DIPLOMA THESIS

"Security and Privacy in the Air interface of cellular networks"

Dimitrios Manos MTE1621

Supervisor: Prof. Christos Xenakis

Table of Contents

Table of Figures

Acronyms

3GPP Third Generation Partnership Project
AES Advanced Encryption Standard
AK Anonymity Key
AKA Authentication and key agreement
AMF Authentication management field
AN Access Network
ASME Access Security Management Entity
AUTN Authentication Token
AV Authentication Vector
BTS Base Transceiver Station
Cell-ID Cell Identity
CK Cipher Key
CKSN Cipher key sequence number
CS Circuit Switched
ECM EPS Connection Management
EEA EPS Encryption Algorithm
EIA EPS Integrity Algorithm
eKSI Key Set Identifier in E-UTRAN
EMM EPS Mobility Management
eNB Evolved Node-B
EPC Evolved Packet Core
EPS Evolved Packet System
EPS-AV EPS authentication vector
E-UTRAN Evolved UTRAN
GERAN GSM EDGE Radio Access Network
E-UTRAN Evolved Universal Terrestrial Radio Access Network
ETSI European Telecommunications Standards Institute
GERAN GSM/EDGE Radio Access Network
GUTI Globally Unique Temporary UE Identity
HE Home Environment
HLR Home Location Register
HSS Home Subscriber Server
IK Integrity Key
IMEI International Mobile Station Equipment Identity
IMEISV International Mobile Station Equipment Identity and Software Version number
IMSI International Mobile Subscriber Identity
KDF Key Derivation Function
KSI Key Set Identifier
KSS Key Stream Segment
LAI Location Area Identity
LSB Least Significant Bit
LSM Limited Service Mode
LTE Long Term Evolution
MAC The message authentication code included in AUTN, computed using f1
MME Mobility Management Entity
MME-RN MME serving the RN
ME Mobile Equipment
MS Mobile Station

MSIN Mobile Station Identification Number
MSC Mobile Services Switching Centre
NAS Non Access Stratum
NAS-MAC Message Authentication Code for NAS for Integrity (called MAC in TS24.301 [9])
NCC Next hop Chaining Counter
NH Next Hop
PDCP Packet Data Convergence Protocol
PLMN Public Land Mobile Network
PS Packet Switched
PSK Pre-shared Key
P-TMSI Packet-TMSI
Q Quintet, UMTS authentication vector
RAI Routing Area Identifier
RAND Random challenge
RN Relay Node
RRC Radio Resource Control
SEG Security Gateway
SQN Sequence number
SQNHE Individual sequence number for each user maintained in the HLR/AuC
SQNMS The highest sequence number the USIM has accepted
SGSN Serving GPRS Support Node
SIM (GSM) Subscriber Identity Module
SMC Security Mode Command
SN Serving Network
SN id Serving Network identity
SRVCC Single Radio Voice Call Continuity
T Triplet, GSM authentication vector
TMSI Temporary Mobile Subscriber Identity
UE User Equipment
UEA UMTS Encryption Algorithm
UIA UMTS Integrity Algorithm
UICC Universal Integrated Circuit Card
UMTS Universal Mobile Telecommunications Systems
UP User Plane
USIM Universal Subscriber Identity Module
UTRAN Universal Terrestrial Radio Access Network
VLR Visitor Location Register
XRES Expected Response

1. Introduction

Cellular telephony is a part of our everyday life for more than 20 years. During this period mobile devices that begun as heavy handheld transponders in 1G systems have been turned to smart devices built on small powerful computers in 4G systems.

This technological transformation in conjunction to the parallel networking revolution through Internet's expansion provided to the telecom industry the competitive advantage of producing devices that form not just a communication platform but even more a means of information, of financial exchanges and entertainment. The involvement of mobile telephony networks in our networked social-economic world is obvious and in this context the importance of protecting these networks against potential threats is more than critical.

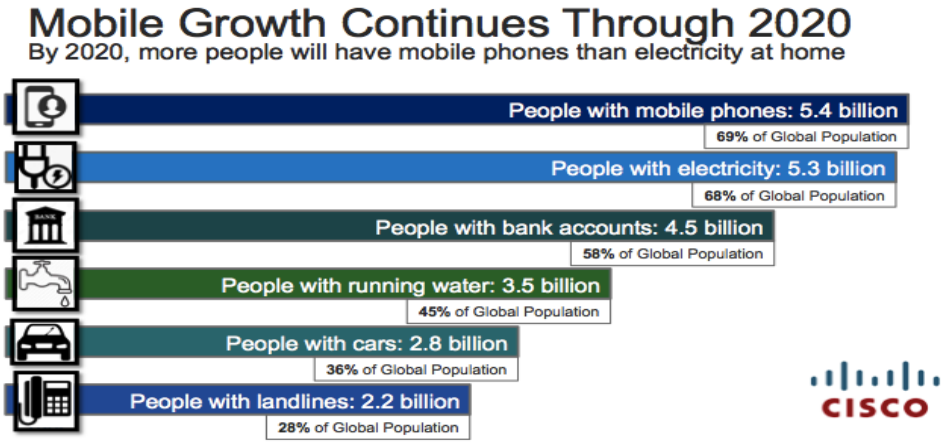

Figure 1 Mobile growth

In this thesis we will focus on the radio Air Interface between the user's equipment (UE) and the access network (AN) of the mobile operator's network. This interface shows high interest in terms of security due to its wireless nature; since every mobile user is immediately exposed.

All the critical information regarding security and privacy such as identification and authentication parameters for both the device and the user, the security parameters for establishing the communication channel as well as the user' mobility data and the voice data itself, are transferred through the Air Interface.

Consequently, the study of implementing the communication in this link can reveal crucial issues of protecting the security and privacy of the user. In the context of our work, we have studied the evolution of implementing the communication in Air Interface for the basic mobile generations in use (2G, 3G, 4G).

The purpose of this thesis is to study the implementation of the Cellular Network Access Security and the deviations that took place during the mobile technologies' evolution. In this work, we focused on the basic mobile technologies for each generation: GSM for 2G, UMTS for 3G and LTE for 4G.

2. Problem confrontation

While wireless communications provide great flexibility and mobility, they often come at the expense of security. Indeed, wireless communications rely on open and public transmission media that raise further vulnerabilities in addition to the security threats found in wired networks. A number of specific open issues and even inherent dangers are yet to be solved. With wireless communications, important and vital information is often placed on a mobile device that is vulnerable to theft and loss. In addition, this information is transmitted over the unprotected airwaves.

The problems begun from first generation (1G) analog mobile phones that relied on an electronic serial number to confirm that the terminal should be allowed access to the service. Thus, in 2G systems were designed with security in mind. Each subscriber to a 2G service receives a Subscriber Identity Module (SIM) card which contains the user's identity and a long-life authentication key (technically speaking, a shared secret key) supposed to last for the whole duration of the subscription. The SIM is a removable security module which is issued and managed by the users' home service operator (even when the user is roaming) and is independent of the terminal. SIM-based authentication does not require any user action, other than entering the familiar 4-digit Personal Identification Number (PIN) into the terminal.

Global System for Mobile Communications (GSM), perhaps the best known 2G system, provides a range of security features, including authentication of the mobile user to the network, data confidentiality across the air interface, and a degree of user pseudonymity through the use of temporary identities. Third and fourth generation (3G and 4G) systems, such as UMTS/3GPP and Long-Term Evolution (LTE), have enhanced these security features, however, user privacy protection has remained largely unchanged.

This 2G technology has created an avenue for the development of 3G and 4G cellular network standards such as Universal Mobile Telecommunications System (UMTS) and Long Term Evolution (LTE). With the deployment of new technologies still on the way, the era beyond 3G and 4G requires that cellular and radio technologies need to work together forming highly heterogeneous networks.

The problem we want to address has to do with possible security and privacy leakages due to protocols' design. Under this scope we study the evolution of implementing the Air Interface in the most used systems of the three major mobile technologies, particularly GSM for 2G, UMTS for 3G and LTE for 4G.

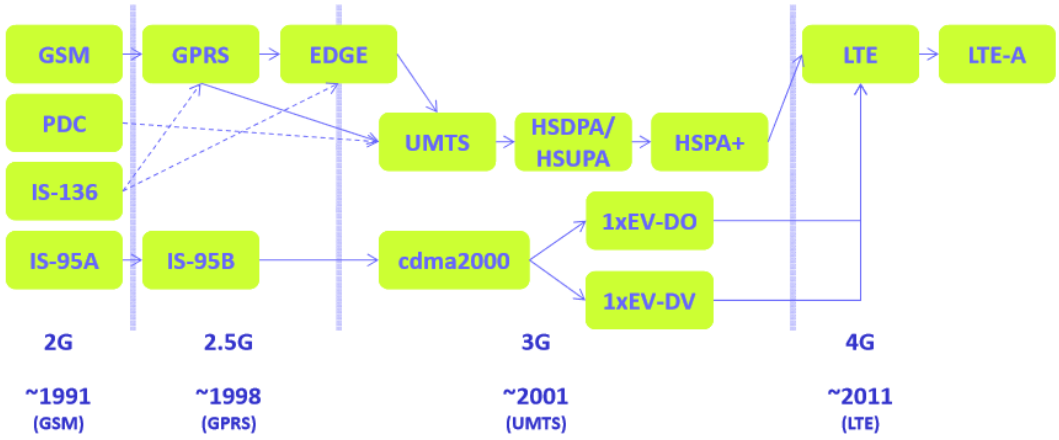

Figure 2 Evolution of cellular technologies

3. Definitions

A brief introduction is needed in order to obtain the core security modules used in cellular systems. To begin with, a complete mobile phone is referred to as a user equipment (UE), where the term encapsulates not only the mobile equipment (ME) or mobile station (MS), i.e. the phone, but also the subscriber's identity module (SIM or USIM) within it, where the (U)SIM takes the form of a cut-down smart card. The (U)SIM embodies the relationship between the human user and the issuing home network, including the International Mobile Subscriber Identity (IMSI), the telephone number of the UE (MSISDN), and other user (subscriber's) data, together with a secret key shared with the issuing network which forms the basis for all the air interface security features.

The Third Generation Partnership Project (3GPP) plays a vital role in our discussion, since 3GPP is the worldwide standards development organization who is responsible of the security and privacy in cellular systems, specifying the security architecture and protocols. 3GPP unites six telecommunications standard development organizations Association of Radio Industries and Businesses (ARIB-Japan), Alliance for Telecommunications Industry Solutions (ATIS-USA), China Communications Standards Association (CCSA-China), European Telecommunications Standards Institute (ETSI-Europe), Telecommunications Technology Association (TTA-Korea), and Telecommunication Technology Committee (TTA-Japan), known as "Organizational Partners" and provides their members with a stable environment to produce the highly successful Reports and Specifications that define 3GPP technologies. The major cellular technologies evolved by 3GPP are:

- GSM
  Security was a major driver for the success of GSM. Specifications were developed to prevent terminal equipment theft, to allow encryption and authentication, to control payment for copyright material downloading and to respond to many other security threats.

- UMTS
  The extension of GSM into GPRS/EDGE allowed the introduction of Internet-based technologies, which were further utilized in the next mobile generation, UMTS. The UMTS security specifications developed in 3GPP built on the mechanisms used in GSM. In addition, they offered numerous security enhancements, including: Authentication, public safety, location services, cell broadcast services, IP Multimedia Subsystem (IMS) and Selective disabling of user equipment.

- LTE
  The LTE networks provide consistent Internet Protocol connectivity between the end user and the network, resulting in the redesign of the whole security architecture to provide more robustness. LTE improves greatly the security of UMTS with stronger cryptographic algorithms for a more secure connection, and introduces a new SIM card as foundation of its security architecture.

3.1 3GPP Security Areas

According to 3GPP, five security feature groups are defined. Each of these feature groups meets certain threats and accomplishes certain security objectives:

1. Network access security (I): the set of security features that provide users with secure access to services, and which in particular protect against attacks on the (radio) access link.

2. Network domain security (II): the set of security features that enable nodes to securely exchange signaling data, user data (between AN and SN and within AN), and protect against attacks on the wireline network.

3. User domain security (III): the set of security features that secure access to mobile stations.

4. Application domain security (IV): the set of security features that enable applications in the user and in the provider domain to securely exchange messages.

5. Visibility and configurability of security (V): the set of features that enables the user to inform himself whether a security feature is in operation or not and whether the use and provision of services should depend on the security feature.

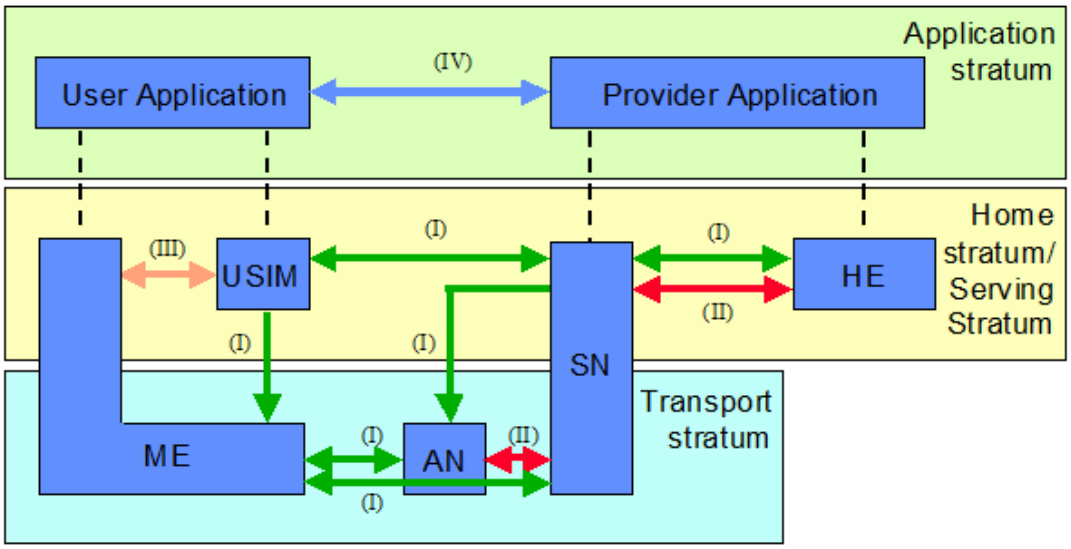

Figure 3 Overview of 3GPP security architecture

In this study we will focus on the Network Access domain where the radio interface is implemented. We will examine the implementation of the Over The Air communication in terms of the security and privacy aspects. More specifically we will study Entity Authentication and User's Privacy through the evolution of 2G, 3G and 4G cellular networks.

Before proceeding we need to establish some terminology. Network access security features can be further classified into the following categories:

1. entity authentication

2. confidentiality

3. data integrity

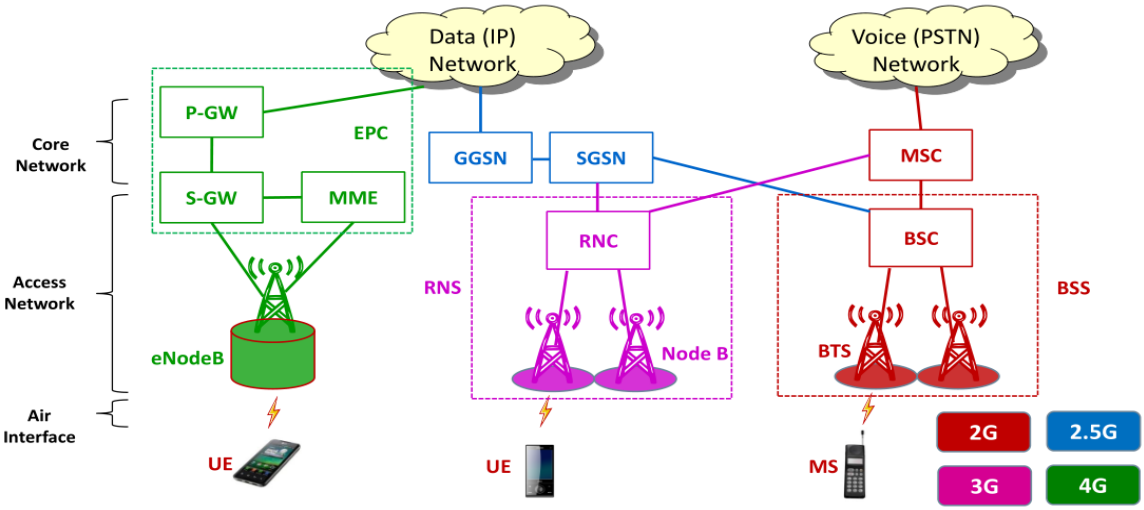

Figure 4 Simplified Cellular Network Architecture for 2G, 3G, 4G

3.2 Entity Authentication

Authentication is to verify everyone is who they claim to be. Authentication is performed via Authentication and Key Agreement Procedure (AKA). AKA protocol implements this procedure and is at the core of mobile telephony air interface security. AKA can be triggered by the initial network attach request, by the Routing Area Update (RAU) request, or by the service request.

This procedure is regularly performed between the entities of access network and the user equipment. The involved parties are the core network (CN) -that issued the subscriber's security module of (U)SIM-, the access network (AN) and the user equipment (UE). The Authentication Center (AuC) of the core network generates authentication vectors (used by the access network in AKA) and sends them to the access network. The AKA protocol starts with the access network sending a user authentication request to the UE. Then UE checks -in mutual authentication systems 3G and 4G- the validity of this request (thereby authenticating the network), and then sends a user authentication response. The AN checks this response to authenticate the UE. As a result, if successful, the UE and the network have authenticated each other, and at the same time they establish two shared secret keys. In order to participate in the protocol, the UE, in fact the USIM installed inside the UE, must possess two values:

1. a long term secret key K, known only to the USIM and to the USIM's home network
2. a sequence number SQN, maintained by both the USIM and the home network -in 3G and 4G-

In 2G, we only have UE authentication and there is no SQN number; whereas in 3G & 4G, we perform mutual authentication to verify the handset as well as the base station and SQN number is used in order to identify replay attacks.

The following is a description of the security features classified into the category of entity authentication:

- User authentication: The property that the network that provides the service (serving network) corroborates the identity of the user.

- Network authentication: The property that the user corroborates that he is connected to a serving network that is authorized by the user's home network to provide him services; this includes the guarantee that this authorization is recent.

The following security features deal with the confidentiality of data on the network access link:

- Cipher algorithm agreement: The property that the mobile station and the serving network can securely negotiate the algorithm that they shall use subsequently.
- Cipher key agreement: The property that the mobile station and the serving network agree on a cipher key that they may use subsequently.
- Confidentiality of user data: The property that user data can not be overheard on the radio interface.
- Confidentiality of signaling data: The property that signaling data can not be overheard on the radio interface.

The features provided to achieve integrity of data on the network access link are the following:

- Integrity algorithm agreement: The property that the mobile station and the serving network can securely negotiate the integrity algorithm that they shall use subsequently.
- Integrity key agreement: The property that the mobile station and the serving network agree on an integrity key they may use subsequently.
- Data integrity and origin authentication of signaling data: The property that the receiving entity (mobile station or serving network) is able to verify that signaling has not been modified in an unauthorized way since it was sent by the sending entity (serving network or mobile station) and that the origin of the signaling data received is indeed the one claimed.

3.3 User's Privacy

User's privacy has to do with the disclosure of his/her identity. In the cellular world each Mobile device contains an IMEI (International Mobile Equipment Identity) as well as the (U)SIM card that contains an IMSI (International Mobile Subscriber Identity). During the operation, IMSI has to be hidden with help of temporary identities in order to provide:

- user identity confidentiality
- user location confidentiality
- user untraceability

An IMSI is a 15-digit decimal number (see Fig. 5). Of the 15 digits, the first three form the mobile country code (MCC). The next two or three digits identify the network operator, and are known as the mobile network code (MNC). The length of the MNC, i.e. whether it contains two or three digits, is a national matter. The remaining nine or ten digits, known as the mobile subscriber identification number (MSIN), are administered by the relevant operator in accordance with the national policy. IMSIs therefore have geographical significance, and their use is typically managed by the network operator in blocks. The

combination of the MCC and the MNC can be used to uniquely identify the core network of the IMSI. The MSIN is used by the operator to identify the subscriber for billing and other operational purposes. Each IMSI uniquely identifies the mobile user, as well as the user's core network and home country.

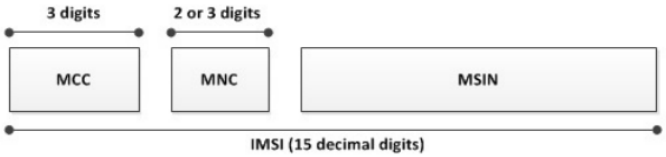

Figure 5 IMSI structure

The core network also assigns a temporary mobile subscriber identity (TMSI) and sends the TMSI to the UE in encrypted form. The TMSI is unique to the location area in which the subscriber is currently located. Accordingly, whenever the subscriber visits a new location area, the core network must update the TMSI value.

In 2G, 3G and 4G the following temporary identities have been introduced:

- 2G:
  - TMSI (Temporary Mobile Subscriber Identity)
- 3G:
  - P-TMSI (Packet TMSI)
- 4G:
  - GUTI (Globally Unique Temporary UE Identity)

IMSI is used to identify the subscriber for authentication and access provision, limiting the degree to which its use compromises user privacy is the main focus of this paper. When a subscriber is roaming, i.e. accessing service from a network other than its core network, the IMSI is sent from the UE via the visited network to the home core network. Since the IMSI is a permanent user identity, the air interface protocols are designed to minimize the number of circumstances in which it is sent across the air interface.

Providing user privacy requires that the permanent user identity cannot be intercepted when sent across the radio link. A level of identity confidentiality is provided by use of the TMSI instead of the IMSI. However, on certain occasions a UE needs to send its IMSI across the air interface in clear text. One such case is when a UE is switched on and wishes to connect to a new network, and hence will not have an assigned TMSI. Another case is where the serving network is unable to identify the IMSI from the TMSI.

3.4 Subscriber Identity Module

The aforementioned sensitive parameters (keys, identities) are being kept in the Subscriber Identity Module, a smart card that accompanies every user of the cellular network. The USIM data storage capabilities are specified in section 10.1 of 3GPP TS 21.111 [5]. Information held within the USIM is stored in files, which can be divided into the following categories: application dedicated files (ADFs), dedicated files (DFs) and elementary files (EFs). The IMSI is stored in the USIM and is normally fixed. The elementary file EF IMSI contains the value of the IMSI.

Subscriber's Identity Module has evolved through the cellular generation in order to achieve higher security performance and storage capacity. In the figure 7 below we present the evolution of SIM to USIM and ISIM for 3G and 4G. Through the technologies evolution SIM

has evolved to USIM and IMS SIM (ISIM), which are the security and subscription applications for 3G and 4G technologies, running on the smartcard (Universal Integrated Circuit Card -UICC). This module could be on a UMTS 3G or IMS LTE network. It contains parameters for identifying and authenticating the user to the IMS. The ISIM application can co-exist with SIM and USIM on the same UICC making it possible to use the same smartcard in both GSM networks and earlier releases of UMTS.

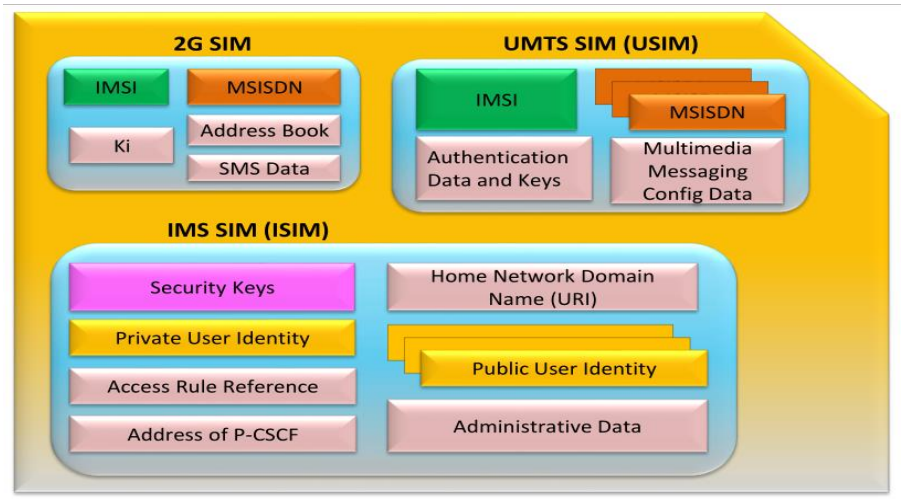

Figure 6 SIM evolution over 2G, 3G, 4G

4. Air interface of 2G, 3G and 4G cellular networks

In order to study the network access security techniques for mobile networking we will pass through the air interface protocols that implement the radio interface. In accordance to the OSI reference protocol stack, radio interface in the cellular network of GSM, which forms the base of all the next technologies, is divided into three protocol layers:

Layer 1 - the physical layer (L1)

Layer 2 - the data link layer (L2)

Layer 3 - network layer (L3)

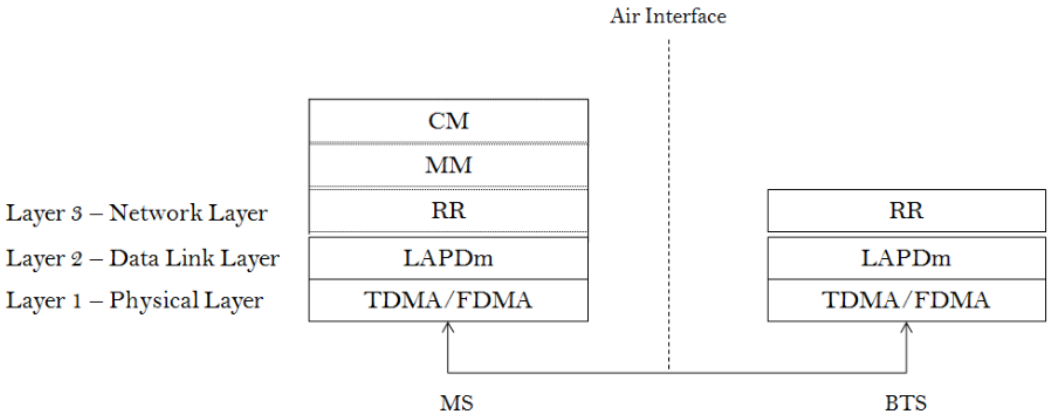

Figure 7 Air Interface layers (GSM)

First of all we have to distinguish Logical and Physical channels in order to study the radio interface implementation. Physical channels are the time slots at given frequencies and have to do with modulation, slot synchronization, multiple access techniques, duplexing, and frequency hopping issues. Logical channels are built on top of physical channels and carry the information that is exchanged between User's Equipment and Access Network, thus they are divided in traffic and control (signaling) channels. Signaling at the Air interface is mainly controlled by Layer 3, the Network Layer. Layers 1 and 2, Physical and Data Link Layer, provide the mechanisms for the protected transmission of signaling messages sent on Layer 3.

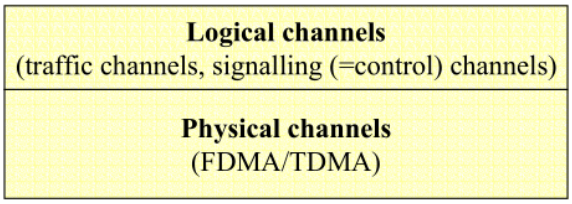

Figure 8 Logical vs Physical channels

In GSM, Layer 3 contains all the functions to establish, maintain and terminate mobile connections as well as control functions for additional services. Messages in Layer 3 are divided into different sublayers, namely, Radio Resource management (RR), Mobility Management (MM) and Connection Management (CM), as depicted in figure 8 above. In UMTS and LTE, Layer 2 is split into following sublayers: Medium Access Control (MAC), Radio Link Control (RLC), Packet Data Convergence Protocol (PDCP) and Broadcast/Multicast Control (BMC). Furthermore, Layer 3 and RLC are divided into Control (C-) and User (U-) planes. PDCP and BMC exist in the U-plane only. In the C-plane, Layer 3 is partitioned into

sublayers where the lowest sublayer, denoted as Radio Resource Control (RRC), interfaces with layer 2 and terminates in the radio access network.

4.1 2G GSM Air interface

We start by considering the widely used 2G system, GSM. The protocol stack of GSM is given below:

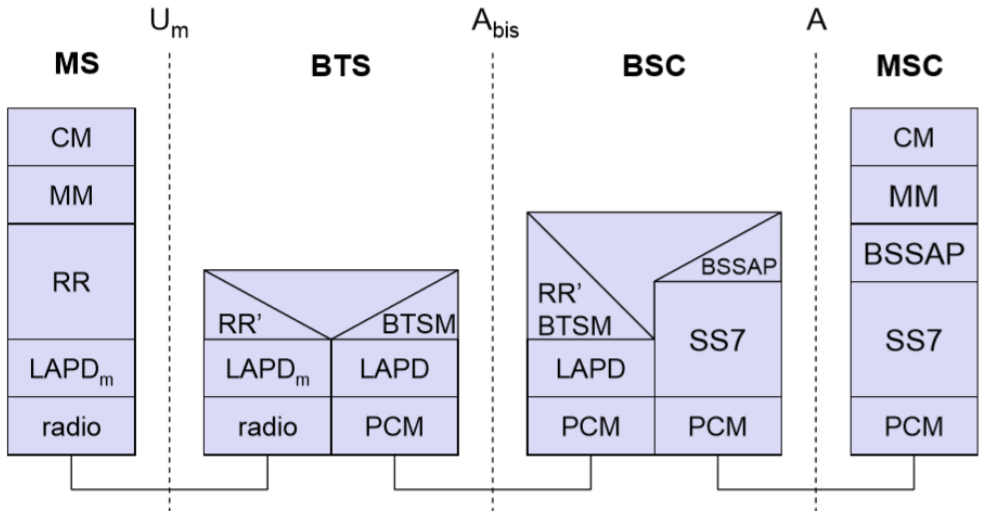

Figure 9 GSM protocol stack

The air interface between the MS and the Base Transceiver Station (BTS) is called Um. The GSM air interface is based on time division multiple access (TDMA) with frequency division multiple access (FDMA). TDMA allows multiple users to share a common RF channel on a time-sharing basis, while FDMA enables different frequencies to be used in uplink (MS to BTS) and downlink (BTS to MS) directions. Most of the implementations use a frequency band of 900 MHz. The other derivative of GSM is called Digital cellular system 1800 (DCS1800).

The radio interface is implemented between MS and BTS, of course Base Station Controller (BSC) and Mobile Switching Center (MSC) take part in the security/privacy procedures, all these elements are depicted below in the overall GSM network architecture.

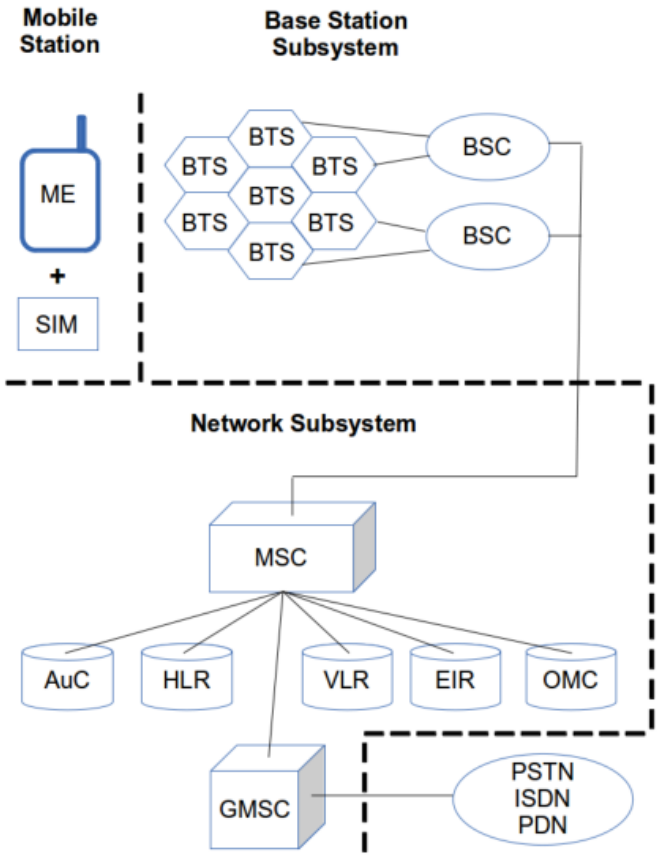

Figure 10 GSM network architecture

4.1.1 Security of 2G GSM Air Interface

According to the relevant technical specification for GSM security (GSM 02.09) by ETSI, GSM is designed to provide the following security features:

- subscriber identity (IMSI) confidentiality;
- subscriber identity (IMSI) authentication;
- user data confidentiality on physical connections;
- connectionless user data confidentiality;
- signaling information element confidentiality.

In the following paragraphs we will see the implementation of these features in Air Interface.

4.1.1.1 GSM Authentication and Key Agreement procedure

The authentication process is implemented via AKA procedure and can be separated in four steps, as presented in figure 11. It begins with the MS requesting access to the network by broadcasting the IMSI/IMEI for identification purposes. This usually occurs when the MS is turned on for the first time or if the MS has been turned off for more than 48 hours. The MSC receives the IMSI/IMEI and sends the IMSI to the Home Location Register (HLR) with a request for an authentication triplet. The IMEI is also sent to the EIR to check its database to make sure the number is not on the black list. The HLR likewise checks the IMSI with its database to ensure the user is a valid subscriber. Once the IMSI/IMEI have been verified the

HLR will send the triplet request along with the IMSI to the AuC. The AuC is the only other place beside the SIM card that stores the 128-bit individual subscribers authentication key, Ki , associated with each IMSI. The IMSI and SIM card are created together when the SIM card is produced. The AuC uses the IMSI to look up the individual subscribers Ki and then produces a 128-bit Random Number (RAND). The AuC also contains the authentication and ciphering key algorithms referred to as A3 and A8 respectively. The RAND and Ki are inputted into the A3/A8 algorithms producing a 32-bit Signed Response (SRES) for authentication and a 64-bit ciphering key Kc used for encryption of data. The SRES is considered the response of the challenge to the MS. The RAND , SRES, and Kc are referred to as the triplet. To reduce the load of authentication requests to the AuC, the AuC produces a set of five triplets at a time and sends them to the MSC for storage. When the MS authenticates again the MSC will use one of the unused stored triplets before requesting another triplet from the AuC. Note that each triplet is only good for the IMSI it was generated for. The MSC chooses one of the five triplets and then sends the RAND or challenge to the MS. The MS uses the Ki stored on the SIM card along with the RAND received from the MSC as inputs into the A3/A8 algorithms located on the SIM card. The A3/A8 algorithms produce the SRES and Kc. The MS then sends the SRES back to the MSC for verification. The MSC then compares the received SRES from the MS with the SRES generated from the AuC. If they agree then the MS is authenticated. If they disagree the connection is terminated and an authentication failure message is sent to the MS.

The encryption process begins after the MS has been authenticated on the GSM network. The MSC transfers the cipher key Kc to the BTS and the unencrypted communication between the MS and BTS is then put into Cipher Mode. The encryption algorithm referred to as A5 is stored on the BTS and the ME. The cipher key Kc generated from the A8 algorithm is then inputted into the A5 algorithm to produce a keystream. This keystream is used for the encryption/decryption of over the air communications between the MS and BTS.

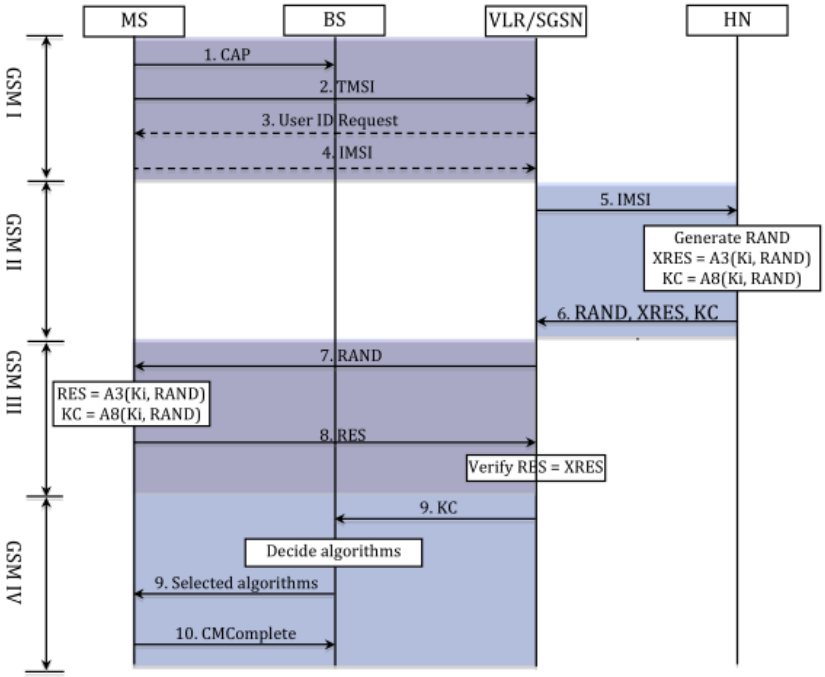

Figure 11 GSM AKA procedure

4.1.1.2 Authentication-Encryption algorithms

One of the biggest problems facing the 1G mobile phones was that no encryption was used with the first generation AMPS network. This brought about a need for a 2G mobile phone where over-the-air communications was encrypted, this begun with the 2G GSM network. There are four main algorithms used in 2G GSM mobile phones:

- A3: Authentication Algorithm
- A8: Cipher key Kc Algorithm
- A5/1: ”Stronger” Over The Air Communication Encryption Algorithm
- A5/2: ”Weaker” Over The Air Communication Encryption Algorithm
- A5/0: “Null” Over The Air Communication No Encryption Algorithm

These algorithms were privately developed by the Security Algorithm Group of Experts (SAGE). The stronger A5/1 algorithm was used in Western Countries and the deliberately weakened version A5/2 was imported to the other countries. It was rumored that a weaker algorithm was used at the request of European intelligence agencies headed by the creator of the A5/2 algorithm France. By 1999 the algorithm was considered extremely weak a month after it was leaked to the public. The A5/2 algorithm was then mandated to be phased out according to the GSM Association by 2006. In 2007 it was then prohibited to implement the A5/2 algorithm in any new phone. Countries that relied only on A5/2 that didn’t add support for the A5/1 algorithm were left with no encryption at all on newer phones. The upgrade from 2G GSM to 2.5G GPRS brought about a new set of security algorithm. These algorithms were again developed by SAGE behind closed doors and never released to the public.

The original GSM network used the COMP128 algorithm for their A3/A8 algorithms. The COMP128 algorithm is a one-way compression function. There have been three versions of the COMP128 algorithm with only the first version publicly known after it was leaked. GPRS used an upgraded COMP128 algorithm referred to as COMP128-2 which supposedly fixed certain security flaws. A year after GPRS was released the algorithm was upgraded to what is referred to as COMP128-3. These COMP128 algorithms all have the same core design but each upgrade brought about security enhancements. Since the designs were never made public the security of GSM mobile phones relied heavily on security by obscurity. The advantage of releasing the algorithm to the public is that the more people you have scrutinizing the algorithm the better chances there are of finding any flaws in its design. The new encryption algorithm for GPRS is referred to as the GPRS Encryption Algorithm (GEA). When the GPRS system had an update a year later GEA was upgraded to an algorithm referred to as GEA2. These algorithms have not been included in this paper as they have not been published. In this chapter we will examine the 2G GSM authentication and cipher key algorithms A3/A8 along with the A5 encryption algorithms.

Authentication of the mobile user and the creation of the Cipher Key Kc is accomplished by means of the A3/A8 algorithms. The A3 algorithm is used to authenticate the mobile user by sending back a response (SRES) to the challenge (RAND) given by network. The A8 algorithm is used to generate the 64-bit Cipher Key Kc which is used in the encryption of over-the-air communications. The A3/A8 algorithms are located on the protected SIM card along with the 128-bit Individual Subscribers Authentication Key Ki . The input for both algorithms are a 256-bit combination of the 128-bit RAND and the 128-bit Ki . Since both algorithms rely on the same input they are executed at the same time and rely on the algorithm called COMP128. The 256-bit input for the COMP128 algorithm produces a 96-bit Output which is composed of the 32-bit SRES and the 64-bit Cipher Key Kc.

The two algorithms used to encrypt over the air communications in 2G networks are referred to as A5/1, A5/2 and A5/0. These algorithms are stream cipher that were designed in secret but leaked to the public. Note that A5/0 refers to the case where no encryption is used.

The A5/1 algorithm consists of three Linear Feedback Shift Registers (LFSR). The first register R1 has a length of 19 bits with the rightmost bit referred to as position 0. The other two registers R2 and R3 have lengths of 22 and 23 bits respectively. Each register has a clocking bit and tapping bits. The tapping bits are XORed together and are used to produce the entry for bit position 0 after a shift has occurred. The tapping bits for R1 are at positions 18, 17, 16, and 13. The tapping bits for R2 are at positions 21 and 20. R3 has tapping bits at position 22, 21, 20, and 7. The clocking bit is responsible for the shifting of the registers. The clocking bit for R1, R2, and R3 occur at positions 8, 10, and 10 respectively. A majority rule function is used on the clocking bits to determine if a register is clocked. This stop/go process is done with each cycle. Since the stop/go process relies on a majority rule function there will always be at least two registers being clocked for each cycle. If a register is clocked the tapping bits are XORed and the value is used to produce the bit in position 0. For example, if the clocking bits for R1, R2, and R3 are 1, 0, 1 then R1 and R3 would be clocked. If the clocking bits were 1, 0, 0 then R2 and R3 would be clocked. Regardless of whether a register is clocked or not the leftmost bit: 18, 21, and 22 are XORed together and the result is used to form the keystream. The input for the A5/1 algorithm consists of a 64-bit Cipher Key Kc generated from the COMP128 algorithm along with a publicly known 22-bit Frame Number Fn . A frame number is used so that a new keystream will be produced for each frame making each keystream unique. The frame number increases incrementally while the Cipher Key remains the same. The Cipher Key is not changed until the MS is authenticated again. This means that the same Cipher Key can be used for days at a time. The output of A5/1 consists of 228-bits. The first 114-bits are the keystream for encrypting communication from the MS to the BTS. The last 114-bits of keystream are used to encrypt the communication from the BTS to the MS. A new keystream is created for every frame in intervals of 4.6 milliseconds. Each one of these frames produces a publicly known frame counter Fn. Once the number of frames reaches 8,388,608 the frame number cycles back to the beginning and resumes counting again.

While A5/1 was used in European countries and North America, a deliberate weakening of the A5/1 algorithm called the A5/2 algorithm was used elsewhere. The A5/1 algorithm is referred to as the ”stronger” algorithm while the A5/2 algorithm is referred to as the ”weaker” algorithm. The A5/2 algorithm was made weaker by requests of the intelligence agencies to ensure it was breakable. The intelligence agencies did not want a strong encryption algorithm to be used in the Middle East and this issue was the main driving force for creating the weaker algorithm. The A5/2 algorithm consists of 4 LFSR referred to as R1, R2, R3, and R4. The first three registers R1, R2, and R3 all share the same tapping bits as in the A5/1 algorithm along with the same lengths of 19, 22, and 23 bits. The new fourth register R4 is 17 bits long with tapping bits at positions 11 and 16. For the A5/1 algorithm the clocking of the registers were controlled by clocking bits located in each register. In A5/2 the clocking of registers R1, R2 and R3 are controlled by clocking bits located in R4. The clocking is determined by performing a majority function for bits R4[3], R4[7], and R4[10]. If the majority bit is the same as R4[10], then the first register R1 is clocked. Similarly R2 and R3 are clocked if the majority bit is the same as R4[3] or R4[7]. The clocking of R4 is done after the majority function is computed and R1, R2, and R3 are checked for clocking. The output bit for A5/2 depends on the XOR of 6 bits. The first three bits come from the original

output source of the A5/1 algorithm: the XOR of the three register bits R1[18]⊕R2[21]⊕R3[22]. The A5/2 algorithm adds an additional 3 bits to be XORed by using a majority function for each register. The majority function for R1 is computed by taking the majority of bits R1[12], R1[14]⊕1, and R1[15]. The majority function for R2 relies on bits R2[9], R2[13], and R2[16]⊕1. The majority function for R3 relies on bits R3[13]⊕1, R3[16], and R3[18]. The input for the A5/2 algorithm is the same as the A5/1 algorithm. It consists of the same 128-bit cipher key produced by the A8 algorithm along with the 22-bit frame number Fn . The output consists of 114+114-bits of keystream. The first 114-bits of keystream are used for encrypting MS to BTS communications. The second block of 114-bits are used to encrypt BTS to MS communications.

4.1.2 Privacy of 2G GSM Air Interface

As seen in the previous paragraph the authentication process begins with the MS requesting access to the network by broadcasting the IMSI/IMEI. This usually occurs when the MS is turned on for the first time or if the MS has been turned off for more than 48 hours. The MSC receives the IMSI/IMEI and sends the IMSI to the HLR with a request for an authentication triplet.

To further protect the privacy of subscriber identity, a security mechanism that utilizes temporary identity information is used. Upon successful registration to the network, a subscriber is assigned a temporary international mobile identification (TMSI) by the visitors location register (VLR) of the serving network. In subsequent transactions between the MS and the serving network, the subscriber is identified by this TMSI instead of the permanent and private international mobile station identification (IMSI). The mapping between a TMSI and the IMSI of a subscriber is known and valid only in the serving network. An attacker who captures TMSI information that is exchanged over the air interface cannot derive the subscriber's identity from TMSI information.

Although TMSI was implemented in order to avoid the mobile equipment clearly using IMSI for identification Maintenance of user anonymity is also a significance weakness of GSM networks since there are cases where this is required. This happens when UE needs to send its IMSI across the air interface in clear text when a it is switched on and wishes to connect to a new network, and hence will not have an assigned TMSI. Another case is where the serving network is unable to identify the IMSI from the TMSI. Thus, the IMSI might be clearly revealed, either by eavesdropping regular message transactions, or by enforcing an IMSI identification by a false BTS.

4.2 3G UMTS Air Interface

We start by considering the first introduced 3G system, UMTS. In UMTS the protocol layering is separated depending on the kind of the information transmitted (user-oriented (U-plane) or control-oriented (C-plane)). The user plane is the logical plane responsible for carrying user data being sent over the network (e.g., voice communication, SMS, application traffic) while the control plane is responsible for carrying all of the signaling communication needed for the UE to be connected. The protocol stack of UMTS for C-plane is given below:

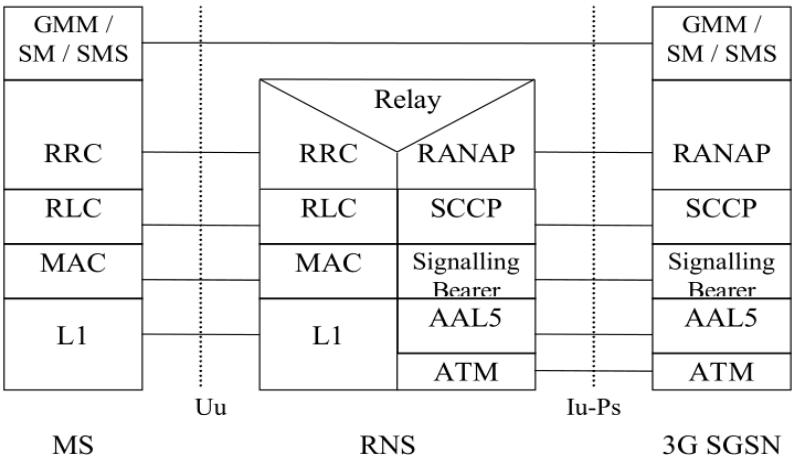

Figure 12 UMTS protocol stack (control plane)

The air interface between the MS and the BTS is called Uu. The major difference between GSM network and UMTS network is in the air interface transmission. Time division multiple access (TDMA) and freqency division multiple access (FDMA) are used in GSM networks. The air interface access method for UMTS networks is wide-band code division multiple access (WCDMA), which has two basic modes of operation: frequency division duplex (FDD) and time division duplex (TDD). This new air interface access method requires a new radio access network (RAN) called the UTMS terrestrial RAN (UTRAN). The core network requires minor modifications to accommodate the UTRAN. Two new network elements are introduced in the UTRAN: the radio network controller (RNC) and Node B. The UTRAN contains multiple radio network systems (RNSs), and each RNS is controlled by an RNC. The RNC connects to one or more Node B elements. Each Node B can provides service to multiple cells.

The RNC in UMTS networks provides functions equivalent to the base station controller (BSC) functions in GSM network. Nodes B in UMTS networks is equivalent to the base transceiver station (BTS) in GSM network. In this way, the UMTS extends existing GSM and GPRS networks, protecting the investment of mobile wireless operators. The UMTS network architecture is given below:

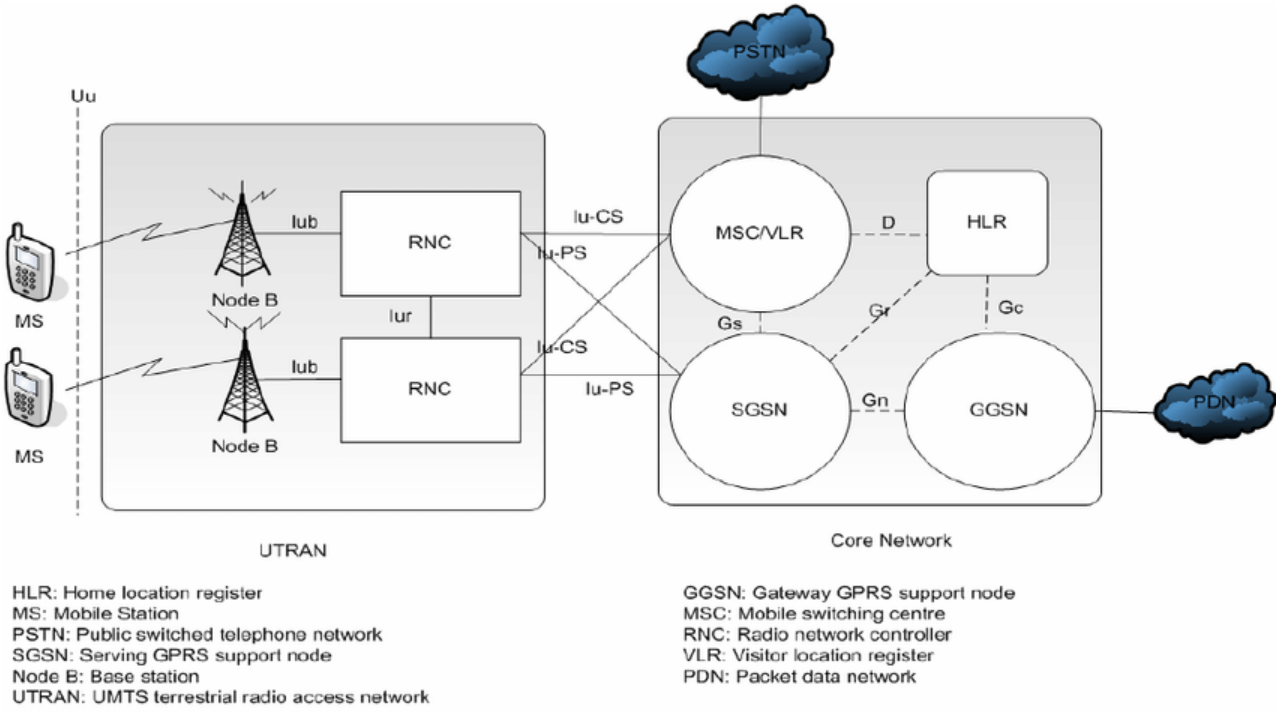

Figure 13 UMTS network architecture

Privacy and authentication mechanisms in the 3GPP based networks are largely based on security mechanisms developed in GSM standards but with many enhancements to address the identified shortcomings, noticeably the lack of mutual authentication, of GSM networks.

4.2.1 Security of 3G UMTS Air Interface

According to the relevant technical specification for UMTS security (TS 33.102) by 3GPP, UMTS is designed to provide the following security features:

- User identity confidentiality
  - user identity confidentiality: the property that the permanent user identity (IMSI) of a user to whom a service is delivered cannot be eavesdropped on the radio access link
  - user location confidentiality: the property that the presence or the arrival of a user in a certain area cannot be determined by eavesdropping on the radio access link
  - user untraceability: the property that an intruder cannot deduce whether different services are delivered to the same user by eavesdropping on the radio access link
- Entity authentication
  - user authentication: the property that the serving network corroborates the user identity of the user
  - network authentication: the property that the user corroborates that he is connected to a serving network that is authorised by the user's HE to provide him services; this includes the guarantee that this authorisation is recent
- Confidentiality
  - cipher algorithm agreement: the property that the MS and the SN can securely negotiate the algorithm that they shall use subsequently

  - cipher key agreement: the property that the MS and the SN agree on a cipher key that they may use subsequently
  - confidentiality of user data: the property that user data cannot be overheard on the radio access interface
  - confidentiality of signalling data: the property that signalling data cannot be overheard on the radio access interface
- Data integrity
  - integrity algorithm agreement: the property that the MS and the SN can securely negotiate the integrity algorithm that they shall use subsequently
  - integrity key agreement: the property that the MS and the SN agree on an integrity key that they may use subsequently
  - data integrity and origin authentication of signalling data: the property that the receiving entity (MS or SN) is able to verify that signalling data has not been modified in an unauthorised way since it was sent by the sending entity (SN or MS) and that the data origin of the signalling data received is indeed the one claimed
- Mobile equipment identification

The 3GPP wanted to design the UMTS system to address some of the security flaws of the original 2G GSM system while using the GSM system as a foundation upon which to build. Unlike the 2G GSM networks where the security algorithms were kept secret the 3GPP called upon the open scientific community to design their cryptographic systems. The 3GPP made public all of their drafts, designs, and algorithms for the scrutiny of the academic community. This created a process for the 3GPP to select the best authentication and encryption algorithms of the time. There were three certain flaws of the 2G system that needed to be addressed such as the short key length of the A5 algorithms, the vulnerability to the false base station attack, and the lack of a message integrity system. The A5 algorithms used a key length of only 64-bits which was considered secure at the time, but with the advent of more powerful computers 64-bit key lengths became a vulnerability. The false base station attack was possible because 2G systems only authenticated the user and not the network. The 3GPP wanted to create a system were both the user and the network were authenticated. A system of checking for message integrity was needed to ensure that there was no tampering of communications between the network and subscriber. With these three ideas in mind the 3GPP designed the UMTS cryptographic system.

UMTS was built on the GSM network to allow for easy migration for the networks and provide backward compatibility for the user. Thus a 2G phone would be able to operate on a 3G network. Some of the terminology has also changed to address the upgraded technology of more advanced cell phones. The 2G GSM system refers to the mobile phone and SIM card as the MS while the UMTS 3G system uses the term User Equipment (UE). The mobile phone is now referred to as the Terminal and 3G phones have larger screens and fewer buttons with the advent of touch screens. The word terminal was chosen for 3G phones because the phone is often thought of as a handheld computer. In the GSM system the authentication algorithm was hard coded into the SIM card. For UMTS the SIM card is referred to as the Universal Subscriber Identity Module (USIM). The difference between a SIM card and a USIM card is that the security algorithms are not hard coded into the USIM but rather implemented as applications. This allows for multiple applications to be placed on the USIM which allows opportunity for the UE to be used on multiple carriers around the world. The HSDPA/HSPA+/LTE networks are all based on UMTS and therefore use the same algorithm set for encryption and authentication.

In the following paragraphs we will see the implementation of the abovementioned features in Air Interface.

4.2.1.1 UMTS Authentication and Key Agreement procedure

Unlike 2G systems the UMTS system authenticates not only the UE, but the UE also authenticates the network. To reduce network traffic this authentication is done with a single pass. The authentication process for UMTS is referred to as Authentication and Key Agreement (AKA) procedure and can be separated in four steps, as presented in figure 15. The AKA process takes place as soon as the UE is detected on the network. The Key Agreement refers to the generation of the Confidentiality Key (CK) used to encrypt over-the-air communications and the Integrity Key (IK). The IK is used to verify that the message has not been tampered with by a man in the middle. The AKA process begins when an Authentication Data Request is made by sending the UE's IMSI number to the users HLR. The HLR then produces a set of n Authentication Data Responses or Authentication Vectors AV(1),AV(2),...Av(n). Each Authentication Vector is a 5-tuple consisting of the following: a RAND , an Expected Response (XRES), a CK, an IK, and an Authentication Token (AUTN). The HLR then sends this 5-tuple to the VLR. The VLR selects the first Authentication Vector AV(1) and sends the two values RAND (1) and AUTN(1) to the UE. The UE authenticates the network by verifying AUTN(1). Then UE the computes the user authentication Response (RES) RES(1) from the RAND(1) and sends this to the VLR. The VLR compares the RES(1) with XRES(1). If the RES(1) and XRES(1) agree, then the authentication of the user is complete and the CK(1) and IK(1) can be used to encrypt and check the integrity of over-the-air communications.

For the AKA process there is a total of seven algorithms. The name used to refer to these seven algorithms is "Milenage". The seven algorithms used for the authentication process are referred to as f1, f1*, f2, f3, f4, f5, and f5*. These set of algorithms are stored on the UE's USIM card along with the networks AuC. The USIM is designed to produce the CK, IK, and the AUTN in less than 500ms. The f1 algorithm is responsible for the network authentication. The f1* algorithm is in charge of the re-synchronization message authentication. The f2 algorithm is the user authentication function. The f3 algorithm produces the CK while the f4 algorithm produces the IK. The f5 function is responsible for the Anonymity Key (AK) derivation while the f5* is used to derive the anonymity key when the re-synchronization message function f1* is used.

The Milenage AKA process begins with the UE receiving the RAND and AUTN computed in the AuC and sent by the VLR. The AUTN is computed by concatenating the XOR of the SQN, a 48-bit sequence number that is an input to either of the functions f1 and f1*, and AK, the AMF, and the MAC − A. Thus AUTN = SQN $\oplus$ AK||AMF||MAC − A. The UE then inputs the 128-bit RAND and the 128-bit shared Subscribers Key K into the f5 function. This produces the 48-bit AK. Since the first 48-bits of the AUTN consists of the SQN $\oplus$ AK, AK is XORed with the first 48-bits of AUTN to produce the SQN. The USIM then checks to make sure that the SQN is in the correct range. If the the 48-bit SQN is in the correct range then it is inputted into the f1 function along with 128-bit RAND, K, and the 16-bit AMF. Note that the AMF is found by taking AUTN[48]...AUTN[64]. The output of f1 is the XMAC − A. If the XMAC – A agrees with the MAC − A, then the UE authenticates the network. Note that if the SQN is found to be out of range, then the normal key generation ceases and the functions f1* and f5* are used in place of f1 and f5. The f1* and f5* are only used when the SQN is found out of range. Once the network has been authenticated the key generation process can begin.

The RAND and K are inputted into f2, f3, and f4 to produce the RES, CK, and IK. For each function f1 − f5*, the Rijndael block cypher is used with input lengths of 128-bits. The

Rijndael block cipher consists of ten rounds after an initial Round Key addition. The first nine rounds consists of the following: a byte substitution transformation, a shift row transformation, a mix column transformation, and then a Round Key addition. The tenth round consists of a byte substitution transformation, a shift row transformation, and a Round Key addition. The result between each round is referred to as the State. The State is a 4x4 rectangular array of bytes. The output, input, and key are all represented as 4x4 arrays. Since each State represents 128-bits and a 4x4 array consists of 16 spaces each space represents a byte. The plaintext and Subscribers Key are divided into P0 , P1 ,..., P15 and K0 , K1 ,...,K15 . These values are then mapped respectively to a 0,0 , a 1,0 , a 2,0 , a 3,0 , a 0,1 , a 1,1 , a 2,1 , a 3,1 ,... and K 0,0 , K 1,0 , K 2,0 , K 3,0 , K 0,1 , K 1,1 , K 2,1 , K 3,1 ,... .

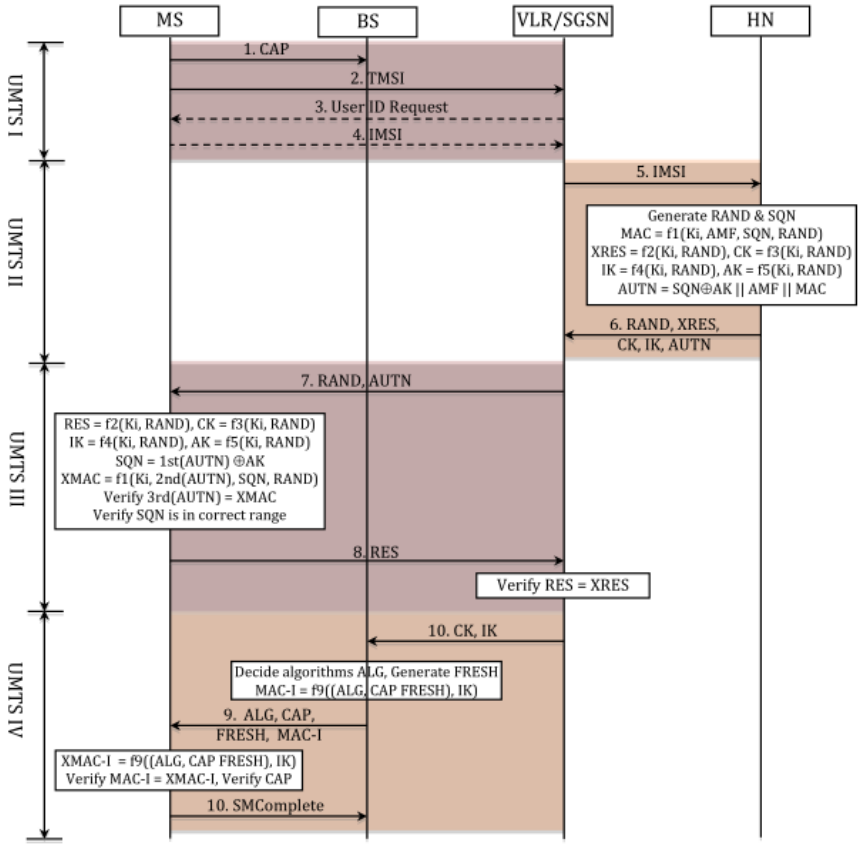

Figure 14 UMTS AKA procedure

4.2.1.2 Authentication – Encryption - Integrity algorithms

The Milenage set of algorithms use a block cipher for encryption called Rijndael, which is the proposed algorithm for the Advanced Encryption Standard. The Rijndael algorithm uses a 128-bit key along with a 128-bit block size input. The symbol E[x] k equals the result of applying the Rijndael encryption algorithm to the 128-bit value x under the 128-bit key k. The Milenage set of algorithms also use a 128-bit value called the Operator Variant Algorithm Configuration Field (OP). Each network is allowed to select its own value. Whether or not it is made public is determined by the operator. Note that the security of the algorithm does not depend on the secrecy of the OP. This value of OP is used by all subscribers on the network. Note that only the value OPc is used in the Milenage algorithms where OPc is computed from the inputs K, the 128-bit subscriber key that is an input to the functions f1,f1*, f2,f3,f4,f5, and f5*, and OP. Thus only the value of OPc is stored on the USIM card as opposed to having OP stored on the USIM card.

The encryption algorithm for the UMTS system is referred to as f8. There were two goals in mind about choosing of this algorithm: resilience and world-wide availability. Resilience in the sense that it will be in use for at least 20 years and provide protection from an exhaustive key search attack through an effective key space. To accomplish the goal of being a world-wide cellular system the algorithm had to be free of any use restrictions and made available publicly for all to implement. Because the time frame for developing an encryption algorithm was limited the 3GPP decided to choose an already proven strong cryptographic system and modify it to their needs. The 3GPP group then limited their search to only algorithms that carried no restrictions on export or use. After a call for open submissions the MISTY algorithm was chosen as the basis for the encryption algorithm f8. The MISTY algorithm was then modified for cellular phone use and renamed to KASUMI which is Japanese for MISTY. The descriptions and diagrams of the f8 algorithm to follow come from the openly available 3GPP documentations: General Report on the Design, Specification and Evaluation of 3GPP Standard Confidentiality and Integrity Algorithms, Specification of the 3GPP Confidentiality and Integrity Algorithms Document 1: f8 and f9 specifications, Document 2: KASUMI Specification, Document 3: Implementers' Test Data, and Document 4: Design Conformance Test Data.

The core of the f8 algorithm is KASUMI which is a block cipher that produces a 64-bit output from a 64-bit input under the control of a 128-bit key. The f8 algorithm itself is a stream cipher that is used to encrypt/decrypt blocks of data under a confidentiality key CK. The f8 algorithm uses KASUMI as a keystream generator with output keystream in multiples of 64-bits.

The integrity algorithm used in the UMTS system is referred to as f9. The f9 algorithm computes a Message Authentication Code (MAC) on an input message under an integrity key IK. There is no limitation on the input message length for the f9 algorithm. For ease of implementation the f9 algorithm is based on the same block cipher KASUMI which is used by the confidentiality algorithm f8 described above.

4.2.2 Privacy of 3G UMTS Air Interface

UMTS standard mandates the use of temporary identifiers to address mobile devices on the air interface, whenever a service is requested. The issue of identity confidentiality protection was already raised in the early GSM networks and the solution that is being adopted ever since has been updated but never substantially revisited. A (P)-TMSI, or (Packet)-Temporary Mobile Subscriber Identity, is the way mobile subscribers are identified over the air on packet/circuit-switched mobile networks in GSM and UMTS. These are attributed only after a successful authentication procedure, which takes place only when the network has the means to identify the home network of the mobile subscriber wishing to access the serving network. If the serving network has no valid credentials of the mobile user, it must establish the user identity and verify its rights before granting any service. This is done by asking the mobile subscriber for its permanent identity (or IMSI, the International Mobile Subscriber Identity), which is sent in cleartext by the mobile device. Once the serving network obtains the authentication information for the mobile subscriber, the mutual authentication procedure is executed and the subscriber is eventually given a TMSI and allowed to use the network.

4.3 4G LTE Air interface

We start by considering the most known 4G system, LTE. In LTE, as well as in UMTS, the protocol layering is separated depending on the kind of the information transmitted (user-oriented or control-oriented). The following protocols are used for communication over the

air interface (the radio link between the UE and the eNodeB). This protocol suite is referred to as the air interface protocol stack, which is generally divided into three layers (L1, L2, L3). Logically, these protocols set the foundation for all TCP/IP traffic operating above it. These protocols are:

- Radio Resource Control (RRC) operating at layer 3
- Packet Data Convergence Protocol (PDCP) operating at layer 2
- Radio Link Control (RLC) operating at layer 2
- Medium Access Control (MAC) operating at layer 2
- Physical Access (PHY) operating at layer 1

Each protocol within the air interface cellular stack performs a series of functions and operates on one of two logical planes: the user plane or the control plane. Below we present the LTE protocol stack designed for the control plane:

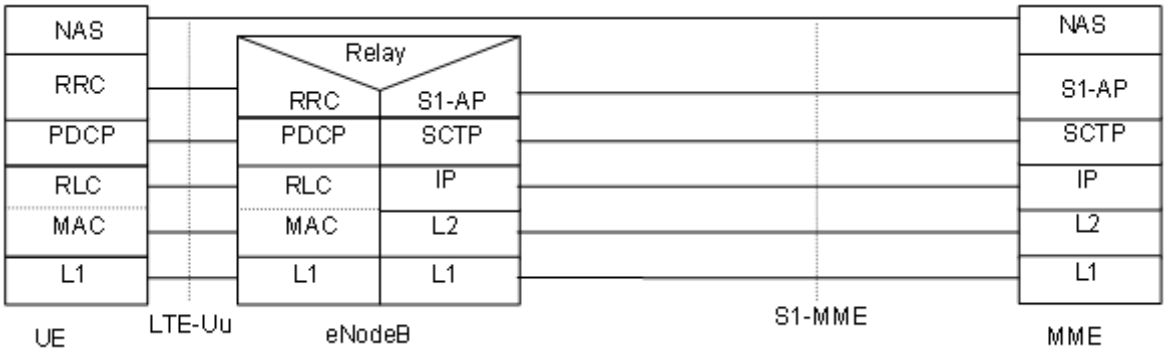

Figure 15 LTE protocol stack (control plane)

The air interface between the MS and the BTS is called LTE-Uu. This air interface is based on an OFDM modulation and utilizes MIMO techniques to increase the data rates. This Radio Access Network (RAN) has evolved over time into the Evolved Universal Terrestrial Radio Access Network (E-UTRAN). UEs connect to the E-UTRAN to send data to the core network. The E-UTRAN is a mesh network composed of base stations. A base station, or Evolved Node B (eNodeB), modulates and demodulates radio signals to communicate with UEs. eNodeBs then act as a relay point to create and send IP packets to and from the core network. Cellular networks are designed to pass connectivity from one radio access device in the E-UTRAN to the next as the connected UE changes location. This seamless handoff ability allows devices to have a constant connection with minimal interruptions providing what is know as 'mobility' within cellular networks. eNodeBs use the X2 interface to communicate with each other, primarily transmitting control signaling to allow for LTE network communication enabling UE mobility. During this handover the serving eNodeB must transfer all UE context 1, cellular parameters and other information about the UE, to the receiving eNodeB. The LTE network architecture is given below:

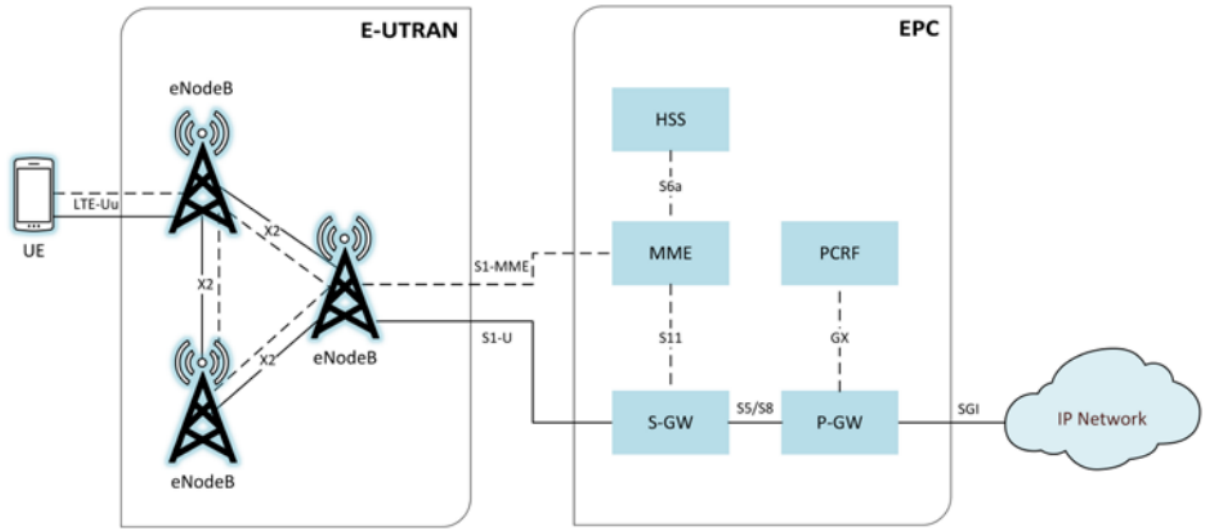

Figure 16 LTE network architecture

4.3.1 Security of 4G LTE Air Interface

According to the relevant technical specification of LTE security (TS 33.401) by 3GPP, LTE is designed to provide the following security features:

- User-to-Network security
  - User identity and device confidentiality
  - Entity authentication
  - User data and signaling data confidentiality
  - User data and signaling data integrity
- Security visibility and configurability
- Security requirements on eNodeB
  - Requirements for eNB setup and configuration
  - Requirements for key management inside eNB
  - Requirements for handling User plane data for the eNB
  - Requirements for handling Control plane data for the eNB
  - Requirements for secure environment of the eNB

In the following paragraphs we will see the implementation of the abovementioned features in Air Interface.

4.3.1.1 LTE Authentication and Key Agreement procedure

The primary LTE authentication mechanism mobile handsets use to authenticate to an LTE network is the known as the Authentication and Key Agreement (AKA) procedure. This procedure can be separated in five steps, as presented in figure 22. The use of AKA in LTE is required by 3GPP TS 33.401. The AKA protocol cryptographically proves that the UICC and the Mobile Network Operator (MNO) have knowledge of the secret key K. The AKA procedure provides mutual authentication between the UICC and the LTE network. AKA is begun by a UE providing its identifier to the appropriate MME. This identifier may be permanent, as is the case with the IMSI, or may be temporary. Examples of temporary identifiers include the Temporary Mobile Subscriber Identity (TMSI) and Globally Unique Temporary UE Identity (GUTI). After the identifier is provided to the core network, the MME provides the identifier, alongside additional cryptographic parameters and the serving network ID (SN id), to the HSS/AuC. These values then are used to generate an

authentication vector (AUTN). To compute an AUTN, the HSS/AuC needs to use a random nonce (RAND), the secret key K, and a Sequence Number (SQN) as inputs to a cryptographic function. This function produces two cryptographic parameters used in the derivation of future cryptographic keys, alongside the expected response (XRES) and authentication token (AUTN). This authentication vector is passed back to the MME for storage. In addition, the MME provides the AUTN and RAND to the UE, which is then passed to the USIM application. The USIM sends AUTN, RAND, the secret key K, and its SQN through the same cryptographic function used by the HSS/AuC. The result is labeled as RES, which is sent back to the MME. If the XRES value is equal to the RES value, authentication is successful and the UE is granted access to the network.

The three main responsibilities of the LTE-AKA are: (1) mutual authentication between a UE and an MME on behalf of the HSS/AuC, (2) key agreement between a UE and an MME as well as between a UE and eNB, and (3) enhancements over the weaknesses of the GSM-AKA, and UMTS-AKA. The handshake behavior of the LTE-AKA is equivalent with that of the UMTS-AKA. The only difference between them is the number of authentication vector, and elaborate key hierarchy. The MME initiates authentication by asking for the identity of an UE that has roamed into the MME's territory. The UE sends the IMSI in clear text as a response. Or the UE can send its GUTI with past location information if the UE already has the GUTI that the old MME transmitted after the LTE-AKA procedure. The MME passes the message to the HSS/AuC with adding the serving network identity and network type. The authentication vector created from the HSS/AuC consists of four components: RAND, XRES, K_{ASME}, AUTN where RAND is the random number generated by the HSS/AuC. K_{ASME} is the unique component of LTE-AKA. K_{ASME} is used to derive the encryption key and integrity check key for signaling, user plane and non-access stratum; and XRES is the expected response from the UE, if the UE is a valid user. The purpose of the is to make it KSI_{ASME} possible or the UE and the MME to identify the cached K_{ASME} without invoking the authentication procedure. This is used to allow re-use of the K_{ASME}.

The rest of the AKA operation is exactly same with the UMTS-AKA. After successful authentication, the K_{RRCenc} and K_{RRCint}, are available to encrypt and to authenticate messages of Radio Resource Control (RRC) signaling data; the K_{NASenc} and K_{NASint} are available to encrypt and to authenticate messages of application data; finally the K_{UPenc} are used to encrypt user plane messages. After the LTE-AKA, the MME assigns to the UE by sending the encrypted message using K_{NASenc}.

Once the UE and MME have completed the LTE-AKA, they can start communicating in a secure manner after the Security Mode Command (SMC) procedure. The purpose of the SMC procedure is to take an LTE security context into use, and signaling and data protection with the designated keys and security algorithms. During the SMC procedure, the UE negotiate the ciphering and integrity algorithms with MME and eNB for NAS protection and AS protection, respectively. The SMC message consists of two-way handshake (i.e., request and complete). The NAS SMC request message invoked by the MME contains the NAS ciphering and integrity protection algorithm, and the KSI_{ASME} for identifying K_{ASME}. The NAS SMC request message is integrity protected with an NAS integrity key. NAS uplink and downlink ciphering at the UE starts after receiving the NAS SMC complete message. The AS SMC request message triggered by the eNB contains the selected AS algorithms for ciphering and integrity protection and KSI_{ASME}. RRC and user plane ciphering at the eNB shall start after receiving the AS SMC complete message.

3GPP's technical specification 33.401 directs that both the NAS and RRC control plane messages must be integrity protected. 3GPP TS 33.401 5.1.4.1 requires that "Integrity protection, and replay protection, shall be provided to NAS and RRC-signaling." It is specified that user plane packets traveling on the Uu interface are not integrity protected. Specifically, 3GPP TS 33.401 5.1.4.1 states "User plane packets between the eNodeB and the UE shall not be integrity protected on the Uu interface."

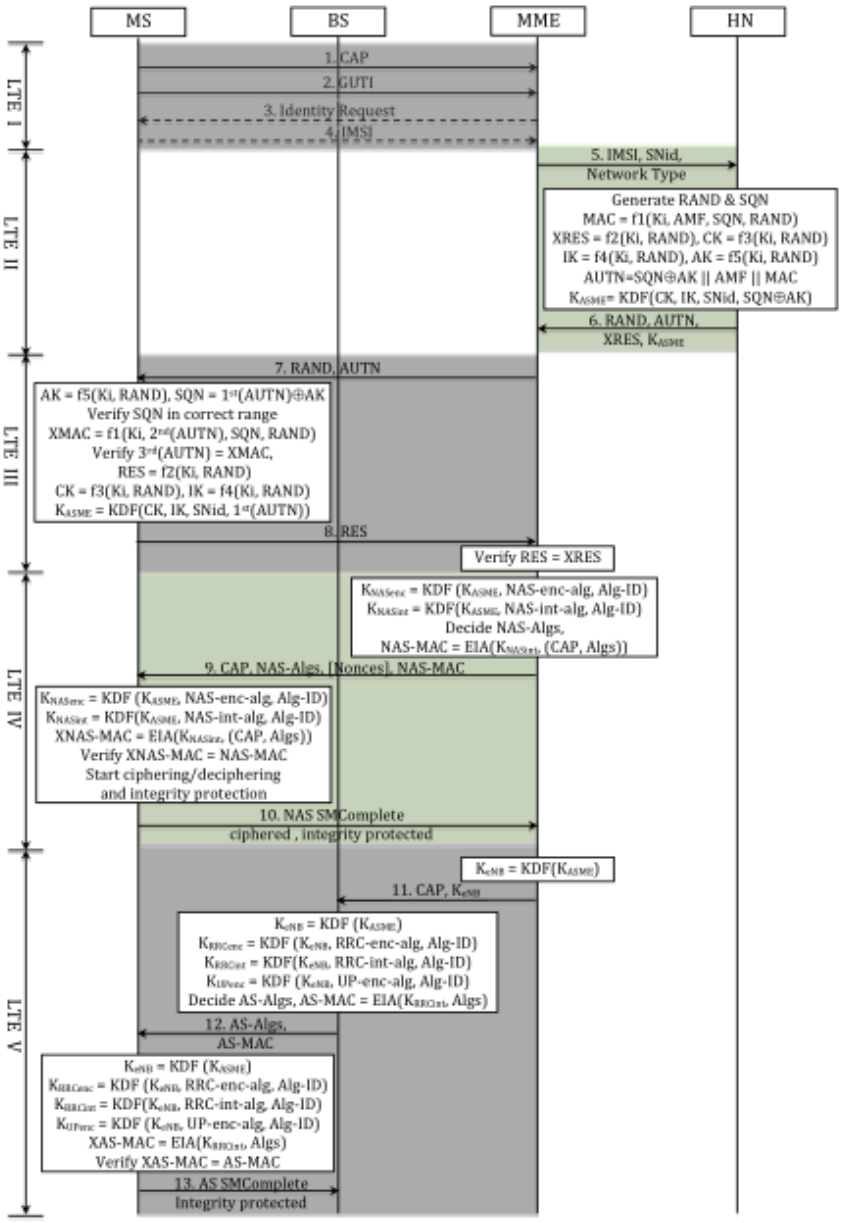

Figure 17 LTE AKA procedure

Since LTE AKA has been designed to provide appropriate encryption and integrity mechanisms for the security of the signaling and user data traffic and each mechanism for the two different types of traffic requires an independent secret key, the LTE-AKA may use numerous secret keys. As a consequence, the LTE-AKA must manage a novel key hierarchy. This key hierarchy is given below:

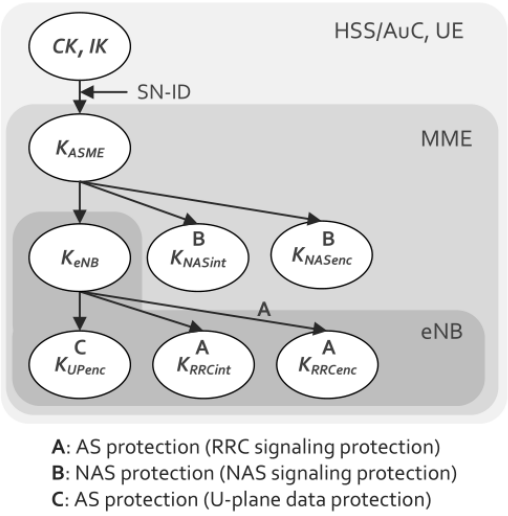

Figure 18 Key hierarchy in LTE

4.3.1.2 Authentication - Encryption - Integrity algorithms

As seen in the previous paragraphs, in older 2G cellular systems, the cryptographic algorithms used to secure the air interface and perform subscriber authentication functions were not publicly disclosed. UMTS introduced the first publicly disclosed cryptographic algorithms used in commercial cellular systems. The terms UEA (UMTS Encryption Algorithm) and UIA (UMTS Integrity Algorithm) are used within UMTS as broad categories. UEA1 is a 128-bit block cipher called KASUMI, which is related to the Japanese cipher MISTY. UIA1 is a message authentication code (MAC), also based on KASUMI. UEA2 is a stream cipher related to SNOW 3G, and UIA2 computes a MAC based on the same algorithm. LTE builds upon the lessons learned from deploying the 2G and 3G cryptographic algorithms.

LTE introduced a new set of cryptographic algorithms and a significantly different key structure than that of GSM and UMTS. There are 3 sets of cryptographic algorithms for both confidentiality and integrity termed EPS Encryption Algorithms (EEA) and EPS Integrity Algorithms (EIA). EEA1 and EIA1 are based on SNOW 3G, very similar to algorithms used in UMTS. EEA2 and EIA2 are based on the Advanced Encryption Standard (AES) with EEA2 defined by AES in CTR mode (e.g., stream cipher) and EIA2 defined by AES-CMAC (Cipher-based MAC). EEA3 and EIA3 are both based on a Chinese cipher ZUC. While these new algorithms have been introduced in LTE, network implementations commonly include older algorithms for backward compatibility for legacy devices and cellular deployments. Many keys in LTE are 256-bits long, but in some current implementations only the 128 least significant bits are used. The specification has allowed for a system-wide upgrade from 128-bit to 256-bit keys. In LTE, the control and user planes may use different algorithms and key sizes.

4.3.2 Privacy of 4G LTE Air Interface

In addition to the IMEI and IMSI, MNO's may utilize other LTE identities, including the Globally Unique Temporary Identity (GUTI) and the Temporary Mobile Subscriber Identity (TMSI). The GUTI can identify a UE to a network without having to send the long-term identity (i.e., IMSI). Different identities are used for various reasons, including limiting the exposure of a permanent identity, to minimize tracking of a device as it accesses multiple services on the network.

As provisioned in LTE TS 33401, the user identification mechanism should be invoked by the serving network whenever the user cannot be identified by means of a temporary identity (GUTI). In particular, it should be used when the serving network cannot retrieve the IMSI based on the GUTI by which the user identifies itself on the radio path. The mechanism allows the identification of a user on the radio path by means of the permanent subscriber identity (IMSI). The mechanism is initiated by the MME that requests the user to send its permanent identity. The user's response contains the IMSI in cleartext. This represents a breach in the provision of user identity confidentiality.

5. Deviations in Security and Privacy

Following the description of security and privacy mechanisms used in GSM, UMTS and LTE, we now come in real world implementations of cellular networks, where plenty of deviations in entity authentication and user's privacy have been revealed. In the following paragraphs we present these deviations in GSM, UMTS and LTE systems.

5.1 Security and Privacy deviations in GSM

GSM represents the first successful globally established digital cellular communication system. Multiple improvements of the original GSM security specification have been achieved with the introduction of newer versions, such as the GPRS (General Packet Radio Service) and the EDGE (Enhanced Data rates for GSM Evolution). The study of the vulnerabilities of GSM networks is highly significant for three reasons. Some operators still maintain the original GSM configuration, especially in developing countries. Additionally, the serving networks and user equipment are designed in order to provide backward compatibility, while the selection of the serving network is achieved based either on the coverage or the capacity of the deployed network, making this way possible the enforcement of degradation to GSM from the new and supported technologies.

Obviously, due to their open nature, communications that utilize the wireless medium are inherently vulnerable to various threats. GSM was designed in order to minimize the risk of interception, incorporating technologies with transitive benefits to security assurance, such as frequency hopping. Yet, interception remained feasible, while even commercial interception systems, specifically designed for GSM, are available. Additionally, authentication was only defined towards the SN to UE entity path. Offering this way open ground for a variety of attacks, including the notorious GSM vulnerability to the man in the middle attack. The complete absence of integrity protection is also noticeable.

The initial GSM protocol did not take under consideration the integrity protection of neither signaling nor data messages, facilitating this way unidentified message tampering. Another security weakness is identified on the fact that encryption mechanisms are only utilised through the wireless link of the air interface, with complete lack of user visibility, and not through the fixed network connections. The backbone network, comprising of all the network elements and connections other than the one among the MS (Mobile Station) and BTS (Base Transceiver Station), is mainly based on unencrypted communication. Regarding the backbone network, the deployed SS7 (Signaling System No. 7) system also carries security vulnerabilities, regarding the feasibility of message modification, due to the amount and complexity of the existing interfaces and the tight relations of SS7 with the internet.

Additionally, various flaws have been identified at the implementation of both the authentication and encryption algorithms, allowing the extraction of the Ki (Subscriber authentication) and Kc (Encryption) keys. The most commonly used A3/A8 algorithms are versions of COMP128, which was based on security by obscurity. This algorithm raised many concerns, while various implementation flaws allowed the extraction of the key, with relatively limited physical access given to the ME and a set of pre-calculated challenges. Similarly, over the air extraction of the Ki is also feasible with the use of a pre-defined loop of challenges towards the mobile equipment, allowing the mathematical calculation of the key. The cryptographic algorithms A5-1/2 were also based on security by obscurity. Various cryptanalysis methods have been discovered, able to identify the ciphering key in real time using a personal computer with moderate capabilities, to analyse encrypted conversation. Since the extraction of security related keys and numbers such as the Ki, Kc and IMSI

(International Mobile Subscriber Identity) is reasonably inexpensive, both financially and computationally, the creation of clone SIM cards is feasible. However, the GSM incorporates a safety valve, that would recognize a duplicate SIM card operating in distinct location area and deactivate the subscriber's account until further action is taken. Maintenance of user anonymity is also a significance weakness of GSM networks. Although TMSI (Temporary Mobile Subscriber Identity) was implemented in order to avoid the mobile equipment clearly using IMSI for identification, there are cases where this is required. Thus, the IMSI might be clearly revealed, either by eavesdropping regular message transactions, or by enforcing an IMSI identification by a false BTS. Finally it must be mentioned that GSM has significant vulnerabilities towards DOS (Denial-of-Service) [12] [13] and Replay attacks [14]. Multiple methods have been identified that make the execution of DOS attacks feasible in GSM networks, mostly utilizing the limited signalling channels and the architectural choice of the MS to connect to the BTS that provides the higher transmitted power.

5.1.1 Possible attacks against GSM networks

The presented vulnerabilities of GSM networks have been exhaustively studied, since they can be exploited in order to launch and successfully achieve a variety of attacks. Since they have been identified, some of these threats have been taken under consideration during the development of GSM enhanced versions and future generation cellular communication systems. In this section the most common and severe of these attacks are presented.

- Eavesdropping on traffic or signaling:

An adversary may eavesdrop user traffic or signaling and control data over the radio path. This may disclose user sensitive information, or provide access to security management information that can be further used in order to conduct additional active attacks. The used A5 ciphering algorithm received extended criticism since various cryptanalysis methods have been identified.

The first known attack of this family was a TMTO (Time Memory Trade Off) based on the birthday paradox. Yet, its successful realization had strict requirements both in the quantity of the required information and processing time. Future TMTO based attacks though, required only a small portion of non-ciphered information and limited computational power. The weaknesses of the A5 algorithm becomes easily identifiable by the amount of the proposed cryptanalysis mechanisms. Although some of them impose unrealistic requirements, many have been used to successfully demonstrate attacks against the all three GSM encryption algorithms.

Early encryption algorithms based on COMP128, which has been broken. The most well-known attack of COMP128 came from the same people who reversed engineered the leaked code: Wagner, Goldberg, and Briceno. Wagner and Goldberg were Berkley researchers and Briceno was the Director of the Smartcard Developers Association. The attack is called the WGB attack deriving its name from the discoverers. The description of the attack comes from Handschuh and Paillier from the ENST Computer Science Department and Gemplus Cryptography Department located in France [21]. This attack is a chosen-plaintext attack which means that it assumes the attacker has the capability to choose any plaintext they want and obtain the resulting ciphertext. This chosen-plaintext attack can be done with a SIM card reader and a computer with the COMP128 algorithm. There are two different situations for carrying out a chosen-plaintext attack on a mobile phone. The first situation being if someone has physical access to your SIM card and secondly using an over-the-air method. The COMP128 algorithm being a one-way compression function, or commonly

referred to as a hash function, is vulnerable to a type of attack called the birthday attack. The birthday attack gets its name from the birthday problem in probability theory. The birthday problem deals with finding the probability of having two individuals with the same birthday out of a set of n randomly chosen individuals. If you count February 29 there are 366 possible birthdays guaranteeing a matching pair when n = 367. The surprising result of the birthday problem is that 99% probability occurs with just n = 57 and 50% probability with n = 23.

This can be interpreted as a 50% probability in a class of 23 people that two random people will share the same birthday. The birthday attack uses this higher probability of finding two matching pairs among a fixed number of permutations to help find a collision. A collision occurs when two randomly chosen inputs produce the same output. Therefore finding a collision in the COMP128 function results in finding two different messages M 1 and M 2 such that COMP128(M 1)=COMP128(M 2). Note though that the attacker usually has no choice over choosing M 1 or M 2 . A collision is possible in a compression function as we are taking a larger domain and mapping it to a smaller range (pigeonhole principal). Using the probability model of the birthday problem and applying it to a compression function a compression of n bits results in a collision occurring in 2 n/2 attempts by brute force. This number is referred to as the birthday bound.

One of the major weaknesses of the A5/1 algorithm was the deliberate setting of the last 10-bits of the Cipher Key produced by the COMP128-1/COMP128-2 algorithm to zero. The result of this is that the Cipher Key used in A5/1 is reduced from 64-bits to effectively 54-bits. The time complexity of a brute force attack is thus reduced from 2 64 to 2 54 . This would take too long for real time eavesdropping but if one were to record the intercepted communication between the MS and BTS it would require about 250 hours with a Pentium based computer from 1998. Later an upgraded version COMP128-3 was released that expanded the 54-bit key to the full 64-bits. The two common types of attacks for the A5/1 algorithm are guess-and-determine attacks and time-memory-data tradeoff attacks. A description of these type of attacks comes from Gendrullis, Novotny, and Rupp from the Horst Gortz Institute for IT-Security in Germany [19]. Both forms of attack assume that at least 64-bits of consecutive keystream bits are known to the attacker. A guess-and-determine attack works as the name suggests by guessing bits of the register and using the known keybits to determine the remaining register bits. Ross Anderson, a professor in Security Engineering at the University of Cambridge, devised an attack against A5/1 by completely guessing the bits in R1, R2, and half of R3. The second half of R3 is then derived from the known keystream bits. This attack had a worst-case scenario of 2 52 operations needed to be performed to recover the Cipher Key. An improved guess-and-determine attack was proposed by J.D. Golic who guessed the lower half of each register and clocked the registers until the guessed bits ran out. Each output bit he was able to produce a linear equation in terms of the bits for the upper half of the three registers. This guessing the lower half process would continue until 64 linearly independent equations were obtained. The system of linear equations were then solved using Gaussian elimination. This attack could be completed in 2 40 steps. Both of these attacks are not very practical in that real time eavesdropping would not be possible with the number of steps required to determine the key.

The other type of attack is a time-memory-data trade-off attack. This attack relies on precomputed data to reduce the time needed to determine the Cipher Key. One of the

weaknesses of A5/1 is that the length of the registers is small enough that it is possible to precompute all the possible states of the three registers. Once a state is given all successive states are stored as well. Each register will have an array indexed by the state number in the succession, the clocking bit, and the output bit. This precomputation stage requires 2 48 operations and memory requirements of about 300GB. Once the precomputation stage is complete the next stage involves observing the keystream output for a given length of time. The tables are then used to try and match the string of keystream bits with the correct states of the registers with a success rate of 60%. The more keystream bits you have the faster the Cipher Key is recovered. If you are able to observe 2 minutes worth of keystream output it is possible to produce the Cipher Key within one second. If you only have 2 seconds worth of keystream bits then the attack will be able to find the Cipher Key within minutes. This type of attack is impractical due to the time it takes to complete the precomputational data and the unlikely fact of being able to observe 2 seconds worth of keystream bits. Observing 2 seconds of keystream bits amounts to having roughly 25,000-bits. There is also special hardware available built specifically to attack stream ciphers with a keysize of 64-bits or less. One publicly known code-breaker machine is called COPACOBANA (Cost-Optimized Parallel Code Breaker) and comes with a price tag of $10,000. The machine is able to find the Cipher Key within 7 to 14 hours with only 64-bits of known Keystream. The COPACOBANA uses a guess-and-determine attack on A5/1. With the length and computation time required to implement these attacks real time eavesdropping is highly unlikely.

Real time eavesdropping is accomplished by taking advantage of a serious security flaw in the GSM security architecture. The ME stores both the A5/1 (stronger) and A5/2 (weaker) algorithms to allow for encryption when roaming on networks that only support A5/2. The security flaw is that the Cipher Key used for the A5/1 algorithm is the same as the Cipher Key used for the A5/2 algorithm. If an attacker is able to force the phone into using the A5/2 algorithm in place of the A5/1 algorithm the Cipher Key can be found with relative ease compared to attacking the A5/1 algorithm.

An instant ciphertext-only attack for A5/2 was developed by Barkan, Biham, and Keller from the Israel Institute of Technology. The attack only requires a few dozen milliseconds of ciphertext and finds the correct key in less than a second using a personal computer. The attack takes advantage of the fact that GSM networks use error-correction codes. This error-correction code is applied to the plaintext and then both are encrypted to be sent over the air. The addition of an error-correction code increases the size of the ciphertext sent. It is the standard today to first encrypt the message and then apply error-correcting codes. The reason being that encrypting error-correcting codes introduces a structured redundancy in the ciphertext which is exploited in the attack. The attack focuses on the error-correction codes used at the beginning of a conversation. The error-correction code has a fixed length of 184-bits and when implemented with the plaintext result in a 456-bit message. The 456-bit message is then divided into four frames and then encrypted and transmitted. The initial internal state of the registers are treated as a variable and every bit of ciphertext is written as a quadratic function of those variables. Each block of 456-bits will produce 450 equations of which only 272 equations are needed. Thus, after two blocks (8 frames) we can use Gaussian elimination on the system of quadratic equations to find all of the original linear variables in the initial state. The setup process of A5/2 is then inverted once the initial state is known to find the Cipher Key. This attack on the weaker A5/2 algorithm is used to eavesdrop in real time on any ME that is equipped with the A5/2 algorithm. As stated in the A5/1 attack section most ME are equipped with the A5/0 (no encryption), A5/1 (strong

encryption), and A5/2 (weak encryption) algorithms to allow for roaming on A5/2 only networks. The attack downgrades the encryption of the ME from A5/1 to the weaker A5/2 algorithm where the Cipher Key is easily found. Once the Cipher Key is found the A5/1 algorithm provides no protection as the same Cipher Key is used for both algorithms. This downgrade attack is implemented by what is referred to as a Man in the Middle Attack. This attack takes advantage of three key weaknesses in the GSM network:

1. Authentication and Key agreement protocol can be executed between the MS and the network at the beginning of a call at the network's discretion.
2. The MS cannot ask for authentication of the network. The network chooses the encryption algorithm to use (A5/0, A5/1 or A5/2). The MS only lists the available ciphers it supports. Note though that if no encryption is selected (A5/0) a message will display on the MS indicating no encryption.
3. There is no Cipher Key separation between A5/1 and A5/2. The Cipher Key only depends on the RAND.

- Masquerading:

Various masquerading attacks are possible within the GSM network. An adversary can impersonate a network element, in order to intercept or passively analyze user traffic or related signaling and control messages. Signaling and control information can be further used, so that the adversary can masquerade as another network subscriber, gaining access to services on behalf of the legitimate user. The main steps of such an attack require the adversary to first masquerade as a legitimate BTS towards the subscriber, and after authentication has been achieved, use the extracted information to masquerade as the subscriber towards the network. Such attacks are feasible due to the lack of mutual entity authentication, allowing this way the introduction of fake BTS. Such equipment can be further used for IMSI/IMEI catching attacks or selective jamming attacks. Similarly, an adversary can manipulate the behaviour of the terminal or the SIM card by masquerading as a legitimate originator of applications and data. Finally, an intruder can impersonate a legitimate user and utilise the authorised services, simply (assuming that he has the required access privileges) by receiving the required information by other entities such as the serving network or the user.

- Man in the middle:

As referred above, an adversary using similar methodologies can get nested between a subscriber and the serving network. This can allow him to execute a wide variety of illegitimate actions, such as deleting, modifying, spoofing and replaying signaling messages or user data.

The Man in the Middle Attack uses a fake base station to impersonate the network for the MS while impersonating the MS for the network. When the network asks for authentication an authentication request is sent to the attacker who forwards it to the MS. The MS computes the SRES and returns it to the attacker. The attacker holds on to the SRES and does not send it to the Network. The attacker requests the MS to start encryption using the A5/2 algorithm. The MS replies with a 456-bit Cipher Mode Complete Message encrypted using A5/2 to verify it is encrypting the message. The attacker does not have the Cipher Key so he is not able to respond. Immediately after not receiving a response the MS resends the Cipher Mode Complete Message of 456-bits in another frame set. Now the attacker has the required amount of ciphertext needed to begin the A5/2 attack and within a second has the

Cipher Key. As an added bonus to the attacker the same Cipher Mode Complete Message is sent with a different frame number and the only difference being that one bit is changed from a 0 to a 1 to indicate retransmission. The attacker then sends the SRES computed by the MS back to the network and is authenticated. The network starts encrypting using the stronger A5/1 algorithm. The attacker knowing the Cipher Key is able to send and respond to encrypted messages using A5/1. The network views the attacker as the authenticated subscriber and the attacker is able to forward messages to the MS while eavesdropping in real time or use the subscribers account to place calls at their expense. The attacker is also able to receive SMS and data messages and change them at will before forwarding them to the network or MS. The whole attack causes only a one second delay in authentication and the GSM network gives a 12 second window for the MS to authenticate itself. This Man in the Middle attack using the downgraded A5/2 algorithm provides a real time security threat to any MS equipped with the A5/2 algorithm. This attack was negated in the 3G UMTS system by requiring authentication both ways between the MS and network. Note that the removal of the A5/2 algorithm took some time to implement. The attacks against A5/2 began to be published in 1999 but were not taken seriously until 2003. In 2003 the GSM Association recognized that there was a problem with the downgrade attack and the only solution was to remove the algorithm from new equipment. Although it wasn't until 2007 that the removal of the A5/2 algorithm was implemented.

- Privilege abuse:

An inherit threat towards commercial cellular communication systems, is the subscriber's capability to misuse their privileges and gain unauthorized access to services or intensively overuse their subscriptions, even if such cases can only have a short and limited duration. Similar types of subscriber originated attacks may include use of stolen terminals, IMEI manipulation, and terminal/SIM data modification or extraction. Additionally the commercial cellular networks become increasingly interconnected with the internet, inheriting this way some of its security threats such as un-traceability or identity theft, including the widespread of viruses and malware.

- Denial of service and Jamming:

These are two of the most severe attacks against GSM networks, due to the extensive network sensitivity towards them and the variety of methods that can be used, in order to execute such attacks. The most common and successful method of realising a DOS attack against a GSM network, requires the exhaustion of the network resources, aiming mostly on the limited signalling channels. Such an attack makes use of the fact that initial communication among a MS and the network, is executed before authentication. Thus, a MS may repeatedly follow the appropriate protocol steps, requesting additional signalling channels without ever completing the protocol cycle and release them. Regarding jamming attacks, these can exploit the broadcast channels of the network and more precisely the synchronization bursts or registration identifiers, forcing a MS to lose signalling interconnection with the network.

- The downgrade attack

In the 2G GSM section we demonstrated the security threat brought about by having a weaker encryption algorithm included in the ME. If the ME was equipped with both the A5/1 and A5/2 algorithms a downgrade attack was used to force the phone into using the weaker encryption mode A5/2 to discover the Cipher Key. The same threat would apply to UE with

the A5/2 and f8 algorithms. A 3G phone using the f8 algorithm could potentially be downgraded to using the A5/2 algorithm where an attack could be mounted to recover the Cipher Key. To prevent this downgrade attack along with enhancing the security of the GSM/GPRS/EDGE networks each of their respective security algorithms were upgraded and based on the f8 algorithm. Note that 3G UMTS UE needed to contain 2G encryption algorithms to allow for backwards compatibility while roaming. The new algorithms for the GSM and GPRS/EDGE networks are named A5/3 and GEA3. Since the 2G networks use a Cipher Key of only 64-bits and 3G networks use 128-bits, a modification of the inputs had to be made. The A5/3 and GEA3 algorithms are based off of f8 and have the same core structure. Therefore the core part of the f8 algorithm that they all share is referred to as the Core Keystream Generator (KGCORE)

Using a rogue base station broadcasting at a high power level, an attacker can force a user to downgrade to either GSM or UMTS. As of the time of this writing, there are no significant, publicly known weaknesses in the cryptographic algorithms used to protect the confidentiality and integrity of the UMTS air interface. Unfortunately, significant weaknesses exist for the 2G GSM cryptographic algorithms used to protect the confidentiality and integrity of the air interface. Examples of broken 2G cryptographic algorithms are A5/1 and A5/2. Depending on the algorithm negotiated while attaching to the rogue base station, the air interface cryptographic algorithms chosen to protect the air interface may be cryptographically broken, leading to a loss of call and data confidentiality. Real world deployments utilize GSM networks to connect with LTE networks, which bring this into scope.

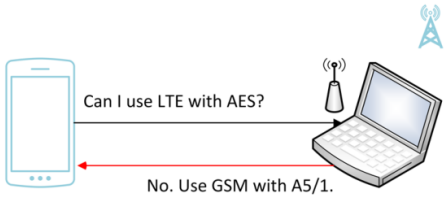

Figure 19 Simplified downgrade attack

This problem relays in the fact that the network not and the BS decides when to turn on encryption (A5/0 or A5/2 or A5/1).

The over-the-air attack would work by impersonating a BTS. Thus it is referred to as the False Base Station attack. Programing codes for the GSM network were leaked when an Italian carrier in the 1990s went bankrupt and posted the codes on the SourceForge website leaving it there for over 4 years before being taken down. With this code attackers were able to build the software necessary to make a fake base station. The ME is designed to connect to the BTS that emits the strongest signal. Thus an attacker would have to build a false base station that emits a stronger signal than the real BTS. To emit a signal strong enough to take over the real signal the attack would have to take advantages of places were signal strength from the BTS was already weak. These places of attack include subways, elevators, and inside offices. A sign to the user that an over-the-air attack was taking place would be a rapid battery drain as the phone would be communicating for around 8 hours with the attacker. These attacks were not very practical due to the time required to extract the key but they did demonstrate the weaknesses of the COMP128 algorithm. SAGE, the closed door designers of COMP128, released a press announcement after the attack was published noting that COMP128 was just a recommended algorithm for the carriers to use and not a requirement. Although the Smartcard Developer Association announced that they have not

been able to identify one network in Europe or the U.S. that used an algorithm other than COMP128. By 1998 the GSM Association estimated that there were 80 million GSM users worldwide so a change in authentication algorithms would require the replacement of 80 million SIM cards along with a software upgrade to the networks. Because of this reason no changes to the system were made until 2001 when a patched version of COMP128 referred to as COMP128-2 was released and implemented in newer SIM cards. The algorithm changed the way the SRES was calculated but was never released to the public.

- Deny tracking of stolen mobiles

Since IMEI is not authenticated, this fact can lead to denial of tracking stolen devices.

5.2 Security and Privacy deviations in UMTS

In the design of 3G systems like UMTS, a new security architecture was specified. However, the approach that was taken was rather conservative. Indeed, the new approach maintained backward compatibility with GSM, while trying overcoming some perceived weaknesses of 2G systems. A main heritage of GSM still present in 3G systems is the automatic integrated roaming. 3G systems retain the basic idea of the GSM radio signaling system, that is, the concept that each user has a “home” cell and may be currently visiting another, operated by the home operator (telecom company) of by a local one. In order to find the location of its users (and bill them accordingly) the mobile network relies on distributed location registers, respectively called the Home Location and Visited Location Register (HLR/VLR). The HLR/VLR solution ensures that 3G calls can be set up with the same speed users experienced (and liked) in 2G networks. On the other hand, it preserves operator-based management of user authentication via shared authentication keys stored in SIMs. Like in 2G systems, 3G systems’ users identify themselves by providing the identity stored in their SIM and known to their home service operator, just like users accessing a computer system. 3G authentication was designed with the following requirements in mind.

Despite the given attention to the aforementioned vulnerabilities of GSM networks, during the design of the UMTS, various vulnerabilities have been identified over the 3G systems as well. The official ETSI report over the design and evaluation of the MILENAGE algorithm set, identifies that the prime attack point against the implemented algorithms, is the USIM. For the f2 to f5* functions, it is mathematically proven that, no combination of significantly less than 2 64 output values can be used in order to predict any new output value.

Furthermore, the f1/f1* functions are equivalent to a standard Cipher Block Chaining-Message Authentication Code, while they use distinct output bits acquiring this way sufficient cryptographic separation. A simple internal collision attack exists against the CBC MAC, which requires about 2**64 values. Thus, the report summarises that the f1/f1* functions appear to be sound. The same report also investigates the independence between the f1/f1* and f2 to f5* algorithms, since a connection among them can be exploited by an adversary for the execution of a variety of active attacks. A mathematical analysis of these algorithms supports that a sufficient separation exists, among these two groups of functions.

5.2.1 Possible attacks against UMTS networks

Several attacks against the f1/f1* functions, combinations of f2-f5 and combinations of f1-f1*and f2-f5*have been defined. Theseattacksrequireabout2 64 queries and among them is the well-known CBC-MAC internal collision attack. Yet, these attacks at the time of the report where considered to be impractical. Similarly, a variety of attacks against the Rijndael-AES algorithm have been identified.

Furthermore, attacks based on camping on a false BS and camping on a false BS/MS are not resolved by 3G security architecture. Such an attack requires the use of modified BS or MS in order to entice a subscriber to connect a false BS that acts as a repeater, being able to relay, modify or ignore certain messages between the SN and the subscriber. Similarly, the UMTS security architecture only partly counteract attacks aiming to hijacking incoming and outgoing calls when encryption is disabled, which also require a modified BS/MS. Such attacks are feasible among the periodic integrity protection messages. The same report recognizes exploitable vulnerabilities regarding the impersonation of both the SN and the user and eavesdropping of user data. This is achievable by the enforced use of compromised key vectors. Additional vulnerabilities that have been identified in UMTS networks include:

1. Unencrypted IMSI transmission during new TMSI allocation.
2. The IMEI is not considered a security feature. Thus, it is not protected.
3. No protection against jamming attacks.
4. A subscriber is possible to be enticed to connect on a false BS.
5. GEA0 is supported, meaning that unencrypted communication is possible and acceptable, allowing hijacking and Man-in-the-middle attacks.
6. It is possible to enforce a ME to fall back to GPRS/EDGE, if support of UMTS/HSPA services is not available.

The described vulnerabilities can be exploited for an execution of a wide variety of attacks, these include:

- Denial of service:

Can be achieved by an attacker with a modified ME who is able to send a de-registration request to the SN regarding the legitimate user. The same result can be achieved if an attacker sends a location update request from a different LA than that of a legitimate user. Furthermore, an attacker may entice the legitimate user to connect to a false BTS, completing the attack by blocking the traffic towards the SN.

- Identity catching:

Some explicitly defined cases exist where the IMSI is requested by the network to be transmitted unencrypted. Such cases allow passive or active identity catching, by an attacker who uses a modified MS (passive) or BS (active). SN impersonation and eavesdropping: An attacker is possible to masquerade as the legitimate SN towards the subscriber. This can be achieved by an intruder who sets his modified equipment between the SN and the subscriber, being able at this point to control the level of encryption used. Additionally, an adversary can modify the ciphering capabilities of the ME, causing this way a mismatch with the SN, possibly enforcing it to select the lowest level or complete deactivation of encryption. Finally, an attacker may entice a subscriber to make use of a compromised authentication vector and a modified BS.

- Subscriber's impersonation:

Many feasible methods have been identified in order to successfully achieve such an attack. These include the use of compromised authentication vectors, eavesdropped authentication responses and hijacking of ingoing/outgoing calls with both disabled and enabled encryption

5.3 Security and Privacy deviations in LTE

Despite the fact that LTE was designed based on the security and privacy gaps found in UMTS, there are several deviations yet to be met.

In fact, the key management scheme used by the LTE-AKA evolved from its predecessor, the UMTS-AKA, by overcoming some of the latter's critical weaknesses in generation and delivery of secret keys. In the UMTS-AKA, multiple sets of secret keys are generated in advance by an authentication server, delivered to a serving network, then used one key at a time at the time of authentication by a UE in the serving network. A weakness in this system is that the secret key is generated without being bound to its serving network; the result is that neither a UE nor any other element in the serving network is able to tell if a current secret key was designated for a specific serving network. Furthermore, the UMTS-AKA is inefficient in using network resources because multiple sets of secret keys require storage overhead in the serving network and their delivery wastes bandwidth in the core network. This inefficiency increases in proportion to the increase in the number of UEs in the serving network.

The most prominent enhancement of LTE-AKA over the GSM/UMTS-AKA is to prevent the redirection attacks using false base station by including to authentication vectors. However, the LTE-AKA still lacks privacy of subscribers and the prevention mechanism against DoS attack.

Security and privacy deviations exist also in 4G LTE and this fact can lead to the threats towards the UE:

5.3.1 Possible attacks against LTE networks

Following the afore mentioned, we present below several attacks against LTE.

- Rogue Base Stations

Rogue base stations are unlicensed base stations that are not owned and operated by an authentic MNO. They broadcast a cellular network masquerading as a legitimate carrier network. The necessary hardware to construct these devices can be inexpensively obtained using commercial off-the-shelf (COTS) hardware. The software required to operate a 2G (GSM) base station is open source and freely available, and can be configured to operate as a rogue base station.

Rogue base stations exploit the fact that mobile handsets will attach to whichever base station is broadcasting as its preferred carrier network and is transmitting at the highest power level. Therefore, when a rogue base station is physically proximate to a mobile handset while transmitting at very high power levels, the handset may attempt to connect to the malicious network. Mobile handsets are engineered to be backwards compatible with older cellular systems providing a consistent user experience during mobility. Rogue base stations take advantage of this backward compatibility and exploit weaknesses in these older cellular systems.

At the time of this writing, a majority of rogue base stations broadcast a 2G GSM cellular network. The security protections offered by GSM lack mutual authentication between the handset and cellular network, and strong cryptographic algorithms with keys of sufficient length. Additionally, there is no requirement mandating that the 2G GSM air interface is encrypted.

- Device and Identity Tracking

As previously stated, both the IMSI (UICC) and IMEI (handset) act as unique identifiers. Both of these identifiers can be indicators of who owns a mobile handset and where a device is physically located. It is commonplace today for individuals to constantly keep their mobile

devices physically near them, and if a rogue base station is used to intercept traffic in an area, such as where you reside, the operator of the rogue network may be able to identify whether a specific individual is, or is not, residing within a specific location. This poses a threat to privacy because an eavesdropper can determine if the subscriber is in a given location. Data needed for geolocation is available via signaling channels, and is sent over the air interface during handset attach and authentication.

- Downgrade Attacks

Using a rogue base station broadcasting at a high power level, an attacker can force a user to downgrade to either GSM or UMTS. As of the time of this writing, there are no significant, publicly known weaknesses in the cryptographic algorithms used to protect the confidentiality and integrity of the UMTS air interface. Unfortunately, significant weaknesses exist for the 2G GSM cryptographic algorithms used to protect the confidentiality and integrity of the air interface. Examples of broken 2G cryptographic algorithms are A5/1 and A5/2. Depending on the algorithm negotiated while attaching to the rogue base station, the air interface cryptographic algorithms chosen to protect the air interface may be cryptographically broken, leading to a loss of call and data confidentiality.

Real world deployments utilize GSM networks to connect with LTE networks, which bring this into scope.

- Unauthenticated REJECT Messages

During the UE attach procedure certain messages can be sent before security parameters are negotiated. One of these unauthenticated messages is the ATTACH REJECT message, which prevents a UE from completing the attach procedure. A rogue base station coercing a UE to participate in a UE attach procedure can send this unauthenticated ATTACH REJECT message. In response to receiving this message, a UE will no longer attempt to attach to this, or other LTE networks. Since the ATTACH REJECT message is sent even before the UE can authenticate the network, it is unable to distinguish the rogue base station from a real one. This can cause a Denial of Service (DoS) that may persist until a hard reboot of the UE is performed. Certain baseband implementations will not automatically try to reconnect if this ATTACH REJECT message is received. Similarly, the TRACKING AREA UPDATE REJECT message can be sent by a rogue base station in the same manner, and may have the same effect as the ATTACH REJECT message.

Also femtocells, small extensions of the cellular network - often for personal or business use, technically the standard refers to them as Home Node B (HeNB), introduce many new threats since customers retain physical control and can perform offline attacks on the core network through the femtocell or jamming requires less power because an attacker can be closer. Moreover femtocell attackers can quickly set one up in new location to attract UEs.

6. Conclusion

GSM, UMTS and LTE are similar in many ways in the sense that they inherit most of the elements implemented in the GSM/GPRS and EDGE architecture while changing the names of the elements.

A major similarity is the fact that they all implement radio access points and they use cellular technology; another similarity is that they all employ the use of the HLR as the subscriber database although it is called the Home Subscriber Server (HSS) in LTE.

The differences between most of the technologies are based on the evolvement of the core elements and the access methodologies, bandwidths and modulation types. The table below is used to give a better comparative analysis between the technologies.

	GSM	UMTS	LTE
Access Methodology	TDMA/FDMA	WCDMA	OFDMA/ SC-FDMA
Maximum downlink speed	10-150Kbps	384Kbps	100Mbps
Maximum uplink speed	10-150Kbps	128Kbps	50Mbps
Bandwidth	200 KHz	5 MHz	1.4 to 20 MHz
Modulation types supported	GMSK	QPSK	QPSK, 16QAM, 64QAM
Core Network type	Circuit Switched	Circuit/packet Switched	Fully IP based

Figure 20 Comparison of GSM, UMTS, LTE features

In terms of security and privacy, GSM, perhaps the best known 2G system, provides a range of security features, including authentication of the mobile user to the network, data confidentiality across the air interface, and a degree of user pseudonymity through the use of temporary identities. Third and fourth generation (3G and 4G) systems, such as UMTS/3GPP and Long-Term Evolution (LTE), have enhanced these security features, notably by providing mutual authentication between network and phone, and integrity protection for signaling commands sent across the air interface. However, user privacy protection has remained largely unchanged, relying in all cases on the use of temporary identities, and it has long been known that the existing measures do not provide complete protection for the user identity.

In terms of security, UMTS and LTE also inherit most of the procedures implemented in the GSM/GPRS and EDGE architecture. Although major modifications have been taken place during the evolution from 2G to 3G and 4G. In the figure below we present the security features of GSM, UMTS and LTE.

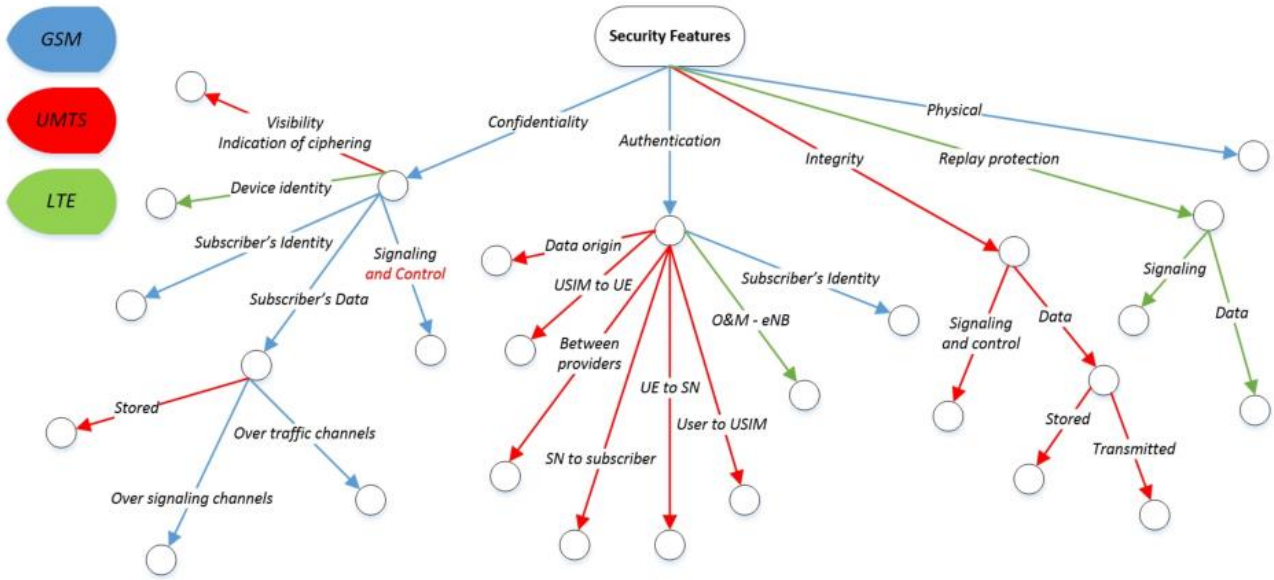

Figure 21 Security features of GSM, UMTS, LTE

In the following pictures we summarize the security processes established in GSM, UMTS and LTE where the evolution of cellular technologies is obvious in terms of security.

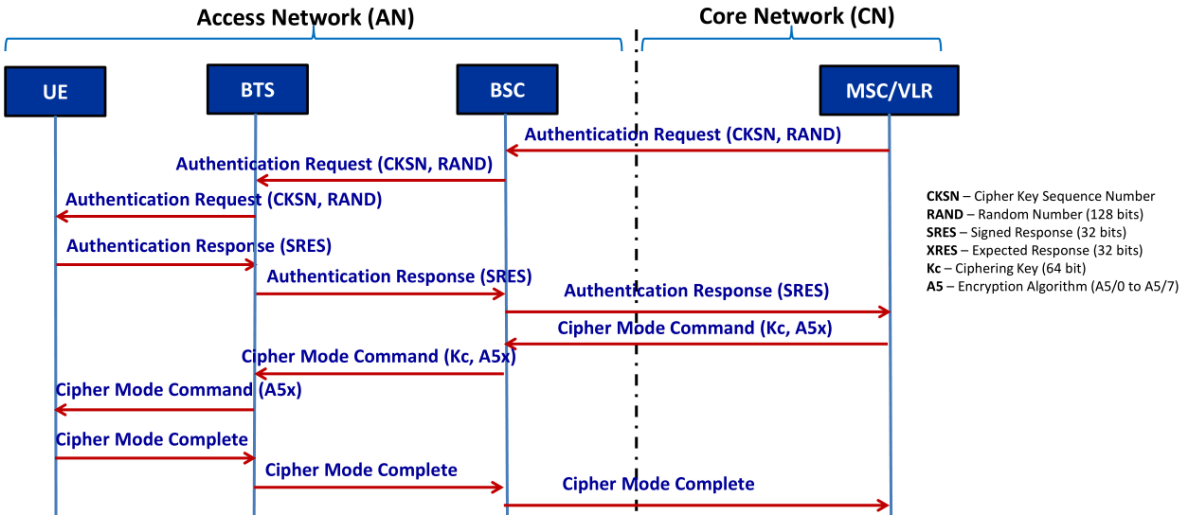

Figure 22 GSM Security process

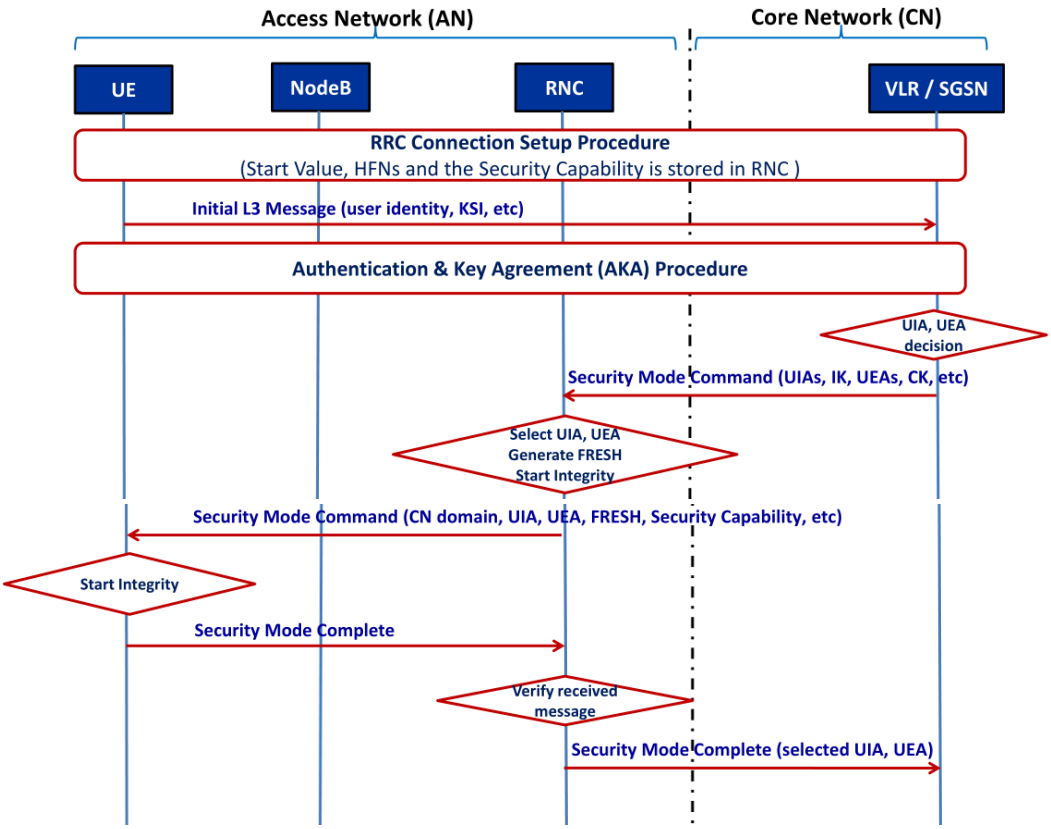

Figure 23 UMTS Security proccess

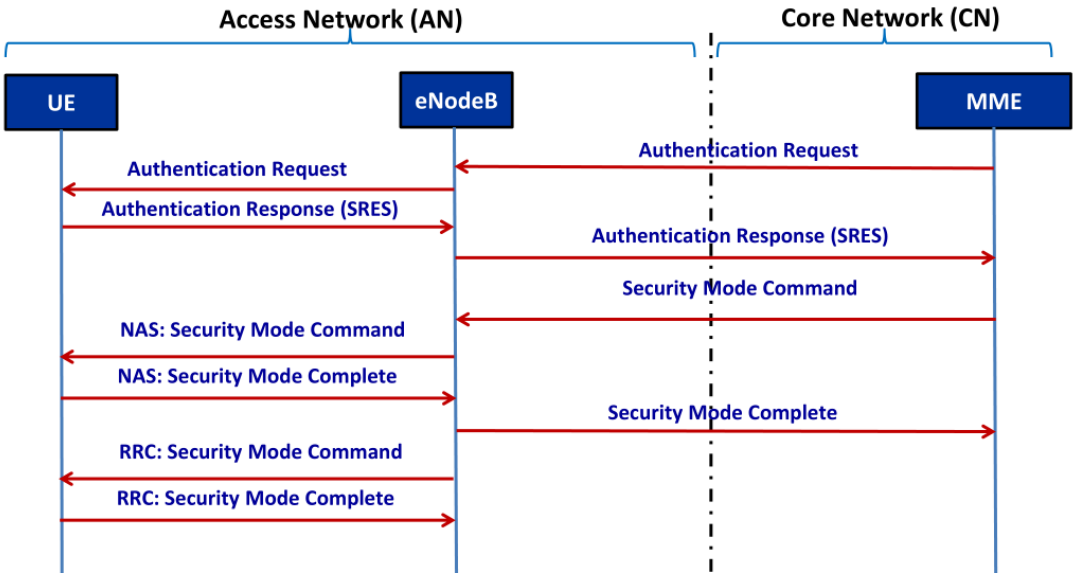

Figure 24 LTE Security process

So, on the one hand, we see that regarding security a lot of improvements have been taken place since the initial 2G concept has been enhanced with mutual authentication, stronger encryption algorithms and integrity algorithms. This evolution through GSM/GPRS to UMTS and LTE is given below:

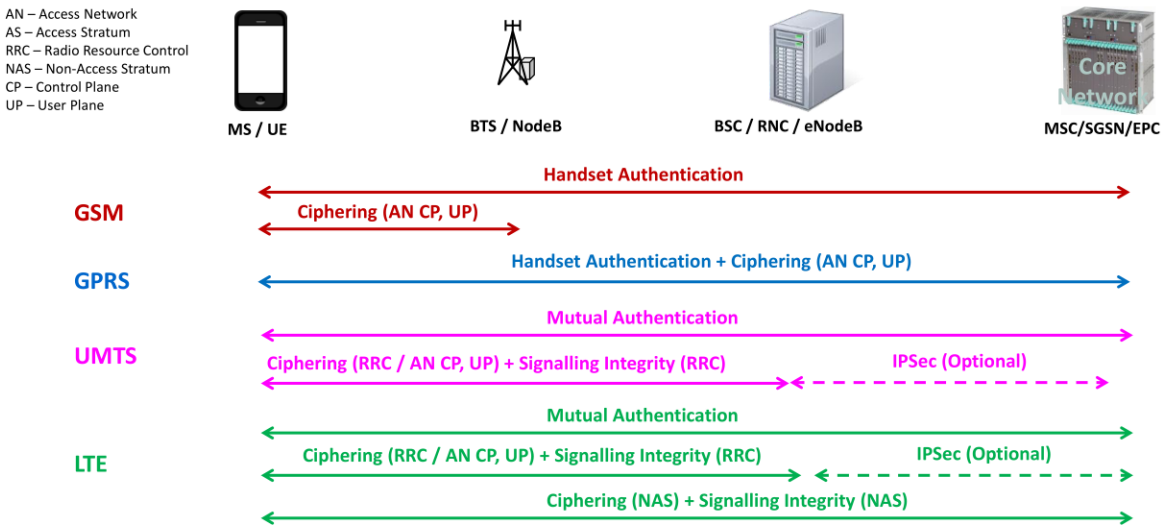

Figure 25 Evolution of Security Architecture

On the other hand, regarding privacy the status has not changed from 2G and the problem of IMSI cleartext transmission still remains.

As discussed above in mobile telephony, each User Subscriber Identity Module (USIM) has a unique International Mobile Subscriber Identity (IMSI). If the IMSI is sent in cleartext across the Air Interface, and the mapping to the phone user is known by an adversary, then a particular user could be tracked using the IMSI, since it is fixed for the lifetime of the USIM.

To avoid this major privacy weakness, the visited network assigns the phone a Temporary Mobile Subscriber Identity (TMSI), which is used for addressing purposes across the Air Interface, and which changes frequently. Of course, if an adversary could link a TMSI to the IMSI then this would represent a breach of user privacy. Furthermore, even though subscribers are given temporary identities for using the network, the issues of identity privacy and location tracking are still open. TMSIs are usually unchanged in a given location (or tracking) area, which is composed by up to a hundred adjacent cells.

Unfortunately, it is necessary to send the IMSI across the air interface in certain circumstances, e.g. when registering with the network after switching on a phone. Despite the use of TMSI or GUTI, the UE must transmit the permanent identity (IMSI) in plaintext when registering for the first time. Traditionally, the 3GPP mobile network does not provide adequate mechanisms for user identity confidentiality protection against active attack. An active attacker equipped with a rogue base station (e.g., IMSI catcher) can send an identity request message in the GSM/UMTS/LTE-AKA to the UE. The UE always respond with its IMSI. Thus, the subscriber identity is not confidentially treated. Disclosure of the IMSI brings about the chance that an adversary could acquire subscriber information, location information, and even conversation information.

An active attacker can exploit this mechanism to perform so-called IMSI catching, i.e. harvest the IMSIs of all subscribers in the vicinity of the attacker's false base station. In GSM, the

attacker can even go further and eavesdrop on the subscriber's traffic by downgrading to weak or no encryption. The latter is not possible in UMTS or LTE due to the mandatory use of signaling integrity. IMSI catching attacks were known already when UMTS was designed, and they were re-discussed during the LTE design phase.

Public key methods (with the public key owned by the visited or the home network), symmetric key methods and the use of pseudonyms are mentioned as potential countermeasures. All these countermeasures have issues, and none of them has finally been adopted in UMTS or LTE, but it may be worth revisiting the arguments. Three of these methods are suggested in section 6.1.

Attacker type		2G	3G	4G
Attacker is outside RAN	Passive	Yes	Yes	Yes
	IMSI catcher	No	No	No
	MitM	No	Yes	Yes
RAN=Attacker	Passive	No	No	No
	Active	No	No	No

Figure 26 Identity protection in 2G, 3G, 4G

Furthermore, the necessity of backwards compatibility and interoperability between the cellular technologies due to coverage limitations of newer technologies, leads to the downgrade to 2G status in terms of security and privacy.

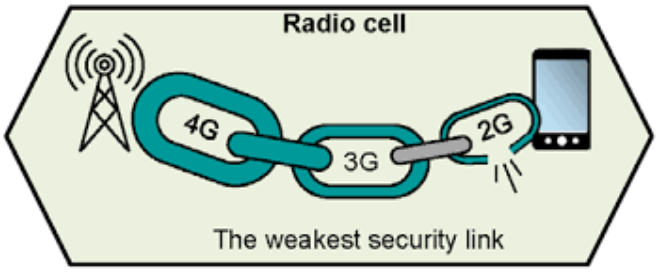

Figure 27 Weakest link in Air Interface

6.1 Related work - suggestions

Several studies have been issued during the last decade concerning Security and Privacy issues in cellular networking. We have noticed and present below the most privacy-oriented of them since we believe that the biggest gap exists in the implementation of user's privacy and not in the implementation of entity authentication where the relevant algorithms can be directly improved.

Major modifications to the air interface protocol, which would require changes to the operation of all the serving networks as well as all the deployed phones. It seems likely that making the necessary major modifications to the operation of the air interface after deployment is essentially infeasible. Many of the proposed schemes also involve the use of

public key cryptography, which has a high computational cost, although there do exist schemes which only use symmetric cryptography. It would therefore be extremely valuable if a scheme offering greater user privacy could be devised which did not involve making significant changes to the existing mobile telecommunications infrastructures, and had minimal computational cost.

Through related work we have noted the following three approaches:

1. A 3-Way Authentication and Key Agreement (PE3WAKA) protocol [31] has been proposed by Geir M. Køien, a researcher in the Network Technologies group of Telenor R&I in order to provide substantially improved user privacy and a 3-way security context.

2. A Secure Mobile Identities (SMI) protocol [11] as a repetitive key-exchange protocol that uses this weak SIM authentication as a foundation to enable mobile users to establish stronger identity authenticity, has been proposed by Varun Chandrasekaran, Fareeha Amjad, Ashlesh Sharma and Lakshminarayanan Subramanian by Entrupy Inc. and New York University.

3. A Pseudonymity approach [22] through the possible use of multiple IMSIs for a single account to provide a form of pseudonymity on the air interface, has been proposed by Mohammed Shafiul Alam Khan and Chris J Mitchell by Royal Holloway, University of London.

Though, we have to take into account that any improvements proposed should not require significant changes to the existing network infrastructure, because obviously significant changes to widely deployed infrastructure are unlikely to be feasible; only realistic and practical proposals have to be made.

7. Bibliography

[1] ETSI, GSM 02.09, Version 8.1.0, 2006

[2] 3GPP, TS 33.102, Version 15.1.0, 2018

[3] 3GPP, TS 33.401, Version 11.5.0, 2012

[4] ETSI, GSM 04.08, Version 5.3.0, 1996.

[5] 3GPP, TS 21.111, Version 12.0.0, 2014

[6] J. Eberspächer, H.-J. Vögel, C. Bettstetter, and C. Hartmann, "GSM Architecture, Protocols and Services", 3rd ed. Wiley, 2009

[7] Gunnar Heine, "GSM Networks: Protocols, Terminology and Implementation", Artech House 1999

[8] Cristina Parra Fernandez, "Development of a tool for GSM networks performance analysis and evaluation", Barcelona 2013

[9] Thomas M. Chen, Nhut Nguyen, "Authentication and Privacy"

[10] T. Halonen, J. Romero, and J. Melero, "GSM, GPRS and EDGE performance. Evolution towards 3G/UMTS", 2nd ed. Wiley, 2003

[11] Varun Chandrasekaran, Fareeha Amjad, Ashlesh Sharma, Lakshminarayanan Subramanian, "Secure Mobile Identities", Entrupy Inc., New York University 2016

[12] Chunyu Tang, David A. Naumann, and Susanne Wetzel, "Analysis of authentication and key establishment in inter-generational mobile telephony", Stevens Institute of Technology 2013

[13] Ravishankar Borgaonkar, Lucca Hirschi, Shinjo Park, and Altaf Shaik, "New Privacy Threat on 3G, 4G, and Upcoming 5G AKA Protocols", sciendo 2018

[14] Chris J. Mitchell, "The security of the GSM air interface protocol", Technical Report RHUL-MA-2001-3, Royal Holloway University, 2001

[15] Dr.-Ing. Andreas Willig, "The GSM Air Interface Fundamentals and Protocols", Communication Networks Group Hasso-Plattner-Institute University of Potsdam, 2003

[16] S. Pütz, R. Schmitz, and T. Martin, "Security Mechanisms in UMTS", 2001

[17] Igor Bilogrevic, "Security and Privacy in Next Generation Mobile Networks: LTE and Femtocells", Laboratory for computer Communications and Applications (LCA1) EPFL, Lausanne, Switzerland 2010

[18] Jeffrey Cichonski, Joshua M. Franklin, Michael Bartock, "Guide to LTE Security", NIST Special Publication 800-187, NIST 2017

[19] "UTRAN Radio Interface protocols, Department of Communications and Networking", Aalto University, Finland

[20] Christos Xenakis, Christoforos Ntantogian, Orestis Panos, "(U)SimMonitor: A Mobile Application for Security Evaluation of Cellular Networks", Department of Digital Systems, University of Piraeus , Piraeus, Greece

[21] Vesa Lehtovirta, "3GPP SECURITY STANDARDIZATION", Ericsson Security Research, NomadicLab, 2016

[22] Mohammed Shafiul Alam Khan, Chris J Mitchell, "Improving Air Interface User Privacy in Mobile Telephony", Information Security Group, Royal Holloway, University of London

[23] Adam Kostrzewa, "Development of a man in the middle attack on the GSM Um-Interface", Technische Universität Berlin Fakultät IV, Institut für Softwaretechnik und Theoretische Informatik, 2011

[24] Asha Mehrotra, "GSM System Engineering", Artech House, 1997

[25] Dr Bhaskar Ramamurthi, “GSM Network Architecture, Channelisation, Signaling and Call Processing”, Department of Electrical Engineering, IIT Madras

[26] Chen Dan, Liu Yijun, Tang Jiali, “Analysis of Security Based on CDMA Air Interface System”, School of Computer Engineering, Jiangsu Teachers University of Technology, 2012

[27] Ralf Kreher and Torsten Rüdebusch, “UMTS Signaling, UMTS Interfaces, Protocols, Message Flows and Procedures
Analyzed and Explained”, John Wiley & Sons Ltd, 2007

[28] Altaf Shaik, Ravishankar Borgaonkar, N. Asokan, Valtteri Niemi and Jean-Pierre Seifert, “Practical Attacks Against Privacy and Availability in 4G/LTE Mobile Communication Systems”, Technische Universität Berlin and Telekom Innovation Laboratories, Aalto University and University of Helsinki, 2017

[29] “Openair LTE Core Network User Plane Training session, OAI workshop, BUPT”, Beijing, 04.27.2017

[30] V. Srinivasa Rao, Rambabu Gajula, Lead Engineer, “Signaling Procedures in LTE”, published in webbuyersguide.com, March 12, 2010

[31] Geir M. Køien, “Subscriber Privacy in Cellular Systems”, 2014