The ISC Int'l Journal of Information Security

January 2009, Volume 1, Number 1 (pp. 3–15)

http://www.isecure-journal.org

A TESLA-Based Mutual Authentication Protocol for GSM Networks

Ali Fanian [a,*], Mehdi Berenjkoub [a], T. Aaron Gulliver [b]

[a] *Department of Electrical and Computer Engineering, Isfahan University of Technology (IUT), Isfahan, Iran*

[b] *Department of Electrical and Computer Engineering, University of Victoria, Victoria, BC Canada*

ARTICLE INFO

Article history:
Received: 26 May 2008
Revised: 26 October 2008
Accepted: 4 December 2008
Published Online: 28 January 2009

Keywords:
GSM, Entity Authentication, Bilateral Authentication, Unilateral Authentication, Man-in-the-Middle Attack, TESLA Protocol

ABSTRACT

The widespread use of wireless cellular networks has made security an ever increasing concern. GSM is the most popular wireless cellular standard, but security is an issue. The most critical weakness in the GSM protocol is the use of one-way entity authentication, i.e., only the mobile station is authenticated by the network. This creates many security problems including vulnerability against man-in-the-middle attacks. Several solutions have been proposed to establish mutual entity authentication. However, none provide a flaw-free bilateral authentication protocol. In this paper, we show that a recently proposed solution is vulnerable to a "type attack". Then, we propose a novel mutual entity authentication using the TESLA protocol. The proposed solution not only provides secure bilateral authentication, but also decreases the call setup time and the required connection bandwidth. An important feature of the proposed protocol is that it is compatible with the GSM standard.

1 Introduction

The Global System for Mobile communications (GSM) is widely used internationally as a standard for wireless cellular networks [1]. The GSM architecture consists of mobile stations (MS) and base transceiver stations (BTS), which communicate with each other through radio links. Each BTS is connected to the mobile switching center (MSC), which is responsible for routing signals to and from fixed networks [1, 2]. GSM uses several databases for management and authentication purposes. The Home Location Register (HLR) contains information about every user in the system and their location. The Visitor Location Register (VLR) is a database located in every MSC which contains information about visiting users. The Authentication Center (AuC) is the user authentication center, and contains the secret key K_i that subscriber i shares with the system. The secret key K_i is used to generate the set of security parameters for user authentication. Each GSM subscriber saves his secret key K_i in the Subscriber Identity Module (SIM) card. During the personalization process, every SIM card gets a unique identity and an International Mobile Subscriber Identity (IMSI) from the AuC. The GSM architecture is shown in Figure 1.

GSM authentication is based on a challenge/response protocol in which the network sends a challenge to each MS and the MS must send back an appropriate response. This is a unilateral authentication protocol in which the network cannot be authenticated by the MS. Thus an adversary can introduce himself as a valid network to a user via a man-in-the-middle attack. In this way, an adversary can capture the

* Corresponding author.
Email addresses: Fanian@ec.iut.ac.ir (A. Fanian), Brnjkb@cc.iut.ac.ir (M. Berenjkoub), a.gulliver@ieee.org (T.A. Gulliver).

ISSN: 2008-2045

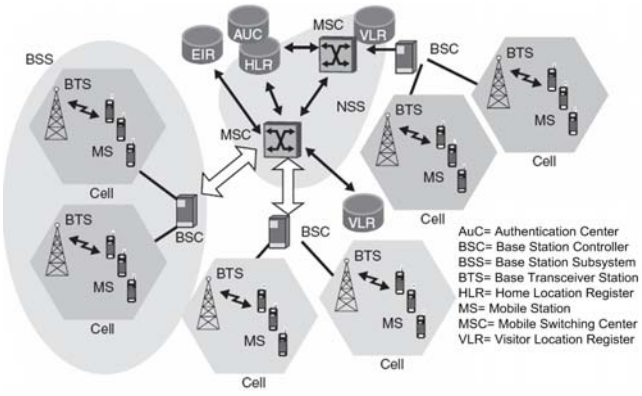

Figure 1. The GSM architecture

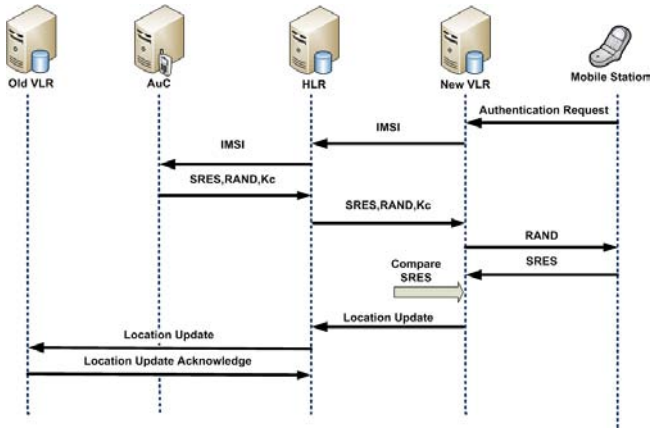

Figure 2. The GSM authentication protocol

IMSI of an MS and even discover K_i using a collision attack [3]. To do this, an adversary must send a large number of requests to an MS and analyze the responses.

Although confidentiality and authentication services have been improved in the Universal Mobile Telecommunication System (UMTS) to solve the main security problems of GSM, GSM systems will exist for a long time to come. This is the reason that new authentication schemes are still being proposed to improve the original GSM authentication process [4].

In this paper, we propose a new mutual authentication protocol based on the TESLA. This protocol increases authentication efficiency and decreases the delay due to the authentication process. The rest of the paper is organized as follows. Section 2 reviews the GSM authentication protocol and related work. Section 3 discusses the drawbacks of a recently proposed authentication protocol by Chang *et al.* [5]. Details of our scheme are described in Section 4. Performance evaluation of the proposed approach and an analysis of the message overhead are conducted in Section 5. Finally, some conclusions are given in Section 6.

2 Preliminaries

From the security perspective, the most important component in GSM is the authentication protocol. We discuss the GSM authentication protocol and its main drawbacks. Then some approaches to improving this protocol are presented. The notation used throughout the paper is given in Table 1 and the current GSM authentication protocol is schematically depicted in Figure 2.

Table 1. Notation

Notation	Description
HLR	The home location register
VLR	The visitor location register
AuC	Authentication center
MSC	Mobile switching center
BSC	Base station controller
BTS	Base transceiver stations
MS	Mobile station
IMSI	The international mobile subscriber identity
TMSI	The temporary mobile subscriber identity
LAI	The location area identity
VLR_ID	The identification of VLR
Ki	The temporary secret key shared between MS and HLR
Tv	The timestamp generated by MS
SRES	The signed result for the first time of the authentication
$SRES_j$	The signed result for the j'th (j>1) time of the authentication
MAC	Message authentication code
RAND	The random number generated by HLR
A3,A8	The two algorithms which are used for authentication and key generation
A5	The data encryption algorithm
PKI	Public key infrastructure
SIM	Subscriber Identity Module
SDCCH	Standalone Dedicated Control Channel
||	Message Concatenation

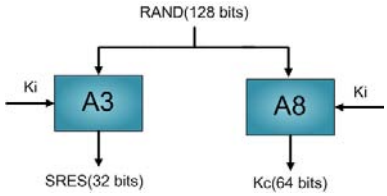

Figure 3. Method of derivating $SRES$ and K_c.

2.1 Review of the Current Authentication Protocol for GSM

Based on Figure 1, the GSM authentication protocol can be described by the following steps:

- Step1: When the MS enters a new visiting area and requests a communication service, an authentication request is sent to the BTS via the SDCCH channel. This request includes TMSI and LAI. The latter, in fact, is an old location area identity which introduces the old VLR to the new VLR.
- Step 2: After receiving the request, the BTS sends it to the corresponding VLR and the new VLR uses the received TMSI to get the IMSI from the old VLR. It then sends IMSI to HLR.
- Step 3: With the assistance of the AuC, HLR generates n separate sets of authenticating parameters $\{SRES, RAND, K_c\}$ and sends them to the VLR. The value of n is typically 5. The $RAND$ parameter is a network challenge, $SRES$ is an appropriate response to it and K_c is the session key. Figure 3 shows how $SRES$ and K_c are derived.
- Step 4: After receiving the sets of authenticating parameters, the VLR adds them to its database and selects one to authenticate the mobile station for each authentication request. VLR sends the selected $RAND_i$ to the MS
- Step 5: Once MS receives $RAND_i$ from the VLR, it computes $SRES' = A3(RAND_i, K_i)$ and the temporary session key $K_c' = A8(RAND_i, K_i)$, respectively, where K_c' is kept secret for consequent communications. Then $SRES'$ is sent back to the VLR.
- Step 6: Upon receiving $SRES'$ from the MS, the VLR compares it with the selected $SRES$ kept in its own database. If they are not the same, the authentication fails; otherwise, VLR has authenticated the MS. In the first authentication, VLR sends the location update message to HLR. HLR, in turn, sends the same message to the old VLR and receives from it a confirmation.

2.2 GSM Authentication Protocol Problems and Related Work

Several drawbacks to the GSM authentication protocol have been identified [6, 7, 4, 5]. The main ones are the following:

(1) Mutual authentication between the MS and VLR is not provided by the GSM protocol. Only the MS is authenticated by the VLR, and mobile users cannot authenticate the network.
(2) Excessive bandwidth consumption between the VLR and HLR.
(3) The VLR suffers from space overhead. For each MS in the visiting VLR, there are n sets of authenticating parameters stored in the VLR database.
(4) If the MS stays in the same VLR for a long time and consumes all of the authenticating parameters, the VLR will again request the HLR for n sets of authenticating parameters. This increases the bandwidth consumption and the HLR load significantly.

In the last decade, many alternative authentication protocols for GSM have been proposed. Among them, there is an important group based on asymmetric cryptography. Aydemir [8] presented a strong user authentication protocol for GSM, which permits users to access their accounts remembering only a password without being limited to their SIM card. In this protocol, users have to enter their password for every use of the mobile device. The proposed protocol uses asymmetric cryptography which is not suitable for mobile devices due to the high computation load. Lin and Jan [9] proposed a mutual authentication protocol in which a user takes a ticket during user authentication in the first phase of the protocol. The second phase of the protocol allows the MS to anonymously access the system for a limited period of time. This is done via prepaid tickets generated by the HLR, which can be used with the VLR. This protocol also uses asymmetric cryptography. Peinado [4] proposed another PKI based on mutual authentication similar to that in [9]. This protocol also has two phases. In the first phase, a user takes a ticket from the HLR during initial authentication and uses it for subsequent authentications. Despite the lack of any need to change the GSM architecture, the protocol is still based on PKI and the MS has a heavy computational load. Bocan [6] proposed a new security mechanism to reinforce the protocol against a DoS (Denial of Service) attack. This proposal, however, requires significant processing time and memory from both the MS and the network. In addition, the network has to continuously send broadcast messages and wait for the MS to respond.

The other group of proposed improvements retains the MS restrictions and employs symmetric cryptography. Al-Tawil *et al.* [7] proposed a new authentication protocol with less signaling traffic and a better call set-up time. This protocol cannot solve drawbacks 1 to 4, but adds a "mobile user events counter" $CountM$ into the HLR and the MS SIM card. Hwang *et al.* [10] proposed a mutual authentication technique for the GSM network. The main idea is for the HLR to issue a ticket for the new VLR the first time the MS is authenticated in the new location. The VLR and MS authenticate each other using the ticket without any need to refer to the HLR. Detailed analysis shows that this technique only works correctly for the first MS authentication, and cannot subsequently verify the network [5]. Chang *et al.* [5] developed a modified version of [10] with the removal of this security flaw. However, this scheme suffers from message indistinguishability which is introduced in the next section.

3 The Chang *et al.* Protocol and a New Attack Against it

In this section, the protocol of Chang *et al.* [5] is presented, and a new "type attack" against it is introduced.

3.1 The Chang *et al.* Protocol

After the MS joins a new visiting area, protocol phase I, depicted in Figure 4, is used for the initial authentication. In the MS request, the timestamp T_1 is increased to the initial message. After obtaining the corresponding IMSI from the received TMSI, the new VLR forwards the MS request to the HLR. Afterwards, to authenticate the VLR by the MS, the HLR generates a certificate for the VLR as follows: $Cert\_VLR = A3(T_1, K_i)$. The HLR also generates a random number, R, and a temporary key, K_T, for future MS authentication as follows: $K_T = A3(R, K_i)$. The HLR sends $Cert\_VLR$, R and K_T to the VLR through a secure channel. The VLR sends $Cert\_VLR$, T_1 and R to the MS and computes $SRES$ as follows: $SRES = A5(R\|T_1, K_T)$.

After receiving a message from the VLR, the MS checks the validity of T_1, and if it is valid, computes $A3(T_1, K_i)$ and compares it with $Cert\_VLR$. If they match, the network is authenticated to the MS. Also, the MS computes K_T and then $SRES' = A5(R'\|T_1, K_T)$ and sends the latter to the VLR. Finally, the protocol ends with a comparison of $SRES$ and $SRES'$ by the VLR. If they are equal, the MS is authenticated; otherwise, the MS request is denied.

To achieve mutual authentication in subsequent

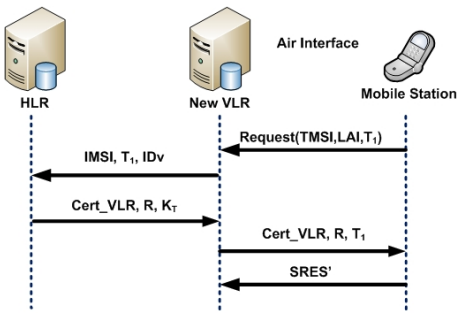

Figure 4. Phase I of the protocol in [5]

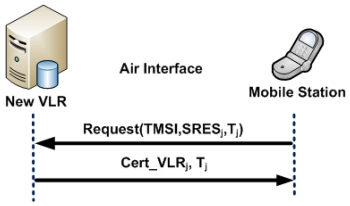

Figure 5. Phase II of the protocol in [5]

connections, Chang *et al.* in [5] proposed protocol phase II as shown in Figure 5. In this phase, the MS request includes $SRES_j$ which authenticates the MS to the VLR. $SRES_j$ is obtained by concatenation of T_j, T_{j-1} and K_T as the input sequence for algorithm A5 as follows: $SRES_j = A5(T_j\|T_{j-1}, K_T)$, where T_j and T_{j-1} are the current and previous timestamps, respectively, generated by the MS. After receiving the MS request, the VLR first verifies it by checking $SRES_j$ and then generates a new version of the certificate, namely $Cert\_VLR_j$, in order to authenticate itself to the MS as follows: $Cert\_VLR_j = A3(T_j, K_T)$.

3.2 The New Attack

Phase I of the Chang *et al.* authentication protocol is vulnerable to attack. As shown in Figure 6, an adversary can change the message in order to change the temporary key K_T generated by the HLR, and use this for MS authentication and the generation of a new VLR certificate. If an adversary can replace the temporary key, he can execute a man-in-the-middle attack, and impersonate the BTS as an authenticated network. To modify the temporary key, an adversary who masquerades as a BTS receives the MS authentication request and forwards this request to the VLR via the genuine BTS. MS authentication will continue between the VLR and HLR until the VLR sends back the VLR certificate, R and T_1 to the adversary. When the adversary receives this information from the VLR, he modifies the received message and sends it back to the MS. However, in this new message R is replaced by T_1 which causes K_T to change without the MS noticing that this message has been forged. When the MS receives the information from the masquerading adversary, he checks the validity of $Cert\_VLR$ and finds

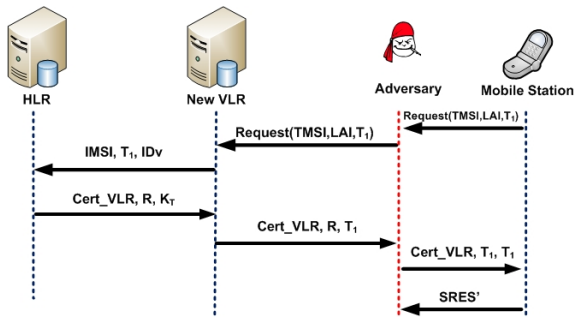

Figure 6. An attack on Chang's protocol

that the network is legal. Then K_T is computed from the received parameters. Due to R being changed by the masquerading BTS, the MS generates K'_T which is identical to $Cert\_VLR$, and makes $SRES'$ with use of K'_T and sends it to the VLR. The adversary, however, does not allow this message to be received by the VLR, and instead, diverts the consequent traffic to the fake network. In this stage, the masquerading BTS and the MS have a shared temporary key, K'_T. Afterward, whenever the MS wants to start a connection with sending a request to the network based on phase II of the protocol, the masquerading BTS will be able to respond to it by generating the acceptable $Cert\_VLR_j$. So, an adversary can introduce itself as an authorized BTS and modify and exchange MS traffic. Using the same algorithm for generating $Cert\_VLR$ and K_T allows the adversary to make a "type attack". In order to prevent this attack, it is sufficient to modify the algorithm to generate K_T, which is different from the structure of $Cert\_VLR$. Instead, we consider a different approach in the next section in order to achieve a more efficient solution.

4 The Proposed TESLA-Based Mutual Authentication Protocol For GSM

In this section, an efficient mutual authentication protocol for GSM is introduced. The main objective of the protocol is to enable the MS to authenticate the network. This is achieved using the TESLA protocol. In addition, we solve other GSM authentication problems presented in section 2 related to efficiency. In fact, our protocol employs concurrent acquisition to decrease memory overhead in the VLR, to reduce the processing loads for the MS and HLR, to reduce the control messages for authentication, and to improve call setup time. In our protocol, the HLR creates a temporary key for the MS and VLR in order to achieve MS authentication over a limited period of time. We next introduce the TESLA protocol and then describe our proposed protocol.

4.1 The TESLA Broadcast Authentication Protocol

One of the main challenges in securing broadcast communications is source authentication. This mechanism allows receivers of broadcast data to verify that the data really originates from the claimed source and is not modified during transmission. A typical solution for source authentication is based on PKI. In this solution, the source signs data using his private key. Today, the most widely used certification systems are PGP [11] and X.509 [12]. Both rely on public key cryptography, which makes them unsuitable for low-powered, computationally constrained devices such as mobile devices [13]. Thus, PKI is not considered here.

TESLA [13, 14] is a broadcast authentication technique that provides asymmetric properties in spite of using symmetric cryptographic functions. It uses a MAC function as a security engine that is very suitable for low-power devices. The asymmetric property is achieved by delaying key disclosure. This approach has been used for authentication in network environments [15]. A requirement of TESLA is loose synchronization between nodes in the network.

TESLA divides time into intervals of equal duration and assigns a corresponding key (tK_n) to each time slot t_n [16]. To broadcast data during time interval n, the sender appends to every packet a MAC, computed by the secret key tK_n. Each receiver buffers the packets but can't verify their originality until the sender discloses the key tK_n by broadcasting the corresponding key seed s_n. Note that after s_n is disclosed, anyone can compute tK_n and impersonate the genuine sender by forging MACs. Therefore, the use of tK_n in computing MACs is restricted to time interval n. In TESLA, the sender discloses s_n after passing d time slots, where d is called the "detection delay" and is selected to equal the maximum network delay for all receivers. The keys tK_n are derived from the key-seeds using a publicly available one-way function F'. The key-seeds are related to each other via a reverse-time chain of one-way functions. To create the chain of key-seeds, the sender chooses a terminal seed s_l, and generates the previous seed s_{l-1} using a one-way function F. Similarly, the remaining seeds $\{s_0, s_1, s_2, ..., s_l\}$ are derived as follows: $S_{k-1} = F(S_k), k = l, l-1, ..., 1$. The sender uses the seed-chain in the opposite direction (starting with seed s_0) to derive the TESLA keys by applying the one-way function F' as shown in Figure 7 .

When a user receives a packet, he first checks whether the packet is fresh (i.e. it was sent in a timeslot whose TESLA key hasn't been disclosed yet) or not. The receiver discards all dated packets

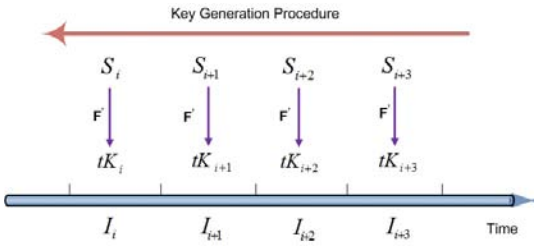

Figure 7. The seed-chain and derivation of the TESLA keys

and buffers only the fresh ones. Once the user receives a TESLA-seed s_n, he checks $F(s_n) = s_{n-1}$ to be sure of s_n's authenticity. He derives tK_n by $tK_n = F'(s_n)$; and authenticates the packets that were sent in timeslot n. It should be mentioned that correct functionality of TESLA requires that at least one of the seeds is formerly authenticated in some way by the receivers, for instance by means of PKI or pre-shared key. Afterwards, each TESLA key can be verified using the previous one.

4.2 An Efficient Authentication Protocol for GSM

The goal of our protocol is mutual authentication which not only overcomes the drawbacks described in Section 2, but also makes the authentication mechanism more efficient. Mutual authentication in our protocol consists of two unilateral authentication protocols which are independent from each other. One is used for network authentication, and the other performs MS authentication. Thus our proposed scheme consists of three parts. Part one is executed for network authentication. Part two is executed once the MS joins the visiting VLR and requests the first authentication. Part three is employed for the jth ($j > 1$) MS authentication request. These parts are described below.

4.2.1 Part 1: Network Authentication

As mentioned in the previous section, the TESLA protocol is used to authenticate a node in broadcast applications. In GSM, there are some broadcast channels that the BTS and MS can use for different tasks. The broadcast channel in the downlink direction contains some logical channels which can be used to broadcast control messages. These logical channels include Synchronization Channels (SCHs), Frequency Correction Channels (FCCHs), and Broadcast Control Channels (BCCHs) for transmitting system information, etc. These channels are arranged in carrier number zero. As with the other GSM channels, this channel is divided into 8 time slots via TDMA, and the zero time slot is used for this purpose [17]. Each carrier has a bandwidth of 200 KHz. Further, GSM

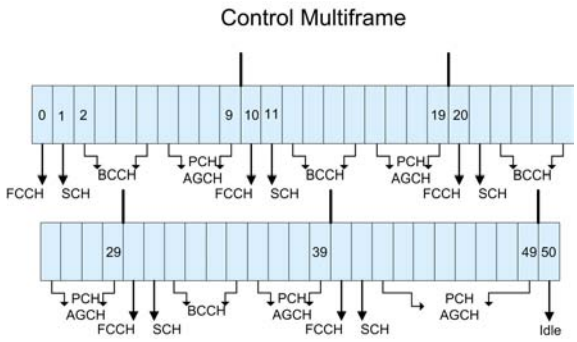

Figure 8. The arrangement of logical broadcast channels in GSM control multiframe

defines both control multiframe and data multiframe. A control multiframe is composed of 51 frames with time duration 235.4 ms but a data multiframe contains 26 frames with time duration 120 ms.

In Figure 8, the arrangement of logical broadcast channels is shown for the control multiframe. This figure shows that the 51st slot is free and the GSM does not use it to send special information. Thus, this channel can be utilized for broadcast transmission of the TESLA protocol. So the time period will be 235.4 ms if an idle logical channel is used.

In order to authenticate each BTS to the mobile phones in the coverage area of its corresponding cell, the Tesla message is applied as shown in Figure 9. This message is broadcast to all mobile phones. As a matter of fact, TESLA in the proposed protocol is used for network authentication not for authentication of data transmission.

A BSC generates a TESLA message and sends it to all BTSs which it supports, as shown in Figure 9. In this message, j is the TESLA time interval, and tK_{j-d} and tK_j are the TESLA keys for the dth previous and current intervals, respectively. As mentioned previously, the TELSA message is sent by the control multiframe burst and this burst is sent to different cells asynchronously. Hence, we should select d so that the disclosed key is not acceptable in all cells. Considering cells in the same location area, it is possible for two BTS to have the delay for sending the same TESLA message in their control multiframe burst. Therefore, we should not disclose TESLA keys until the next control multiframe interval. Thus, to eliminate this security flaw, we select a delay disclosure time equal to two.

In ordinary use, TESLA message has at least 4 parameters as follows: $\{j \| M_j \| MAC_{K_j}(M_j) \| K_{j-d}\}$ [14]. But our TESLA message has only two parameters (Figure 9). In fact, the number of available bits in an idle time slot is restricted to 114 bits and we must use this space effectively. So instead of using separate

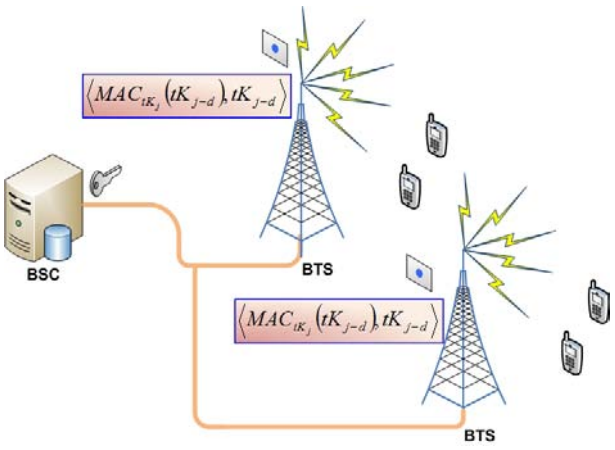

Figure 9. Broadcast messages for network authentication (Part 1)

parameter for j we used the frame number which is broadcast to every MS on the SCH channel. Also the disclosed TESLA key K_{j-d} could also be used as a nonce number. In fact, as shown in Figure 9, we use tK_{j-d} for two purposes; one as a nonce and the other as a former TESLA key interval. Through this mechanism, it suffices to send only two parameters in a TESLA message. Due to the 114-bit restriction, each parameter can only be 57 bits. For cryptographic purposes, a 57 bit key length is too short to provide sufficient security. On the other hand, the required lifetime for a TESLA key is less than one second (235.4 ms), and an adversary cannot accomplish an exhaustive key search in less than a second. However, as a more secure alternative, one can deploy two consecutive idle slots to send both a TESLA key and a MAC. In this case, the TESLA key and its answer can be 114 bits long, while the TESLA duration increases to 470.8 ms. In the remainder of the paper, we use a key length of 57 bits for illustration purposes, but it is easy to expand this to 114 bits.

Considering that the TESLA key in each interval is utilized as a random number in the next interval, we pad the disclosed key with a 7 bit HMAC [18] pattern, which is referred to as R_j. As a result, we have $R_j = tK_{j-d} \| \text{"0110110"}$ in the jth time interval. In order to calculate the MAC, the A8 algorithm is applied as follows: $MAC_{tK_j}(tK_{j-d}) = A8(tK_j, tK_{j-d})$. In view of the fact that algorithm A8 requires two 128 bit inputs, the inputs are padded with the HMAC pattern. Only 57 bits of the 64 bit output are used. Also, BSC generates TESLA key chain via A8 algorithms. In order to generate each TESLA key in the chain, the former generated key will be repeated to reach 128 bit input and the other input is a repeated of pattern "0110110" and then A8 algorithm is applied to them. Since the length of A8 algorithm output is 64 bits, only 57 bits of the output is used as a new TESLA key. The generation of TESLA key chain can be shown as follows:

$tK_{j-1} = A8(tK_j \| tK_j \| tK_j, \text{"0110110"} \| ... \text{"0110110"} \| \text{"01"})$

As mentioned in Section 4.1, to verify a TESLA message, one of the TESLA keys must be authenticated by an MS node. Then each MS will be able to verify the chain of the disclosed TESLA key using the authenticated key. In fact, a new MS to the service area has to authenticate at least one member of the TESLA key chain. For this purpose, the proposed solution takes advantage of the first MS authentication protocol described in Part 2 below.

Finally, it should be noted that the proposed network authentication protocol does not create excessive signaling during the MS authentication in the GSM network. Actually, using the TESLA protocol allows every MS located in the service area to continuously authenticate the network via the available broadcast channel regardless of the MS authentication.

4.2.2 Part 2: The First MS Authentication

The main task of this part is MS authentication in the new visiting VLR. For initial MS authentication, since the VLR does not access the pre-shared key K_i, the structure called "the authenticated temporary key" is applied both to authenticate the subscriber and to decrease the number of control messages. For this reason, the key K_i is not directly used to generate $SRES$ and K_c, but taking advantage of the A5 algorithm the "temporary key" ($TKey$) is formed. $TKey_i$ is computed through algorithm A5 by both the HLR and MS using K_i and timestamp T_v as follows:$Tkey_i = A5(k_i, T_v)$. T_v is a 128 bit timestamp generated by the MS and sent to the HLR by means of an authentication request. The secret key is used as a temporary key between the MS and VLR for MS authentication. As the VLR cannot produce $TKey_i$, the HLR generates and delivers it to the VLR securely. For MS authentication, MS must generate a proper $SRES$. The MS uses $TKey_i$ and $RAND^c$ as inputs through A3 to compute $SRES$. $RAND^c$ is generated by the MS using the current $RAND$ number which broadcast in a TESLA message (R_j) and a counter ($CountU$) in the MS. We use a 64 bit counter which counts during mobile activation such as MS registration, call origination, call termination and each activation which needs MS authentication. This counter is saved in the SIM, HLR and current VLR and is used for user authentication. Actually, the idea of using a counter is taken from the UMTS standard, which is one of the third-generation (3G) cell phone technologies [19]. Every MS in the current service area has the same random number R_j so $CountU$ is suitable to counter a replay

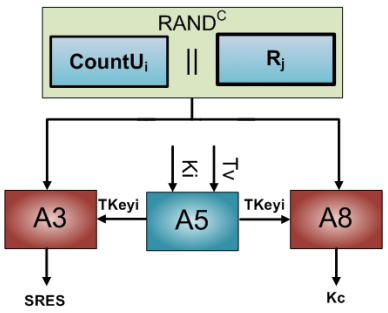

Figure 10. Generation of K_c and $SRES$

attack. In a GSM network, a session cipher key (K_c) is used for data confidentiality which is produced by the HLR during MS authentication. Thus, we have to generate this session key in the VLR and MS without employing the HLR. The MS uses $TKey_i$ and $RAND^c$ again as the inputs to A8 to compute K_c. K_c can be generated by the VLR directly. The generation of K_c and $SRES$ is shown in Figure 10.

Now we describe MS authentication in the new service area. The authentication process is as follows:

- Step 1: The MS generates the following parameters: timestamp T_v, $TKey_i$, $RAND^c$, $SERS$, and K_c. Then authentication is requested from the VLR. In the authentication request message, TMSI, LAI and $SRES$ are sent to the BSC.
- Step 2: After receiving the authentication request, the new VLR uses the received TMSI to get the IMSI from the old VLR and then retrieve the latest sent R_j (created from tK_{j-d}). IMSI is sent along with its identification ID_v, T_v and R_j to the HLR through a secure channel.
- Step 3: After receiving the information from the VLR, the HLR verifies whether ID_v is legitimate or not. Then $CountU$ is retrieved and $TKey_i$ is computed from the received parameters. Later the HLR computes $SRES'$ and can subsequently verify the validity of the MS by comparing the received $SRES$ with the computed $SRES'$. If they are identical, the MS is legitimate, otherwise HLR will terminate the authentication protocol. Then the HLR computes the network authentication parameter for the MS as follows: $Net\_Auth = A8(k_i, R_j \bigoplus T \bigoplus T_v)$. T is the $TKey_i$ life time produced by the HLR. The latest disclosed TESLA key affects the Net_Auth parameter, R_j, so the MS can verify that the disclosed TESLA key is legitimate.
- Step 4: The HLR sends K_c, $CountU_i$, Net_Auth, T and $TKey_i$ to the VLR. Receiving this message from the HLR means that the MS has been authenticated and the temporary key can be used for future authentication. Then the VLR sends Net_Auth and T to the MS and increments $CountU_i$.
- Step 5: When receiving information from the VLR, the MS first checks whether Net_Auth is valid or not. If it is valid, the MS saves TK_{TS_n} (the TESLA key for interval TS_n) as an authenticated seed so that later TESLA keys can be verified. Then the MS increment $CountU_i$. Figure 11 illustrates MS authentication in the new service area.

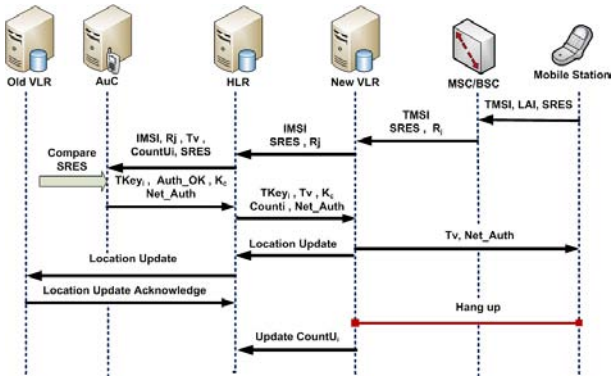

Figure 11. MS authentication in the new service area

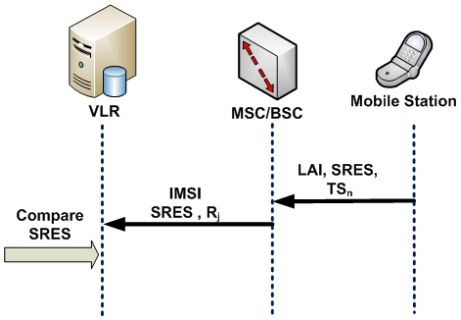

Figure 12. MS authentication message in a location

4.2.3 Part 3: The jth MS authentication in the visiting VLR

For the jth communication, $j > 1$, the VLR can authenticate the MS directly. MS authentication in this situation can be done much more efficiently using our protocol as compared to GSM authentication and other protocols. As shown in section 2, MS authentication in the same visiting VLR has been done with 3 messages. Transmission of these messages leads to wasted bandwidth and increased authentication delay. The MS uses the RACC channel which is a global channel in the cell for sending control messages. Usage of this channel must be very efficient. In our protocol, by sending one message, the MS can be authenticated to the VLR.

The MS computes $SRES_j$ using the last received R_j and $CountU_i$ as follows: $SRES_j = A3(TKey, RAND^c)$. The MS sends $SRES$ as the

authentication request and the VLR can verify the MS as shown in Figure 12.

One drawback of GSM networks is loss of integrity service. An attacker can modify GSM commands and fake transmitted messages. For example, an attacker can change cipher message command issued by the MS to a message that states the MS is unable to support secure communication. By faking this message, the MS and BTS send their data as plaintext so an attacker can easily obtain them. With our protocol, a GSM network can be modified to avoid this problem at a very low cost. For this modification, the BTS must compute the integrity of the control message sent to a specific MS or every MS in the cell by a TESLA key, and the MS must compute the integrity of sending control messages by his $TKey_i$ key.

5 Analysis of the Proposed Authentication Protocol

The authentication protocol is the main concern of the security architecture in a GSM network. This protocol must be not only secure but also efficient. In this section we first consider the performance of our protocol and then present a security analysis.

5.1 GSM Performance Analysis

The authentication process may be accomplished during subscriber registration in a new location area (LA) (and its elimination from the old one), during a conversation (in which the subscriber is either the origin or the destination), or when changing system parameters. In these situations, the authentication process leads to the exchange of a large number of control messages. To evaluate the performance of the proposed authentication protocol, we must determine the resources needed for execution. In our analysis, we consider the required bandwidth and delay for mutual authentication. We use a framework which was proposed in [20] and compare the performance of our protocol against the performance of GSM and other approaches. Stach in [20] introduced a fluid flow model for MS mobility and used the network model proposed in [21]. Our model is based on these results [20]. The average mobile speed is v and the direction of movement is uniformly distributed over the interval $[0, 2\pi]$. The network subscribers are distributed in the network uniformly with density ρ , and each Registration Area has perimeter equal to L. The entire GSM network contains 128 VLR or 128 Registration Areas, each with an area equal to 57.4 square kilometers. The boundary length of each zone is 30.3 kilometers, the mean rate for subscriber contact is 1.4 contacts per hour. The mean density,ρ , is equal to 390 subscribers per square kilometer, and their mean velocity, v, is 5.6 kilometers per hour. The total number of network subscribers is 2.865×10^6 [21].

The computations in [20] were done for different call origination rates such as 1.4, 2.8 and 5.6 calls per hour per handset. However, in our analysis we compute network parameters only for an origination rate of 1.4. It is assumed that call origination and call termination are symmetric with respect to the number of messages. Using these assumptions, the rate of crossing registrations is equal to $R = \frac{\rho \times L \times v}{\pi}$.

As mentioned in section 2, when a mobile enters a new service area, the registration protocol is executed. Hence the rate of crossing registrations in the new service area is equal to the MS registrations rate. Therefore, we can compute the registration rate as $R_{Reg.,LA} = \frac{390 \times 30.3 \times 5.6}{3600\pi} = 5.85(s^{-1})$.

In order to achieve network equilibrium, the rate of MS egress from the service area must equal the MS ingress to another service area, and should be equal to $R_{DeReg.,LA} = 5.85(s^{-1})$. Thus, the total number of messages received at the HLR for registration (number of authentication requests) is $R_{Reg.,HLR} = R_{Reg.,VLR} \times 128 = 5.85 \times 128 = 749.8(s^{-1})$

Considering the call origination rate of 1.4 per hour, and the number of users in the network, the call origination rate is
$R_{CallOrig.} = 1.4 \times 2.865 \times 10^6/3600 = 1114.2(s^{-1})$

However, in a GSM network the VLR gets five triplets from the HLR, which can be used for five authentications in the best case where all triplets are used in the service area. In this situation, the VLR will be sent MS authentication once for every five requests, so the call origination rate for the HLR is $R_{CallOrig._{HLR}} = 1114.2/5 = 222.8(s^{-1})$.

It should be noted that the call origination rate in the HLR is more than 222.8 because of handset mobility, but we assume all triplets are used in the service area for simplicity. Similarly, the conversation establishment rate (where the subscriber is the destination) is obtained as above. As a result, the conversation establishment rate for each LA per second is
$R_{CallOrig./VLR} = R_{CallTerm./VLR} = 1114.2/128 = 8.7$

First, the GSM protocol efficiency is analyzed according to the given framework. The authentication request rate for each VLR and HLR and for various operations is shown in Table 2.

The number of control messages needed for authenticating operations when the subscriber enters a new area is calculated based on Figure 2. The same process can be deployed for the case when the subscriber

Table 2. Authentication request rates for a VLR and HLR

	$VLR(persec.)$	$HLR(persec.)$
Registration	5.85	748.9
Call origination	8.7	222.8
Call termination	8.7	222.8
Sum	23.25	1194.5

Table 3. The number of control messages exchanged in a GSM network

	HLR	VLR
Registration	4[16]	4[16]
Call origination	0.8	3.4
Call termination	0.8	3.4

Table 4. Control messages rate for long term presence in an area

	$VLR(persec.)$	$HLR(persec.)$
Registration	23.4	2995.6
Call origination	29.58	891.36
Call termination	29.58	891.36
Sum	82.56	4778.32

is in the same area, whether it is contacted or initiates the contact, except when no message has been sent to the old VLR according to the "location update". The number of messages for each network parameter is shown in Table 3.

As we have shown, for the registration operations of a new subscriber in an area, 5.85 authentication requests reach the VLR each second, and each time 4 messages must be processed. As a result, the number of control messages needed for registration per second for each VLR is $4 \times 5.85 = 23.4$. Also, authentication of call origination directly occupies the VLR directly until all triplets are used. This occurs once for every five authentication requests, and the HLR processes 4 messages to generate a new set of triplets, so the number of messages per authentication is 0.8. The total number of messages which the HLR must process is then $1114.2 \times 0.8 = 891.36$. The GSM message rates are given in Table 4. The message rates can similarly be calculated for other networks and types of authentication requests.

Table 5. The number of control messages when the proposed protocol is deployed

	HLR	VLR
Registration	4	4
Call origination	0	1
Call termination	0	1

Table 6. Control message rate for the proposed protocol

	$VLR(persec.)$	$HLR(persec.)$
Registration	23.4	2995.6
Call origination	8.7	0
Call termination	8.7	0
Sum	40.8	2995.6

5.2 The Proposed Protocol Performance Analysis

Now we analyze the efficiency of the proposed protocol, which as was shown, requires fewer control messages in comparison with the GSM protocol. Only one control message is needed to authenticate a subscriber provided that the subscriber stays in the LA for a long time (Figure 11). While it is not easy to increase n in the GSM protocol in order to increase the efficiency, with the proposed protocol, only one authentication temporary key and one counter $CountU_i$ are required in the VLR for each subscriber. The number of control messages exchanged for network registration is shown in Table 4. It is assumed that the keys are valid for a sufficient time (equal to the time a subscriber stays in an area). In this case, it is not necessary to contact the HLR or AuC during establishment. In fact, the VLR can authenticate the subscriber as often as necessary. Note that the long term validity assumption will not cause any security problem. Therefore, the control message numbers and rates for the VRL and HLR with the proposed protocol corresponding to Table 1 are given in Table 5 and 6.

5.3 Delay Comparison

The authentication delay in the GSM network consists of the time that the subscriber begins the authentication process until the network decides to either accept or reject the subscriber. In fact, the exchange of authentication messages between the subscriber and the BTS results in a transmission delay. This delay is called TRF (note that both BTS and VLR delays are ignored). In the core network, the time delay due to the message exchange between various databases is called TDB (Figure 13) Taking these definitions into consideration, the authentication delay in a GSM

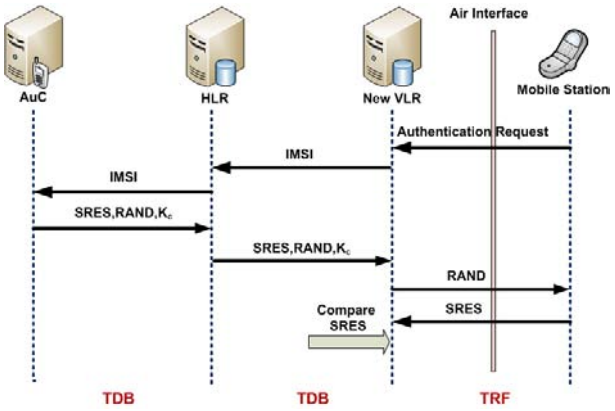

Figure 13. The delay of control messages related to the authentication in the GSM network

Table 8. Comparison between control messages with both protocols

	GSM	Proposed Protocol	Improvement Percentage
VLR	82.56	40.8	50.6%
HLR	4778.32	2995.6	37.3%
Sum	4860.88	3036.4	37.6%

network equals $Auth.Delay = 4 \times TDB + 3 \times TRF$. Then $Auth.Delay = 0.8 \times TDB + 3 \times TRF$ provided that the subscriber stays in an LA for a long time. In essence, the authentication delay with the proposed protocol is obtained upon Figure 11 as $Auth.Delay = TRF$.

Note that the proposed protocol creates at least a three time reduction in the delay resulting from the authentication process. Of course, the higher the delays arising from both the radio band and the connection between various data bases, the higher would be the efficiency of the proposed protocol.

5.4 Comparison of the Proposed Protocol and Related Results

As shown in Section 2, many authentication protocols for GSM have been proposed. In this section, we analyze the performance of the proposed protocol and compare it with protocols suggested in [10,4,5]. Table 7 compares these protocols based on service delivery and bandwidth usage. The results in this table were computed for the MS using the following parameters (in bits): Key Length = 128, $SRES = 32$, timestamp = 32, LAI = 16, RAND = 128, TMSI = 32. We also assume the TDB and TRB delays are identical and equal to one hop. Table 8 shows the improvement in network control message rate of the network (messages per second). It is clear that the proposed protocol has significantly reduced the control message rate.

To analyze the memory consumption with each method, we have to compare memory usage in the MS, VLR and BSC. We suppose each VLR manages N_i MSs, so the memory usage in the VLR for the original GSM authentication protocol is about $5 \times (TMSI + K_C + SRES) \times N_i$. The corresponding value for the protocol in [5] is $(TMSI + K_T + T_{j-1}) \times N_i$, and for the proposed protocol is $(TMSI + CountU_i + TKey_i) \times N_i$. The memory usage in the MS for original GSM is $(TMSI + K_c)$, for the protocol in [5] is $(TMSI + K_T + T_{j-1})$ and for the proposed protocol is $(TMSI + TKey_i + 2 \times MAC + CountU_i + tK_{j-d})$. Moreover, the proposed protocol uses BSC memory to store the TESLA keys chain. If we suppose each BSC generates a chain of TESLA keys for use within a day, 345600 keys must be stored, or 2.4 Mbytes of memory. Table 7 shows the memory usage with the five protocols. It is assumed that each VLR handles 500,000 MSs. In this table, the corresponding values for the protocols in [10] and [4] are also given.

5.5 Security Analysis

The proposed solution for mutual authentication in GSM networks, in fact, includes two protocols, one for MS authentication and the other for network authentication. The MS authentication protocol, in turn, includes two versions, one for MS authentication in a new service area, and the other for MS authentication for subsequent connections in the same service area.

The first MS authentication protocol initiates in the same manner as the original GSM protocol, except that the first message includes the response to an implicit network challenge. This challenge is the latest disclosed key that was received by the MS via the TESLA-based network authentication protocol. So, the network receives the MS response along with the corresponding challenge sent by the BSC in the first round of the protocol. As a result, the network can verify the MS authentication and how live it is. Since the broadcast TESLA message refreshes every 235.4 ms, a replay attack is practically impossible. The other protocols achieve this in the second round of the protocol. Since this MS authentication is now ended, the first MS authenticated must obtain the network response to the MS. The HLR response has two purposes. The first is to establish a common temporary key between the VLR and the MS for subsequent authentication without on-line mediation by the HLR. The second is to authenticate the latest TESLA key (R_j) to synchronize new MSs with the TESLA key chain broadcast in the corresponding service area. In subsequent MS authentications, the MS can directly authenticate the latest disclosed key via the chain of hash functions. The MS can also employ the common session key to communicate directly with

Table 7. Comparison between the proposed protocol and existing GSM authentication protocols

	GSM	[10]	[5]	[4]	Our Protocol
Capacity of transferring information during Ms authentication (bits)	224	224	272	576	208
Authentication Delay (hops)	3.8	3	2	3	1
Asymmetric Operation in MS	-	-	-	×	-
Memory Usage in MS (bit)	160	160	288	256	488
Memory Usage in VLR (MByte)	40	10	18	10	11
Memory Usage in BSC (MByte)	-	-	-	-	2.4
Processing in MS	$A3$	$A5$	$A5 + A3$	$2 \times A5$	$A5$
Mutual Authentication in First Connection	-	×	×	✓	✓
Mutual Authentication in Later Connection	-	×	×	✓	✓
Control Channel Modification	-	-	×	-	×
Additional Processing	-	-	-	-	One A8 every 235.4 ms

the VLR. Thus, the MS does not need to receive a response from the HLR.

The security of the network authentication protocol is based on the TESLA protocol, which relies on one-way functions. An essential requirement for the TESLA protocol is time synchronization of the MS nodes in each service area, with the corresponding BSC as a broadcaster. This requirement is crucial only when a MS node joins a new service area. In this situation, based on the proposed protocol, the MS first assumes the latest disclosed key broadcast via the signaling channel is authentic. However, the MS can verify this assumption by running the first MS authentication protocol when establishing any contact. In summary, although the MS authentication protocol and network authentication protocol are seemingly independent, they subtly co-operate with each other to concurrently meet the required security criteria.

6 Conclusions

In recent years, GSM has become widespread throughout the world. Due to GSM authentication protocol problems, many improved protocols have been proposed. However, some cannot solve all of the drawbacks, and the remainder require that the GSM architecture be altered. In this paper, we propose mutual entity authentication using the TESLA protocol. The proposed solution not only provides a secure bilateral authentication mechanism, but also decreases the call setup time delay and the required connection bandwidth. To evaluate the performance of the proposed authentication protocol, we replicated the traffic analysis in [20]. The relationships between density, velocity, call origination rate and call termination were shown for each call phase. The analysis shows that the proposed protocol reduces the rate of control messages for GSM authentication. This protocol achieves a 62% improvement compared to the original GSM authentication protocol.

References

[1] M. Rahnema. Overview of the GSM System and Protocol Architecture. *IEEE Communnications Magazine*, pages 92–100, 1993.

[2] B. Mallinder. An Overview of the GSM System. In *Proceedings of the Nordic Seminar on Digital Band Mobile Radio Communnications*, pages 12–15, 1988.

[3] K. Schramm. *DES Sidechannel Collision Attacks on Smartcard Implementations*. M.Sc. thesis, Ruhr-Universit Bochum, 2002.

[4] A. Peinado. Privacy and Authentication Protocol Providing Anonymous Channels in GSM. *Computer Communication*, 27:1709–1715, 2004.

[5] C.C. Chang, J.S. Lee, and Y.F. Chang. Efficient Authentication Protocols of GSM. *Computer Communication*, 28:921–928, 2005.

[6] V. Bocan and V. Cretu. Mitigating Denial of Service Threats in GSM Networks. In *Proceedings of the International Conference on Availability, Reliability and Security (ARES2006)*, pages 523–528, 2006.

[7] K. Al-Tawil, A. Akrami, and H. Youssef. A New Authentication Protocol for GSM Network. In *Proceedings of the IEEE 23rd Annual Conference on Local Computer Networks*, pages 21–30, Boston, 1998.

[8] . Aydemir and A. Aydin Selçuk. A Strong User Authentication Protocol for GSM. In *Proceedings of the IEEE International Workshop on Enabling Technologies*, pages 150–153, 2005.

[9] W.D. Lin and J.-K. Jan. A Wireless-Based Authentication and Anonymous Channels for Large Scale Area. In *Proceedings of the IEEE Symposium on Computers and Communnications (ISCC)*, page 3641, 2001.

[10] C. Lee, M. Hwang, and W. Yang. Extension of Authentication Protocol for GSM. *IEE Proceedings Communications*, 150(2):91–95, 2003.

[11] P. R. Zimmermann. *The Official PGP User's Guide*, volume 1995. MIT Press, 1995.

[12] ITUT. The directory: Authentication framework. Technical report.

[13] D. Brown. Techniques for Privacy and Authentication in Personal Communication Systems. *IEEE Personal Communnications*, pages 6–10, 1995.

[14] A. Perrig, R. Canetti, J.D. Tygar, and D. Song. The TESLA Broadcast Authentication Protocol. *UC Berkeley and IBM Research*, 5(2), 2002.

[15] R.J. Anderson, F. Bergadano, B. Crispo, J.H. Lee, C. Manifavas, and R.M. Needham. A New Family of Authentication Protocols. *Operating Systems Review*, 32(4):9–20, 1998.

[16] M. Bohge and W. Trappe. TESLA Certificates: An Authentication Tool for Networks of Compute-Constrained Devices. In *Proceedings of the ACM Workshop on Security (WiSE'03)*, San Diego, CA, USA, 2003.

[17] M. Schwartz. *Mobile Wireless Communications*. Cambridge University Press, 2005.

[18] W. Stalling. *Cryptography and Network Security*. Prentice Hall, 2003.

[19] C. Blanchard. Security for the Third Generation (3G) Mobile System. *InformationSecurity Technical Report*, 5(3):55–65, 2000.

[20] J.F. Stach, E.K. Parka, and K. Makkib. Performance of an Enhanced GSM Protocol Supporting Non-Repudiation of Service. *Computer Communication*, 22:675–680, 1999.

[21] R. Thomas, H. Gilbert, and G. Mazziotto. Influence of the Mobile Station on the Performance of a Radio Mobile Cellular Network. In *Proceedings of the 3rd Nordic Seminar*, Copenhagen, Denmark, 1988.

Ali Fanian received the BS and MS degrees in computer engineering (Hardware and Computer Systems Architecture) in 1999 and 2001, respectively from Isfahan University of Technology (IUT), Isfahan, Iran. He is currently Ph.D student in IUT. Different aspects of computer architecture and network security are Mr. Fanian research interests; specially, adhoc networks, wireless network security and hardware design.

Mehdi Berenjkoub received the Ph.D. degree from Department of Electrical and Computer Engineering, Isfahan University of Technology in 2000. The title of his dissertation is two-party key distribution protocols in cryptography. He started his work in the same department as an assistant professor from that time. Graduate courses presented by him include Fundamentals of Cryptography, Cryptographic Protocols, Network Security, and Speech Processing. He has supervised more than a dozen M.Sc. students and a Ph.D. candidate in related areas. He also was one of the founder members for Iranian Society of Cryptology in 2001. He has continued his cooperation with the society as an active member. He along with his colleagues recently established a research group on Security in Networks and Systems in IUT. He also is responsible for a newly established academic CSIRT in IUT. His current interested research topics are wireless network security and authentication protocols.

T. Aaron Gulliver received the Ph.D. degree in Electrical Engineering from the University of Victoria, Victoria, BC, Canada in 1989. From 1989 to 1991 he was employed as a Defence Scientist at Defence Research Establishment Ottawa, Ottawa, ON, Canada. He has held academic positions at Carleton University, Ottawa, and the University of Canterbury, Christchurch, New Zealand. He joined the University of Victoria in 1999 and is a Professor in the Department of Electrical and Computer Engineering. In 2002, he became a Fellow of the Engineering Institute of Canada. He is currently an Editor for IEEE Transactions on Wireless Communications. From 2000-2003, he was Secretary and a member of the Board of Governors of the IEEE Information Theory Society. His research interests include information theory and communication theory, algebraic coding theory, MIMO systems and ultra-wideband communications.