The security of the GSM air interface protocol

Chris J. Mitchell

Technical Report
RHUL–MA–2001–3
18 August 2001

Royal Holloway
University of London

Department of Mathematics
Royal Holloway, University of London
Egham, Surrey TW20 0EX, England
http://www.rhul.ac.uk/mathematics/techreports

Abstract

This paper describes the level of security offered by the GSM air interface protocol[1]. All known attacks against this protocol are described, and their feasibility is assessed. Consideration of the level of security offered by particular algorithms which may be used to help implement the protocol are outside the scope of this paper. The main objectives of the paper are to provide a benchmark against which the security features of future mobile telecommunications networks can be judged, and also to help determine the security requirements for future networks.

1 Introduction

1.1 Scope and purpose

This paper is primarily concerned with the level of security offered by the GSM air interface protocol. In order to assess the security level achieved, all known attacks against the GSM protocol are described, and their feasibility is assessed. However, consideration of the level of security offered by particular algorithms which may be used to help implement the protocol are outside the scope of this paper. Note that a very brief presentation of some of the possible protocol attacks was previously given in [6]. An analogous description of protocol weaknesses in the North American IS-41 scheme was given in [5].

The main objectives of the paper are to provide a benchmark against which the security features of future mobile telecommunications networks can be judged, and also to help determine the security requirements for future networks.

1.2 Terminology

In order to describe security threats to the GSM air interface, we need to outline how the GSM security features operate. This we do in Section 2 immediately below. However, before giving this description we first outline the terminology we will use.

A *Mobile Station* (MS), i.e. a mobile telephone with its *Subscriber Identity Module* (SIM), visits a network (more formally a *Public Land Mobile Network* or PLMN) by communicating with a *Base Station* (BS) belonging to that network, and receives an entry in a *Visitor Location Register* (VLR) maintained by that network. The MS communicates with the BS across a

[1] This report was originally written in June 2000

radio path, also known as the *Air Interface* or the *Radio Interface*. The MS is said to be *registered* with the network which it is currently visiting.

Every MS has a *Home Network*, or *Home PLMN* (HPLMN), with which it has a contractual relationship. Every MS has a permanent identifier, the *International Mobile Subscriber Identity* (IMSI), which is shared with, and assigned by, its HPLMN. For every PLMN we refer to the MSs which 'belong' to it, meaning those MSs for which it is the HPLMN. Each PLMN maintains a *Home Location Register* (HLR) in which the current (most recent known) location of all its mobiles is recorded. Each PLMN also operates an *Authentication Centre* (AuC), used to store secret key information relating to each of its mobiles. The AuC can be integrated with other network functions, e.g. with the HLR.

2 The GSM air interface protocol

We start the paper by giving a brief description of the operation of the GSM air interface security features. GSM security features are standardised in GSM 02.09, [2].

GSM provides the following security features for the air interface:

- International Mobile Subscriber Identity (IMSI) confidentiality,
- IMSI authentication,
- user data confidentiality on physical connections,
- connectionless user data confidentiality, and
- signalling information element confidentiality.

We consider the provision of these features in turn.

2.1 Subscriber identity (IMSI) confidentiality

The purpose of this function is to avoid an interceptor of the mobile traffic being able to identify which subscriber is using a given resource on the radio path. The provision of this function implies that the IMSI should not normally be transmitted in clear text in any signalling message on the radio path.

Instead of using the IMSI, a *Temporary Mobile Subscriber Identity (TMSI)* is used to identify a mobile subscriber on the radio path. The TMSI is allocated by the VLR where the MS is registered. A new TMSI is allocated by the VLR at least on every location update. In certain special circumstances, the fixed part of the network can require the MS to send the IMSI in clear. A new TMSI is then allocated and sent to the MS in ciphertext.

These special circumstances include the very first time that a MS registers in a network, when a MS registers with a new VLR and the VLR with which the MS was previously registered cannot be contacted, and after a software failure in the current VLR.

2.2 Subscriber identity (IMSI) authentication

Subscriber identity (IMSI) authentication is the corroboration by the land-based part of the system that the subscriber identity (IMSI or TMSI), transferred by the mobile subscriber within the identification procedure across the radio path, is the one claimed.

The authentication of the GSM PLMN subscriber identity may be triggered by the network when the subscriber applies for:

- a change of subscriber-related information element in the VLR or HLR (including some or all of: location updating involving change of VLR, registration of a supplementary service, or erasure of a supplementary service);
- an access to a service (including one or both of: set-up of mobile originating or terminated calls, activation or deactivation of a supplementary service);
- first network access after restart of MS or VLR; or
- in the event of cipher key sequence number mismatch.

The authentication procedure consists of the following steps.

1. The fixed subsystem transmits a non-predictable number RAND to the MS.
2. The MS computes the 'signature' of RAND, labelled SRES, using algorithm A3 and the secret Individual Subscriber Authentication Key, K_i, i.e.
$$\mathrm{SRES} = \mathrm{A3}_{K_i}(\mathrm{RAND})$$
where $\mathrm{A3}_K(D)$ denotes the output of algorithm A3 given key input K and data input D. At the same time the MS uses agorithm A8 to compute the encryption key K_c as
$$K_c = \mathrm{A8}_{K_i}(\mathrm{RAND})$$

Note that algorithms A3 and A8 are PLMN operator specific — only the input/output formats are standardised.

3. The MS transmits the signature SRES to the fixed subsystem.

4. The fixed subsystem tests SRES for validity.

The key K_i is allocated, together with the IMSI, at subscription time. K_i is stored both in the subscribers SIM and in an AuC belonging to the MS's HPLMN. A PLMN may contain one or more AuCs.

Several scenarios are possible when a VLR wants to perform an authentication, depending on whether the TMSI or IMSI is used for identification, and whether the TMSI can be used to retrieve security information already available in the old VLR. This security related information consists of a Random (RAND), a Signed Response (SRES) and a ciphering key K_c. These three together are called a triplet.

When no triplets are available in the VLR, and/or triplets cannot be retrieved from the old VLR (in the event that the MS performs a location update in a new VLR), then new triplets are requested from the HLR/AuC. Typically the HLR/AuC will supply a number of triplets to the VLR, so that several authentications can be performed without requiring further internetwork communications.

2.3 Confidentiality of signalling information elements, connectionless data and user information elements

The following information is considered sensitive and must be protected against eavesdropping:

- Signalling information elements

- To ensure identity confidentiality, the TMSI must be transferred in a protected mode at allocation time and at other times when the signalling procedures permit it.

- The confidentiality of connection-less user data requires at least the protection of the message part pertaining to OSI layers 4 and above.

- User data.

Confidentiality is achieved by the use of encryption at OSI layer 1.

Stream cipher encryption is performed using algorithm A5 with the key K_c. Algorithm A5 is specified in Annex C of GSM 03.20, [3]. A5 is not

PLMN specific. However several A5 versions are possible and negotiation between the MS and the network is carried out to decide on which version of A5 to use. The key K_c is generated by the MS during the authentication procedure (see Section 2.2). K_c is calculated at the same time as SRES, and is transmitted from the AuC to the VLR together with SRES (as part of a 'triplet').

A distinction is made between data on a Dedicated Control Channel (DCCH) and data on a Traffic Channel (TCH).

- On a DCCH the start of enciphering is under control of the network. The BS sends in clear text a message 'Start cipher' and deciphering is started in the BS. The MS starts enciphering and deciphering and sends its next message (any message) enciphered. When this message is deciphered correctly in the BS, enciphering is started in the BS.
- On a TCH, enciphering and deciphering are started as soon as a key is present, unless 'Null Cipher' mode is selected.

The enciphering stream at one end and the deciphering stream at the other end must be synchronised for the enciphering bit stream and the deciphering bit streams to coincide. The underlying Synchronisation scheme is described in Annex C of [3].

3 The false base station threat

As should be clear from the description of the GSM security scheme, the MS is authenticated to the BS, but the BS is not authenticated to the MS. That is, the scheme provides unilateral rather than mutual authentication. This allows the possibility of attacks where a malicious third party masquerades as a BS to one or more MSs.

Active attacks involving the impersonation of network elements were considered when GSM was originally designed, but were not deemed to be worth addressing. The perceived complexity of building BS devices was deemed such that the risk arising from the threat was assessed as rather small. Further, the fact that traffic exchanged between BS and MS is encrypted, reduces the risks arising from the lack of full mutual authentication (see also [6]).

However, a level of threat does remain, not least because of the following factors.

- The cost of BS devices has fallen rapidly, and 'testing' devices capable of emulating a genuine BS are readily available.

- The use of encryption on the air interface is completely controlled by the BS, and some networks do not ‘switch on’ data encryption.

In the next section we explore the precise implications of the false base station threat in more detail. Note that we consider a variety of different ‘active attacks’ on the radio path between MS and BS. In some cases we consider attacks where we assume that an attacker can modify some of the signalling information exchanged between an MS and the BS with which it is registered. In practice, such an attack will probably be easiest to implement by having a false BS ‘capture’ the MS, and then act as a simple relay between the MS and a genuine BS, except that key items of data will be changed.

4 Specific threats resulting from base station impersonation

4.1 Loss of IMSI confidentiality

The fact that there is a provision to enable a BS to request a MS to send its IMSI across the air interface means that there is a very straightforward way for a false BS to compromise IMSI confidentiality. The false BS simply transmits the ‘Identity request’ message to the MS, which responds with the IMSI. This attack has been discussed previously in a number of places — see, for example, [1, 6].

In fact, unless public key cryptography is employed, such a threat is very difficult to neutralise, even when mutual authentication between BS and MS is provided.

Note also that, even if the ’Identity request’ message did not exist, there is another simple threat to IMSI confidentiality. This stems from the observation that if the BS sends a particular RAND value to an MS, then that MS will always give the same SRES value in response. Hence if an attacker with a false BS has at some time in the past seen a RAND/SRES pair for a particular MS, then the attacker can determine whether an MS is the same as the previously identified MS by sending it the same RAND value and observing the response.

4.2 Loss of data confidentiality

The threat of loss of confidendiality of data transmitted across the air interface can arise in a number of ways. We consider them in turn.

4.2.1 Non-confidential networks

Probably the most straightforward threat of this type arises from the fact that it is entirely up to the network whether or not to use data encryption. This should be clear from the description in Section 2.3, since radio path encryption starts only when the BS sends the 'Start cipher' command to the MS. Moreover the (human) user of the MS typically has no indication of whether or not their call is being encrypted, although there is no reason why a handset manufacturer could not arrange for a visual or audible indication to be given to the user when encryption is enabled/disabled. Indeed, an encryption indicator is specified in GSM, although there is a flag on the SIM which allows the operator to suppress the indicator. Terminals which provide such an indication do exist, but are unfortunately relatively uncommon at present.

Some networks never switch on encryption. The decision as to whether or not to use encryption will typically be a matter governed by the licencing arrangements prevailing in the country within which the network is located, and some countries prevent the use of encryption. Hence, in some locations all mobile traffic may be vulnerable to unauthorised interception.

Note that, whilst the situation described in the previous paragraph may seem undesirable, in other mobile networks the situation can be substantially worse! For example, in IS-41 networks, encryption of the air interface is considered as a 'Value Added Service', and hence is very rarely enabled. Of course it is true that, in this latter case, the user might at least be aware that he or she has not subscribed to the encryption service and hence cannot expect to have it enabled.

4.2.2 Mobile capture by false base station

The next threat to data confidentiality arises in a somewhat more complex way. We summarise the attack below. Before starting the description we briefly explain the resources assumed to be available to the attacker.

We assume that the attacker has a valid subscription to a GSM network, and an MS device (presumably incorporating the attacker's SIM) capable of communicating with a genuine BS. We also assume that the attacker has a device capable of emulating a BS, which is integrated with the attacker's MS device. This integration should permit transparent routing of call information from a legitimate MS 'captured' by the false BS to a genuine BS, via the attacker's own MS device. The full nature of the required integration should be clear from the attack which we now outline.

1. The attacker uses his false BS to 'capture' the target MS. That is,

the target MS, presumably belonging to an individual whose calls the attacker wishes to intercept, registers with the attacker's false BS, believing it to be a BS belonging to a legitimate network.

2. When authentication needs to take place between the captured target MS and the false BS, the false BS can send an arbitrary RAND value and can ignore the SRES returned in response. Of course, the false BS will not know the correct value of K_c, but this will not matter, as we see below.

3. The genuine BS may require the call to be enciphered, and the encryption will take place in the 'normal' way between the genuine BS and the attacker's MS, where it will be decrypted. The false BS will never send the 'Start cipher' command to the captured MS, and hence the fact that the false BS does not know the value of K_c for the mobile does not matter.

4. When the target MS places a call, the false BS detects this, and can read the dialled digits since the MS will not be encrypting. The false BS then uses its integrated MS to place a call to the same destination using the IMSI/TMSI belonging to the attacker via the genuine BS with which the attacker's MS is registered.

5. The false BS can now monitor the entire call and all associated signalling information.

It is interesting to note that the attacker has to pay for all the calls which it intercepts, but this may be a small price to pay for the confidential conversations which it may intercept. Note that the attack is described above for the case where the legitimate connection to the network is in the form of a legitimate MS, although it would equally be possible to use other connections, e.g. to the fixed network.

The attack may be detected in three possible ways.

- The caller (target MS owner) may detect the attack *after the event* since he/she will not be billed for the call.

- If the caller is in possession of a handset which is capable of indicating whether or not encryption is in operation, he/she may observe that all calls are being transmitted unencrypted.

- The recipient of the call may observe that the call comes from an unexpected caller, by making use of the Calling Line Identifier (CLI) information.

However, the risks (to the attacker) associated with all of these methods of detection can be minimised. The first method of detection only works after the event, and will only be noticed by those rare users checking their bills for *missing* calls. The second method of detection will indeed work, although handsets which indicate whether or not a call is encrypted are relatively rare, and detection depends on the user checking every call. Finally, CLI does not always work, notably with calls made internationally and with calls placed from a PABX (private automatic branch exchange). In fact the display of CLI information can be suppressed by the attacker; indeed, it is often a legal requirement that operators allow callers to inhibit the display of CLI information. Moreover, the fact that the CLI is 'wrong' will not be detectable by the call recipient if he/she is using a terminal incapable of displaying CLI information, as is the case for most terminals connected to the fixed network.

4.2.3 Mobile capture by false base station (variations)

The attack described in the immediately previous section has a number of variations with very similar properties. We briefly mention a few of them.

- **Capture of called number only**. In this attack the false BS does not need to maintain a connection with a genuine network. The false BS captures the target MS and suppresses encryption on the link between MS and false BS. Whenever the MS tries to place a call, the false BS captures the called number but does not actually attempt to make the connection. As a result the false BS obtains information about who the target MS user is trying to call, although the false BS does not actually enable calls themselves to be intercepted.

- **Spoof answer/redirection of calls from target MS**. Once a false BS has captured a target MS (and suppressed encryption on the radio path), whenever the target MS user attempts to place a call, the operator of the false BS can answer the call as if he/she were the requested end-point of the call. If the false BS has the means to route calls through to a genuine BS, then the call can be routed to any destination chosen by the attacker.

- **Spoof calls to target MS**. Once a false BS has captured a target MS (and suppressed encryption on the radio path), the false BS can set up spoof calls at will to the target MS.

- **Late capture variants**. The three variant attacks just described can work just as well even if the false BS does not 'permanently' capture the target MS. The false BS can wait for the target MS to place a call to the genuine BS, and then ‘take over’ the connection to the MS once the authentication process is complete (and before encryption is switched on).

Finally, all the attack descriptions so far apply to calls originated by the target MS. However, exactly analogous attacks will also apply to calls made to the target MS. Such calls are at an equal risk of interception using a false BS.

4.2.4 Encryption suppression

Now consider the situation where the target MS has registered with a genuine BS, and an attacker has the means to manipulate the communications between the MS and BS. As part of the registration process, the MS sends the BS a ‘terminal capability message’ which informs the network as to the encryption capabilities of the MS. The attacker might attempt to modify this mesage so that the BS believes that there is an incompatibility between the encryption capabilities of the MS and network, and hence encryption is not enabled.

If such an attack were realised, then this would be a simple way of performing attacks on both the data confidentiality and the authentication process (by a process similar to that described in Section 4.3 below). However, such an attack is actually prohibited by the GSM standards — in GSM 02.09, [2], it is stated that a network should deny service to all user equipment that does not support both of the encryption algorthms A5/1 and A5/2. Since no other algorithms are in use, incompatibilites cannot arise if the GSM standards are adhered to.

Of course, if GSM is carelessly implemented, then an attack of this type could be realised and could be a major threat.

4.3 Bypassing IMSI authentication on unencrypted networks

We now consider again the situation arising in a network which does not use encryption, i.e. in a network where the BS never sends a ‘Start cipher’ message to an MS. If the attacker has the means to manipulate traffic between the MS and a genuine BS with which it is registered, then the means exist to bypass the MS authentication process and obtain calls fraudulently. This attack operates as follows.

We suppose the target MS is registered with a genuine BS. The attacker (by some means) causes the MS to initiate the call set-up procedure. This could, for example, be caused by the attacker sending a false indication of an incoming call to the target MS. The BS successfully authenticates the MS, and the attacker now takes over the (unencrypted) communications channel and uses it to place a call of their choice.

Note that this attack is not possible if the network activates encryption, since the attacker would not be able to 'take over' the genuine 'session' between MS and BS.

4.4 Call stealing on unencrypted networks

A second method for fraudulently obtaining service (in a network without encryption) operates as follows. As before we assume that the attacker has the means to manipulate traffic between the MS and a genuine BS with which it is registered. We also assume that the target MS is roaming in a network other than its HPLMN.

An associate of the attacker now makes a call to the target MS. Once the call is connected, and authentication is complete, the attacker takes over the connection and answers the call placed by his associate. Because of the way that GSM charging operates, the target MS owner will now be charged for the roaming leg of the call.

5 Other threats

5.1 Compromise of authentication data

As described in Sections 2.2 and 2.3 above, the authentication process requires the BS to know a triplet, consisting of a random value RAND (sent to the MS), the correct response SRES, and the associated cipher key K_c. It should be clear that there is nothing in the cryptographic provisions to stop the BS using the same triplet for repeated authentication processes.

This observation leads to an associated threat. If a single triplet is ever compromised, then this can be used by a false BS to impersonate a genuine network to a MS indefinitely. Moreover, because the BS has the key K_c, it will not be necessary for the BS to suppress encryption on the radio path, and the 'captured' MS will lose the only reliable method of detecting the fact that it has been captured by a false BS.

More serious still, compromise of triplets could enable an attacker to impersonate a genuine MS, and obtain calls at a genuine user's expense. Of course, this is a short term threat, since once the genuine BS stops using the compromised triplets then the attacker can no longer impersonate the genuine MS.

5.2 Denial of service threats

Denial of service threats always exist in a mobile telecommunications network. Probably the simplest is for the attacker to simply 'jam' the radio path by sending a signal at high power on the frequencies used by GSM. However such a threat would at least be simple to detect. A little more worrying are denial of service threats which are not so easily detectable, and the false base station attack enables such threats to be realised. One possibility is simply for a false BS to 'capture' the target MS, and then simply prevent the MS making or receiving any calls.

6 Lessons for third generation mobile systems

We conclude the main part of this paper by briefly reviewing how the attacks described above have influenced the approach to security taken by the emerging 3GPP specifications for third generation mobile systems[2]. We consider the main attacks in turn, and describe countermeasures built into 3GPP.

6.1 Threats to identity confidentiality

This threat was outlined in Section 4.1. Given that the 3GPP specifications do not use public key cryptography for authentication, it is more difficult to justify its use for user identity confidentiality. Without public key cryptography it is very difficult to counteract this threat. Solutions using 'symmetric' (conventional) cryptography have been proposed (see, for example, [4]), typically involving the use of multiple temporary identities. Unfortunately, such solutions are vulnerable to major database failures at the HPLMN, recovery from which requires a mechanism to force the MS to send its permanent identity across the air interface in clear text. Other mechanisms exist which seek to protect the secrecy of the MS's permanent identity by encrypting it using a 'common key' when it is sent across the air interface, and such a solution has been considered for adoption in 3GPP. Of course, this latter solution still possesses vulnerabilities, since compromise of a single 'IMSI hiding' key shared by a number of MSs will compromise identity confidentiality for all these MSs.

[2] see `www.3gpp.org`.

It is generally accepted that completely defeating *active* attacks on IMSI confidentiality is extremely difficult, and hence the decision has been made to adopt a GSM-like mechanism within 3GPP, based on a single 'level' of temporary identities. Such a mechanism offers a good measure of identity confidentiality against passive (monitoring only) attacks. However, the risk therefore remains in 3GPP of active attacks compromising the permanent MS identity.

6.2 Threats to data confidentiality

The problems identified in GSM (see Section 4.2) all stem from the fact that the use of encryption is not mandatory, and also that the integrity of commands sent to a MS by the BS is not protected. Whilst it would be attractive to solve the problem by making encryption mandatory, this has not been adopted in 3GPP because of problems with the regulatory regimes in some countries. Instead within 3GPP there is a mandatory 'cipher mode' command from the BS to the MS, which is integrity protected using a shared secret key. This command must be sent whether or not encryption is to be used on the air interface, and hence there is no way a false BS can successfully imitate a genuine BS.

6.3 Threats to mobile authentication

The threats described in Section 4.3 arise primarily because not all networks are encrypted. The solution adopted in 3GPP is to permit an MS and a BS to engage in an authentication process at intervals during a call, as well as providing for authentication of called numbers. This latter measure prevents an intruder changing the call destination.

6.4 Threats to authentication data

The main reason for the threat in GSM (see Section 5.1) is because an MS has no way of detecting whether or not an authentication challenge (RAND) has been used before. The threat is therefore avoided in 3GPP by including a 'freshness indicator', in the form of a serial number, with every authentication challenge provided to the visited network by the HPLMN's AuC. This enables an MS to check every challenge for 'freshness' before proceeding with the authentication procedure.

7 Conclusions

We have reviewed the main security threats to the GSM air interface. The following conclusions can be drawn regarding the relative robustness of the air interface security services offered by GSM.

First and foremost the confidentiality services are at much higher risk of attack than the authentication service. As we have seen, the only known attacks to 'break' the MS authentication process (and hence enable the fraudulent use of the mobile network) are only feasible in networks which do not make use of air interface encryption. Thus, given that fraud was probably the primary threat as far as the GSM system designers were concerned, the system still meets its main objective, as long as the encryption feature is used.

As far as the confidentiality services are concerned, the threat to IMSI confidentiality is greatest. A relatively unsophisticated false BS should be able to induce an MS to transmit its IMSI. The threat to data confidentiality is a little less, since, as we have seen, attacks to reveal the contents of calls require a rather more sophisticated false BS capable of integration with an MS device.

Finally we note that the range of threats to the GSM air interface (and not just to confidentiality) is significantly greater when the network does not employ encryption, and when GSM is not properly implemented (allowing encryption to be suppressed).

Acknowledgements

The author would like to acknowledge the invaluable advice and encouragement of Peter Howard and Tim Wright, without which this paper would not have been written. The research reported in this paper was carried out with the financial support of Vodafone Ltd. and the EU-funded ACTS project AC336 USECA (`http://www.useca.freeserve.co.uk`). However, the author alone is responsible for any errors or omissions which may exist in this paper.

References

[1] L. Chen, D. Gollmann, and C. Mitchell. Tailoring authentication protocols to match underlying mechanisms. In J. Pieprzyk and J. Seberry, editors, *Information Security and Privacy — Proceedings: First Aus-*

tralasian Conference, Wollongong, NSW, Australia, June 1996, pages 121–133. Springer-Verlag, Berlin, 1996.

[2] ETSI. *GSM 02.09, Security aspects.* Version 7.0.0 (Release 98).

[3] ETSI. *GSM 03.20, Security related network functions.* Version 7.0.0 (Release 98).

[4] C.J. Mitchell. Security in future mobile networks. In *Proceedings of the Second International Workshop on Mobile Multi-Media Communications (MoMuC-2)*, Bristol, UK, April 1995.

[5] S. Patel. Weaknesses of North American wireless authentication protocol. *IEEE Personal Communications Magazine*, 4(3):40–44, June 1997.

[6] F.C. Piper and M. Walker. Cryptographic solutions for voice telephony and GSM. In *Proceedings of COMPSEC '98*, London, UK, November 1998. Elsevier.