A Man-in-the-Middle Attack on UMTS

Ulrike Meyer
Darmstadt University of Technology
Department of Computer Science
Hochschulstrasse 10
D-64283 Darmstadt
Germany
umeyer@cdc.informatik.tu-darmstadt.de

Susanne Wetzel
Stevens Institute of Technology
Department of Computer Science
Castle Point on Hudson
Hoboken, NJ 07030
USA
swetzel@stevens.edu

ABSTRACT

In this paper we present a man-in-the-middle attack on the Universal Mobile Telecommunication Standard (UMTS), one of the newly emerging 3G mobile technologies. The attack allows an intruder to impersonate a valid GSM base station to a UMTS subscriber regardless of the fact that UMTS authentication and key agreement are used. As a result, an intruder can eavesdrop on all mobile-station-initiated traffic.

Since the UMTS standard requires mutual authentication between the mobile station and the network, so far UMTS networks were considered to be secure against man-in-the-middle attacks. The network authentication defined in the UMTS standard depends on both the validity of the authentication token and the integrity protection of the subsequent security mode command.

We show that both of these mechanisms are necessary in order to prevent a man-in-the-middle attack. As a consequence we show that an attacker can mount an impersonation attack since GSM base stations do not support integrity protection. Possible victims to our attack are all mobile stations that support the UTRAN and the GSM air interface simultaneously. In particular, this is the case for most of the equipment used during the transition phase from 2G (GSM) to 3G (UMTS) technology.

Categories and Subject Descriptors

C.2.0 [**Computer-Communications Networks**]: Security and Protection; C.2.1 [**Network Architecture and Design**]: Wireless Communication

General Terms

Security

Keywords

UMTS, GSM, man-in-the-middle attack, authentication, mobile communication

1. INTRODUCTION

The Universal Mobile Telecommunications System (UMTS) is one of the third generation mobile technologies which is currently being launched in many countries, particularly in Europe.

Unlike its European predecessor, the Global System for Mobile Communication (GSM), UMTS provides high data rates and has a relatively low cost for data transmission. Because of these improved technology features, UMTS is expected to revolutionize the current state of the art in mobile services and applications. This is even more so as handsets equipped with large-size, high resolution displays are currently being developed.

According to market analysts, services mobile users are currently most interested in include e-mail applications followed by mobile payment systems, mobile online banking and mobile shopping [2]. Consequently, many mobile operators and banks have begun collaborating in order to develop these mobile banking, payment and shopping applications for their future UMTS customers.

By nature, these applications are security-critical, particularly with respect to the authenticity of the user as well as to the confidentiality and integrity protection of the data traffic. Security of these applications depends on the security of the underlying mobile communication system. While GSM suffers from many security weaknesses, UMTS was designed to be secure against the known GSM attacks. In this paper we show that this alone, however, is not sufficient as there are security weaknesses in the interface of the two technologies GSM and UMTS.

In particular, while in the process of building up the new UMTS network, the system will not provide area wide UMTS coverage. Since sufficient network coverage, however, is crucial with respect to usability and customer satisfaction, the UMTS standard allows for inter-operation with GSM, ranging from roaming support to seamless handover procedures between the two systems [12, 14]: UMTS subscribers can roam in areas with GSM coverage only and GSM subscribers with UMTS-equipped mobile phones can in turn access the UMTS radio network.

However, this improved functionality is a double-edged sword as roaming between UMTS and GSM is in fact roaming between two systems that are not equally well protected against malicious attacks. GSM only supports subscriber authentication and encryption of the radio interface between a mobile station and a base station. In contrast, UMTS also provides authentication of the serving network as well as integrity protection of the signaling traffic on the air interface between the mobile station and a radio access network.

The UMTS authentication and key agreement procedure was designed to be secure against man-in-the-middle attacks [11]. In order to protect against network impersonation, UMTS applies a combination of two mechanisms: the validity of an authentication token and the integrity protection of the signaling messages. The authentication token guarantees the freshness and the origin (home network) of the authentication challenge, protecting against replay of authentication data. The integrity protection prevents the possibility that a man in the middle can simply forward correct and timely authentication messages and fool both the mobile station and the base station into not using encryption for subsequent communication.

In this paper we show that both mechanisms must be employed in order to prevent man-in-the-middle attacks. In particular, we describe a scenario in which an attacker can impersonate a valid GSM base station with respect to a UMTS subscriber even when UMTS authentication and key agreement are used. Consequently, the attack allows an intruder to eavesdrop on all mobile-station-initiated communication.

Our attack requires that the mobile equipment of the victim supports both the GSM as well as the UTRAN radio interface. As long as UMTS does not provide area-wide coverage, this will be the case for most of the available user equipment.

In our attack an intruder first obtains a valid authentication token from any real network and then uses this token to impersonate a GSM base station subsystem to the user. The attack succeeds since one of the two necessary security components in UMTS authentication can be circumvented as a GSM base station is not expected to support integrity protection. As a consequence, an attacker can fool the mobile station into not using encryption and can thus eavesdrop on all future communication. While the attack does not allow the intruder to impersonate the mobile station to a real network, the traffic from the victim device can be forwarded to a valid network by simply having the attacker establish a regular connection to a valid network.

Outline:

The remainder of the paper is organized as follows: We first provide an overview of related work. In Section 3 we briefly review the security mechanisms defined in the UMTS standard, with a focus on the authentication and key agreement procedure of UMTS subscribers to UTRAN and to GSM base station subsystems. In Section 4 we present the details of our man-in-the-middle attack. We close the paper with some remarks on the feasibility and consequences of our attack and discuss possible countermeasures.

2. RELATED WORK

It is well known that the missing network authentication in GSM networks allows for man-in-the-middle attacks [4, 6]: An attacker impersonates a valid base station with respect to the mobile station and at the same time impersonates the victim mobile station to a real base station by simply forwarding the authentication traffic. As a consequence, the attacker can, for example, eavesdrop on all communication by fooling both sides into the use of no encryption on the radio interface.[1] Moreover, the scenario allows for call theft and other active attacks [4, 6].

Naturally, UMTS subscribers that roam to a GSM (part of a) network are also vulnerable to such false base station attacks during GSM authentication. It has previously been recognized in an internal but publicly available document of the 3GPP standardization organization [9] that in order to prevent the "false base station attacks" for UMTS subscribers roaming to GSM, the integrity protection mechanism must be modified. To date, this change has not been adopted in the standard [13]. Furthermore, the attack described in this paper goes far beyond the attack foreseen by 3GPP in that UMTS subscribers are vulnerable to what 3GPP calls a "false base station attack" even if subscribers are roaming in a pure UMTS network and even though UMTS authentication is applied.

Our attack can be categorized as a "roll-back attack". These attacks exploit weaknesses of old versions of algorithms and protocols by means of the mechanisms defined to ensure backward compatibility of newer and stronger versions. Previous findings of roll-back attacks include the SSL Protocol [7] and the encryption mechanisms in GSM [5].

3. UMTS SECURITY

In UMTS networks, a mobile station is connected to a visited network by means of a radio link to a particular base station (Node B). Multiple base stations of the network are connected to a Radio Network Controller (RNC) and multiple RNCs are controlled by a GPRS[2] Support Node (GSN) in the packet-switched case or a Mobile Switching Center (MSC) in the circuit-switched case (see Figure 1). The Visitor Location Register (VLR) and the serving GSN keep track of all mobile stations that are currently connected to the network. Every subscriber can be identified by its International Mobile Subscriber Identity (IMSI). In order to protect against profiling attacks, this permanent identifier is sent over the air interface as infrequently as possible. Instead, locally valid Temporary Mobile Subscriber Identities (TMSI) are used to identify a subscriber whenever possible.

Every UMTS subscriber has a dedicated home network with which he shares a long term secret key Ki. The Home Location Register (HLR) keeps track of the current location of all subscribers of the home network. Mutual authentication between a mobile station and a visited network is carried out with the support of the current Serving GSN (SGSN) or the MSC/VLR respectively. UMTS supports encryption of the radio interface as well as integrity protection of the signaling messages. For a detailed description we refer to [13].

3.1 Authentication and Key Agreement

UMTS is designed to inter-operate with GSM in order to facilitate the evolution from GSM to UMTS networks. This results in different versions of the authentication and key agreement procedure depending on the type of the sub-

[1] In order to protect GSM networks against man-in-the-middle attacks, 3GPP currently discusses to add structure to the authentication challenge RAND [8]. This structure shall indicate to the mobile station which encryption algorithms it is allowed to use with the key generated from this particular RAND. Adapting this kind of mechanism to UMTS networks has not yet been discussed.

[2] General Packet Radio Service

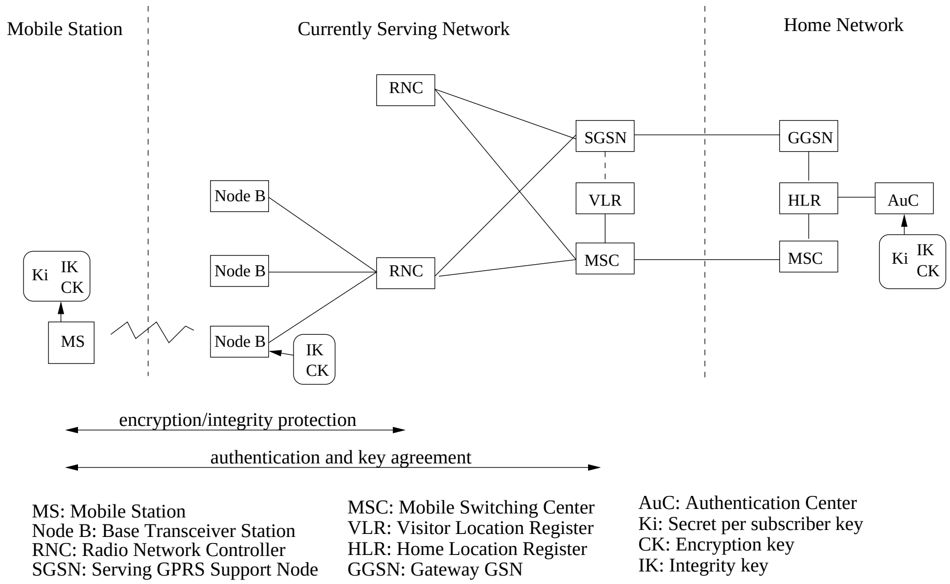

Figure 1: UMTS architecture and storage of secret keys

scriber, the capabilities of his mobile equipment and the capabilities of the currently serving network. A comprehensive description of all scenarios can be found in [14]. In the following we will focus only on those two versions of the authentication and key agreement protocols that are of relevance for our attack. The first one is the standard scenario where a UMTS subscriber wants to connect to a UMTS network. In the second one, a UMTS subscriber (with a mobile device that allows him to roam between GSM and UMTS networks) connects to a serving network operating a GSM base station subsystem, while the backbone components SGSN and MSC are standard UMTS components.

3.1.1 Case 1: UMTS Subscriber - UMTS Network

Both the network and the mobile station support all UMTS security mechanisms. The authentication and key agreement is as follows (see also Figure 2):

1. The mobile station and the base station establish a Radio Resource Control connection (RRC connection). During the connection establishment the mobile station sends its security capabilities to the base station. The security capabilities include the supported UMTS integrity and encryption algorithms and optionally the GSM encryption capabilities as well.
2. The mobile station sends its current temporary identity TMSI to the network.
3. If the network cannot resolve the TMSI it requests the mobile station to send its permanent identity and the mobile stations answers the request with the IMSI.
4. The visited network requests authentication data from the home network of the mobile station.
5. The home network returns a random challenge RAND, the corresponding authentication token AUTN, authentication response XRES, integrity key IK and the encryption key CK.
6. The visited network sends the authentication challenge RAND and the authentication token AUTN to the mobile station.
7. The mobile station verifies AUTN and computes the authentication response. If AUTN is not correct the mobile station discards the message.
8. The mobile station sends its authentication response RES to the visited network.
9. The visited network checks whether RES=XRES and decides which security algorithms the radio subsystem is allowed to use.
10. The visited network sends the allowed algorithms to the radio subsystem.
11. The radio access network decides which (of the allowed) algorithms to use.
12. The radio access network informs the mobile station of its choice in the security mode command message. The message also includes the security capabilities the network received from the mobile station in step 1. This message is integrity protected with the integrity key IK.
13. The mobile station validates the integrity protection and checks the correctness of the security capabilities.

3.1.2 Case 2: UMTS Subscriber - GSM Base Station

The mobile unit (UMTS subscriber) supports both USIM and SIM application. The base station system uses GSM technology while the VLR/MSC respectively the SGSN are UMTS components. The mobile station and the core network both support all UMTS security mechanisms. However, the GSM Base Station System (BSS) does not support integrity protection and only uses the GSM encryption algorithms. The first eight steps of the authentication protocol

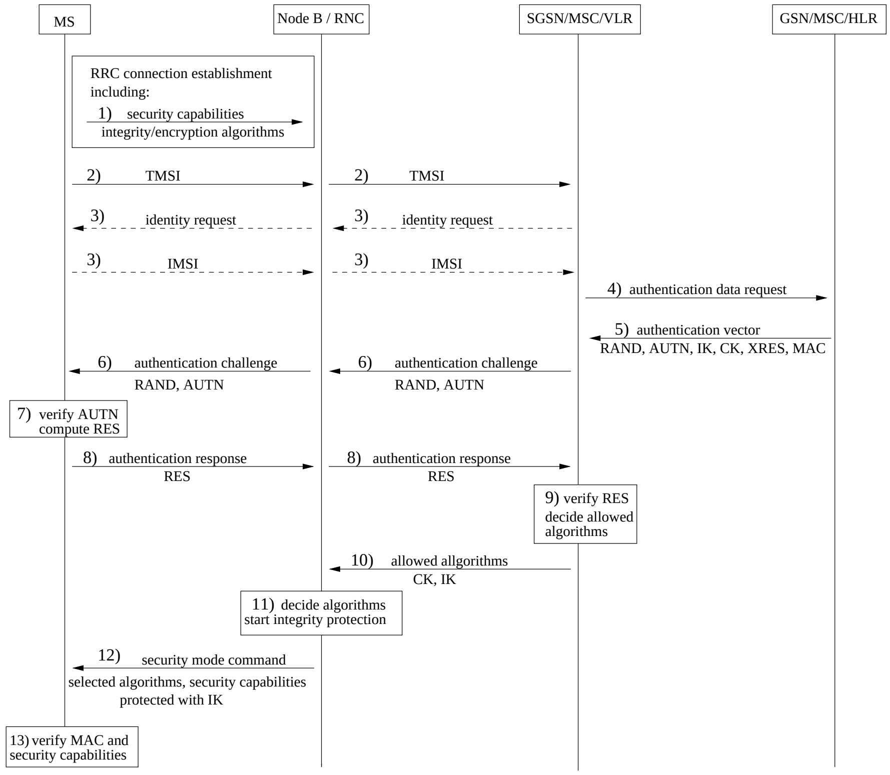

Figure 2: Authentication and key agreement in standard UMTS networks

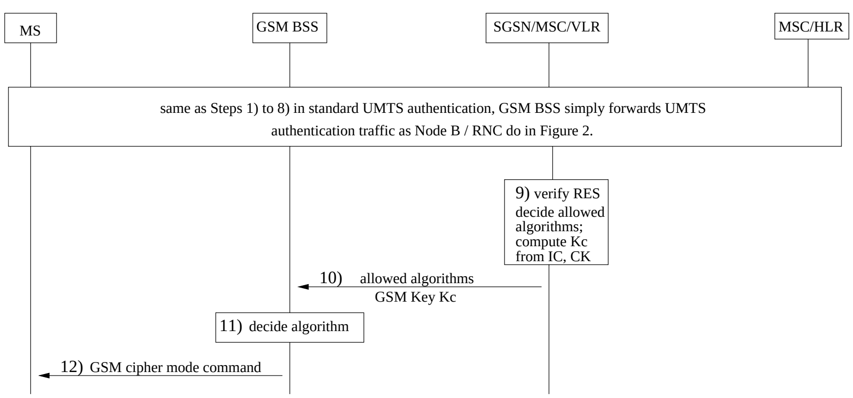

Figure 3: Authentication and key agreement in UMTS networks with GSM components

are carried out as in the standard case (see Figure 2). The GSM BSS simply forwards the UMTS authentication traffic.

9. The MSC/SGSN decides which GSM encryption algorithms are allowed and computes the GSM key Kc from the UMTS keys IK, CK.

10. The MSC/SGSN advises the GSM BSS of the allowed algorithms and forwards the GSM encryption key Kc.

11. The GSM BSS decides which of the allowed encryption algorithms to use depending on the ciphering capabilities of the mobile station.

12. The GSM BSS sends the GSM cipher mode command to the mobile station.

The GSM cipher mode command message includes the encryption algorithm to be used to encrypt the radio interface between the mobile unit and the network. Currently "no encryption" and three encryption algorithms are defined [3].

4. MAN-IN-THE-MIDDLE ATTACK ON UMTS

4.1 Protection of UMTS-Only User Equipment against Man-in-the-Middle Attacks

In order to mount a man-in-the-middle attack against a user of a UMTS-only mobile station (see Case 1 in Section 3.1), an attacker would have to impersonate a valid network to the user. However, in the UMTS-only equipment case, the combination of two specific security mechanisms protects the mobile station from this attack: the authentication token AUTN and the integrity protection of the security mode command message in Step 12 of Figure 2. The authentication token ensures the timeliness and origin of the authentication challenge and as such protects against replay of authentication data. The integrity protection prevents an attacker from simply relaying correct authentication information while fooling the respective parties into not using encryption for subsequent communication.

In particular, AUTN contains a sequence number SQN and a message authentication code MAC. On receipt of AUTN (Step 6), the mobile station first checks the message authentication code MAC. A correct MAC indicates that the authentication token was originally generated by the home network. The mobile station then extracts the sequence number SQN. If the sequence number is in the right range, the mobile station is assured that AUTN was issued recently by its home network. Otherwise, the mobile station knows that either AUTN is a replay of an old value or the synchronization of the sequence number failed. (A more detailed description of the procedure is given in [13].)

It is important to note that the correctness of the MAC and AUTN alone do not provide assurance to the mobile unit that the token was in fact received directly from the authorized network and not relayed by an attacker. It is only the combination with an additional integrity protection of the signaling messages that prevents network impersonation: The message in Step 12 is not only integrity protected but more importantly also includes the security mode capabilities that the mobile unit originally announced in Step 1. By checking the correctness of the integrity protection, the mobile station is assured that this message was generated by a network entity that is in possession of the right integrity key. Furthermore, including the security capabilities of the mobile station in the integrity protected message in Step 12 is crucial in that it prevents both the mobile unit and the network from being fooled into using no encryption (or weak encryption) by an attacker. In order to succeed, an attacker would have to forge the integrity protection on the security mode command message, which is assumed to be infeasible [10]. If the security capabilities of the mobile station were not repeated back in Step 12, the attacker could easily forge the protection, as message 1 is not integrity protected. An attacker could therefore request no or weak encryption on behalf of the victim mobile station (instead of its original security capabilities). In turn, the attacker would inform the mobile station of the choice of no (weak) encryption by the network in Step 12.

4.2 Vulnerability of Combined UMTS/GSM User Equipment

Unlike in standard UMTS networks, the message sent in Step 12 in hybrid GSM/UMTS networks, as described in Case 2 (see Section 3.1), is neither integrity protected nor does it repeat the parameters previously announced by the mobile unit in Step 1. As such, the message can be easily forged by an attacker. This limitation is due to the fact that GSM does not support integrity protection. In the following we detail a man-in-the-middle attack which exploits this shortcoming, i.e., we show that an attacker can impersonate a GSM base station to a UMTS subscriber using the UMTS authentication procedure of the hybrid GSM/UMTS scenario.

In order to mount our attack, we assume that the attacker knows the IMSI of his victim. This is a reasonable assumption as the attacker can easily obtain the IMSI from the mobile station by initiating an authentication procedure prior to the attack (see Step 3 of Figure 2) and disconnecting from the mobile station after receiving the IMSI. Note that by doing so, the attacker also learns the security capabilities of the mobile station.[3] A mobile station always connects to the base station which provides the best reception. An attacker can easily enforce this kind of a setting, i.e., make a victim device connect to him instead of a real base station by drowning the real base stations that are present by sending its beacons with higher transmitting power. Our attack works in two phases:

Phase 1:

The attacker acts on behalf of the victim mobile station in order to obtain a valid authentication token AUTN from any real network by executing the following protocol:

1. During the connection setup the attacker sends the security capabilities of the victim mobile station to the visited network.

2. The attacker sends the TMSI of the victim mobile station to the visited network. If the current TMSI is unknown to the attacker, he sends a faked TMSI (which eventually cannot be resolved by the network).

3. If the network cannot resolve the TMSI, it sends an

[3]The feasibility of this attack is already recognized in the UMTS specification [11].

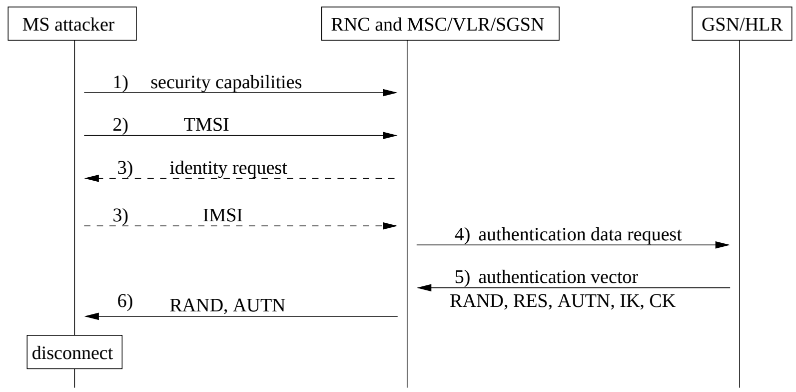

Figure 4: Phase 1 – Attacker obtains currently valid AUTN

identity request to the attacker. The attacker replies with the IMSI of the victim.

4. The visited network requests the authentication information for the victim device from its home network.

5. The home network provides the authentication information to the visited network.

6. The network sends RAND and AUTN to the attacker.

7. The attacker disconnects from the visited network.

Since none of the messages sent in Steps 1 to 7 are protected by any means, the network cannot recognize the presence of the attacker. Consequently, the attacker obtains an authentication token which he in turn can use in Phase 2 of the attack to impersonate a network to the victim device.

Phase 2:

The attacker impersonates a valid GSM base station to the victim mobile station.

1. The victim mobile station and the attacker establish a connection and the mobile station sends its security capabilities to the attacker.

2. The victim mobile station sends its TMSI or IMSI to the attacker.

3. The attacker sends the mobile station the authentication challenge RAND and the authentication token AUTN he obtained from the real network in Phase 1 of the attack.[4]

4. The victim mobile station successfully verifies the authentication token.

5. The victim mobile station replies with the authentication response.

6. The attackers decides to use "no encryption" (or weak encryption, e.g., a broken version of the GSM encryption algorithms [3]).

7. The attacker sends the mobile station the GSM cipher mode command including the chosen encryption algorithm.

Note that the attack does not allow the intruder to impersonate the mobile station to the network at the same time. In order to allow for a regular use of the connection by the victim unit, the attacker has to establish a regular connection to a real network to forward traffic it receives from the mobile station. As a side effect the attacker has to pay the cost for this connection.

Feasibility of the Attack:

An attacker trying to impersonate a valid network to a UMTS subscriber has to overcome two difficulties: he has to send (or forward) a valid authentication token to the victim mobile station and he has to ensure that no encryption is used after the authentication.

In our attack, the intruder solves the first problem by impersonating the victim mobile station to a real network in order to obtain a valid authentication token. This is possible since none of the respective messages are (integrity) protected.

Requesting no encryption or weak encryption is more difficult as it is the radio access network deciding which encryption algorithm is used (in both the GSM and the UTRAN case). The decision strongly depends on the security capabilities of the mobile unit which are sent to the network during connection setup (see Step 1 in Figures 2 and 3). In principle, both radio access networks (i.e., GSM and UTRAN) allow "no encryption".

In UTRAN, the security mode command message that informs the mobile station which algorithm to use is integrity protected. This alone, however, does not protect against network impersonation. The attacker could still fool a valid network into integrity protecting a "no encryption" message with the right key by sending false information about the encryption capabilities of the victim mobile station. However, in the integrity protected security mode command message the network sends back the security capabilities it received to the mobile station. Therefore, unless the attacker can forge the integrity check, the mobile station would thus detect the attack.

In the GSM case, though, integrity protection is not supported. As a consequence, the corresponding cipher mode

[4]The mobile station accepts the authentication token if the token is fresh, i.e., not too much time has elapsed between Phase 1 and Phase 2.

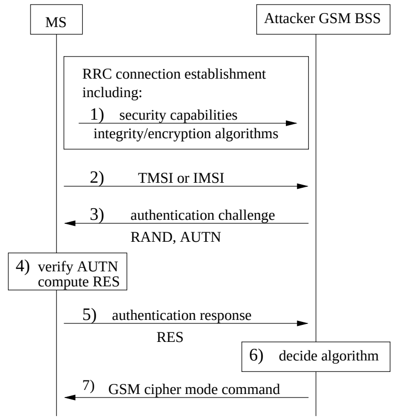

Figure 5: Phase 2 – Attacker impersonates valid GSM base station to the victim

command message is not integrity protected, thus allowing an attacker to easily forge this message and fool the victim mobile station into using either no encryption or a weak encryption algorithm. Eventually, the attacker is able to eavesdrop on all mobile station initiated communication.

Our attack only works as long as the time gap between Phase 1 and Phase 2 is small enough so that no other authentication between the victim mobile unit and another network takes place. Otherwise, the sequence number within the authentication token might be out of range.

As stated before, our attack does not work against mobile equipment that is capable of the UTRAN interface only. Yet, in the transition phase from GSM to UMTS, most users are expected to use equipment that is capable of both the UTRAN radio interface and GSM.

The UMTS specification includes an optional (and not fully specified) display of the current encryption and integrity protection state [13]. If this is implemented in the victim's mobile station, the victim may be able to suspect the attack, depending on the details provided to the user. If, for example, the mobile station only displays encryption on/off and the attacker uses a broken algorithm like A5/2 for encryption [1], the subscriber will not be able to detect the attack. On the other hand, if the mobile station displays UMTS/GSM authentication only the victim may be mislead about his current level of protection.

4.3 Countermeasures

While in theory it would be possible to avoid this attack by disallowing roaming to GSM, this is not a practical option, primarily because of economical reasons. Instead, the authentication procedure would need to be changed in order to protect against our man-in-the-middle attack. The problem can be fixed by not only placing the generation of the integrity check on the cipher mode command message back in the MSC/VLR, but also mandating the inclusion of the security mode capabilities of the mobile station in the integrity protected message.

5. CONCLUSION AND FUTURE WORK

The UMTS specification allows for a variety of different combinations of UMTS and GSM user equipment, subscriber identity modules and radio access networks. In order to protect UMTS subscribers from attacks known to GSM networks, the UMTS specification does not allow UMTS-capable equipment to carry out GSM authentication and key agreement unless the network is not capable of UMTS authentication and key agreement [13]. However, in this paper we have detailed an attack which shows that the use of the currently specified UMTS authentication and key agreement procedures are not sufficient in order to protect UMTS subscribers from man-in-the middle attacks. Implementing countermeasures to thwart the attack will require modification of the UMTS standard.

A future direction of research is to investigate whether the mechanism currently discussed by 3GPP to introduce structure to RAND in order to prevent man-in-the-middle attacks in GSM can be adapted for UMTS networks in order to thwart man-in-the-middle attacks.

6. REFERENCES

[1] E. Barkan, E. Biham, and N. Keller. Instant Ciphertext-Only Cryptanalysis of GSM Encrypted Communication. In *Advances in Cryptology - CRYPTO 2003*, LNCS 2729, August 2003.

[2] Deutsche Bank Research. Mobile Banking's Banana Problem: Too Little Business in Sight. *E-Banking snapshot*, 2002. http://www.dbresearch.com.

[3] ETSI Technical Specification. ETSI TS 100.929, V8.0.0, Digital Cellular Telecommunications System (phase 2+)(GSM); Security Related Network Functions. 2000.

[4] D. Fox. Der IMSI-Catcher. *DuD, Datenschutz und Datensicherheit*, 2002.

[5] G. Horn and P. Howard. Review of Third Generation Mobile System Security Architecture. In *ISSE 2000*, September 2000.

[6] C. J. Mitchell. The Security of the GSM Air Interface Protocol. Technical Report RHUL-MA-2001-3, Royal Holloway University of London, 2001.

[7] D. Wagner and B. Schneier. Analysis of the SSL 3.0 Protocol. In *2nd USENIX Workshop on Electronic Commerce*, November 1996.

[8] 3GPP. S3-030651: Further Development of the Special RAND Mechanism. `http://www.3gpp.org/TB/SA/SA3/SA3.htm`.

[9] 3GPP. S3-99206: Response to "CR to TS 25.301 - Integrity Control Mechanism". http://www.3gpp.org/TB/SA/SA3/SA3.htm.

[10] 3GPP Technical Report. 3GPP TR 33.909 V1.0.0, Third Generation Partnership Project; Technical Specification Group Services and System Aspects; Report on the evaluation of 3GPP Standard Confidentiality and Integrity Algorithms. December 2000.

[11] 3GPP Technical Specification. 3GPP TS 21.133, V4.1.0, Third Generation Partnership Project; 3G Security; Security Threats and Requirements. December 2001.

[12] 3GPP Technical Specification. 3GPP TS 23.009, V5.6.0, Third Generation Partnership Project; Handover Procedures. October 2003.

[13] 3GPP Technical Specification. 3GPP TS 33.102, V5.3.0, Third Generation Partnership Project; Technical Specifications Group Services and System Aspects; 3G Security; Security Architecture. September 2003.

[14] 3GPP Technical Report. 3GPP TR 31.900, V5.3.0., Third Generation Partnership Project; SIM/USIM Internal and External Interworking Aspects. 2003.