gateway MSC (GMSC), which interfaces with the Home Location Register (HLR) to determine subscriber locations. The Interworking Function (IWF) ensures seamless protocol mapping between cellular and fixed networks, which is crucial for network interoperability [3][17]. Critical databases such as the HLR and Visited Location Register (VLR) manage user locations and profiles, while the Authentication Center (AUC) and Equipment Identity Register (EIR) enhance security. The network's operational aspects are centralized in the Operation and Maintenance Center (OMC), facilitating efficient management [17].

Figure 17 illustrates the GSM system's hierarchical structure. Each cell group is assigned to a BSC, and at least one BSC is in a Location Area (LA). The GSM infrastructure is categorized into the radio access, core, and management networks. Communication initiates from the mobile station, passes through the BTS and BSC, and is processed by the MSC. This system supports secure data transmission to the intended recipient [17].

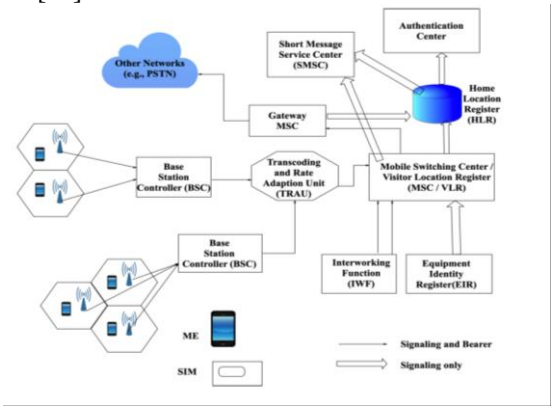


Figure 12 GSM System Architecture

The Subscriber Identity Module (SIM) card enables network access and personalization of the mobile station, supporting features like international roaming. The International Mobile Station Equipment Identity (IMEI) uniquely identifies each mobile device, aiding in security and management. The International Mobile Subscriber Identity (IMSI) stored on the SIM card is essential for accurate subscriber billing. Additional identifiers within the GSM network facilitate the proper functioning and billing of mobile services [17].

Inter-network connections, including those with international networks, are facilitated by the International Switching Center (ISC), which plays a pivotal role in global communication interoperability [3]. The GSM network's comprehensive architecture ensures robust and secure mobile communication, supporting a wide range of services and international connectivity.

SIGTRAN Protocol

The Sigtran protocol is a significant advancement in mobile networks as it enables SS7 signaling over PSTN and IP networks. It utilizes a protocol stack to channel messages from the MSC to the HLR over PSTN, incorporating MTP3, SCCP, TCAP, and MAP for mobile exchanges. Using SCTP at the transport layer, Sigtran integrates SS7 protocols with the TCP/IP stack, enhancing communication over IP and utilizing existing network infrastructures. Protocols like M2PA, M2UA, and M3UA are crucial in adapting SS7 for IP transmission and ensuring messages are properly encapsulated within SCTP packets.

Sigtran is designed to convert SS7 signaling between IP and TDM setups using signaling gateway processes. M3UA, ASP, and NIF are used for this purpose. These protocols enable direct IP connections between network elements but require accommodation for multiple connection types to ensure redundancy and load balancing. Sigtran implementation presents commercial and technical challenges, requiring specialized hardware at Signaling Transfer Points (STPs). However, it offers benefits over traditional TDM signaling, including increased network redundancy, expanded bandwidth, and greater routing flexibility. The evolution of Sigtran aims for cost efficiency, scalability, and simpler deployment to enhance SS7 signaling over IP. These phases range from:

Phase 1: replacing expensive SS7 TDM links with IP

Phase 2: *introducing* IP-based Signal Transfer Points (STPs).

Phase 3: integrating IP-based applications for direct SS7 message handling to supporting traditional switches over IP and merging IP telephony with legacy SS7 networks.

Phase 4: The Sigtran protocol suite, including IP, SCTP, and various adaptation modules like M2UA, M2PA, M3UA, and SUA.

Phase 5: facilitates SS7 protocol emulation and signaling encapsulation for communication over IP networks, catering to diverse operational requirements and scenarios.

C. 3G Security Protocols & Architecture

General Packet Radio Service (GPRS) Architecture

GPRS (General Packet Radio Service) significantly enhanced GSM networks by introducing packet-switched data services, addressing the limitations of the earlier circuit-switched framework, which was capped at speeds of up to 9.6 kbps. This

advancement allowed for data transmission speeds up to 100 kbps, with typical speeds ranging from 40 to 53 kbps, leveraging the same 200-kHz channel and eight timeslot structure as GSM. However, GPRS introduced different channel coding schemes, like Coding Scheme 2 (CS-2), to reduce overhead and improve data transmission reliability through error correction without necessarily increasing bandwidth.

The shift to packet-switched data was a key innovation of GPRS. It optimized the use of RF resources by allowing users to access them only when actively transmitting or receiving data. This change markedly enhanced network resource efficiency, which is particularly beneficial for activities such as web browsing. It enabled better resource allocation and set the stage for the development of 3G technologies [19]. GPRS users can be categorized into three classes:

Class A supports subscribers to concurrently utilize voice and data services on the GPRS network.

Class B enables simultaneous GPRS and GSM attachment but does not allow the concurrent use of voice and data services.

Class C facilitates simultaneous provision of voice and data services but allows attachment to GSM or GPRS.

The following diagram denotes the architecture of GPRS:

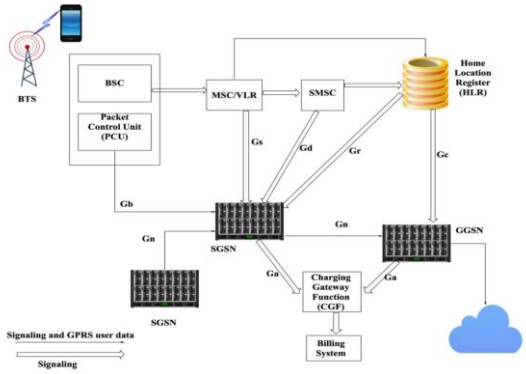


Figure 13 GPRS Network Architecture

In the GPRS network, the Packet Control Unit (PCU) is key to managing the air interface. It takes care of access control, packet scheduling, and packet assembly/reassembly. It's usually part of the Base Station Controller (BSC) and aids in the efficient flow of packets through the network [19].

The Serving GPRS Support Node (SGSN) in the GSM network serves functions similar to the Mobile Switching Center (MSC)/Visitor Location Register (VLR) in circuit-switched systems, managing mobility, security, and access control. It facilitates seamless station registration and uninterrupted data sessions through Packet Data Protocol (PDP) contexts. The SGSN communicates with the Base Station Controller (BSC) over the GB interface and with the Home Location Register/Visitor Location Register (HLR/VLR) over the GS interface. It also connects to the Short Message Service Center (SMSC) for SMS messaging and the Gateway GPRS Support Node (GGSN) for external internet access. The GGSN manages data flows and assigns fixed IPv4 addresses to mobile stations for consistent connectivity. [19].

The EDGE Network Architecture

As indicated and applied in GSM and GPRS, EDGE an evolution within the GSM and GPRS frameworks, employs the same 200-kHz channel bandwidth divided into eight timeslots. It enhances data transmission through 8-PSK modulation alongside the 0.3 GMSK modulation of GSM, improving bandwidth efficiency significantly. This advancement allows for the transmission of more data within the same spectral bandwidth. In EDGE, referred to as Enhanced GPRS (EGPRS), new coding schemes (MCS-1 to MCS-9) are introduced to optimize data service speeds further, reaching practical speeds of up to 100 kbps [20].

High-Speed Circuit Switched Data (HSCSD)

Before the introduction of GPRS or EDGE, there was a recognized need for higher data service speeds. At that time, GSM enabled data services only up to 9.6 kbps, utilizing a single time slot. The straightforward solution was to implement HSCSD, was the precursor to GPRS and EDGE, addressing the need for higher data speeds within GSM, initially capped at 9.6 kbps. HSCSD increased capacity by enabling multiple time slots, with enhancements allowing up to 57.6 kbps through channel coding modifications. The adoption of 8-PSK modulation in EDGE further amplified throughput with fewer timeslots, significantly improving data transmission efficiency.

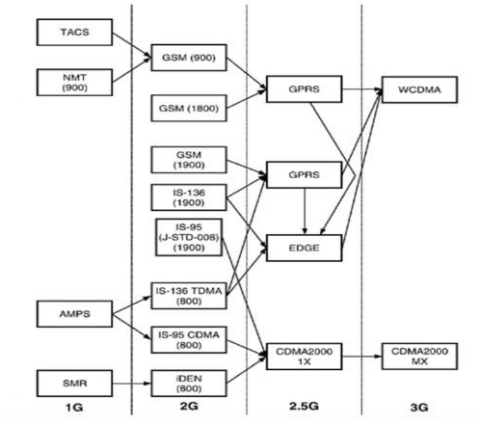


Figure 14 Migration Path [20]

Universal Mobile Telecommunication Service (UMTS)

UMTS represents a significant leap from 2.5G to 3G, primarily in the access network segment, adopting WCDMA as its air interface technology. The core network evolution followed, with notable GSM architectural enhancements leading to various 3GPP releases. Each release introduced substantial upgrades from 1996 through 2000, including the shift to UTRA and extensive core network changes. UMTS aims to provide superior performance, with data rates up to 2 Mbps and services tailored for diverse user needs, from low-delay conversational services to interactive and streaming services [21].

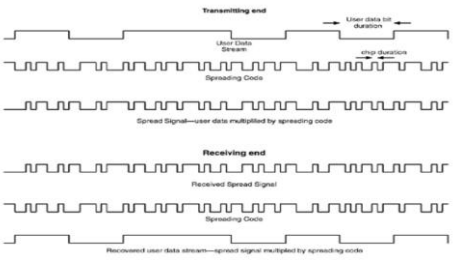


Figure 15 CDMA Basic Concepts

3GPP Release 1999 and 3GPP Release 4 Network Architectures

In 3GPP Release 1999, the foundational architecture for UMTS was established, setting up interfaces for WCDMA connectivity between User Equipment (UE) and the network. This setup introduced Node B as the base station and the Radio Network Controller (RNC) to manage network resources, forming the radio network subsystem. These components, connected through standardized interfaces, support interoperability and mobility management across the network. The evolution to 3GPP Release 4 marked a significant transition towards a distributed core network architecture. This change aimed to enhance efficiency and support the shift towards packet-based voice and data transport, optimizing network operations for the evolving demands of mobile communications [11][21].

CDMA 2000 Architecture

The CDMA2000 architecture, evolving from CDMA-One, incorporates packet data services and focuses on efficient data management. The system architecture can vary between centralized and distributed models, depending on specific design and operational considerations. It shares evolutionary traits with WCDMA in transitioning from 2G to 3G technologies [11].

D. 4G Security Protocols & Architecture

OFDM Technology

The transition to Long-Term Evolution (LTE) marked a significant evolution from traditional 3G systems based on UMTS and CDMA 2000 technologies using Code Division Multiplexing (CDM) techniques. LTE's integration of Orthogonal Frequency Division Multiplexing (OFDM) was a game-changer, offering substantially enhanced data rates and benefits over its 3G predecessors [12].

Orthogonal Frequency Division Multiple Access (OFDMA), a key innovation within LTE, effectively addresses interferences caused by multipath scenarios. By dividing high-bit-rate data streams into parallel lower-bit-rate streams, OFDMA mitigates the issues arising from multipath propagation. This method not only simplifies handling multipath scenarios but also reduces computational complexity, thanks in part to the use of the Fast Fourier Transform (FFT). Moreover, adopting multi-antenna techniques significantly improves link capacity, robustness against interference, and spectral efficiency. This is crucial for combating multipath fading and enhancing the quality and reliability of wireless communications. One of the standout features of LTE is implementing multiuser Multiple Input Multiple Output (MIMO) technology. This allows for the concurrent support of multiple users on the uplink, further optimizing network capacity and user experience [12].

Long Term Evolution (LTE): Architectural Evolution and Decentralization

The architecture of LTE introduces significant changes compared to UMTS, as depicted in the figure below. Focused on delivering high data rates, low latency, and efficient packet-optimized radio access, LTE employs a decentralized approach, streamlining tasks and resources. This departure from central control enhances bandwidth utilization. Key distinctions include eliminating centralized controllers like RNC, SGSN, and GGSN. LTE's flattened network structure ensures a simplified, interconnected base station system without a central controller, enhancing real-time service responsiveness.

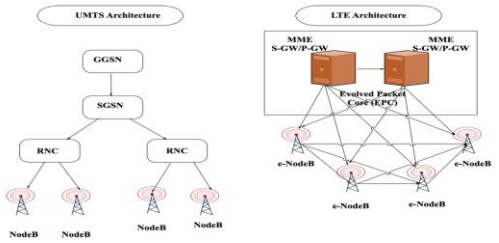


Figure 16 LTE Architectural Evolution and Decentralization

The cornerstone of LTE architecture is the e-Node B (evolved-Node B), which comprises the Remote Radio Unit (RRU) for signal modulation and demodulation and the Baseband Unit (BBU) for digital signal processing. This innovative structure consolidates functionalities, managing tasks such as radio bearer control, admission control, mobility control, and resource scheduling. The e-Node B plays a pivotal role in user data management, handling responsibilities like IP header compression, user data stream encryption, and routing data to the serving gateway. It interfaces with the Serving Gateway (S-GW) for 3GPP radio access connectivity and the Packet Data Network Gateway (P-GW) for IP data service control. The Mobility Management Entity (MME) ensures user authentication, while the S-GW and P-GW form the Evolved Packet Core (EPC) together.

Diameter Protocol

Initially designed as a successor to Signaling System 7 (SS7), the Diameter protocol focused on authentication, authorization, and accounting, essential functions in telecommunications. Over time, Diameter evolved into a versatile signalling protocol for various provider interactions, extending beyond its original scope to support the telecommunications industry's expanding needs [22]. Its robust, self-contained routing logic streamlines network communication and enhances security by reducing external exposure risks, making Diameter a pivotal component in modern telecommunications networks [23][24][25]. This evolution highlights Diameter's significance and adaptability within the industry, underlining its role in advancing network communications.

E. 5G Security Protocols & Architecture

5G RADIO ACCESS Network

The key technologies supporting the 5G radio Access Network (RAN) encompass mmWave Communication, massive Multiple Input Multiple Output (MIMO), ultra-dense cell deployment, Machine-to-Machine (M2M) and Device-to-Device (D2D) communications, cloud-RAN, and mobile edge and fog computing.

mmWave Communication and massive Multiple Input Multiple Output (MIMO)

Millimeter-wave (mmWave) Communication, pivotal for 5G, leverages high-frequency bands above 10 GHz, unlocking substantial bandwidth and enabling data rates up to tens of Gbps.

This leap requires venturing into higher spectrum bands, which are traditionally used less due to challenges like interference and propagation loss. The mmWave spectrum, spanning 30 to 300 GHz, was initially pinpointed by Jagadis Chandra Bose in 1897. Despite its historical use in limited applications due to high loss, recent advancements are making it viable for broader 5G network applications, aiming to overcome hurdles such as interference and fragmentation.

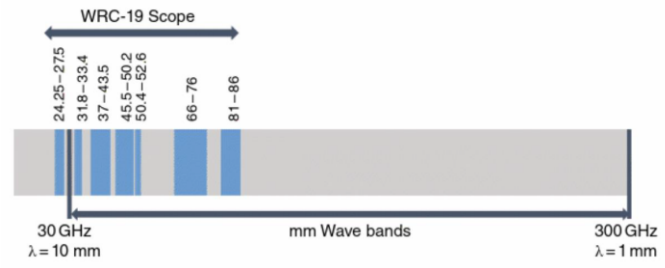


Figure 17 mmWave Communication

To address the 5G network's density and capacity demands, the adoption of massive MIMO (Multiple Input Multiple Output) technology is essential. This method, an evolution of MIMO techniques used in 4G for multi-user support, significantly scales up the number of antennas at base stations, promising enhanced network efficiency and capacity. However, this technology's deployment faces challenges, including signal contamination mitigation and channel estimation.

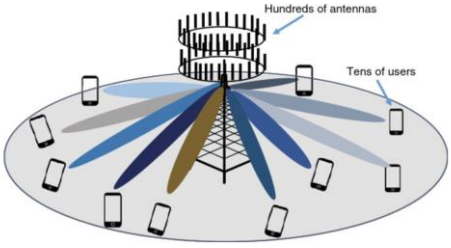


Figure 18 massive MIMO Concept

Massive MIMO and related strategies like spatial diversity and antenna beamforming are crucial for mitigating signal degradation caused by multi-path propagation. Spatial diversity ensures multiple signal paths to the receiver, enhancing transmission reliability, while massive MIMO uses numerous antennas to boost spatial diversity, capacity, and link robustness. Antenna beamforming further strengthens signal quality and network performance by focusing transmissions. Collectively, these advancements in antenna technology underpin the high data rates, improved signal reliability, and network robustness essential for fulfilling 5G's promises

Ultra-dense Small Cells

To enhance network density and throughput, leveraging small cells—a technology defining low-powered radio access nodes with a short coverage range—emerges as a critical strategy. Unlike the broader coverage offered by traditional macro-cell base stations in 3G and 4G networks, small cells cover areas from ten to several hundred meters, suitable for high-traffic zones like indoor spaces and hotspots. Operating across licensed and unlicensed spectra, including Wi-Fi frequencies, small cells help bring the network closer to the end-users, optimizing spectral efficiency through the densification of wireless nodes.

Integrating small cells within the 5G infrastructure forms a heterogeneous network that combines them with macro cells, often linked via wireless backhaul. This setup aims to extend network capacity by offloading traffic to small cells and introduces complexity in managing interference and mobility across the diverse cell types. Research in small cell technology for 5G focuses on challenges such as load balancing, efficient wireless backhauling, exploiting mmWave frequencies, and implementing massive MIMO within small cell configurations to ensure the network's future scalability and performance [1].

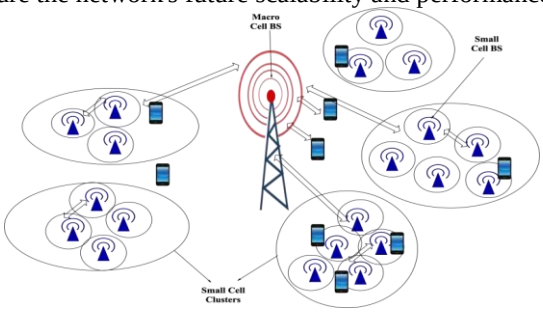


Figure 19 Illustration of Ultra-dense Small Cells

M2M and D2D Communications

About two-thirds of 5G use cases are linked to IoT and Machine-Type Communication (MTC), covering a broad spectrum of essential communications. Initially conceptualized in the 4G LTE era by 3GPP, M2M or MTC communication continues to be a critical driver for 5G development. This type of communication facilitates the autonomous data exchange between devices and the network infrastructure, supporting interactions from an MTC device to a server or directly between MTC devices.

The array of services and applications powered by M2M communication is vast, including but not limited to monitoring and metering, home and industrial automation, healthcare applications, and automotive technologies. As we move forward, addressing M2M communication's challenges, including scalability, privacy and security, and energy efficiency, will be crucial in unlocking its full potential within the 5G ecosystem [1].

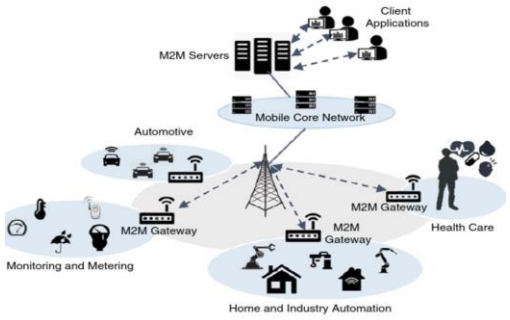


Figure 20 M2M Communication

D2D communication defined in LTE Release 12 by 3GPP, enables mobile devices and users to communicate directly without network infrastructure. This communication model, integral to the 5G framework, aims to enhance spectrum efficiency, improve user data rates, and reduce latency and energy usage. D2D communication can operate within licensed

cellular spectrums, like LTE, or utilize unlicensed spectrums, including Wi-Fi, for out-of-band D2D scenarios.

The figure demonstrates various use cases for D2D, which span proximity-based services, gaming, public safety, vehicular communications, and data offloading. While D2D offers numerous advantages, it also introduces challenges such as interference management, device discovery, and ensuring security and privacy. Future research directions include exploring D2D's integration with mmWave and massive MIMO technologies, highlighting its potential to enrich the 5G landscape further [1].

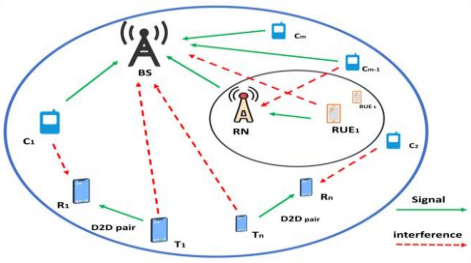


Figure 21 D2D Communication

IV. Attacks and Security Implementation on Mobile Network Protocols

Mobile networks face many challenges, such as facing security threats from attacks on their network. The first common attack is the man-in-the-middle attack. It works by interfering with the communication pattern between the infrastructure of the networks and the devices. Moreover, a denial-of-service attack operates by hacking network resources, making services unavailable to users. Packet sniffing works by obtaining information through the eavesdropping method and obtaining crucial information such as the username and financial data of legitimate users [26]. Additionally, spoofing attacks cause threats through the illegal impermeability of networks to gain unauthorized access. Lastly, phishing attacks cause threats by convincing users through illegal means to disclose crucial information, which allows hackers to cause attacks.

Most importantly, these attacks pose serious threats to mobile networks by exploiting their weaknesses. As a result, the integrity and confidentiality of information shared are compromised, hindering the service delivery offered by the networks [27]. Additionally, there are many protocols that face attacks, and they include the Global System for Mobile Communication (GSM), the Universal Mobile Telecommunication Service (UMTS), LTE, and the 5G protocols [26]. The attacks pose serious threats and undermine the effectiveness and reliability of these networks.

A. Security Implementations and Attacks on 1G

The security of wireless cellular networks is crucial for managing and protecting sensitive information in our technologically evolving world. Despite facing challenges like open access and limited bandwidth, providing features such as authentication and confidentiality remains essential. In the first generation (1G) of mobile technology, security concerns emerged as criminals exploited vulnerabilities to engage in mobile fraud, including phone cloning and eavesdropping, using handheld scanners readily available in commercial outlets such as Radio Shack [1].

B. Security Implementations and Attacks on 2G

Due to the massive cellular network architecture of 2G, it is open to several attacks, such as [1].

i. ***Man-in-the-Middle Attack:*** In a Man-in-the-Middle attack, the attacker positions themselves between two communicating entities, intercepting their messages, and potentially manipulating the communication.
ii. ***Channel Jamming:*** This tactic disrupts legitimate users' access to a wireless channel within a particular network, causing a denial of service.
iii. ***Unauthorized Access:*** Weaknesses in the authorization mechanism can permit attackers to infiltrate the network, enabling the performance of unauthorized activities.
iv. ***Eavesdropping:*** The absence of encryption in communication channels allows attackers to listen in on sensitive calls or text messages surreptitiously.
v. ***Session Hijacking:*** Session Hijacking involves a malicious user taking control of an already established session. The attacker can impersonate the base station, leading to potential unauthorized access and manipulation of the ongoing communication.
vi. ***Denial of Service attack (DOS):*** This prevalent attack involves overwhelming network resources by inundating the network with excessive data, surpassing its capacity. The consequence is that subscribers are unable to access the network resources.
vii. ***Distributed Denial of Service attack (DDOS):*** Like DOS, this attack employs multiple hosts to launch a large-scale assault, as it is impractical for a single host to achieve such an impact.
viii. ***Channel Jamming:*** This tactic disrupts legitimate users' access to a wireless channel within a particular network, causing a denial of service.
ix. ***Unauthorized Access:*** Weaknesses in the authorization mechanism can permit attackers to infiltrate the network, enabling the performance of unauthorized activities.
x. ***Eavesdropping:*** The absence of encryption in communication channels allows attackers to listen in on sensitive calls or text messages surreptitiously.
xi. ***Message Forgery:*** If the communication channel is insecure, attackers can intercept messages in transit and alter their content without the users' knowledge.
xii. ***Message Replay:*** Even if the communication is secured, attackers can intercept an encrypted message, store it, and replay it later. Users may need to be informed that the received packet is replayed.

2G employs security measures to safeguard against various attacks. The A3 algorithm ensures the authentication of legitimate network users. For encrypting the communication flow between devices, the A5 algorithm is employed, while the A8 algorithm generates the cipher key. In the authentication process, the Visitor Location Register (VLR) sends a random value (RAND) to the Subscriber Identity Module (SIM).

Subsequently, the mobile station sends the Cipher Key (SRES) generated by the A8 algorithm to the VLR. The VLR compares the two values, accepting the subscriber if they match and rejecting the authentication if they do not [1].

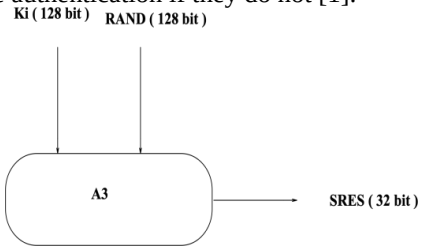


Figure 22 A3 Algorithm

After authentication, as illustrated in the Figure below, the base transceiver station and the mobile station implement voice, data, and signalling encryption by utilizing the cipher key Kc [1]. Kc is generated by multiplexing a random value (RAND) and the key Ki using the A8 algorithm. This encryption for confidentiality is specific to the communication between the mobile station and the base station transceiver and does not extend to the entire GSM network. Throughout the process, both the Subscriber Identity Module (SIM) in the mobile station and the Visitor Location Register (VLR) calculate the same Kc, facilitating the encryption or decryption of messages using the A5 algorithm.

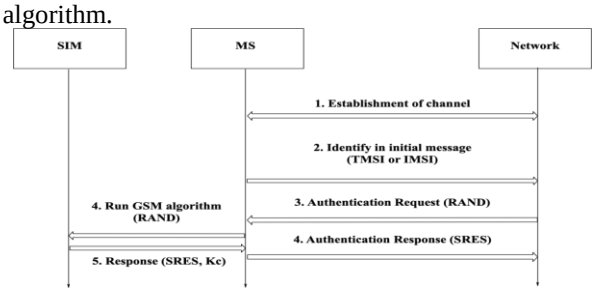


Figure 23 A3 Algorithm Working Principle

C. *Security Implementations and Attacks on 3G*

The 3G mobile network is installed with the protocol of the Universal Mobile Telecommunication Service (UMTS). Moreover, it is vulnerable to various attacks, such as protocol exploitation, hijacking sessions, SIM card cloning, and packet sniffing. Protocol exploitation interferes with the 3G network through intercepting mobile signals to gain illegal access to phone conversations [27]. On the other hand, hijacking sessions operate by hijacking the establishment procedures of mobile networks, interfering with their connectivity. Packet sniffing exploits the network by accessing sensitive content from mobile network users, such as their usernames and passwords. Most importantly, network operators try their best to implement security measures that reduce these attacks.

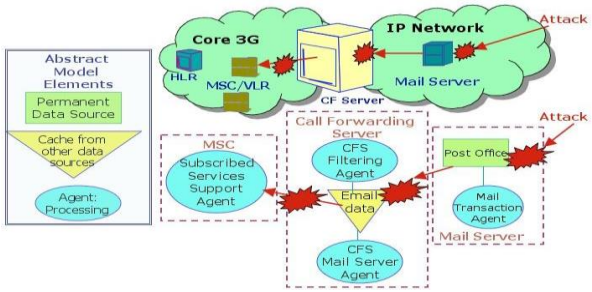


Figure 24 Attack Propagation in Call forwarding services (CFS) with simplified abstract model [43]

D. *Security Implementations and Attacks on 4G*

The 4G mobile network is installed with the LTE protocol. Furthermore, it is vulnerable to various security attacks, such as the LTE IMSI Catcher and DOS attacks. These attacks exploit this network by using unique features that are not easy to identify. For instance, LTE IMSI Catcher causes damage by interfering with user numbers, and as a result, they can track the movement of the users [28]. Denial-of-service attacks cause damage by causing traffic in the system and disrupting the original flow of services to users. Nevertheless, mobile networks should play a critical role in regularly conducting system updates on their networks to establish security threats in their systems.

The Diameter protocol, a successor to the SS7, has vulnerabilities and potential attacks, especially within the context of 4G and 5G networks [29]. Many attacks possible in SS7 have equivalents in Diameter, with some differences. Authentication in Diameter is primarily through certificates, but its usage is limited due to operational implications. Security is bolstered by other factors such as origin checks and certain hiding functions [24]. Threats such as subscriber tracking via specific messages like IDR (Interrogation Data Request) are prevalent. However, intercepting Diameter traffic doesn't necessarily mean intercepting actual voice and message traffic due to its role in signaling rather than data transport [25].

Potential attack scenarios include location tracking and interception of voice traffic, which involve complexities and require multiple factors for successful interception. Subscriber privacy, particularly regarding location, is a significant concern, emphasizing the need for robust security measures. Diameter messages incorporate checks for legitimate message origins, but these checks may be susceptible to manipulation by attackers due to implementation flaws and lack of thorough checking in practice [29]. Mitigating attacks involves practical considerations such as cross-checking values within Diameter messages and maintaining state tables for subscriber devices. Tools like "Diameter Enum" are designed to assess and test Diameter setups for vulnerabilities and misconfigurations, providing insights into exposed functionality and potential attack vectors.

E. *Security Implementations and Attacks on 5G*

The 5G mobile network is vulnerable to various attacks such as network slicing, denial of service attacks, man-in-the-middle attacks, authentication problems, and jamming problems. These attacks disrupt the normal service delivery to legal users,

causing great harm to them. The man in the middle interferes with the communication pattern through the illegal access of sensitive data from the devices of the users [28]. On the other hand, network slicing operates by compromising the available resources and, as a result, contributing to the network's depreciation. Jamming works by causing traffic in the network by disrupting the network signals. Most importantly, the 5G network is the latest version and is updated with security measures that reduce the number of vulnerabilities in its network. These vulnerabilities to attacks have various impacts on the security landscape. Increased exploitation of mobile network services causes some of the security protocols to depreciate, making them lose customer trust and creditworthiness. Most importantly, mobile network operators should install robust security measures that prevent access by third parties [30].

Pre-authentication Message Exploits which enable unauthorised network access and security lapses can result from pre-authentication message exploits. These communications might be intercepted by attackers looking to carry out attacks or get private data [31]. Null Encryption and Null Integrity where attackers may utilise 5G networks' support for these features to launch bidding down assaults and rogue base stations. The integrity and confidentiality of communications might be jeopardised by this [31]. Key Management Issues where there is a security risk due to the absence of explicit standards for key management of operator public keys in subscribers' USIMs. Unauthorised access and data breaches might be caused by poor key management procedures.

Insecure Protocol Implementation leads to vulnerabilities that attackers can exploit may be introduced by insecure protocol implementations in 5G networks.

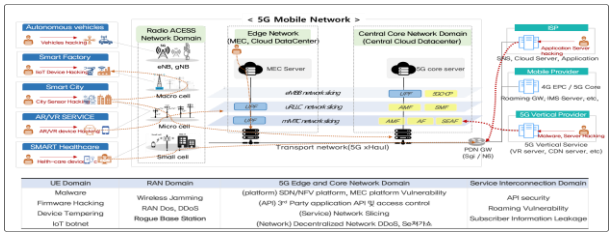


Figure 25 5G Security Attacks Propagation [27]

5G network can safeguard against these attacks and vulnerabilities by implementing a global PKI Architecture, by offering a foundation of trust for digital certificate-based message and communication peer authentication, the implementation of a worldwide 5G Certificate Authority (CA) can improve security [31]. This method helps with important management issues and provides a more adaptable design. Robust Key Management will stop unwanted access and data breaches, secure key management procedures must be established. These procedures must include appropriate key rotation and protection [31]. Managing and storing operator public keys securely is essential to preserving the network's integrity. Enhanced Cryptographic Functions on the 5G network security may be reinforced against possible intrusions by enhancing cryptographic functions and algorithms. Vulnerabilities can be reduced by routinely upgrading integrity protection and encryption systems.

Most importantly, these attacks cause serious regulatory and legal challenges that disrupt the services offered to customers. As a result, they contribute to the depreciation in value of mobile networks, which is mainly caused by various factors such as poor security measures and poor regulatory control measures [30]. Additionally, mobile operators should create awareness among their customers to reduce cases of cyber security. There are many reasons why these mobile network protocols are vulnerable to attacks. Some mobile network protocols are installed with complex systems that are hard for users to understand. Weak authentication mechanisms installed in mobile networks increase their vulnerability to attacks [30]. Human factors such as committing mistakes and errors increase the vulnerability of these networks. Moreover, some are fitted with weak encryption mechanisms, making it hard to protect their data. They are highly interconnected with features that resemble other network systems, increasing their vulnerability to attacks.

V. SIGNALING SYSTEM NO.7 (SS7)

Signaling System No. 7 (SS7) protocol is a remarkable testament to the evolution of telecommunications technology, with its roots tracing back over a century. This protocol has been a significant player in the development and advancement of telecommunications networks globally. In the early days of telecommunication, manual telephone switching was the norm. Human operators were tasked with manually establishing connections between telephone lines through large ports. However, as the demand for telecommunication services grew, the need for a more efficient and automated signaling mechanism became increasingly apparent [18]. This growing necessity led to the development of automatic switching systems. The initial systems relied heavily on mechanical switches such as rotary machines. Over time, these evolved into more sophisticated electro-mechanical devices like the Strowger switch and Hebditch and Preece (HP) exchanges [1]. These systems, which were based on pulse dialing and in-line signaling, paved the way for the transition to digital signaling technologies.

The introduction of digital switching marked a significant milestone in the evolution of telephone signalling. It enabled faster and more reliable communication, revolutionizing telecommunications [21]. With the advent of SS7, initially called C5 and later C6, a paradigm shift occurred in the industry. SS7 introduced digital messaging for signalling information relay between telephone exchanges, transitioning from in-line signalling to common channel signalling [21]. Common channel signalling separated signalling from voice circuits, leading to more efficient bandwidth utilization and enhanced security. By establishing a dedicated signalling path, SS7 facilitated faster call setup times and improved network performance [21]. This transition was instrumental in addressing the inefficiencies and vulnerabilities associated with in-line signalling, making telecommunications networks more robust and secure [21]. The architecture of SS7 is centred around Service Switching Points

(SSPs), which are telephone switches responsible for providing voice services to subscribers [15]. These SSPs communicate using the SS7 protocol, with signalling paths established between them for call setup and management. SS7 nodes are identified by point codes, serving as unique identifiers similar to IP addresses in IP networks [15]. Network managers configure routes statically, ensuring deterministic routing within the SS7 network.

Despite its historical roots, the SS7 protocol remains a cornerstone of modern telecommunications infrastructure. Its widespread adoption and standardized implementation have made it an indispensable component of global telecommunications networks [21]. Service providers and enterprises worldwide rely on SS7 for secure and efficient signalling, enabling seamless communication between subscribers and ensuring the integrity of telecommunications services [21]. The history and importance of the SS7 protocol underscore its pivotal role in shaping the evolution of telecommunications networks [15]. From its origins in manual telephone switching to its modern-day implementation as a digital signalling standard, SS7 is crucial in facilitating reliable and secure communication across global telecommunications networks. Its impact is felt every day, in every corner of the globe, as people connect and communicate with each other. This is the legacy of SS7, a protocol that has genuinely shaped the world as we know it [15].

A. SS7 Terminologies

Signal Switching Point (SSP): A crucial element of the SS7 network, is a telephone switch responsible for call processing and registration. It analyzes dialed numbers, establishes call routing, and manages call setup and teardown processes. Serving as the central hub in traditional telephone networks, the SSP ensures the successful connection of voice calls between subscribers, making it essential to ensure call delivery to intended destinations.

***Signal Transfer Point (STP*)**: This point acts as a router within the SS7 network, directing signalling messages between Service Switching Points (SSPs) based on the source and destination point codes (SPC and DPC). Its role is akin to that of a traffic controller, ensuring efficient and reliable communication by guiding messages along the most effective routes.

Signal Control Point (SCP): Serves as the intelligence hub of the SS7 network, providing instructions and services to SSPs. It performs various functions, including number translation, balance confirmation, and executing tasks within the Intelligent Network (IN) architecture. It can offer specialized services such as location-based services or manage prepaid billing systems, essentially acting as the brain of the network that optimizes performance through intelligent decisions.

Mobile Switching Center (MSC): Functions as an SSP specifically for cellular wireless networks, managing call processing in circuit-switched mobile networks. It oversees call routing, handovers between cell sites, and subscriber authentication. The MSC is pivotal in managing voice calls and mobility for mobile subscribers, ensuring seamless communication within mobile networks and acting as the mobile counterpart to the SSP.

A Number and B Number: In SS7 signaling, the A Number represents the source phone number, while the B Number signifies the destination phone number. These numbers are used to identify the calling and called parties during call setup and routing processes. In a telephone call, the A Number corresponds to the caller's phone number, while the B Number corresponds to the recipient's phone number. They're like the addresses on a letter, guiding the call from the caller to the recipient.

Point Code: Serves as a unique identifier at the MTP3 layer, analogous to an IP address in the SS7 network. It is used for routing signaling messages between network elements and plays a crucial role in establishing communication paths. Point Codes are used by STPs to determine the next hop for signaling messages within the SS7 network. They're like the GPS coordinates of the network, guiding messages to their destinations.

Global Title (GT): A unique identifier at the SCCP layer, often associated with specific applications or services within the SS7 network. It provides addressing information for routing messages to specific destinations based on application or service requirements. In an SS7 network, a GT may be used to route messages to a specific service such as voicemail or SMS messaging. It's like the name of a business, directing messages to the right service.

Link, Link Set, and Route Set: Links represent direct connections between Point Codes, Link Sets consist of multiple links between the same Point Codes, and Route Sets consist of redundant routes to a destination. Links enable direct communication between network elements, Link Sets provide redundancy and fault tolerance, and Route Sets offer multiple paths for message routing. In an SS7 network, Link Sets and Route Sets ensure reliable communication by providing backup routes and alternate paths for signaling messages. They're like the roads, highways, and detours of the network, ensuring that every message reaches its destination, even if a link fails.

Each of these components plays a vital role in the functioning of the SS7 network, ensuring that every call is routed correctly and efficiently. From the SSP, which handles the basic call processing, to the SCP, which provides intelligence to the network, each component contributes to the overall performance and reliability of the network. The A Number and B Number guide each call, while the Point Code and Global Title ensure that each message reaches its intended destination. And through it all, the Link, Link Set, and Route Set provide the infrastructure that keeps the network running smoothly, ensuring that every message has a path to follow, even in the event of a failure. This is the power and complexity of the SS7 network, a testament to the ingenuity and innovation of telecommunications technology.

B. SS7 Protocols and Architecture

The Signaling System No. 7 (SS7) protocol stack is a complex hierarchical structure that consists of multiple layers, each serving specific functions within the signaling network. This structure is designed to facilitate efficient and reliable signaling communication between network elements, ensuring the smooth transmission of signaling messages [32]. The SS7 protocol stack is commonly organized into eight layers [11]:

***MTP Level 1 (MTP1 - Physical Layer*):** This is the groundwork of the SS7 network, focusing on the physical aspects of connectivity. It defines the electrical, mechanical, and functional specifications required for signalling points to interface with each other over physical media, such as copper wires or fibre optic cables. It ensures the integrity of the network's backbone's physical connections.

MTP Level 2 (MTP2 - Data Link Layer): Acting as the network's safeguard, this layer is tasked with framing, error detection, and flow control. By implementing error correction and sequencing mechanisms, MTP2 ensures that signalling messages traverse the physical link reliably and efficiently.

MTP Level 3 (MTP3 - Network Layer): MTP3 serves as the traffic controller of the SS7 network. It manages the routing of signalling messages between signalling points by determining the most efficient paths based on destination point codes (DPC) and signalling link selection.

The SCCP (Signaling Connection Control Part): Operating at the network layer, the SCCP enhances routing and connection capabilities within the SS7 signalling network. It supports efficient network management by facilitating establishing and maintaining signaling connections, using various addressing schemes and routing strategies for reliable message delivery.

The TCAP (Transaction Capabilities Application Part): operates at the application layer and facilitates the exchange of at the application layer, TCAP enables the exchange of non-circuit-related information, supporting the implementation of advanced signalling services like database queries and service activation. This layer allows for dynamic communication scenarios, facilitating the network's handling of complex transactions and services.

The ISUP (ISDN User Part): At the application layer, ISUP is tailored to manage signalling related to ISDN services and oversee call setup, teardown, and signalling for call-related functions. It is essential to exchange call control information, highlighting its role as the ISDN services specialist within the SS7 network.

The Mobile Application Part (MAP)*:* Critical for mobile networks, MAP facilitates communication between network elements to provide vital services such as subscriber authentication, call routing, SMS delivery, and mobility management. Its functions are integral to the GSM and UMTS frameworks, underscoring its importance in mobile telecommunications.

Each layer and component of the SS7 protocol stack contributes to the network's capability to route calls efficiently and reliably, from foundational call processing to the provision of specialized services. The architecture's depth—from the physical connections enabled by MTP to the advanced service facilitation by TCAP, ISUP, and MAP—demonstrates the ingenuity and complexity of telecommunications technology. This system supports today's telecommunication services' dynamic requirements. It ensures that every message and call is processed with high reliability and precision, marking a significant achievement in the evolution of communication networks [32].

C. DENIAL OF SERVICE on SS7 PROTOCOL

DoS attacks threaten telecommunications, particularly within the Home Location Register (HLR) component [13]. Attackers exploit network weaknesses, disrupting services and compromising user communication. Vulnerabilities, like bypassing authentication in emergency location services, such as the L2 Map feature in the United States, allow unauthorized access to precise subscriber location information. This poses significant privacy and security risks, particularly in emergencies, potentially hindering response efforts and endangering lives [18].

Attackers can also manipulate subscriber data within the Mobile Switching Center (MSC) or Visitor Location Register (VLR) [18]. This unauthorized control allows them to alter calling, SMS, or data services and even delete subscriber data from the VLR. Such manipulation compromises network integrity and violates subscriber privacy, emphasizing the need for robust security measures [11]. Additionally, the Home Location Register (HLR), storing subscriber information, is vulnerable to attacks [18]. Exploiting HLR vulnerabilities, attackers can reroute calls, intercept SMS messages, or engage in fraudulent activities, posing privacy risks [16]. Furthermore, attackers can exploit Unstructured Supplementary Service Data (USSD) codes for malicious actions like transferring credits or activating call forwarding without user consent, underscoring the need for comprehensive security measures to mitigate such risks.

D. Tracking and Location Attacks on SS7 Protocol (IMSI CATCHER)

IMSI catchers, MC catchers or Salt Tower spoofers deceive smartphones into connecting to them instead of actual cell towers. Initially used by law enforcement, their accessibility has led hackers to adopt them. Terms like cell site simulators, fake cell towers, Stingray, dirt box, and cell spoofers are used interchangeably. These devices emit signals mimicking legitimate towers, tricking devices into connecting. Once connected, they can intercept, block, or modify calls and messages covertly, posing a significant threat to user privacy and security [12][32][16][34].

IMSI catchers can be built using inexpensive online parts, open-source software, and a portable computing device like a laptop [33]. Concerns arise due to their affordability and simplicity, raising alarms about potential misuse by malicious entities. The ease of construction underscores the urgent need for robust security measures [33]. An emerging trend involves converting Wi-Fi routers into IMSI catchers, creating open Wi-Fi networks to lure unsuspecting smartphone users with auto Wi-Fi connect enabled [34]. IMSI catchers can exploit mobile phones' tendency to connect to the strongest signal at a higher priority frequency or physically position themselves closer to the target than the nearest legitimate tower [34].

This manipulation of mobile phone behavior highlights the sophistication of IMSI catcher attacks and the challenges in defending against them. Despite advancements in mobile technology security, such as 4G LTE and 5G protocols, IMSI catchers can downgrade connections to less secure 3G or 2G networks, bypassing security measures [21]. This exposes users to potential attacks, emphasizing the ongoing struggle to secure mobile communications [35]. Criminals employ IMSI catchers for various nefarious activities, including location tracking, data

extraction, interception, and spyware delivery to targeted smartphones. These devices enable monitoring of individuals, gathering information, intercepting communications, and compromising devices with malware, underscoring their threat to user privacy and security.

VI. Implementation

A significant security concern for the next generation of wireless technology, particularly 5G, is its reliance on 4G/LTE components. However, what is more alarming is the reliance on 2G and 3G components. This is because an adversary with network access maliciously could target 4G/LTE systems with a downgrade attack to use 2G/3G protocols. SS7, the signalling protocol responsible for setting up and terminating calls, is the main issue with 2G/3G systems. Over time, the protocol improved with additional features like SMS, prepaid billing, call waiting/forwarding, and more [44]. SS7 contains several vulnerabilities, such as denial of service, fraudulent activity, breach of user privacy, and interception [1], and some of these vulnerabilities will be exploited during this implementation.

Sigploit

A few years back, two researchers, Rosalia D'Alessandro and Ilario Dal Grande, came together to create a telecommunications security testing framework called Sigploit. Furthermore, Sigploit was created because access to the SS7 network needed specialized knowledge and authorization, and this is usually restricted to telecommunication operators, service providers and authorized personnel with the necessary credentials, permissions and technical expertise. It is also worth noting that unauthorized access to the SS7 network is illegal and can result in severe legal consequences [45]. The setup of this framework is in a Linux environment, and it provides simulation executables that allow users to test attacks that target SS7 in a controlled environment. It offers attacks that compromise user privacy and fraud attacks [44].

Setup and Execution

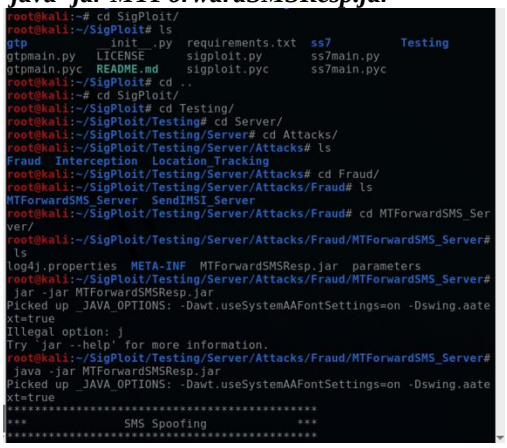
All setups below were executed using a distribution of Debian Linux. However, this can also be done on other operating systems but would require slight modification of the commands used to work with their various packet or network managers.

1. To begin, we install all necessary system dependencies using the command below. However, if you already have python and/or Wireshark installed, you can exclude them from the command:
   Sudo apt-get install git python openjdk-8-jdk maven lkscpt-tools wireshark
2. To run SigPloit, download the python tool and install the python dependencies with the code below:
   git clone https://github.com/SigPloiter/SigPloit
   cd SigPloit
   sudo pip2 install -r requirements.txt
3. Add the following IP addresses to simulate attacker and target for the test environment to leverage. Then launch SigPloit to verify functionality from root/home depending on your installation path.
   sudo ip address add 192.168.56.101/32 dev lo
   sudo ip address add 192.168.56.102/32 dev lo
   cd SigPloit
   python sigploit.py

SMS Routing Attack

The SS7 has several vulnerabilities that can be grouped under fraudulent attacks, and SMS routing is one of them. This attack is possible because a text message is sent from the MSC () to the destination MSC, and the sender is not authenticated by the local HLR () in that transmission. An attacker can send a spoofed SMS to an MSC claiming to be another entity, whether or not they belong to the network. In our simulation, we impersonated the police using 911, making this vulnerability very useful in executing SMS-based phishing attacks [1].

4. To execute this exploit, we must launch the simulation binary in the respective directory using the following command:
   cd SigPloit/Testing/Server/Attacks/Fraud/MTForwardSMS_Server/
   java -jar MTForwardSMSResp.jar

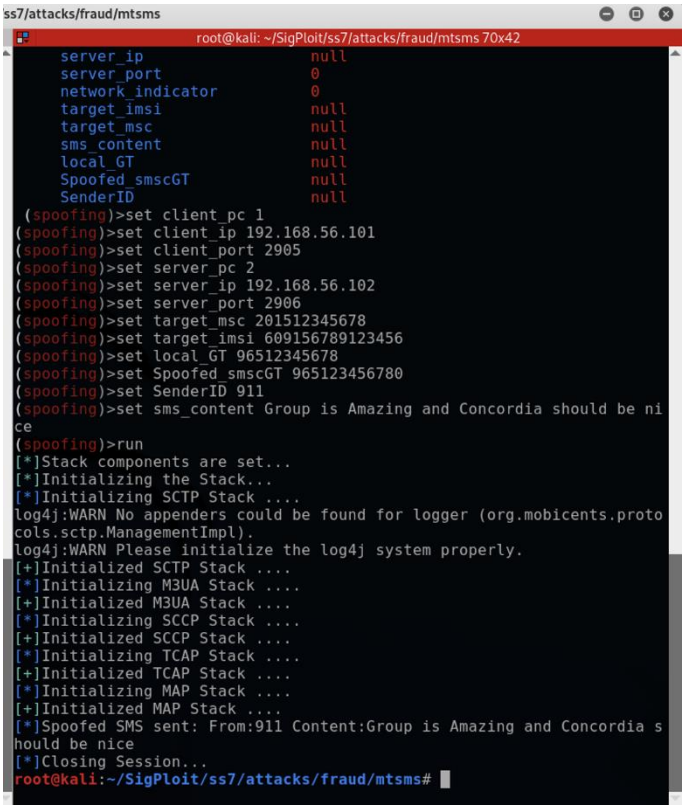


5. Next, we can initiate the exploit from SigPloit by executing the Fraud JAR file from the directory with the following command:
   cd SigPloit/ss7/attacks/fraud/mtsms
   java -jar MTForwardSMS.jar

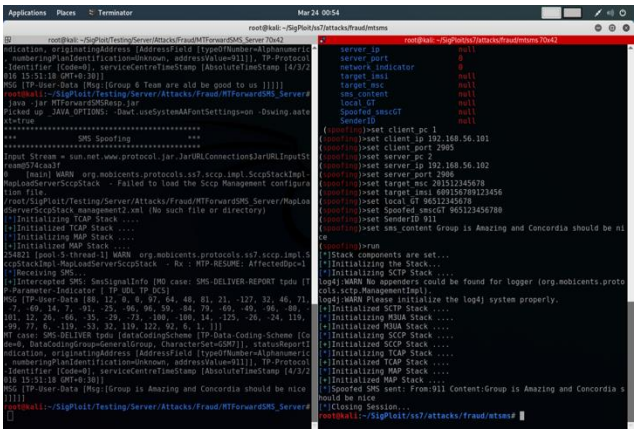
6. The following parameters would need to be configured as the client and server (Attacker and Target) reside on the hardcoded IP addresses we had earlier added alongside their respective ports. From the parameters to be configured, the target MSC and IMSI indicates which target MSC server to attack and the victim IMSI which the message should be sent. The sestination of the attacker is configured as the local_GT sets the global title, and this is the destination/number of the attacker while the spoofed_smsGT is an arbitrary 10-digit number. The SenderID is the spoofed user that the attacker is impersonating ansd the sms_content contains the content of the spoofed message [1].

   set client_pc 1

set client_ip 192.168.56.101
set client_port 2905
set server_pc 2
set server_ip 192.168.56.102
set server_port 2906
set target_msc 201512345678
set target_imsi 609156789123456
set local_GT 96512345678
set Spoofed_smsGT 965123456780
set SenderID 911
set sms_content Group 6 is Amazing and Concordia should be nice



7. As soon as all these parameters are set and ready, enter 'run' and the attack will execute as seen in the snapshots.



Note: More Attacks can be found in the [Appendix]

VII. Limitations and Future Consideration

Signaling System 7 (SS7) vulnerabilities pose significant risks to global telecommunications networks, as they can be exploited to track user locations, intercept messages, and bypass encryption. Despite its critical role in facilitating worldwide communication, the SS7 protocol's inherent security flaws underscore the urgent need for robust protective measures and system updates. One of the most significant vulnerabilities lies in the exploitation of the 'anytime interrogation' feature [36]. This feature, originally intended for internal network use, allows attackers to request location information about mobile subscribers without their knowledge or consent. Attackers can manipulate this feature to obtain a subscriber's Cell ID and International Mobile Subscriber Identity (IMSI), both of which are critical pieces of information for tracking their movements [37] [38]. Moreover, the absence of plausibility checks in most Mobile Switching Centers (MSCs) allows attackers to access location information across different networks, irrespective of their affiliation. By requesting IMSI information from the Home Database (HR), attackers can ascertain the subscriber's location for SMS routing, thereby further facilitating tracking activities.

Individuals can be tracked based on SS7 vulnerabilities solely with their phone number [39]. This highlights the alarming ease with which location information can be remotely obtained. Such tracking capability poses significant privacy concerns, as individuals may be oblivious to the fact that their movements are being monitored [38]. Beyond the implications for privacy, SS7 vulnerabilities have also been exploited for commercial purposes. For instance, shipping companies utilize SS7 to track their vehicles, while banks employ it to verify SIM card swaps in an effort to prevent fraudulent transactions [40][41]. These commercial use cases underscore the widespread exploitation of SS7 vulnerabilities, extending beyond merely malicious activities. In an attempt to mitigate SS7 vulnerabilities, some network operators have implemented filters to reduce attack traffic [42]. However, despite these countermeasures, some attacks persist, indicating ongoing security challenges. The involvement of state actors or other network operators in orchestrating these attacks further complicates the security landscape surrounding SS7. This highlights the need for continued vigilance and robust security measures to protect against these vulnerabilities.

The implementation findings on the SS7 protocol reveal critical vulnerabilities that stem primarily from its insufficient authentication mechanisms across numerous operations. This fundamental flaw renders the protocol vulnerable to unauthorized access and subsequent privacy invasions. Such vulnerabilities not only jeopardize the integrity of global telecommunications but also pose significant risks to individual privacy. The situation is further complicated by challenges encountered with obsolete software during the research phase. Specifically, the reliance on outdated versions of essential tools like Python Pip3 and Maven, necessary for running applications such as SigPloit, highlights a broader issue within network security: the struggle to maintain up-to-date and secure systems amidst rapidly evolving technological landscapes.

Additionally, there are important lessons to be learned from the evolution of mobile networks and the challenges they face. First, there is a need for mobile network operators to conduct continuous security updates to prevent threats. Moreover, industry stakeholders should collaborate with security researchers to identify weaknesses in the industry and come up with solutions to the problems. Some important recommendations include investing in technology that promotes research and development to identify weaknesses in the industry and prevent future security threats. Additionally, mobile network operators should invest heavily in authentication and encryption measures to promote privacy and integrity levels in the networks.

VIII. Summary and Conclusion

Exploring mobile network security, particularly through the lens of the Signaling System No. 7 (SS7) protocol vulnerabilities, illuminates the dynamic and evolving nature of mobile telecommunications. As mobile networks advance from one generation to the next, they are consistently equipped with newer security protocols designed to counteract emerging threats. However, the persistence of vulnerabilities within systems like SS7 illustrates a critical gap in the ongoing battle for network security.

The responsibility of safeguarding mobile networks against security threats extends beyond implementing advanced security protocols. It necessitates a proactive and continuous effort from mobile network operators to conduct regular system updates and vulnerability assessments. These actions are vital for identifying and mitigating security threats that could compromise the network's integrity and the privacy of its users.

Collaboration emerges as a critical theme in addressing mobile networks' security challenges. Partnerships between industry stakeholders and security researchers are indispensable. Such collaborations can drive the identification of vulnerabilities within the network infrastructure and foster the development of robust solutions to address these issues. Engaging in this cooperative approach allows for a deeper understanding of the unique challenges presented by each generation of mobile networks, from security threats to potential exploitation by unauthorized entities.

In conclusion, the journey through the complexities of mobile network security underscores the importance of vigilance, innovation, and collaboration in protecting telecommunications infrastructure. While each generation of mobile networks brings forth new capabilities and services, it also introduces specific challenges that require targeted security measures. Addressing these challenges effectively requires a commitment to regular system updates, stakeholder collaboration, and the continuous evolution of security protocols. As mobile networks evolve, so must the strategies employed to safeguard them, ensuring that they remain resilient against tomorrow's threats.

Team Contributions

Team Member	Contribution
Anita Francis Archibong	*Abstract, Introduction and mobile networks evolution over time.* *Security protocols used at different stages/versions.*
Josephine Famiyeh	*Security protocols used at different stages/versions.* *Various attacks on these security protocols and explain why they were deprecated.*
Elvis Okoye	*Signaling System No. 7 protocol and its vulnerabilities that allow one to track mobile phone users and intercept 2FA texts.*
Valentine Ozonyia	*Related work/ literature review.*
Ugochukwu Kizito Ugwu	*Implement at least two (but the more the better) of the attacks on SS7 protocols in an emulated environment.* *Lessons learned, future considerations & conclusion.*
Ihekweazu Samuel	*Implement at least two (but the more the better) of the attacks on SS7 protocols in an emulated environment.* *Lessons learned, future considerations & conclusion.*
Tweneboath Kodua	*Implement at least two (but the more the better) of the attacks on SS7 protocols in an emulated environment.* *Lessons learned, future considerations & conclusion.*

References

[1] Liyanage, Madhusanka. Comprehensive Guide to 5g Security.
[2] B. G. Gopal and P. G. Kuppusamy, "A Comparative Study on 4G and 5G Technology for Wireless Applications."
[3] E. Koivusalo, Converged Communications: Evolution from Telephony to 5G Mobile Internet, 1st ed., Wiley, 2022. doi: 10.1002/9781119867531.
[4] B. A. Kumar and P. T. Rao, "Overview of advances in communication technologies," in Proc. 2015 13th International Conference on Electromagnetic Interference and Compatibility (INCEMIC), Visakhapatnam, India, Jul. 2015, pp. 102–106. doi: 10.1109/INCEMIC.2015.8055856.
[5] J. Pavia, D. Lopes, P. Cristovao, P. Sebastiao, and A. Correia, "The evolution and future perspective of security in mobile communications networks," in Proc. 2017 9th International Congress on Ultra-Modern Telecommunications and

Control Systems and Workshops (ICUMT), Munich, Germany, Nov. 2017, pp. 267–276. doi: 10.1109/ICUMT.2017.8255180.
[6] U. B. Shukurillaevich, R. O. Sattorivich, and R. U. Amrillojonovich, "5g Technology Evolution," in Proc. 2019 International Conference on Information Science and Communications Technologies (ICISCT), Tashkent, Uzbekistan, Nov. 2019, pp. 1–5. doi: 10.1109/ICISCT47635.2019.9011957.
[7] R. Yadav, "Challenges and Evolution of Next generation Wireless Communication," Hong Kong, 2017.
[8] "2G," Wikipedia, Feb. 28, 2024. [Online]. Available: https://en.wikipedia.org/w/index.php?title=2G&oldid=1210853617. Accessed Mar. 13, 2024.
[9] "4G," Wikipedia, Jan. 30, 2024. [Online]. Available: https://en.wikipedia.org/w/index.php?title=4G&oldid=1201013483. Accessed Mar. 13, 2024.
[10] A. A. Salih, S. R. M. Zeebaree, A. S. Abdulraheem, R. R. Zebari, M. A. M. Sadeeq, and O. M. Ahmed, "Evolution of Mobile Wireless Communication to 5G Revolution," 2020, vol. 62, no. 05.
[11] Smith, Clint, and Daniel Collins. 3g Wireless Networks. New York: McGraw-Hill, 2002.
[12] Khan, Farooq. Lte for 4g Mobile Broadband: Air Interface Technologies and Performance. .
[13] X. Lin, J. G. Andrews, A. Ghosh and R. Ratasuk, "An overview of 3GPP device-to-device proximity services," in IEEE Communications Magazine, vol. 52, no. 4, pp. 40-48, April 2014, doi: 10.1109/MCOM.2014.6807945
[14] Ruparelia, Nayan. Cloud Computing. .
[15] NFV, Network Functions Virtualisation. "ETSI GS NFV 001 V1. 1.1 (2013-10)." (2013).
[16] Kazmi, S. M. Ahsan, et al. Network Slicing for 5g and Beyond Networks.
[17] H. Sengar, D. Wijesekera and S. Jajodia, "Authentication and Integrity in Telecommunication Signaling Network," 2005 12th IEEE International Conference and Workshops on Engineering of Computer-Based Systems (ECBS'05), 2005, pp. 3-10, doi: 10.1109/ECBS.2005.18.
[18] Frenzel, Louis E. Handbook of Serial Communications Interfaces: a Comprehensive Compendium of Serial Digital Input/output (i/o) Standards. .
[19] Sanders, Geoffrey. Gprs Networks.
[20] Smith, Clint, and Daniel Collins. 3g Wireless Networks. New York: McGraw-Hill, 2002.
[21] Seidenberg, P., M. P Althoff, and Bernhard H Walke. Umts : the Fundamentals. Hoboken: Wiley [Imprint], n.20.
[22] S. Mavoungou, G. Kaddoum, M. Taha, and G. Matar, "Survey on threats and attacks on mobile networks," in IEEE Access, vol. 4, pp. 454-467, 2016.
[23] I. Singh, "Signaling Security in LTE Roaming," Master's thesis, Aalto University, Espoo, Finland, 2019
[24] B. T. Kotte. Analysis and Experimental Verification of Diameter Attacks in Long Term Evolution Networks. Espoo: Aalto University, 2016
[25] L. He, Z. Yan, and M. Atiquzzaman, "LTE/LTE-A network security data collection and analysis for security measurement: A survey," IEEE Access, vol. 6, pp. 4220-4241, Jan. 2018
[26] H. de Carballo Macedo, Luzia, O. H., & Campista, M. E. M. "Attacks to mobile networks using SS7 vulnerabilities: a real traffic analysis". p. 1-13. April 2023.
[27] Kim, H. "5G core network security issues and attack classification from network protocol perspective", p.10(2), 1-15. Jan 2020.
[28] Singla, A., Hussain, S. R., Chowdhury, O., Bertino, E., & Li, N. "Protecting the 4G and 5G cellular paging protocols against security and privacy attacks". Proceedings on Privacy Enhancing Technologies, 2020.
[29] S. Mavoungou, G. Kaddoum, M. Taha, and G. Matar, "Survey on threats and attacks on mobile networks," in IEEE Access, vol. 4, pp. 454-467, 2016.
[30] Karthigha, M., Latha, L., & Sripriyan, K. "A comprehensive survey of routing attacks in wireless mobile ad hoc networks". In 2020 International Conference on Inventive Computation Technologies (ICICT). p. 396-402. Feb 2020.
[31] R. Piqueras Jover and V. Marojevic, "Security and Protocol exploit analysis of the 5G specifications," IEEE Access, vol. 7, pp. 24956–24963, 2019.
[32] Eberspächer, J. Gsm : Architecture, Protocols and Services. 3rd ed., English lang. ed. Chichester, U.K.: Wiley, 2009.
[33] R. Buyya and A. V. Dastjerdi, Internet of Things: Principles and Paradigms, San Francisco, CA: Morgan Kaufmann Publishers, 2016. 171
[34] Zhang, Ying. Network Function Virtualization: Concepts and Applicability In 5g Networks.
[35] Mumtaz, Shahid, Jonathan Rodriguez, and Linglong Dai. Mmwave Massive Mimo : a Paradigm for 5g. First edition. .
<massive mimo>
[36] A. Khan, M. A. Khan, and S. A. Khan, "Proceedings of 4th International Bhurban Conference on Applied Sciences & Technology (IBCAST)," in IEEE Xplore, Islamabad, Pakistan, 4-7 Jan. 2007, pp. 1-2.
[37] B. Kamwendo, "Vulnerabilities of signaling system number 7 (SS7) to cyber-attacks and how to mitigate against these vulnerabilities," M.S. thesis, Univ. of the Witwatersrand, Johannesburg, South Africa, 2015.
[38] S. J. Mullender and A. S. Tanenbaum, "The design of a capability-based distributed operating system," The Computer Journal, vol. 29, no. 4, pp. 289-299, 1986.
[39] H. Sengar, D. Wijesekera and S. Jajodia, "Authentication and Integrity in Telecommunication Signaling Network," 2005 12th IEEE International Conference and Workshops on Engineering of Computer-Based Systems (ECBS'05), 2005, pp. 3-10, doi: 10.1109/ECBS.2005.18.
[40] G. Miller and C. Parsons, "Finding You: The Network Effect of Telecommunications Vulnerabilities for Location Disclosure," Citizen Lab Research Report No. 171, University of Toronto, October 2023
[41] A. Klinger, "Technical report on SS7 vulnerabilities and mitigation measures for digital financial services transactions," Report of Security Workstream, Security, Infrastructure and Trust Working Group, International Telecommunication Union, Geneva, Switzerland, 2020, pp. 1-31.
[42] A. Khan, M. A. Khan, and S. A. Khan, "Proceedings of 4th International Bhurban Conference on Applied Sciences & Technology (IBCAST)," in IEEE Xplore, Islamabad, Pakistan, 4-7 Jan. 2007, pp. 1-2.
[43] K. Kotapati, P. Liu, Y. Sun, and T. F. LaPorta, "A taxonomy of cyber-attacks on 3G networks," *Intelligence and Security Informatics*, p. 631–633, 2005.
[44] https://www.kroll.com/en/insights/publications/cyber/3g-practical-attacks-against-the-ss7-signaling-protocol
[45]https://www.techtarget.com/searchnetworking/definition/Signaling-System-7

TABLE OF FIGURES

APPENDIX

LITERATURE REVIEW

Mobile communication has revolutionized the way we interact with each other and how we lead our daily lives. Before the advent of second-generation (2G) digital networks, there was an array of pioneering first-generation (1G) networks upon

which the foundations of today's advanced communication systems were built. Cellular networks can be traced back to the 1940s in the United States with the deployment of mobile telephone networks that relied on a single, high-powered base station with an omnidirectional antenna [3]. Due to frequency band limitation, the number of calls that can be handled simultaneously is restricted. This system used a manual call connection by operators, which limited scalability and automation [3]. During the 1960s, Bell Labs conceptualized the approach, a massive breakthrough in mobile network capacity; this reused frequency allowed different devices within a large geographical area to use the same frequency without causing interference [3]. This led to creating multiple base stations with assigned coverage known as a cell. In the early days of cellular networks, the systems introduced were primarily national, with countries developing their standards. The Nordic Mobile Telephone (NMT) network stands out as it later supported roaming within Scandinavia, allowing mobile phone access across multiple countries [3]. Other examples were Advanced Mobile Phone Service (AMPS) in the United States and Total Access Communication Systems (TACS) in Great Britain. The first-generation systems had several challenges, including limited capacity, lack of interoperability due to no globalized standards, and the use of analog signals, which are vulnerable to noise and eavesdropping.

The second-generation (2G) mobile technology changed the second stage of mobile communication technology. This was a radical change from the first generation of analog systems into standardized digital systems that include networks; it offers enhanced bandwidth efficiency and helps the burgeoning expansion of mobile phone deployment considerably [3]. First implemented in 1991 by Radiolinja (currently part of Elisa Oyj), 2G networks utilize digital radio signalling of the cellular base stations towards the mobile network infrastructure, hence achieving better efficiency and security [3] [8]. One of the most dominant standards in managing the 2G network is the GSM (Global System for Mobile Communications) standard, which applies Time Division Multiple Access (TDMA) technology.
This gave way for voice and data services on a digital platform, introducing even the subscriber-identity module (SIM) for user anonymity and a secure way for billing, thereby getting rid of the vulnerabilities that the analog systems used to face [3][5]. It protected against radio channel disturbances, improved voice quality, and strengthened security through encryption. This put digital communication on the scene as a crucial point for future advancements in mobile networks [3]. GSM networks operated on dedicated frequency bands, such as GSM 900 and DCS 1800, which are globally reserved for GSM systems.
The bands used this way employed frequency hopping to mitigate interference, while additional bands like 1900 and 850 GSM bands were used in North America due to pre-allocated global GSM bands [3]. Its architecture was complete, with mobile equipment, base transceiver stations, controllers, mobile switching centers, and databases. Among these, voice calls, SMS, and data transport services were offered, along with supplementary services, including forwarding and barring [3].
Signalling protocols played a significant role, as they are essential in the function of GSM networks to interchange control information from the mobile stations to the network. Link Access Procedures on the **Dm** channel (LAPDm) is the protocol that carries the upper layer signalling messages. In the radio interface, layer 3 (RIL3) signalling protocols exist between the core and the user plane and are used by the signalling for radio resources, mobility, and call control [3].

It largely adhered to the layered protocol stacks of the Open Systems Interconnection (OSI) model, and its form and structure of using communication technology were glaringly evident. Another example is the adaptation by GSM to its requirements for mobile communication while clinging to the traditions then-prevailing: "The use of established telecommunication protocols, such as SS7 within the network's core [3]. The watershed moment was developing and deploying GSM as a 2G cellular system. Phases like the GSM Phase 2 and GSM Phase 2+ followed, conducted standard international adoptions, and developed further, pointing to their suitability and adaptability to new conditions. This set the foundation for the modern mobile-first world, where GSM can support different services and set up international roaming agreements [3]. This factor revolutionized the telecommunication era and laid down the steppingstone for other generations in mobile communication technology.

Before the introduction of the third generation (3G), there was an improvement in data to 2.5G (GPRS), 2.75G (EDGE), and 2.875 (EDGE Evolution) [5]. This improvement in data was first introduced: Universal Mobile Telecommunications Systems (UMTS) provided mutual authentication and higher-grade encryption, with which many users could communicate effectively and securely through this platform [5]. Securing mobile networks was a top concern in the 2G (GSM) days since basic encryption and user authentication mechanisms were introduced to protect the users' data and communications. However, it had many weaknesses that affected its security. Among other things are the need for end-to-end encryption, the insufficiency of its encryption algorithms, and vulnerability to various possible attacks. The raised vulnerabilities thus warranted more robust security and, therefore, gave birth to improvement in the subsequent generations.

With the coming of 3G networks, it was a significant leap in mobile security, with the ability to take care of many pitfalls not covered at the 2G stage. Universal Mobile Telecommunications System (UMTS) is a 3G standard that increases customer privacy through more robust encryption algorithms and mutual user-network authentication [3]. Among other aspirations, standardization effort documents like ITU-R M.687 focused on quality of service, global deployability, and compatibility with fixed networks [3]. However, it was not accessible from a host of challenges; the first was the vulnerability to new attacks arising from the increased data rates and the introduction of packet-switched technologies. The third generation (3G) provides a deep technical grounding and security considerations for UMTS. It handles standardization of the process and aims to assure that the supported voice and data services give mobile users the right quality of service as for

fixed networks [3]. This generation takes a more holistic view of security since new services have emerged, with data rates, among other factors. This means that the 2G to 5G transition continues to make efforts and improve security mechanisms for more secure and reliable mobile communications. However, each new generation brought considerable improvements in the emerging security mechanisms, which shows that the evolution is a running battle with new threats and vulnerabilities. 3G architecture had many weaknesses, which made this network an accessible platform for man-in-the-middle attacks.

The fourth generation (4G) networks, set on Long-term Evolution (LTE), came with a massive leap in mobile communication technology, with speeds up to 100Mbps. With an architecture designed to support high-quality video streaming, high-speed internet, and connectivity, the average user experience is improved. “The first-release LTE standard was commercially deployed in Oslo, Norway, and Stockholm, Sweden, in 2009” [9]. 4G LTE came with massive bandwidth speeds and network capacity; as of 2022, 4G accounted for 80% of mobile connectivity worldwide [9]. 4G LTE’s open platform introduced a new set of vulnerabilities to Distributed denial of services (DDOS) attacks and IP spoofing [5]. Because of these identified risks, LTE introduced a more robust security framework, such as stronger encryption, protection of identity privacy, and integrity protection mechanisms.

4G technology is a real thrust toward furthering mobile communication due to its speed, capacity, and connectivity compared to the earlier technologies. Showing continuous improvement, as reflected in the network’s move to an all-IP-based architecture. This was an achievement in the sense that it primarily improved data throughput and network efficiency. The architecture brought about by the 4G network would bring in aspects of high-quality video streaming, fast internet, and connectedness in a bid to improve the experience of the average user.

4G refers to the fourth generation of mobile communication, marking a massive change toward IP-based services, very high data rates, and enhanced spectral efficiency. These systems, including 4G, use the base stations of this network, based mainly on Long Term Evolution (LTE) and WiMAX technologies, providing their users with data rates up to 1 Gbps for stationary and 100 Mbps for mobile users [2]. It would be a great speed and, as such, can enable many possibilities: high-quality video streaming, advanced gaming, and even live communication. Among some outstanding features, technologies such as the use of MIMO (Multiple Input Multiple Output) significantly increase data throughput and network capacity [3]. 4G networks also use OFDM (Orthogonal Frequency Division Multiplexing) and OFDMA (Orthogonal Frequency Division Multiple Access) in the process of their operation while trying to ensure that the frequency spectrum is used duly and effectively at high data rates, though some attention has been paid toward keeping low latency. Despite this, the development of 4G technology is experiencing some restraints, such as frequency spectrum scarcity and interoperability with already existing infrastructure at the level of the network. This 4G technology also demands colossal resources for the procurement of new network equipment and technologies against the economic effectiveness and feasibility of 4G services, most especially in remote areas [9]. The 4G technology has filled a spot in the evolution of mobile networks towards the possibility of 5G because it has been a basis that permits further innovations in wireless communication. That is, it underscores high-speed data connectivity and supports a range of IoT (Internet of Things) devices in the generations that assure much better speeds, less latency, and reliable connections [9]. 4G technology has had a significant impact on mobile communication, providing unprecedented data speeds, improved network efficiency, and the ability to support a range of applications. However, its application and adoption in an area face technical, economic, and regulatory challenges. 4G still remains an essential area of research and development for the provision of ubiquitous, high-speed mobile connectivity bridging to 5G and beyond.

The rollout of 5G is expected to provide a secure, resilient, and flexible mobile network infrastructure to support the next wave of digital transformation.

As the capabilities and applications of mobile networks increased, so did the vulnerabilities. 5G networks with data transfer rates of up to 1Gbps per second were launched by major networks in 2019 [10]. 5G still needs to be fully rolled out everywhere in the world, but it is currently being used at an increased price in developed nations across Europe and North America. 5G architecture is designed to make it even more flexible and efficient, using technologies like Network Functions Virtualization (NFV) and Software Defined Networking (SDN) to manage its network dynamically [5] with the introduction of Massive Multiple-input Multiple Output (MIMO) and Orthogonal Frequency Division Multiplexing (OFDM) in 5G [2]. The fundamental difference between 5G and its predecessors is that it is designed to operate on a broader spectrum, ranging between 3 to 300 GHz, with speeds up to 1Gbps or even higher [2]. It has the potential to support an even more comprehensive array of devices, sustaining multiple connections at ultra-high-speed data transfer rates while maintaining very low latency.

Mobile wireless communication began with 1G, an analog communication network that was meant for voice communication only. Later, 2G came into play in which, apart from digital communication, text messaging services were also made possible. 3G and 4G technologies increased the rate of data transmission, supported multimedia services, and improved the quality of service and bandwidth, hence enabling a vast array of Internet-based applications and services. As we stand at the threshold of the 5G revolution, much betterment and transformation in wireless communication are promising because they provide higher speed, low latency, and the capability to connect many devices simultaneously [10].

5G technology is envisaged to become a keystone for the future digital society with more advanced features such as higher data rates, energy efficiency, cost reduction, and significant improvements in connectivity [2][6][10]. The other constraints and challenges that 5G aims to address include those in 4G, such as coverage, flexibility, and interconnection, to support an ever-growing demand for data and connectivity [2]. 5G technology, on the other hand, is expected across smart cities, automated

industries, and digital health care. It will have its footprint in enabling smooth communication of devices and sensors across the Internet of Things.

One of the defining features of 5G technology is its reliance on higher frequency bands, ranging from 3 GHz to 300 GHz. In effect, it is this shift to higher frequencies, the millimeter waves, as they are called, that is the critical change necessary for attaining the high data rates and low latency promised by 5G. However, with this increase comes new problems, notably in terms of network coverage and signal reliability, since at higher frequencies, electromagnetic signals will tend to have shorter ranges and will be more prone to being blocked by physical obstacles [6][10]. The 5G network will deploy superior technology so that it can offer sufficient signal coverage and capacity and efficiently use the spectrum. In this case, these will exploit Massive MIMO (Multiple Input Multiple Output), beamforming, and small cell technologies [2][6].

Despite the technological advances, deploying 5G networks faces several hurdles. Another big challenge has been in infrastructure development, which requires substantial investment and spectrum allocation approval by regulators. In regard to the above observation, the 5G network will have to transit from the current centralized model to that of a decentralized nature, which caters to different requirements required by a myriad of applications and services [2][6][10]. It is, therefore, relevant to secure interoperability with all the generations of mobile technologies to enable global standardization towards the smooth adoption and success of 5G. The potential applications of 5G technology are vast and varied, extending well beyond enhanced mobile broadband. 5G is expected to enable breakthrough applications in telemedicine, driverless cars, and virtual and augmented reality. 5G has been characterized by ultra-reliable, low-latency communication in that it will allow for real-time device and system controlling and monitoring, therefore building efficiencies and spurring innovations in various segments. The progression towards 5G constitutes a critical threshold in the development of mobile wireless communication, with the potential to revolutionize telecommunication, among other important societal and economic sectors. The deployment and adoption of the 5G networks will be successful only if a host of daunting technical, regulatory, and economic challenges is surmounted as the technology underpinning 5G races to keep pace with the burgeoning demand for reliable and fast connectivity. They argue that this, however, faces a continual effort in research and development without which the industry stakeholders, in collaboration with regulators and governments, cannot reap to the fullest the benefits offered by this technology [2][6][10].

The slow march of mobile communication technologies from 1 G's basic analog systems to the promised 5G and beyond, an ultra-high-speed, fully interlaced world of the future is symbolic of humanity's journey, marked by both the pioneering achievements and formidable challenges, points not only towards our technological ambitions but also to the way our society has had to evolve. We must take a look at our past as we look into the future; when the promise of 5G is the great catalyst of the new wave of digital transformation, we also need to take a look back over what brought us here, looking ahead with a sober view that both opportunity and responsibility lay in front of us.

The underlying development is continuous pressure for more efficiency, security, and accessibility of mobile networks. Every leap in generation has brought with it faster speeds and more robust connectivity and opened new digital landscapes for innovation. That has moved from simple voice 1G networks to multimedia-rich possibilities of 4G, which are nothing less than a revolution and make a digital economy possible, where information is currency, and connectivity is the spinal cord.

Nevertheless, even as we stand at this 5G dawn, whose transformative promise of ultra-reliable, low-latency communications for applications such as telemedicine, autonomous vehicles, and immersive augmented and virtual reality experiences cannot be better realized. The heft of responsibility could not have been greater. 5G networks will depend on excellent technology for deployment and require a rethink of the network infrastructure. Carefully balanced between regulatory and economic considerations, this should be coupled with an understanding of ensuring commitments toward equitable access to the benefits of this new wave of connectivity.

Such an evolution toward 5G and the hope of many more mobile communication generations are drivers underpinning the paramount importance of cybersecurity and data privacy on the front line in such a connected world. With the complexity and increasing importance of networks in our daily lives, never before have the protection of user's data and the resilience of the communication infrastructure been so high. This begets a united call from global industry, policymakers, and the public to forge standards and practices that protect our digital future.

As we write the following chapters of the mobile communication saga, it is incumbent on us to do so with celebration not only of the technological marvels that have brought us to this point but also with guidance through the shoals of technical, ethical, and societal minefields that are going to define how the world will connect in the future. The migration from 2G to 5G and beyond is much more than a story of technical invention; it is a saga of human ambition, creativity, and the insatiable drive to connect, understand, and make the world around them better.

While we look to the horizons 5G will reveal, we must ensure its use is a force for good: bridging divides and leveraging sustainability in an inclusive, connected, and resilient society. The advent of mobile communication is a vivid reminder of what we can realize when we dare to think of something that is supposed to be impossible yet try ever so hard to make it possible for a better connected and inclusive future.

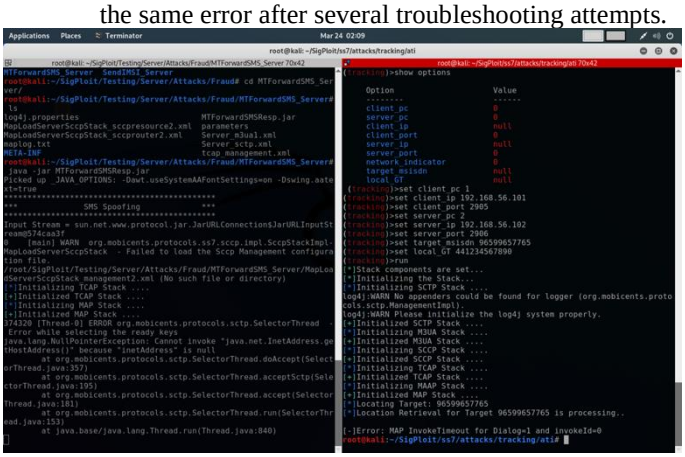
More SS7 Implementation Attacks

Location Tracking Attack

There are also attacks that exploit vulnerabilities on user privacy and these also leverage on the lack of authentication. An attacker with network access could breach the privacy of subscribers

within the environment and this can be done through the use of AnyTime Interogation environment. All subscriber information which includes their location is stored in the HLR (Home Location Register). An attacker only needs to draft an intended spoofing message asking the HLR for this information and it will tell them because there is no authentication of the sender [44]. During our simulation, we attempted this attacke using AnyTime Interogation but we encountered an error message which we were unable to resolve after several attempts and research.

1. To execute this exploit, we launch the simulation binary in the directory as seen below:
   cd SigPloit/Testing/Server/Attacks/Fraud/MTForwardSMS_Server/
   java -jar MTForwardSMSResp.jar
2. Next, we launch the attack by executing the AnyTime Interogation JAR file on SigPloit:
   cd SigPloit/ss7/attacks/tracking/ati/
   Java -jar AnyTimeInterrogation.jar
3. Just like the SMS routing attack, the following parameters would need to be set, and we will also make use of the existing hardcoded IP addresses and ports. The MSISDN here represents the victims phone number while the local_GT parameter serves as the attacker's destination.
   set client_pc 1
   set client_port 2905
   set client_ip 192.168.56.102
   set server_pc 2
   set server_port 2906
   set server_ip 192.168.56.102
   set target_msisdn 96599657765
   set local_GT 441234567890
4. Once these parameters are set and ready, enter 'run' for the attack to simulate however, we kept encountering the same error after several troubleshooting attempts.



Interception Attack

Similar to location tracking where user privacy is breached, intercepting voice and SMS texts via ss7 follows the same concept. A spoofed message is sent to the Home Location register by an attacker which provides the location of the given user. In this attack, whenever the target receives a text or call, they will be routed to the attacker and this is also the basis of the vulnerability existing in SMS based multi-factor authentication as opposed to application based multi-factor authentication [44]. For this simulation, we made use of the jss7-attack-simulator and Wireshark however, it is worth npting that we were not able to fully replicate this attack due to suspected obsolated software versions for openjdk-8 but the documentations are seen below.

5. Before running the jss7-simulator, we start up Wireshark with administrative privileges and configure it to listen on the local adapter. We also added a filter for the sccp traffic to streamline the visual entries.
6. Next, we execute the attack with the commands below:
   cd restcomm-jss7-${jSS7.release.version}/ss7/restcomm-ss7-simulator/bin
   java -jar run.sh attack_simulator -a simple -m intercept:sms
7. For this attack to work, the attacker requests for an update of the targets location from the HLR which is done with an updateLocation request.
   The simulated targets IMSI identity is given as 24201111111110. A sppofed text is then sent to the vistim and the local MSC would reach out to the HLR and request for the victim's location but because this has already been tampered with by the attacker, the HLR will send the MSC the attacker's location in place of the victim. The MSC then routes the request to the attacker's location and subsequently, the attacker will be able to take and receive the targets calls and messages. Below is an error message gotten while trying to implement this attack [44].