

NEAR EAST UNIVERSITY



Faculty of Engineering

Department of Electrical and Electronic
Engineering

GSM Security

Graduation Project
EE- 400

Student: Mohammed Kamal Zorob (20010839)

supervtsor: Pr:of.. Dr.. Fakhradclin Mamedov

Nicosia - 2003



NEU



Dedicated to my :Mom el <.Daa

CONTENTS



DEDICATED	
AKNOWLEDEMENTS	
ABSTRACT	ii
CONTENTS	iii
INTRODUCTION	1
1. AN OVERVIEW OF CELLULAR SYSTEMS	2
1.1 Introduction	2
1.2 Concept of Cellular Communication	2
1.3 First-Generation Cellular System	3
1.4 Technologies for Second-Generation Cellular Systems	4
1.4.1 Narrow-Band Advanced Mobile Phones Services (N-AMPS)	5
1.4.2 Extended Time-Division Multiple Access (E-TDMA)	6
1.5 Cordless Phones and Telepoint Systems	6
1.6 Second-Generation Cellular Systems	7
1.6.1 CT-2	8
1.6.2 DCT 900/DECT/CT 3	9
1.6.3 Global System Mobile (GSM)	10
1.6.4 IS-54	11
1.6.5 IS-95	12
1.6.6 Japanese Digital Cellular (JDC)	12
1.7 Second-Generation-Plus PCs Systems	12
1.8 Vision of the Third-Generation Systems	15
2. CELLULAR COMMUNICATIONS	17
2.1 Overview	17
2.2 Mobile Communications Principles	17
2.2.1 Early Mobile Telephone System Architecture	18
2.3 Mobile Telephone System Using the Cellular Concept	19
2.4 Cellular System Architecture	21
2.4.1 Cells	21
2.4.2 Clusters	22

2.4.3	Frequency Reuse	22
2.4.4	Cell Splitting	23
2.4.5	Handoff	24
2.5	North American Analog Cellular Systems	26
2.5.1	The Advanced Mobile Phone Service (AMPS)	26
2.5.2	Narrowband Analog Mobile Phone Service (NAMPS)	27
2.6	Cellular System Components	27
2.6.1	PSTN	28
2.6.2	Mobile Telephone Switching Office (MTSO)	28
2.6.3	The Cell Site	28
2.6.4	Mobile Subscriber Units (MSUs)	29
2.7	Digital Systems	29
2.7.1	Time Division Multiple Access (TDMA)	31
2.7.2	Extended Time Division Multiple Access (E-TDMA)	32
2.7.3	Fixed Wireless Access (FWA)	32
2.7.4	Personal Communications Service (PCS)	33
2.7.5	Code Division Multiple Access (CDMA)	34
3.	THE GSM NETWORK	35
3.1	Overview	35
3.2	Services Provided by GSM	36
3.3	Architecture of the GSM Network	37
3.3.1	Mobile Station	38
3.3.2	Base Station Subsystem	39
3.3.3	Network Subsystem	39
3.4	Radio Link Aspects	40
3.4.1	Multiple Access and Channel Structure	41
3.4.1.1	Traffic Channels	41
3.4.1.2	Control Channels	42
3.4.1.3	Burst Structure	43
3.4.2	Speech Coding	44
3.4.3	Channel Coding and Modulation	44
3.4.4	Multipath Equalization	46

3.4.5 Frequency Hopping	46
3.4.6 Discontinuous Transmission	46
3.4.7 Discontinuous Reception	47
3.4.8 Power Control	47
4. GSM SECURITY	48
4.1 Overview	48
4.2 Origins of GSM Security	49
4.3 Security Services	49
4.4 Security Inside the GSM Network	50
4.4.1 Subscriber Identity Authentication	51
4.4.2 User and Signaling Data Confidentiality	52
4.4.3 Subscriber Identity Confidentiality	53
4.5 GSM Security Model	55
4.5.1 The Purpose of GSM Security	55
4.5.2 GSM's Security Limitations	56
4.5.3 Descriptions of the Functions of the Services	57
4.5.4 World-Wide Use of the Algorithms	60
4.6 Design Goals	62
4.7 Choosing the Security Architecture for GSM	64
4.8 The Architecture Chosen	66
4.9 Authentication Algorithm Design	69
4.10 Other GSM Security Mechanisms	71
4.10.1 SIM Card	71
4.10.2 IMEI	71
4.10.3 Roaming	72
4.11 Problems with GSM Security	72
4.11.1 The Limitation and Problems with GSM Security	72
4.11.2 Possible Improvement	73
5. CONCLUSION	74
6. REFERENCES	75

ACKNOWLEDGMENTS

I could not have prepared this project without the generous help of my supervisor, colleagues, friends, and family.

First, I would like to thank my supervisor Prof. Dr. Fakhreddin Mamedov for his invaluable advice, and belief in my work and myself over all the courses of this Degree. Prof. Dr. Fakhreddin Mamedov supplied the warmth, enthusiasm, and clarity of judgement that every student hopes for. Going beyond the limited role of literary agent, he provided valuable advice at each stage of the preparation of this project.

I would like to express my gratitude to Assoc. Prof. Dr. Adnan Khashman for him because he provided valuable advice at each stage of the preparation of this project also.

My thanks to Assist. Prof. Dr. Kadri Buruncuk for his help and making this project possible.

I will never forget the help that i got from this university for continueing my education especially from Prof. Dr Şenol Bektaş, so my regards and my love to him.

My depest thanks are to my family. I could never have prepared this project without the encouragement and support of my parents, brothers, and sister.

I would also like to thank all my friends for their help and for their patience also for their support, Mr Tayseer Alshanableh and Mr izzet Ağoren for their supporting.

ABSTRACT

The GSM standard was designed to be a secure mobile phone system with strong subscriber authentication and over-the-air transmission encryption. The security model and algorithms were developed in secrecy and were never published. Eventually some of the algorithms and specifications have leaked out. The algorithms have been studied since and critical errors have been found. Although the GSM standard was supposed to prevent phone cloning and over-the-air eavesdropping, both of these are possible with little additional work compared to the analog mobile phone systems and can be implemented through various attacks.

Nowadays the GSM service is sufficient to every single person in this whole wide world, a subscriber only needs a good mobile station and a subscriber identity given by local GSM network. Security of each subscriber, so his speech never be intercepted or no one would clone his identity and calls in his account, was one of the biggest problems security does face, and especially with that huge number of subscribers. That's what I tried to present in this project.

Subscriber Authentication in a GSM network first relies on a unique secret value (PIN) inserted into each SIM card. Once the mobile equipment has validated the user, the network then uses a "challenge-response" technique to authenticate the subscriber.

A 128 bit random number R is issued as a challenge to the mobile station, the MS encrypts the challenge using the authentication algorithm (A3), and a secret key k_i assigned to the mobile. The keys and algorithms are stored on the SIM card, and in an Authentication Centre.

The 32 bit signed response SRES is compared to the results of running the stored key through the same algorithm at the Authentication Centre. If the SRES is equal to the Authentications Centres result then the subscriber is authenticated. A mobile may be "challenged" at any time to ensure authentication.

INTRODUCTION

Securing information from unauthorized access is a major problem for any network wire line or wireless. Security, in a broad sense, focuses on network security, system security, information security, and physical security. The mobile telephone network is also subject to other unwanted security breaches, which cost hundreds of million dollars a year in the mobile phone business. Security services are needed to offer for mobile users without possibilities to violate the general trust.

The aim of this project to provide an idea of GSM security and how can design a security for GSM, and shows how we can choose the algorithm using for this purpose.

The project consists of the introduction and four chapters.

Chapters 1 present the concept of cellular communication and discuss the first- and second-generation cellular systems used in United States and Europe. We outline the problems associated with the second-generation-plus PCS system and provide the vision of a third-generation system.

Chapter 2 discusses the basics of radio telephony systems, including both analog and digital systems.

Chapter 3 shows an overview GSM including history, system, architecture, functions, radio interface.

Chapter 4 shows that GSM security met its design goals in a simple and elegant way, and has provided more than adequate security for most of its users, and problems with GSM security, possible improvement.

Finally conclusion section presents the obtained important results and gives summery about the idea of this project.

1. AN OVERVIEW OF CELLULAR SYSTEMS

1.1 Introduction

In this chapter we briefly present the concept of cellular communication and discuss the first- and second-generation cellular systems used in United States and Europe. We outline the problems associated with the second-generation-plus PCS system and provide the vision of a third-generation system.

1.2 Concept Of Cellular Communication

The idea of cellular communication is simple. During the late the bell system proposed to alleviate the problem of spectrum congestion by restructuring the coverage areas of mobile radio systems. The traditional approach to mobile radio involved setting up a high-power transmitter located on the top of the highest point in the coverage area. The mobile telephone needed to have a line of sight to the base station for adequate radio coverage. Line of sight transmissions limited to the distance to horizon (as much as 40 or 50 miles away for a high base station antenna). The result adequate coverage over a large area, it also implied that a few available radio channels were locked up over a large area by small number of users. In 1970 the bell system in New York City could support just 12 simultaneous mobile conversations. The 13th caller was blocked. The cellular concept handles the coverage problems differently. It does not use the broadcasting method; it uses a large number of low-power transmitters designed to serve only a small area. Thus, insted of an area like New York City being covered by a single transmitter, the city was divided into the smaller coverage areas called [cells, By reducing the total coverage of area into smaller cells, it with small cells was that not all mobile calls would now be completed within a single cell. To deal with this problem, the idea ofhandoff was used.

It is enormously expensive to build a system with thousands of cells right from the beginning.however; larage-radius cells can evolve gracefully into small-radius cells over a period of time using cell-splitting. When the traffic reaches a point in a particular cell such that the existing allocations of channels in that cell can no large support a good

grade of service, that cell is subdivided into smaller cells with lower transmitter power to fit within the area of the former cell.

Thus the essential elements of a cellular system are:

1. Low-power transmitter and small coverage areas or cells
2. Frequency reuse
3. Handoff and central control
4. Cell splitting to increase call capacity

1.3 First-Generation Cellular System

As the United States was planning its cellular network in the 1970s, England, Japan, Germany, and Scandinavian countries were also planning their systems. Each system used a different frequency band and different protocols for signaling between mobile units and its base stations. They all used analog FM (with during the 1970s the FCC forced the TV broadcasters off the little-used UHF channels 70-84 and made the frequencies available for two-way radio and the new cellular technology. During that time, the Bell system and Motorola actively pursued support for a 900-MHz cellular system using different designs for channel reuse and protocols.

In the late 1970s, the FCC mandated that a single nationwide standard must be developed before licenses for cellular systems would be awarded. The Electronics Industry Association (EIA) formed a cellular standards committee and standardized the Advanced Mobile Phone System (AMPS) protocol for the United States.

In 1985 the total access communication system (TACS) was introduced in the United Kingdom. TACS is a close relative of North America's Nordic Mobile telephone (NMT). The cellular approach promised virtually capacity through cell splitting. As the popularity of wireless communications escalated in the 1980, the cellular industry faced practical limitations. For a fixed allocation of spectrum, a large increase in capacity implies corresponding reduction in cell size. For example, the U.S. AMPS design allows for cells as small as 1,600 meters (m) (1 mile). As the cells get smaller, it becomes increasingly difficult to place base station at the locations that offer necessary radio coverage. Also, reduction cell size demand increased signaling activity as more rapid handoffs occur; in addition, base station as required to handle more access requests and

registrations from the mobile stations. The problem becomes particularly difficult in large urban areas where capacity requirements are most pressing. In addition to the capacity bottleneck, the utility of the first generation analog systems was diminished by proliferation in of incompatible standards in Europe. The same mobile Telephone frequencies cannot be used in different European countries. The limitations of first-generation analog systems provided motivations to the second- generation systems. The principal goals of the second-generation systems were: higher capacity and hence lower cost, and a continental system with full international roaming and handoff capabilities. In Europe, these goals are served by new spectrum allocations and by the formulation of a Pan-European cellular Standard GSM.

In North America, where one standard (United States, Canada, and Mexico) existed and covered a region as large as Europe, the push for a new system was not as strong. In the new digital systems, higher capacity is derived from applications of advanced transmission techniques including efficient speech coding. Error correcting channels codes, and band with-efficient modulation techniques. In Europe, the approach was to open new- frequency bands for a pan-European system and not to have compatibility with existing cellular systems. In the United States, the same frequency bands, were shared with new digital systems, and the standards supported dual-mode telephones that could be used in both analog and digital systems.

1.4 Technologies for Second-Generation Cellular Systems

Standards and system designs exist for several new and competing technologies for the second-generation cellular systems. They are:

1. Narrow-band Advanced Mobile Phone Service (N-AMPS)
2. Time-Division Multiple Access (TDMA) (already explained in chapter2 section 2.9.2)
3. Extended Time-Division Multiple Access (E-TDMA)
4. Spread Spectrum (SS)-Code-Division Multiple Access (CDMA)

NOT all systems have seen widespread use, and some may disappear. The following sections provide a brief description and status of these technologies

1.4.1 Narrow-Band Advanced Mobile Phones Service (N-AMPS)

Motorola developed N-AMPS by dividing an analog channel into three parts; thereby tripling the present analog channel capacity. Bandwidth per user is decreased from 30 kHz to 10 kHz. Each new channel is capable of handling its own calls. N-AMPS act primarily as a bridge to digital communications that allows cellular systems to increase capacity at a low cost.

A smaller bandwidth per user in N-AMPS results in a slight degradation in speech quality that is compensated for with the addition of an interference avoidance scheme called Mobile Reported Interference (MRI). This capability, along with full call control (e.g., conference call, call waiting, call transfer, handoff, and power control) is provided using a (new) continuous 100 bits per second (bps) in-band, sub-audible, signaling control channel. This scheme has the additional benefit of eliminating the audio gaps typical in the AMPS blank and burst-signaling scheme. The associated control channel can also be used for sending alphanumeric characters when not actively managing call control. It has been typically used for features such as displaying calling line identification numbers, as well as features similar to those provided by an alphanumeric pager. This capability has allowed for the combining of cellular and paging applications in a single device. N-AMPS was standardized as EIA/TIA, IS-88, IS-89, and IS-90 in late 1992. In 1993, IS-88 and IS-553 (AMPS) were combined to form a single analog standard called IS-91. N-AMPS has been implemented in both U.S. and international markets.

Although there has been some concern about the ultimate capacity of N-AMPS, many operators have chosen to implement N-AMPS in the identical reuse pattern as the original AMPS design. With trunking efficiencies excluded, this results in three times AMPS capacity. In a typical international market, up to 90% N-AMPS penetration has been achieved using a four-cell reuse pattern with very small (500 m) cell radii. It has been claimed by those operators that audio quality is at least as good as the AMPS systems they replaced. In addition, they report that overall dropped call performance is considerably less than AMPS due to the signaling enhancements made to the N-AMPS air interface.

1.4.2 Extended Time-Division Multiple Access (E-TDMA)

General Motors' (GM's) effort to enter the cellular market uses E-TDMA. E-TDMA uses half rate voice coding at 4.5 KB that requires only one IS-54 time slot and thus allows six calls per frequency. Digital Speech Interpolation (DSI) permits deleting silence on calls. It thus reduces activity by 55-65% and allows more calls to be handled by the same number of time slots. E-TDMA is claimed to increase capacity by 12 times that of AMPS and 4 times that of IS-54. GM plans to include E-TDMA mobile phones in a future automobile model. E-TDMA supporters must solve many technical problems.

Voice quality being the most important one, Furthermore, none of the cellular carriers seem to be interested in deploying it. E-TDMA does not appear to be a serious contender in the digital technology.

1.5 Cordless Phones and Telepoint Systems

Closely related to cellular and often confused with the mobility aspects of cellular and PCS are cordless phones and Telepoint systems. Cordless phones are the low-power, low-range phones that enable an individual to move around a house or apartment and still place and receive phone calls. First-generation cordless telephones are stand-alone consumer products they do not require any interoperability specifications at all. Each cordless telephone comes with its own base station and needs to be compatible only with that base station. The billing, security, and privacy are achieved (to a limited degree) by preventing the phone from operating with any other base station. Because of the popularity of cordless phones and the inability of some telephone companies to maintain public telephones in large cities, a hybrid approach was conceived, called "Telepoint".

In the Telepoint system, the user owns a small low-power phone (similar in size and functionality to a cordless phone). The Telepoint phone works within 100 m of a public base station. The phone typically cannot receive calls and can place calls only when in range of the base station. The user would walk or drive within range of the base station and place the call. Roaming is not supported; the user must remain within range of the base station or the call is dropped. This concept, based on a limited frequency

allocation, was implemented as the Cordless Telephone-I (CT- 1) 900-MHz analog system. It was implemented or proposed in 13 European and Scandinavian countries. Different incarnations of the design can serve residential (wireless local loop) or public pay phone markets. CT-1 was intended to serve the residential market. An enhanced version; CT -1+, is similar to CT- 1, but has added Telepoint capabilities. CT-I uses FDMA, in which a single channel per radio carrier frequency is employed. CT-1 carries multiple narrowband carriers within a frequency band. Duplex operation, i.e., the simultaneous transmission and reception of voice signals, is implemented using separate frequencies.

Once the concept of a small low-power phone was introduced, designs were evolved to support wireless PBXs, cellular phones, PCS, and neighborhood wireless local loops. Often the system designs for second-generation cordless phones and Telepoint systems and the designs for digital cellular systems overlap.

The culmination of research in digital technologies resulted in CT- 2, CT- 3/ DCT-900 and DECT standards for cordless telephones/Telepoint in Europe. These systems are being offered for a variety of uses-cellular, PCS, cordless phones, Telepoint, and wireless PBXs.

Cordless phones in the United States have either used FM in the 46/49-MHz band or SS in the 902-928-MHz band. Telepoint systems have not seen widespread use in the United States.

1.6 Second-Generation Cellular Systems

First-generation cellular systems were designed to satisfy the needs of business customers and some residential customers. With the increased demand of cellular telephones in Europe, several manufacturers began to look for new technologies that could overcome the problems of poor signals and battery performance. Poor signals resulted in poor performance for the user and a high frequency of false handoffs for the system operator. Better battery performance was needed to reduce size and cost of self contained handheld units (handsets). Research efforts were directed toward wireless technologies to provide high-quality, interference-free speech and decent battery performance. The size of handset and better battery performance led to low-power

designs and performance targets possible only with fully digital technologies. Digital cellular systems based on the GSM (TDMA) standard have emerged in Europe, while systems based on IS-54 (TDMA) and IS-95 (CDMA) are being developed in the United States.

The following sections describe the CT-2, DCT 900/DECT/CT-3, GSM, IS-54, and IS-95 systems and point out the main differences between them.

1.6.1 CT-2

The handsets used in offices and homes these systems were developed for residential, business, and Telepoint applications in the United Kingdom. The handsets used in offices and homes were provided with a "value-added" public service from base stations located in railway stations, airports, and shopping centers. Although business and residential use of CT-2 offered full incoming and outgoing call facilities, Telepoint service was limited to outgoing calls only. The United Kingdom chose FDMA for CT-2 to meet the original goal of a simple, single-user, home mobile telephone that avoids interference at call setup and supports multichannel multiplexing or handoffs. FDMA/TDD meets the needs for simple single-user channelization and simple measurement of signal power for a frequency channel from both ends of a radio link.

With the introduction of Telepoint and Wireless Private Branch Exchange (WPBX) applications in the United Kingdom, there was a need for a handset user to roam between different Telepoint operators' base stations and WPBX products. Therefore, the message protocols across the air interface needed to be well defined and common to all users. This resulted in the Common Air Interface (CAI) concept for CT-2.

CT-2 uses 4 MHz of spectrum, from 864 to 868 MHz, divided into 40-100 kHz channels. On each channel, the base station and mobile station alternate in the transmission of TDD data packets. The TDD rate, the rate at which base station and mobile stations switch between "send" and "receive" is set at 500 Hz-1 ms each for send and receive. Power from both the mobile and base stations is restricted to 10 mW, and the mobile station selects the channel with the lowest noise. The mobile station is

frequency agile during a call if the bit error rates on the selected channel reach unacceptable levels. Within this burst structure, there is a data rate of 72 kbs. In each burst, 72 bits of data are available for speech, control, signaling, and base/mobile station synchronization purposes. There is an allowance for a guard time between bursts to allow the sender to turn off its transmitter and settle into receiver mode and for the receiver to turn on its transmitter and settle at its center frequency. The guard time is nominally 4 bits long. By use of the guard period, both ends of the links are sure that the receiver is able to decode accurately the first- and later-transmitted bits in the burst.

The CT-2 modulation technique is binary frequency shift keying. With a channel spacing of 100 kHz, the bandwidth efficiency of CT-2 is 0.72 bps/Hz about half of that of GSM. The speech coder is a standard Adaptive Differential Pulse Code Modulator (ADPCM) operating at 32 kbs.

1.6.2 DCT 900/DECT/CT 3

The application of wireless technology, particularly for a large business complex with WPBXs that support roaming and handoffs between different cells is demanding. The system should manage the traffic in the cells in real time as the handsets move throughout the complex. This dramatically increases the complexity of the call processing software over that of a standard PBX. The software must also account for the three-dimensional environment of the system with the overlap of radio waves through different floors. Furthermore, the building environment affects the propagation of radio waves with the reflection and absorption of radio energy dependent on the construction materials. Ideally, a large building should be designed with a WPBX in mind; in practice, real buildings will have been designed before the WPBX, and compromises will be needed.

The capacity needs of a large modern office building can be met only with a high frequency reuse achieved by use of Pico-cells with an indoor cell size of less than 50 m. Low-output power enables the handsets to be small and provides a talk time that exceeds the possibilities of other technologies. The most important requirement of the business PBX user is that the voice quality of the call is comparable to that of existing wired extensions.

The DCT 900/DECT/CT-3 choice of TDMA/TDD was dictated by the needs of multiple mobile telephones accessing multiple base units and connected to a PBX and by the shortage of paired frequency bands in Europe. The solution to this problem required the multiplexing of multiple users at a base unit and support for handoff and was readily implemented in a single-frequency TDMA/TDD. With only one frequency, TDD permitted simple rapid monitoring of power in all channels from both ends of a radio link.

Dynamic time slot allocation algorithms for Dynamic Channel Allocation (DCA), with continuous transmission in at least one time slot as a "beacon" from all base units, provided a convenient mechanism for initial base-unit and time-slot selection.

The emerging standard for the large WPBX is DECT. This standard was frozen by ETSI in 1991, and the first system appeared in 1993. DECT standards do not compete directly with CT 2 because they are not oriented on the same market. DECT standards are more dedicated to PBX with large capacity, whereas CT- 2 fulfills the requirements of the PBX with small capacities.

The modulation technique of DECT is Gaussian Minimum Shift Keying. The relative bandwidth of the Gaussian filter is wider (0.5 times of the bit rate) than in GSM. The bandwidth efficiency is 0.67-bps/Hz, which is comparable to that of CT-2. The speech coder of DECT is ADPCM with bit rate of 32 Kbs.

The DECT standard enables the development of systems specially designed to handle high capacity in a stationary environment. DECT cannot compete with cellular technology for use in vehicles, but it will be considerably cheaper in the applications it has been designed for. DECT's TDMA broadband solution may more adequately cover the businessperson's demands, such as high voice quality and data transmission capability

1.6.3 Global System Mobile (GSM)

GSM was driven by the need for a common mobile standard throughout Europe and the desire for digital transmission compatible with data and privacy. Spectrum was reallocated near 900 MHz throughout much of Europe and the surrounding region so that

completely new technology could be developed by GSM. The GSM effort in the early to mid-1980s considered several system implementations including TDMA, CDMA, and FDMA technologies. A TDMNFDMN FDD technology was chosen with a radio link bit rate of 270 kbs.

The GSM modulation is Gaussian Minimum Shift Keying (GMSK). The bandwidth efficiency of 270-kbs signals operating with 200-kHz carrier spacing is 1.35 bps/Hz. The GSM's speech coder is referred to linear predictive coding with regular pulse excitation. The source rate is 13 kbs and transmission rate, including error detecting and correcting codes; is 22.8 kbs.

1.6.4 IS-54

In North America, where a common analog air interface was available and roaming anywhere in Canada, the United States or Mexico was possible; there was no need to replace the existing analog systems. Therefore, the Cellular Telecommunication Industry Association (CTIA) requested the TIA to specify a system that could be retrofitted into the existing AMPS system. The high cost of the cell sites was the major driving force. Thus, the important factor in the IS-54 was to maximize the number of voice channels that can be supported by a cell site within the available cellular spectrum. Several TDMNFDMA and pure FDMA system proposals were considered before the IS-54 standard was selected. IS-54 fits three TDMA 8-kbs encoded speech channels into each 30 kHz AMPS channel.

IS-54 uses a linear modulation technique, Differential Quadrature Phase Shift Keying (DQPSK) to provide a better bandwidth efficiency: The transmission rate is 48.6 kbs with a channel spacing of 30 kHz. This gives bandwidth efficiency of 1.62 bps/Hz, a *20% improvement* over GSM. The main penalty of linear modulation is power efficiency that affects the weight of handsets and time between battery charging. The IS-54 speech coder is a type of code book excited linear predictive coding referred to as Vector Sum Excited Linear Prediction. The source rate is 1.95 kbs and the transmission rate is 13 kbs.

1.6.5 IS-95

Recently a CDMA protocol has been proposed by QUALCOMM and standardized in the United States as IS-95. IS-95 is aimed at the dual-mode operation with the existing analog cellular system. The basic idea behind increased capacity in IS-95 is the use of a wideband channel the proposed channel width is 1.25 MHz in each cell or 42-30 kHz channels where many subscribers can talk together without interfering with each other. Many users with different codes share each channel. IS-95 proposes soft. Handoff to improve voice quality and RAKE receiver to take advantage of multipath fading and to lower signal-to- interference (S_{il}) ratio. Other factors that affect the channel capacity include use of variable rate vocoder, voice activity factor, a power control in the forward and reverse channel.

1.6.6 Japanese Digital Cellular (JDC)

Japanese Digital Cellular standards are aimed to replace the three incompatible analog cellular systems in Japan. The basic radio channel design defined in the JDC standard is comparable with the North American IS-54 TDMA,) digital and European GSM system. The JDC systems use three-channel TDMA. Two required Frequency bands have been reserved: 800-MHz band with 130 MHz of duplex separation and 1.5-GHz band with 48 MHz of duplex separation. The 800-MHz band will be used first, whereas the 1.5-GHz band is for future use. The modulation scheme is $\pi/4$ -QPSK with interleaved carrier spacing of 25 kHz. The speech coder uses 11.2-kbs VSELP including channel coding.

1.7 Second-Generation-Plus PCS Systems

Although many people describe PCS as a third-generation system, the U.S. implementation uses modified cellular protocols. The opening of the 2-GHz band by the FCC has generated a flurry of activity to develop new systems. Unfortunately, in the race to deploy systems, most work has been to up band the existing cellular systems to the new 2-GHz band: Whether the protocol is GSM, IS-54, or IS-95, each proponent wants to make minimum changes in its protocol to win the PCS race.

It may not be until later in the 1990s before true third-generation systems offering Wireless multimedia access emerges. The initial offering may be tailored to the environment and the need for rapid entry into the marketplace.

A further factor is the need to support wireless residential service (wireless CENTEREX), cordless phones, Telepoint, wireless PBXs, low-mobility (on-street) portable phones, and high-mobility (in-vehicle) mobile phones. Although there is a desire for one protocol to support all needs, cost constraints may result in several solutions. Each one optimized to a particular need. This need and demand for wireless communications in several environments has been shown by the rapid growth of different technologies that are optimized for particular applications and environments. Examples are:

1. Residential mobile telephones and their evolution to digital technology in CT-2 and to DCT 900/DECT/CT 3 for in-building PBX environments
2. Analog; cellular telephones for widespread mobile service and their digital evolutions to GSM, E-TDMA, IS-54, and CDMA
3. Wireless data networks both for low-rate wide-area coverage and higher- rate wide Local Area networks (WLANs)

Basic needs for PCS include standardized low-power technology to provide voice and moderate-rate data to small, lightweight, economical, pocket-size personal handsets that can be used for tens of hours without attention to batteries and to be able to provide such communications economically over wide areas, including in homes and other buildings, outdoors for pedestrians in neighborhoods and urban areas, and anywhere there are reasonable densities of people.

The CT- 2 and DCT 900/DECT/CT- 3 technologies look attractive for providing low-priced personal communications services with volume penetration. To permit widespread use of these technologies in outside environments where base stations have less attenuation between themselves than between mobile stations and base stations, time synchronization of base station transmissions is to achieve good performance with TDD. While the DCT 900DECT/CT-3 technology was appropriate for WPBXs, it needs modification for more widespread PCS applications, for which it also incurs synchronization requirements and additional complexity .

The DCS1800 is a standard for PCN that has been developed by ETSI. It is a derivative of the GSM 900 MHz cellular standard. In Europe DCS has been allocated frequencies from 1710 to 1785 MHz and 1805 to 1880 Hz to provide a maximum theoretical capacity of 375 radio carriers, each with 8 or 16 (half rate) voice/data channels. In DCS1800 there are provisions for national roaming between operators with overlapping coverage.

These modifications have enabled the GSM cellular standard to be enhanced to provide a high capacity, quality PCN system that can be optimized for handheld operation. The 1800-112Hz operating band in the DCS results in a small cell structure that is compatible with the PCN concept. The 1800-MHz band is occupied by fixed radio links for which alternative technologies exist, and clearance of the band can be more readily effected than attempting to manage coexistence and transition between the first- and second-generation cellular systems at 800/900 MHz.

The initial implementation of European PCN is based on the provision of a high-quality small cell network (cell radius Less than 1 km in a dense urban environment to 5 km in the rural environment). Radio cover age and system parameters are optimized-for low-power handsets, and emphasis is placed on providing a significantly higher statistical call success and quality level for the handheld portable than current cellular networks provide.

The future evolution of DCS1800 may include microcell structure for cover- age and capacity enhancement into buildings such as airport terminals, railway stations, and shopping centers, where. Large numbers of people gather. A further development would then be in "private" cells within offices to provide business communications. Ubiquitous deployment of microcells in a PCN environment will require a very fast handoff processing capability that is not currently avail- able on DCS1800. How successfully the DCS1800 technology can be implemented in office environments to replace DCT 900/DECT/CT- 3 are in question.

1.8 Vision of the Third-Generation Systems

First-generation analog and second-generation digital systems are designed to support voice communication with limited data communication capabilities. Third-generation systems are targeted to offer a wide variety of services listed in Table 1.1. Most of the services are wireless extensions of Integrated Services Digital Network (ISDN), whereas services such as navigation and location information are mobile specific. Wireless network users will expect a quality of service similar to that provided by the wire line networks such as ISDN. Service providers will require higher-complexity protocols in the physical link layer because of the unpredictable nature of the radio propagation environment and the inherent terminal mobility in a wireless network. These protocols will use powerful forward error correction and digital speech interpolation techniques to match the quality of service of the fixed network.

Because of the multitude of teleservices offered in different operating scenarios, the teletraffic density generated will depend on the environment, the mix of terminal types, and the terminal density. Teletraffic density will vary substantially for high-bit-rate services provided in business areas, whereas basic services such as speech and video telephony will be offered in all other environments.

The third-generation network will concentrate on the service quality, system capacity, and personal and terminal mobility issues. The system capacity will be improved by using smaller cells and the reuse of frequency channels in a geographically ordered fashion. A third-generation network will use different cell structures according to the operational environment. Cell structures will range from conventional macro-cells to indoor pico-cells. In particular, micro-cells with low transmission power will be widely deployed in urban areas, while other cell structures will be used according to the environment to provide ubiquitous coverage. It is expected that the cost of base station equipment for micro-cells will be significantly reduced because of the elimination of costly high-power amplifiers and the economies of scale in micro-cell base station manufacturing. Nevertheless, the system's cost will still play a dominating role in the design of the network infrastructure because more micro-cellular base stations will be required to provide adequate radio coverage. Micro-cells with a radius less than 1,000 m will be used extensively to provide coverage in metropolitan areas. Micro-cell base

Stations will be mounted on lampposts or on buildings where electric supply is readily available. For high-user-density area such as airport terminals, railway stations, and shopping malls, pico-cells with coverage of tens of meters will be used. To facilitate efficient handoff when the vehicle-based user

Cells at high speed, these calls will be handled by umbrella cells (overlay macro-cells) whose coverage areas will contain several to tens of micro-cells.

The planning of third-generation systems will be more complicated than the design of present speech-oriented, micro-cell-based mobile systems and will require a more advanced and intelligent network-planning tool.

Table1.1 Proposed Teleservices A Third Generation System

Teleservices	Throughput (kbs)	Target bit error
Telephony	8 - 32	10^{-7}
Voice mail	32	10^{-7}
Program sound	32	10^{-7} ,
Video telephony	128	10^{-7}
Video conference	64	10^{-7}
Remote terminal	384 - 768	10^{-6}
User profile editing	1.2 - 9.6	10^{-6}
Telefax(group4)	1.2 - 9.6	10^{-6}
Voiceband data	64	10^{-6}
Database access	2.4 - 768.Q.	10^{-6}
Message broadcast	2.4	10^{-6}
Unrestricted digital information	64 - 1,920	10^{-6}
Navigation	2.4 - 64.0	10^{-6}
Location	2.4 - 64.0	10^{-6}

2. CELLULAR COMMUNICATIONS

2.1 Overview

A cellular mobile communications system uses a large number of low-power wireless transmitters to create cells, the basic geographic service area of a wireless communications system. Variable power levels allow cells to be sized according to the subscriber density and demand within a particular region. As mobile users travel from cell to cell, their conversations are handed off between cells to maintain seamless service. Channels (frequencies) used in one cell can be reused in another cell some distance away. Cells can be added to accommodate growth, creating new cells in un-served areas or overlaying cells in existing areas.

This chapter discusses the basics of radio telephony systems, including both analog and digital systems. Upon completion of this chapter, you should be able to describe the basic components of a cellular system and identify digital wireless technologies.

2.2 Mobile Communications Principles

Each mobile uses a separate, temporary radio channel to talk to the cell site. The cell site talks to many mobiles at once, using one channel per mobile. Channels use a pair of frequencies for communication--one frequency (the forward link) for transmitting from the cell site and one frequency (the reverse link) for the cell site to receive calls from the users. Radio energy dissipates over distance, so mobiles must stay near the base station to maintain communications. The basic structure of mobile networks includes telephone systems and radio services. Where mobile radio service operates in a closed network and has no access to the telephone system, mobile telephone service allows interconnection to the telephone network.

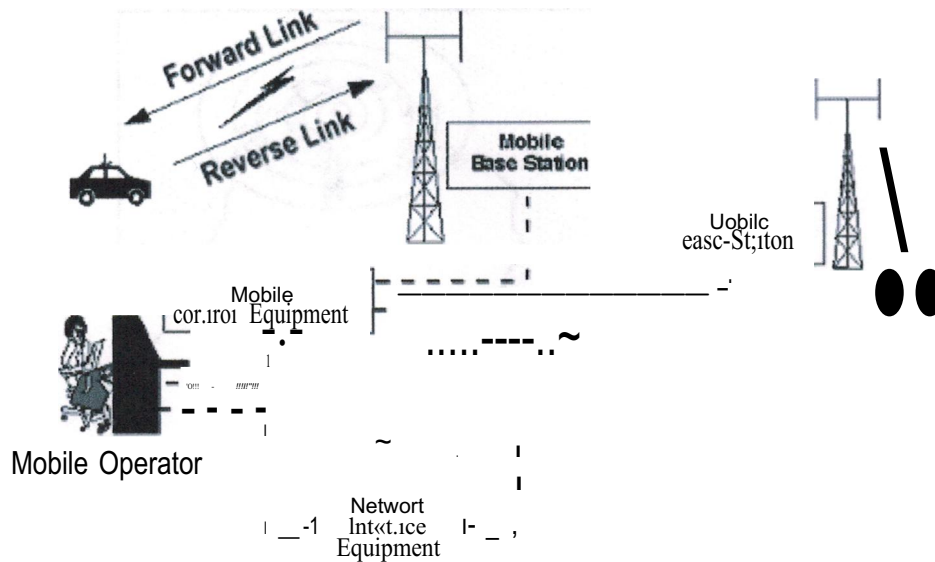


Figure 2.1 Basic Mobile Telephone Service Network

2.2.1 Early Mobile Telephone System Architecture

Traditional mobile service was structured in a fashion similar to television broadcasting: One very powerful transmitter located at the highest spot in an area would broadcast in a radius of up to 50 kilometers. The cellular concept structured the mobile telephone network in a different way. Instead of using one powerful transmitter, many low-power transmitters were placed throughout a coverage area. For example, by dividing a metropolitan region into one hundred different areas (cells) with low-power transmitters using 12 conversations (channels) each, the system capacity theoretically could be increased from 12 conversations or voice channels using one powerful transmitter to 1,200 conversations (channels) using one hundred low-power transmitters. Shows a metropolitan area configured as a traditional mobile telephone network with one high-power transmitter.

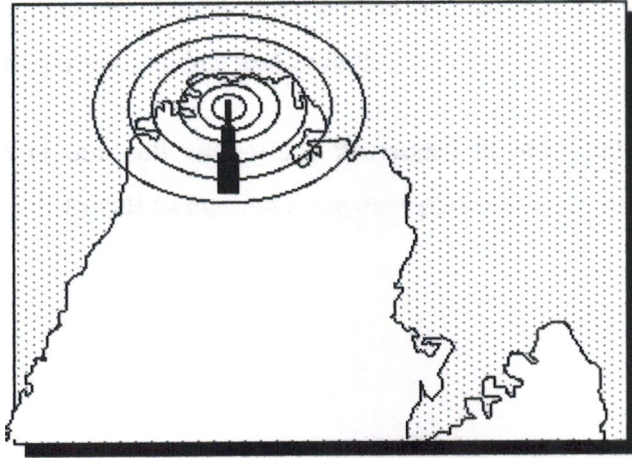


Figure 2.2. Early Mobile Telephone System Architecture

2.3 Mobile Telephone System Using the Cellular Concept

Interference problems caused by mobile units using the same channel in adjacent areas proved that all channels could not be reused in every cell. Areas had to be skipped before the same channel could be reused. Even though this affected the efficiency of the original concept, frequency reuse was still a viable solution to the problems of mobile telephony systems.

Engineers discovered that the interference effects were not due to the distance between areas, but to the ratio of the distance between areas to the transmitter power (radius) of the areas. By reducing the radius of an area by 50 percent, service providers could increase the number of potential customers in an area fourfold. Systems based on areas with a one-kilometer radius would have one hundred times more channels than systems with areas 10 kilometers in radius. Speculation led to the conclusion that by reducing the radius of areas to a few hundred meters, millions of calls could be served.

The cellular concept employs variable low-power levels, which allow cells to be sized according to the subscriber density and demand of a given area. As the population grows, cells can be added to accommodate that growth.

Frequencies used in one cell cluster can be reused in other cells. Conversations can be handed off from cell to cell to maintain constant phone service as the user moves between cells.

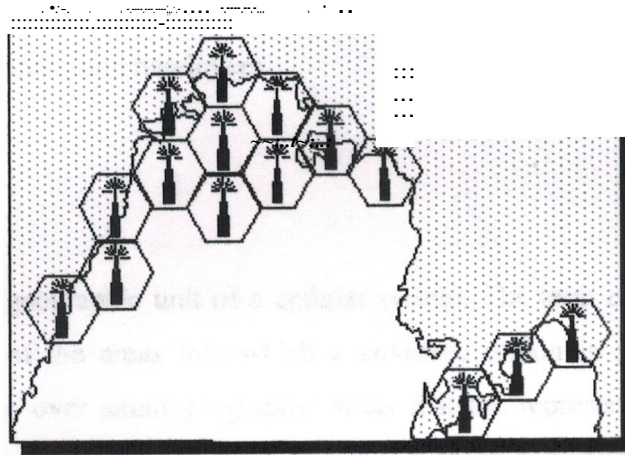


Figure 2.3. Mobile Telephone System Using a Cellular Architecture

The cellular radio equipment (base station) can communicate with mobiles as long as they are within range. Radio energy dissipates over distance, so the mobiles must be within the operating range of the base station. Like the early mobile radio system, the base station communicates with mobiles via a channel. The channel is made of two frequencies, one for transmitting to the base station and one to receive information from the base station.

2.4 Cellular System Architecture

Increases in demand and the poor quality of existing service led mobile service providers to research ways to improve the quality of service and to support more users in their systems. Because the amount of frequency spectrum available for mobile cellular use was limited, efficient use of the required frequencies was needed for mobile cellular coverage.

In modern cellular telephony, rural and urban regions are divided into areas according to specific provisioning guidelines. Deployment parameters, such as amount of cell-splitting and cell sizes, are determined by engineers experienced in cellular system architecture.

Provisioning for each region is planned according to an engineering plan that includes cells, clusters, frequency reuse, and handovers.

2.4.1 Cells

A cell is the basic geographic unit of a cellular system. The term *cellular* comes from the honeycomb shape of the areas into which a coverage region is divided. Cells are base stations transmitting over small geographic areas that are represented as hexagons. Each cell size varies depending on the landscape. Because of constraints imposed by natural terrain and man-made structures, the true shape of cells is not a perfect hexagon.

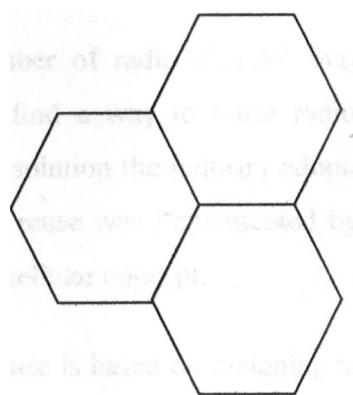


Figure 2.4. Cells

2.4.2 Clusters

A cluster is a group of cells. No channels are reused within a cluster. Figure 2.5 illustrates a seven-cell cluster.

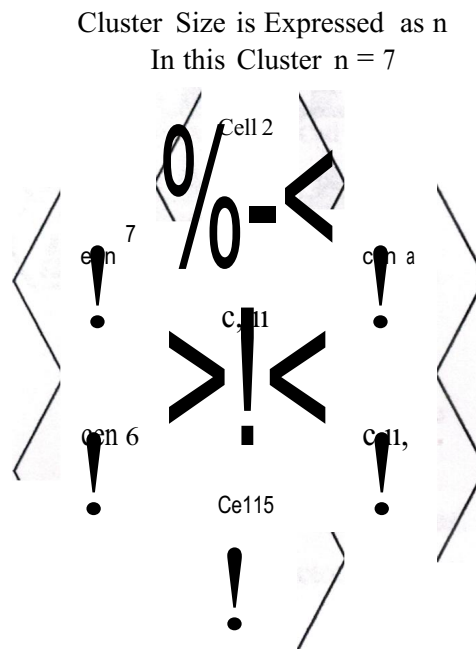


Figure 2.5. A Seven-Cell Cluster

2.4.3 Frequency Reuse

Because only a small number of radio channel frequencies were available for mobile systems, engineers had to find a way to reuse radio channels to carry more than one conversation at a time. The solution the industry adopted was called frequency planning or frequency reuse. Frequency reuse was implemented by restructuring the mobile telephone system architecture into the cellular concept.

The concept of frequency reuse is based on assigning to each cell a group of radio channels used within a small geographic area. Cells are assigned a group of channels that is completely different from neighboring cells. The coverage area of cells is called the footprint. This footprint is limited by a boundary so that the same group of channels can be

used in different cells that are far enough away from each other so that their frequencies do not interfere.

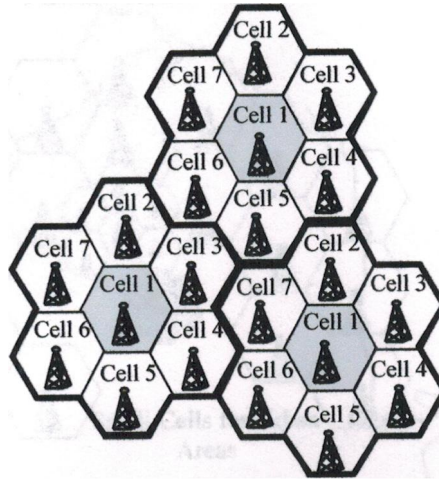


Figure 2.6. Frequency Reuse

Cells with the same number have the same set of frequencies. Here, because the number of available frequencies is 7, the frequency reuse factor is $1/7$. That is, each cell is using $1/7$ of available cellular channels.

2.4.4 Cell Splitting

Unfortunately, economic considerations made the concept of creating full systems with many small areas impractical. To overcome this difficulty, system operators developed the idea of cell splitting. As a service area becomes full of users, this approach is used to split a single area into smaller ones. In this way, urban centers can be split into as many areas as necessary to provide acceptable service levels in heavy-traffic regions, while larger, less expensive cells can be used to cover remote rural regions.

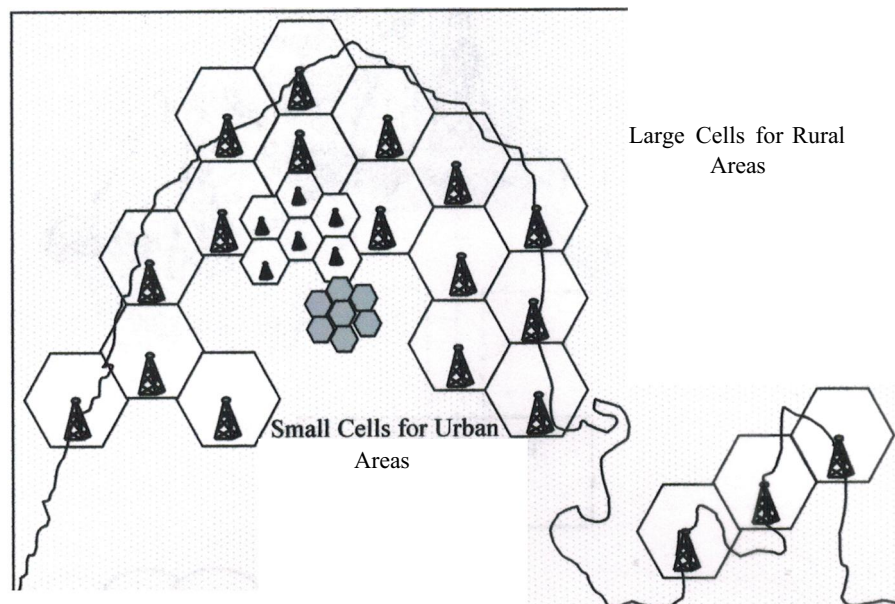


Figure 2.6. Cell Splitting

2.4.5 Handoff

The final obstacle in the development of the cellular network involved the problem created when a mobile subscriber traveled from one cell to another during a call. As adjacent areas do not use the same radio channels, a call must either be dropped or transferred from one radio channel to another when a user crosses the line between adjacent cells. Because dropping the call is unacceptable, the process of handoff was created. Handoff occurs when the mobile telephone network automatically transfers a call from radio channel to radio channel as mobile crosses adjacent cells.

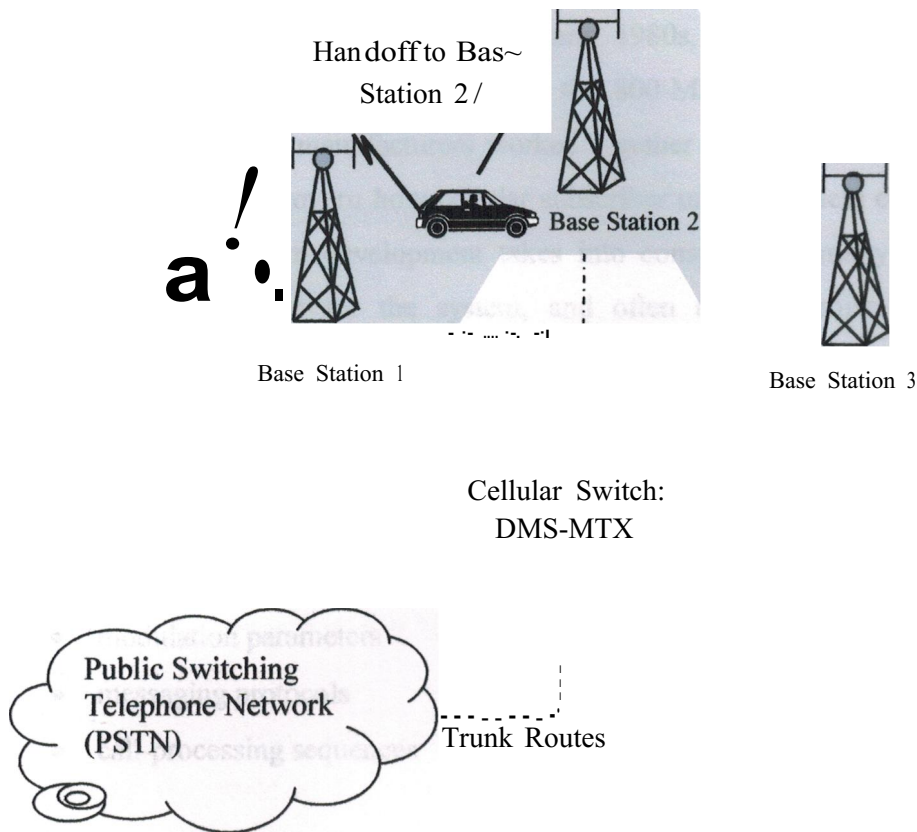


Figure 2.7. Handoff between Adjacent Cells

During a call, two parties are on one voice channel. When the mobile unit moves out of the coverage area of a given cell site, the reception becomes weak. At this point, the cell site in use requests a handoff. The system switches the call to a stronger-frequency channel in a new site without interrupting the call or alerting the user. The call continues as long as the user is talking, and the user does not notice the handoff at all.

2.5 North American Analog Cellular Systems

Originally devised in the late 1970s to early 1980s, analog systems have been revised somewhat since that time and operate in the 800-MHz range. A group of government, Telco, and equipment manufacturers worked together as a committee to develop a set of rules (protocols) that govern how cellular subscriber units (mobiles) communicate with the cellular system. System development takes into consideration many different, and often opposing, requirements for the system, and often a compromise between conflicting requirements results. Cellular development involves the following basic topics:

- frequency and channel assignments
- type of radio modulation
- maximum power levels
- modulation parameters
- messaging protocols
- call-processing sequences

2.5.1 The Advanced Mobile Phone Service (AMPS)

AMPS was released in 1983 using the 800-MHz to 900-MHz frequency band and the 30-kHz bandwidth for each channel as a fully automated mobile telephone service. It was the first standardized cellular service in the world and is currently the most widely used standard for cellular communications. Designed for use in cities, AMPS later expanded to rural areas. It maximized the cellular concept of frequency reuse by reducing radio power output. The AMPS telephones (or handsets) have the familiar telephone-style user interface and are compatible with any AMPS base station. This makes mobility between service providers (roaming) simpler for subscribers. Limitations associated with AMPS include the following:

- low calling capacity
- limited spectrum
- no room for spectrum growth
- poor data communications
- minimal privacy
- inadequate fraud protection

AMPS is used throughout the world and is particularly popular in the United States, South America, China, and Australia. AMPS uses Frequency Modulation (FM) for radio transmission. In the United States, transmissions from mobile to cell site use separate frequencies from the base station to the mobile subscriber.

2.5.2 Narrowband Analog Mobile Phone Service (NAMPS)

Since analog cellular was developed, systems have been implemented extensively throughout the world as first-generation cellular technology. In the second generation of analog cellular systems, NAMPS was designed to solve the problem of low calling capacity. NAMPS is now operational in 35 U.S. and overseas markets, and NAMPS was introduced as an interim solution to capacity problems. NAMPS is a U.S. cellular radio system that combines existing voice processing with digital signaling, tripling the capacity of today's AMPS systems. The NAMPS concept uses frequency division to get 3 channels in the AMPS 30-kHz single channel bandwidth. NAMPS provides 3 users in an AMPS channel by dividing the 30-kHz AMPS bandwidth into 3-10 kHz channels. This increases the possibility of interference because channel bandwidth is reduced.

2.6 Cellular System Components

The cellular system offers mobile and portable telephone stations the same service provided fixed stations over conventional wired loops. It has the capacity to serve tens of thousands of subscribers in a major metropolitan area. The cellular communications system consists

of the following four major components that work together to provide mobile service to subscribers.

- public switched telephone network (PSTN)
- mobile telephone switching office (MTSO)
- cell site with antenna system
- mobile subscriber unit (MSU)

2.6.1 PSTN

The PSTN is made up of local networks, the exchange area networks, and the long-haul network that interconnect telephones and other communication devices on a worldwide basis.

2.6.2 Mobile Telephone Switching Office (MTSO)

The MTSO is the central office for mobile switching. It houses the mobile switching center (MSC), field monitoring, and relay stations for switching calls from cell sites to wire line central offices (PSTN). In analog cellular networks, the MSC controls the system operation. The MSC controls calls, tracks billing information, and locates cellular subscribers.

2.6.3 The Cell Site

The term *cell site* is used to refer to the physical location of radio equipment that provides coverage within a cell. A list of hardware located at a cell site includes power sources, interface equipment, radio frequency transmitters and receivers, and antenna systems.

2.6.4 Mobile Subscriber Units (MSUs)

The mobile subscriber unit consists of a control unit and a transceiver that transmits and receives radio transmissions to and from a cell site. The following three types of MSUs are available:

- the mobile telephone (typical transmit power is 4.0 watts)
- the portable (typical transmit power is 0.6 watts)
- the transportable (typical transmit power is 1.6 watts)

The mobile telephone is installed in the trunk of a car, and the handset is installed in a convenient location to the driver. Portable and transportable telephones are hand-held and can be used anywhere. The use of portable and transportable telephones is limited to the charge life of the internal battery.

2.7 Digital Systems

As demand for mobile telephone service has increased, service providers found that basic engineering assumptions borrowed from wire line (landline) networks did not hold true in mobile systems. While the average landline phone call lasts at least 10 minutes, mobile calls usually run 90 seconds. Engineers who expected to assign 50 or more mobile phones to the same radio channel found that by doing so they increased the probability that a user would not get dial tone—this is known as call-blocking probability. As a consequence, the early systems quickly became saturated, and the quality of service decreased rapidly. The critical problem was capacity. The general characteristics of time division multiple access (TDMA), Global System for Mobile Communications (GSM), personal communications service (PCS) 1900, and code division multiple access (CDMA) promise to significantly increase the efficiency of cellular telephone systems to allow a greater number of simultaneous conversations. Figure 2.8 shows the components of a typical digital cellular system.

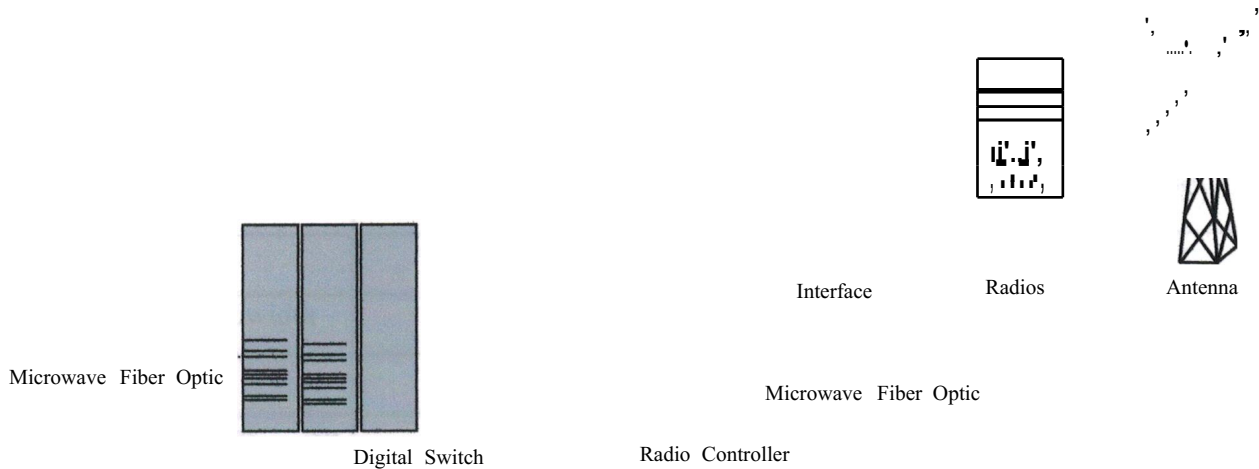


Figure 2.8. Digital Cellular System

The advantages of digital cellular technologies over analog cellular networks include increased capacity and security. Technology options such as TDMA and CDMA offer more channels in the same analog cellular bandwidth and encrypted voice and data. Because of the enormous amount of money that service providers have invested in AMPS hardware and software, providers look for a migration from AMPS to digital analog mobile phone service (DAMPS) by overlaying their existing networks with TDMA architectures.

Table 2.1. AMPS/DAMPS Comparison

	Analog	Digital
standard	EIA-553 (AMPS)	IS-54 (TDMA + AMPS)
spectrum	824 MHz to 891 MHz	824 MHz to 891 MHz
channel bandwidth	30kHz	30kHz
channels	21 cc/395 Vc	21 cc; 395 Vc
Conversations per channel	1	3 or 6
subscriber capacity	40 to 50 conversations per cell	125 to 300 conversations per cell
TXIRCVtype	continuous	time shared bursts
carrier type	constant phase variable frequency	constant frequency variable phase
mobile/base relationship	mobile slaved to base	authority shared cooperatively
privacy	poor	better--easily scrambled
noise immunity	poor	high
fraud detection	ESN plus optional password (PIN)	ESN plus optional password (PIN)

2.7.1 Time Division Multiple Access (TDMA)

North American digital cellular (NADC) is called DAMPS and TDMA. Because AMPS preceded digital cellular systems, DAMPS uses the same setup protocols as analog AMPS.

TDMA has the following characteristics:

1. IS-54 standard specifies traffic on digital voice channels
2. initial implementation triples the calling capacity of AMPS systems
3. capacity improvements of 6 to 15 times that of AMPS are possible
4. many blocks of spectrum in 800 MHz and 1900 MHz are used

5. all transmissions are digital
6. TDMA/FDMA application 7. 3 callers per radio carrier (6 callers on half rate later), providing 3 times the AMPS capacity

TDMA is one of several technologies used in wireless communications. TDMA provides each call with time slots so that several calls can occupy one bandwidth. Each caller is assigned a specific time slot. In some cellular systems, digital packets of information are sent during each time slot and reassembled by the receiving equipment into the original voice components. TDMA uses the same frequency band and channel allocations as AMPS. Like NAMPS, TDMA provides three to six time channels in the same bandwidth as a single AMPS channel. Unlike NAMPS, digital systems have the means to compress the spectrum used to transmit voice information by compressing idle time and redundancy of normal speech. TDMA is the digital standard and has 30-kHz bandwidth. Using digital voice encoders, TDMA is able to use up to six channels in the same bandwidth where AMPS uses one channel.

2.7.2 Extended Time Division Multiple Access (E-TDMA)

The E-TDMA standard claims a capacity of fifteen times that of analog cellular systems. This capacity is achieved by compressing quiet time during conversations. E-TDMA divides the finite number of cellular frequencies into more time slots than TDMA. This allows the system to support more simultaneous cellular calls.

2.7.3 Fixed Wireless Access (FWA)

FWA is a radio-based local exchange service in which telephone service is provided by common carriers. It is primarily a rural application—that is, it reduces the cost of conventional wire line. FWA extends telephone service to rural areas by replacing a wire line local loop with radio communications. Other labels for wireless access include fixed loop, fixed radio access, wireless telephony, radio loop, fixed wireless, radio access, and Ionic. FWA systems employ TDMA or CDMA access technologies.

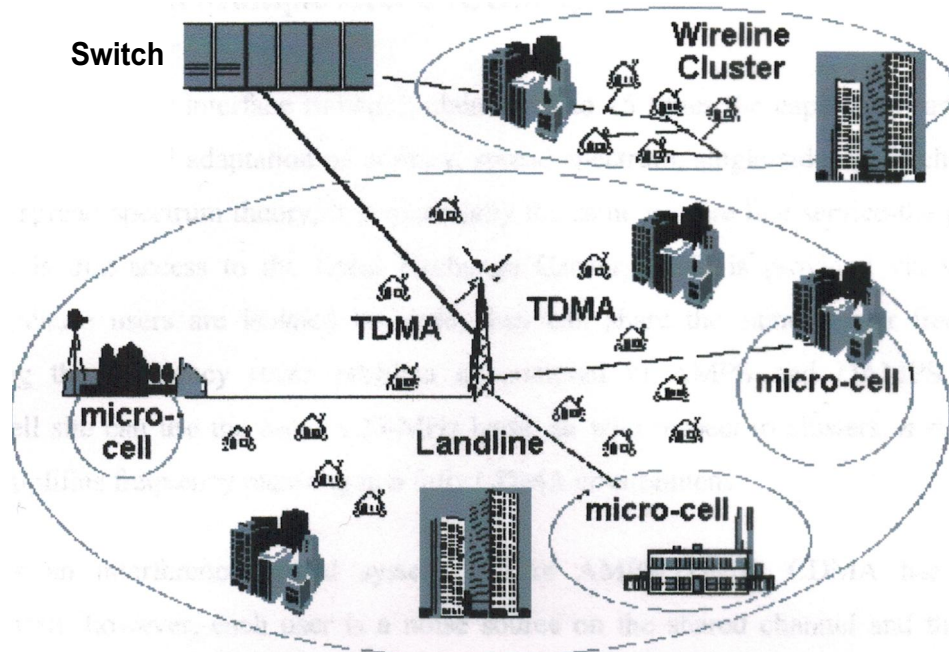


Figure 2.9 Fixed Wireless Access

2.7.4 Personal Communications Service (PCS)

The future of telecommunications includes PCS. PCS at 1900 MHz (PCS 1900) is the North American implementation of digital cellular system (DCS) 1800 (GSM). Trial networks were operational in the United States by 1993, and in 1994 the Federal Communications Commission (FCC) began spectrum auctions. As of 1995, the FCC auctioned commercial licenses. In the PCS frequency spectrum, the operator's authorized frequency block contains a definite number of channels. The frequency plan assigns, specific channels to specific cells, following a reuse pattern that restarts with each n th cell. The uplink and downlink bands are paired mirror images. As with AMPS, a channel number implies one uplink and one downlink frequency (e.g., Channel 512 = 1850.2-MHz uplink paired with 1930.2-MHz downlink).

2.7.5 Code.Division.Multiple Access (CDMA)

CDMA is a digital air interface standard, claiming 8 to 15 times the capacity of analog. It employs a commercial adaptation of military, spread-spectrum, single-sideband technology. Based on spread spectrum theory, it is essentially the same as wire line service-the primary difference is that access to the Local Exchange Carrier (LEC) is provided via wireless phone. Because users are isolated by code, they can share the same carrier frequency, eliminating the frequency reuse problem encountered in AMPS and DAMPS. Every CDMA cell site can use the same 1.25-MHz band, so with respect to clusters, $n = 1$. This greatly simplifies frequency planning in a fully CDMA environment.

CDMA is an interference-limited system. Unlike AMPS/TDMA, CDMA has a soft capacity limit; however, each user is a noise source on the shared channel and the noise contributed by users accumulates. This creates a practical limit to how many users a system will sustain. Mobiles that transmit excessive power increase interference to other mobiles. For CDMA, precise power control of mobiles is critical in maximizing the system's capacity and increasing battery life of the mobiles. The goal is to keep each mobile at the absolute minimum power level that is necessary to ensure acceptable service quality. Ideally, the power received at the base station from each mobile should be the same (minimum signal to interference).

3. THE GSM NETWORK

3.1 Overview

During the early 1980s, analog cellular telephone systems were experiencing rapid growth in Europe, particularly in Scandinavia and the United Kingdom, but also in France and Germany. Each country developed its own system, which was incompatible with everyone else's in equipment and operation. This was an undesirable situation, because not only was the mobile equipment limited to operation within national boundaries, which in a unified Europe were increasingly unimportant, but there was also a very limited market for each type of equipment, so economies of scale and the subsequent savings could not be realized.

The Europeans realized this early on, and in 1982 the Conference of European Posts and Telegraphs (CEPT) formed a study group called the Group Special Mobile (GSM) to study and develop a pan-European public land mobile system. The proposed system had to meet certain criteria:

1. Good subjective speech quality
2. Low terminal and service cost
3. Support for international roaming
4. Ability to support handheld terminals
5. Support for range of new services and facilities
6. Spectral efficiency
7. ISDN compatibility

In 1989, GSM responsibility was transferred to the European Telecommunication Standards Institute (ETSI), and phase I of the GSM specifications were published in 1990. Commercial service was started in mid-1991, and by 1993 there were 36 GSM networks in 22 countries. Although standardized in Europe, GSM is not only a European standard. Over 200 GSM networks (including DCS1800 and PCS1900) are operational in 110 countries around the world. In the beginning of 1994, there were 1.3 million subscribers worldwide, which became more than 55 million by October 1997.

With North America making a delayed entry into the GSM field with a derivative of GSM called PCS 1900, GSM systems exist on every continent, and the acronym GSM now aptly stands for Global System for Mobile communications.

The developers of GSM chose an unproven (at the time) digital system, as opposed to the then-standard analog cellular systems like AMPS in the United States and TACS in the United Kingdom. They had faith that advancements in compression algorithms and digital signal processors would allow the fulfillment of the original criteria and the continual improvement of the system in terms of quality and cost. The over 8000 pages of GSM recommendations try to allow flexibility and competitive innovation among suppliers, but provide enough standardization to guarantee proper inter working between the components of the system. This is done by providing functional and interface descriptions for each of the functional entities defined in the system.

3.2. Services provided by GSM

From the beginning, the planners of GSM wanted ISDN compatibility in terms of the services offered and the control signaling used. However, radio transmission limitations, in terms of bandwidth and cost, do not allow the standard ISDN B-channel bit rate of 64 kbps to be practically achieved.

Using the ITU-T definitions, telecommunication services can be divided into bearer services, tele-services, and supplementary services. The most basic tele-service supported by GSM is telephony. As with all other communications, speech is digitally encoded and transmitted through the GSM network as a digital stream. There is also an emergency service, where the nearest emergency-service provider is notified by dialing three <igits (similar to 91 1).

A variety of data services is offered. GSM users can send and receive data, at rates up to 9600 bps, to users on POTS (Plain Old Telephone Service), ISDN, Packet Switched Public Data Networks, and Circuit Switched Public Data Networks using a variety of access methods and protocols, such as X.25 or X.32. Since GSM is a digital network, a modem is not required between the user and GSM network, although an audio modem is required inside the GSM network to inter work with POTS.

Other data services include Group 3 facsimile, as described in ITU-T recommendation T.30, which is supported by use of an appropriate fax adaptor. A unique feature of GSM, not found in older analog systems, is the Short Message Service (SMS). SMS is a bi directional service for short alphanumeric (up to 160 bytes) messages. Messages are transported in a store-and-forward fashion. For point-to-point SMS, a message can be sent to another subscriber to the service, and an acknowledgement of receipt is provided to the sender. SMS can also be used in a cell-broadcast mode, for sending messages such as traffic updates or news updates. Messages can also be stored in the SIM card for later retrieval.

Supplementary services are provided on top of tele-services or bearer services. In the current (Phase I) specifications, they include several forms of call forward (such as call forwarding when the mobile subscriber is unreachable by the network), and call barring of outgoing or incoming calls, for example when roaming in another country. Many additional supplementary services will be provided in the Phase 2 specifications, such as caller identification, call waiting, multi-party conversations.

3.3. Architecture of the GSM network

A GSM network is composed of several functional entities, whose functions and interfaces are specified. Figure 3.1 shows the layout of a generic GSM network. The GSM network can be divided into three broad parts. The Mobile Station is carried by the subscriber. The Base Station Subsystem controls the radio link with the Mobile Station. The Network Subsystem, the main part of which is the Mobile services Switching Center (MSC), performs the switching of calls between the mobile users, and between mobile and fixed network users. The MSC also handles the mobility management operations. Not shown is the Operations and Maintenance Center, which oversees the proper operation and setup of the network. The Mobile Station and the Base Station Subsystem communicate across the Um interface, also known as the air interface or radio link. The Base Station Subsystem communicates with the Mobile services Switching Center across the A interface.

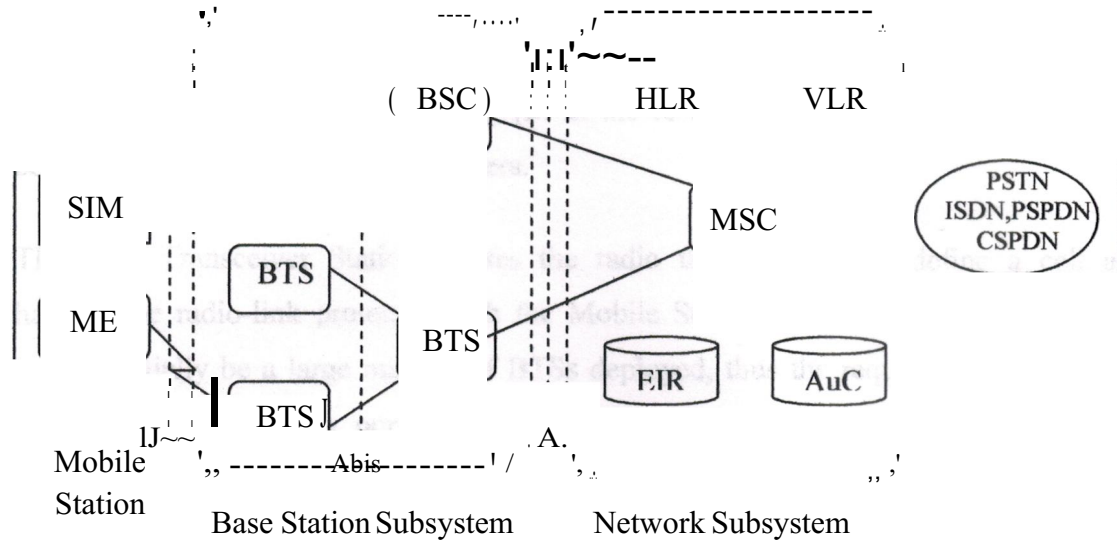


Figure 3.1. General Architecture of a GSM Network

3.3.1 Mobile Station

The mobile station (MS) consists of the mobile equipment (the terminal) and a smart card called the Subscriber Identity Module (SIM). The SIM provides personal mobility, so that the user can have access to subscribed services irrespective of a specific terminal. By inserting the SIM card into another GSM terminal, the user is able to receive calls at that terminal, make calls from that terminal, and receive other subscribed services.

The mobile equipment is uniquely identified by the International Mobile Equipment Identity (IMEI). The SIM card contains the International Mobile Subscriber Identity (IMSI) used to identify the subscriber to the system, a secret key for authentication, and other information. The IMEI and the IMSI are independent, thereby allowing personal mobility. The SIM card may be protected against unauthorized use by a password or personal identity number.



3.3.2 Base Station Subsystem

The Base Station Subsystem is composed of two parts, the Base Transceiver Station (BTS) and the Base Station Controller (BSC). These communicate across the standardized Abis interface, allowing (as in the rest of the system) operation between components made by different suppliers.

The Base Transceiver Station houses the radio transceivers that define a cell and handles the radio-link protocols with the Mobile Station. In a large urban area, there will potentially be a large number of BTSs deployed, thus the requirements for a BTS are ruggedness, reliability, portability, and minimum cost.

The Base Station Controller manages the radio resources for one or more BTSs. It handles radio-channel setup, frequency hopping, and handovers, as described below. The BSC is the connection between the mobile station and the Mobile service Switching Center (MSC).

3.3.3 Network Subsystem

The central component of the Network Subsystem is the Mobile services Switching Center (MSC). It acts like a normal switching node of the PSTN or ISDN, and additionally provides all the functionality needed to handle a mobile subscriber, such as registration, authentication, location updating, handovers, and call routing to a roaming subscriber. These services are provided in conjunction with several functional entities, which together form the Network Subsystem. The MSC provides the connection to the fixed networks (such as the PSTN or ISDN). Signaling between functional entities in the Network Subsystem uses Signaling System Number 7 (SS7), used for trunk signaling in ISDN and widely used in current public networks.

The Home Location Register (HLR) and Visitor Location Register (VLR), together with the MSC, provide the call-routing and roaming capabilities of GSM. The HLR contains all the administrative information of each subscriber registered in the corresponding GSM network, along with the current location of the mobile. The location of the mobile is typically in the form of the signaling address of the VLR associated with the mobile

station. The actual routing procedure will be described later. There is logically one HLR per GSM network, although it may be implemented as a distributed database.

The Visitor Location Register (VLR) contains selected administrative information from the HLR, necessary for call control and provision of the subscribed services, for each mobile currently located in the geographical area controlled by the VLR. Although each functional entity can be implemented as an independent unit, all manufacturers of switching equipment to date implement the VLR together with the MSC, so that the geographical area controlled by the MSC corresponds to that controlled by the VLR, thus simplifying the signaling required. Note that the MSC contains no information about particular mobile stations --- this information is stored in the location registers.

The other two registers are used for authentication and security purposes. The Equipment Identity Register (EIR) is a database that contains a list of all valid mobile equipment on the network, where each mobile station is identified by its International Mobile Equipment Identity (IMEI). An IMEI is marked as invalid if it has been reported stolen or is not type approved. The Authentication Center (AuC) is a protected database that stores a copy of the secret key stored in each subscriber's SIM card, which is used for authentication and encryption over the radio channel.

3.4 Radio Link Aspects

The International Telecommunication Union (ITU), which manages the international allocation of radio spectrum (among many other functions), allocated the bands 890-915 MHz for the uplink (mobile station to base station) and 935-960 MHz for the downlink (base station to mobile station) for mobile networks in Europe. Since this range was already being used in the early 1980s by the analog systems of the day, the CEPT had the foresight to reserve the top 10 MHz of each band for the GSM network that was still being developed. Eventually, GSM will be allocated the entire 2x25 MHz bandwidth.

3.4.1 Multiple Access and Channel Structure

Since radio spectrum is a limited resource shared by all users, a method must be devised to divide up the bandwidth among as many users as possible. The method chosen by GSM is a combination of Time- and Frequency-Division Multiple Access (TDMA/FDMA). The FDMA part involves the division by frequency of the (maximum) 25 MHz bandwidth into 124 carrier frequencies spaced 200 kHz apart. One or more carrier frequencies are assigned to each base station. Each of these carrier frequencies is then divided in time, using a TDMA scheme. The fundamental unit of time in this TDMA scheme is called a *burst period* and it lasts $15/26$ ms (or approx. 0.577 ms). Eight burst periods are grouped into a *TDMA frame* ($120/26$ ms, or approx. 4.615 ms), which forms the basic unit for the definition of logical channels. One physical channel is one burst period per TDMA frame.

Channels are defined by the number and position of their corresponding burst periods. All these definitions are cyclic, and the entire pattern repeats approximately every 3 hours. Channels can be divided into *dedicated channels*, which are allocated to a mobile station, and *common channels*, which are used by mobile stations in idle mode.

3.4.1.1 Traffic Channels

A traffic channel (TCH) is used to carry speech and data traffic. Traffic channels are defined using a 26-frame multi-frame, or group of 26 TDMA frames. The length of a 26-frame multi-frame is 120 ms, which is how the length of a burst period is defined (120 ms divided by 26 frames divided by 8 burst periods per frame). Out of the 26 frames, 24 are used for traffic, 1 is used for the Slow Associated Control Channel (SACCH) and 1 is currently unused (see Figure 3.2). TCHs for the uplink and downlink are separated in time by 3 burst periods, so that the mobile station does not have to transmit and receive simultaneously, thus simplifying the electronics.

In addition to these *full-rate* TCHs, there are also *half-rate* TCHs defined, although they are not yet implemented. Half-rate TCHs will effectively double the capacity of a system once half-rate speech coders are specified (i.e., speech coding at around 7 kbps, instead of 13 kbps). Eighth-rate TCHs are also specified, and are used for signaling. In

the recommendations, they are called Stand-alone Dedicated Control Channels (SDCCH).

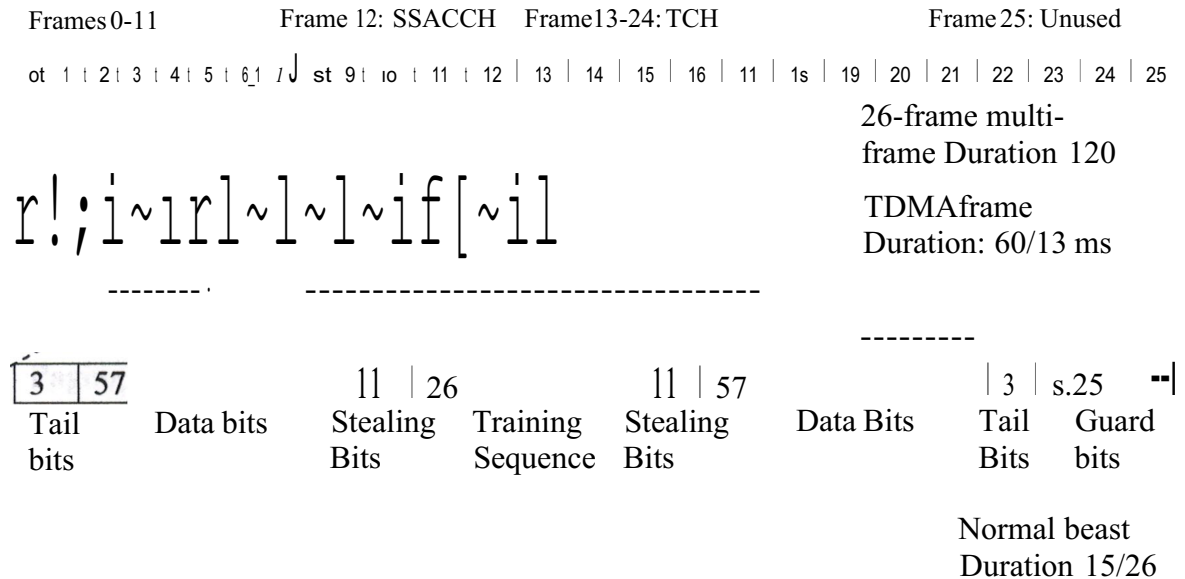


Figure 3.2. Organization of bursts, TDMA frames, and multi-frames for speech and data

3.4.1.2 Control Channels

Common channels can be accessed both by idle mode and dedicated mode mobiles. The common channels are used by idle mode mobiles to exchange the signaling information required to change to dedicated mode. Mobiles already in dedicated mode monitor the surrounding base stations for handover and other information. The common channels are defined within a 51-frame multi-frame, so that dedicated mobiles using the 26-frame multi-frame TCH structure can still monitor control channels. The common channels include:

Broadcast Control Channel (BCCH)

Continually broadcasts, on the downlink, information including base station identity, frequency allocations, and frequency-hopping sequences.

Frequency Correction Channel (FCCH) and Synchronization Channel (SCH)

Used to synchronize the mobile to the time slot structure of a cell by defining the boundaries of burst periods, and the time slot numbering. Every cell in a GSM network broadcasts exactly one FCCH and one SCH, which are by definition on time slot number 0 (within a TDMA frame).

Random Access Channel (RACH)

Slotted Aloha channel used by the mobile to request access to the network.

Paging Channel (PCH)

Used to alert the mobile station of an incoming call.

Access Grant Channel (AGCH)

Used to allocate an SDCCH to a mobile for signaling (in order to obtain a dedicated channel), following a request on the RACH.

3.4.1.3 Burst Structure

There are four different types of bursts used for transmission in GSM. The normal burst is used to carry data and most signaling. It has a total length of 156.25 bits, made up of two 57 bit information bits, a 26 bit training sequence used for equalization, 1 stealing bit for each information block (used for FACCH), 3 tail bits at each end, and an 8.25 bit guard sequence, as shown in Figure 3.2 The 156.25 bits are transmitted in 0.577 ms, giving a gross bit rate of 270.833 kbps.

The F burst, used on the FCCH, and the S burst, used on the SCH, have the same length as a normal burst, but a different internal structure, which differentiates them from normal bursts (thus allowing synchronization). The access burst is shorter than the normal burst, and is used only on the RACH.

3.4.2 Speech Coding

GSM is a digital system, so speech which is inherently analog, has to be digitized. The method employed by ISDN, and by current telephone systems for multiplexing voice lines over high speed trunks and optical fiber lines, is Pulse Coded Modulation (PCM). The output stream from PCM is 64 kbps, too high a rate to be feasible over a radio link. The 64 kbps signal, although simple to implement, contains much redundancy. The GSM group studied several speech coding algorithms on the basis of subjective speech quality and complexity (which is related to cost, processing delay, and power consumption once implemented) before arriving at the choice of a Regular Pulse Excited -- Linear Predictive Coder (RPE--LPC) with a Long Term Predictor loop. Basically, information from previous samples, which does not change very quickly, is used to predict the current sample. The coefficients of the linear combination of the previous samples, plus an encoded form of the residual, the difference between the predicted and actual sample, represent the signal. Speech is divided into 20 millisecond samples, each of which is encoded as 260 bits, giving a total bit rate of 13 kbps. This is the so-called Full-Rate speech coding. Recently, an Enhanced Full-Rate (EFR) speech coding algorithm has been implemented by some North American GSM1900 operators. This is said to provide improved speech quality using the existing 13 kbps bit rate.

3.4.3 Channel Coding and Modulation

Because of natural and man-made electromagnetic interference, the encoded speech or data signal transmitted over the radio interface must be protected from errors. GSM uses convolution encoding and block interleaving to achieve this protection. The exact algorithms used differ for speech and for different data rates. The method used for speech blocks will be described below.

Recall that the speech code produces a 260 bit block for every 20 ms speech sample. From subjective testing, it was found that some bits of this block were more important for perceived speech quality than others.

Tue bits are thus divided into three classes:

- Class Ia 50 bits - most sensitive to bit errors
- Class Ib 132 bits - moderately sensitive to bit errors
- Class II 78 bits - least sensitive to bit errors

Class Ia bits have a 3 bit Cyclic Redundancy Code added for error detection. If an error is detected, the frame is judged too damaged to be comprehensible and it is discarded. It is replaced by a slightly attenuated version of the previous correctly received frame. These 53 bits, together with the 132 Class Ib bits and a 4 bit tail sequence (a total of 189 bits), are input into a $1/2$ rate convolution encoder of constraint length 4. Each input bit is encoded as two output bits, based on a combination of the previous 4 input bits. The convolution encoder thus outputs 378 bits, to which are added the 78 remaining Class II bits, which are unprotected. Thus every 20 ms speech sample is encoded as 456 bits, giving a bit rate of 22.8 kbps.

To further protect against the burst errors common to the radio interface, each sample is interleaved. The 456 bits output by the convolution encoder are divided into 8 blocks of 57 bits, and these blocks are transmitted in eight consecutive time-slot bursts. Since each time-slot burst can carry two 57 bit blocks, each burst carries traffic from two different speech samples.

Recall that each time-slot burst is transmitted at a gross bit rate of 270.833 kbps. This digital signal is modulated onto the analog carrier frequency using Gaussian-filtered Minimum Shift Keying (GMSK). GMSK was selected over other modulation schemes as a compromise between spectral efficiency, complexity of the transmitter, and limited spurious emissions. The complexity of the transmitter is related to power consumption, which should be minimized for the mobile station. The spurious radio emissions, outside of the allotted bandwidth, must be strictly controlled so as to limit adjacent channel interference, and allow for the co-existence of GSM and the older analog systems (at least for the time being).

3.4.4 Multipath Equalization

At the 900 MHz range, radio waves bounce off everything—buildings, hills, cars, airplanes, etc. Thus many reflected signals, each with a different phase, can reach an antenna. Equalization is used to extract the desired signal from the unwanted reflections. It works by finding out how a known transmitted signal is modified by multipath fading, and constructing an inverse filter to extract the rest of the desired signal. This known signal is the 26-bit training sequence transmitted in the middle of every time-slot burst. The actual implementation of the equalizer is not specified in the GSM specifications.

3.4.5 Frequency Hopping

The mobile station already has to be frequency agile, meaning it can move between a transmit, receive, and monitor time slot within one TDMA frame, which normally are on different frequencies. GSM makes use of this inherent frequency agility to implement slow frequency hopping, where the mobile and BTS transmit each TDMA frame on a different carrier frequency. The frequency hopping algorithm is broadcast on the Broadcast Control Channel. Since multipath fading is dependent on carrier frequency, slow frequency hopping helps alleviate the problem. In addition, co-channel interference is in effect randomized.

3.4.6 Discontinuous Transmission

Minimizing co-channel interference is a goal in any cellular system, since it allows better service for a given cell size, or the use of smaller cells, thus increasing the overall capacity of the system. Discontinuous transmission (DTX) is a method that takes advantage of the fact that a person speaks less than 40 percent of the time in normal conversation, by turning the transmitter off during silence periods. An added benefit of DTX is that power is conserved at the mobile unit.

The most important component of DTX is, of course, Voice Activity Detection. It must distinguish between voice and noise inputs, a task that is not as trivial as it appears, considering background noise. If a voice signal is misinterpreted as noise, the transmitter is turned off and a very annoying effect called clipping is heard at the

receiving end. If, on the other hand, noise is misinterpreted as a voice signal too often, the efficiency of DTX is dramatically decreased. Another factor to consider is that when the transmitter is turned off, there is total silence heard at the receiving end, due to the digital nature of GSM. To assure the receiver that the connection is not dead, *comfort noise* is created at the receiving end by trying to match the characteristics of the transmitting end's background noise.

3.4.7 Discontinuous Reception

Another method used to conserve power at the mobile station is discontinuous reception. The paging channel, used by the base station to signal an incoming call, is structured into sub-channels. Each mobile station needs to listen only to its own sub-channel. In the time between successive paging sub-channels, the mobile can go into sleep mode, when almost no power is used.

3.4.8 Power Control

There are five classes of mobile stations defined, according to their peak transmitter power, rated at 20, 8, 5, 2, and 0.8 watts. To minimize co-channel interference and to conserve power, both the mobiles and the Base Transceiver Stations operate at the lowest power level that will maintain an acceptable signal quality. Power levels can be stepped up or down in steps of 2 dB from the peak power for the class down to a minimum of 13 dBm (20 milliwatts).

The mobile station measures the signal strength or signal quality (based on the Bit Error Ratio), and passes the information to the Base Station Controller, which ultimately decides if and when the power level should be changed. Power control should be handled carefully, since there is the possibility of instability.

This arises from having mobiles in co-channel cells alternating increase their power in response to increased co-channel interference caused by the other mobile increasing its power. This is unlikely to occur in practice but it is (or was as of 1991) under study.

4. GSM SECURITY

4.1 Overview

Security was perceived by some in the first days of GSM as an unnecessary expense. Certainly, initially, all involved considered protection of user data from eavesdropping as more important than authentication of the user, though some questioned whether the perceived complexity of introducing encryption over the radio interface was justified. However, as fraud losses from cloning of analogue phones rocketed in the US and the UK, as the Dutch PTT withdrew all its NMT phones so a form of authentication could be added, and as Germans introduced simple authentication on its C-Netz system, it became apparent that authentication of user identity was also very important.

In recent times, the security of GSM has been attacked as too weak. These criticisms are often made without knowledge of neither, the design goals of GSM security nor the regulatory context in which the designers had to work. This chapter aims to show that GSM security met its design goals in a simple and elegant way, and has provided more than adequate security for most of its users. Indeed, GSM offers more "access network" security than fixed phones in most countries (taking the phone to the local exchange link as the "access network" for fixed line systems). GSM has never been subject to commercial cloning that was visited upon analogue NMT, AMPS and TACS system. Moreover, GSM represented the first time ever that encryption functionality had been provided in a consumer device, and played its part in the liberalization of policy on encryption that today's security designers enjoy.

Securing information from unauthorized access is a major problem for any network wire line or wireless. Security, in a broad sense, focuses on network security, system security, information security, and physical security. It is made up of a suite of multiple technologies that solve numerous authentication, information integrity, and identification problems. The mobile telephone has become a necessity for the business professionals or even the general public. Although it provides the convenience of being able to contact people anywhere, a major drawback cannot be ignored - the mobile telephone communication uses radio signal, which is subject to eavesdropping. The mobile telephone network is also subject to other unwanted security breaches, which cost hundreds of million dollars a year in the mobile phone business. There are new

technologies emerging which provides a data transfer between different kinds of devices that are close to each other. Security services are needed to offer for mobile users without possibilities to violate the general trust.

4.2 Origins of GSM Security

The security of GSM was developed by the Security Experts Group (SEG) which was formed by CEPT in 1984. There was a lot of concern in CEPT regarding protection of communications systems in general at that time. The origin of the SEG could be said to be a joint meeting of the three CEPT groups CD (data), CS (signaling) and SF (services and facilities) in Berne in January 1984, in which land mobile systems were discussed for the first time. In November, 1984, a proposal from CD to set up a joint CD-GSM group on security (SEG) was accepted by GSM and the first meeting of SEG was held in Malmoe, Sweden, in May 1985. This was a memorable meeting for the delegates as the Swedish air traffic controllers were on strike at that time, forcing the delegates to fly to Copenhagen and travel by boat to Malmoe. The SEG was initially a joint CD/GSM activity, but gradually, the CD part vanished so it was in fact a subgroup of GSM.

The SEG was chaired by Thomas Haug of Swedish Telecommunications Administration, now called Telia. The membership, like that of CEPT and GSM was drawn from national PTTs and from those organizations that had won a mobile network licence in their country.

4.3 Security Services

When two or more parties exchange information in a session, then authenticity, confidentiality and integrity are needed. Firstly, the parties need to be sure about the identities of each other. Authentication means that the participants somehow prove that their identities are what they claim them to be. Secondly, an outsider should not be able to read the messages. This is meant by confidentiality. Thirdly, integrity requires that messages should not be altered during transmissions. Sometimes also non-repudiation is needed. That is, the sender of a message cannot later deny sending the information, or the receiver cannot deny the reception.

Authentication, integrity, and confidentiality are connected to each other. An authentic message originates from the party it claims to come from, and it arrives unchanged, so this implies integrity. On the other hand, authentication of an intended receiver of a message is used to provide confidentiality. The services in the network may contain classified information, or their computational resources may be limited. Therefore the users may have different access rights. The users need to be authorized to access the services, and access control is needed to prevent unauthorized access.

This way access control provides confidentiality and integrity of the resources. Sometimes the users may need the service to proof the authenticity of the resources as well. A security need for both the sessions and the resources is availability. For the users to be able to communicate with each other and access the resources at all times, high availability is desired. Most security threats fall into the following categories: disclosure of information, loss of integrity, unauthorized access and denial of service.

4.4 Security inside the GSM Network

The GSM security requirements addressed are: data confidentiality over the radio channel, subscriber authentication, and confidentiality of the subscriber's identity and location. The GSM specifications were designed by the GSM Consortium in secrecy and were distributed only on a need-to-know basis to hardware and software manufacturers and to GSM network operators. The specifications were never exposed to the public, thus preventing the open science community around the world from studying the enclosed authentication and enciphering algorithms as well as the whole GSM security model.

The algorithm in question should be publicly available, so that the algorithm is exposed to the scrutiny of the public. According to the general opinion no single entity can employ enough experts to compete with the open scientific community in crypt analyzing an algorithm. Thus, the algorithms designed and implemented in secrecy will probably be somehow cryptographically weak and contain design faults. Eventually, the GSM algorithms leaked out and have been studied extensively ever since by the open scientific community. Interesting facts have been discovered since then, during the cryptanalysis of the A3, A5 and A8 algorithms.

Three distinct security services are provided. These are subscriber identity authentication, user and signalling data confidentiality, and subscriber identity confidentiality. Each of these is considered in turn, and the mechanisms used to provide them outlined. Actually the second of the services is a grouping of three GSM features: user data confidentiality on physical connections, connectionless user data confidentiality, and signalling information element confidentiality. The reason for combining them into one service is that they are all provided by one and the same mechanism.

4.4.1 Subscriber Identity Authentication

This subscriber identity authentication service is the heart of the GSM security system. It is used to enable the fixed network to authenticate the identity of mobile subscribers, and to establish and manage the encryption keys needed to provide the confidentiality services. All networks and mobiles must support the service, although the frequency of application is at the discretion of the network.

Authentication is initiated by the fixed network, and is based upon a simple challenge-response protocol. When a mobile subscriber (MS) attempts to access the system, the network issues it a random challenge RAND. The MS computes a response SRES to RAND using a one-way function A3 under control of a subscriber authentication key Ki. The key Ki is unique to the subscriber, and is shared only by the subscriber and an authentication centre that serves the subscriber's home network. The value SRES computed by the MS is signaled to the network, where it is compared with a pre-computed value. If the two values of SRES agree, the mobile subscriber has been authenticated, and the call is allowed to proceed. If the values are different, then access is denied.

The same mechanism is also used to establish a cipher key Kc for encrypting user and signaling data on the radio path. The key is computed by the MS using a one-way function A8, again under control of the subscriber authentication key Ki, and is pre-computed for the network by the authentication centre that serves the subscriber's home network. Thus at the end of a successful authentication exchange, both parties possess a fresh cipher key Kc.

The pre-computed triples (RAND, SRES, Kc), held by the fixed networks for a particular subscriber, and is passed from the home network's authentication centre to visited networks upon demand. The challenges are used just once. Thus the authentication centre never sends the same triple to two distinct networks, and a network never re-uses a challenge.

In practice the two functions A3 and A8 are combined into a single algorithm, called A38, which is used to simultaneously compute SRES and Kc from RAND and Ki. In this project this combined algorithm is referred to as the authentication algorithm. The protocol described above makes it quite clear that this algorithm need only be available to an authentication centre and the mobile subscribers that that authentication centre serves. In particular, there is no need for a common GSM authentication algorithm and different networks may use different algorithms. (The algorithms do, however, need to have the same input and output parameters; in particular, the length of Kc is determined by the GSM cipher algorithm). Nevertheless it is desirable that there is a GSM standard authentication algorithm, which may be used by all networks, which do not wish to develop a proprietary algorithm. The operators may free to design their A3 algorithm.

4.4.2 User and Signalling Data Confidentiality

As mentioned earlier, this service consists of three elements: user data confidentiality and signaling information on physical connections, connectionless user data confidentiality and signaling information element confidentiality. The first element provides for privacy of all user-generated data, both voice and non-voice, transferred over the radio path on traffic channels. The second element provides for privacy of user data transferred in packet mode over the radio path on a dedicated signaling channel, whilst the third element provides for privacy of certain user related signaling elements transferred over the radio path on dedicated signaling channels.

All of these elements of service are provided using the same layer 1 encryption mechanism, and must be supported and used by all networks and mobiles. The mechanism is now briefly described. Encryption is achieved by means of a ciphering algorithm *A5* that produces a key stream under control of a cipher key Kc. This key stream is then bit-for-bit exclusive-or and with the data transferred over the radio path between the MS and the base station (BS). The cipher key is established at the MS as

part of the authentication procedure, as described in the last section, and is transferred through the fixed network to the BS after the MS has been identified.

It is essential that the MS and BS synchronize the starting of their cipher algorithms, but this only directly addresses the situation when the network initiates an authentication check. The procedures still need to be specified in detail to cover the situation when the network does not authenticate the MS. When the network intends to issue an authentication challenge, the BS starts deciphering all data immediately after the MS has been identified using the cipher key K_c that the MS will derive upon receipt of the challenge RAND. The MS starts ciphering and deciphering the moment it has computed K_c (and SRES) from RAND, as described in the last section, and before SRES is transmitted. On the BS side, enciphering starts as soon as SRES has been received, deciphered and found to be correct. To cope with possible transmission loss or errors, the authentication request and response message are repeated under the control of timers.

Synchronization of the ciphering key stream is maintained by using the TDMA frame structure of the radio sub-system. The TDMA frame number is used as a message key for the cipher algorithm A5, and the algorithm produces a synchronized key stream for enciphering and deciphering the data bits in the frame. For each frame, a total of 114 bits are produced for enciphering / deciphering data transferred from the MS to the BS, and an additional 114 bits are produced for deciphering / enciphering data received at the MS from the BS. A frame lasts for 4.6 ms, so that the cipher has to produce the 228 bits in this time. The cipher algorithm A5 must be common to all GSM networks, and three algorithms have been proposed as candidates, for the GSM standard: a French algorithm, a Swedish algorithm and a UK algorithm.

4.4.3 Subscriber Identity Confidentiality

Subscriber identity confidentiality on the radio interface was one of the security requirements of GSM.

This service allows mobile subscribers to originate calls, update their location, etc, without revealing their International Mobile Subscriber Identity (iMSi) to an eavesdropper on the radio path. It thus prevents location tracing of individual mobile

subscribers by listening to the signaling exchanges on the radio path. All mobiles and networks must be capable of supporting the service, but its use is not mandatory.

A robust identity confidentiality mechanism is in fact quite a difficult thing to achieve. Confidentiality usually involves encryption and encryption requires, for symmetric ciphering, a shared secret key. However, generation of a shared secret key should generally be done in combination with, or after authentication, or how does one entity know who they are sharing a secret key with and who therefore they are revealing their identity too? Authentication, however, is authentication of a particular identity, but identities have not yet been exchanged!

Mechanisms involving public key cryptography can be used but only if one of the entities (e.g. a "server" or a network operator) does not mind revealing its identity to eavesdroppers.

A simple mechanism involving public keys might be that one entity (the "server") transmits a certificate for its public key and the other entity encrypts its identity using the received public key. The transmitted identity can then be authenticated by a variety of means that do not reveal the identity to passive eavesdroppers.

Public key cryptography was not available to the GSM designers, so a simple mechanism, using temporary identities and the basic facilities of GSM security was designed.

In order to provide the subscriber identity confidentiality service it is necessary to ensure that the IMSI, or any information which allows an eavesdropper to derive the IMSI, is not (normally) transmitted in clear in any signaling message on the radio path. The mechanism used to provide this service is based on the use of a temporary mobile subscriber identity (TMSI), which is securely updated after each successful access to the system. Thus, in principle, the [MSI] need only be transmitted in clear over the radio path at registration. When a subscriber attempts access with an operator with which it is not presently registered (so, first access in a roamed to network, or the first access for time in its home network) it must reveal its identity, and request access using its permanent identity, the International Mobile Subscriber Identity, or IMSI. The [MSI] is then authenticated, a process which results in the sharing of Kc. The subscriber is then assigned a Temporary Mobile Subscriber Identity (TMSI, pronounced "timsy") which is

sent to the subscriber encrypted with K_c . The next time the user attempts access in that network, it uses the TMSI to identify itself and the network looks up its table of TMSI to IMSI mapping to find the subscriber's permanent identity and the triplets with which it can authenticate the subscriber and begin encryption. So that a subscriber cannot be followed around, it is frequently given a new TMSI (if the same TMSI were used for a while, a subscriber previously identified by some out of band means could be recognized by the TMSI). Theoretically, the IMSI should only have to be used on a subscriber's first ever registration with any network, and it should be possible for the TMSI to be used even across different networks. In practice, however, the IMSI must be revealed on first registration in a new network at least, and in some networks, more frequently than this.

The GSM identity confidentiality is simple and efficient, but is not robust. The TMSI must be revealed on first registration with a network, and the mechanism as a whole can be compromised using a "false base station".

The TMSI updating mechanism functions in the following manner. For simplicity, assume the MS has been allocated a TMSI, denoted by TMS_{io} , and the network knows the association between TMS_{io} and the subscriber's IMSI. The MS identifies itself to the network by sending TMS_{fo} . Immediately after authentication (if this takes place), the network generates a new TMSI, denoted TMS_{in} , and sends this to the MS encrypted under the cipher key k_c . Upon receipt of the message, the MS deciphers and replaces TMS_{io} by TMS_{in} .

4.5 GSM Security Model

4.5.1 The Purpose of GSM Security:

The use of radio communications for transmission to the mobile subscribers makes GSM Public Land Mobile Networks (PLMN) particularly sensitive to misuse of their resources by unauthorized persons using manipulated Mobile Stations, who try to impersonate authorized subscribers and eavesdropping of the various information, which are exchanged on the radio path.

So the security features in GSM PLMN is implemented to protect:

- The access to the mobile services.
- Any relevant item from being disclosed at the radio path, mainly in order to ensure the privacy of user-related information.

Since all cases of GSM fraud against a specific wireless carrier will result in a substantial loss to the operator. This substantial loss may include the following:

- No direct financial loss, where the result is lost customers and increase in use of the system with no revenue.
- Direct financial loss, where money is paid out to others, such as other networks, carriers and operators of Value Added Networks' such as Premium Rate service lines.
- Potential embarrassment, where customers may move to another service because of the lack of security.
- Failure to meet legal and regulatory requirements, such as License conditions, Companies Acts or Data Protection Legislation

4.5.2 GSM's Security Limitations

Existing cellular systems have a number of potential weaknesses that were considered in the security requirements for GSM.

The security for GSM has to be appropriate for the system operator and customer:

- The operators of the system wish to ensure, that they could issue bills to the right people, and that the services cannot be compromised.
- The customer requires some privacy against traffic being overheard,

The countermeasures are designed to:

- make the radio path as 'Secure as the fixed network, which implies anonymity and confidentiality to protect against eavesdropping;
- have strong authentication, to protect the operator against billing fraud;
- prevent operators from compromising each others' security, whether inadvertently or because of competitive pressures.

The security processes must not:

- significantly add to the delay of the initial can set up or subsequent communication;
- increase the bandwidth of the channel,
- allow for increased error rates, or error propagation;
- add excessive complexity to the rest of the system,
- must be cost effective.

The designs of an operator's GSM system should take into account, the environment and have secure procedures such as:

- the generation and distribution of keys,
- exchange of information between operators,
- the confidentiality of the algorithms.

4.5.3 Descriptions of the functions of the services

The security services provided by GSM are:

- **Anonymity** so that it is not easy to identify the user of the system,
- **Authentication** so the operator knows who is using the system for billing purposes.
- **Signalling Protection** so that sensitive information on the signaling channel, such as telephone numbers, is protected over the radio path.
- **User Data Protection** so that user data passing over the radio path is protected.

Anonymity

Anonymity is provided by using temporary identifiers. When a user first switches on his/her radio set, the real identity is used, and a temporary identifier is then issued. From then on the temporary identifier is used. Only by tracking the user is it possible to determine the temporary identity being used.

Authentication

Authentication is used to identify the user (or holder of a Smart Card] to the network operator. It uses a technique that can be described as a "Challenge and Response", based on encryption,

Authentication is performed by a challenge and response mechanism. A random challenge is issued to the mobile, the mobile encrypts the challenge using the authentication algorithm (A3) and the key assigned to the mobile, and sends a response back. The operator can check that, given the key of the mobile, the response to the challenge is correct.

Eavesdropping the radio channel reveals no useful information, as the next time a new random challenge will be used. Authentication can be provided using this process. A random number is generated by the network and sent to the mobile. The mobile uses the Random number R as the input (Plaintext) to the encryption, and, using a secret key unique to the mobile K_i , transforms this into a response Signed Response (SRES) (Cipher text) which is sent back to the network.

The network can check that the mobile really has the secret key by performing the same SRES process and comparing the responses with what it receives from the mobile.

User Data and Signalling Protection

The response is then passed through an algorithm A8 by both the mobile and the network to derive the key K_c used for encrypting the signalling and messages to provide privacy (A5 series algorithms).

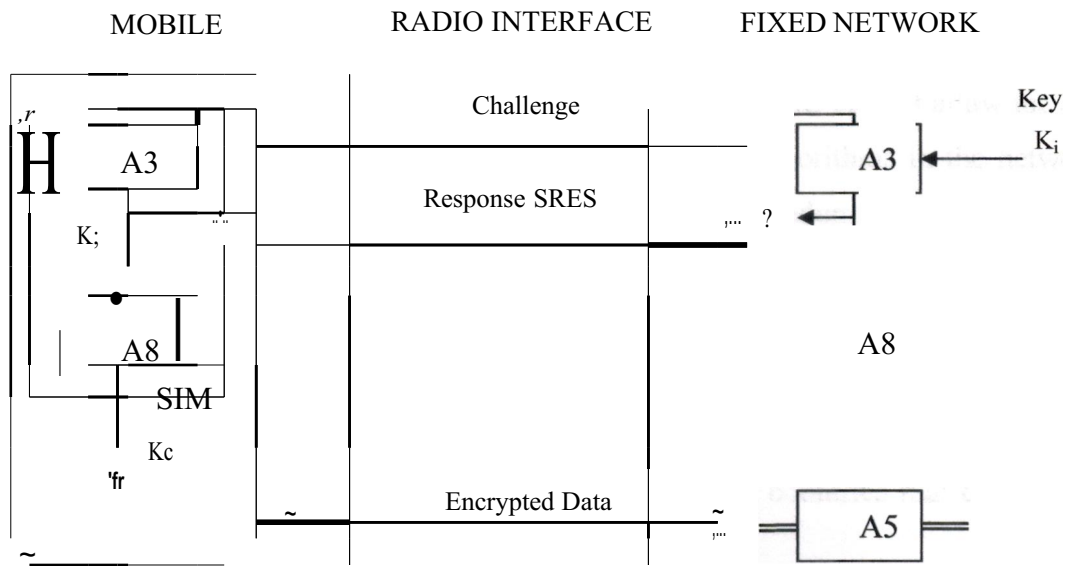


Figure 4.1 Encryption for GSM

Implementation and Roaming

The authentication algorithm A_3 is an operator option, and is implemented within the smart card (known as the Subscriber Interface Module or SIM). So that the operators may inter-work without revealing the authentication algorithms and mobile keys (K_i) to each other, GSM allows triplets of challenges (R), responses ($SRES$) and communication keys (K_c) to be sent between operators over the connecting networks.

The A_5 series algorithms are contained within the mobile equipment, as they have to be sufficiently fast and are therefore hardware. There are two defined algorithms used in GSM known as $A_5/1$ and $A_5/2$. The enhanced Phase 1 specifications developed by ETSI allows for inter-working between mobiles containing $A_5/1$, $A_5/2$ and unencrypted networks. These algorithms can all be built using a few thousand transistors, and usually takes a small area of a chip within the mobile.

4.5.4 World-Wide Use of the Algorithms

There are now three different possibilities for GSM, unencrypted, and use of the A5/1 algorithm or the A5/2 algorithm to secure the data. This arose because the GSM standard was designed for Western Europe, and export regulations did not allow the use of the original technology outside Europe. The uses of the algorithms in the network operator's infrastructure are controlled by the GSM Memorandum of Understanding Group (MoU) according to the formula below:

- The present A5/1 algorithm can be used by countries which are members of CEPT.
- The algorithm A5/2 is intended for any operators in countries that do not fall into the above category.

Export controls on mobiles are minimal, and the next generation of mobiles will support A5/1, A5/2 and no encryption. The protocols to support the various forms of AS (up to seven) are available in GSM.

Loss areas

There are a number of areas that can be exploited; the most likely intention of all the techniques is the ability to make money at the lowest cost possible.

Technical fraud

Technical fraud is where a weakness of the system is exploited to make free calls. For example, Call Forwarding or Conference Call facilities may be used to give reduced price services to customers from a stolen mobile. These are often known as 'Call Sales Offices'. Hackers and phreakers are often able to gain access and exploit a weakness in the switching or billing system and gain the ability to make calls or financial advantage. In some cases hackers and phreakers can take over the entire billing system and routing system; thus causing inconvenience for customers and carriers.

Procedural fraud

Procedural fraud results from the exploitation of business processes, where a flaw or weakness can be used to gain money. It may be possible for example to get free calls

from a stolen mobile, and sell the calls on for a lower cost than any legitimate network operator. This can be minimized by designing processes so that losses can be stopped by the use of correct and up to date policies, and by taking the opportunity to create a fraud away from the attacker or employee.

Comparison with other frauds

Many of the techniques that can be used to commit fraud on telecommunications networks can also be used for a mobile network. Analogue mobile phone systems (AMPS) were subject to being eavesdropped (with conventional RF-Scanners available at electronics shops and Radio Shack), and the phones could be cloned (ESN snarling over thin-air) so that bills were paid by the owner of the original mobile phone. Existing cellular systems have a number of potential weaknesses that were considered in the security requirements for GSM. Networks such as GSM, with international roaming and interactions with other operators (carriers), offer other opportunities for exploitation. GSM has been designed to offer various technical solutions to prevent misuse, such as strong authentication, together with anonymity and encryption of the signaling and data over the radio. However, all systems are dependent on secure management deployment and special procedures; lapses in these areas have severe impact on the resilience of the business process to fraud. For example; many carriers still make use of the COMP128 encryption algorithm for both A3 (the authentication algorithm to prevent phone cloning) and A8 (the voice-privacy key-generation algorithm), which is fine for securing against simple over-the-air attacks.

However we have determined that the COMP128's voice-encryption algorithms only encrypt voice between the GSM wireless phone and the base station. It does not encrypt voice within the phone network, nor does it encrypt end to end. It only encrypts the over-the-air portion of the transmission. The attack on COMP128 takes just 2A19 queries to the GSM smart-card chip, which takes approximately 8 hours over the air. This attack can be tested on as many simultaneous phones in radio range as your rogue base station has channels.

4.6 Design Goals

The security functionality within any system is a balance between the likelihood and impact of threats, user demand for certain security features and the cost and complexity of security measures. A security mechanism that is impervious to attack by any organization over any timescale would generally not be appropriate goals for a system transporting public, largely non-sensitive data. System designers must therefore set appropriate goals for the security of their system prior to beginning detailed design. SEG undertook this task and came up with the following simple goal for GSM security:

It would provide a degree of protection on the radio path which was approximately the same as that provided in the fixed network.

SEG were concerned with security on the radio interface only - there was no attempt to provide security on the fixed network part of GSM.

Before describing how this simple goal was translated into more formal security requirements, a few definitions are given:

Confidentiality is the property data has when it cannot be read by parties not authorized to read it. Confidentiality is provided by encryption in GSM.

Authentication of user identity is the property of establishing that the claimed identity of an entity really is their identity.

Integrity protection is the property of data whereby modification to the data can be detected. This is not explicitly provided by GSM but is provided implicitly by the use of ciphering along with the use of non-linear checksums (as stream ciphers are used in P-M, stream ciphering alone does not provide integrity protection).

The following requirements for the GSM security were developed over the course of the design exercise. These are listed in:

- Subscriber identity authentication. This protects the network from unauthorized use.
- Subscriber identity confidentiality. This provides protection against the tracing of a user's location by listening to exchanges on the radio interface.

- User data confidentiality across the radio interface. This protects the user's connection orientated data from eavesdropping on the radio interface.
- Connectionless user data confidentiality across the radio interface. This protects user information sent in connectionless packet mode in a signaling channel from eavesdropping on the radio interface.
- Signaling information element confidentiality across the radio interface. This protects selected fields in signaling messages from eavesdropping on the radio interface.

There was not general agreement on the issue of identity confidentiality within the group. Some members felt it was very important, particularly the German delegates. Others felt it was not a real requirement, and the subscriber must in some circumstances reveal their identity anyway, that the requirement could not be robustly met in any case.

There was also some debate during the design of GSM security as to whether user data should be given "privacy" or "confidentiality". "Privacy" was taken to mean protection from a determined "amateur" attacker but not necessarily a large organization - "confidentiality" was taken to mean protection from attack by the latter. The final conclusion was to try and provide confidentiality.

It should be noted that right from the start, there was concern within the group of providing too much security and thereby bringing unnecessary export problems upon GSM. The security was therefore designed with this constraint in mind, and also two further constraints:

- GSM did not have to be resistant to "active attacks" where the attacker interferes with the operation of the system, perhaps masquerading as a system entity. Active attacks are in contrast to passive attacks where the attacker merely monitors inter-system communication and does not interfere.
- The trust that must exist between operators for the operation of the security should be minimized.

4.7 Choosing the Security Architecture for GSM

Many people see security for communications systems as a matter of algorithms and attacks on the security of communications systems as a matter of attacks on algorithms. However, in designing security for a system, the choice of algorithms is often one of the last choices. The first task is to decide the goals of the security, which for GSM we have already talking about.

The second choice is determine the security protocols that will be used to achieve these goals. Usually, only after this point can the algorithms be decided. However, if the choice of algorithm is going to involve the choice between the use of secret or public key cryptography, then this basic choice must be made sooner, as it may dictate the whole security architecture.

The choice of public or secret key cryptography should still not be taken until the security goals have been decided though, in all the interesting debates about algorithm security, the designers may lose sight of the goals and why they are actually engaged in a design process at all. Having said all this, there were contributions at SEQ meetings which proposed particular architectures without any rationale or reference to claimed goals.

A security protocol is an interaction between two or more (but usually only two) entities following pre-determined steps that achieve some security goal. These protocols may involve proving that certain parties have certain items of secret information (this occurs during authentication or proof of claimed identity) and also may involve distribution or generation of secret keys for protection of communication. For instance, the widely known protocol, SSL achieves authentication of the server (generally a web server), generation of a shared secret key of protect the communication of data between the client (usually a browser on a PC) and the server, and the subsequent protection of that data for the duration of the SSL session. The emphasis on server authentication (though client authentication is possible) in SSL was provided so that users would have confidence in who they were sending data, e.g. credit card numbers, to. Client authentication is not mandatory in SSL as the use of the channel is generally "free" or the user has already been authenticated for charging purposes prior to the start of the SSL session (and as authentication is public key based in SSL, there is the complexity

of provisioning clients with key pairs and certificates). This emphasis on server authentication can be contrasted with the emphasis on client or user authentication in GSM, where the use of the channel is not free, and the user must therefore be authenticated so that they can be charged.

There are many ways of satisfying the goals for GSM security and the process of designing the GSM security architecture reflected the many possibilities open to the designers - many candidate architectures were proposed by the participating parties. BT, for instance, proposed the use of public key cryptography along with their own secret, symmetric encryption algorithm, BeCrypt. The reader might be interested to know why public Key Cryptography' (PKC) was not used. There were three main reasons:

- Implementations at the time were immature, the impact on the terminal for the provision of PKC functionality was therefore not accurately known;
- Messages would be longer, as PKC requires longer keys than symmetric cryptography, for provision of the same cryptographic strength;
- There was no real gain from the use of PKC. The authentication protocol runs between a subscriber and the network operator the subscriber has chosen to use. There is therefore a well established relationship and the one to many authentication possibility of PKC is not therefore required.

The large number of proposals caused problems for the group's progress, in that there were just too many protocols to examine properly. A security protocol should not be accepted until it has been examined thoroughly by a good number of experts. Flaws in communications systems security often occur because of a weakness in the protocols involved, and not in the strength of algorithms. However, these flaws are often subtle and take time and careful analysis to uncover. A small group therefore decided to trim the number of proposals down to a manageable level. This small group was, as is often the case in such situations, not elected or formally tasked with trimming the number; it was just a small number of people taking an initiative.

4.8 The Architecture Chosen

The architecture finally chosen was a simple and elegant one. It is based on secret key and not public key cryptography as stated above.

The architecture is centred on a long-term secret key, K_i , which is possessed by both the subscriber's mobile phone and subscriber's operator only. Authentication of the mobile phone by the network consists of proof by the mobile phone that it possesses the K_i . As part of this process cipher keys used for encryption during a call are also derived from K_i .

Before describing these operations in detail, a couple of design principles /constraints must be given. The first is that it was decided that the K_i must remain with the subscriber's home operator and must not be passed to other operator if the subscriber's home operator and must not be passed to another operator if the subscriber roams to that network. This is because the K_i is such a sensitive piece of information. With the K_i of a particular subscriber an impostor can pretend to be ("masquerade as") that subscriber and they can eavesdrop on all subscriber's calls. K_i should therefore not be revealed to more entities than is strictly necessary and SEG found a way for it only to be known by the minimum number of entities, the mobile phone (actually the "SIM", see below) and the home network. The second constraint is that long distance signalling should be minimized. It was therefore not acceptable that authentication process should involve the home operator for every call made by a roaming subscriber of that operator.

K_i should not every be known by the user themselves either, as this would allow the self-cloning of phone, and subsequent denial of the calls made by the cloned phone that had already occurred in analogue networks. A secure module within the phone that could be programmed by or under the control of the operator and in which K_i was stored and all operation involving K_i carried out was therefore required. A smart card was the obvious choice for such a security module, and the GSM Subscriber Identity Module (SIM) was born. The K_i is stored and used in the SIM and not in the terminal (or to use GSM terminology, the mobile equipment (ME)).

The security architecture is now described, in stages.

Ki in the home operator is held in the operator Authentication Center (AuC). (GSM phase 1 did allow the ki to be sent from the AuC to VLR for use there, but this way only allowed for VLRs in the same PLMN as the AuC, the specification advised against the option, and the option was dropped in GSM Phase2.) The AuC generates a random number, RAND for each subscriber. Random challenges are commonly used in security protocols to guarantee that a particular run of the protocol is "fresh" and entirely new and that an impostor who has captured some parameters from a previous run of the protocol cannot masquerade as the genuine subscriber or operator or interfere (either actively or passively) with the current run of the protocol. As shown in figure 4.2, for a particular subscriber, each RAND is passed as a parameter, along with the ki for that subscriber, through an algorithm named A3. A3 produces as an output, an expected response, XRES. The use of a challenge-response mechanism was not a proposal of a particular delegate in SEG. Once it was decided that a secret key mechanism would be, a challenge-response mechanism was the obvious choice.

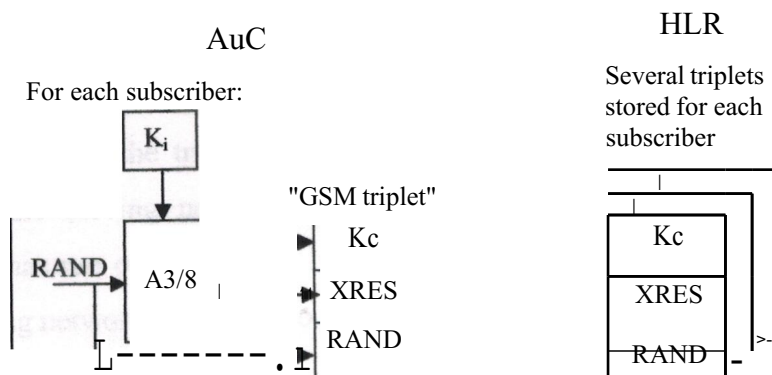


Figure 4.2 Diagram of Triplet Generation

RAND and k_i are also passed to another algorithm A8 which produces a cipher key, Kc. Typically, algorithms A3 and A8 are combined into one, called A3/8, and we shall consider them as such from now on. A RAND and the resulting XRES and Kc produced by A3/8 are called a "triplets". An AuC will normally produce a batch of triplets for a particular subscriber all at once and pass these for distribution to the HLR. This separation of triplet generation in the AuC from triplet distribution and subscriber management in the HLR, means that the AuC need only communicate with the HLR. In theory, therefore, greater access control can be placed on the AuC since it only ever

communicates with one, known, entity, the associated HLR of the same operator.

When a subscriber attempts to make a call or a location update in either its home operator's network, or in a network it has roamed to, the SIM passes its identity to the VLR serving that subscriber. The VLR makes a request to the subscriber's HLR for a batch of triplets for the identity claimed by the subscriber (i.e. the SIM) and the HLR responds with a batch of triplets for that claimed identity. The VLR authenticates the SIM by sending a RAND from the batch to the mobile phone, as shown in Figure 4.2. The ME passes RAND to the SIM where K_i and A3/8 are held. The SIM passes RAND and its K through algorithm(s) A3/8 residing within the SIM as was done in the AuC. The "signed response" produced by the SIM, SRES, is passed back to the VLR. The VLR compares SRES with the expected response, XRES, for that RAND, and if they match, the SIM/mobile phone's claimed identity is deemed to be authenticated. A3/8 in the SIM also produces K_c and if the SIM is authenticated, the VLR passes the K_c from the triplet to the Base Transceiver Station (BTS, the "abase station") serving the mobile. The SIM passes K_c to the ME and the BTS and mobile can then begin ciphering communications using K_c . The algorithm used for ciphering is termed *AS*.

With the use of the triplets, authentication can be performed in the serving network without the serving network operator having knowledge of K_i . When the serving network has run out of triplets, it should request more from the home operator (though the serving network is allowed to re-use triplets if it cannot obtain more).

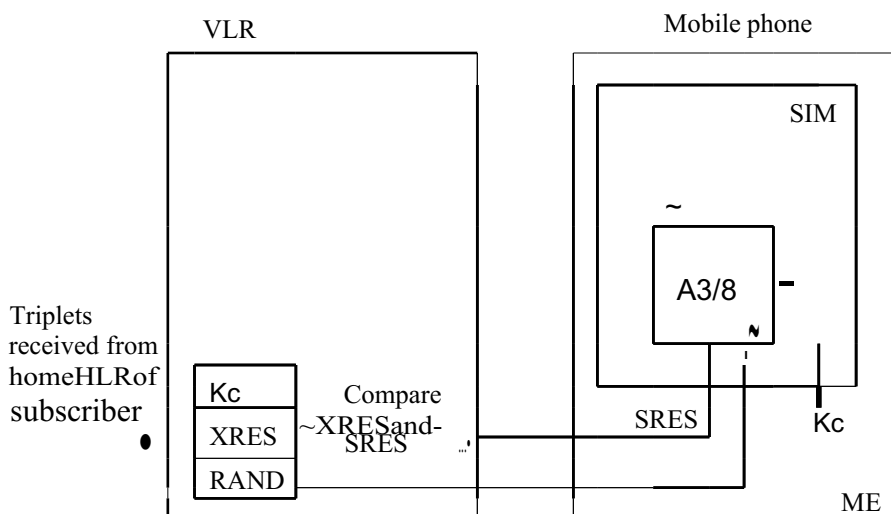


Figure 4.3 Diagram for authentication in the serving VLR

The inquisitive reader may be wondering when the use of algorithms A1-4, A6 and A7 is to be described. Their use, will not be described though, as they, and also algorithms A9 to A12 came up in initial versions of the architecture, but were not required in the final version. A3, A5 and A8 were not renamed A1 to A3. A4 and K4 are used by some operators to denote the encryption algorithm and key protecting the personalization data of a SIM (including the IMSI and Ki) between the personalization centre and the AuC, but they are not specified in any GSM specification.

The security architecture described above was specified in GSM specification 03.20

4.9 Authentication Algorithm Design

The effectiveness of authentication relies on a number of algorithm requirements not yet given. The first is that it is statistically near impossible for an impostor to guess what the correct SRES should be and therefore masquerade as another subscriber. As parameters SRES and RAND are 32 bits long, and the mobile has only one chance to return SRES for a particular RAND, provided that the algorithm has been so designed that SRES is indistinguishable from any other 32 bit number that might be returned instead of SRES, such an impostor has only a 1 in 2^{32} , or 1 in approximately 10 billion chance of guessing SRES correctly. This was felt sufficiently improbable as to not represent a realistic attack.

The second assumption is that, as RAND and SRES are passed un-encrypted between the mobile and the base station, an impostor cannot derive Ki from collecting a number of RAND-SRES pairs. This means that A3/8 must be designed to resist a known plaintext attack where the attacker knows what is ciphered as well as the ciphered result. Further, as an attacker could steal a SIM for some time, and send whatever challenges he liked to the SIM, and collect the SRESs given, A3/8 must be resistant to a chosen plaintext attack. This latter requirement was shown not to be satisfied by the algorithm COMP128, used as A3/8 by many operators.

A third requirement is that, again as RAND and SRES are passed un-encrypted between the mobile and the base station, an impostor cannot derive a particular Kc from the RAND and SRES in the same triplet as that Kc or by collecting a number of RAND-SRES pairs. This means that SRES and Kc must be completely unrelated though derived from the same RAND and Ki.

It has been mentioned that an important design consideration was that K_i was not to be shared with the serving network. A by-product of this decision is that algorithm A3/8 does not need to be known by the serving VLR, as A3/8 is only used where K_i is present, that is, in the AuC and the SIM. It should be noted that the VLR does not need any cryptographic algorithms, as A5 is not used in the VLR either but in the BTS, the security functionality in the VLR is therefore a simple comparison and distribution of parameters. This differs from systems based on ANSI-41, as used in many US networks, where the VLR must possess cryptographic capability. As A3/8 is only present in the SIM and AuC and the use of A3/8 is a protocol between a subscriber's SIM and the AuC of that subscriber's operator (albeit with the HLR and VLR as intermediaries), A3/8 does not have to be standardized. However, as the parameters of the triplet are passed via the HLR, VLR and ME as well as the SIM and AuC, the lengths of the parameters in the triplets must be standardized in the absence of a flexible encoding method. Each operator can therefore have a different A3/8 and operators were encouraged to take advantage of this possibility.

SEG felt it was an advantage that A3/8 did not need to be standardized. One claimed advantage of this is that less standardization work must be done. However, in response to this it could be said that now each operator must develop their own A3/8, so though the amount of standardization has gone down, the amount of development will go up. A second purported advantage is that each operator can also keep their A3/8 secret. However, this is also a moot point, because, as has been mentioned previously with regard to protocols, flaws in algorithms can be very subtle, and keeping an algorithm secret necessarily means there will be less potential examination of the algorithm. A clear advantage is that operators can gracefully bring in a new A3/8 on a SIM by SIM basis - the AuC knows which subscriber a request for triplets is for and can therefore use an updated A3/8. A clear disadvantage of there being different A3/8 is that AuC manufacturers must cope with different requirements from different operators.

However, in spite of the arguments for and against standardization of A3/8, it was recognized that an example algorithm would be required, for implementation tests, and for those operators that did not possess or wish to possess the capability to obtain such an algorithm.

This algorithm was COMP128, designed by a research wing of Deutsches Telecom. The use of COMP128 amply illustrates the disadvantages mentioned above.

4.10 Other GSM Security Mechanisms

4.10.1 SIM Card

There is always the possibility (we have no idea) that the SIM card can be compromised. This is considered unlikely. esocci:lh" as some use them version of A3. Keys Ki and the matching Th ISI could be selling the information for money.

4.10.2 IMEI

In GSM the customer subscription and authentication capability is contained within a smart card (SIM, Subscriber Identity Module). Any mobile will take on the identity of a subscriber by insertion of a smart card. The mobiles now become attractive items to steal, as they can be used with another SIM card.

To prevent this, GSM has specified an International Mobile Equipment Identifier (IMEI). Although to an operator, at first evaluation, it may seem as the stolen mobiles have no effect, as they do not affect a subscription, there will be problems with an increase in customer facing staff to handle esquires, and a possibility that GSM handsets are expensive to insure.

An Equipment Identity Register (EIR) exists in each network, with Black, White and Grey Lists for stolen or non type approved mobiles, valid mobiles and mobiles that need tracking respectively. Grey lists are for local tracking of mobiles within a network.

GSM has defined a procedure so that approved, lost or stolen mobile IMEIs can be communicated to all other operators. A Central Equipment Identity Register has been (CEIR) proposed. Type approval authorities' issue white list numbers (random ranges of valid IMEIs) to mobile manufacturers and manufacturers inform the CEIR when the mobiles are released to market. All operators are able to post their black lists to the CEIR, and in return collect a consolidated list of all operators black and white lists. By this method stolen or invalid mobiles can be quickly barred throughout the world.

4.10.3 Roaming

International roaming problems are minimised by the use of two procedures:

- Rapid exchange of billing information by means of EDI
- Notification of the home network of the visitor when the visitor has exceeded a certain billing limit.

4.11 Problems with GSM Security

4.11.1 The Limitation and Problems with GSM Security

Problems with GSM security

- Security by obscurity, which means that all of the algorithms used are not available to the public. Most security analysts believe any system that is not subject to the scrutiny of the world's best minds can't be as secure.
- Only provides access security. All communication between the Mobile Station and the Base Transceiver Station are encrypted. But all communications and signalling is generally transmitted in plain text in the fixed network.
- Difficult to upgrade the cryptographic mechanisms
- Lack of user visibility (e.g. doesn't know if encrypted or not)
- The flaw of the algorithms.

4.11.2 Possible Improvement

Security could be improved in some areas with relatively simple measures. One solution is to use another cryptographically secure algorithm for A3. This would require issuing new SIM-cards to all subscribers and updating HLR software. This would effectively disable the attacker from cloning SIM-cards, the most dangerous attack, which is discussed above. This solution is easy to be implemented because the network operators can make the changes themselves and do not need the support of hardware or software manufacturers or the GSM Consortium. There is now a new algorithms available called COMP128-2.

The operator can employ a new *A5* implementation with strong encryption too. A new A5/3 algorithm has also been agreed upon to replace the aging A5/2 algorithm. This improvement would require the co-operation of the hardware and software manufacturers because they will have to release new versions of their software and hardware that would comprise with the new algorithm.

Third solution would be to encrypt the traffic on the operator's backbone network between the network components. This would disable the attacker from wiretapping the backbone network.

This solution could probably also be implemented without the blessings of the GSM Consortium, but the co-operation of the hardware manufacturers would still be required.

CONCLUSION

GSM security is simple and elegant and, unlike many security protocols in consumer equipment, has not had inter-operability problems. Though weaknesses have been found in GSM security, it must be recalled that GSM was designed as a piece of consumer equipment and these weaknesses are sophisticated and will not cause ordinary "consumers" to be at risk. Further there was severe restriction on the use of cryptography at the time GSM was designed.

Overall then it can be said that GSM security has been a success.

The security in GSM is modelled around anonymity, confidentiality, and authentication.

Three proprietary algorithms are used to achieve authentication and confidentiality. The first algorithm A3 is used to encode a random challenge issued by the network to the user. The second algorithm A8 takes the output of the first algorithm and creates a key, which will be used by another algorithm A5 in the encryption process. Anonymity is achieved by using temporary identifiers when the user logs onto the system.

The weaknesses of the GSM system include:

- The use of Radio as a transmission media is inherently flawed.
- The GSM algorithms have not been made available to cryptanalysts, and as such their strength is unknown.
- The key used for encryption is only 40bits long.
- The authentication process does reveal a users identity and location.

Security in GSM is only provided over the radio link, whilst the security needs in the fixed network is ignored. As such GSM can only hope to be as secure or worse than today's PSTN's (Public Switched Telephone Networks). Members of the public should use discretion when transmitting sensitive data over the GSM network.

REFERENCE

- [1]. Bellcore Technical Advisories, "Generic Framework Criteria for Universal Digital Personal Communications Systems (PCS)," FA-TSY-001013, Issue 1 (March 1990), and FA-NWT-001013, Issue 2 (December 1990); "Generic Criteria for Version 0.1 Wireless Access Communications Systems (WACS)," TA-NTWT-001313, Issue 1 (July 1992).

- [2]. Cheung et al., "Network Planning for Third-Generation Mobile Radio Systems," *IEEE Communications Magazine* 32, no. 11 (November 1994): 54-69.

- [3]. Chia, "Beyond the Second-Generation Mobile Radio Systems," *British Telecom Engineering* 10 (January 1992): 326-335.

- [4]. Üx., D. C., "'Perncma\ Comm.unicatim.'-1-1\ \Tie'NI)omt," *IEEE Communication Magazine* (November 1990): 8-20.

- [5]. European Telecommunications Standards Institution (ETSI i, "Recommendations for G5M900/DC51800" (ETSI, Cedex, France).

- [6]. Gardiner, J. D. "Second Generation Cordless Telephony in U.K: Telepoint Services and the Common Air Interface," *IEE Electronics and Communications Engineering Journal* (April 1992).

- [7]. Gilhousen, K. S., et al., "On the Capacity of a Cellular CDMA system." *IEEE transaction on Velzicular technology* VT-40, no. 2 (May 1991): 303-12.

- [8]. Goodman, D. J.. "Trends in Cellular and Cordless Communications; " *IEEE Communiications JUagatine.*" (June 1991): 31-40.

- [9]. Grillo European Perspective on Third Generation Personal Communication Systems," Proceedings of the IEEE VTC Conference. Orlando, Florida. May 199~.

- [10]. Mehro: A Cellular Radio Using 4 Digital Systems, Boston-London: Artech House. 199

- [11]. Owen and C. Pudney, "DECT'-Integrated services for cordless communications," Proceedings of the Fifth International Conference on Mobile Radio and Personal Communications. Institution of Electrical Engineers, Warwick. United Kingdom. December 1989.
- [12]. Tuttle bee, *Cordless Telecommunication in Europe*, London-Heidelberg-NEW York: Springer-Verlag, 1990.
- [13]. Viterbi. A. J., and Roberto Padovani, "Implications of Mobile Cellular CDZ'IA." *IEEE communication magazine* (December 1992): 30 no. 12 3841.
- [14]. Digital cellular telecommunication system (phase 2); security aspects (GSM 02.09 version 4.5.1), ETS 300 506.
- [15]. Digital cellular telecommunication system (phase 2); security related network functions (GSM 03.20 version 4.4.1), ETS 300 534.
- [16]. Golie J.Dj. Cryptanalysis of alleged AS stream cipher. Euro crypt 97, published as lecture note in computer science # 1233., Springer Verlag, Berlin, 1997.
- [17]. Biryukov A., Shamir A. Real time cryptanalysis of the alleged AS/1 on a PC, preliminary draft, December 1999.
- [18]. Babbage S.H. improved exhaustive search attacks on stream ciphers.ECOS95 (European Convention on Security and Detection), {EE Conference Publication no.408, May 1995.
- [19]. Biryukov A., Shamir A. ,Wagner !D. Real time cryp:tarra:lysis of :a5/1 on a PC .FSE200 .puplished as lecture notes in Computer science #1978, Spring Verlag, Berlin, 2000.
- [20]. Piper F.C., Walker M. cryptographic solution for voice telephony and GSM .In: proc. COMPSEC '98. London: Elsevier, 1'998.