A Contemporary Foreword on GSM Security

Paulo S. Pagliusi

Information Security Group, Royal Holloway, University of London
Egham, Surrey TW20 0EX, UK
p.s.pagliusi@rhul.ac.uk
http://www.isg.rhul.ac.uk

Abstract. This article contains a current outline of the GSM system security, with focus on the air interface protocol. It presents the terminology and describes the GSM security operation, including its principles and features. This document also discusses the effectiveness of GSM authentication and the strength of GSM encryption. It includes therefore the most significant physical and cryptanalytic attacks on GSM security mechanisms, such as the up to date optical fault induction and partitioning attacks. GSM security features retained and enhanced for the 3G Security and further applications in network (Internet) remote access are also contemplated. This article aims primarily at contributing to a progressive research in mobile systems security and at reviewing the security solutions implemented in this area for further applications.

1 Introduction

This article contains an up to date overview of the European GSM[1] cellular phone system security, with focus on the air interface protocol. It presents the terminology and describes the GSM security operation, including its principles and features, such as subscriber identity confidentiality and authentication, stream ciphering of user traffic and user-related control data, and use of triplets and SIM module.

This document also discusses the effectiveness of GSM authentication and the strength of GSM encryption. It includes therefore the most significant physical and cryptanalytic attacks against GSM security mechanisms, for instance the new optical fault induction and partitioning attacks, and the GSM features retained and enhanced for the Third Generation Security (3GPP)[2]. Further applications in network remote access are also contemplated. This article aims primarily at contributing to a progressive research in mobile systems security and at reviewing the security solutions implemented in this area for further applications in other correlated areas, such as authentication for Internet remote access supporting ubiquitous mobility.

[1] GSM was formerly acronym for Groupe Spéciale Mobile (founded 1982). Now is acronym for Global System for Mobile Communication.

[2] 3GPP (3rd Generation Partnership Project) is a partnership project including: ETSI (Europe), ARIB & TTA (Japan), TTC (Korea) and T1P1 (USA).

1.1 Terminology

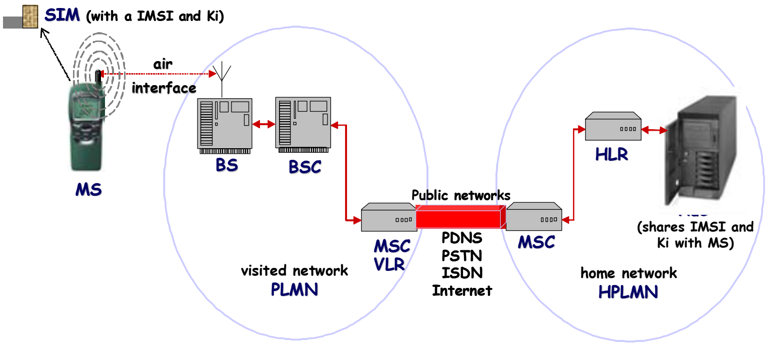

Figure 1 – GSM System Overview.

- Mobile Station (MS): a Mobile Equipment (ME or "mobile telephone") with its GSM Subscriber Identity Module (SIM). Each MS has a contractual relationship with a network, called the home network but may be allowed to roam in other visited networks when outside home network coverage area (**Figure 1**).
- International Mobile Subscriber Identity (IMSI) and Authentication Key (Ki): at the subscription time, the home network assigns the MS a unique and permanent identifier, the IMSI, together with a unique 128-bit secret key (Ki). Each customer's Ki is also stored in an Authentication Centre (AuC) in the home network. Ki plays two roles in GSM: authentication consists of proof that MS possesses Ki and encryption is performed with the use of a cipher key derived from Ki.
- GSM Subscriber Identity Module (SIM): module implemented on a smart card that must be inserted into the ME for service access. The IMSI and the authentication key Ki of the MS should be "securely stored" in the SIM.
- Public Land Mobile Network (PLMN): network that currently provides service or "is visited" by a MS. A MS is registered with the PLMN which it is currently visiting. A PLMN contains, among others components: a Base Station (BS) and a Visited Location Register (VLR).
- Base Station (BS): the Base Transceiver Station belonging to a PLMN serving the MS. Base stations form a patchwork of radio cells over a given geographic coverage area. Base Stations are connected to base station controllers (BSC).
- Base Station Controller (BSC): is a node controlling a number of BS, coordinating handovers and performing BS co-ordination not related to switching. The BSC to BS link is in many cases a point to point microwave link. BSC are also connected to mobile switching centres (MSC) via fixed or microware links. MSC are connected to the public networks (e.g. PSTN, PDNS, ISDN and Internet) [6].
- Visited Location Register (VLR): used to record information about all MS "visiting" a specific PLMN.
- Home PLMN (HPLMN): each MS has a home PLMN with which shares an IMSI and a Ki. The HPLMN and the visited PLMN have a bilateral agreement, under which the visited PLMN trusts the HPLMN to pay for the services that the visited

PLMN provides to the MS. Each HPLMN maintains a Home Location Register (HLR) and operates an Authentication Centre (AuC) to support its MS.

- Home Location Register (HLR): used to record the most recent known location of all MS belonging to a specific HPLMN.
- Authentication Centre (AuC): used by a HPLMN to generate random challenges (RAND) and to store secret key information (Ki) relating to each of its MS. The AuC can be integrated with other network functions, e.g. with the HLR.
- Air Interface: synonym for radio path. The MS 'visits' a PLMN by communicating with the serving BS across an air interface and receiving an entry in the VLR.

2 Security Features for GSM

The purpose of security for GSM system is to make the system as secure as the public switched telephone network and to prevent phone cloning. The use of air interface at the transmission media allows a number of potential threats from eavesdropping. As stated by [4], "it was soon apparent in the threat analysis that the weakest part of the system was the radio path, as this can be easily intercepted". In fact, there was no attempt to provide security on the fixed network part of GSM. And it should be noted that the GSM security was designed with three constraints in mind [13]:

- Concern of grant too much security and so bringing export problems upon GSM;
- GSM did not have to be resistant to "active attacks" where the attacker interferes with the operation of the system, perhaps masquerading as a system entity; and
- The trust between operators for the security operation should be minimized.

The technical features for security are only a small part of the GSM security requirements; the greatest threat is from simpler attacks such as disclosure of the encryption keys, insecure billing systems or even corruption. A balance is required to ensure that these security processes meet these requirements. At the same time a judgment must be made of the cost and effectiveness of the GSM security measures.

The principles of GSM security are, according to [6] and [13]:

- Subscriber identity confidentiality;
- Subscriber identity authentication;
- Stream ciphering of user traffic and user-related control data; and
- Use of SIM as security module.

It is also important to emphasize the GSM feature of use of triplets. The GSM principles and this special feature are described in the following sections.

2.1 Subscriber Identity Confidentiality

- Purpose: to avoid an interceptor of the mobile traffic being able to identify which subscriber is using a given resource on the air interface.
- Description: The use of temporary identifiers provides anonymity, so that it is not easy to identify the GSM user. This protects against the tracing of a user's loca-

tion by listening to exchanges on the radio path. Instead of using the IMSI, a new temporary mobile subscriber identity (TMSI) is allocated by the PLMN at least on every location update and used to identify a MS on the air interface.

- Operation: When a MS attempts access with a PLMN with which it is not presently registered, the MS uses its IMSI to identify itself. The IMSI is then authenticated by the PLMN, which results in the sharing of a cipher key (Kc). When the PLMN switch on encryption, the VLR generates a TMSI to the MS, storing the association of TMSI and IMSI in its database. The TMSI is then sent to the MS, encrypted with Kc. The next time the MS attempts access in that PLMN, it uses the TMSI previously allocated by the VLR instead of its IMSI. Then the PLMN looks up its table of TMSI to IMSI mapping to find the MS permanent identity. After a successful authentication and once an encrypted channel has been established, the PLMN assigns to the MS another TMSI. It is frequently given a new TMSI to that a MS cannot be previously identified and followed around. After a handover to a new VLR, or a successful re-authentication with the same VLR, the PLMN always sends a new TMSI to the MS. Then the MS stores the new TMSI and removes the association with any previously allocated TMSI. In turn, the VLR removes the association with the old TMSI and the IMSI from its database.

2.2 Subscriber Identity Authentication

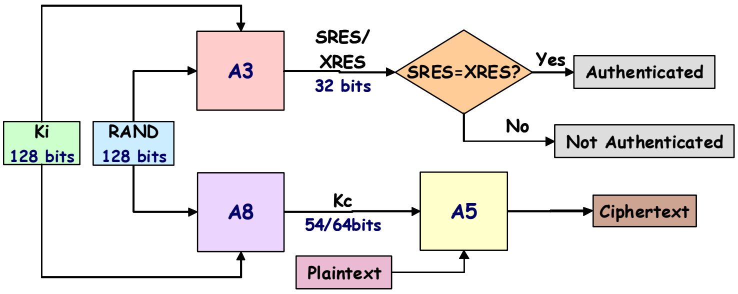

Figure 2 - GSM Authentication, Cipher Key Generation and Encryption.

- Purpose: The authentication is used to identify the MS to the PLMN operator.
- Description: It consists in the guarantee by the land based part of the system that the MS identity presented across the air interface is the real one originally embedded in the SIM. The PLMN then knows who is using the system for billing purposes. This protects the PLMN from illicit use. GSM authentication is a one-way process, i.e., the visited PLMN is not authenticated.
- Operation: Authentication is performed by a challenge and response mechanism. Ki in the HPLMN is held in the AuC. A random challenge (RAND) is generated by the AuC and issued to the MS, via PLMN. The MS encrypts RAND using Ki and the authentication algorithm A3 implemented within the SIM, and send a signed response (SRES) back to the PLMN. AuC performs the same process with RAND to compute the expected response (XRES), which is sent to the PLMN.

The PLMN now can check that the MS has Ki and that the response received is correct by comparing the value received from the HPLMN (XRES) with what it receives from MS (SRES). Eavesdropping of the radio channel should reveal no useful information, as the next time a new RAND will be used (Figure 2).

$$\begin{gathered} SRES = A3_{Ki}(RAND) \\ XRES = SRES? \end{gathered} \tag{1}$$

2.3 Stream Ciphering of User Traffic and User-Related Control Data

- Purpose: stream cipher encryption is used in GSM system to protect sensitive information against eavesdropping on the air interface.
- Description: This provides protection to the user data passing over the radio path on physical connections or connectionless and to the sensitive information on the signalling channel (e.g. phone numbers, TMSI) from eavesdropping on the air interface. Confidentiality is achieved by the use of encryption at physical layer, the choice being influenced by: speech coder, error propagation, delay and handover.
- Operation: At the same time that XRES and SRES are calculated, RAND and Ki are passed through algorithm A8 by both the MS (SIM) and the HPLMN (AuC) to derive the cipher key Kc. Then Kc is delivered from the HPLMN (AuC→HLR) to the serving PLMN (VLR→BS). Typically, algorithms A3 and A8 are combined into one called A3/8 that is residing within the SIM and the AuC. Kc is used for encrypting the signalling and messages to provide privacy through the use of A5 series algorithms (Figure 2). The BS tells ME which A5 algorithm it has (if it has one) and sends a cipher command. At the same time, the BS starts decrypting. The ME starts encrypting and decrypting when it receives the cipher command. The BS starts encrypting when it receives back the cipher command acknowledged. A fresh cipher key Kc is generated for each call. When a handover occurs during a call, the necessary information is transferred by the PLMN to the new BS, and the encryption continues using the same Kc.

$$\begin{gathered} Kc = A8_{Ki}(RAND) \\ Ciphertext = A5_{Kc}(Plaintext) \end{gathered} \tag{2}$$

2.4 Use of Triplets

- Purpose: with the use of the triplets, authentication can be performed in the 'visited' PLMN without the network operator (BS, VLR) having knowledge of Ki.
- Description: A random challenge (RAND) and the resulting expected response (XRES) and the cipher key (Kc) produced by A3/8 form a "triplet" (224 bits).
- Operation: An AuC will produce a batch of triplets for a MS, each entry with a different RAND, all at once and pass these for distribution to the associated HLR of the same HPLMN. When a MS attempts to make a call or a location update in either its HPLMN or in a visited PLMN, the SIM passes its identity to the VLR. The VLR makes a request to the subscriber's HPLMN for a batch of triplets for

the identity claimed by the MS (i.e. SIM) and the HLR of the HPLMN responds with a batch of (n) triplets for that claimed identity (Figure 3).

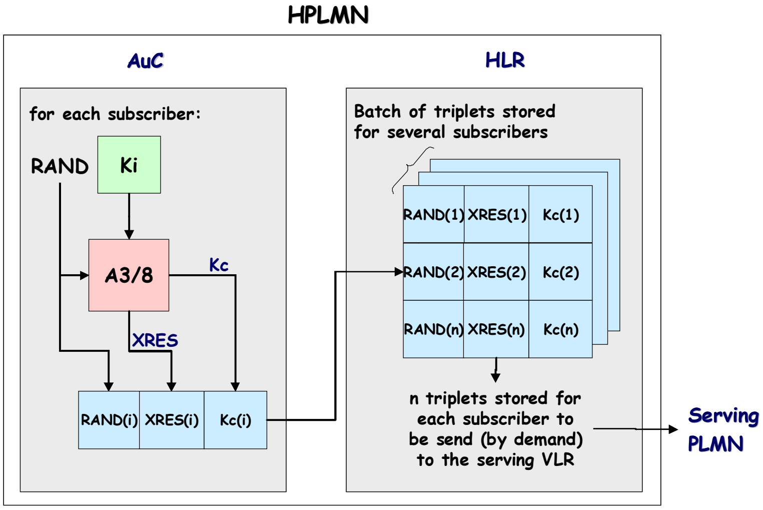

Figure 3 – GSM Triplet Generation, Distribution and Subscriber Management.

The serving VLR then authenticates the SIM by sending RAND(i) of the batch of triplets to the MS, via BS, and by comparing the value of the stored expected response XRES(i) for that RAND(i) with the received SRES produced by the SIM. If they match, the MS claimed identity is deemed to be authenticated and so the VLR can pass the cipher key Kc(i) from the triplet to the serving BS. Kc(i) and the secret key Kc calculated by the SIM, both with the same value, can be used respectively by the BS and the MS to protect the air interface (Figure 4). When the PLMN has run out of triplets, it should request more from the home HLR, though the PLMN is allowed to re-use triplets if it cannot obtain more from the HPLMN.

$$\text{Triplet}(i)=(\text{RAND}(i)\ \text{XRES}(i),\text{Kc}(i));\ 1 \le i \le n. \tag{3}$$

Where: n=number of entries stored in a batch of triplets for a subscriber.
i=entry chosen by the serving VLR to be send to the MS via BS.

2.5 Use of SIM as Security Module

- Purpose: Key distribution, authentication and cipher key generation.
- Description: SIM is implemented on a smart card (which should be "tamper proof") to make it "infeasible" to extract the Ki. Although the SIM is required at the start of a call only, In GSM a call must close if the SIM is removed from the ME during a call to avoid parallel calls using a unique SIM (i.e., a stolen SIM).
- Operation: As described before, the ME passes the RAND received from the VLR to the SIM. Then SIM passes its Ki value and the received RAND through algorithm(s) A3/8. The resulting SRES produced by the SIM is passed back to the ME

and then to the VLR, that verify if the SIM claimed identity can be authenticated. If the SIM is authenticated, the VLR passes Kc from the triplet to the serving BS. Then SIM passes Kc to the ME and as a result the BS and the ME can begin ciphering communication using Kc and the A5 algorithm (Figure 4).

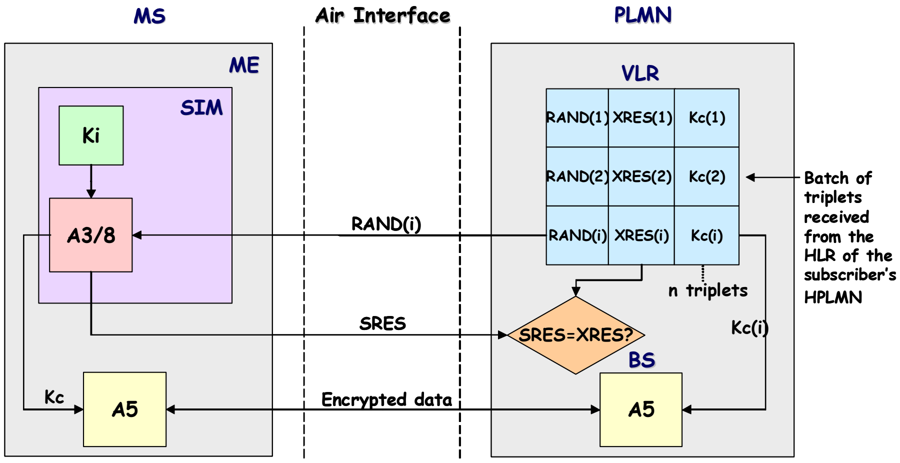

Figure 4 – Authentication and Encryption for GSM using triplets and SIM.

3 Effectiveness of GSM Authentication

The efficacy of GSM authentication relies on a number of algorithm requirements:

1st) it is statistically near impossible for an impostor to guess what the correct SRES should be and therefore masquerade as another subscriber. As the MS has only one chance to return SRES for a particular RAND, and the parameters SRES/XRES are 32 bits long, such an impostor has only a 1 in 2^{32} chance of guessing SRES correctly. Since SRES must be indistinguishable from any other 32 bit number that might be returned instead of SRES, than this is not a realistic attack.

2nd) an impostor cannot derive Ki from collecting a number of RAND-SRES pairs obtained from eavesdropping the air interface. This means that A3/8 must resist a known plaintext attack. Further, as an attacker could steal a SIM, send chosen RAND to the SIM and collect the SRES returned A3/8 must be resistant to a chosen plaintext attack (this latter requirement was shown not to be satisfied by the algorithm COMP128, used as A3/8 by many operators).

3rd) an impostor cannot derive a particular Kc from the RAND and SRES in the same triplet or by collecting a number as RAND-SRES pairs. This means that SRES and Kc, though derived from the same RAND and Ki, must be completely unrelated.

4th) Ki was not to be shared with the serving PLMN. Even A3/8 does not need to be known by the VLR, as this algorithm is used only where Ki is present (i.e. in the AuC and SIM). The same occurs with A5, not used in the VLR but in the BS.

4 Strength of GSM Encryption

A5 series algorithms are contained within the BS and the ME, but not within the SIM, as they have to be sufficiently fast and are therefore hardware. There are two defined algorithms used in GSM known as A5/1 (only members of CEPT[3]) and A5/2 (exportable). Neither A5/1 nor A5/2 has been officially published. The cipher key Kc, related with algorithm A5, is 64 bits long but the top 10 bits are forced to 0 in SIM and AuC. Then there are only 54 bits of effective key length. There is a new Kc in each call and 22 bit message key changes every frame (4.615 ms), giving a 5.4 hour repeat cycle. Encryption operates at the physical layer (layer 1) in the protocol stack. Ciphering only exists between MS (ME) and BS, as it was assumed that most other links afterwards would be along fixed lines. The decision to put encryption at the layer 1 had some consequences, as described below:

- The maximum amount of data, both user and signalling data is encrypted;
- Since ciphering takes place after error correction and deciphering takes place before error correction, than a stream cipher must be used for GSM because of the relatively high uncorrected error rate (of about 10^{-3}) in wireless environments;
- The frame counter, normally used for synchronisation at layer 1, is used with an expanded length as an input to the key stream generator. The frame counter is 1024 times longer than the longest frame aggregation required for non-encryption purposes to avoid repetition during a call and so causes encryption weakness; and
- The encryption algorithm can be implemented in hardware.

As a stream cipher, A5 works on a bit by bit basis (and not on blocks, as DES and AES). So, an error in the received ciphertext will only result in the corresponding plaintext bit being in error, as shown in the diagram for A5 operation (**Figure 5**).

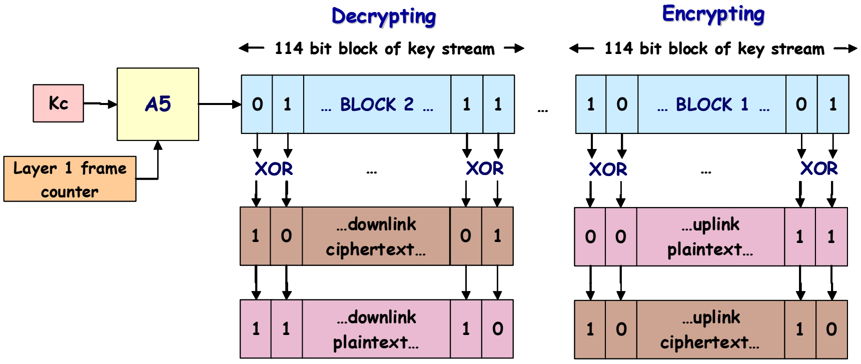
Decrypting
Encrypting
114 bit block of key stream
114 bit block of key stream
Kc
A5
Layer 1 frame counter
0 1 ... BLOCK 2 ... 1 1
1 0 ... BLOCK 1 ... 0 1
XOR ... XOR
XOR ... XOR
1 0 ...downlink ciphertext... 0 1
0 0 ...uplink plaintext... 1 1
1 1 ...downlink plaintext... 1 0
1 0 ...uplink ciphertext... 1 0

Figure 5 - Operation of A5 at the mobile station [13].

As a function of Kc and frame counter, a key stream generator produces a string of pseudo-random bits. This string of bits is XORed with the plaintext to produce the ciphertext. At the decrypting end, the same key stream is produced and XORed with the ciphertext to produce the plaintext. In GSM both sides can transmit simultane-

[3] CEPT - Conférence Européenne des administrations des Postes et des Télécommunications.

ously (it is full duplex), so within a frame, a MS or BS both transmits and receives a frame. The first 114 bit block (BLOCK1) of the string is used to encrypt the plaintext data being transmitted. The second 114 bit block (BLOCK2) is used to decrypt the data received in that frame, as shown in the diagram (**Figure 5**). At the other end of the air interface, the first block is used to decrypt the received ciphertext and the second block of the same string is used to encrypt the plaintext to be transmitted.

5 Attacks on GSM Security

The most significant physical and cryptanalytic attacks on GSM are given below; [9], [10] and [13] are also good references on this subject.

5.1 Microwave Links

The fact that the BS to BSC link is in many cases a point to point microwave link is a clear gap in GSM security. This link can be eavesdropped upon as data is at this point un-encrypted. At the time of GSM design, it was expected that the BS to BSC link would be across fixed links and therefore that encryption would not be required. In 3GPP, the encryption extends as far as the Radio Network Controller (RNC), the 3GPP equivalents of the BSC - microwave links are therefore protected [13].

5.2 SIM/ME Interface

When a call is not in progress, it is possible for the SIM to be removed from one ME and put in another with which it has no previous association. As the SIM-ME interface is unprotected, it is possible for the SIM to be connected to terminal emulators instead of “genuine” ME and consequently for messages on the SIM-ME interface to be tapped. However, unless the algorithms on the SIM are sub-standard, there is no advantage in compromising the SIM-ME interface in this way.

5.3 Attacks on the Algorithm A3/8

Wagner and Goldberg announced in April 1998 that they had cracked COMP128, an algorithm taking the function of A3/8 in the SIM of many operators. COMP128 had a weakness which would allow complete knowledge of Ki if around 160 000 chosen RAND-SRES pairs could be collected (“chosen plaintext” attack). There are active attacks that can be used to obtain these pairs. The quickest attack would be to steal the user’s mobile phone, remove the SIM and connect it to a phone emulator that can be used to send 160 000 chosen RAND to the SIM and receive the SRES. SIM tend to have relatively slow clock speeds and it can therefore take up to 10 hours to obtain the 160 000 pairs (with faster SIM, it would take 2 and a half hours).

Another method is to use a false BS to send the RAND over the air interface. The rate at which pairs can be collected is slower and would take a number of days; however the attacker does not need physical possession of the SIM. After these efforts, the attacker has the Ki and can masquerade as the user and run calls on her bill, and also determine the Kc for the user's calls and therefore eavesdrop upon them [13].

5.4 Attacks on the Algorithm A5/1

The only attack on an algorithm that has been confirmed to be A5/1 was that by Biryukov and Shamir [2], later improved by Wagner [3]. The technique used is known as "time-memory trade-off". In a pre-processing phase, a large database of algorithm states and related key stream sequences is created. In the attack phase, the database is searched for a match with sub-sequences of the known key stream. If a match is found, then with high probability the database gives the correct algorithm state. It is then simple to compute Kc and decipher the rest of the call. Shamir and Biryukov made an attack feasible in practice, if 2 minutes of known key stream could be obtained. Wagner spotted a further optimization which would allow Kc to be obtained with only 2 seconds of known plaintext (from both uplink and downlink).

This is not trivial, because the precise sequence of bits that is encrypted must be obtained. It is a fine piece of cryptanalytic work, but in reality it would probably not be used, as the false BS attack represents an easier method of eavesdropping [13].

5.5 Attacks on the SIM Card: Optical Fault Induction and Partitioning Attacks

Since the SIM module is implemented on a smart card, than any discovery of a common vulnerability in smart cards immediately affects the security of the information stored in the SIM (e.g., IMSI and Ki). Thus, it is vital to emphasize a new class of attacks on smart cards called optical fault induction, revealed by Skorobogatov and Anderson [12]. They discovered the flaw after Skorobogatov found that he could interrupt the operation of the smart card's microprocessor by exposing it to an electronic camera flashbulb. Such attacks are practical and cheap. They have carried them out using a camera flashgun (bought second-hand for $30) and a microscope to extract secret information from smart cards. According to them, illumination of a target transistor causes it to conduct, thereby inducing a transient fault.

They were able to expose the circuit to the light by scraping most of the protective coating from the surface of the microprocessor circuit embedded in each smart card. With more study, they were able to focus the flash on individual transistors within the chip by beaming the flash through a microscope. By sequentially changing the values of the transistors used to store data, they were able to 'reverse engineer' the memory address map, allowing them to extract the secret data from the smart card. The authors asserted that they have developed a technology to block these attacks.

Another relevant SIM card weakness was lately exposed in an IBM paper entitled "Partitioning Attacks: Or How to Rapidly Clone Some GSM Cards" [10]. The IBM team launched a new class of side channel (i.e. timing of operations, power consumption, electromagnetic emanations, etc.) attacks called partitioning attacks. The partitioning attack exploits vulnerabilities in the execution of COMP128 table lookups. They have launched a version of the attack on several implementations of COMP128 in various types of SIM cards. Thus, Ki can be recovered from a SIM card with less than 1000 invocations with random inputs, or 255 chosen inputs, or only 8 adaptively chosen inputs. So an adversary who has possession of a SIM card for a minute can easily extract Ki. In contrast, the previously best technique to attack SIM cards was to use an attack on COMP128 with 160 000 chosen inputs (section 5.3).

The new IBM approach seems to be a much more useful method than either breaking the cryptographic algorithms (COMP128) used by the SIM card or by intrusive attacks, such as the optical fault induction. The IBM researchers' report also offers advice to the smart card industry on how to protect against vulnerabilities.

5.6 False Base Station

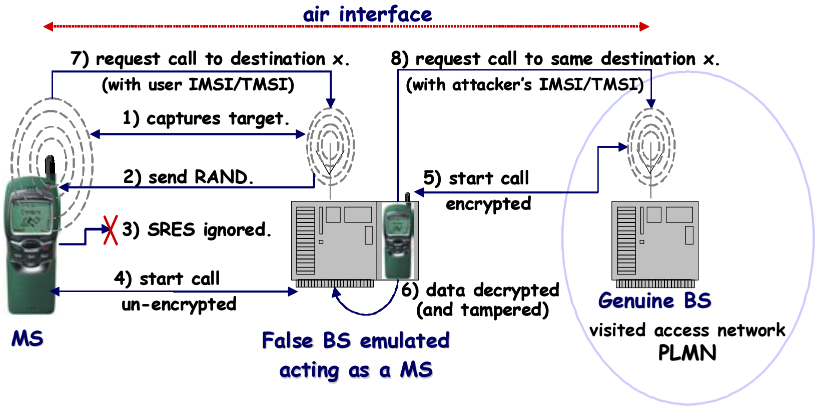

Figure 6 - The False BS Threat.

According to [9], “the MS is authenticated to the BS, but the BS is not authenticated to the MS”. That is, GSM provides unilateral authentication. This allows attacks where a malicious third party masquerades as a BS to one or more MS. One of the assumptions when GSM was designed was that the system would not be subject to active attacks, where the attacker could interfere with the operation or impersonate one or more entities. It was believed such attacks would be too expensive compared to other methods of attacking GSM as a whole and wiretapping the fixed links or even just bugging the target [13]. But the cost of BS devices has fallen quickly and it is easy to find BS emulators. Moreover the use of encryption on a call does not happen automatically – the call begins un-encrypted and the MS is instructed to begin (or not) ciphering at a particular point in call set-up. The start encryption instruction is given un-encrypted and with no origin authentication by the PLMN and it can be subject to tampering in transit; i.e., the BS instructs the MS to begin encryption but this is manipulated in transit to be an instruction not to cipher. The problem of the PLMN expecting encryption and the genuine MS not is realized by the false BS acting as a MS and setting up a call to the genuine BS itself. The un-encrypted call from the target MS to the false BS is connected to the encrypted call from the false BS to the PLMN, so it seems to the caller that they have the call they requested. But the call from the target MS to the false BS is not encrypted so the link can be eavesdropped at this point. And since the call between the false BS (but true SIM) and the PLMN is an encrypted genuine call, the PLMN does not see that anything is awry.

The Figure 6 shows the steps to be followed in order to achieve the false BS attack. One effect of this attack is that the call is made on the false BS subscription and not that of the MS's. So the BS attack can be detected later if an itemized bill is checked. The 3G prevention of this attack is carried out by integrity protection of the start encryption command, and cannot be achieved merely by mutual authentication.

5.7 Other Threats

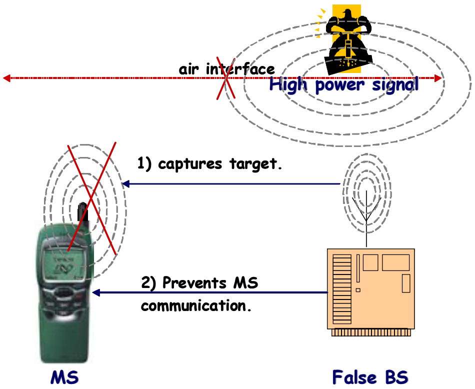

Figure 7 – Denial of service threats to GSM.

As mentioned in section 5.3, the PLMN is allowed to re-use triplets if it cannot obtain more from the HPLMN. But there is nothing to stop the BS using the same triplet repeated times. Thus if a single triplet is compromised, a false BS can impersonate a genuine PLMN to a MS indefinitely. Moreover, as the false BS has Kc, it will not be necessary for the false BS to suppress encryption on the air interface. While the genuine BS is using the compromised triplet, an attacker could impersonate a MS and obtain calls at a valid user's expense [9]. Another threat that always exists is denial of service (D.O.S), which can be achieved if an attacker sends a signal at high power on the GSM frequencies to 'jam' the radio path. Another kind of D.O.S is for a false base station to capture a MS and prevent the MS communication (Figure 7).

5.8 Measurement and fraud detection

As illustrated in [4], a properly designed billing system can be used to detect fraud patterns from normal GSM usage. Different types of fraud often produce a distinct pattern that can be detected. Therefore what is necessary to detect are: (i) multiple calls at the same time; (ii) large variations in revenue being paid to other parties; (iii) large variations in the duration of calls, such as very short or long calls; (iv) changes in client usage, indicating that a MS has been stolen or is being abused; and (v) monitor the usage of a GSM client closely during a "probationary period".

There are some "Fraud Engines" on the market that can provide these features, enabling patterns in billing data to be analyzed, and give time for swift effective action. With fraud detection capability, and security procedures in place, it is possible to minimise the effect of fraud on a billing system.

6 Third Generation Security – 3GPP

The 3G radio access link security was built on the security of GSM. 3GPP, that developed the 3G/UMTS standards, adopts the security features from GSM that have proved to be needed and robust and try to ensure compatibility with GSM to ease inter-working and handover. 3G security tries to correct the problems with GSM by addressing security weaknesses and by adding new features. The 3G security has the following security features: mutual authentication and key agreement between MS and network; encryption of user traffic and signalling data over the air interface; and integrity protection of signalling data over the air interface. The GSM security features to retain and enhance in 3GPP are, according to [1] and [6]: (i) maintain a smart card as a subscriber identity module (UMTS SIM or USIM); (ii) authentication of the MS to the network; (iii) encryption of the user traffic and the signalling data over the air interface; and (iv) user identity confidentiality over the air interface.

The new security features required for 3GPP includes mandatory integrity protection for critical signalling commands (e.g. for the start encryption command), which provides enhanced protection against false BS attacks (section 5.6) by allowing the MS to check the authenticity of certain signalling messages. This feature also extends the influence of MS authentication when encryption is not applied by allowing the PLMN to check the authenticity of certain signalling messages. Mutual authentication and key agreement also provides enhanced protection against false BS attacks by allowing the MS to authenticate the PLMN. 3G authentication provides authentication of MS (USIM) to PLMN and PLMN to MS, establishes a cipher key (CK) and an integrity key (IK), and assures to the MS that the keys were not used before.

In 3G systems a new sequence number (SQN), generated in the AuC, is attached to the triplets to prevent the threat resulting from the triplets re-use (section 5.7). USIM has a scheme to verify freshness of received SQN (the SQN attached to a triplet must exceed the most recently formerly received subset). A Message Authentication Code (MAC) is also attached to show that the 'quintet' really came from the HPLMN and to integrity protect the SQN recently attached.

The encryption of the user traffic and the signalling data over the air interface is performed through an algorithm called KASUMI [5] (based on the Japanese MISTY1 [7]), which had an open design process, taking a longer cipher key length (128-bit) derived during authentication. The encryption terminates at the RNC, a 3G entity similar to the BSC. The links BS-RNC that may be over microwave are thus ciphered (section 5.1). KASUMI is also used for the integrity protection of commands (critical signalling) between MS and RNC [13].

3G specifications also begun to include protection of network signalling (e.g. quintets) transmitted between and within networks, which can be used to eavesdrop upon MS traffic or to masquerade as valid MS. Further, to prevent false messages transmitted along the network, it is vital that the origin of such commands can be authenticated so that only authorised parties can send them.

7 Further Applications in Network (Internet) Remote Access

The use of triplets is an significant GSM feature that permits delegation of the authentication function as well as cipher key distribution from the HPLMN to the serving PLMN through a minimal trust relationship between operators. That is, the home network does not need to reveal the information most sensitive, such as the secret key (Ki), to any intermediate entity in the PLMN currently 'visited' by the MS. This kind of solution (enhanced in the 3GPP with the use of quintets), can be "exported" to applications like network remote access supporting ubiquitous client mobility.

In particular, in the scenario where a user's access device wishes to access the Internet via different multiple access media and network interfaces (e.g. PPP, Bluetooth, IEEE 802.1x), leading to the use of a number of network operators, such as the recently inaugurated IETF PANA work[4]. In this scenario, the backend authentication server (home AAA[5] server) can make use of triplets (or 3G quintets) to delegate client (e.g. PANA client) authentication function to any intermediate authentication agent (e.g. PANA authentication agent) in the access network, achieving minimal trust relationship between home and access network operators.

Moreover, the same triplet (or quintet) distribution operation can include cipher keys distribution to any network access point, such as NAS (Network Access Server), wishing to establish an encrypted channel with a visiting access device (e.g. PANA Client). We can imagine a feasible scenario where the interface between the client and the NAS is wireless and the NAS must use a signalling message, such as a start encryption command, to activate the encryption in the air interface. In this case, beyond the use of triplets, the introduction of mandatory integrity protection for critical signalling protocol instructions is crucial to avoid the typical masquerading or 'false entity in-the-middle' attack, as occurs in the false BS attack, of which prevention cannot be achieved solely by mutual authentication. Thus, the mandatory integrity protection for critical signalling messages is another solution arising from the mobile telecommunications sphere that may be clearly adapted on security mechanisms for the Internet remote access area.

[4] PANA is acronym for "Protocol for carrying Authentication for Network Access". See more details in the PANA workgroup link: http://www.ietf.org/ html.charters/pana-charter.html.

[5] AAA is acronym for Authentication, Authorization and Accounting.

8 Conclusions

This article reviewed the GSM cellular phone system security, with focus in the air interface protocol. This document described the GSM security operation, the effectiveness of GSM authentication and the strength of GSM encryption. The GSM security threats and attacks were included in this paper for a better understanding of the GSM features retained and enhanced for the Third Generation Security (3GPP).

Some simple but useful and robust security solutions implemented in GSM and enhanced in the 3G security, such as the use of triplets and quintets and the mandatory integrity protection for critical signalling commands, can be "exported" and adapted for application in other correlated areas, such as authentication for network (Internet) remote access supporting ubiquitous mobility.

9 Acknowledgements

The author would like to acknowledge the many helpful insights and corrections provided by Chris Mitchell, Kenny Paterson and Peter Howard. The author would also like to thank all the staff members of the Information Security Group of Royal Holloway, University of London and of the Information Security Division of CASNAV, Brazilian Navy for their advice and encouragement.

References

[1] 3GPP TS 33.102 V3.11.0, "Security Architecture", 3rd Generation Partnership Project, Technical Specification Group, 3G Security, Valbonne, France, 2002, http://www.3gpp.org/ftp/Specs/2002-03/R1999/33_series/33102-3b0.zip.

[2] A. BIRYUKOV, A. SHAMIR, "Real time cryptanalysis of the alleged A5/1 on a PC", preliminary draft, December 1999.

[3] A. BIRYUKOV, A. SHAMIR, D. WAGNER, "Real time cryptanalysis of A5/1 on a PC", in *FSE 2000*, LNCS No. 1978, Springer Verlag, Berlin, 2000.

[4] C. BROOKSON, "GSM (and PCN) Security and Encryption", 1994, http://www.brookson.com/gsm/gsmdoc.htm.

[5] ETSI TS 135 202 V4.0.0, "Universal Mobile Telecommunications System (UMTS); Specification of the 3GPP confidentiality and integrity algorithms; Document 2: Kasumi algorithm specification", http://www.etsi.org/ dvbandca/3GPP/3gppspecs.htm.

[6] P. HOWARD, "GSM and 3G Security", lecture notes, Royal Holloway, University of London, 19 Nov 2001, http://www.isg.rhbnc.ac.uk/msc/teaching/is3/is3.shtml.

[7] M. MATSUI, "New block encryption algorithm MISTY", in *Fast Software Encryption '97*, Lecture Notes in Computer Science No. 1267, Springer-Verlag, 1997, pp. 54–68.

[8] C. MITCHELL et. al., "Link 3GS3 Technical Report 2: Security Mechanisms for Third Generation Systems", Vodafone, GPT and RHUL, 15/05/96, pp. 25 and 92.

[9] C. MITCHELL, "The security of the GSM air interface protocol", Technical Report, RHUL-MA-2001-3, 18 August 2001.

[10] J. R. RAO, P. ROHATGI AND H. SCHERZER, "Partitioning Attacks: Or How to Rapidly Clone Some GSM Cards", IBM Watson Research Center, *in 2002 IEEE Symposium on Security and Privacy*, Oakland, CA, May 2002.

[11] R. SCHMITZ, "SHAMAN Deliverable D02 – Intermediate Report: Results of review, Requirements and reference Architecture", Information Society Technologies, 08 November 2001, pp. 41-42.

[12] S. SKOROBOGATOV, R. ANDERSON, "Optical Fault Induction Attacks", University of Cambridge, in *IEEE Symposium on Security and Privacy*, Oakland, May 2002.

[13] M. WALKER AND T. WRIGHT, "Security", in F. Hillebrand, editor, *GSM and UMTS: The Creation of Global Mobile Communication*, pp. 385-406, John Wiley & Sons, New York, 2002.