NEAR EAST UNIVERSITY

Faculty of Engineering

Department of Electrical and Electronic Engineering

AUTHENTICATION AND MESSAGE SECURITY (GSM)

Graduation Project
EE- 400

Student: Suleman Bakhsh (981243)

Supervisor: Prof. Dr. Fakhreddin Mamedov

Lefkoşa – 2002

TABLE OF CONTENTS

ACKNOWLEDGMENT

This project is dedicated to Allah Subhanahu wa Ta`aalaa whose guidance, help and grace was instrumental in making this humble work a reality. I have many people to thank for their assistance.

First of all I would like to thank Prof. F.M. for his support and guidance as being my advisor through my project. Because of his patience and kindness I was able to present my project as a completion of a four month work. He always welcomes any question or enquiry. He was also generous with his time and provided me with critical comments and expert advice.

Great thanks go to my family, who were there for me every time and were praying for me day and night. They paved my way to achieve this position.

Finally, I want to thank all my dear friends for everything they did for me, from the first day in the university till now. I really enjoyed being with them and I wish them the best in their future.

One thing to remember here is that "there is no powerful magic than the power of hard working".

LIST OF ABBREVIATIONS

A3	The authentication algorithm used in the GSM system.
A5	The encryption algorithm used in the GSM system.
A8	The key generation algorithm used in the GSM system.
AuC	Authentication Center.
BSC	Base Station Controller.
BSS	Base Station Subsystem. The BSS can be divided in two parts: • The Base Transceiver Station (BTS) or Base Station. • The Base Station Controller (BSC).
BTS	Base Transceiver Station.
COMP128	A one-way function that is currently used in most GSM networks for A3 and A8.
GPRS	General Packet Radio Service.
GSM	Global System for Mobile communications.
HLR	Home Location Register.
Kc	The secret session key used to encrypt over-the-air traffic between the BTS and the MS.
Ki	Ki is the secret key shared between the SIM and the HLR of the subscriber's home network.
LSB	Least Significant Bit.
LSFR	Linear Shift Feedback Register.
MS	Mobile Station, the mobile phone.
MSC	Mobile services Switching Center.

NSS	Network and Switching Subsystem.
SGSN	Serving GPRS Support Node.
SIM	Subscriber Identity Module.
SRES	Signed RESponse.
VLR	Visitor Location Register.

ABSTRACT

The radio communications aspect of the GSM system makes it particularly sensitive to unauthorized use.

For this reason, security mechanisms are defined for the GSM system:

- Subscriber identity (IMSI) confidentiality.
- Subscriber identity (IMSI) authentication.
- Data confidentiality over the air interface.
- Mobile equipment security.

A number of security parameters have been defined in the core specifications to support these security features. The IMSI is used to uniquely identify subscribers and the TMSI to provide subscriber identity confidentiality. The authentication vectors (Kc,RAND,SRES) are used in the authentication process and the ciphering key (Kc) is used to encrypt signaling and user data over the air interface. Finally the IMEI can be used to establish whether a piece of mobile equipment is suitable to be used on the network, i.e., approved and neither stolen nor faulty.

INTRODUCTION

GSM is the most widely used cellular mobile phone system in the world with over 100 million GSM subscribers. GSM was one of the first digital mobile phone systems to follow the analog era. Widely known problems with GSM's analog counter parts were the possibility of phone fraud through cloning phones and thus calling in someone else's expense, and the possibility of someone intercepting the phone call over the air and eavesdropping on the discussion. The GSM system was supposed to correct these problems by implementing strong authentication between the MS and the MSC, as well as implementing strong data encryption for the over-the-air transmission channel between the MS and the BTS.

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 GSM Consortium relied on Security by Obscurity, i.e. the algorithms would be harder to crack if they were not publicly available. According to the open scientific community, one of the basic requirements for secure cryptographic algorithms is that the security of the crypto system lies solely on the key. This is known as Kerckhoffs' assumption. 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 cryptanalysing 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.

The rest of the projcet is organized as follows:

- Chapter 1 is an overview of GSM system including all its features in general.
- Chapter 2 introduces the GSM security model.

- Chapter 3 introduces the security related services and functions in GSM.
- Chapter 4 compares the GPRS security model to the GSM security model. And gives some suggestions about possible improvements to the GSM security model.

1. OVERVEIW OF GSM SYSTEM

The Global System for Mobile communications is a digital cellular communications system. It was developed in order to create a common European mobile telephone standard but it has been rapidly accepted worldwide. GSM was designed to be compatible with ISDN services.

1.1 History of the cellular mobile radio and GSM

The idea of cell-based mobile radio systems appeared at Bell Laboratories (in USA) in the early 1970s. However, mobile cellular systems were not introduced for commercial use until the 1980s. During the early 1980s, analog cellular telephone systems experienced a very rapid growth in Europe, particularly in Scandinavia and the United Kingdom. Today cellular systems still represent one of the fastest growing telecommunications systems.

But in the beginnings of cellular systems, each country developed its own system, which was an undesirable situation for the following reasons.

- The equipment was limited to operate only within the boundaries of each country.
- The market for each mobile equipment was limited.

In order to overcome these problems, the Conference of European Posts and Telecommunications (CEPT) formed, in 1982, the Groupe Spécial Mobile (GSM) in order to develop a pan-European mobile cellular radio system (the GSM acronym became later the acronym for Global System for Mobile communications). The standardized system had to meet certain criteria:

- Spectrum efficiency
- International roaming
- Low mobile and base stations costs
- Good subjective voice quality
- Compatibility with other systems such as ISDN (Integrated Services Digital Network) Ability to support new services

Unlike the existing cellular systems, which were developed using an analog technology, the GSM system was developed using a digital technology.

In 1989 the responsibility for the GSM specifications passed from the CEPT to the European Telecommunications Standards Institute (ETSI). The aim of the GSM specifications is to describe the functionality and the interface for each component of the system, and to provide guidance on the design of the system. These specifications will then standardize the system in order to guarantee the proper interworking between the different elements of the GSM system. The most important events in the development of the GSM system are presented in the table 1.1

Table 1.1 Events in the development of GSM

Year	Events
1982	CEPT establishes a GSM group in order to develop the standards for a pan-European cellular mobile system
1985	Adoption of a list of recommendations to be generated by the group
1986	Field tests were performed in order to test the different radio techniques proposed for the air interface
1987	TDMA is chosen as access method (in fact, it will be used with FDMA) Initial Memorandum of Understanding (MoU) signed by telecommunication operators (representing 12 countries)
1988	Validation of the GSM system
1989	The responsability of the GSM specifications is passed to the ETSI
1990	Appearance of the phase 1 of the GSM specifications
1991	Commercial launch of the GSM service
1992	Enlargement of the countries that signed the GSM- MoU> Coverage of larger cities/airports
1993	Coverage of main roads GSM services start outside Europe
1995	Phase 2 of the GSM specifications Coverage of rural areas

From the evolution of GSM, it is clear that GSM is not anymore only a European standard. GSM networks are operational or planned in over 80 countries around the world. The rapid and increasing acceptance of the GSM system is illustrated with the following figures:

- 1.3 million GSM subscribers worldwide in the beginning of 1994.
- Over 5 million GSM subscribers worldwide in the beginning of 1995.
- Over 10 million GSM subscribers only in Europe by December 1995.

Since the appearance of GSM, other digital mobile systems have been developed. The table 1.2 charts the different mobile cellular systems developed since the commercial launch of cellular systems.

1.2 Cellular systems

1.2.1 The cellular structure

In a cellular system, the covering area of an operator is divided into cells. A cell corresponds to the covering area of one transmitter or a small collection of transmitters. The size of a cell is determined by the transmitter's power.

The concept of cellular systems is the use of low power transmitters in order to enable the efficient reuse of the frequencies. In fact, if the transmitters used are very powerful, the frequencies can not be reused for hundred of kilometers as they are limited to the covering area of the transmitter.

The frequency band allocated to a cellular mobile radio system is distributed over a group of cells and this distribution is repeated in all the covering area of an operator. The whole number of radio channels available can then be used in each group of cells that form the covering area of an operator. Frequencies used in a cell will be reused several cells away. The distance between the cells using the same frequency must be sufficient to avoid interference. The frequency reuse will increase considerably the capacity in number of users.

Table 1.2 Mobile cellular systems

Year	Mobile Cellular System
1981	Nordic Mobile Telephony (NMT), 450>
1983	American Mobile Phone System (AMPS)
1985	Total Access Communication System (TACS) Radiocom 2000 C-Netz
1986	Nordic Mobile Telephony (NMT), 900>
1991	Global System for Mobile communications> North American Digital Cellular (NADC)
1992	Digital Cellular System (DCS) 1800
1994	Personal Digital Cellular (PDC) or Japanese Digital Cellular (JDC)
1995	Personal Communications Systems (PCS) 1900-Canada>
1996	PCS-United States of America>

In order to work properly, a cellular system must verify the following two main conditions:

- The power level of a transmitter within a single cell must be limited in order to reduce the interference with the transmitters of neighboring cells. The interference will not produce any damage to the system if a distance of about 2.5 to 3 times the diameter of a cell is reserved between transmitters.
- Neighboring cells can not share the same channels. In order to reduce the interference, the frequencies must be reused only within a certain pattern.

In order to exchange the information needed to maintain the communication links within the cellular network, several radio channels are reserved for the signaling information.

1.2.2 Cluster

The cells are grouped into clusters. The number of cells in a cluster must be determined so that the cluster can be repeated continuously within the covering area of an operator. The typical clusters contain 4, 7, 12 or 21 cells. The number of cells in each cluster is very important. The smaller the number of cells per cluster is, the bigger the number of channels per cell will be. The capacity of each cell will be therefore increased. However a balance must be found in order to avoid the interference that could occur between neighboring clusters. This interference is produced by the small size of the clusters (the size of the cluster is defined by the number of cells per cluster). The total number of channels per cell depends on the number of available channels and the type of cluster used.

1.2.3 Types of cells

The density of population in a country is so varied that different types of cells are used:

- Macrocells
- Microcells
- Selective cells
- Umbrella cells

- **Macrocells**

The macrocells are large cells for remote and sparsely populated areas.

- **Microcells**

These cells are used for densely populated areas. By splitting the existing areas into smaller cells, the number of channels available is increased as well as the capacity of the cells. The power level of the transmitters used in these cells is then decreased, reducing the possibility of interference between neighboring cells.

- **Selective cells**

It is not always useful to define a cell with a full coverage of 360 degrees. In some cases, cells with a particular shape and coverage are needed. These cells are called

selective cells. A typical example of selective cells is the cells that may be located at the entrances of tunnels where coverage of 360 degrees is not needed. In this case, a selective cell with coverage of 120 degrees is used.

- **Umbrella cells**

A freeway crossing very small cells produces an important number of handovers among the different small neighboring cells. In order to solve this problem, the concept of umbrella cells is introduced. An umbrella cell covers several microcells. The power level inside an umbrella cell is increased comparing to the power levels used in the microcells that form the umbrella cell. When the speed of the mobile is too high, the mobile is handed off to the umbrella cell. The mobile will then stay longer in the same cell (in this case the umbrella cell). This will reduce the number of handovers and the work of the network. The number of handover demands and the propagation characteristics of a mobile can help to detect its high speed.

1.3 The transition from analog to digital technology

In the 1980s most mobile cellular systems were based on analog systems. The GSM system can be considered as the first digital cellular system. The different reasons that explain this transition from analog to digital technology are presented below:

1.3.1 The capacity of the system

Cellular systems have experienced a very important growth and Analog systems were not able to cope with this increasing demand. In order to overcome this problem, new frequency bands and new technologies were proposed. But the possibility of using new frequency bands was rejected by a big number of countries because of the restricted spectrum (even if later on, other frequency bands have been allocated for the development of mobile cellular radio). The new analog technologies proposed were able to overcome the problem to a certain degree but the costs were too important.

The digital radio was, therefore, the best option (but not the perfect one) to handle the capacity needs in a cost-efficiency way.

1.3.2 Compatibility with other systems such as ISDN

The decision of adopting a digital technology for GSM was made in the course of developing the standard. During the development of GSM, the telecommunications industry converted to digital methods. The ISDN network is an example of this evolution. In order to make GSM compatible with the services offered by ISDN, it was decided that the digital technology was the best option.

Additionally, a digital system allows, easily than an analog one, the implementation of future improvements and the change of its own characteristics.

1.3.3 Aspects of quality

The quality of the service can be considerably improved using a digital technology rather than an analog one. In fact, analog systems pass the physical disturbances in radio transmission (such as fades, multipath reception, spurious signals or interferences) to the receiver. These disturbances decrease the quality of the communication because they produce effects such as fadeouts, crosstalk, hisses, etc. On the other hand, digital systems avoid these effects transforming the signal into bits. This transformation combined with other techniques, such as digital coding, improves the quality of the transmission. The improvement of digital systems comparing to analog systems is more noticeable under difficult reception conditions than under good reception conditions.

1.4 The GSM network

1.4.1 Architecture of the GSM network

The GSM network can be divided into four main parts:

- The Mobile Station (MS).
- The Base Station Subsystem (BSS).
- The Network and Switching Subsystem (NSS).
- The Operation and Support Subsystem (OSS).

The architecture of the GSM network is presented in figure 1.1

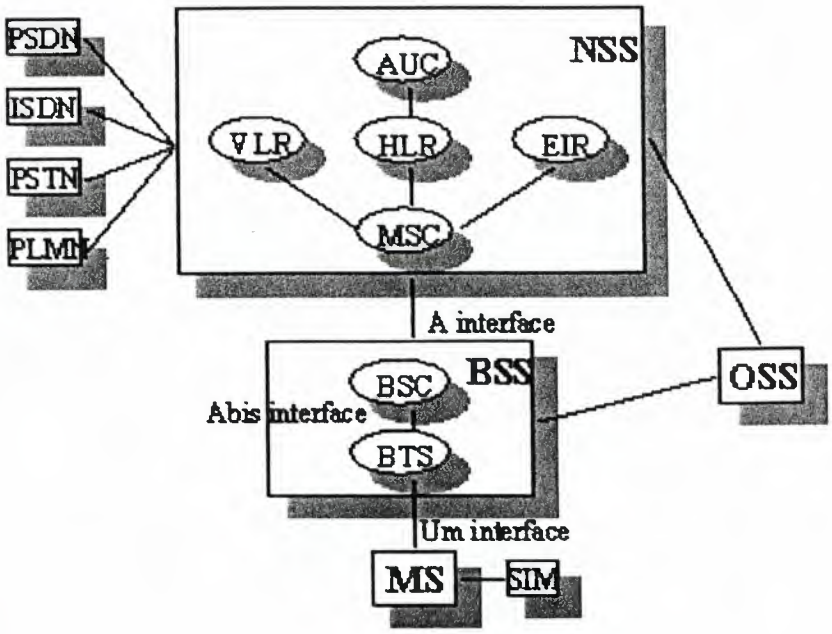

Figure 1.1 Architecture of the GSM network

- **Mobile Station**

A Mobile Station consists of two main elements:

- The mobile equipment or terminal.
- The Subscriber Identity Module (SIM).

1) The Terminal

There are different types of terminals distinguished principally by their power and application:

- The `fixed' terminals are the ones installed in cars. Their maximum allowed output power is 20 W.
- The GSM portable terminals can also be installed in vehicles. Their maximum allowed output power is 8W.
- The handhelds terminals have experienced the biggest success thanks to their weight and volume, which are continuously decreasing. These terminals can emit up to 2 W. The evolution of technologies allows decreasing the maximum allowed power to 0.8 W.

2) The SIM

The SIM is a smart card that identifies the terminal. By inserting the SIM card into the terminal, the user can have access to all the subscribed services. Without the SIM card, the terminal is not operational.

The SIM card is protected by a four-digit Personal Identification Number (PIN). In order to identify the subscriber to the system, the SIM card contains some parameters of the user such as its International Mobile Subscriber Identity (IMSI).

Another advantage of the SIM card is the mobility of the users. In fact, the only element that personalizes a terminal is the SIM card. Therefore, the user can have access to its subscribed services in any terminal using its SIM card.

- **The Base Station Subsystem**

The BSS connects the Mobile Station and the NSS. It is in charge of the transmission and reception. The BSS can be divided into two parts:

- The Base Transceiver Station (BTS) or Base Station.
- The Base Station Controller (BSC).

1) The Base Transceiver Station

The BTS corresponds to the transceivers and antennas used in each cell of the network. A BTS is usually placed in the center of a cell.

Its transmitting power defines the size of a cell. Each BTS has between one and sixteen transceivers depending on the density of users in the cell.

2) The Base Station Controller

The BSC controls a group of BTS and manages their radio resources.

A BSC is principally in charge of handovers, frequency hopping, exchange functions and control of the radio frequency power levels of the BTSs.

- **The Network and Switching Subsystem**

Its main role is to manage the communications between the mobile users and other users, such as mobile users, ISDN users, fixed telephony users, etc. It also includes data bases needed in order to store information about the subscribers and to manage their mobility. The different components of the NSS are described below.

1) The Mobile services Switching Center (MSC)

It is the central component of the NSS. The MSC performs the switching functions of the network. It also provides connection to other networks.

2) The Gateway Mobile services Switching Center (GMSC)

A gateway is a node interconnecting two networks. The GMSC is the interface between the mobile cellular network and the PSTN. It is in charge of routing calls from the fixed network towards a GSM user. The GMSC is often implemented in the same machines as the MSC.

3) Home Location Register (HLR)

The HLR is considered as a very important database that stores information of the subscribers belonging to the covering area of a MSC. It also stores the current location of these subscribers and the services to which they have access. The location of the subscriber corresponds to the SS7 address of the Visitor Location Register (VLR) associated to the terminal.

4) Visitor Location Register (VLR)

The VLR contains information from a subscriber's HLR necessary in order to provide the subscribed services to visiting users. When a subscriber enters the covering area of a new MSC, the VLR associated to this MSC will request information about the new subscriber to its corresponding HLR. The VLR will then have enough information in order to assure the subscribed services without needing to ask the HLR each time a communication is established. The VLR is always implemented together with a MSC; so the area under control of the MSC is also the area under control of the VLR.

5) **The Authentication Center (AuC)**

The AuC register is used for security purposes. It provides the parameters needed for authentication and encryption functions. These parameters help to verify the user's identity.

6) **The Equipment Identity Register (EIR)**

The EIR is also used for security purposes. It is a register containing information about the mobile equipments. More particularly, it contains a list of all valid terminals. A terminal is identified by its International Mobile Equipment Identity (IMEI). The EIR allows then to forbid calls from stolen or unauthorized terminals.

7) **The GSM Interworking Unit (GIWU)**

The GIWU corresponds to an interface to various networks for data communications. During these communications, the transmission of speech and data can be alternated.

- **The Operation and Support Subsystem (OSS)**

The OSS is connected to the different components of the NSS and to the BSC, in order to control and monitor the GSM system. It is also in charge of controlling the traffic load of the BSS.

However, the increasing number of base stations, due to the development of cellular radio networks, has provoked that some of the maintenance tasks are transferred to the BTS. This transfer decreases considerably the costs of the maintenance of the system.

1.4.2 The geographical areas of the GSM network

A cell, identified by its Cell Global Identity number (CGI), corresponds to the radio coverage of a base transceiver station. A Location Area (LA), identified by its Location Area Identity (LAI) number, is a group of cells served by a single MSC/VLR. A group of location areas under the control of the same MSC/VLR defines the

MSC/VLR area. A Public Land Mobile Network (PLMN) is the area served by one network operator. The figure1.2 presents the different areas that form a GSM network.

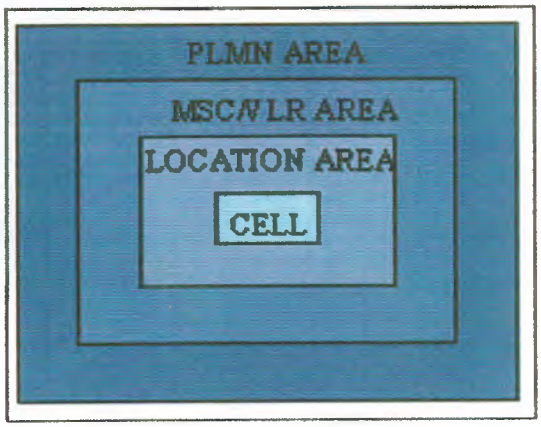

Figure 1.2 GSM network areas

1.4.3 The GSM functions

In this paragraph, the description of the GSM network is focused on the different functions to fulfill by the network and not on its physical components. In GSM, five main functions can be defined:

- Transmission.
- Radio Resources management (RR).
- Mobility Management (MM).
- Communication Management (CM).
- Operation, Administration and Maintenance (OAM).

1) Transmission

The transmission function includes two sub-functions:

- The first one is related to the means needed for the transmission of user information.
- The second one is related to the means needed for the transmission of signaling information.

Not all the components of the GSM network are strongly related with the transmission functions. The MS, the BTS and the BSC, among others, are deeply concerned with transmission. But other components, such as the registers HLR, VLR or EIR, are only concerned with the transmission for their signaling needs with other components of the GSM network.

2) Radio Resources management (RR)

The role of the RR function is to establish, maintain and release communication links between mobile stations and the MSC. The elements that are mainly concerned with the RR function are the mobile station and the base station. However, as the RR function is also in charge of maintaining a connection even if the user moves from one cell to another, the MSC, in charge of handovers, is also concerned with the RR functions.

The RR is also responsible for the management of the frequency spectrum and the reaction of the network to changing radio environment conditions. Some of the main RR procedures that assure its responsibilities are:

- Channel assignment, change and release.
- Handover.
- Frequency hopping.
- Power-level control.
- Discontinuous transmission and reception.
- Timing advance.

- **Handover**

The user movements can produce the need to change the channel or cell, especially when the quality of the communication is decreasing. This procedure of changing the resources is called handover. Four different types of handovers can be distinguished:

- Handover of channels in the same cell.
- Handover of cells controlled by the same BSC.
- Handover of cells belonging to the same MSC but controlled by different BSCs.

- Handover of cells controlled by different MSCs.

Handovers are mainly controlled by the MSC. However in order to avoid unnecessary signaling information, the first two types of handovers are managed by the concerned BSC (in this case, the MSC is only notified of the handover).

The mobile station is the active participant in this procedure. In order to perform the handover, the mobile station controls continuously its own signal strength and the signal strength of the neighboring cells. The list of cells that must be monitored by the mobile station is given by the base station. The power measurements allow deciding which the best cell is in order to maintain the quality of the communication link. Two basic algorithms are used for the handover:

- The `minimum acceptable performance' algorithm. When the quality of the transmission decreases (i.e. the signal is deteriorated), the power level of the mobile is increased. This is done until the increase of the power level has no effect on the quality of the signal. When this happens, a handover is performed.
- The `power budget' algorithm. This algorithm performs a handover, instead of continuously increasing the power level, in order to obtain a good communication quality.

3) Mobility Management

The MM function is in charge of all the aspects related with the mobility of the user, specially the location management and the authentication and security.

- **Location management**

When a mobile station is powered on, it performs a location update procedure by indicating its IMSI to the network. The first location update procedure is called the IMSI attach procedure.

The mobile station also performs location updating, in order to indicate its current location, when it moves to a new Location Area or a different PLMN. This location updating message is sent to the new MSC/VLR, which gives the location information to the subscriber's HLR. If the mobile station is authorized in the new

MSC/VLR, the subscriber's HLR cancels the registration of the mobile station with the old MSC/VLR.

A location updating is also performed periodically. If after the updating time period, the mobile station has not registered, it is then deregistered. When a mobile station is powered off, it performs an IMSI detach procedure in order to tell the network that it is no longer connected.

- **Authentication and security**

The authentication procedure involves the SIM card and the Authentication Center. A secret key, stored in the SIM card and the AuC, and a ciphering algorithm called A3 are used in order to verify the authenticity of the user. The mobile station and the AuC compute a SRES using the secret key, the algorithm A3 and a random number generated by the AuC. If the two computed SRES are the same, the subscriber is authenticated. The different services to which the subscriber has access are also checked.

Another security procedure is to check the equipment identity. If the IMEI number of the mobile is authorized in the EIR, the mobile station is allowed to connect the network.

In order to assure user confidentiality, the user is registered with a Temporary Mobile Subscriber Identity (TMSI) after its first location update procedure.

4) Communication Management (CM)

The CM function is responsible for:

- Call control.
- Supplementary Services management.
- Short Message Services management.

a) Call Control (CC)

The CC is responsible for call establishing, maintaining and releasing as well as for selecting the type of service. One of the most important functions of the CC is the

call routing. In order to reach a mobile subscriber, a user dials the Mobile Subscriber ISDN (MSISDN) number which includes:

- a country code
- a national destination code identifying the subscriber's operator
- a code corresponding to the subscriber's HLR

The call is then passed to the GMSC (if the call is originated from a fixed network) which knows the HLR corresponding to a certain MISDN number. The GMSC asks the HLR for information helping to the call routing. The HLR requests this information from the subscriber's current VLR. This VLR allocates temporarily a Mobile Station Roaming Number (MSRN) for the call. The MSRN number is the information returned by the HLR to the GMSC. Thanks to the MSRN number, the call is routed to subscriber's current MSC/VLR. In the subscriber's current LA, the mobile is paged.

b) Supplementary Services management

The mobile station and the HLR are the only components of the GSM network involved with this function. The different Supplementary Services (SS) to which the users have access are presented later on this paper.

c) Short Message Services management

In order to support these services, a GSM network is in contact with a Short Message Service Center through the two following interfaces:

- The SMS-GMSC for Mobile Terminating Short Messages (SMS-MT/PP). It has the same role as the GMSC.
- The SMS-IWMSC for Mobile Originating Short Messages (SMS-MO/PP).

5) Operation, Administration and Maintenance (OAM)

The OAM function allows the operator to monitor and control the system as well as to modify the configuration of the elements of the system. Not only the OSS is part of the OAM, also the BSS and NSS participate in its functions as it is shown in the following examples:

- The components of the BSS and NSS provide the operator with all the information it needs. This information is then passed to the OSS which is in charge of analyzing it and controlling the network.
- The self test tasks, usually incorporated in the components of the BSS and NSS, also contribute to the OAM functions.
- The BSC, in charge of controlling several BTSs, is another example of an OAM function performed outside the OSS.

1.5 The GSM radio interface

The radio interface is the interface between the mobile stations and the fixed infrastructure. It is one of the most important interfaces of the GSM system.

One of the main objectives of GSM is roaming. Therefore, in order to obtain a complete compatibility between mobile stations and networks of different manufacturers and operators, the radio interface must be completely defined.

The spectrum efficiency depends on the radio interface and the transmission, more particularly in aspects such as the capacity of the system and the techniques used in order to decrease the interference and to improve the frequency reuse scheme. The specification of the radio interface has then an important influence on the spectrum efficiency.

1.5.1 Frequency allocation

Two frequency bands, of 25 MHz each one, have been allocated for the GSM system:

- The band 890-915 MHz has been allocated for the uplink direction (transmitting from the mobile station to the base station).
- The band 935-960 MHz has been allocated for the downlink direction (transmitting from the base station to the mobile station).

But not all the countries can use the whole GSM frequency bands. This is due principally to military reasons and to the existence of previous analog systems using part of the two 25 MHz frequency bands.

1.5.2 Multiple access scheme

The multiple access scheme defines how different simultaneous communications, between different mobile stations situated in different cells, share the GSM radio spectrum. A mix of Frequency Division Multiple Access (FDMA) and Time Division Multiple Access (TDMA), combined with frequency hopping, has been adopted as the multiple access scheme for GSM.

- **FDMA and TDMA**

Using FDMA, a frequency is assigned to a user. So the larger the number of users in a FDMA system, the larger the number of available frequencies must be. The limited available radio spectrum and the fact that a user will not free its assigned frequency until he does not need it anymore, explain why the number of users in a FDMA system can be "quickly" limited.

On the other hand, TDMA allows several users to share the same channel. Each of the users, sharing the common channel, is assigned his own burst within a group of bursts called a frame. Usually TDMA is used with a FDMA structure.

In GSM, a 25 MHz frequency band is divided, using a FDMA scheme, into 124 carrier frequencies spaced one from each other by a 200 kHz frequency band. Normally a 25 MHz frequency band can provide 125 carrier frequencies but the first carrier frequency is used as a guard band between GSM and other services working on lower frequencies. Each carrier frequency is then divided in time using a TDMA scheme. This scheme splits the radio channel, with a width of 200 kHz, into 8 bursts. A burst is the unit of time in a TDMA system, and it lasts approximately 0.577 ms. A TDMA frame is formed with 8 bursts and lasts, consequently, 4.615 ms. Each of the eight bursts, that form a TDMA frame, are then assigned to a single user.

- **Channel structure**

A channel corresponds to the recurrence of one burst every frame. It is defined by its frequency and the position of its corresponding burst within a TDMA frame. In GSM there are two types of channels:

- The traffic channels used to transport speech and data information.
- The control channels used for network management messages and some channel maintenance tasks.

a) Traffic channels (TCH)

Full-rate traffic channels (TCH/F) are defined using a group of 26 TDMA frames called a 26-Multiframe. The 26-Multiframe lasts consequently 120 ms. In this 26-Multiframe structure, the traffic channels for the downlink and uplink are separated by 3 bursts. As a consequence, the mobiles will not need to transmit and receive at the same time which simplifies considerably the electronics of the system.

The frames that form the 26-Multiframe structure have different functions:

- 24 frames are reserved to traffic.
- 1 frame is used for the Slow Associated Control Channel (SACCH).
- The last frame is unused. This idle frame allows the mobile station to perform other functions, such as measuring the signal strength of neighboring cells.

Half-rate traffic channels (TCH/H), which doubles the capacity of the system, are also grouped in a 26-Multiframe but the internal structure is different.

b) Control channels

According to their functions, four different classes of control channels are defined:

- Broadcast channels.
- Common control channels.
- Dedicated control channels.
- Associated control channels.

-Broadcast channels (BCH)

The BCH channels are used, by the base station, to provide the mobile station with the sufficient information it needs to synchronize with the network. Three different types of BCHs can be distinguished:

- The Broadcast Control Channel (BCCH), which gives to the mobile station the parameters needed in order to identify and access the network
- The Synchronization Channel (SCH), which gives to the mobile station the training sequence needed in order to demodulate the information transmitted by the base station
- The Frequency-Correction Channel (FCCH), which supplies the mobile station with the frequency reference of the system in order to synchronize it with the network

- Common Control Channels (CCCH)

The CCCH channels help to establish the calls from the mobile station or the network. Three different types of CCCH can be defined:

- The Paging Channel (PCH). It is used to alert the mobile station of an incoming cal
- The Random Access Channel (RACH), which is used by the mobile station to request access to the network
- The Access Grant Channel (AGCH). It is used, by the base station, to inform the mobile station about which channel it should use. This channel is the answer of a base station to a RACH from the mobile station

- Dedicated Control Channels (DCCH)

The DCCH channels are used for message exchange between several mobiles or a mobile and the network. Two different types of DCCH can be defined:

- The Standalone Dedicated Control Channel (SDCCH), which is used in order to exchange signaling information in the downlink and uplink directions.
- The Slow Associated Control Channel (SACCH). It is used for channel maintenance and channel control.

- Associated Control Channels

The Fast Associated Control Channels (FACCH) replace all or part of a traffic channel when urgent signaling information must be transmitted.

The FACCH channels carry the same information as the SDCCH channels.

- **Burst structure**

As it has been stated before, the burst is the unit in time of a TDMA system. Four different types of bursts can be distinguished in GSM:

- The frequency-correction burst is used on the FCCH. It has the same length as the normal burst but a different structure.
- The synchronization burst is used on the SCH. It has the same length as the normal burst but a different structure.
- The random access burst is used on the RACH and is shorter than the normal burst.
- The normal burst is used to carry speech or data information. It lasts approximately 0.577 ms and has a length of 156.25 bits. Its structure is presented in figure1.3.

The tail bits (T) are a group of three bits set to zero and placed at the beginning and the end of a burst. They are used to cover the periods of ramping up and down of the mobile's power.

The coded data bits correspond to two groups, of 57 bits each, containing signaling or user data.

The stealing flags (S) indicate, to the receiver, whether the information carried by a burst corresponds to traffic or signaling data.

The training sequence has a length of 26 bits. It is used to synchronize the receiver with the incoming information, avoiding then the negative effects produced by a multipath propagation.

The guard period (GP), with a length of 8.25 bits, is used to avoid a possible overlap of two mobiles during the ramping time.

- **Frequency hopping**

The propagation conditions and therefore the multipath fading depend on the radio frequency. In order to avoid important differences in the quality of the channels,

the slow Frequency hopping is introduced. The slow frequency hopping changes the frequency with every TDMA frame. A fast frequency hopping changes the frequency many times per frame but it is not used in GSM. The frequency hopping also reduces the effects of co-channel interference.

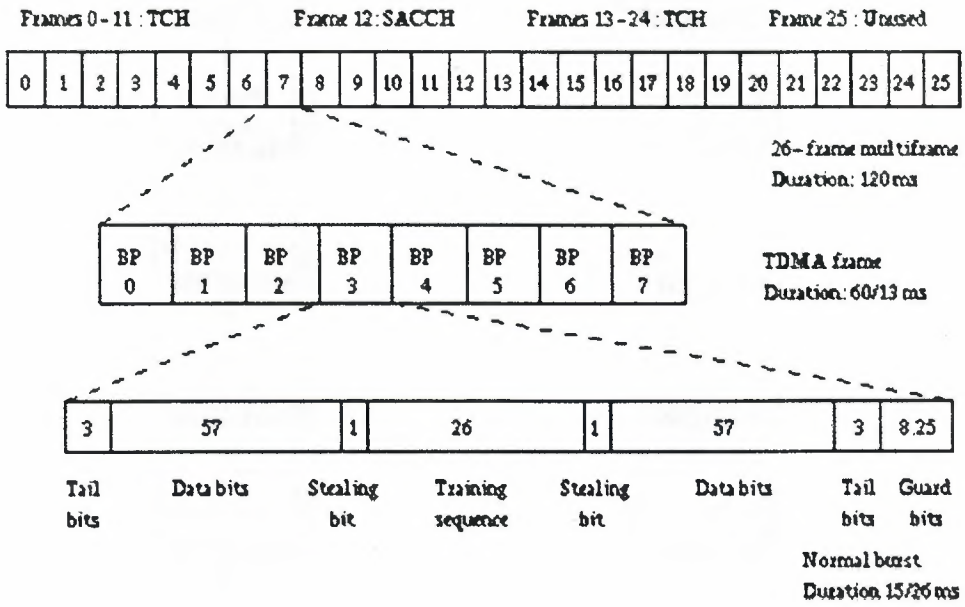

Figure 1.3 Structure of the 26-Multiframe, the TDMA frame and the normal burst

There are different types of frequency hopping algorithms. The algorithm selected is sent through the Broadcast Control Channels.

Even if frequency hopping can be very useful for the system, a base station does not have to support it necessarily on the other hand, a mobile station has to accept frequency hopping when a base station decides to use it.

1.5.3 from source information to radio waves

The figure1.4 presents the different operations that have to be performed in order to pass from the speech source to radio waves and vice versa.

- **Speech coding**

The transmission of speech is, at the moment, the most important service of a mobile cellular system. The GSM speech codec, which will transform the analog signal (voice) into a digital representation, has to meet the following criteria:

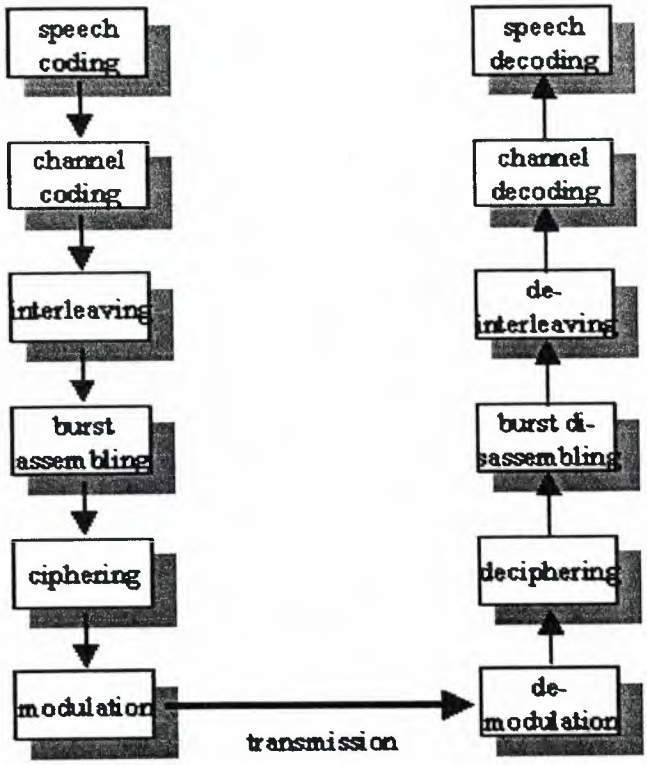

Figure 1.4 From speech source to radio waves

- A good speech quality, at least as good as the one obtained with previous cellular systems.
- To reduce the redundancy in the sounds of the voice. This reduction is essential due to the limited capacity of transmission of a radio channel.
- The speech codec must not be very complex because complexity is equivalent to high costs.

The final choice for the GSM speech codec is a codec named RPE-LTP (Regular Pulse Excitation Long-Term Prediction). This codec uses the information from previous samples (this information does not change very quickly) in order to predict the current sample. The speech signal is divided into blocks of 20 ms. These blocks are then passed to the speech codec, which has a rate of 13 kbps, in order to obtain blocks of 260 bits.

- **Channel coding**

Channel coding adds redundancy bits to the original information in order to detect and correct, if possible, errors occurred during the transmission.

a) Channel coding for the GSM data TCH channels

The channel coding is performed using two codes: a block code and a convolutional code.

The block code corresponds to the block code defined in the GSM Recommendations 05.03. The block code receives an input block of 240 bits and adds four zero tail bits at the end of the input block. The output of the block code is consequently a block of 244 bits.

A convolutional code adds redundancy bits in order to protect the information. A convolutional encoder contains memory. This property differentiates a convolutional code from a block code. A convolutional code can be defined by three variables: n, k and K. The value n corresponds to the number of bits at the output of the encoder, k to the number of bits at the input of the block and K to the memory of the encoder. The ratio, R, of the code is defined as follows: R = k/n. Let's consider a convolutional code with the following values: k is equal to 1, n to 2 and K to 5. This convolutional code uses then a rate of R = 1/2 and a delay of K = 5, which means that it will add a redundant bit for each input bit. The convolutional code uses 5 consecutive bits in order to compute the redundancy bit. As the convolutional code is a 1/2 rate convolutional code, a block of 488 bits is generated. These 488 bits are punctured in order to produce a block of 456 bits. Thirty two bits, obtained as follows, are not transmitted:

C (11 + 15 j) for j = 0, 1, ..., 31

The block of 456 bits produced by the convolutional code is then passed to the interleaver.

b) Channel coding for the GSM speech channels

Before applying the channel coding, the 260 bits of a GSM speech frame are divided in three different classes according to their function and importance. The most

important class is the class Ia containing 50 bits. Next in importance is the class Ib, which contains 132 bits. The least important is the class II, which contains the remaining 78 bits. The different classes are coded differently. First of all, the class Ia bits are block-coded. Three parity bits, used for error detection, are added to the 50 class Ia bits. The resultant 53 bits are added to the class Ib bits. Four zero bits are added to this block of 185 bits (50+3+132). A convolutional code, with r = 1/2 and K = 5, is then applied, obtaining an output block of 378 bits. The class II bits are added, without any protection, to the output block of the convolutional coder. An output block of 456 bits is finally obtained.

c) Channel coding for the GSM control channels

In GSM the signaling information is just contained in 184 bits. Forty parity bits, obtained using a fire code, and four zero bits are added to the 184 bits before applying the convolutional code (r = 1/2 and K = 5). The output of the convolutional code is then a block of 456 bits, which does not need to be punctured.

- **Interleaving**

An interleaving rearranges a group of bits in a particular way. It is used in combination with FEC codes in order to improve the performance of the error correction mechanisms. The interleaving decreases the possibility of losing whole bursts during the transmission, by dispersing the errors. Being the errors less concentrated, it is then easier to correct them.

a) Interleaving for the GSM control channels

A burst in GSM transmits two blocks of 57 data bits each. Therefore the 456 bits corresponding to the output of the channel coder fit into four bursts (4*114 = 456).

The 456 bits are divided into eight blocks of 57 bits. The first block of 57 bits contains the bit numbers (0, 8, 16, ...448), the second one the bit numbers (1, 9, 17, ...449), etc. The last block of 57 bits will then contain the bit numbers (7, 15, ...455). The first four blocks of 57 bits are placed in the even-numbered bits of four bursts. The other four blocks of 57 bits are placed in the odd-numbered bits of the same four bursts.

Therefore the interleaving depth of the GSM interleaving for control channels is four and a new data block starts every four bursts. The interleaver for control channels is called a block rectangular interleaver.

b) Interleaving for the GSM speech channels

The block of 456 bits, obtained after the channel coding, is then divided in eight blocks of 57 bits in the same way as it is explained in the previous paragraph. But these eight blocks of 57 bits are distributed differently. The first four blocks of 57 bits are placed in the even-numbered bits of four consecutive bursts. The other four blocks of 57 bits are placed in the odd-numbered bits of the next four bursts. The interleaving depth of the GSM interleaving for speech channels is then eight. A new data block also starts every four bursts. The interleaver for speech channels is called a block diagonal interleaver.

c) Interleaving for the GSM data TCH channels

A particular interleaving scheme, with an interleaving depth equal to 22, is applied to the block of 456 bits obtained after the channel coding. The block is divided into 16 blocks of 24 bits each, 2 blocks of 18 bits each, 2 blocks of 12 bits each and 2 blocks of 6 bits each. It is spread over 22 bursts in the following way:

- the first and the twenty-second bursts carry one block of 6 bits each
- the second and the twenty-first bursts carry one block of 12 bits each
- the third and the twentieth bursts carry one block of 18 bits each
- from the fourth to the nineteenth burst, a block of 24 bits is placed in each burst

A burst will then carry information from five or six consecutive data blocks. The data blocks are said to be interleaved diagonally. A new data block starts every four bursts.

- **Burst assembling**

The burst assembling procedure is in charge of grouping the bits into bursts. Section 1.5.2.3 presents the different bursts structures and describes in detail the structure of the normal burst.

- **Ciphering**

Ciphering is used to protect signaling and user data. First of all, a ciphering key is computed using the algorithm A8 stored on the SIM card, the subscriber key and a random number delivered by the network (this random number is the same as the one used for the authentication procedure). Secondly, a 114 bit sequence is produced using the ciphering key, an algorithm called A5 and the burst numbers. This bit sequence is then XORed with the two 57 bit blocks of data included in a normal burst.

In order to decipher correctly, the receiver has to use the same algorithm A5 for the deciphering procedure.

- **Modulation**

The modulation chosen for the GSM system is the Gaussian Minimum Shift Keying (GMSK).

The aim of this section is not to describe precisely the GMSK modulation as it is too long and it implies the presentation of too many mathematical concepts. Therefore, only brief aspects of the GMSK modulation are presented in this section.

The GMSK modulation has been chosen as a compromise between spectrum efficiency, complexity and low spurious radiations (that reduce the possibilities of adjacent channel interference). The GMSK modulation has a rate of 270 5/6 kbauds and a BT product equal to 0.3. Figure1.5 presents the principle of a GMSK modulator.

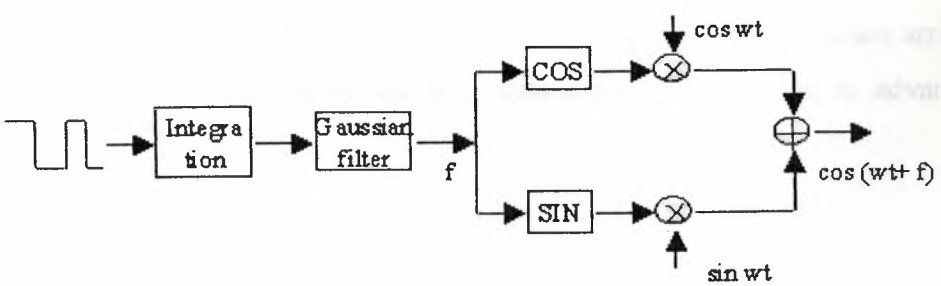

Figure 1.5 GMSK modulator

1.5.4 Discontinuous transmission (DTX)

This is another aspect of GSM that could have been included as one of the requirements of the GSM speech codec. The function of the DTX is to suspend the radio transmission during the silence periods. This can become quite interesting if we take into consideration the fact that a person speaks less than 40 or 50 percent during a conversation. The DTX helps then to reduce interference between different cells and to increase the capacity of the system. It also extends the life of a mobile's battery. The DTX function is performed thanks to two main features:

- The Voice Activity Detection (VAD), which has to determine whether the sound represents speech or noise, even if the background noise is very important. If the voice signal is considered as noise, the transmitter is turned off producing then, an unpleasant effect called clipping.
- The comfort noise. An inconvenient of the DTX function is that when the signal is considered as noise, the transmitter is turned off and therefore, a total silence is heard at the receiver. This can be very annoying to the user at the reception because it seems that the connection is dead. In order to overcome this problem, the receiver creates a minimum of background noise called comfort noise. The comfort noise eliminates the impression that the connection is dead.

1.5.5 Timing advance

The timing of the bursts transmissions is very important. Mobiles are at different distances from the base stations. Their delay depends, consequently, on their distance. The aim of the timing advance is that the signals coming from the different mobile stations arrive to the base station at the right time. The base station measures the timing delay of the mobile stations. If the bursts corresponding to a mobile station arrive too late and overlap with other bursts, the base station tells, this mobile, to advance the transmission of its bursts.

1.5.6 Power control

At the same time the base stations perform the timing measurements, they also perform measurements on the power level of the different mobile stations. These power levels are adjusted so that the power is nearly the same for each burst.

A base station also controls its power level. The mobile station measures the strength and the quality of the signal between itself and the base station. If the mobile station does not receive correctly the signal, the base station changes its power level.

1.5.7 Discontinuous reception

It is a method used to conserve the mobile station's power. The paging channel is divided into sub channels corresponding to single mobile stations. Each mobile station will then only 'listen' to its sub channel and will stay in the sleep mode during the other sub channels of the paging channel.

1.5.8 Multipath and equalization

At the GSM frequency bands, radio waves reflect from buildings, cars, hills, etc. So not only the 'right' signal (the output signal of the emitter) is received by an antenna, but also many reflected signals, which corrupt the information, with different phases.

An equalizer is in charge of extracting the 'right' signal from the received signal. It estimates the channel impulse response of the GSM system and then constructs an inverse filter. The receiver knows which training sequence it must wait for. The equalizer will then, comparing the received training sequence with the training sequence it was expecting, compute the coefficients of the channel impulse response. In order to extract the 'right' signal, the received signal is passed through the inverse filter.

1.6 GSM services

It is important to note that all the GSM services were not introduced since the appearance of GSM but they have been introduced in a regular way. The GSM Memorandum of Understanding (MoU) defined four classes for the introduction of the different GSM services:

Three categories of services can be distinguished:

- Teleservices.
- Bearer services.
- Supplementary Services.

1.6.1 Teleservices

- Telephony.
- Facsimile group 3.
- Emergency calls.
- Teletex.
- Short Message Services. Using these services, a message of a maximum of 160 alphanumeric characters can be sent to or from a mobile station. If the mobile is powered off, the message is stored. With the SMS Cell Broadcast (SMS-CB), a message of a maximum of 93 characters can be broadcast to all mobiles in a certain geographical area.
- Fax mail. Thanks to this service, the subscriber can receive fax messages at any fax machine.
- Voice mail. This service corresponds to an answering machine.

1.6.2 Bearer services

A bearer service is used for transporting user data. Some of the bearer services are listed below:

- Asynchronous and synchronous data, 300-9600 bps.
- Alternate speech and data, 300-9600 bps.
- Asynchronous PAD (packet-switched, packet assembler/disassembler) access, 300-9600 bps.
- Synchronous dedicated packet data access, 2400-9600 bps.

2. GSM SECURITY MODEL

2.1 Introduction to the GSM Security Model

The GSM Security Model is based on a shared secret between the subscriber's home network's HLR and the subscriber's SIM. The shared secret, called Ki, is a 128-bit key used to generate a 32-bit signed response, called SRES, to a Random Challenge, called RAND, made by the MSC, and a 64-bit session key, called Kc, used for the encryption of the over-the-air channel. When a MS first signs on to a network, the HLR provides the MSC with five triples containing a RAND, a SRES to that particular RAND based on the Ki and a Kc based again on the same Ki. Each of the triples is used for one authentication of the specific MS. When all triples have been used the HLR provides a new set of five triples for the MSC.

When the MS first comes to the the area of a particular MSC, the MSC sends the Challenge of the first triple to the MS. The MS calculates a SRES with the A3 algorithm using the given Challenge and the Ki residing in the SIM. The MS then sends the SRES to the MSC, which can confirm that the SRES really corresponds to the Challenge sent by comparing the SRES from the MS and the SRES in the triple from the HLR. Thus, the MS has authenticated itself to the MSC.

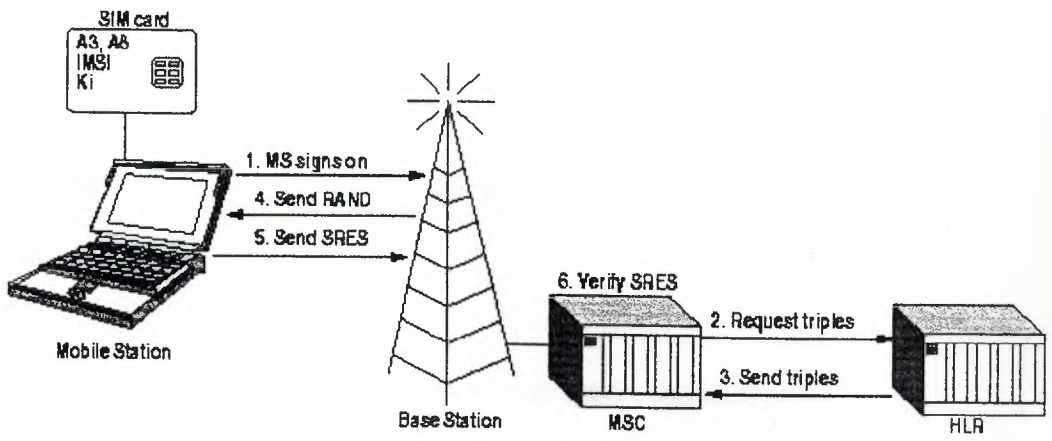

Figure 2.1 Mobile station authentication

The MS then generates a Session Key, Kc, with the A8 algorithm using, again, the Challenge from the MSC and the Ki from the SIM. The BTS, which is used to communicate with the MS, receives the same Kc from the MSC, which has received it in the triple from the HLR. Now the over-the-air communication channel between the BTS and MS can be encrypted.

Each frame in the over-the-air traffic is encrypted with a different keystream. This keystream is generated with the A5 algorithm. The A5 algorithm is initialized with the Kc and the number of the frame to be encrypted, thus generating a different keystream for every frame. This means that one call can be decrypted when the attacker knows the Kc and the frame numbers. The frame numbers are generated implicitly, which means that anybody can find out the frame number at hand. The same Kc is used as long as the MSC does not authenticate the MS again, in which case a new Kc is generated. In practice, the same Kc may be in use for days. The MS authentication is an optional procedure in the beginning of a call, but it is usually not performed. Thus, the Kc is not changed during calls.

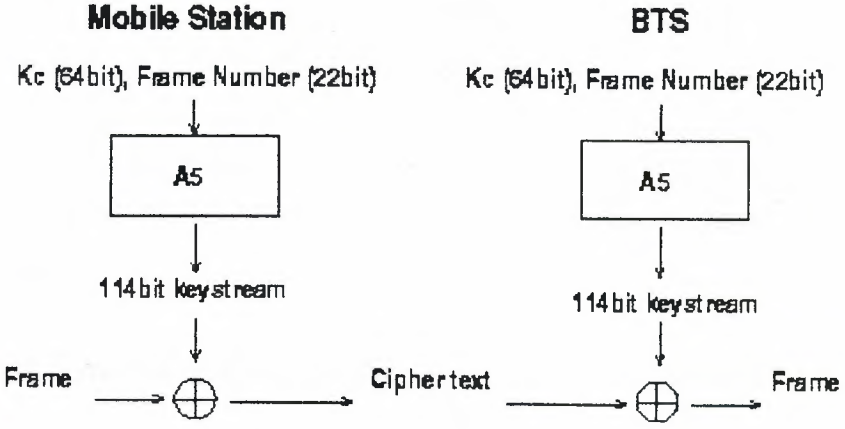

Figure 2.2 Frame encryption and decryption

Only the over-the-air traffic is encrypted in a GSM network. Once the frames have been received by the BTS, it decrypts them and sends them in plaintext to the operator's backbone network.

2.2 A3, The MS Authentication Algorithm

The A3 is the authentication algorithm in the GSM security model. Its function is to generate the SRES response to the MSC's random challenge, RAND, which the MSC has received from the HLR. The A3 algorithm gets the RAND from the MSC and the secret key Ki from the SIM as input and generates a 32-bit output, which is the SRES response. Both the RAND and the Ki secret are 128 bits long.

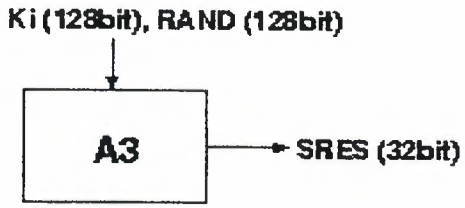

Figure 2.3 Signed response (SRES) calculation

Nearly every GSM operator in the world uses an algorithm called COMP128 for both A3 and A8 algorithms. COMP128 is the reference algorithm for the tasks pointed out by the GSM Consortium. Other algorithms have been named as well, but almost every operator uses the COMP128 except a couple of exceptions. See Figure 2.5.

The COMP128 takes the RAND and the Ki as input, but it generates 128 bits of output, instead of the 32-bit SRES. The first 32 bits of the 128 bits form the SRES response.

2.3 A8, The Voice-Privacy Key Generation Algorithm

The A8 algorithm is the key generation algorithm in the GSM security model. The A8 generates the session key, Kc, from the random challenge, RAND, received from the MSC and from the secret key Ki. The A8 algorithm takes the two 128-bit inputs and generates a 64-bit output from them. This output is the 64-bit session key Kc. See Figure 2.4. The BTS received the same Kc from the MSC. HLR was able to generate the Kc, because the HLR knows both the RAND (the HLR generated it) and the secret key Ki,

which it holds for all the GSM subscribers of this network operator. One session key, Kc, is used until the MSC decides to authenticate the MS again. This might take days.

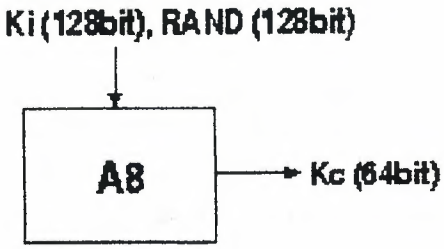

Figure 2.4 Session key (Kc) calculation

As stated, COMP128 is used for both the A3 and A8 algorithms in most GSM networks. The COMP128 generates both the SRES response and the session key, Kc, on one run. The last 54 bits of the COMP128 output form the session key, Kc, until the MS is authenticated again. See Figure 2.5.

Note that the key length at this point is 54 bits instead of 64 bits, which is the length of the key given as input to the A5 algorithm. Ten zero-bits are appended to the key generated by the COMP128 algorithm. Thus, we have a key of 64 bits with the last ten bits zeroed out. This effectively reduces the keyspace from 64 bits to 54 bits. This is done in all A8 implementations, including those that do not use COMP128 for key generation, and seems to be a deliberate feature of the A8 algorithm implementations.

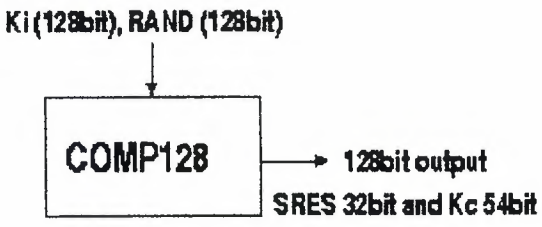

Figure 2.5 COMP128 calculation

Both the A3 and A8 algorithms are stored in the SIM in order to prevent people from tampering with them. This means that the operator can decide which algorithms to use independently from hardware manufacturers and other network operators. The authentication works in other countries as well, because the local network asks the HLR

of the subscriber's home network for the five triples. Thus, the local network does not have to know anything about the A3 and A8 algorithms used.

2.4 A5/1, The Strong Over-the-Air Voice-Privacy Algorithm

The A5 algorithm is the stream cipher used to encrypt over-the-air transmissions. The stream cipher is initialized all over again for every frame sent. The stream cipher is initialized with the session key, Kc, and the number of the frame being de/encrypted. The same Kc is used throughout the call, but the 22-bit frame number changes during the call, thus generating a unique keystream for every frame. See Figure 2.6.

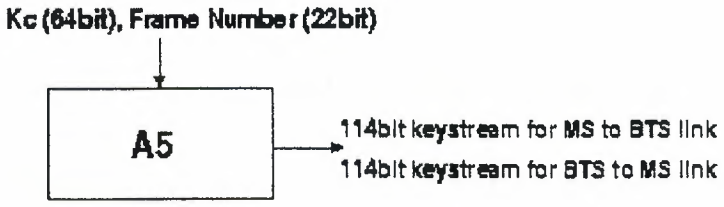

Figure 2.6 Keystream generation

The A5 algorithm used in European countries consists of three LSFRs of different lengths. See Figure 2.7. The combined length of the three LSFRs is 64 bits. The outputs of the three registers are XORred together and the XOR represents one keystream bit. The LSFRs are 19, 22 and 23 bits long with sparse feedback polynomials. All three registers are clocked, based on the middle bit of the register. A register is clocked if its middle bit agrees with the majority value of the three middle bits. For example, if the middle bits of the three registers are 1, 1 and 0, the first two register are clocked or if the middle bits are 0, 1 and 0, then the first and third register are clocked.

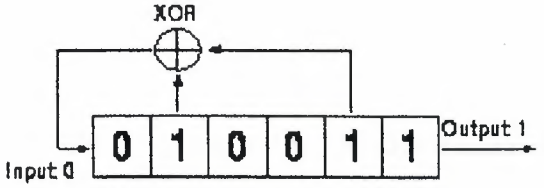

Figure 2.7 An example LSFR with feedback polynomial of $x^6 + x^4 + x$

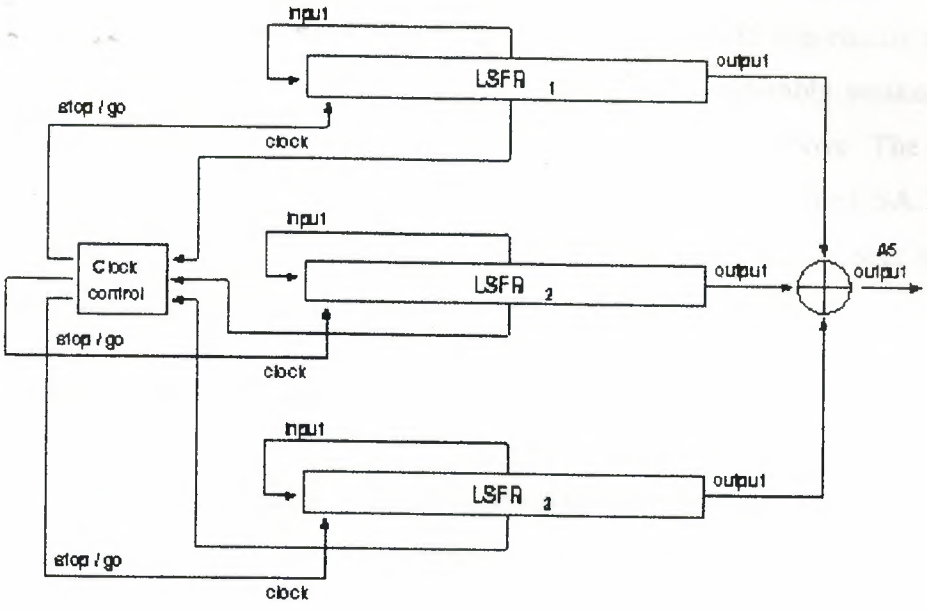

Figure 2.8 A5 LSFR construction

The three LSFRs are initialized with the session key, Kc, and the frame number. The 64-bit Kc is first loaded into the register bit by bit. The LSB of the key is XORred into each of the LSFRs. The registers are then all clocked (the majority clocking rule is disabled).

All 64 bits of the key are loaded into the registers the same way. The 22-bit frame number is also loaded into the register in the same way except that the majority clocking rule applies from now on. After the registers have been initialized with the Kc and the current frame number, they are clocked one hundred times and the generated keystream bits are discarded.

This is done in order to mix the frame number and keying material together. Now 228 bits of keystream output are generated. The first 114 bits are used to encrypt the frame from MS to BTS and the next 114 bits are used to encrypt the frame from BTS to MS. After this, the A5 algorithm is initialized again with the same Kc and the number of the next frame.

Since the first GSM systems, other A5 algorithms have been designed and implemented. The main motivation has been that the original A5 encryption algorithm is too strong to export to the Middle East. Thus, the first 'original' A5 algorithm was

renamed A5/1. Other algorithms include A5/0, which means no encryption at all, and A5/2, a weaker over-the-air privacy algorithm. Generally, the A5 algorithms after A5/1 have been named A5/x. Most of the A5/x algorithms are considerably weaker than the A5/1, which has the time complexity of 2^54 at most as, shown above. The estimated time complexity of A5/2 is as low as 2^16. This encryption is used in the USA. The other A5 implementations have not leaked. Thus, there are no real facts about them, just guesses and assumptions.

2.5 Possible Interception Attacks

The interesting question about the GSM security model is whether a call can be eavesdropped, now that at least one of the algorithms it depends on has been proven faulty.

Scientist around the world seem to be unanimous that the over-the-air interception and real time decoding of a call is still impossible regardless of the reduced key space. But there seem to be other ways of attacking the system that are feasible and seem to be very real threats. There are also many attacks that are realistic, yet do not abuse any of the faults in the security algorithms.

2.5.1 Brute-Force Attack against A5

A real-time brute-force attack against the GSM security system is not feasible, as stated above. The time complexity of the attack is 2^54 (2^64 if the ten bits were not zeroed out). This requires too much time in order to be feasible in eavesdropping on GSM calls in real time. It might be possible to record the frames between the MS and the BTS and launch the attack afterwards though.

If we have a Pentium III class chip with approximately 20 million transistors and the implementation of one set of LSFRs (A5/1) would require about 2000 transistors, we would have a set of 10,000 parallel A5/1 implementations on one chip.

If the chip was clocked to 600 MHz and each A5 implementation would generate one output bit for each clock cycle and we would need to generate 100+114+114 output bits, we could try approximately 2M keys per second per A5/1 implementation. A keyspace of 2^54 keys would thus require about 900,000 seconds, 250 hours, with one

chip. The attack can be optimized by giving up on a specific key after the first invalid keystream bit. This would cut the required time down by one third. The attack can also be distributed between multiple chips, thus drastically decreasing the time required.

2.5.2 Divide-and-Conquer Attack against A5

A divide-and-conquer attack manages to reduce the complexity from 2^54 of the brute-force attack to 2^45, which is a relatively dramatic change (2^9 = 512 times faster). The divide-and-conquer attack is based on a known-plain-text attack. The attacker tries to determine the initial states of the LSFRs from a known keystream sequence. The attacker needs to know 64 successive keystream bits that can be retrieved if the attacker knows some cipher text and the corresponding plain text. This depends largely on the format of the GSM frames sent back and forth. The GSM frames contain a lot of constant information, e.g. frame headers. The required 64 bits might not always be known, but 32 to 48 bits are usually known, sometimes even more. Keep in mind that the attacker needs only one 64-bit plain text segment.

In short the divide-and-conquer attack is implemented by guessing the content of the two shorter LSFRs and then computing the third LSFR from the known keystream. This would be a 2^40 attack, if the clockings of the first two registers were not dependent on the third register. Because the middle bit of the third register is used for clocking, we have to guess about half of the bits in the third register between the clock bit and the LSB as well. This fact increases the time complexity from 2^40 to 2^45 [2].

However, J. Golic has proposed another divide-and-conquer attack based on the same assumptions with the average complexity of 2^40.16. Golic showed that only 2^62.32 internal states could be reached from the 2^64 initial states. Based on this assumption, he describes how to obtain linear equations by guessing n bits in the LSFRs. By solving these linear equations, one could recover the initial states of the three LSFRs. The complexity of solving the linear equations is 2^41.16. On average, one would resolve the internal state with 50 per cent chance in 2^40.16 operations.

Golic also proposed a Time-Memory Trade-Off Attack based on the Birthday Paradox in the same paper. The objective of the attack is to recover the internal states of

the three LSFRs at a known time for a known keystream sequence corresponding to a known frame number, thus reconstructing the session key, Kc.

2.5.3 Accessing the Signaling Network

As the two examples above clearly state, the A5 algorithm is not secure cryptographically, as there is another more feasible attack than the brute-force attack and it is not secure in practice either, because the brute-force attack in itself is not very hard to implement with current hardware. Yet, the algorithm is secure enough to prevent over-the-air call interception and real-time encryption cracking. Unfortunately, the air waves between the MS and the BTS are not the only vulnerable point in the GSM system.

As stated earlier, the transmissions are encrypted only between the MS and the BTS. After the BTS, the traffic is transmitted in plain text within the operator's network.

This opens up new possibilities. If the attacker can access the operator's signaling network, he will be able to listen to everything that is transmitted, including the actual phone call as well as the RAND, SRES and Kc. The SS7 signaling network used in the operator's GSM network is completely insecure if the attacker gains direct access to it.

In another scenario, the attacker could attack the HLR of a particular network. If the attacker can access the HLR, he will be able to retrieve the Kis for all the subscribers of that particular network. Luckily the HLR is usually a bit more secure than the rest of the network, thus making it a slightly less probable point of entry, yet not completely improbable either keeping in mind the potential gain involved.

Accessing the signaling network is not very difficult. Although the BTSs are usually connected to the BSC through a cable, some of them are connected to the BSC through a microwave or even a satellite link. This link would be relatively easy to access with the right kind of equipment. Most of the commercially available equipment for GSM eavesdropping seem to use this particular vulnerability. Unfortunately we cannot verify this, because the equipment and specifications are available only to law enforcement personnel and such. The microwave link might be encrypted, however, depending on the hardware manufacturer, thus making it slightly more difficult to monitor it. It is really a question about whether the attacker wants to crack the A5 encryption protecting the session of a specific MS or the encryption between the BTS

and the BSC and gaining access to the backbone network. The possibility of accessing the cable leaving the BTS should not be ruled out either. This might be a very real threat and an attack could go undetected for a long time, if implemented carefully. The ability to tap on to the data transmitted between the BTS and BSC would enable the attacker to either monitor the call by eavesdropping on the channel throughout the call or he could retrieve the session key, Kc, by monitoring the channel, intercept the call over the air and decrypt it on the fly. Now that he knows the Kc, the real-time encryption is not a problem.

2.5.4 Retrieving the Key from the SIM

The security of the whole GSM security model is based on the secret Ki. If this key is compromised the whole account is compromised. Once the attacker is able to retrieve the Ki, he can not only listen to the subscribers calls, but also place calls billed to the original subscriber's account, because he can now impersonate the legitimate subscriber. The GSM network has trip wires for this: If two phones with the same ID are powered at the same time, the GSM network notices this, makes a location query for the phones, notices that the 'same' phone is in two different locations at the same time, and closes the account, thus preventing the attacker and the legitimate subscriber from placing calls. But this is not relevant if the attacker is only interested in listening to the calls of the subscriber. In this case, the attacker can stay passive and just listen to the call, thus staying invisible to the GSM network.

The Smartcard Developer Association and the ISAAC security research group discovered a flaw in the COMP128 algorithm that effectively enabled them to retrieve the secret key, Ki, from a SIM. The attack was performed on a SIM they had physical access to, but the same attack is applicable when launched over-the-air as well.

The attack is based on a chosen-challenge attack that works, because the COMP128 algorithm is broken in such a way that it reveals information about the Ki when the appropriate RANDs are given as arguments to the A8 algorithm. The SIM was accessed through a smartcard reader connected to a PC. The PC made about 150.000 challenges to the SIM and the SIM generated the SRES and the session key, Kc, based on the challenge and the secret key. The secret key could be deduced from the SRES responses through differential cryptanalysis. The smartcard reader used in implementing

the attack could make 6.25 queries per second to the SIM card. So the attack required about eight hours to conduct. The results had to be analyzed as well, but this was apparently very quick, compared to the actual attack. Thus, the attacker needs to have physical access to the target SIM for at least eight hours. This is still very reasonable.

2.5.5 Retrieving the Key from the SIM over the Air

The SDA and ISAAC researchers are confident that the same SIM-cloning attack could be launched over the air as well. Unfortunately, they can probably not confirm their suspicions, because the necessary equipment is illegal. The over-the-air attack is based on the fact that the MS is required to respond to every challenge made by the GSM network. If the signal of the legitimate BTS is over powered by a rogue BTS of the attacker, the attacker can bomb the target MS with challenges and re-construct the secret key from these responses. Again the MS has to be available to the attacker over the air for the whole time it takes to conduct the attack. It is not known how long the attack would take when conducted over the air. Estimates vary from eight to thirteen hours.

The attack might be conducted in a subway, where the signal of the legitimate BTS is not available, but the phone is still turned on. The subscriber would be unaware of such an attack though the fact that the battery of the phone has run out slightly quicker than usual might make him suspicious.

The attack can also be performed in parts: instead of performing an eight-hour attack, the attacker could tease the phone for twenty minutes every day on the victim's way to work. Once the SIM is cloned, the SIM-clone is usable until the subscriber gets a new SIM, which in practice does not happen very often.

In another scenario, the subscriber is on a business trip in another country. The attacker has somehow bullied the local GSM operator to perform this attack on the subscriber's phone.

The attacker would again be able to reconstruct the Ki based on the MS's SRES answers and the attack would probably go unnoticed, because the challenges originate from a legitimate network. Keep in mind that the local network does not know anything about the Ki, because the triples originate from the HLR of the subscribers home network. Thus, the local network has to deduce the Ki from the A3 responses.

2.5.6 Retrieving the Key from the AuC

The same attack used in retrieving the Ki from a SIM card can be used to retrieve the Ki from the AuC. The AuC has to answer to requests made by the GSM network and return valid triples to be used in MS authentication. The procedure is basicly identical to the procedure used in the MS to access the SIM card. The difference is that the AuC is a lot faster in processing requests than a SIM card is, because it needs to process a lot more requests compared to one SIM card.

2.5.7 Cracking the A8 Algorithm

Another possibility is that someone will be able to crack the A8 key generation algorithm and retrieve the secret key, Ki, based on the random challenge, RAND, the session key, Kc, and the SRES response (assuming the same algorithm is used for both A3 and A8 as is the case with COMP128) with a minimal amount of work. For example, the attacker may find a RAND that produces the Ki as a result (an over simplified example). All three variables are obtained relatively easily. The RAND and SRES are sent over the air in plain text and the session key Kc can be relatively easily deduced from the encrypted frames and the known plain text given enough time. A vulnerability like this in the key generation algorithm would of course devastate the whole GSM security model and give the GSM Consortium something to think about when designing their next security algorithm.

and the BSC and gaining access to the backbone network. The possibility of accessing the cable leaving the BTS should not be ruled out either. This might be a very real threat and an attack could go undetected for a long time, if implemented carefully. The ability to tap on to the data transmitted between the BTS and BSC would enable the attacker to either monitor the call by eavesdropping on the channel throughout the call or he could retrieve the session key, Kc, by monitoring the channel, intercept the call over the air and decrypt it on the fly. Now that he knows the Kc, the real-time encryption is not a problem.

2.5.4 Retrieving the Key from the SIM

The security of the whole GSM security model is based on the secret Ki. If this key is compromised the whole account is compromised. Once the attacker is able to retrieve the Ki, he can not only listen to the subscribers calls, but also place calls billed to the original subscriber's account, because he can now impersonate the legitimate subscriber. The GSM network has trip wires for this: If two phones with the same ID are powered at the same time, the GSM network notices this, makes a location query for the phones, notices that the 'same' phone is in two different locations at the same time, and closes the account, thus preventing the attacker and the legitimate subscriber from placing calls. But this is not relevant if the attacker is only interested in listening to the calls of the subscriber. In this case, the attacker can stay passive and just listen to the call, thus staying invisible to the GSM network.

The Smartcard Developer Association and the ISAAC security research group discovered a flaw in the COMP128 algorithm that effectively enabled them to retrieve the secret key, Ki, from a SIM. The attack was performed on a SIM they had physical access to, but the same attack is applicable when launched over-the-air as well.

The attack is based on a chosen-challenge attack that works, because the COMP128 algorithm is broken in such a way that it reveals information about the Ki when the appropriate RANDs are given as arguments to the A8 algorithm. The SIM was accessed through a smartcard reader connected to a PC. The PC made about 150.000 challenges to the SIM and the SIM generated the SRES and the session key, Kc, based on the challenge and the secret key. The secret key could be deduced from the SRES responses through differential cryptanalysis. The smartcard reader used in implementing

the attack could make 6.25 queries per second to the SIM card. So the attack required about eight hours to conduct. The results had to be analyzed as well, but this was apparently very quick, compared to the actual attack. Thus, the attacker needs to have physical access to the target SIM for at least eight hours. This is still very reasonable.

2.5.5 Retrieving the Key from the SIM over the Air

The SDA and ISAAC researchers are confident that the same SIM-cloning attack could be launched over the air as well. Unfortunately, they can probably not confirm their suspicions, because the necessary equipment is illegal. The over-the-air attack is based on the fact that the MS is required to respond to every challenge made by the GSM network. If the signal of the legitimate BTS is over powered by a rogue BTS of the attacker, the attacker can bomb the target MS with challenges and re-construct the secret key from these responses. Again the MS has to be available to the attacker over the air for the whole time it takes to conduct the attack. It is not known how long the attack would take when conducted over the air. Estimates vary from eight to thirteen hours.

The attack might be conducted in a subway, where the signal of the legitimate BTS is not available, but the phone is still turned on. The subscriber would be unaware of such an attack though the fact that the battery of the phone has run out slightly quicker than usual might make him suspicious.

The attack can also be performed in parts: instead of performing an eight-hour attack, the attacker could tease the phone for twenty minutes every day on the victim's way to work. Once the SIM is cloned, the SIM-clone is usable until the subscriber gets a new SIM, which in practice does not happen very often.

In another scenario, the subscriber is on a business trip in another country. The attacker has somehow bullied the local GSM operator to perform this attack on the subscriber's phone.

The attacker would again be able to reconstruct the Ki based on the MS's SRES answers and the attack would probably go unnoticed, because the challenges originate from a legitimate network. Keep in mind that the local network does not know anything about the Ki, because the triples originate from the HLR of the subscribers home network. Thus, the local network has to deduce the Ki from the A3 responses.

2.5.6 Retrieving the Key from the AuC

The same attack used in retrieving the Ki from a SIM card can be used to retrieve the Ki from the AuC. The AuC has to answer to requests made by the GSM network and return valid triples to be used in MS authentication. The procedure is basicly identical to the procedure used in the MS to access the SIM card. The difference is that the AuC is a lot faster in processing requests than a SIM card is, because it needs to process a lot more requests compared to one SIM card.

2.5.7 Cracking the A8 Algorithm

Another possibility is that someone will be able to crack the A8 key generation algorithm and retrieve the secret key, Ki, based on the random challenge, RAND, the session key, Kc, and the SRES response (assuming the same algorithm is used for both A3 and A8 as is the case with COMP128) with a minimal amount of work. For example, the attacker may find a RAND that produces the Ki as a result (an over simplified example). All three variables are obtained relatively easily. The RAND and SRES are sent over the air in plain text and the session key Kc can be relatively easily deduced from the encrypted frames and the known plain text given enough time. A vulnerability like this in the key generation algorithm would of course devastate the whole GSM security model and give the GSM Consortium something to think about when designing their next security algorithm.

and the BSC and gaining access to the backbone network. The possibility of accessing the cable leaving the BTS should not be ruled out either. This might be a very real threat and an attack could go undetected for a long time, if implemented carefully. The ability to tap on to the data transmitted between the BTS and BSC would enable the attacker to either monitor the call by eavesdropping on the channel throughout the call or he could retrieve the session key, Kc, by monitoring the channel, intercept the call over the air and decrypt it on the fly. Now that he knows the Kc, the real-time encryption is not a problem.

2.5.4 Retrieving the Key from the SIM

The security of the whole GSM security model is based on the secret Ki. If this key is compromised the whole account is compromised. Once the attacker is able to retrieve the Ki, he can not only listen to the subscribers calls, but also place calls billed to the original subscriber's account, because he can now impersonate the legitimate subscriber. The GSM network has trip wires for this: If two phones with the same ID are powered at the same time, the GSM network notices this, makes a location query for the phones, notices that the 'same' phone is in two different locations at the same time, and closes the account, thus preventing the attacker and the legitimate subscriber from placing calls. But this is not relevant if the attacker is only interested in listening to the calls of the subscriber. In this case, the attacker can stay passive and just listen to the call, thus staying invisible to the GSM network.

The Smartcard Developer Association and the ISAAC security research group discovered a flaw in the COMP128 algorithm that effectively enabled them to retrieve the secret key, Ki, from a SIM. The attack was performed on a SIM they had physical access to, but the same attack is applicable when launched over-the-air as well.

The attack is based on a chosen-challenge attack that works, because the COMP128 algorithm is broken in such a way that it reveals information about the Ki when the appropriate RANDs are given as arguments to the A8 algorithm. The SIM was accessed through a smartcard reader connected to a PC. The PC made about 150.000 challenges to the SIM and the SIM generated the SRES and the session key, Kc, based on the challenge and the secret key. The secret key could be deduced from the SRES responses through differential cryptanalysis. The smartcard reader used in implementing

the attack could make 6.25 queries per second to the SIM card. So the attack required about eight hours to conduct. The results had to be analyzed as well, but this was apparently very quick, compared to the actual attack. Thus, the attacker needs to have physical access to the target SIM for at least eight hours. This is still very reasonable.

2.5.5 Retrieving the Key from the SIM over the Air

The SDA and ISAAC researchers are confident that the same SIM-cloning attack could be launched over the air as well. Unfortunately, they can probably not confirm their suspicions, because the necessary equipment is illegal. The over-the-air attack is based on the fact that the MS is required to respond to every challenge made by the GSM network. If the signal of the legitimate BTS is over powered by a rogue BTS of the attacker, the attacker can bomb the target MS with challenges and re-construct the secret key from these responses. Again the MS has to be available to the attacker over the air for the whole time it takes to conduct the attack. It is not known how long the attack would take when conducted over the air. Estimates vary from eight to thirteen hours.

The attack might be conducted in a subway, where the signal of the legitimate BTS is not available, but the phone is still turned on. The subscriber would be unaware of such an attack though the fact that the battery of the phone has run out slightly quicker than usual might make him suspicious.

The attack can also be performed in parts: instead of performing an eight-hour attack, the attacker could tease the phone for twenty minutes every day on the victim's way to work. Once the SIM is cloned, the SIM-clone is usable until the subscriber gets a new SIM, which in practice does not happen very often.

In another scenario, the subscriber is on a business trip in another country. The attacker has somehow bullied the local GSM operator to perform this attack on the subscriber's phone.

The attacker would again be able to reconstruct the Ki based on the MS's SRES answers and the attack would probably go unnoticed, because the challenges originate from a legitimate network. Keep in mind that the local network does not know anything about the Ki, because the triples originate from the HLR of the subscribers home network. Thus, the local network has to deduce the Ki from the A3 responses.

2.5.6 Retrieving the Key from the AuC

The same attack used in retrieving the Ki from a SIM card can be used to retrieve the Ki from the AuC. The AuC has to answer to requests made by the GSM network and return valid triples to be used in MS authentication. The procedure is basicly identical to the procedure used in the MS to access the SIM card. The difference is that the AuC is a lot faster in processing requests than a SIM card is, because it needs to process a lot more requests compared to one SIM card.

2.5.7 Cracking the A8 Algorithm

Another possibility is that someone will be able to crack the A8 key generation algorithm and retrieve the secret key, Ki, based on the random challenge, RAND, the session key, Kc, and the SRES response (assuming the same algorithm is used for both A3 and A8 as is the case with COMP128) with a minimal amount of work. For example, the attacker may find a RAND that produces the Ki as a result (an over simplified example). All three variables are obtained relatively easily. The RAND and SRES are sent over the air in plain text and the session key Kc can be relatively easily deduced from the encrypted frames and the known plain text given enough time. A vulnerability like this in the key generation algorithm would of course devastate the whole GSM security model and give the GSM Consortium something to think about when designing their next security algorithm.

3. SECURITY FUNCTIONS IN GSM

The different security related services and functions in GSM are grouped as follows:

- Subscriber identity confidentiality.
- Subscriber identity authentication.
- Signaling information element and connectionless user data confidentiality and data confidentiality for physical connections (ciphering).

It shall be possible to introduce new authentication and ciphering algorithms during the systems life time. The fixed network may support more than one authentication and ciphering algorithm.

The security procedures include mechanisms to enable recovery in event of signaling failures. These recovery procedures are designed to minimize the risk of a breach in the security of the system.

3.1 Subscriber identity confidentiality

3.1.1 Generality

The purpose of this function is to avoid the possibility for an intruder to identify which subscriber is using a given resource on the radio path (e.g. TCH (Traffic Channel) or signaling resources) by listening to the signaling exchanges on the radio path. This allows both a high level of confidentiality for user data and signaling and protection against the tracing of a user's location.

The provision of this function implies that the IMSI (International Mobile Subscriber Identity), or any information allowing a listener to derive the IMSI easily, should not normally be transmitted in clear text in any signaling message on the radio path.

Consequently, to obtain the required level of protection, it is necessary that:

- a protected identifying method is normally used instead of the IMSI on the radio path; and

- the IMSI is not normally used as addressing means on the radio path ;

- When the signaling procedures permit it, signaling information elements that convey information about the mobile subscriber identity must be ciphered for transmission on the radio path.

3.1.2 Identifying method

The means used to identify a mobile subscriber on the radio path consists of a TMSI (Temporary Mobile Subscriber Identity). This TMSI is a local number, having a meaning only in a given location area; the TMSI must be accompanied by the LAI (Location Area Identification) to avoid ambiguities.

The network (e.g. a VLR) manages suitable data bases to keep the relation between TMSIs and IMSIs. When a TMSI is received with an LAI that does not correspond to the current VLR, the IMSI of the MS must be requested from the VLR in charge of the indicated location area if its address is known; otherwise the IMSI is requested from the MS.

A new TMSI must be allocated at least in each location updating procedure. The allocation of a new TMSI corresponds implicitly for the MS to the de-allocation of the previous one. In the fixed part of the network, the cancellation of the record for an MS in a VLR implies the de-allocation of the corresponding TMSI.

To cope with some malfunctioning, e.g. arising from a software failure, the fixed part of the network can require the identification of the MS in clear. This procedure is a breach in the provision of the service, and should be used only when necessary.

When a new TMSI is allocated to an MS, it is transmitted to the MS in a ciphered mode. The MS must store its current TMSI in a non volatile memory, together with the LAI, so that these data are not lost when the MS is switched off.

3.1.3 Procedures

This presents the procedures, or elements of procedures, pertaining to the management of TMSIs.

3.1.3.1 Location updating in the same MSC area

This procedure is part of the location updating procedure which takes place when the original location area and the new location area depend on the same MSC. The part of this procedure relative to TMSI management is reduced to a TMSI re-allocation (from TMSIo with "o" for "old" to TMSIn with "n" for "new").

The MS sends TMSIo as an identifying field at the beginning of the location updating procedure. The procedure is schematized in figure 3.1.

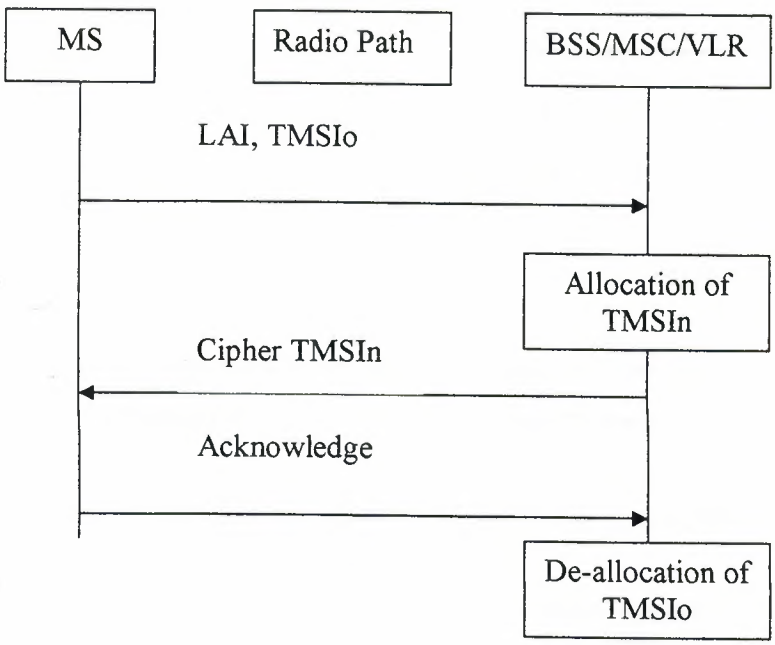

Figure 3.1 Location updating in the same MSC area

Signaling Functionalities: The MS and BSS/MSC/VLR agree on means for ciphering signaling information elements, in particular to transmit TMSIn.

3.1.3.2 Location updating in a new MSCs area, within the same VLR area

This procedure is part of the location updating procedure which takes place when the original location area and the new location area depend on different MSCs, but on the same VLR. The procedure is schematized on figure 3.2.

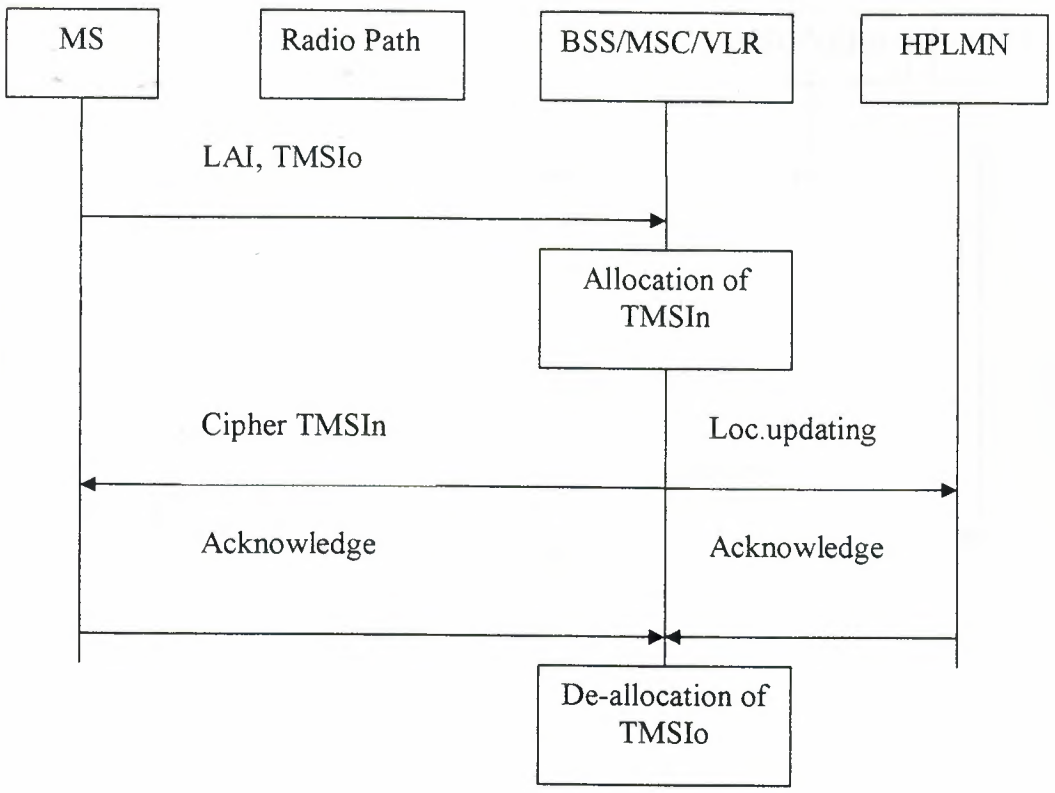

Figure 3.2 Location updating in a new MSCs area, within the same VLR area

Signaling functionalities: Loc.Updating: stands for Location Updating The BSS/MSC/VLR indicates that the location of the MS must be updated.

3.1.3.3 Location updating in a new VLR; old VLR reachable

This procedure is part of the normal location updating procedure, using TMSI and LAI, when the original location area and the new location area depend on different VLRs. The MS is still registered in VLRo ("o" for old or original) and requests registration in VLRn ("n" for new).

LAI and TMSIo are sent by MS as identifying fields during the location updating procedure. The procedure is schematized in figure 3.3.

The MSC/VLRn needs some information for authentication and ciphering; this information is obtained from MSC/VLRo.

Cancellation: The HLR indicates to VLRo that the MS is now under control of another VLR. The "old" TMSI is free for allocation.

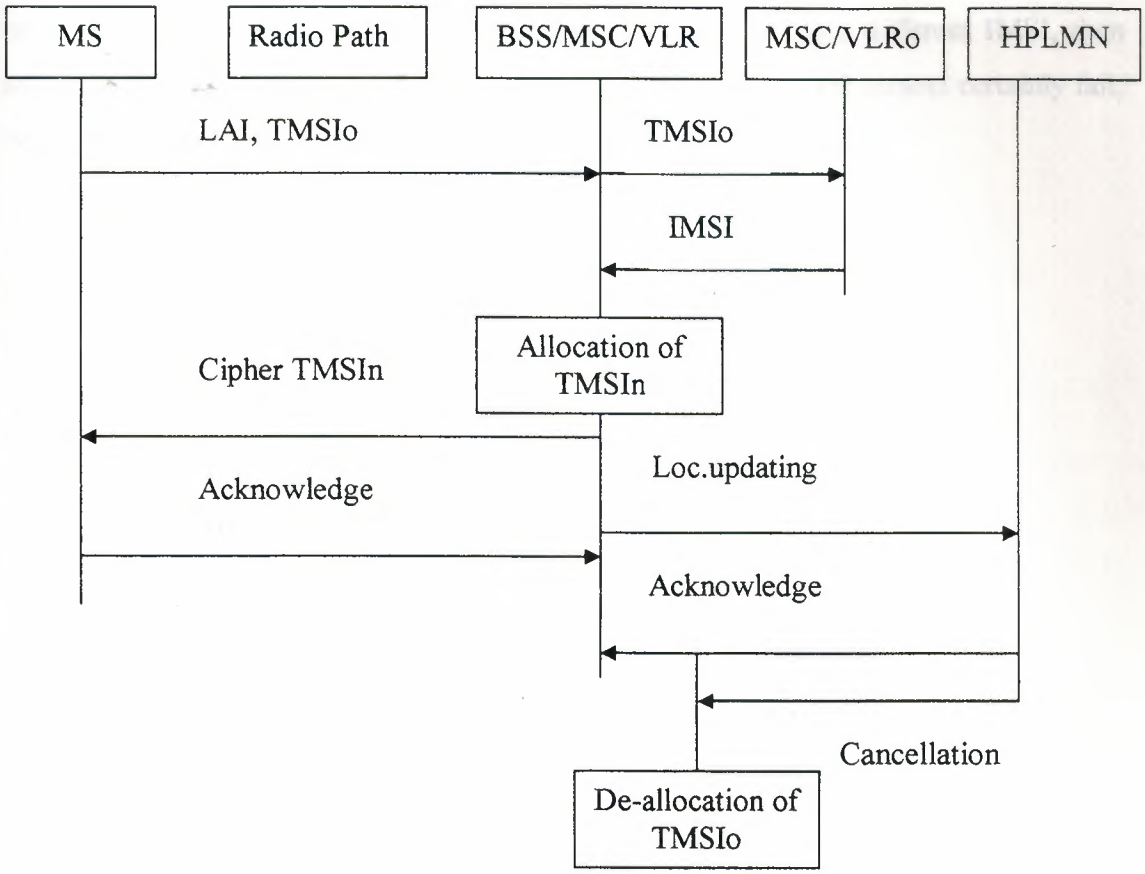

Figure 3.3 Location updating in a new VLR; old VLR reachable

3.1.3.4 Reallocation of a new TMSI

This function can be initiated by the network whenever a radio connection exists. The procedure can be included in other procedures, e.g. through the means of optional parameters. The execution of this function is left to the network operator.

When a new TMSI is allocated to an MS the network must prevent the old TMSI from being allocated again until the MS has acknowledged the allocation of the new TMSI. If an IMSI record is deleted in the VLR by O&M action, the network must prevent any TMSI associated with the deleted IMSI record from being allocated again until a new TMSI is successfully allocated to that IMSI.

If an IMSI record is deleted in the HLR by O&M action, it is not possible to prevent any TMSI associated with the IMSI record from being allocated again. However, if the MS whose IMSI record was deleted should attempt to access the

network using the TMSI after the TMSI has been allocated to a different IMSI, then authentication or ciphering of the MS whose IMSI was deleted will almost certainly fail, which will cause the TMSI to be deleted from the MS.

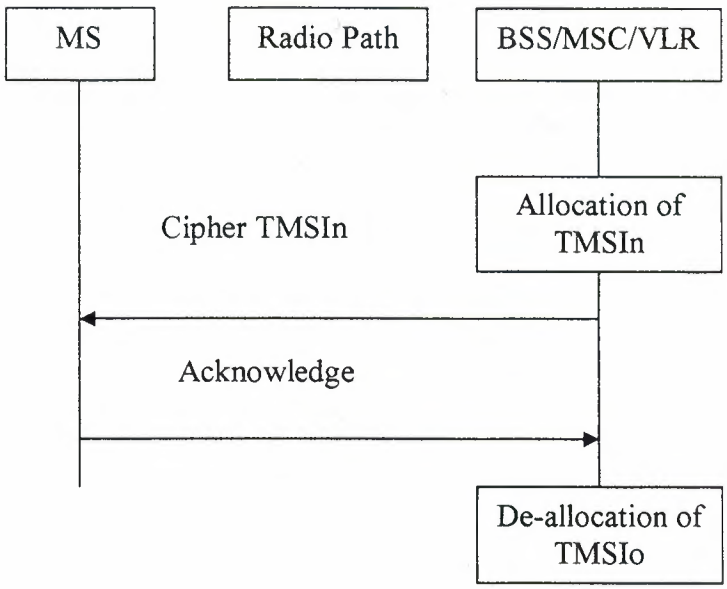

Figure 3.4 Reallocation of a new TMSI.

3.1.3.5 Unsuccessful TMSI allocation

If the MS does not acknowledge the allocation of a new TMSI, the network shall maintain the association between the old TMSI and the IMSI and between the new TMSI and the IMSI.

For an MS-originated transaction, the network shall allow the MS to identify itself by either the old TMSI or the new TMSI. This will allow the network to determine the TMSI stored in the MS; the association between the other TMSI and the IMSI shall then be deleted, to allow the unused TMSI to be allocated to another MS.

For a network-originated transaction, the network shall identify the MS by its IMSI. When radio contact has been established, the network shall instruct the MS to delete any stored TMSI. When the MS has acknowledged this instruction, the network shall delete the association between the IMSI of the MS and any TMSI; this will allow the released TMSIs to be allocated to another MS.

In either of the cases above, the network may initiate the normal TMSI reallocation procedure. Repeated failure of TMSI reallocation (passing a limit set by the operator) may be reported for O&M action.

3.1.3.6 Combined location area updating with the routing area updating

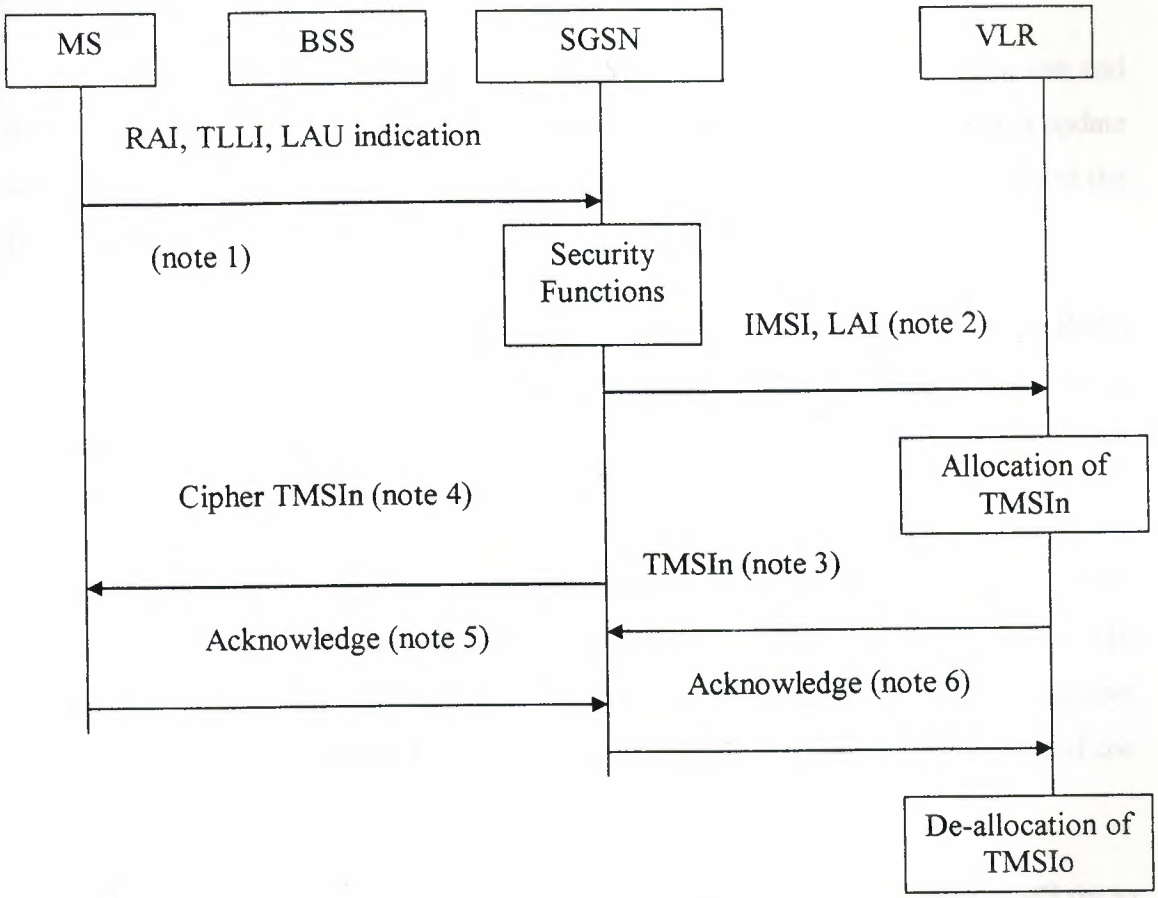

Figure 3.5 Combined routing area and location updating in the same VLR.

This subclause is only applicable if GPRS is supported.

NOTE 1: The Routeing Area Update Request message including the old Routing Area Identifier (RAI), the Temporary Logical Link Identifier (TLLI), and an indication that a combined Location Area Update (LAU) is performed.

NOTE 2: Location Updating message.

NOTE 3: Location Updating Accept message including the new TMSI.

NOTE 4: Routing Area Update Accept message including the new TMSI and the new TLLI (if any).

NOTE 5: Routing Area Update Complete message including the TLLI and TMSI.

NOTE 6: TMSI Reallocation Complete message including the TMSI.

When the VLR does not change the TMSI, the old TMSI will stay in use and there is no need to send any TMSI to the MS. In case of combined routing area update and inter-VLR location area updating procedure, the old TMSI will be cancelled and the HLR is updated.

This procedure is part of the location updating of a General Packet Radio Service (GPRS) class A or B mobile when the Gs-interface (SGSN MSC/VLR signaling interface) is implemented. This procedure is not relevant if the Gs-interface is not implemented.

The location area updating procedure and the routing area updating procedure are combined to one MS Serving GPRS Support Node (SGSN) procedure. The MS includes a Location Area Update (LAU) indication in the Routing Area Update Request message. The SGSN performs the location updating towards the VLR on behalf of the MS.

If the Location Updating message indicates a reject (if for example the MS try to enter a forbidden location area), then this should be indicated to the MS and the MS shall not access non-GPRS service until a successful Location Update is performed. For the combined location and routing area update and the combined GPRS Attach and IMSI Attach for GPRS class A and B mobiles, the authentication is performed by the SGSN. The MSC/VLR relies on the SGSN authentication. This authentication procedure generates no ciphering key for circuit switched ciphering.

The ciphering key for circuit switched operation is allocated through an authentication by MSC/VLR when the circuit switched service is requested. Also, the MSC/VLR may use the old ciphering key if existing.

3.1.3.7 Local TMSI unknown

This procedure happens when a data loss has occurred in a VLR and when a MS uses an unknown TMSI, e.g. for a communication request or for a location updating request in a location area managed by the same VLR. This procedure is schematized in figure 3.6.

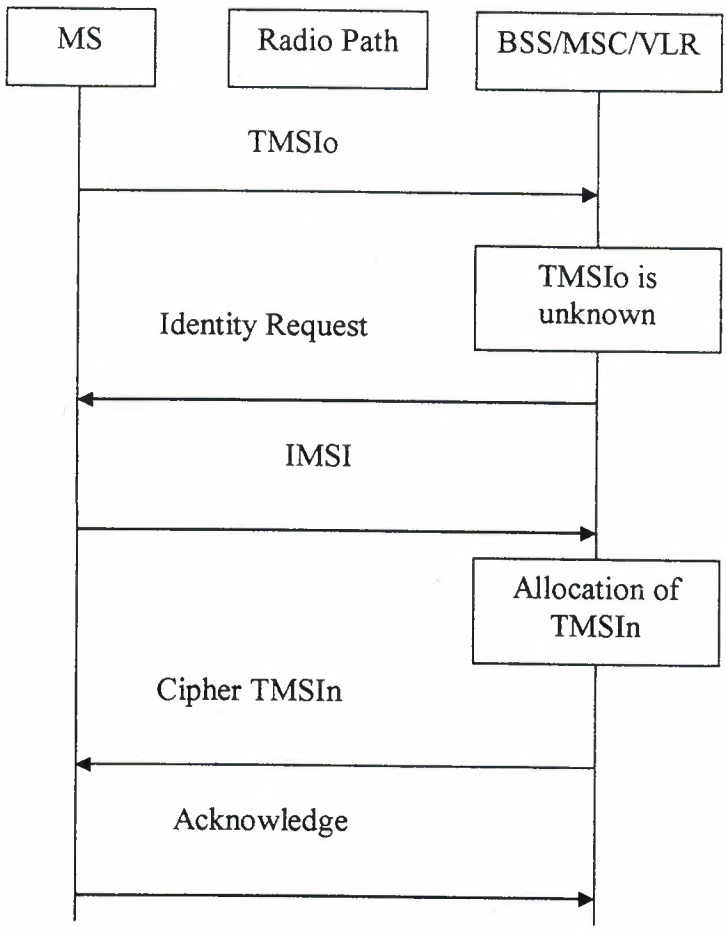

Figure 3.6 Location updating in the same MSC area; local TMSI unknown

3.1.3.8 Location Updating in a new VLR; old VLR not reachable

This procedure arises when the VLR receiving the LAI and TMSIo cannot identify the VLRo. In that case the relation between TMSIo and IMSI is lost, and the identification of the MS in clear is necessary.

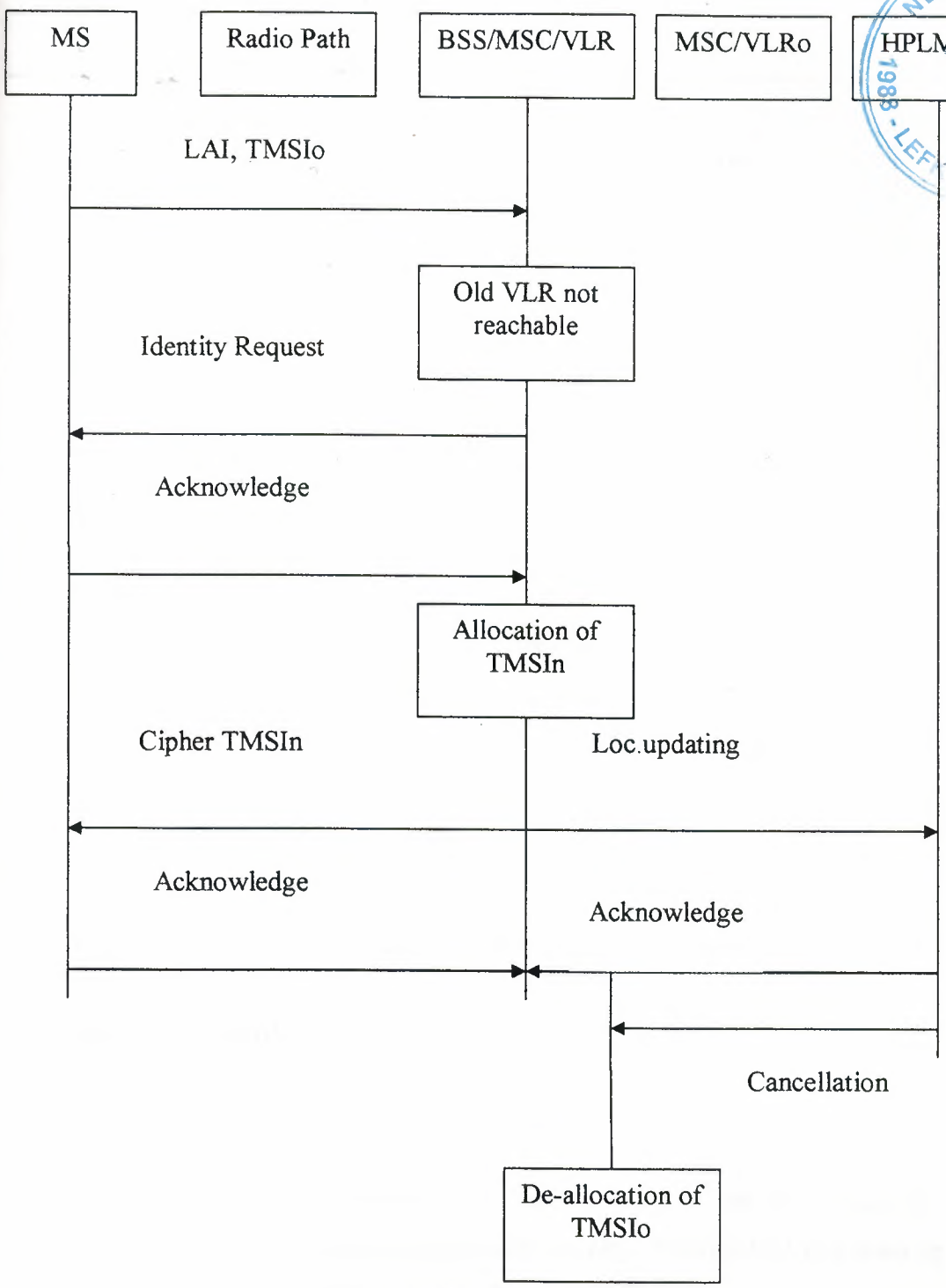

Figure 3.7 Location Updating in a new VLR; old VLR not reachable.

3.1.3.9 Location updating in a new VLR in case of a loss of information

This procedure arises when the VLR in charge of the MS has suffered a loss of data. In that case the relation between TMSIo and IMSI is lost, and the identification of the MS in clear is necessary. The procedure is schematized in figure 3.8.

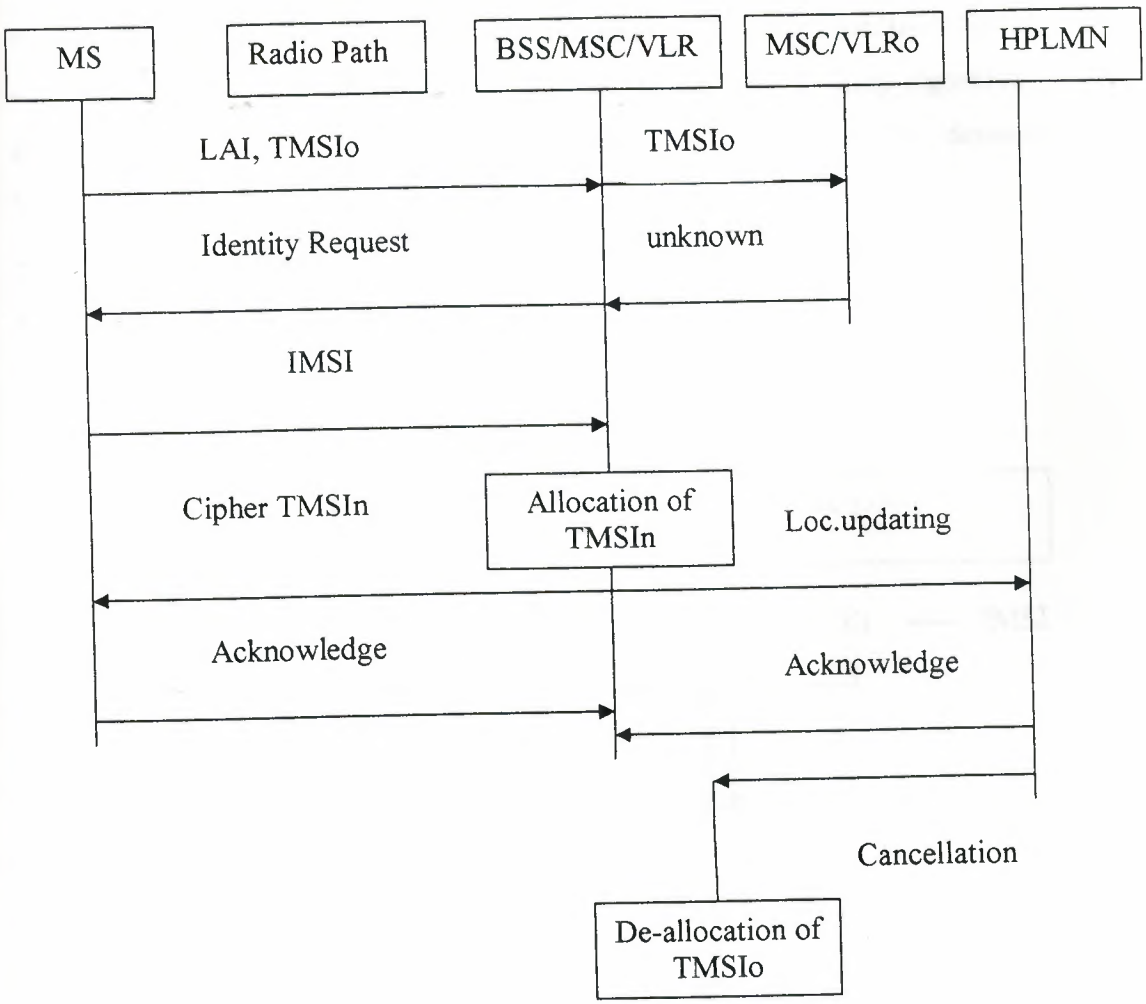

Figure 3.8 Location updating in a new VLR in case of a loss of information

3.2 Subscriber identity authentication

3.2.1 Generality

The authentication procedure will also be used to set the ciphering key. Therefore, it is performed after the subscriber identity (TMSI/IMSI) is known by the network and before the channel is encrypted.

Two network functions are necessary: the authentication procedure itself, and the key management inside the fixed subsystem.

3.2.2 The authentication procedure

The authentication procedure consists of the following exchange between the fixed subsystem and the MS.

- The fixed subsystem transmits a non-predictable number RAND to the MS.
- The MS computes the signature of RAND, say SRES, using algorithm A3 and some secret information: the Individual Subscriber Authentication Key, denoted below by Ki.
- The MS transmits the signature SRES to the fixed subsystem.
- The fixed subsystem tests SRES for validity.

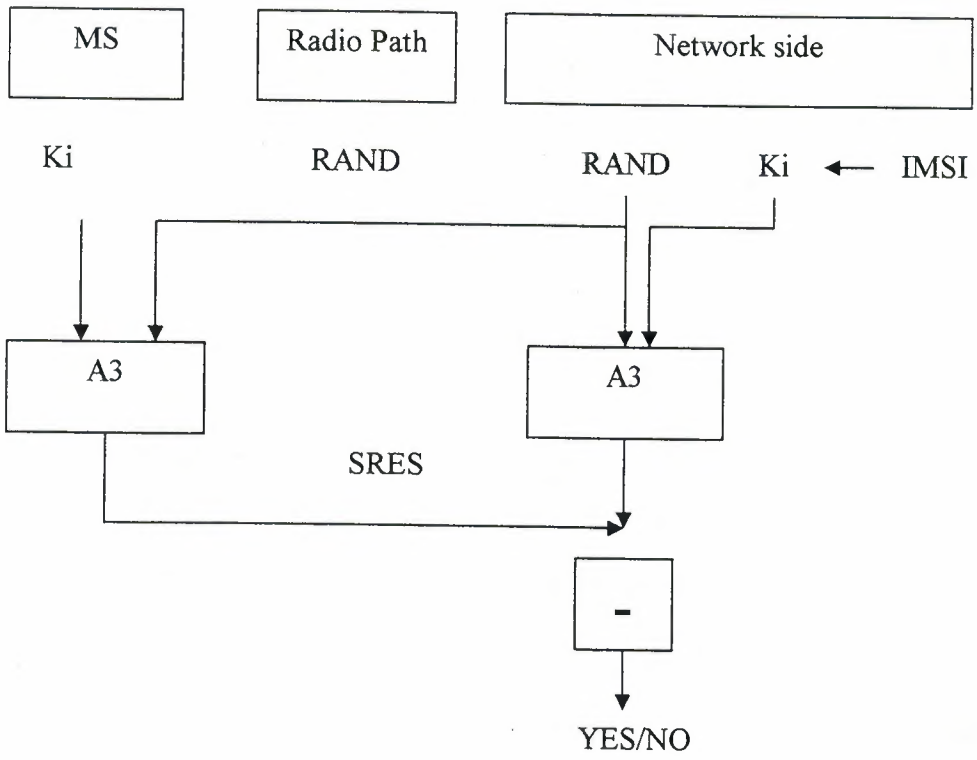

Figure 3.9 The authentication procedure

3.2.3 Subscriber Authentication Key management

The Subscriber Authentication Key Ki is allocated, together with the IMSI, at subscription time.

Ki is stored on the network side in the Home Public Land Mobile Network (HPLMN), in an Authentication Centre (AuC). A PLMN may contain one or more AuC. An AuC can be physically integrated with other functions, e.g. in a Home Location Register (HLR).

3.2.3.1 General authentication procedure

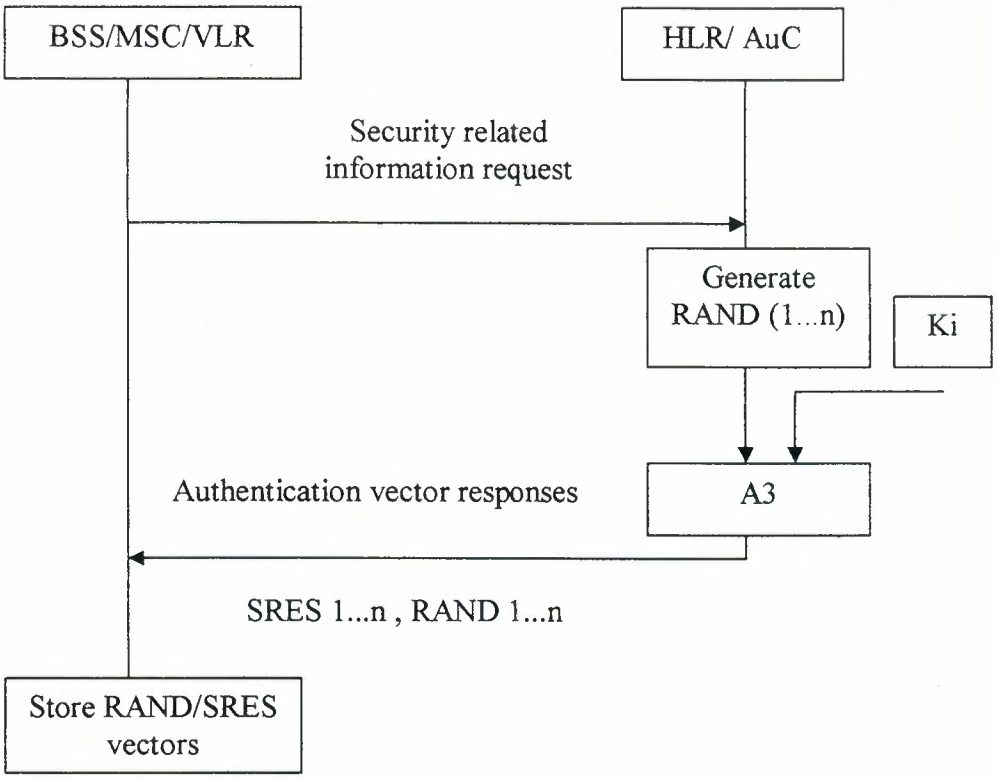

Figure 3.10 Procedure for updating the vectors RAND/SRES.

When needed for each MS, the BSS/MSC/VLR requests security related information from the HLR/AuC corresponding to the MS. This includes an array of pairs of corresponding RAND and SRES. These pairs are obtained by applying Algorithm A3 to each RAND and the key Ki as shown in figure 3.10. The pairs are stored in the VLR as part of the security related information. The procedure used for updating the vectors RAND/SRES is schematized in figure 3.10.

NOTE: The Authentication Vector Response contains also Kc(1..n) which is not shown in this and the following figures.

When an MSC/VLR performs an authentication, including the case of a location updating within the same VLR area, it chooses a RAND value in the array corresponding to the MS. It then tests the answer from the MS by comparing it with the corresponding SRES, as schematized in figure 3.11

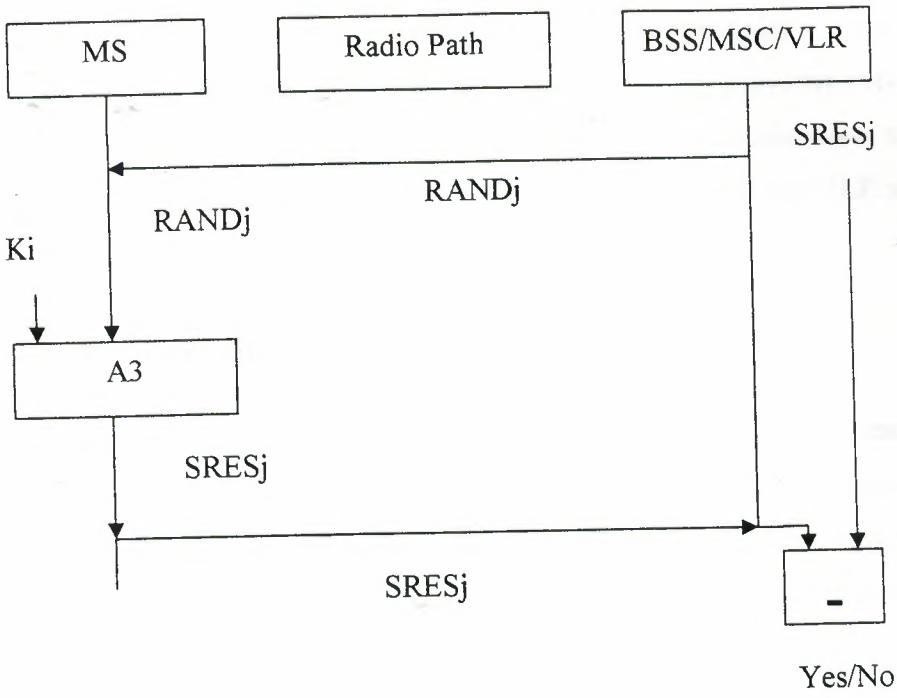

Figure 3.11 General authentication procedure

3.2.3.2 Authentication at location updating in a new VLR, using TMSI

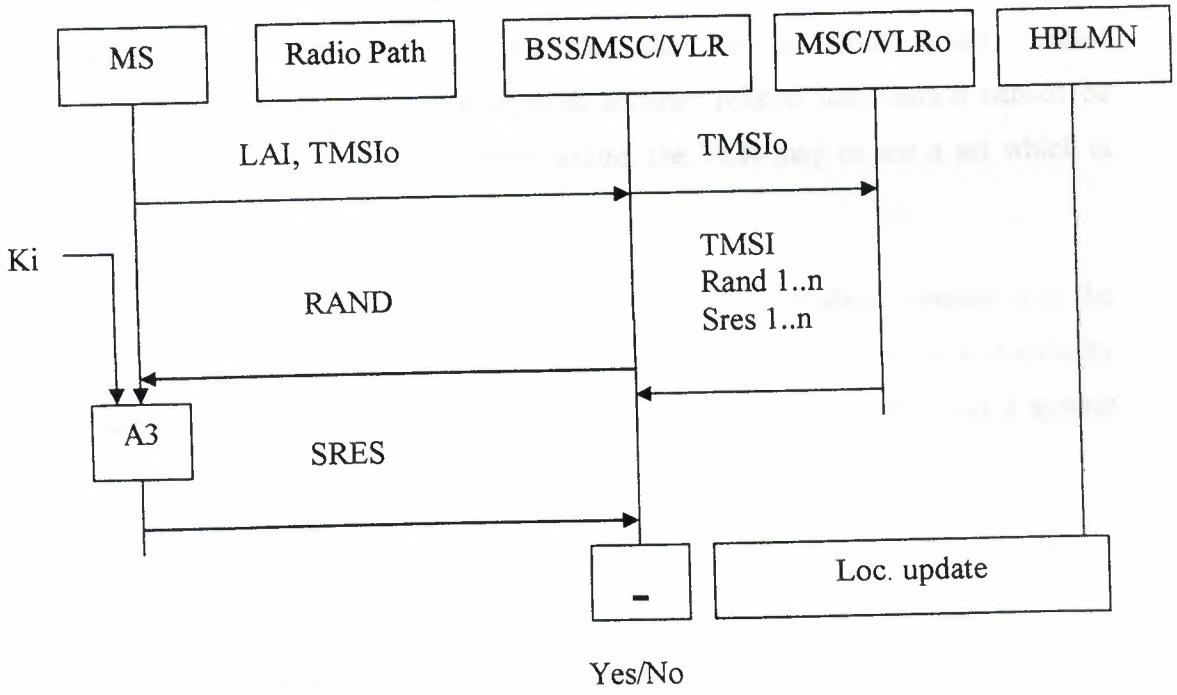

Figure 3.12 Authentication at location updating in a new VLR, using TMSI.

During location updating in a new VLR (VLRn), the procedure to get pairs for subsequent authentication may differ from that described in the previous subclause. In the case when identification is done using TMSI, pairs for authentication as part of security related information are given by the old VLR (VLRo). The old VLR shall send to the new VLR only those pairs which have not been used

3.2.3.3 Authentication with IMSI if authentication with TMSI fails

If authentication of an MS which identifies itself with a TMSI is unsuccessful, the network requests the IMSI from the MS, and repeats the authentication using the IMSI. Optionally, if authentication using the TMSI fails the network may reject the access request or location registration request which triggered the authentication.

3.2.3.4 Re-use of security related information in failure situations

Security related information consisting of sets of RAND, SRES and Kc is stored in the VLR and in the HLR. When a VLR has used a set of security related information to authenticate an MS, it shall delete the set of security related information or mark it as used. When a VLR needs to use security related information, it shall use a set which is not marked as used in preference to a set which is marked as used; if there are no sets which are not marked as used then the VLR shall request fresh security related information from the HLR. If a set of fresh security related information cannot be obtained in this case because of a system failure, the VLR may re-use a set which is marked as used.

"System failure" means that the VLR was unable to establish contact with the HLR, or the HLR returned a positive acknowledgement containing no sets of security related information, or the HLR returned an error indicating that there was a system failure or that the request was badly formatted.

If the HLR responds to a request for security related information with an indication that the subscriber is unknown or barred in the HLR, the VLR shall not re-use security information which has been marked as used. It is an operator option to define how many times a set of security related information may be re-used in the VLR; when a set of security related information has been re-used as many times as is permitted by the operator, it shall be deleted.

If a VLR successfully requests security related information from the HLR, it shall discard any security related information which is marked as used in the VLR.

If a VLR receives from another VLR a request for security related information, it shall send only the sets which are not marked as used.

If an HLR receives a request for security related information, it shall send any sets which are not marked as used; those sets shall then be deleted or marked as used. If there are no sets which are not marked as used, the HLR may as an operator option send sets which are marked as used. It is an operator option to define how many times a set of security related information may be re-sent by the HLR; when a set of security related information has been sent as many times as is permitted by the operator, it shall be deleted.

3.3 Confidentiality of signaling information elements

Connectionless data and user information elements on physical connections.

3.3.1 Generality

In GSM, some signaling information elements are considered sensitive and must be protected.

To ensure identity confidentiality, the Temporary Subscriber Identity must be transferred in a protected mode at allocation time and at other times when the signaling procedures permit it.

The confidentiality of connectionless user data requires at least the protection of the message part pertaining to OSI layers 4 and above. The user information confidentiality of user information on physical connections concerns the information transmitted on a traffic channel on the MS-BSS interface (e.g. for speech). It is not an end-to-end confidentiality service.

These needs for a protected mode of transmission are fulfilled with the same mechanism where the confidentiality function is an OSI layer 1 function. The scheme described below assumes that the main part of the signaling information elements is

transmitted on DCCH (Dedicated Control Channel), and that the CCCH (Common Control Channel) is only used for the allocation of a DCCH.

Four points have to be specified:

- the ciphering method
- the key setting
- the starting of the enciphering and deciphering processes
- the synchronization

3.3.2 The ciphering method

The layer 1 data flow (transmitted on DCCH or TCH) is ciphered by a bit per bit or stream cipher, i.e. the data flow on the radio path is obtained by the bit per bit binary addition of the user data flow and a ciphering bit stream, generated by algorithm A5 using a key determined. The key is denoted below by Kc, and is called "Ciphering Key".

For multislot configurations (e.g. HSCSD) different ciphering bit streams are used on the different timeslots. On timeslot "n" a ciphering bit stream, generated by algorithm A5, using a key Kcn is used. Kcn is derived from Kc as follows:

- Let BN denote a binary encoding onto 64 bits of the timeslot number "n" (range 0-7). Bit "i" of Kcn, Kcn(i), is then calculated as Kc(i) xor (BN<<32(i)) ("xor" indicates: "bit per bit binary addition" and "<<32" indicates: "32 bit circular shift"), the number convention being such that the lsb of Kc is xored with the lsb of the shifted BN.

Deciphering is performed by exactly the same method.

3.3.3 Key setting

Mutual key setting is the procedure that allows the mobile station and the network to agree on the key Kc to use in the ciphering and deciphering algorithms A5.

A key setting is triggered by the authentication procedure. Key setting may be initiated by the network as often as the network operator wishes.Key setting must occur

on a DCCH not yet encrypted and as soon as the identity of the mobile subscriber (i.e. TMSI or IMSI) is known by the network.

The transmission of Kc to the MS is indirect and uses the authentication RAND value; Kc is derived from RAND by using algorithm A8 and the Subscriber Authentication key Ki.

The values Kc are computed together with the SRES values. The security related information consists of RAND, SRES and Kc. The key Kc is stored by the mobile station until it is updated at the next authentication.

Key setting is schematized in figure 3.13

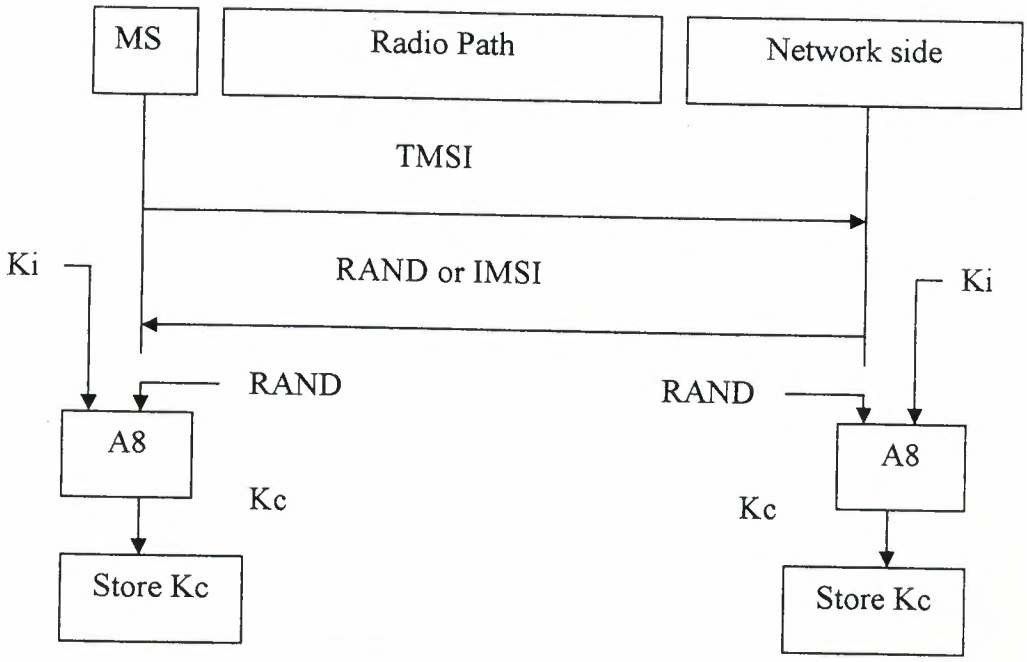

Figure 3.13 Key setting.

3.3.4 Ciphering key sequence number

The ciphering key sequence number is a number which is associated with the ciphering key Kc and they are stored together in the mobile station and in the network.

However it is not directly involved in any security mechanism.

3.3.5 Starting of the ciphering and deciphering processes

The MS and the BSS must co-ordinate the instants at which the enciphering and deciphering processes start on DCCH and TCH. On DCCH, this procedure takes place under the control of the network some time after the completion of the authentication procedure (if any), or after the key Kc has been made available at the BSS.

No information elements for which protection is needed must be sent before the ciphering and deciphering processes are operating.

The transition from clear text mode to ciphered mode proceeds as follows: deciphering starts in the BSS, which sends in clear text to the MS a specific message, here called "Start cipher".

Both the enciphering and deciphering start on the MS side after the message "Start cipher" has been correctly received by the MS. Finally, enciphering on the BSS side starts as soon as a frame or a message from the MS has been correctly deciphered at the BSS.

The starting of enciphering and deciphering processes is schematized in figure 3.14

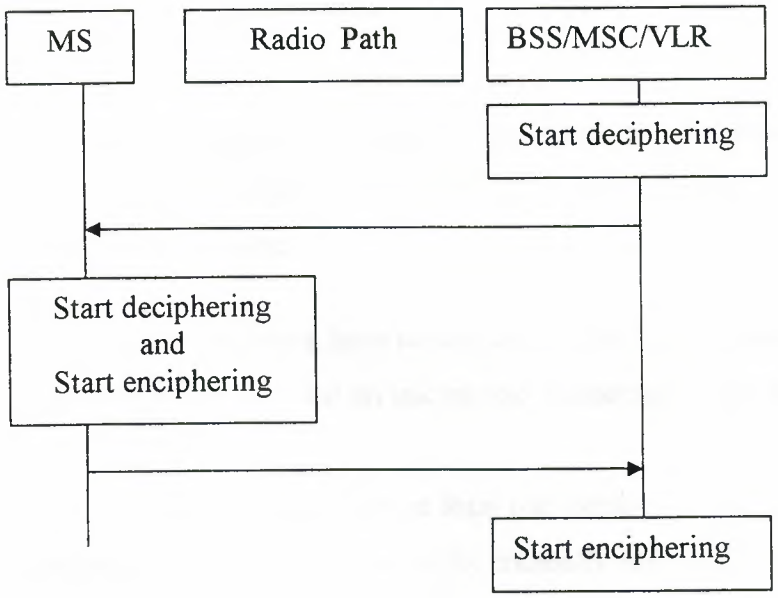

Figure 3.14 Starting of the enciphering and deciphering processes

When a TCH is allocated for user data transmission, the key used is the one set during the preceding DCCH session(Call Set-up). The enciphering and deciphering processes start immediately.

3.3.6 Synchronization

The enciphering stream at one end and the deciphering stream at the other end must be synchronized, for the enciphering bit stream and the deciphering bit streams to coincide.

3.3.7 Handover

When a handover occurs, the necessary information (e.g. key Kc, initialization data) is transmitted within the system infrastructure to enable the communication to proceed from the old BSS to the new one, and the Synchronization procedure is resumed. The key Kc remains unchanged at handover.

3.3.8 Negotiation of A5 algorithm

There are seven versions of the A5 algorithm defined. When an MS wishes to establish a connection with the network, the MS shall indicate to the network which of the seven versions of the A5 algorithm it supports. The network shall not provide service to an MS which indicates that it does not support the ciphering algorithm(s) required by GSM.

The network shall compare its ciphering capabilities and preferences, and any special requirements of the subscription of the MS, with those indicated by the MS and act according to the following rules:

- If the MS and the network have no versions of the A5 algorithm in common and the network is not prepared to use an unciphered connection, then the connection shall be released.
- If the MS and the network have at least one version of the A5 algorithm in common, then the network shall select one of the mutually acceptable versions of the A5 algorithm for use on that connection.

- If the MS and the network have no versions of the A5 algorithm in common and the network is willing to use an unciphered connection, then an unciphered connection shall be used.

4. GPRS SECURITY VS. GSM SECURITY

4.1 GPRS Security

In the GPRS system, the frames are transmitted as cipher text from the MS to the SGSN. This is done because the GPRS system uses multiple timeslots in parallel in order to achieve a greater transmission rate. One GPRS phone can be allocated multiple timeslots by the network, thus increasing the transmission rate of that MS. The frames can be sent in 'parallel' timeslots to the same BTS or to two different BTSs if the MS is handed over from one BTS to another.

To a BTS the use of one timeslot is seen as a separate call. Thus, the BTS is unable to put the frames from different timeslots together. This means that there has to be a network component that is able to receive the frames from one MS, defragment them and send them onwards to the actual destination. The BTSs are also unable to decrypt the frames, because consecutive frames on one channel have not got consecutive frame numbers. See Figure 4.1.

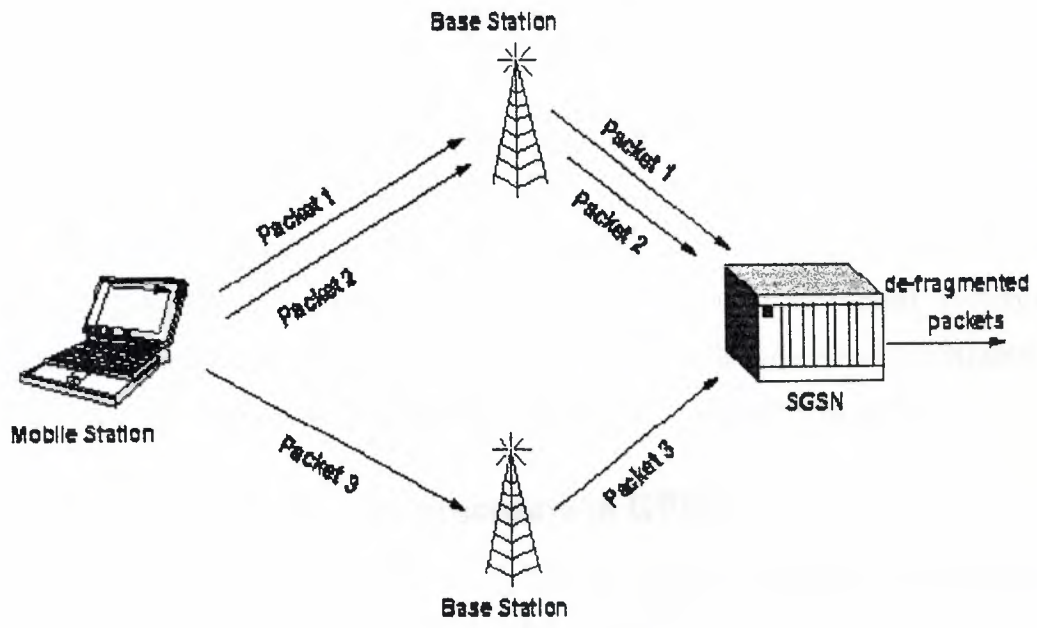

Figure 4.1 GPRS architecture

To simplify the implementation, the frames are decrypted at the SGSN where all of the frames end up and it is thus easy to keep track of frame numbers. The solution is

based on the ease of implementation and has not been implemented in order to increase system security. As a side effect, the GPRS system effectively prevents eavesdropping on the backbone between the BTS and SGSN, because the frames are still encrypted at this point. In GPRS, the triples from the HLR are transmitted to the SGSN and not to the MSC. Thus, security of GPRS depends largely on the placement and security of the SGSNs.

The GPRS system uses a new A5 implementation as well, which is not known publicly. This and the fact that the frames are not decrypted at the BTS, but at the SGSN, rules out a couple of attacks. First, it is very hard to attack the 5 implementation when it is not known. Secondly, the Kc is not transmitted to the BTSs and the transmission channel between the BTS and the SGSN is encrypted making it thus useless to monitor the backbone between the BTS and the SGSN.

This does not mean that the GPRS security model would somehow be more secure than the GSM-only security model. It means that identical attacks do not work with GPRS that work with a GSM-only network. As soon as the A5 implementation used in GPRS leaks out, the GPRS security model is vulnerable to new attacks. And the implementation will leak out eventually or the design is successfully reverse-engineered. As was states above, the security of a crypto system should be based solely on the key.

However the majority of the attacks against the GSM-only system are applicable against GPRS as well. E.g. the SIM-cloning attack. Additionally, the GPRS model introduces another security threat through the use of SGSNs, which know the triples from the HLR. This means that the security of the GPRS network depends largely on the positions of the SGSNs in the network architecture and the security of the SGSNs. If the SGSNs are vulnerable to an attack, then the triples are vulnerable as well.

4.2 General authentication procedure in GPRS

When needed, the SGSN requests security related information for a MS from the HLR/AuC corresponding to the IMSI of the MS. This includes an array of pairs of corresponding RAND and SRES. These pairs are obtained by applying Algorithm A3 to each RAND and the key Ki. The pairs are stored in the SGSN as part of the security related information. See figure 4.2.

When an SGSN performs an authentication, including the case of a routing area updating within the same SGSN area, it chooses a RAND value in the array corresponding to the MS. It then tests the answer from the MS by comparing it with the corresponding SRES. See figure 4.3.

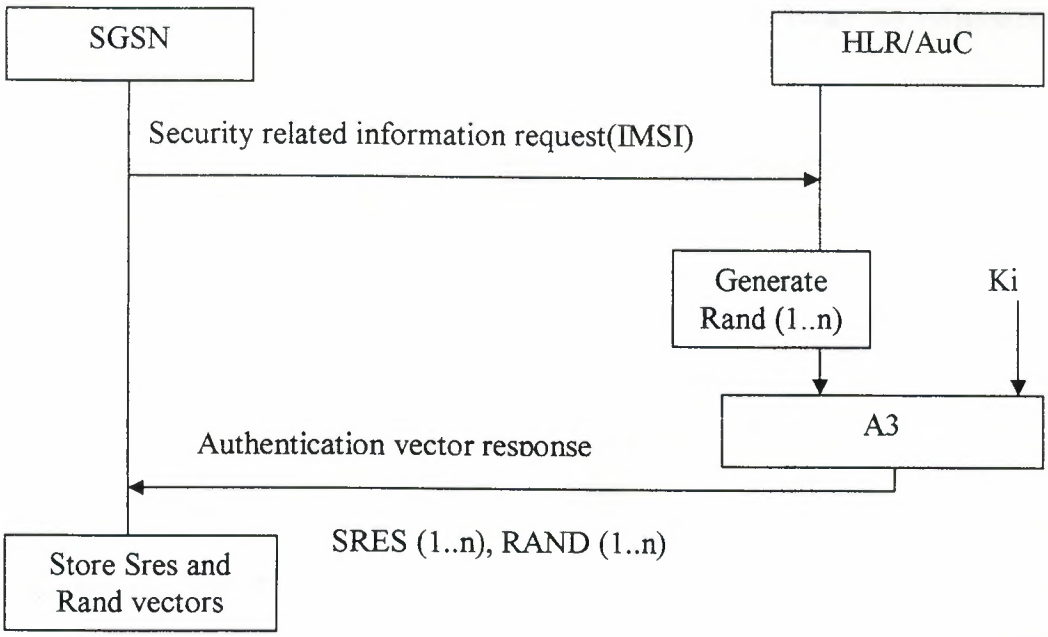

Figure 4.2 Procedure for updating the vectors RAND/SRES

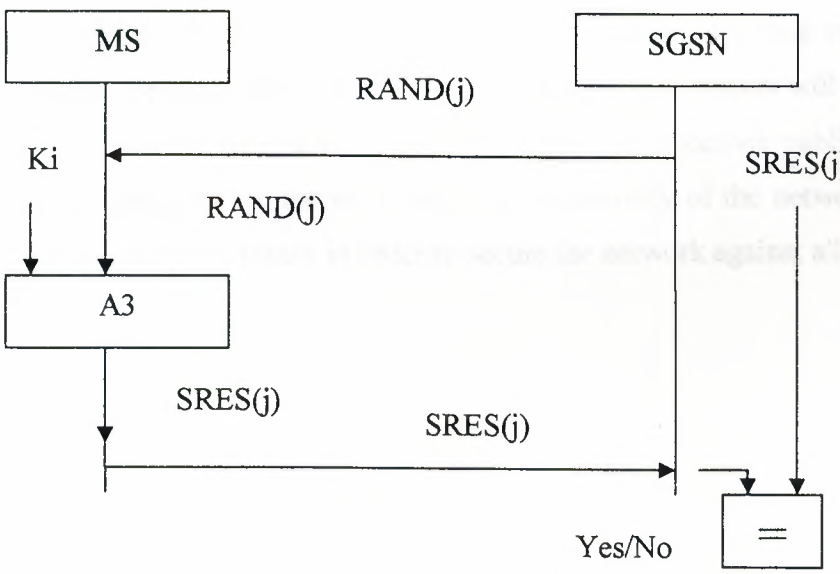

Figure 4.3 General authentication procedure

4.3 Possible Improvements

Security could be improved in some areas with relatively simple measures. The operator could 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. This would also be the easiest improvement introduced here, because the network operator can make the changes itself and does not need the support of hardware or software manufacturers or the GSM Consortium.

Another solution would be to employ a new A5 implementation with strong encryption so that a brute-force attack is not feasible in any case. This would disable the attacker from recording transmitted frames and cracking them in his spare time. This improvement would require the cooperation of the GSM Consortium. The hardware and software manufacturers would have to release new versions of their software and hardware that would comprise with the new A5 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 wire tapping the backbone network. This solution could probably also be implemented without the blessings of the GSM Consortium, but the cooperation of the hardware manufacturers would still be required.

In sum, none of the improvements above are too hard to implement. They all present new expenses mostly to the network operator and are not thus very attractive from the network operator's point of view. Thus, these improvements will probably not be implemented until the insecurity of the GSM networks becomes public knowledge and the network operators are forced to improve the security of the network. All three improvements would be necessary in order to secure the network against all attacks.

CONCLUSION

The security mechanisms specified in the GSM standard make it the most secure telecommunications system available.

The use of authentication, encryption, and temporary identification numbers ensures the privacy and anonymity of the system's users as well as safeguarding the system against fraudulent use.

Even if no encryption is used the digital GSM system is inherently more secure than analog systems due to their use of speech coding, digital modulation and TDMA channel access.

Researches are still going on to produce the better algorithms possible to enhance the security of the current systems. Codes and cryptography have been taken care of them; the GSM Consortium is always looking for the right tools to use for a complete security.

The GSM system is a first approach at a true personal communication system. GSM provides a basic range of security features to ensure adequate protection for both the operator and customer.

REFERENCE

[1] Anderson Ross, A5 - The GSM Encryption Algorithm, 17.6.1994, [referred 30.9.1999] "http://chem.leeds.ac.uk/ICAMS/people/jon/a5.html".

[2] Anon., Crack A5, [referred 29.9.1999] "http://jya.com/crack-a5.htm".

[3] Anon., GSM Alliance Clarifies False & Misleading Reports of Digital Phone Cloning, [referred 29.9.1999]."http://jya.com/gsm042098.txt".

[4] Anon., GSM Cell phones Cloned, [referred 29.9.1999]. " http://jya.com/gsm-cloned.htm".

[5] Anon., GSM Cloning, [referred 24.10.1999] "http://www.isaac.cs.berkeley.edu/isaac/gsm-faq.html".

[6] Briceno M. & Goldberg I. & Wagner D., A Pedagogical Implementation of A5/1, [referred 29.9.1999]"http://www.scard.org/gsm/a51.html".

[7] Golic J. Dj., Cryptanalysis of Alleged A5 Stream Cipher, [referred 29.9.1999] " http://jya.com/a5-hack.htm".

[8] Margrave David, GSM Security and Encryption, [referred 30.9.1999] " http://www.net-security.sk/telekom/phreak/radiophone/gsm/gsm-secur/gsm-secur.html".

[9] Racal Research Ltd., GSM System Security Study, 10.6.1988, [referred 29.9.1999] " http://jya.com/gsm061088.htm".

[10] Schneier B., Applied Cryptography, 2nd Ed., Wiley, New York, 1996, 758, [referred 29.9.1999]

[11] The Telecoms Virtual Library about mobile communications. You can find information about GSM but also about other mobile communications systems. "http://www.analysys.co.uk/vlib/mobile.htm".