1

Evolution of Cellular Systems

Shahriar Shahabuddin[1], Sadiqur Rahaman[1], Faisal Rehman[1], Ijaz Ahmad[1], and Zaheer Khan[2]

[1] University of Oulu, Finland
[2] University of Liverpool, UK

1.1 Introduction

Wireless communication technologies are essential parts of our lives. From WiFi home networks to sophisticated machine-to-machine communication in the robotics industry, we live in a world of wireless connectivity and it is impossible to imagine a single day without using any wireless devices. The blessings of cellular technologies provided us with a great deal of mobility and thus made it possible to listen to the radio while travelling in a car or on the beach. The cellular devices are also convenient in that we no longer have to worry about the size of the cables to connect to the networks. We are now living in a world where conferences for business meetings, distance and online courses from universities, and medical help over long distances are considered as part and parcel of our daily lives. We have greater access to information than ever before and it is all possible due to the advancements and inventions in cellular communication.

The number of cellular users increased dramatically over the last decade compared to the other technologies and are still increasing. We can see from Figure 1.1 that the fixed broadband or fixed wired subscription did not increase that much in a last decade, while the mobile cellular subscriptions are increasing day by day. With the advent of sophisticated technologies, such as tactile computing, autonomous vehicles, wireless charging, smart living, etc., we can only envision how the use of cellular technologies will grow in the future.

This chapter is dedicated towards the evolution of cellular communication. In that respect, we start by discussing the initial developments and history of cellular systems. We subsequently go through the different generations of cellular systems and have a brief discussion about them. As the topic is broad, we try to confine ourselves to the basic information related to the radio interfaces and network architecture of different generations. We align the chapter with the focus of the book by discussing the evolution of security measurements during each generation.

A Comprehensive Guide to 5G Security, First Edition. Edited by Madhusanka Liyanage, Ijaz Ahmad, Ahmed Bux Abro, Andrei Gurtov, and Mika Ylianttila.

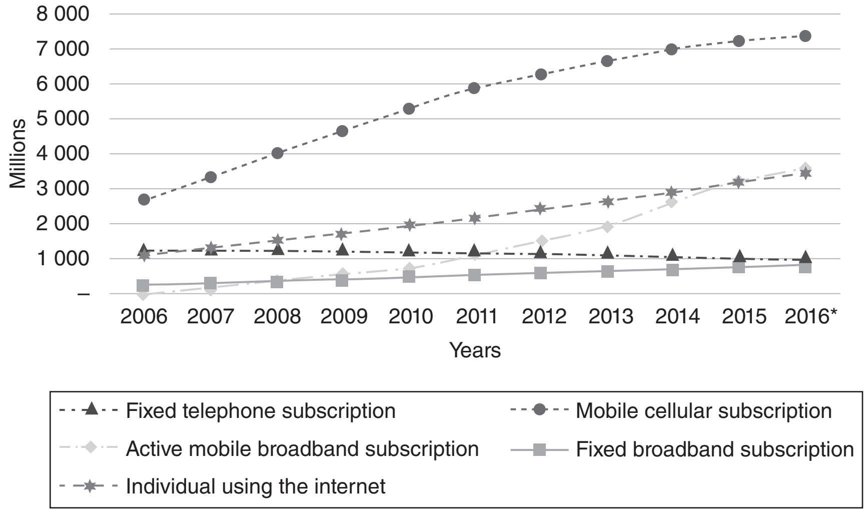

Figure 1.1 Growth of communication services encompassing the last decade.

1.2 Early Development

Wireless communication in its current practice is a very sophisticated technology, making long distance voice, data and multimedia communication possible between people, no matter which part of the world they reside in. The kind of evolution that wireless cellular technologies went through, in particular over the last three decades, and over the last two hundred years in general, makes for a fascinating journey. If we try to trace the initial efforts that became the foundation of the wireless communications of today, we have to go back as early as the ancient Greek, Roma and Chinese cultures, where electrical and magnetic properties of materials were experimented on. The early experiments on electrical and magnetic properties were not intended for wireless communication, since that sort of vision was not present as a motivation for these experiments.

We see that even in the 19th century, when the connection between electricity and magnetism was first developed, the intuition and imagination of what it could achieve was naturally missing amongst the researchers. It is good to say that it was mostly the random experiments that eventually led to the kind of communication systems we have now, and that is something which makes this journey more interesting. Even though, as mentioned above, the experiments towards trying to find electrical and magnetic properties in various ancient cultures, and considered as one of the foremost steps in this journey, it is also important to keep in mind that the last two hundred years present a more coherent and consistent picture that is paved with ground-breaking discoveries.

So, in our analysis the last two hundred years are of primary importance. We have to try to coherently present the connection of all those discoveries as to how one discovery led to another, and what became the motivation to carry out further discoveries. Until this decade, the story is not as linear and direct as it might appear when looking back to its destination. But as far as wireless communications are concerned, it would be unfair and unimaginative to consider this point in time as the final destination, because as far as wireless communication is concerned, the sky is the limit, or even beyond [9].

Starting with the last two hundred years, say the year 1820, the Danish physicist Hans Christian Ørsted, during one of his lectures noticed that when the current from a battery was switched on and off, a compass needle showed the deflection. This observation led him to discover that an electric field creates a magnetic field; more particularly, an electric current produces a circular magnetic field as it flows through a wire.

The connection between electricity and magnetism was of immense importance that rapidly led to further developments. However, it is sometimes claimed that it was Gian Domenico Romagnosi who discovered this connection around two decades before, but the importance of this discovery cannot be considered insignificant. From the years 1823 to 1826, Dominique François Jean Arago, a French mathematician and physicist, discovered something called rotary magnetism, which was termed Arago's rotation. In simple words, he showed that a wire can become a magnet when current flows through it, and that most bodies could be magnetized. These discoveries were further explained by Michael Faraday later. André-Marie Ampère, another French physicist and mathematician, discovered electrodynamics. Ampère showed that two parallel wires carrying electric currents attract or repel each other, depending on whether the currents flow in the same or opposite directions. Ampere's initial plan was to gain more understanding between electricity and magnetism, and this had led him to these discoveries.

Michael Faraday's contributions are very significant in this journey, and he deserves all the credit that we can give him. After Ørsted had discovered the phenomenon of electromagnetism, it motivated many scientists to study this further, the efforts which helped Ampere in his discoveries. Similar motivation led Michael Faraday to carry out experiments, whereby he successfully managed to build two devices to produce electromagnetic rotation. Not only did he discover electromagnetic induction, but also predicted that electromagnetic forces extended the empty space around the conductor. In simple words, he predicted the existence of electromagnetic waves, which proved to be a true prediction later.

Samuel Finley Breese Morse, an American painter, invented the single-wired telegraph system. He was also a co-developer of the Morse code. This discovery also became possible because of the discovery of electromagnetism. The telegraph was important because it was a first attempt to use electromagnetism in an effort to communicate. The list of discoveries continued in the rest of the 19th century, and the German physiologist and physicist Hermann Ludwig Ferdinand von Helmholtz, worked on the phenomenon of electrical oscillation in 1847, which in itself was not a major contribution, but led to the major contribution by Heinrich Rudolf Hertz, one of his students, who later demonstrated electromagnetic radiations. In 1853, William Thomson also contributed in the form of calculating the period, damping and intensity, as the function of the capacity, self-inductance and resistance of an oscillatory circuit. Another proof of Helmholtz's work came from a discovery by Feddersen, who verified the resonant frequency of the tuned circuit, which was suggested by Helmholtz earlier.

James Maxwell is a prominent and influential name in the progression of wireless communication. He proved the existence of electromagnetic waves by formulating the electromagnetic theory of light and developed the general equations of the electromagnetic field, known as Maxwell equations. The most significant aspect of his work was that for the first time it was demonstrated that electricity, magnetism and also light are manifestations of the same phenomenon. This discovery is of absolute importance, because it led to the prediction that radio waves exist, which was a very significant finding for the development of wireless communication. In 1866, the first transatlantic telegraph cable was installed and operated by using the Morse code, with a speed of five words per minute.

The first description of transmission of a wireless signal came in the form of a patent by the American dentist Dr Mahlon Loomis, in 1866. It was the idea of the wireless telegraph, from which he supposedly demonstrated the transmission of a wireless signal between two mountains. In 1882, another patent appeared in terms of wireless signal transmission, when American physicist Amos Emerson Dolbeam, transmitted a wireless signal using an induction coil, microphone, telephone receiver and a battery. In 1887, Hertz, a student of Helmholtz, sent and received wireless waves, using a spark transmitter and a resonator receiver. In 1895, Morse coded wireless signals were transmitted for more than over a mile by Guglielmo Marconi, and he carried out successful reception of a Morse coded wireless signal in 1901, which was sent across the Atlantic. In 1904, the patent of the diode came from J.A. Fleming. The triode amplifier was patented in 1906 by Lee DeForest. In the same year, Fessenden transmitted the first speech signal wirelessly. In 1907, the commercial Trans-Atlantic wireless service was started, which used huge ground stations. In 1915, wireless transmission of voice signals was carried out between New York and San Francisco.

Marconi carried out other ground-breaking and pioneering work in wireless communications by transmitting radio signals over long distances in 1920. Prior to that, Marconi was already working on the concept of wireless telegraphy. The breakthrough in his work came with his conclusion that if the height of the antenna could be raised, then the range of radio signal transmission could be extended, which he developed based on wireless telegraphy, where he grounded his transmitter and receiver. With these improvements, he managed to transmit a signal over 2 miles. He discovered short-wave radio, with wavelengths between the 10 and 100 meters range.

In 1920, we had our first commercial radio broadcast. In 1921, the police car dispatch radios came on the scene. In 1930, the television broadcast experiments were started by the BBC. In 1935, the first telephone call was made around the world. World War II led to rapid advancements in radio technology. In 1947, W. Tyrell proposed hybrid circuits for microwaves, and H.E. Kallaman constructed the VSWR indictor meter. In 1955, John R. Pierce proposed using satellites for communications. Sony marketed the first transistor radio. In 1957, the Soviet Union launched Sputnik I, which transmitted telemetry signals for about five months. The carterfone was a device invented in 1968 by Thomas Carter, which connected a two-way radio to the telephone system, letting one person on the radio talk to another person on the phone.

1.3 First Generation Cellular Systems

The prime developers of the first generation (1G) cellular network were the United States, Japan and some parts of Europe. It was based on analog modulation to provide voice services. In 1979, commercial cellular systems were implemented by Nippon Telephone and Telegraph Company (NTT) in Japan. Nordic Mobile Telephone (NMT-400) is a system developed in 1981 that supports international roaming and automatic handover.

Some European countries implemented this system at that time. Subscribers of NMT-400 were able to transmit up to 15 watts of power using car phones. Six countries – namely Finland, Sweden, Norway, Austria, Spain, and Denmark – adopted NMT-400.

The advanced mobile phone service (AMPS) and its alternative total access communication systems (ETACS and NTACS) were more successful for 1G. From the radio standpoint these above systems were identical. The main difference was the length of the channel bandwidth.

1.3.1 Advanced Mobile Phone Service

The advance mobile phone service (AMPS) was more advanced in comparison to the other 1G systems in the United States. It was deployed in Europe and Japan by an organization named Total Access Communication Systems (ETACS). As mentioned above, from the radio standpoint, the above-mentioned systems were identical, only differing in the length of channel bandwidth. For example, AMPS was based on a 30 kHz bandwidth, while the ETACS and NTACS used 20 kHz and 12.5 kHz for the channel bandwidth, respectively [11].

AT&T and Bell Labs first implemented the AMPS for commercial use in the year of 1983 in Chicago and its neighboring areas, then later in Israel in 1986, in Australia in 1987, and in Pakistan in 1990. By the mid-2000s, all commercial companies discontinued this system from the market around the world. This system was constructed using long base stations (height from 150 ft to 550 ft) with omnidirectional antennas. In the beginning, the carrier to interference ratio (CIR) was kept to 18 dB for better voice quality. Spectrum was assigned by FCC in the USA to two operators in each market, one for the incumbent telecommunications carrier and another for the non-incumbent operator. 20 MHz of spectrum was assigned for each operator, which could support a total of 416 channels. For voice communication, 395 channels were used and the remaining 21 channels were for control information. There were 7-cell frequency re-use patterns, where each sector consisted of 3 sectors per cell. The AMPS is based on the Frequency Modulation for voice communication and used Frequency Shift Keying (FSK) for managing the control channel. After the availability of 2G systems, AMPS were continued by the operators in North America for the purpose of a common fallback service for the entire region and for the roaming service between multiple operators that had implemented 2G systems.

1.3.2 Security in 1G

The first generation (1G) cellular system used analog communication, as stated before. Due to the vulnerable nature of analog signal processing, it was difficult to provide efficient security services for 1G. For example, eavesdropping was a pressing concern for 1G phones, as it was possible for anyone to listen in to a private communication between two users, because all it required was a simple receiver operating at the similar frequencies. There was absolutely no confidentiality in communication in 1G networks. Also, the identity of the cellphone could easily be duplicated, and all the call charges made from the duplicate phone could be directed to the original owner. Since the scale of the network was small, and a small number of users needed servicing, the 1G cellular networks had a limited risk of mass cloning of the mobile sets. Although attempts had been made to completely get rid of mobile set cloning, they were proven to be unsuccessful. Even though the information about the number being dialed could be encrypted, the major problem was transmission through the air, as signals could easily be received by using any FM receiver, since the transmission used frequency modulation [16].

1.4 Second Generation Cellular Systems

The improvement of the processing abilities of hardware platforms made the development of 2G wireless systems possible. Digital modulation scheme was implemented in 2G, targeting the voice market. The overall system performance rapidly improved due to shifting from analog to digital modulation schemes. The total capacity in 2G was improved by using digital speech codecs, implementing time division and Code Division Multiplexing (CDM) techniques for multiplexing several users using a single channel. In 2G, stronger security systems were also introduced by applying encryption algorithms that were absent in the 1G.

Another attractive feature of the second generation along with other new applications was the short messaging service (SMS). The first SMS was sent using Vodafone GSM network on 3 December 1992 in the United Kingdom. Gradually, some European countries implemented this service to notify the users about the voice mail. Nokia released their first SMS supporting mobile phone, which was capable of sending SMS from one user to another. Today, over 23 billion SMS messages are sent from the mobile operator per day, all over the world. The SMS are used for news updates, business alerts, various payments, blogging, voting and for many other uses.

2G systems evolved to support packet data services, while the previous method was the circuit switched data service, which was similar in concept of dial-up modems. Wireless Access Protocol (WAP) was introduced to provide internet contents to handheld devices.

1.4.1 Global System for Mobile Communications

As soon as it became obvious that long-term economic goals in Europe had to be fixed, the CEPT was formed in 1982 by the "Conference Des Administrations Europeans Des Posts et Telecommunications" to address sector needs. The CEPT successively established the "Groupe Spéciale Mobile" (GSM), to develop the specification for a pan-European mobile communications network. The standardized system targeted spectrum efficiency, low mobile and base stations costs, international roaming, better voice quality, compatibility with other systems such as Integrated Services Digital Networks (ISDN), and the ability to support new services. Before GSM, the cellular market was scattered with a variety of mutually incompatible systems implemented in different countries. For example, Scandinavian countries had NMT-400 and NMT-900, the United Kingdom had TACS, Germany had C-450, and France had Radiocom.

The European telecommunications standards institute (ESTI) released the first version of the GSM standard, called the GSM Phase I in 1990. Consequently, many operators implemented GSM and this standard gained acceptance outside of Europe. The standard was eventually renamed as the Global System for Mobile Communications.

The TDMA scheme is used in GSM air interface with a capability of multiplexing eight users in a single 200 KHz channel bandwidth, where the users were separated by different time slots. Gaussian minimum shift keying (GMSK) was introduced as a modulation technique of GSM. Because of the constant envelope property and significant power and spectral efficiency, the GMSK was convenient [1].

A circuit switched data of 9.6 kbps rate was also supported by GSM, along with the voice and SMS service. GSM packet radio systems (GPRS) were introduced by ETSI in the mid-1990s. It was an evolutionary step of GSM systems towards higher data rates. The GPRS and GSM systems both share the same frequency bands, signaling link and time slots. There were four different channel coding schemes to support the data, at the rates of 8 kbps to 20 kbps per slot. Theoretically, the GPRS was able to provide 160 kbps rate, where the 20–40 kbps rate was found in practice.

1.4.2 GSM Network Architecture

The GSM Network architecture is comprised of two major sub-components. This architecture forms the basis of the next generation (3G) systems and LTE. In Figure 1.2, the base station subsystem is comprised of the base-station transceiver (BTS) unit, with which the mobile stations (MS) and the base station controller (BSC) are connected over the air interface. BSC manages the traffic from several BTSs to the switching core. It also manages Mobility across BTSs. Another sub-component is Network Switching

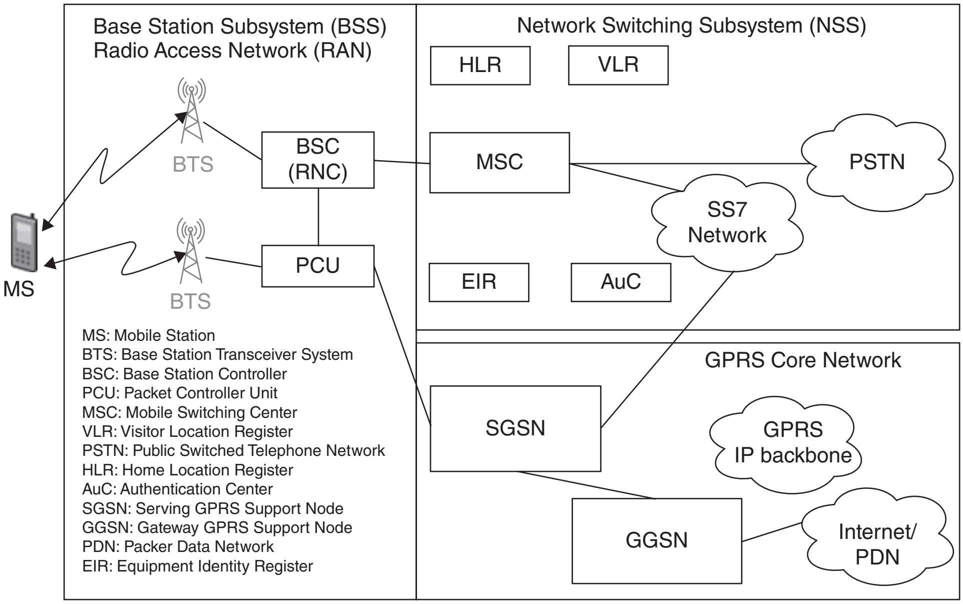
Base Station Subsystem (BSS)
Radio Access Network (RAN)
Network Switching Subsystem (NSS)
HLR
VLR
BTS
BSC (RNC)
MSC
PSTN
SS7 Network
BTS
PCU
EIR
AuC
MS
MS: Mobile Station
BTS: Base Station Transceiver System
BSC: Base Station Controller
PCU: Packet Controller Unit
MSC: Mobile Switching Center
VLR: Visitor Location Register
PSTN: Public Switched Telephone Network
HLR: Home Location Register
AuC: Authentication Center
SGSN: Serving GPRS Support Node
GGSN: Gateway GPRS Support Node
PDN: Packer Data Network
EIR: Equipment Identity Register
GPRS Core Network
SGSN
GPRS IP backbone
GGSN
Internet/ PDN

Figure 1.2 GSM network architecture.

Sub-system. Mobile Switching Center (MSC) and subscriber databases are parts of it. MSC carries out the switching to connect the calling party with the called party. MSC is connected with the Public Switched Telephone Network (PSTN), as shown in Figure 1.2. Home Location Register (HLR) and Visitor Location Resister (VLR) are used to determine the suggested identity of the subscriber for the MSC.

The GPRS system can be upgraded from a GSM system by introducing new components, such as serving GPRS support node (SGSN), and gateway GPRS support node (GGSN), shown in Figure 1.2. For handling data, the packet control unit (PCU) is necessary in the BTS. SGSN was designed to provide location and mobility management. Providing IP access router functionality and connecting the GPRS network to the internet and other IP are the two tasks of GGSN [1].

The data rate of GSM was further increased with the introduction of an enhanced data rate for GSM evolution (EDGE) back in 1997. EDGE specified the use of 8PSK modulation that allows almost three times as much throughput compared to GPRS. An EDGE user could enjoy 80 kbps to 120 kbps of data rate.

1.4.3 Code Division Multiple Access

Code Division Multiple Access (CDMA)-based digital cellular technology was first proposed by Qualcomm in 1989. In 1993, Qualcomm obtained the acceptance of telecommunication industry association (TIA) to embrace their proposal as an IS-95 standard, which was the alternative to IS-54 TDMA, which was adopted earlier as the digital evolution of AMPS. Unlike GSM, multiple users share the same frequency band at the same time in IS-95 CDMA. A unique orthogonal spreading code is assigned for each user that helps to distinguish between different users on the receiver side. Spread signals showed noticeable improvement to multipath fading and interference. The channel bandwidth of IS-95 CDMA is 1.25 MHz for transmitting 9.2 kbps of lower voice signal.

The technical advantages of IS-95 CDMA were more capacity in per MHz of bandwidth, there was no limitation of built-in limit of number of users, power consumption was low so cell size of IS-95 was larger, and soft handoff was introduced. Another interesting feature was the ability to detect the period of silence so that transmission of data could be paused to save energy and increase overall efficiency. The above features gave CDMA systems a huge commercial and user acceptance.

Supplemental Code Channel (SCH) was introduced in the version of IS-95B. It is also known as packet mode transmission for increased efficiency. For example, it supports 14.4 kbps, which is allowed to combine 7 SCH to maintain the peak data rate of 115.2 kbps.

1.4.4 Security in 2G

The 2G cellular network was developed due to an increasing need for improved transmission quality, capacity and coverage. The advancements in semiconductor technology and microwave devices made digital transmission possible in mobile communications. 2G cellular networks incorporated data communications, unlike 1G, amongst other kinds of digital services such as text messages, picture messages and MMS (multimedia messages). With digitized services coming into play, data confidentiality and security became of major concern. 2G cellular systems, in general, comprises of GSM, digital AMPS (D-AMPS), CDMA, and personal digital communication (PDC).

GSM is the most successful and widely-used standard in cellular communications throughout the world, as part of 2G cellular networks. It includes GSM900, GSM-railway (GSM-R), GSM1800, GSM1900, and GSM400. 2G phones using GSM were first introduced around 1990, first deployed in Finland in July 1991. IS-95, or CDMAONE, another technology under the 2G umbrella, based on CDMA, unlike GSM, which is Time Division Multiple Access (TDMA)-based. However, the use of GSM is much wider in scale than IS-95. The successor of GSM is wideband CDMA (W-CDMA), while the successor of IS-95 is CDMA 2000. In order to understand the security measures in 2G cellular networks, it is convenient to first focus on the security in the GSM. Supplemental Code Channel (SCH) was introduced in the version of IS-95B. It is also known as packet mode transmission for increased efficiency.

1.4.5 Security in GSM

GSM tries to focus on four aspects of security that include authentication of a user, ciphering of data and signaling, confidentiality of user identity, and the use of subscriber identity module (SIM) as a security module. SIM is another distinguishing feature of 2G cellular networks. The SIM is basically a detachable smart card containing subscriber information, and used for proving its identity with the operator along with the information regarding the kinds of services it is allowed to access. It plays a vital role in the security process. Authentication requires any particular user to prove that they are a valid customer requesting the service from a particular operator. Ciphering takes care of the interception of all the data and signaling. In order to handle confidentiality, GSM uses international mobile subscriber identity (IMSIs) and, more particularly, uses Temporary Mobile Subscriber Identity(TMSI) to provide confidentiality for the user, by making sure that the information of any particular user being in any particular area is not disclosed to anyone to avoid any intrusion of confidentiality. The SIM card uses algorithms to develop a secure connection with the operator to carry out safe communication. In case the SIM card is taken by an unauthorized person, there is still a PIN code security measure in place.

GSM uses A3 and A8 algorithms between a mobile station and the GSM operator. These are the symmetric algorithms where the same key is used for the encryption and decryption. These algorithms have a one-way function, meaning that output can be found if the inputs are known, but the opposite is not possible. These algorithms are implemented in the SIM card. The technical details of these algorithms are further explained in [12].

1.4.5.1 IMSI

International Mobile Subscriber Identity (IMSI) represents the unique number for every subscriber in the world, and carries the information regarding the home network of the subscriber and country it belongs to. This particular information can be read from the SIM if local access to the SIM exists. It basically comprises of up to 15 decimal

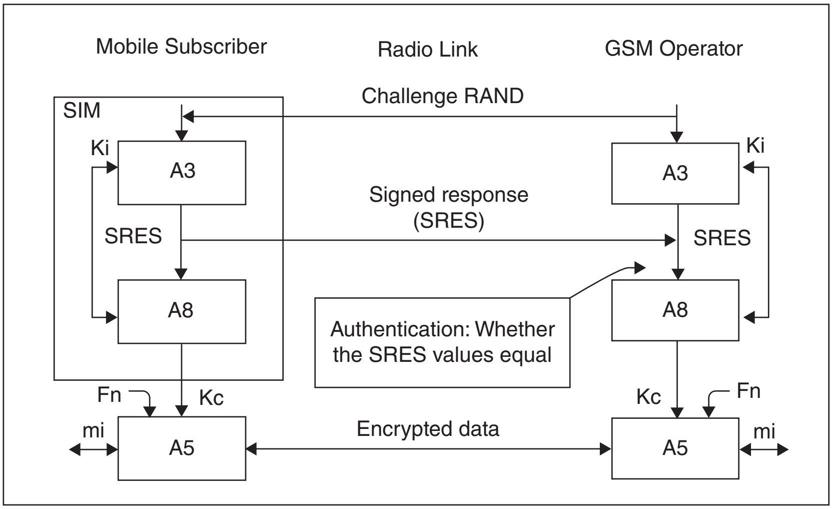

Figure 1.3 GSM authentication process.

digits, out of which the first 5 or 6 specify the network and the country. In order to prevent the eavesdropping, the IMSI is rarely sent, as instead the randomly-generated TMSI is used [14].

1.4.5.2 Ki

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

1.4.5.3 A3 Algorithm

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

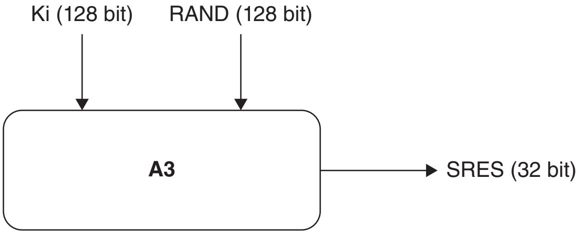

Figure 1.4 The A3 algorithm.

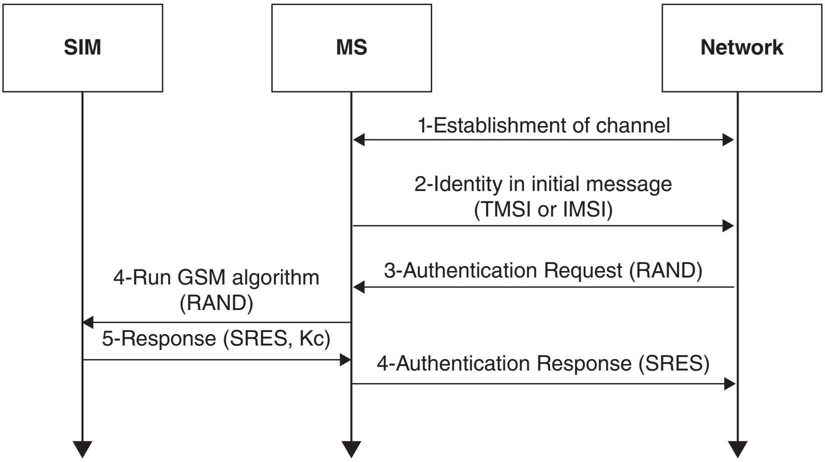
SIM
MS
Network
1-Establishment of channel
2-Identity in initial message (TMSI or IMSI)
3-Authentication Request (RAND)
4-Run GSM algorithm (RAND)
5-Response (SRES, Kc)
4-Authentication Response (SRES)

Figure 1.5 Working principle of A3 algorithm.

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

1.4.5.4 A8 Algorithm

GSM uses ciphering to protect both user data and signaling at an air interface. Once the authentication has been successfully carried out, the RAND coming from the network together with the Ki coming from the SIM, are sent through an A8 ciphering key generating algorithm to create a ciphering key (Kc). This Kc created by the A8 algorithm is used with the A5 ciphering algorithm to cipher or decipher the data. The A5 algorithm is implemented in the hardware of the mobile phone as it encrypts and decrypts data in the air. Whenever the A3 algorithm is run to generate the SRES, the A8 algorithm also runs. Other than the A8 generating the ciphering key Kc, the network also generates the Kc, and shares it with the base stations handling the connection.

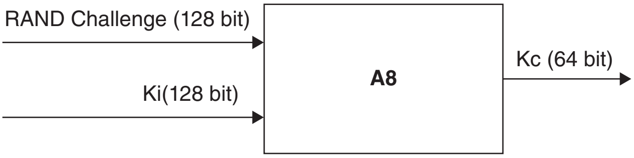

Figure 1.6 The A8 algorithm.

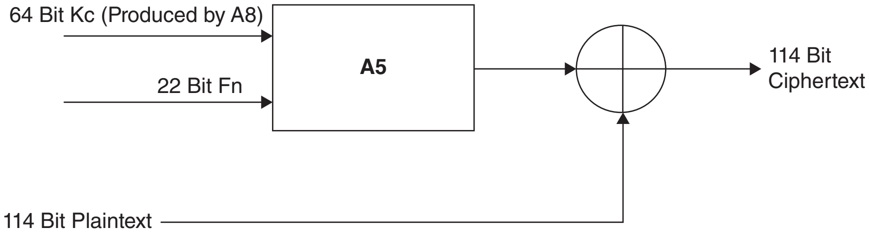

Figure 1.7 The A5 algorithm.

1.4.5.5 COMP128

COMP128 is technically a hash function, which is the implementation of A3 and A8 algorithms in the GSM standard. It is used to provide authentication and helps derive the cipher key (A3/8). GSM allows every operator to use their own A3/8 algorithm, and all the systems support this without the need for a transfer between the networks, even during roaming. However, the operators normally use, for example, COMP128 design, because it requires certain levels of expertise to make their own A3/8 algorithm.

1.4.5.6 A5 Algorithm

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

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

1.4.6 Security in IS-95

The procedures for authentication and security used in IS-95 are the same as in GSM; however, IS-95 uses an additional security technique known as the "private long code mask".

For authentication, both the subscriber and the network use a secret key code. When any subscriber wishes to access the network, the network generates a random code and sends it to the subscriber. The secret key and the random code are used by the subscriber and network to generate another signal. This signed response is then sent to the network by the subscriber, where it is compared to the signed response stored in the network. If the signal matches, access is given to the system.

The additional feature of IS-95, private long code mask, just like the authentication key, is stored in both the subscriber and the network. It is like the public long code mask, which is an electronic serial number transmitted without protection used in analog mode, except that it is more secure. The mobile or the system can initiate operation with a private long code mask by transmitting a "Long Code Transition" order after the call is set up.

1.5 Third Generation Cellular Systems

Third generation (3G) systems provided the higher data rates along with the higher voice capacity and also the advanced features such as applications like multimedia. The planning for the 3G was started in the early 1990s, with the invitation of proposals by International Telecommunications Union (ITU) known as IMT-2000. They started with the investigation of spectrum for these systems. The goal was to implement specifications for global harmony for mobile communication, which is able to initiate global interoperability by providing lower costs. ITU set the requirements for the data rates as the criterion for IMT-2000:

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

Apart from the above requirements, the 3G systems were also intended to provide better quality of Service (QoS) for voice telephony and interactive gaming to internet browsing, e-mailing, and streaming multimedia applications.

1.5.1 CDMA 2000

The 3G standard for IS-95 was known as CDMA 2000 by the CDMA community. In 1999, the standard committee named as the third generation partnership project 2 (3GPP2), took the responsibility of official standardization process of CDMA 2000 from the development group Qualcomm and CDMA. CDMA 2000-1X was the first version of IS-95, where the channel bandwidth of 1.25 MHz was the same as IS-95. The data capability was enhanced by adding supplemental channels, which were actually separate logical channel. The capacity of each individual of fundamental channel was 9.6 kbps, where the capacity increased to 307 kbps by using the multiple supplemental channels. As this specification of channel capacity was in accordance with 3G requirements, it was instead called 2.5G. Gradually, in the version of CDMA 2000-3X, the data rate increased to 2 Mbps by using multiple carriers. Coherent modulation was introduced to improve the uplink channel quality. The capabilities of antennas were advanced by using transmit diversity and incorporating beam steering option. The key point of these upgrades was the backward compatibility. Both A and B versions of IS-95 and CDMA could be implemented in the same carrier, which is convenient for migration of those technologies [20].

1.5.2 UMTS WCDMA

As the popularity of GSM was at its peak, a joint collaboration was formed named 3GPP in 1998 by six regional telecommunication bodies from all over the world. The purpose was to continue the development of UMTS along with other standards of GSM.

The first UMTS standards of 3G were published in 1999, which is known as UMTS Release 99. It brought global success, which can be seen in the statistics of 3G Americas, and the UMTS Forum in May 2010 recorded that the total number of operators of the UMTS network were 346 in over 148 countries. The number of subscriber at that time was 450 million [8].

The architectural design of UMTS was kept the same as the GSM/GPRS network, but the 3G air-interface known as Wide-band CDMA (WCDMA) was a huge modification compared to the 2G air-interface. The design of WCDMA was provoked due to the success of IS-95. WCDMA is actually Direct Spread Spectrum CDMA systems, where user data is multiplied with pseudo random codes to provide synchronization, channelization and scrambling. This system is designed to operate in the 5 MHz bandwidth, which can support 100 different voice calls simultaneously. The peak data rate then varies from 384 to 2048 kbps. In addition, WCDMA supports using multi-code for increasing data rate for a single user.

1.5.3 UMTS Network Architecture

The UMTS network architecture consists of User Equipment (UE), UMTS Terrestrial Radio Access Network (UTRAN) and Core Network (CN). These three major subsections are shown in Figure 1.8. Looking at the first subsection, the components are similar to the 2G network. In UTRAN, Node Bs are connected to the RNC to manage radio resources in data link layer. This section has been developed for the service access point, which was absent in 2G. The CN part is also similar to the 2G, which controls different location registers. The function of SGSN and GGSN are kept the same as its ancestor, 2G.

There are a few differences in the architecture of 2G and 3G. For example, in 3G the base stations are known as Node B, which are controlled by the RNC, then the connection is made with a core switching network for voice calls and data traffic. RNC are connected to one another to confirm soft handover with lowest call drop rate. Except the RNC and Node B in the 3G architecture, there are no major differences with 2G in the overall design. It meant that those two architectures could connect with each other and work with the same core, packet switching and charging network. Only the modulation scheme is the major difference in 3G architecture.

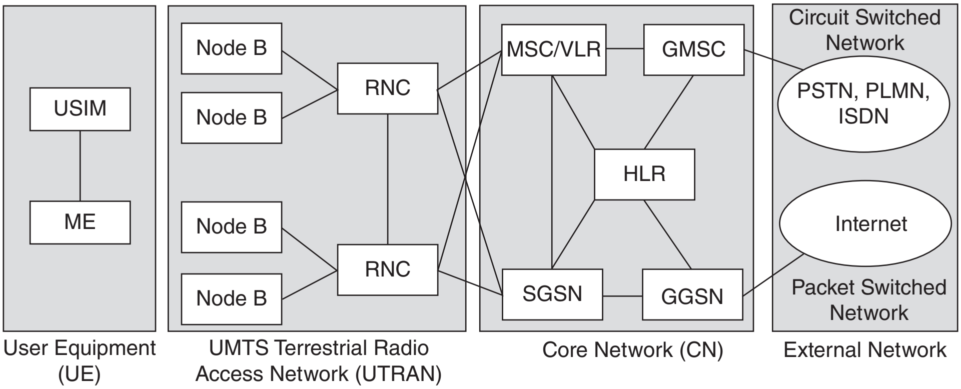

Figure 1.8 UMTS Radio Access Network.

1.5.4 HSPA

High-speed packet access is shortend HSPA. There are two key enhancements combined by 3GPP to UMTS-WCDMA, which are:

1) High-Speed Downlink Packet Access (HSDPA), introduced in the Release 5 in 2002; and
2) High-Speed Uplink Packet Access (HSUPA), which was available in Release 6 in 2004.

AT&T implemented the first HSDPA network, after which around 303 operators from 130 countries worldwide deployed HSDPA. Mostly, the HSPA was deployed as an upgrade of software to the existing UMTS systems [5].

According to internet usage patterns in the late 1990, the users demanded higher speed of downloads. So UMTS evolution focused on improving the downlink. Later, the HSDPA introduced a new downlink transport channel that was capable of providing up to 14.4 Mbps theoretically. It is named the High-Speed Downlink Shared Channel (HS-DSCH). Here the multiplexing technique was time division multiplexing with a limited use of CDM. HSDPA consists of 16 Walsh codes, 15 of which were used for traffic; 5, 10 or 15 codes could be used for a single user to gain higher throughput. The channel frame length was 2 ms, unlike WCDMA of frame length of 15, 20, 40 or 80 ms. In practice, the user of HSDPA was able to obtain throughputs in the range of 500 kbps to 2 Mbps.

1.5.5 Security in 3G

As mentioned above, 3rd generation cellular networks introduced services such as video, audio and graphics applications. It also introduced video telephony and video streaming via cellular networks communication. It was an attractive feature of mobile cellular networks, when looking at the evolution it went through. Extrapolating from the limitations of the 1st generation cellular networks, it was a landmark of sorts. CDMA 2000 and UMTS CDMA came under the 3G umbrella [13].

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

1.5.6 Security in CDMA2000

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

The authentication and key management (AKA) protocol used in CDMA 2000 is the UMTS AKA mechanism described in the next section. The AKA procedure is executed in two stages. The first stage involves transfer of security credentials (authentication vector, AV) from the home environment (HE) to the serving network (SN).

The HE mainly contains the HLR and AC, and the SN consists of the parts of the core network that are directly involved in setting up connections.

In terms of access security, the SN network elements of interest are the PDSN, which handles packet-switched traffic, and the circuit switched nodes VLR/MSC. An operator with a physical access infrastructure will normally have both HE and SN nodes [8].

1.5.7 Security in UMTS

The UMTS security architecture is grouped together in five different sets of features, as shown in the Figure 1.9. The description of these groups of features is given as:

1) *Network access security*: provides the subscriber with secure access to the 3G services, and gives protection against attacks to the radio interface;
2) *Network domain security*: allows all the subscribers to be able to securely exchange signaling data and provides protection against attacks to the wireline network;
3) *User domain security*: deals with secure access to mobile stations;
4) *Application domain security*: makes sure that applications in the user and provider domain are able to communicate with each other securely;
5) *Visibility and configurability of security*: provides security information to the users, as to which security features are in place, and whether a certain security feature requires activation or not.

The network access security features described above can be further classified into the following two categories. The categories with their description are explained below:

1) *User authentication*: is the property of the network that provides service confirming the validity of the identity of the user, and
2) *Network authentication*: is the property that the user validates, which is connected to a serving network with is authorized by the user's home network.

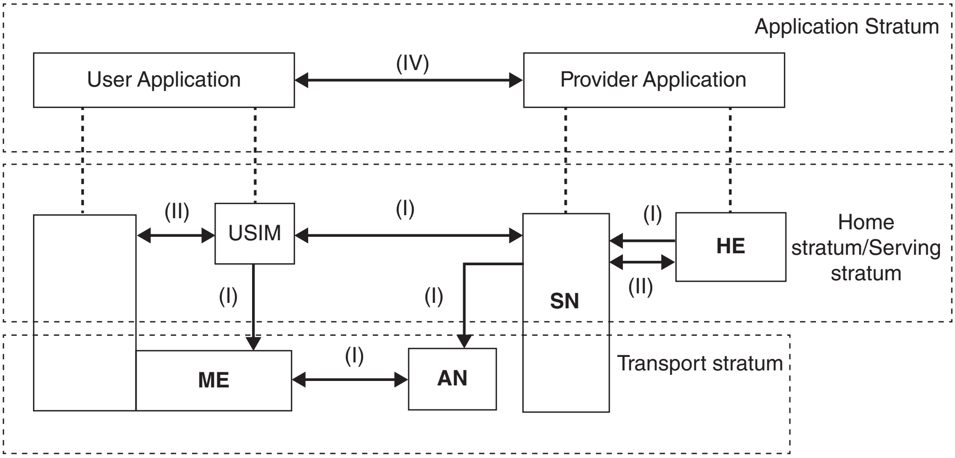

Figure 1.9 Overview of UMTS security architecture.

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

- *Cipher algorithm agreement*: the property that ensures that subscriber and serving network can securely decide on the algorithm that should be subsequently used;
- *Cipher key agreement*: the property that subscriber and the serving network mutually decide on the cipher key that should be subsequently used;
- *Confidentiality of user data*: the property that ensures that user data is protected on the overhead, such that it cannot be overheard; and
- *Confidentiality of signaling data*: the property that ensures that signaling data cannot be overheard on the radio interface.

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

- *Integrity algorithm agreement*: the property that the subscriber and the serving network can securely decide on the integrity algorithm that should be subsequently used;
- *Integrity key agreement*: the property that the subscriber and the serving network agree on an integrity key that shall be subsequently used; and
- *Data integrity and origin authentication of signaling data*: the property that the subscriber or serving network is able to verify that signaling has not been modified later after it was sent by the sending entity, and that the origin of the signaling data received is the valid one.

UMTS AKA is a security mechanism used to accomplish the authentication features and all of the key agreement features described above. This mechanism is based on a challenge/response authentication protocol implemented in such a way as to achieve maximum compatibility with GSM's subscriber authentication and key establishment protocol, so that the transition from GSM to UMTS can be made. A challenge/response protocol is a security measure used by one entity to verify the identity of another entity, without revealing a secret password shared by the two entities involved [23]. Each entity must prove to the other that it knows the password without actually revealing the information that it has knowledge of the password.

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

Table 1.1 Structure of an authentication vector.

Field	Description
RAND	Random Challenge
Ck	Cipher key
Ik	Integrity Key
AUTN	Authentication Token
XRES	Expected Response

Table 1.2 Structure of AUTN field of an authentication vector.

Field	Description
SQN	Sequence Number
AMF	Authentication management Field
MAC-A	Message authentication code

1.6 Cellular Systems beyond 3G

We present an overview of HSPA+, WiMAX and LTE in the following subsections. Although many industries started marketing the WiMAX as 4G systems, technically that was not the case. From an engineering perspective, both WiMAX and LTE represent a break from conventional 3G systems in terms of air-interface technology and network architecture both. These systems are capable of providing throughput level in megabit per second by using advanced signal procession techniques [6,22].

1.6.1 HSPA+

In June 2007, the Release 7 of 3GPP made an enhancement as a further evolution of HSPA. It is sometimes referred to as HSPA+. The key technical enhancements of HSPA+ are achieving higher-order modulation with multiple input multiple output (MIMO) by gaining higher peak rates, operation in dual-carrier downlink, packet connectivity when required to improve battery life, improved mobile receivers for capacity enhancement, and improved data rate and using single frequency network for better performance in multi-cast and broadcast. In May 2010, 56 operators in 34 countries deployed HSPA+ [7,25].

1.6.2 Mobile WiMAX

The institute of Electrical and Electronic Engineers (IEEE) established a group, named 802.16, to develop a standard for wireless metropolitan area network (WMAN). They introduced a standard for fixed wireless application in 2001, and gradually this was enhanced to support mobility. This revised version was known as 802.16e in 2005 and renamed as Mobile WiMAX. The industry consortium, named the Worldwide Interoperability for Microwave Access (WiMAX) Forum, was formed in 2001. The main purposes were to promote, develop, perform interoperability, conformance testing and certify end-to-end wireless systems based on the IEEE 802.16 air-interface standards. In 2007, WiMAX gained the approval of ITU as an IMT-2000 terrestrial radio interface option called IP-OFDMA. As the WiMAX network is designed using IP protocols, which does not offer circuit switched voice telephony, voice services can be provided using the VoIP (Voice over Internet Protocol). Within 2010, there were 504 WiMAX network operators in 147 countries. The notable thing is the number of motivational aspects in LTE design; for example, usages of OFDM and OFDMA technology, was inspired by the implementation of WiMAX [6,17,19].

1.6.3 LTE

The drastic growth of use of the Internet was the motivation for mobile broadband. As the mobile devices were continuously integrating various applications dealing with information, communication and medium of entertainments, it was the demand of time to enable the on-demand access to multimedia content from anywhere. Statics showed that by the end of March 2009, the number of mobile broadband subscribers reached 225 million. To meet this huge number of services with higher performance, LTE design integrates some important radio and core network technologies. Amongst those technologies, the key features of LTE are discussed in three subsections below [10,24]:

1.6.3.1 Orthogonal Frequency Division Multiplexing (OFDM)

Orthogonal Frequency Division Multiplexing (OFDM) was the key difference between the existing 3G systems and the LTE. The traditional 3G systems were based on UMTS and CDMA 2000, where CDM techniques were used. OFDMA provides high data rates along with many more advantages. Due to high data rates, there are more probabilities of intersymbol interference because of multipath. OFDMA was the solution for the problem, by using multicarrier modulation, where high bit rate data streams are divided into several parallel lower bit rates. OFDMA also reduced the computational complexity, because of the implementation of Fast Fourier Transform (FFT). There were other advantages such as coding and interleaving diversity, efficient multicarrier scheme, efficient support of broadcast services, etc.

1.6.3.2 SC-FDE and SC-FDMA

To achieve better battery life, the Single Carrier Frequency Equalization (SC-FDE) transmission method used to transmit the data symbols are sent as a sequence of QAM symbols with an added cyclic prefix. For the uplink of LTE implements, SC-FDMA (multiple version of SC-FDE) allows multiple users to use parts of the frequency spectrum. The complexity of the transmitter and receiver is increased for using these systems.

1.6.3.3 Multi-antenna Technique

The multi-antenna technique provides the solutions of system capacity, link robustness and spectral efficiency. It is possible to combat multipath fading and obtain transmit diversity by using multi-antenna. Beamforming is possible by using multi-antenna so that the transmitted signals can be directed towards the most efficient direction of the receiver. It reduces the signal-to-interference ratio. Another important feature is multiuser MIMO, which allows multiple users in the uplink.

1.6.4 LTE Network Architecture

There are a few differences between the architecture of UMTS and the LTE systems architecture, which is depicted in Figure 1.10. Unlike the UMTS architecture, there is no RNC, SGSN and GGSN blocks in the LTE. In LTE, the Node B is known as eNode B, which is connected to Serving Gateway (S-GW) to terminate interface towards the 3GPP radio access network and Packet Data Network Gateway (P-GW) to control IP data services, including routing, allocating of IP address, enforcing policy, and

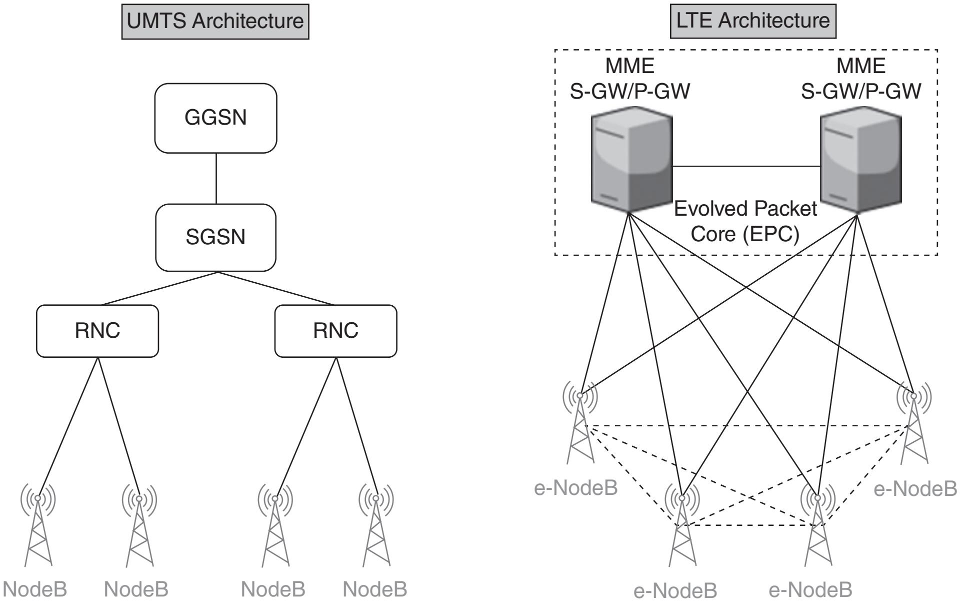

Figure 1.10 UMTS vs. LTE architecture.

providing access to non-3GPP access network. Function of Mobility Management Entity (MME) is to support equipment context and identity by authenticating to the authorized users [21]. The S-GW and P-GW are connected to each other and known as the Evolved Packet Core (EPC). EPC provides the function of access control, packet routing and transfer, mobility management, radio resource management, security and network management [15].

1.7 Fourth Generation Cellular Systems

4G is the fourth generation of mobile telecommunication technology. The requirements of 4G systems are defined by the ITU in IMT Advanced. The requirements are:

- a high degree of sharing features world-wide, which will support a vast range of services and applications with cost efficiency;
- internetworking compatibility within the IMT and also with other radio access networks;
- services compatibility with fixed and IMT networks;
- mobile devices with high quality;
- worldwide roaming capability;
- user-friendly equipment, services and applications; and
- 100 Mbps for high mobility and 1 Gbps for comparatively low mobility devices for supporting advanced services.

These requirements do not quantify the performance requirements, except the last one. In the detailed description of the IMT-Advanced, specific goals were set average and cell-edge performance to the usual peak data rates.

1.7.1 Key Technologies of 4G

1.7.1.1 Enhanced MIMO

Multiple-Input Multiple-Output (MIMO) is a key technique in the modern cellular system, which refers to the use of multiple antennas at both the transmitter and receiver sides. Therefore, base stations and terminals are equipped with multiple antenna elements intended to be used in transmission and reception to make MIMO capabilities available at both the downlink and the uplink.

Enhanced MIMO is considered as one of the main characteristics of LTE-Advanced that will allow the system to meet the IMT-Advanced rate requirements recognized by the ITU-R. The majority of the MIMO technologies already presented in LTE are expected to remain playing a vital role in LTE-Advanced, namely beamforming, spatial multiplexing and spatial diversity. However, further improvements in peak, cell-average and cell-edge throughput need to be obtained to significantly increase performance.

The above-mentioned techniques need some level of channel state information (CSI) at the base station, so that the system can adjust to the radio channel conditions and substantial performance improvement can be attained. For TDD systems, this information is easily collected from the uplink, provided the channel fading is adequately slow, due to the fact that the same carrier frequency is used for transmission and reception. Again, due to the asymmetry of FDD systems, feedback information over the reverse link is required. Full CSI could cause an additional overhead that might be too much, so quantization or statistical CSI are preferable in practice. In addition, terminal mobility can pose serious difficulties to the system performance, as the channel information arriving at the eNB may be outdated.

Multi-antenna techniques in a multi-user situation has the role of delivering streams of data in a spatially multiplexed fashion to the different users in such a way that all the degrees of freedom of a MIMO system are to be used. The idea is to perform an intelligent Space-Division Multiple Access (SDMA), so that the radiation pattern of the base station is adapted to each user to obtain the highest possible gain in the direction of that user.

1.7.1.2 Cooperative Multipoint Transmission and Reception for LTE-Advanced

4G cellular networks have to instantaneously provide a large number of diverse users with very high data rates, and the capacity of the new radio access systems needs to be enlarged. Conventionally, in cellular systems, each user is allocated to a base station on the basis of principles such as signal strength. At the terminal side, all the signals arriving from the rest of the base stations in the form of interference radically limit the performance. The user also connects with a single serving base station while causing interference to the rest of them. Due to the interference limitation of cellular systems, the task of high data delivery cannot be accomplished by simply increasing the signal power of the transmission. Each base station processes in-cell users independently, and the rest of the users are seen as inter-cell interference whose transmission power would also be increased.

CoMP in the framework of LTE-Advanced involves several likely coordinating schemes among the access points. Coordinated beamforming/scheduling is a simpler method, where user data are transmitted only from a single cell. Joint processing techniques require multiple nodes to transmit user data to the UE. Two approaches are being considered as joint transmission, which requires multi-user linear precoding, and dynamic cell selection, where data is transmitted from only one cell that is dynamically selected.

1.7.1.3 Spectrum and Bandwidth Management

To meet the requirements of IMT-Advanced, as well as those of 3GPP operators, LTE-Advanced considers the use of bandwidths of up to 100 MHz in the following spectrum bands:

- 450–470 MHz band (identified in WRC-07 to be used globally for IMT systems);
- 698–862 MHz band (identified in WRC-07 to be used in Region 22 and 9 countries of Region 3);
- 790–862 MHz band (identified in WRC-07 to be used in Regions 1 and 3);
- 2.3–2.4 GHz band (identified in WRC-07 to be used globally for IMT systems);
- 3.4–4.2 GHz band (3.4–3.6 GHz identified in WRC-07 to be used in a large number of countries); and
- 4.4–4.99 GHz band.

1.7.1.4 Carrier Aggregation

In order to utilize the wider bandwidths of up to 100 MHz, a carrier aggregation scheme is required. It is designed in such as way that it consists of several component carriers of 20 MHz bandwidths, so that the LTE-Advanced devices can use greater amounts of data by using several carriers. For example, in a contiguous band, the scenario of using the component carriers for LTE and LTE-Advanced is shown. Also, there are proposed methods available for non-contiguous method with single and multiband operation.

1.7.1.5 Relays

Relay is implemented in the LTE-Advanced technology. The purpose of relaying is to provide coverage in new areas, cell-edge throughput, temporary network deployment, high data rate coverage and group mobility. Also, there are more advantages such as cost reduction, because it requires lower overhead costs than the eNB. The consumption of transmission power can also be reduced by using relay when the location of the relay is appropriate.

1.7.2 Network Architecture

In Release 8, 3GPP specified the rudiments and requirements of the EPS architecture that will serve as a basis for the next-generation networks. The disclaimers contain two major work items, namely LTE and SAE, which led to the specification of the Evolved Packet Core (EPC), Evolved Universal Terrestrial Radio Access Network (E-UTRAN), and Evolved Universal Terrestrial Radio Access (E-UTRA). Each of those corresponds to the core network, radio access network, and air interface of the whole system, respectively. The EPS provides IP connectivity between a UE and an external packet data network using E-UTRAN. Figure 1.12 provides an overview of the EPC, other legacy Packet and Circuit Switched elements and 3GPP RANs, along with the most important interfaces. In the services network, only the Policy and Charging Rules Function (PCRF) and the Home Subscriber Server (HSS) are included, for simplicity [4,28].

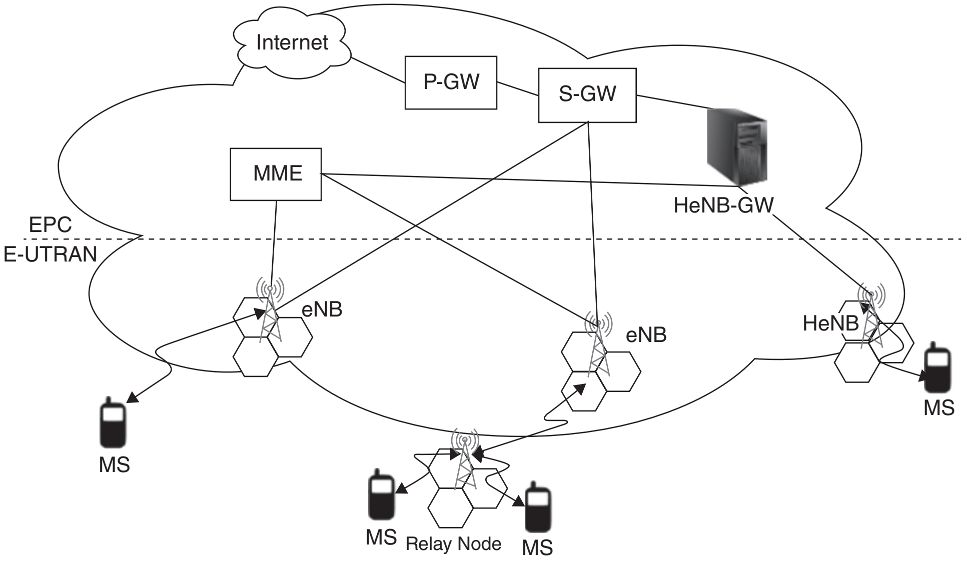

Figure 1.11 LTE-advanced E-UTRAN architecture.

In the framework of 4G systems, both the air interface and the radio access network are being improved or redefined, but so far the core network architecture, that is, the EPC, is not undergoing major changes from the previously standardized SAE architecture. Therefore, in this section, an overview of the E-UTRAN architecture and functionalities is given, which are defined for the LTE-Advanced systems and the main EPC node functionalities, shared by Releases 8, 9 and 10.

Enhanced Node B is the core part of the E-UTRAN architecture. It provides the air interface towards the UE. Each eNB is considered as the logical component that serves one or more several E-UTRAN cells. The target of this technology is to increase coverage, higher data rates, better QoS performance and fairness for the users. The EPC is a flat all-IP based core network. It can be accessed through 3GPP radio access, which allows the handover procedure. The Mobility Management Entity (MME), Serving Gateway (S-GW), and Packet Data Network Gateway (PDN-GW) work in a similar way to the LTE network architecture [26].

1.7.3 Beyond 3G and 4G Cellular Systems Security

Fourth generation cellular networks promise to provide higher user data rates, lower latency and a complete internet protocol (IP)-based network architecture. The major difference between the 3G and 4G cellular networks is that 4G operates entirely on IP protocol and architecture. For this reason, WiMAX is also considered as part of 4G networks. While discussing beyond 3G technologies, the major underlying technology in use is the LTE. Even though a similarity can be drawn between LTE and WiMAX, because of the IP-based protocol and architecture, they differ from each other in network architecture and security. The all IP-based infrastructure brings up increased security issues compared with the previous generation cellular technologies. For this reason, in 4G networks, extra security mechanisms are expected to be carried out to provide the security for reliable communication.

The major concern in 4G security naturally includes that the user who wants to access the network must be authenticated along with the device that will be connected to the network. For this reason, security credentials, identity, certificates, username and password, are used for authentication. If we draw a comparison, starting from 2G when security was started to be taken as a major concern in cellular networks, a unique ID is used on the SIM card in 2G. While in 3G and 4G LTE, temporary ID and further abstraction is used to limit the possibilities of any sort of intrusion. In 4G, further secure signaling between the UE and MME (Mobile Management Entity) is introduced, and also security measures are taken care of between 3GPP and trusted non-3GPP users. As mentioned before, because of operating of open IP-based architecture, security remains the pressing concern in 4G cellular, and is given strong emphasis. It is important to point out that LTE- and LTE-Advanced are the same technologies. The label "Advanced" was primarily added to highlight the relationship between LTE release 10 (LTE-Advanced) and ITU/IMT-Advanced. This does not make LTE-Advanced a different system from LTE [3,28].

1.7.4 LTE Security Model

Figure 1.12 shows the authentication method of LTE with step-by-step details. The authentication process in LTE is initiated by the authentication server when it sends Enhance Authentication Protocol request/Identity message (EAP) to the UE. The UE responds by replying to the EAP-response/Identity message containing the identity message and Network Access Identifier (NAI). Upon receipt of the EAP-response/identity message, the authentication server tried to access the UE's certificate from its record. The authentication server generates the EAP-Request/Authentication and Key Agreement (AKA)-Challenge message using the standard AKA process.

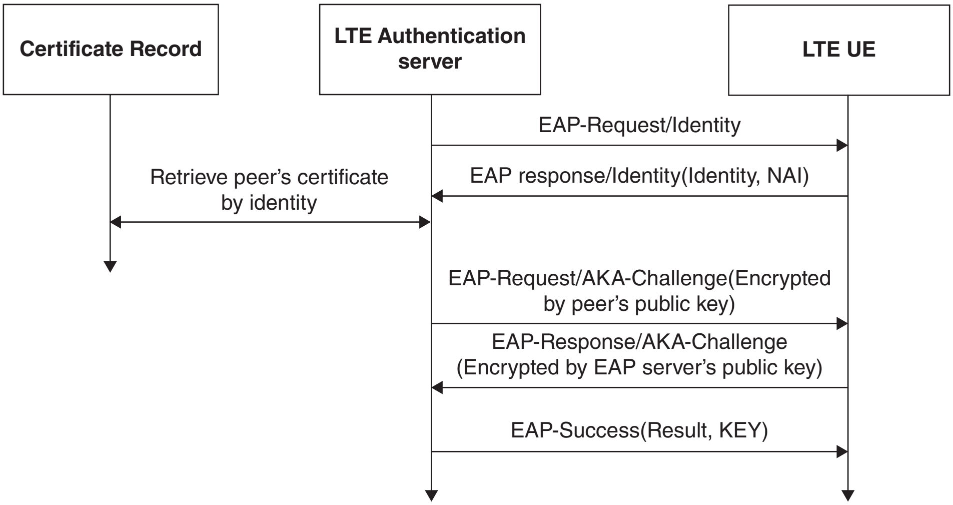
Certificate Record
LTE Authentication server
LTE UE
EAP-Request/Identity
Retrieve peer's certificate by identity
EAP response/Identity(Identity, NAI)
EAP-Request/AKA-Challenge(Encrypted by peer's public key)
EAP-Response/AKA-Challenge (Encrypted by EAP server's public key)
EAP-Success(Result, KEY)

Figure 1.12 Authentication Process in LTE.

The authentication server sends the EAP-Request/AKA-Challenge message encrypted by the UE's public key to the UE. The UE then decrypts the EAP-Request/AKA-Challenge message using its own private key, and then sends the EAP-Response/AKA-Challenge to the authentication server. The authentication server decrypts the information using the server's private key and verifies the EAP-response/AKA-Challenge message using the AKA algorithm. If the message is correct, the EAP server sends the EAP success message to the UE [18].

1.7.5 Security in WiMAX

The IEEE 802.11 security issues were merged by the WiMAX group into IEEE 802.16 standards. This was done because as the WiMAX standard evolved from 802.16 to 802.16a to 802.16e, the requirements evolved from the line-of-sight to mobile WiMAX. Hence, the requirements for security and corresponding standards also evolved to address the changing demands. In order for the security features of the initial IEEE 802.16 standard to work for the IEEE 802.16e standard, additional features are added [29].

The key new features are listed as follows:

1) Privacy Key Management Version2 (PKMv2) protocol;
2) User authentication is carried out using Extensible Authentication Protocol (EAP) method;
3) Message authentication is carried out using Hash-based Message Authentication Code (HMAC) or Cipher-based Message Authentication Code (CMAC) scheme; and
4) Confidentiality is achieved using Advance Encryption Standards (AES) [2].

When it comes to WiMAX, over-the-air security is a major concern ensuring end-to-end network security. While security architecture has been developed to mitigate against threats over the air, there still remain a number of challenges. The main challenge seems to be the balance of security needs with the cost of implementation, performance and interoperability. Since WiMAX uses IP transport mechanisms in handling control/signaling and management traffic, network operators will also have to defend against general IP security threats [30].

1.8 Conclusion

In this chapter, we presented the evolution of cellular systems. We focused on the development related to radio interface, network architecture and security measurements for different generation of cellular systems. The very first 1G system to the most recent 4G system were briefly discussed. The fifth generation (5G) cellular system is the next major phase of cellular communication, also referred to as wireless technologies beyond 2020. The 5G standard has to cope with the demand of a 1000-fold capacity and seamless connectivity for at least 100 billion devices. A stand-alone technology will not be able to cope up with such a demand.

A combination of spectral efficiency, spectrum enhancement and network efficiency, etc. will meet the challenges of 5G. Various efficient networking technologies, such as small cells, device-to-device (D2D), and software-defined networks (SDN) technologies will be adopted. As a part of the spectrum enhancement, the unlicensed bands will be used efficiently in addition to the licensed bands. Different technologies, such as massive MIMO and millimeter-wave MIMO, will play a key part for spectral efficiency in 5G networks and new radio interfaces have to be designed for these technologies.

The 5G system architecture and security models will be discussed in subsequent chapters. While this chapter provides basic information regarding the earlier cellular communications, we invite interested readers to go through the publications referenced in this chapter to gain a comprehensive knowledge of the evolution of cellular systems.

References

1 Al-Tawil, K. (King Fahd University of Petroleum and Minerals), Akrami, A. and Youssef, H. (1998) A new authentication protocol for GSM networks. *IEEE Conference on Local Computer Networks*, 11–14 October.

2 Andrews, J., Ghosh, A. and Muhamed, R. (2007) *Fundamentals of WiMAX*. Upper Saddle River, NJ: Prentice Hall.

3 Sankaran, C.B (2009) Network access security in next- generation 3GPP systems: A tutorial. *Communications Magazine, IEEE*, 47(2), 84–91.

4 Dahlman E. and Parkvall S. (2011) *4G: LTE or LTE-Advanced for Mobile Broadband*. West Sussex, UK: John Wiley & Sons, Ltd.

5 Johnston, D. and Walker, J. (2004) Overview of IEEE 802.16 security. *Security & Privacy, IEEE*, 2(3), 40–48.

6 Chin-Tser, H. and Chang, J.M. (2008) Responding to security issues in WiMAX networks. *IT Professional*, 10(5), 15–21.

7 Holma, H. *et al.* (2007) High-speed packet access evolution in 3GPP release 7. *IEEE Communications Magazine*, 45(12), 29–35.

8 Holma, H. and Toskala, A. (2002) High-speed downlink packet access. In: Chapter 11, *WCDMA for UMTS*. New York: John Wiley & Sons, Inc.

9 Hudderman A.A. (2003) *The Worldwide History of Telecommunications*. West Sussex, UK: John Wiley & Sons, Ltd.

10 IEEE Communications Magazine, Special issue on LTE–LTE *Part I: Core Network*, February 2009.

11 ITU Telecommunications indicators update 2016. www.itu.int/ITU-D/ict/statistics/

12 Josyula, R., Pankaj Rohatgi, R., Scherzer, H. and Tinguely, S. (2002) Partitioning attacks: or how to rapidly clone some GSM cards. *IEEE Symposium on Security and Privacy*, 12–15 May.

13 Korhonen, J. (2001) *Introduction to 3G Mobile Communications*. Norwood, MA: Artech House, Inc.

14 Lo, C.-C. and Chen, Y.-J. (1999) Secure Communication architecture for GSM networks. *IEEE Pacific RIM Conference on Communications, Computers and Signal Processing Proceedings*, 22–24 August.

15 Chang, M.J., Abichar, Z. and Chau-Yun, H. (2010) WiMAX or LTE: who will lead the broadband mobile internet? *IT Professional*, 12(3), 26–32.

16 Shin, M., Ma, J., Mishra, A. and Arbaugh, W.A. (2006) Wireless network security and interworking. *Proceedings of the IEEE*, 94(2), 455–466.

17 Marshall, P. (2008) HSPA+ challenges both WiMAX and LTE on the road to 4G. *Yankee Group Trend Analysis*, 19 September.

18 Seddigh, N., Nandy, B., Makkar, R. and Beaumont, J.F. (2010) Security advances and challenges in 4G wireless networks. In: *Eighth Annual International Conference on Privacy Security and Trust (PST)*, pp. 62–71.

19 Rengaraju, P., Chung-Horng, L., Yi, Q. and Srinivasan, A. (2009) Analysis on mobile WiMAX security. In: *Science and Technology for Humanity (TIC-STH), IEEE Toronto International Conference*, pp. 439–444.

20 Kasera, S. and Narang, N. (2005) *3G Mobile Networks: Architecture, Protocols and Procedures: Based on 3GPP specifications for UMTS WCDMA networks*. New York: McGraw-Hill.

21 Sauter. M. (2005) *From GSM to LTE: An Introduction to Mobile Networks and Mobile Broadbands*. West Sussex, UK: John Wiley & Sons, Ltd.

22 Smith, C. and Collins, D. (2002) *3G Wireless Networks*. Boston, MA: McGraw-Hill.

23 UMTS Forum. www.umts-forum.org

24 Leo, Y., Kai, M. and Liu, A. (2011) A comparative study of WiMAX and LTE as the next generation mobile enterprise network. *Proceedings of the 13th International Conference on Advanced Communication Technology (ICACT)*, pp. 654–658.

25 Muxiang, Z. and Yuguang, F. (2005) Security analysis and enhancements of 3GPP authentication and key agreement protocol. *Proceedings of the Wireless Communications, IEEE Transactions*, 4(2), 734–742.

26 3GPP, *Overview of 3GPP* release 8 v.0.1.1, Tech. Rep., June 2010.

27 3GPP TR 36.913, *Requirements for Further Advancements for E-UTRA*, v8.0.1, March 2009.

28 3rd Generation Partnership Program. *Network Architecture*. Technical Specification 23.002. Release 5. Version 5.5.0.

29 3rd Generation Partnership Program. *Security Architecture*. Technical Specification 33.102. Release 5. Version 5.2.0.

30 3rd Generation Partnership Program. *Cryptographic Algorithm Requirements*. Technical Specification 33.105. Release 4. Version 4.1.0.