CHAPTER 4

Secure Communications in Wireless Networking Environment
Introduction

Section 4.1 has introduced authentication as it relates to the larger domain of computer, communications, and information security in wireless networks, and provided an account of the particular characteristics of the wireless network environment that challenge designers of security systems.

The section 4.2 describes the wireless networking environment is inherently more dynamic, less robust, and more open to intrusion and fraud than are fixed terrestrial infrastructures.

Section 4.3 of the study examines the approaches to authentication taken in three “second-generation” digital cellular networking environments: GSM (Global Systems for Mobile Telecommunications), DICT (Digital Indian (BSNL - Tarang, Reliance etc.) Cordless Telephone).

In Section 4.4, discusses the GSM (Global System for Mobile Communication, or just Global System Mobile) which is the most widespread protocol in the world for digital cellular communications.

In Section 4.5 describes the various authentications in GSM followed by section 4.6 which gives you the basic comparison between GSM and DICT authentication.

In section 4.7 we try to list some Applications of Public-Key Approaches followed by section 4.8 which describes some algorithms for wireless Networks among these are the Modular Square Root (MSR) technique, and some recent variants of Elliptic Curve Cryptography (ECC). These will be briefly described in the following sub-sections. Section 4.9 list some Proposed Enhancements to GSM Security and Authentication Protocols
4.1 Authentication in the Wireless Networking Environment

Webster’s New International Dictionary, 1925 Edition, defines “authentication” to mean: “Act of authenticating, or state of being authenticated; the giving of authority or credit by the necessary formalities; confirmation.” “Authenticate,” the verb, is further defined: “(1) To render authentic; to give authority to, by the proof, attestation, or formalities required by law, or sufficient to entitle to credit; as the document was authenticated by a seal. (2) To prove authentic; to determine as real and true or genuine; as to authenticate a portrait.” Seventy-five years later, in the context of digital computers and communication, these remain serviceable definitions.

4.1.1 Authentication & Security Architecture

In the world of information security, authentication means the act or process of proving that an individual or entity is who, or what, they claim to be. According to Burrows, Abadi, and Needham: “The goal of authentication can be stated rather simply, though informally and imprecisely. After authentication, two principals (peoples, computers, services) should be entitled to believe that they are communicating with each other and not with intruders.” Thus, a corporate IT infrastructure would want to authenticate that the user of a company database system is in fact the director of human resources before granting her privileged access to sensitive employee data (perhaps by means of a user password and a smartcard). Or, the provider of cellular communications system would want authentication of the cellular telephones accessing their wireless system, to establish that the handsets belong to users whose accounts are up to date, and are phones that have not been reported stolen. Authentication is one among a set of services that constitute a security sub-system in a modern computing or communications infrastructure. The particular services that constitute the full set may vary somewhat, depending on the objectives, information content, and level of criticality of the parent system.

William Stallings, in his text Cryptography and Network Security, provides a list of core security services worth quoting at length to place authentication in its proper systems context:

CROSS-REFERENCE

[190x-76]
**Confidentiality:** Ensures that the information in a computer system and transmitted information are accessible only for reading by authorized parties.

**Authentication:** Ensures that the origin of a message or electronic document is correctly identified, with an assurance that the identity is not false.

**Integrity:** Ensures that only authorized parties are able to modify computer system assets and transmitted information.

**Non-repudiation:** Requires that neither the sender nor the receiver of a message be able to deny the transmission.

**Access Control:** Requires that access to information resources may be controlled by or for the target system.

**Availability:** Requires that computer system assets be available to authorized parties when needed.

Stallings’ description suggests that these system security functions apply to human system users. While this is true, it is also important to understand, as indicated by the Burrows, Abadi, and Needham quotation, that the functions can apply as well to physical devices (authentication of a cellular phone), or to a computer system (authentication of a wireless network server).

Authentication in conventional, wireline networks has attracted substantial research work and implementation effort over the last two decades. Among the best-known authentication protocols for distributed computer systems, going back to the 1980s, are Kerberos (developed initially at MIT as part of Project Athena), the Andrew RPC (Remote Procedure Call) handshake, the Needham-Schroeder public-key protocol, and the CCITT X.509 protocol. Detailed discussion of authentication protocols for wireline networking environments is beyond the scope of this thesis.

For a formal analysis of the respective procedures, guarantees, and weaknesses of the four protocols mentioned immediately above, the Burrows, Abai, and Needham article is useful.
4.1.2 Background Concepts in Authentication

While this thesis will avoid detailed exploration of authentication in conventional, non-wireless networks and distributed systems, several general concepts in authentication are important to the discussion in subsequent sections.

These are:

**Authentication Center:** In protocols that involve the use of secret keys for authentication, these keys must be stored by the service provider, together with information about the individual user or subscriber, in a highly secure environment. In the world of cellular telephony in particular, such a system is usually called an Authentication Center.

**Subscriber Authentication:** Much of the discussion concerning authentication in digital cellular networks concerns subscriber authentication. This refers to the authentication of users of cellular telephone services, and would typically take place when a user attempts to place a call, thus registering a request with a network base station for the provision of services. It should be noted that "subscriber authentication" often refers to the authentication of the cellular telephone handset and the information on the smartcard it contains, rather than to the authentication of the actual human user (although the latter authentication, of course, is the ultimate goal).

**Mutual Authentication:** Most authentication protocols involve two "principals," with the possible involvement of trusted third-parties, such as a Certification Authority, depending on the protocol. In mutual authentication, both the principals are authenticated to one another. It is important to note that authentication need not be mutual -- it can be oneway only. In the discussion of authentication in second-generation cellular telephone networks, for instance, we will encounter cases where the network authenticates the cellular phone seeking to use its services, but the network base station is not authenticated to the cell phone.
**Challenge/Response Protocol:** A number of the protocols explored in this study employ challenge/response mechanisms as the basis for authentication. In a challenge/response scenario, the principal seeking to perform authentication on the second principal generates a random number and sends it to the second principal. In many protocols, the random number is simultaneously transmitted to an Authentication Center.

The second principal feeds the random number, together with its secret key, to a commonly agreed upon algorithm. The resulting bit-string, which is determined uniquely by the combination of the random challenge and the secret key of the second principal, is then transmitted back to the first principal. Meanwhile, the Authentication Center – or some similar trusted third-party – which also has access to the secret keys of the principals, performs the same calculation and sends the result back to the first principal.

The first principal compares the two values, and, if they are equal, authenticates the second principal. Note that the challenge/response mechanism does not require that the first principal know the secret key of the second principal, or vice versa.

**Session Key Generation:** Although the generation of a session key is not necessarily part of subscriber authentication in the narrowest sense, it often occurs in the course of the same process, and thus enters the discussion in the following sections. A session key is a digital key that is used in the process of encrypting messages exchanged in a single communication session between principals. It is thus distinct from the private keys or public keys of system users, which typically have a longer lifetime. Communication systems often generate session keys with algorithms that run in parallel with those that resolve challenge/response protocols (see above), and that take the same inputs.

When these terms appear in the subsequent sections of this study, they carry the meanings defined above.
Private-key Versus Public-key Cryptography

Another background concept that underlies the discussion in subsequent sections is the distinction between private-key and public-key cryptography. In brief, with private-key cryptography (also called symmetric key cryptography), two parties who seek to exchange confidential messages share the same “secret key” (usually a random digital bit-string of some mutually agreed upon length).

The keys are “symmetric” in function, in the sense that principal A can use the secret key and an encryption algorithm to produce cipher text (an encoded message) from plain text (the original message). Upon receiving the encoded message, principal B unwinds the process by inputting the same secret key to the algorithm, but this time executing it “in reverse” – in decryption mode. The result of this operation is the original plain text message (“message” here should be taken in a broad sense – it might not be readable text, but the bit-strings in a digitally encoded cellular telephone conversation, or the bytes of a digitally transmitted image file). Common examples of symmetric private-key cryptosystems include DES (Data Encryption Standard), IDEA (International Data Encryption Algorithm), and RC5.

With public-key encryption technology, there is no shared secret key. Each principal who wishes to be able to exchange confidential messages with other principals possesses his or her own secret key, which need not be shared with other principals. In addition, each principal makes publicly available a “public key” (there is no need to conceal this key – in fact, the operation of a public-key cryptographic system requires that other principals be able to easily access this information). Public-key cryptography employs asymmetric encryption algorithms. That is to say, when principal A seeks to send a secure message to principal B, A encrypts the plain text message using B’s public key and the original message as inputs to the algorithm. This requires no special action by B, in that B’s public key is readily available to A. Principal A then transmits the message to Principal B. The public-key encryption algorithm operates in such a fashion that the message
encrypted with B’s public key can only be decrypted with B’s private key. As B has shared this private key with noone, only B can decrypt the message. RSA (named for Ron Rivest, Adi Shamir, and Len Adleman) is perhaps the best-known example of a public-key cryptosystem.

Again, detailed exploration of public-key versus private-key encryption technologies is beyond the scope of this study. The reader is referred to the Stallings text for an excellent, extended discussion. A 1992 paper by Beller, Chang, and Yacobi provides an excellent discussion of the distinction between private-key and public-key systems in the specific context of wireless networks.

In second-generation cellular networks such as GSM (Global Systems Mobile), the use of private-key encryption technology has been universal. A common assumption concerning public-key technologies is that they are too compute-intensive to be practical in wireless networking environments. As we will see in following sections of this study, on “processor-light” public-key encryption algorithms optimized for wireless networks has called this conventional wisdom into question. The ongoing debate about the relative merits of private-key and public-key approaches to security and authentication is a key theme for research related to the operation of wireless networks, and will play itself out in the design of systems for deployment over the coming decade.

4.2 WIRELESS NETWORKING ENVIRONMENT

Wireless networks extend the reach and flexibility of communications and computing in powerful ways. However, the wireless networking environment is inherently more dynamic, less robust, and more open to intrusion and fraud than are fixed terrestrial infrastructures. These factors set the context for security and authentication in wireless networking environments; they pose the challenges that designers of security architectures and systems must overcome.
In an excellent 1994 paper entitled “The Challenges of Mobile Computing” that summarizes the distinctions between wireless and wireline networking environments and the problems wireless networks pose for software engineers, George Forman and John Zahorjan distinguish among factors that derive from “three essential requirements: the use of wireless networking, the ability to change locations, and the need for unencumbered portability.” While Forman and Zahorjan’s analysis is broad—they are examining the impact of wireless networking environments on the whole spectrum of software engineering—the same framework can be used to good advantage in examining the situation as it pertains specifically to security and authentication. The author’s concluding summary is also useful, and as applicable today as it was in 1994:

Wireless communication brings challenging network conditions, making access to remote resources often slow or sometimes temporarily unavailable. Mobility causes greater dynamicism of information.

Portability entails limited resources available on board to handle the mobile computing environment. The challenge for mobile computing designers is how to adapt the system designs that have worked well for traditional computing.

It should be remarked that in the domain of security, the “designs that have worked well for traditional computing” (and communications) are themselves in a state of considerable flux, adding additional uncertainty to the equation.

In the remainder of this section, I will briefly identify the primary challenges of the wireless networking environment for security and authentication protocols, using the three categories proposed by Forman and Zahorjan.
Challenge Area 1: Wireless Network Links

By definition, wireless networks depend on wireless communication links, typically employing radio signals to accomplish information transmission, over at least a significant portion of their infrastructure. The great strength of wireless communications technology, of course, is that it can support ongoing communication with a portable device, such as a cellular phone or a personal digital assistant, that is mobile. On many dimensions, however, the use of wireless links in a network poses problems, in comparison with networks that employ only copper wires, fiber optic cables, or some combination of such fixed infrastructure.

Low Bandwidth: The speed at which wireless networks operate is increasing as technology improves. In general, however, wireless links support data transfer rates several orders of magnitude lower than what is possible over fixed infrastructure. For example, the second-generation cellular phone networks discussed in this study transmit data on a channel at approximately Kbits/second. This will increase to somewhat more than 350Kbits/second in soon to be introduced “third-generation” cellular networks. Currently, wireless LAN systems employing the 802.11b standard can achieve speeds of up to 11 Mbits/second. It should be noted, however, that this rate is for the network as a whole, not for a communications channel to a single machine, and only operates in a small area, such as a floor of a building. In the wired world, Fast Ethernets, operating at 100 Mbits/sec are becoming the norm in building networks, while long-haul Internet backbone circuits operate at multiple Gigabits/sec.

Frequent Data Loss: In comparison with wireline networks, digital data is frequently lost or corrupted when transmitted over a wireless link. Networking protocols that employ mechanisms for data integrity checking can identify these conditions and call for information to be retransmitted, but the impact is to compound the effect of low bandwidth. In addition to slowing the rate at which information is correctly transmitted, data loss can increase the variability of the time required to transmit a given body of data, or to conclude a transaction.

“Openness” of the Air Waves: Fixed-line networks, whether formed from copper or fiber optic cable, can be tapped. This tends to be a technically
challenging procedure, however, and the intrusion can often be detected by network monitoring equipment. By contrast, when a wireless network sends data through the atmosphere using radio signals, anyone can listen, using even inexpensive equipment. Further, such intrusions are inherently passive and difficult to detect. This circumstance poses a fundamental security risk for wireless networks. As we will see in subsequent sections, designers of second generation cellular systems have addressed the most obvious risk, posed when one simply transmits conversations or sensitive data over the wireless link “in the clear,” by using encryption techniques. Derivative exposures are pervasive, however, and have not been thoroughly addressed.

**Challenge Area 2: User Mobility**

As noted, the great advance of wireless networking technology is that the user is free to move about while still maintaining a link to the network. This feature of wireless networking, however, weakens or eliminates some of the basic presuppositions that help insure security in wired networks. For example, in the typical wired network in an office, a user’s desktop computer will be attached to the same port on the same network hub (or equivalent piece of networking equipment) day after day. Further, the set of computers, printers, and other networked devices that is attached to the network at any point in time is known to the system administrator and under his or her control.

In a wireless networking environment, these basic presuppositions no longer apply. Users, not system administrators, determine which network “port,” and even which network, they connect to with their portable devices. Similarly, the set of devices attached to a wireless network at any point in time will depend on the movements and actions of individual users, and is out of the control of the network operator.

**Disconnection and Reconnection:** Users of wireless communication networks are at constant risk of being suddenly, unceremoniously disconnected from the network. This can occur for many reasons: because
the user moves with his or her portable device out of the coverage range of the base station with which they are communicating; because the user’s movement causes a physical obstacle – such as a building or a vehicular tunnel – to intrude itself between the portable device and the base station; or just because of the inherent low reliability of radio links. Also, in the normal course of operation of cellular communication networks, as the user moves from the area of coverage of one base station to another, the network must transfer control of the communications session with a “hand-off,” causing network delay and potential disconnection.

**Heterogeneous Network Connections:** In a typical wired network scenario, a computer is constantly attached to the same home network. The characteristics of this network are a known quantity, while change – say a system upgrade to the file server or firewall – can be carefully planned and monitored. In a wireless networking scenario, however, a mobile station such as a cell phone or PDA is constantly roaming among different host networks.

The characteristics of these networks, and the way that they expect to interact with the user’s home network, can vary considerably.

**Address Migration:** In the conventional wired network, computers and other devices are attached to the same network and retain the same network address (IP address in the Internet world) for long periods of time. If devices are moved between networks, the network administrator can act to update the network address. In a wireless networking environment, network addresses – or at least the networks to which they refer – must be managed in a much more complexity and security risks.

**Location Dependent Information:** The situation with regard to location information is parallel to that in the case of address migration. In the wired network, the locations of computing devices are relatively static and known to the system administrator. In a wireless environment, the locations of computing and communications devices change frequently. Not only must the wireless networking infrastructure track and respond to these location changes in order to provide service to users, it must also provide security
provisions that protect location information. In the wireless environment, protecting user confidentiality includes of course protecting the content of messages and conversations against intrusion, but also requires that the system hold private the location of system users.

**Challenge Area 3: Device Portability**

In order to exploit the potential of wireless networks, users require computing and communication devices they can carry easily. A mobile computing and communications infrastructure is not much use if one has to carry a desktop machine to exploit it. Thus, common electronic products today such as cellular phones, PDAs, laptop computers, digital cameras with networking capabilities, and the like, are designed to be carried by people on the move. As Forman and Zahorjan put it: "Today’s desktop computers are not intended to be carried, so their design is liberal in their use of space, power, cabling, and heat dissipation. In contrast, the design of a hand-held mobile computer should strive for the properties of a wristwatch: small, light weight, durable, water-resistant and long battery life."

One security-related implication of device portability is obvious: any product which is designed to be carried and used on the move is easily stolen. Not only is a cellular phone, say, a simple target for a thief; from the perspective of the system, there is nothing suspicious about the fact that the device is moving from one town to another, though it may now be in the possession of someone other than the owner.

The portable form factor also imposes other limitations on designers of mobile communications and computing products that have implications for security and authentication. These include:

**Processor Speed:** The processing power delivered by the integrated circuits used in such devices as cell phones and PDAs, while increasing over time, does not approach that of desktop machines or network servers. Encryption algorithms and authentication routines require computation, and some
require a great deal of computation. In some security applications in the wireless environment, such as the encryption and decryption of a conversation conducted via cellular phone, the security procedures must execute in near realtime. Thus, the processing power available on a mobile device constrains the choices of designers of security systems for wireless environments.

**Limited Storage Capacity:** For similar reasons, the amount of data that can be stored in a portable communications or computing device is less than the data-storage capacity of a desktop machine or server. While less critical than processor speed, this factor also influences the choices made in designing security systems for wireless networks.

**Low-Power Operation:** Portable electronic products operate on batteries. Any work performed by the processor in a cellular phone or PDA draws power and thus reduces battery life. From the standpoint of a product user, while security may be an essential feature, its implementation represents overhead. Therefore, even if it is possible to execute a processing-intensive security or authentication routine from the technical standpoint, the drain on battery power may be unacceptable.

As can be seen from this list, the challenges confronting designers of security architectures and systems for wireless networks are substantial, and they differ in both kind and degree from the situation found in the conventional wireline network. In effect, these factors explain why a consideration of security in wireless environments differs from an equivalent examination of the wireline network. A trend to consider, though, is that wireless Internet access is growing more pervasive, and that many corporate and home networks are incorporating a wireless component. For this reason, the factors outlined in this section will come in the future to have increased influence on the design of security systems which are not intended for purely wireless environments.

4.3 Authentication in 2nd Generation Digital Cellular Networks

The term "second-generation" cellular networks, often applied to the systems discussed in this section – GSM and DICT.
Prior wireless communications technologies consisted of analog cellular communications systems, so in fact the so-called second-generation systems represented the first generation of viable digital systems for mobile communications.

In most analog cellular systems, subscriber's voices were transmitted "in the clear," making eavesdropping on conversations on the wireless link a major security risk. The second-generation systems addressed this issue through data encryption, which represented a significant advance in security for wireless communications environments. Another security breach that has cost wireless communications service providers with first-generation analog systems hundreds of millions of dollars is "cell-phone cloning." Here, the authentication procedures developed for second-generation digital systems have had a major impact as a deterrent. Nonetheless, as time has passed since their introduction, the security protocols of second-generation systems have come under heightened examination.

Reasons include their exposure to attackers masquerading as legitimate network nodes, and their assumption that the wired network infrastructure can be considered secure.

4.4 SUBSCRIBER AUTHENTICATION IN GSM

GSM (Global System for Mobile Communication, or just Global System Mobile) is the most widespread protocol in the world for digital cellular communications. Developed and first deployed in the European Union, GSM cellular communication systems are now also found in Asia, Africa, and North America. Developed from the ground up as an architecture for digital voice communications, GSM supports a complete set of security protocols that avoid many of the exposures to eavesdropping and subscriber fraud that plagued first generation analog cellular technologies. However, in recent years, GSM has come under criticism from security experts, because, with increasing exposure, certain flaws and security risks in the security architecture have emerged. In addition to the factors noted in the previous section, experts object to the proprietary authentication and encryption algorithms employed by GSM, as these have not been subjected to neutral analysis.
4.4.1 System Components of the GSM Security Architecture

There are four primary sub-components of the overall GSM authentication architecture: the SIM (Subscriber Information Module), the GSM handset, the HLR (Home Location Register)/AuC (Authentication Center), and the VLR (Visitor Location Register).

1. **SIM (Subscriber Information Module):** The SIM is a small smart card provided by the GSM service provider to the individual subscriber. The SIM slips into a GSM cellular phone and carries several elements of data and code: the IMSI (International Mobile Subscriber Identity – a number which is globally unique to the GSM subscriber), an authentication key (usually designated Ki) which is specific to the subscriber, a Personal Identification Number (PIN), a algorithm called A3 used for subscriber authentication, and an algorithm called A8 used for the generation of session keys.

2. **GSM handset:** Within the GSM phone itself is embedded code to implement the A5 algorithm, which handles the encryption and decryption of the information sent between handset and base station during an actual GSM communication session.

3. **HLR/AuC (Home Location Register and Authentication Center):** The HLR and AuC are generally integrated together on the network of the subscriber's GSM service provider, but can be thought of as logically separate entities. The Authentication Center has as a key component a database of identification and authentication information for each subscriber. Attributes in this database include the subscriber's IMSI, authentication key (Ki), LAI (Location Area Identifier), and TMSI (Temporary Mobile Subscriber Identity – a code which, when used in place of the IMSI, can hide the true identity of the subscriber from eavesdroppers). The Authentication Center is responsible for generating triplets of values consisting of the RAND, SRES, and session key (Kc) which are stored in the Home Location Register for each subscriber (the function of the elements of the triplet will be explained in the next section).
4. VLR (Visitor Location Register): The VLR, which, like the HLR, is maintained on the system of each GSM service provider, stores sets of triplets (RAND, SRES, and session key) for each subscriber who is communicating with the base stations of the service provider. When the subscriber roams away from the network of his or her own service provider, of course, information must be forwarded from the HLR to the VLR, in order to complete the authentication process.

A schematic illustrating the way in which these elements of the GSM security architecture are inter-related appears in Figure 4.1.

![Figure 4.1: Schematic illustrating the inter-relationships of the primary components of the GSM security architecture. [Source: Gundersen and Dohmen]](image)

Note that the VLR and HLR/AuC may be under the administrative control of different network service providers. However, the SIM and MS (mobile station) of the subscriber are in effect directly bound to just one HLR/AuC. Thus, the subscriber has a permanent home (unless he or she switches service providers) but may roam geographically among locales served by different GSM providers without interruption of service.

4.4.2 Data Elements in the GSM Authentication Protocol

GSM security protocols, including that for user authentication, are based on symmetric cryptographic technologies, with the SIM (Security Information Module) and Authentication Center both provided with the same IMSI (International Mobile Subscriber Identity) and subscriber authentication key (Ki) for each GSM subscriber. A cornerstone of the GSM security protocols
is that a subscriber's authentication key, while stored in both the SIM and the

Authentication Center, is never transmitted over the network. The key data elements of the GSM authentication protocol appear below, with those data elements which comprise the GSM “triplet” listed first. Recall that the elements of the triplet are generated by the Authentication Center, initially stored in the Home Location Register (HLR), and then forwarded to the Visitor Location Register (VLR) when a subscriber seeks to establish a communication session while roaming.

Elements of the Triplet

RAND: RAND is a 128-bit random number generated by the Authentication Center. It is used during the initial challenge-response phase of the GSM authentication sequence (See the following section for details).

SRES (Signed Response): SRES is a 32-bit number that results when the GSM A3 algorithm is applied to a 128-bit RAND.

Kc (Session Key): Kc is a 64-bit session key, or “ciphering key,” used to encrypt and decrypt data transmitted between the handset and base-station during a single GSM communication session. Kc is generated on the handset side within the SIM by feeding the 128-bit RAND and the subscriber’s unique identification key Ki to the A8 algorithm. Kc is thus unique to both the individual subscriber (given that Ki is used in its generation) and to this particular communication session (given the use of RAND as the second seed element).

Other Essential Data Elements

Ki (Subscriber Authentication Key): Already described, the subscriber authentication key, which is unique to the individual subscriber, is a symmetric key held in both the SIM and the Authentication Center, but never broadcast over the airwaves.

IMSI (International Mobile Subscriber Identification): An identifier that is unique to the individual subscriber.
TMSI (Temporary Mobile Subscriber Identification): A temporary identifier used in GSM communication sessions in place of the IMSI in order to preserve subscriber confidentiality.

Note that subscriber authentication key is protected physically in that it is stored only on the presumably secure Authentication Center server, and embedded in the SIM smartcard.

4.4.3 Operation of the GSM Authentication Protocol

Given the GSM system components and security data elements outlined in the previous two subsections, we can now proceed to outline how subscriber authentication is actually accomplished in the GSM environment. In essence, GSM makes use of a simple challenge-response protocol that takes advantage of the RAND and SRES pre-computed by the Authentication Center and provided to the Visitors Location Register (VLR). The sequence proceeds as follows:

1. When the mobile station (GSM handset) approaches a base-station in a visited network with a request to establish a communications session, it makes known its IMSI (by transmitting it once, in the clear).

2. The visited network submits its own identify, together with the IMSI of the subscriber to the Authentication Center, which responds with a set of RAND, SRES, and Kc (the triplet). The Authentication Center, using the A3 algorithm, has generated the SRES from the random number RAND using the subscriber authentication key Ki. In addition, the AuC has calculated the session key Kc in the same fashion, using the A8 algorithm. Note that while the visited network now knows both the SRES and the session key Kc for this particular communication session, it has no insight into the subscriber's subscriber authentication key.
3. The visited network sends the 128-bit RAND to the mobile station as a challenge.

4. The SIM in the mobile station, having received the RAND from the GSM handset, uses the A3 embedded implementation of the A3 algorithm and the Ki to generate its own SRES. The handset then transmits this SRES back to the visited network (the response to the challenge in step 3).

5. The visited GSM network compares the SRES it has received as a response from the mobile station with the SRES it received earlier from the Authentication Center and stored in the Visitor Location Register. If the two SRES values are identical, the communication session is authorized; otherwise, it is rejected.

6. If the session is authorized, the SIM in the handset also calculates its own version of the session key Kc, using RAND and the subscriber authentication key Ki as inputs to the A8 algorithm. Note that, while the session key has not been transmitted from the base station to the mobile station, both the visited network and the mobile station now possess the same session key.

7. In order to support secure communications within the subsequent exchange between the visited network and the mobile station, both mobile station (handset) and GSM network feed the session key Kc and the TDMA (Time Division Multiple Access – GSM is built on a TDMA communications protocol) round number to the A5 algorithm. The result is a 14-bit sequence that is XOR-ed with the two 57-bit data blocks transmitted in a single TDMA round. Note that, while Kc remains constant through a GSM session, the 114-bit sequence will change in every round due to the changing TDMA round number.

A schematic depicting the flow of data elements in the GSM subscriber authentication sequence, adopted from Gundersen and Dohmen, appears in Figure 4.2.
Figure 4.2: Schematic illustrating the flow of information in the GSM subscriber authentication sequence. [Source: Gundersen and Dohmen]

A schematic illustrating the operation of the A3 and A8 algorithms appears in Figure 4.3
Figure 4.3: Inputs and outputs from the A3 and A8 algorithms in the GSM subscriber authentication sequence. [Follows Gundersen and Dohmen]

Note that, within the GSM handset, the A3 and A8 algorithms execute on the SIM smartcard, meaning that the subscriber authentication key Ki does not leave the SIM.

4.4.4 Evaluation of GSM Authentication and Security

The GSM infrastructure for subscriber authentication and confidentiality of communication sessions represented a major advance over first-generation analog cellular systems. However, as GSM has matured and expanded its reach across Europe and beyond, its basic security mechanisms have come under increasing criticism. Given the strong belief in the security community that only protocols that can be tested should be trusted (that security should depend on the secrecy of keys and not of algorithms), it was inevitable that GSM would be attacked for its dependency on the proprietary A3, A8, and A5 algorithms. Many security analysts view these algorithms as
cryptographically weak, and subject to intrusion by government agencies, in addition to well-equipped hackers.

Among the particular criticisms of the GSM security regime are:

1. Both A3 and A8, the algorithms used respectively for user authentication and the generation of session keys, are typically implemented by GSM service providers with an algorithm called COMP128. COMP128 was reverse engineered at Berkeley in 1998, and cryptanalysis by Berkeley researchers indicates that the protocol can be broken with 219 queries from a rogue base station to the GSM SIM card, which can be achieved in eight hours. Analysis of the GSM application of COMP128 further revealed that it had apparently been deliberately weakened. The algorithm calls for a 64-bit key, but of this total ten key bits are consistently set to zero, dramatically reducing the security offered by the A8 implementation.

2. If the session key Kc is compromised, an intruder can impersonate a legitimate GSM Visitor Location Register (VLR) without having to periodically authenticate itself. Also of concern is the storage of the RAND, SRES, and Kc triplets in the VLR while they are awaiting use, increasing exposure, particularly to insider intrusion.

3. Under the GSM authentication protocol, the GSM base station authenticates the Mobile Station (GSM handset) which seeks to establish a communications session. However, the opposite is not true: thus, the Mobile Station has no guarantee that it is not communicating with a node which is impersonating a GSM base station. To make the situation worse, the same random challenge that is used to authenticate a mobile station, when presented to the A8 algorithm, also becomes the seed for the generation of a session key.

Furthermore, the authentication challenge-response message protocol does not include a time-stamp. Thus, if a rogue station does successfully impersonate a GSM base station, it may secure a session key that will allow the decryption of any messages sent with the same key over a potentially prolonged period.
4. GSM authentication (and security in general) protects the wireless link between the Mobile Station (MS) and the GSM base station that is serving it. The security mechanisms do not protect the transmission of information between the Authentication Center on the user's home network and the serving network. This lack of security in the wired network represents a major exposure for GSM, particularly in light of the fact that communications between GSM base stations and the true wired network are often transmitted over microwave links that make interception easy.

5. Of the two variants of the A5 data encryption algorithm, the weaker, called A5/2, can be exported anywhere in the world without restriction. According to information published by Bruce Schneier, A5/2 was developed with assistance from the NSA, and can be broken in real time with a work factor of approximately 216. A5/1, the stronger of the two variants, is susceptible to attacks that can break it with a work factor of 240, meaning that near real-time compromise is possible if the attacker uses specialized hardware.

4.5 SUBSCRIBER AUTHENTICATION IN DICT

DICT [Digital Indian Cordless Telephone] Protocol, as the name implies, is a standard for cordless telephone-based communications, basically it a European standard but the same is followed in India, like GSM, the basic security mechanisms in DICT are based on a symmetric key architecture in which the subscriber and the home network share a private key.

Subscriber authentication derives from this shared private key, as does the generation of session keys for the encryption of transmitted data. In contrast to GSM, DICT introduces several alternative approaches to how the home network and the serving network interact in the subscriber authentication process.
4.5.1 Data Elements and Algorithms in the DICT Authentication Protocol

The data elements in the DICT authentication protocol include the fundamental private key shared by the subscriber’s handset and the home network, session keys, and several varieties of random numbers used in the authentication and data encryption processes. Key data elements are:

**DICT ID (DICT identification number):** The unique identification number of the individual DICT handset.

**K (Shared Authentication Key):** This is the secret, symmetric key shared by the home network and the subscriber’s handset. Depending on the requirements of the system operator, DICT provides for the generation of K from other keys associated with the subscriber, such as the UPI and UAK (see below).

**UPI (User Personal Identity):** The UPI is a short, personal code that a subscriber can enter into a DICT handset at the beginning of a communication session. The UPI is thus a kind of PIN code. As noted above, as an option, DICT provides for the generation of the K authentication key from a UPI.

**UAK (User Authentication Key):** The UAK is a longer key, unique to a DICT user, which can be embedded within a DICT handset by means of a smartcard-like module called the DICT Authentication Module (DAM).

**RS (Random number):** RS is a random number generated by the subscriber’s home network as part of the authentication sequence.

**Ks (Session Key):** Ks is a key generated from the combination of K and RS in order to provide a master key for a particular DICT communication session.

**RAND_F (Random Number):** A random number distinct from RS.

**DCK (DICT Dynamic Cipher Key):** A key produced from KS and RAND_F for the purpose of providing a data encryption key for use in a DICT communication session.
RES (Authentication Response Code): A value produced by the DICT mobile station from the combination of KS and RAND_F to authenticate the DICT subscriber to the network.

XRES (Authentication Confirmation Code): A value produced by the DICT fixed station from the combination of KS and RAND_F to confirm authentication of the DICT subscriber.

SCK (Static Cipher Key): SCK provides an alternative to the use of the session-specific DCK and KS values. SCK is a static key which is embedded in both the Portable Radio Termination (the handset) and the Fixed Radio Termination (the DICT network module) in order to support data encryption/decryption on an ongoing basis. A single SCK value can be used for an indefinite period.

The DICT authentication protocol involves two algorithms, A11 and A12.

A11: The A11 algorithm takes secret key K and the random number RS as inputs, and produces session key KS.

A12: The A12 algorithm takes KS and random number RAND_F as inputs and produces RES (or XRES on the network side) and the optional session specific encryption key DCK.

4.5.2 Operation of the DICT Authentication Protocol

In the case where the DICT handset or mobile station (Portable Radio Termination in DICT terminology) is interacting directly with its home network (the so-called Fixed Radio Termination), the subscriber authentication protocol closely resembles the challenge/response sequence in GSM. The sequence runs as follows:

1. The mobile station sends its DICT identification number to the fixed station when it seeks to establish a communication session.

2. The fixed station generates two random numbers, RS and RAND_F. Using these values, the private key K corresponding to the mobile station’s ID, and the A11 and A12 algorithms, the fixed station can then generate XRES, KS, and DCK.
3. The fixed station transmits the random number pair (RS, RAND_F) back to the mobile station.

4. The mobile station can now also use A11 and A12 to generate its response to the challenge, RES, and the session keys. The mobile station transmits RES back to the fixed station.

5. The fixed station compares RES and XRES. If they match, the mobile station is authenticated; otherwise, the mobile station's request is rejected.

See Figure 4.4 for a schematic illustration of the operation of the subscriber authentication exchange in DICT.

Figure 4.4: Schematic diagram illustrating the operation of the subscriber authentication message exchange in DICT.

When the DICT subscriber is roaming away from his or her home network, as in the case of GSM, information must be passed from the HLR (Home Location Register) to the VLR (Visitor Location Register). In contrast to GSM, however, DICT provides for three alternative interactions in the
course of the subscriber authentication process, which may be chosen depending on the security standards and requirements of the service provider.

In the first variant, the HLR simply transmits a copy of the secret key $K$ to the VLR, with the visited network then generating its own random numbers and using them in conjunction with $K$ to carry forward the authentication protocol. This minimizes the communication between HLR and VLR, thus achieving efficiency, but it increases the risk that $K$ will be compromised. This variant should only be used if security requirements are low to begin with, and/or (2) the fixed network is known to be extremely secure. In the second variant, the HLR itself generates random numbers $RS$ and $RAND_F$, and then, using $K$, generates $XRES$ and, optionally, the session encryption key $DCK$. These elements are then forwarded to the VLR. This is the alternative most closely comparable to GSM. In the third variant, random number $RS$ and the session key $KS$ are transmitted from HLR to VLR, with the derivation of $XRES$ then taking place at the visited network.

4.6 Comparison of DICT and GSM Authentication

As noted, the GSM and DICT subscriber authentication protocols are similar in that they both depend upon symmetric, private-key techniques and a challenge-response exchange between the mobile node and the network providing service. There are, however, significant differences between the two implementations:

As illustrated in the account in the previous section of the three alternatives for communication of information between HLR and VLR in the subscriber authentication sequence, DICT provides a service provider with more flexibility in the implementation of security protocols than does GSM. In addition to supporting authentication of the subscriber by the network, DICT supports a protocol whereby the subscriber can authenticate the network with which it is communicating. This reduces the possibility that an attacker will successfully impersonate a legitimate base station.
4.7 Applications of Public-Key Approaches

During the 1980s, when the security protocols for GSM were being developed, the most telling criticism of public-key cryptography as far as wireless networks were concerned was that the protocols required too much processing.

RSA, for instance, is commonly estimated to require 1000 times as much computation as common private-key encryption technologies. Given the limitations of cellular telephone handsets, in terms of both processor speed and battery life, designers of cellular networks viewed this as too high a price to pay.

4.8 ALGORITHMS FOR WIRELESS NETWORKS

Beginning in the early 1990s, researchers identified alternative algorithms, requiring less processing power to implement, which can be applied to authentication and security in wireless networking environments. Among these are the Modular Square Root (MSR) technique, and some recent variants of Elliptic Curve Cryptography (ECC). These will be briefly described in the following sub-sections.

4.8.1 The Modular Square Root Algorithm

The Modular Square Root (MSR) algorithm was introduced by M.O. Rabin in 1979, and then researched for its potential in personal communications systems by Beller, Chang, and Yacobi in the early 1990s.

Like most cryptographic algorithms, the approach is based on modular arithmetic and depends on the difficulty of factoring large numbers for its encryption strength.

In brief, MSR operates as follows. The public key is a modulus, N, which is the product of two large prime numbers, p and q (where, in practical implementations, p and q would typically be binary numbers of 75 to 100 bits in length). The combination of p and q constitutes the private key component of the algorithm. If principal A wants to send a confidential
message $M$ to principal $B$, A first calculates $C = M^2 \mod N$, where $C$ is the derived cipher text and $M^2$ is the binary value of the message $M$, squared. Note that this is a modular operation, so the remainder $\mod N$ is the value taken. Upon receipt of the cipher text $C$, principal $B$, who knows $p$ and $q$, is able to reverse the process by extracting the "modular square root" of $C$ to extract $M$ (that is to say $M = \sqrt{C} \mod N$). For a party who does not have access to the values of $p$ and $q$, achieving a solution is precluded because of the difficulty of factoring $N$ – there is no known polynomial-time algorithm. In addition to the fact that it supports a public-key/private-key encryption and message transfer regime, MSR has a second strong advantage when it comes to the wireless environment. The computational load of the algorithm is asymmetric. Calculating the modular square that is required for encryption requires much less computation (just one modular multiplication) than does extracting the modular square root to return the plain text (this requires exponentiation). Thus, if the encrypting function can be placed on the mobile station, and the decrypting function on the base station, MSR ideally matches the constraints imposed by mobile handsets that have slow processors and limited battery reserves.

4.8.2 Elliptic Curve Cryptography

In recent years, Elliptic Curve Cryptography (ECC) has also emerged as a candidate encryption technique for applications in wireless networks. Given the emphasis on minimizing the requirement for processor resources for encryption within the mobile station, "encryption strength per key bit" becomes an important metric. It is generally accepted that encryption with ECC using 160-bit keys offers approximately the same level of security as RSA with 1024-bit keys, and at least one study has indicated that ECC with even a 139-bit key provides this level of security.

Koduri, Mahajan, Montague, and Moseley have proposed an authentication approach that combines short individual passwords with ECC-based cryptography. The authors employ two variants of the basic ECC approach, EC-EKE (Elliptic Curve Encrypted Key Exchange) and SPECKE (Simple Password Elliptic Curve Key Exchange). Both variants require that the communicating principals agree upon a shared password, the mathematical
definition of a specific elliptic curve, and a point on this curve, prior to establishing a communication session (although not explored in the paper, an Authorization Center could provide the necessary information to the principals as the authorization exchange unfolded). In a trial implementation of an authentication procedure for wireless environments using ECDSA (Elliptic Curve Digital Signature Algorithm), Aydos, Yanik, and Koc used an 80MHz Advanced RISC Machines ARM7TDMI as the target processor (the ARM7TDMI is used in numerous applications in portable products designed to communicate via wireless networks). Using an ECC key length of 160 bits, ECDSA signature generation required 46.4 milliseconds, versus 92.4 milliseconds for signature verification. With a key length of 256 bits, these times went to 153.5 milliseconds for signature generation and 313.4 milliseconds for verification. The authors conclude that the ECC-based ECDSA approach to subscriber verification is a practical alternative for wireless environments.

4.8.3 Data Elements in the Improved MSR Protocol

In the Improved Modular Square Root (IMSR) Protocol, both the Serving Network Base Station (SNBS) and the Certification Authority (CA) hold public keys of the type described in the discussion of the MSR algorithm, representing the product of two large primes p and q, which constitute the private keys. Each network base station holds a certificate, obtained from the Certification Authority, which applies a hashing function h to the network ID of the network base station and to its public key. Beller, Chang, and Yacobi use the term “Radio Control Equipment” (RCE) to identify the functional entity that controls communication ports on a wireless network. As we have used “base station” to identify this function in other sections of this study, the same term will be used here for consistency. (Beller, Chang, and Yacobi’s terminology has also been adjusted in several other particulars to maintain consistency across sections).

The key data elements and functions in the IMSR Protocol include the following:
1. **ID$_{BS}$ (Base Station Identifier):** The unique identifier of a wireless Network base station (in this scenario, a base station in the serving, or visited, network).

2. **ID$_{MS}$ (Mobile Station Identifier):** The unique identifier of the mobile station. This corresponds to the IMSI in the GSM authentication protocol.

3. **N$_{BS}$ (Public Key of Base Station):** NBS, the public key of the base station, is the product of two large primes, p$_{BS}$ and q$_{BS}$, known only to the network base station and to the Certification Authority (CA).

4. **N$_{CA}$ (Public Key of CA):** NCA, the public key of the Certification Authority, is similarly the product of two large primes, p$_{CA}$ and q$_{CA}$, known only to the Certification Authority.

5. **K$_s$ (Session Key):** A session key for subsequent data encryption during the communication session, negotiated in the course of the authentication protocol.

6. **RANDX (Random Number):** A random number chosen by the mobile station in the course of determining K$_s$.

7. **h (Hash Function):** h is a one-way hash function, known to all principals, that reduces its arguments to the size of the modulus (i.e., the same length as NBS and NCA).

8. **g (One-way function):** g is a one-way function, known to all principals, that expands its arguments to the size of the modulus.

9. **Cert$_{BS}$ (Certificate for the Base Station):** The certificate CertBS takes the form SQRT(h(ID$_{BS}$, NBS)) mod NCA. Note the use of the hash function h.

10. **Cert$_{MS}$ (Certificate for the Mobile Station):** The certificate CertMS takes the form SQRT(g(ID$_{MS}$) mod NCA. This certificate is embedded in the subscriber’s handset when he or she registers for service, and employs the one-way function g.

11. **m (Mobile Station’s Certificate String):** The string m is generated by concatenating the mobile station’s identifier ID$_{MS}$ and the mobile station’s certificate CertMS.
4.8.4 Operation of the Improved MSR Protocol

Key aspects of the IMSR Protocol for authentication include the use of the certificates and the MSR algorithm. The primary steps in the protocol proceed in this sequence:

1. The mobile station signals the base station on the serving network that it wishes to open a communications session.

2. The base station holds a certificate of the form $\sqrt{h(ID_{BS}, NB_s) \mod N_{ca}}$. The network base station sends a copy of this certificate to the requesting mobile station.

3. The mobile station checks the validity of the certificate by squaring the certificate value, mod $N_{ca}$, and comparing it with the value of $h(ID_{BS}, NB_s)$, which it calculates independently. If the values match, the mobile station proceeds; otherwise, it aborts the communication session.

4. The mobile station chooses a random number, labeled RANDX, which is to serve as the session key $K_s$. The mobile station then calculates a value, call it $a$, where $a = RANDX^2 \mod N_{BS}$. The mobile station then sends $a$ to the base station.

5. The network server calculates the value of RANDX (which is in fact the session key $K_s$) by calculating $RANDX = \sqrt{a} \mod N_{BS}$. Note that this calculation would be impossible for an eavesdropper to perform, because the eavesdropper would not have access to the base station’s prime factors $p$ and $q$. Both mobile station and base station now share the same session key $K_s$.

6. The mobile station now employs the agreed upon session key $K_s$, function $f$, and a string $m$ to calculate a value, call it $b$, where $b = f(K_s, m)$. String $m$, as noted above, concatenates IDMS and CertMS. The mobile station transmits $b$ to the network base station.

7. The mobile station uses its knowledge of the session key $K_s$ to decrypt $b$ and extract $m$. From string $m$, the base station extracts the certificate for the mobile station CertMS, and calculates $Cert_{MS} \mod N_{CA}$. This value it compares to $g(ID_{MS}) \mod N_{CA}$, which it can also calculate. If a match results, the mobile station is in fact the unit it claims to be, and the session
key is confirmed. The operation of the IMSR Protocol is depicted schematically in Figure 4.5. Note that, while the figure depicts only the interaction between the mobile station and the network base station, a certification authority is also a critical piece of the infrastructure. Given the structure of the IMSR Protocol, however, the CA is involved when the base station is established and when a subscriber registers for service, but not at the time of the individual session. This produces the benefit of reducing the need for long-haul communications from the serving network to the home network in the course of establishing a communication session.

Figure 4.5: Schematic diagram illustrating the operation of the Improved Modular Square Root (IMSR) Protocol. [Source: Beller, Chang, Yacobi]
4.9 Proposed Enhancements to GSM Security and Authentication Protocols

Since the introduction of GSM (Global System for Mobile Communication) in Europe, the technology has spread to become a pervasive infrastructure for wireless communications. As noted in the previous section, GSM security mechanisms represent a significant advance on those of first generation analog systems. Nevertheless, they have come under increasing criticism as time has passed, as the security community has gained more familiarity with the workings (and deficiencies) of GSM security, and as real world attacks have taken place on the infrastructure.

These developments have encouraged researchers to investigate potential enhancements to the security protocols of GSM, including its authentication mechanisms. In a broad sense, the proposed enhancements can be classified in two groups: (1) proposals that seek to enhance the level of security and confidentiality provided by GSM systems, and (2) proposals that seek to improve the performance of the security mechanisms of GSM. This section will examine several examples of these investigations.

4.9.1 PROPOSED ENHANCEMENTS TO THE SECURITY LEVEL OF GSM

Among the weaknesses of the GSM authentication protocols pointed out in previous sections are

1. while the GSM network providing service authenticates the mobile station (MS), the mobile station does not reciprocally authenticate the network.

2. The home network must trust the visited network to serve only legitimate users, and cannot distinguish between fraudulent calls which might be initiated, for example, by dishonest insiders in the visited network from legitimate calls placed by its own subscribers.

The pervasive assumption in GSM is thus, that while the service provider must verify the identity and credentials of the individual subscriber before
privileges are extended, the network infrastructure itself is to be trusted without authentication. Most analysts in the security community consider this assumption naive.

4.9.1 ADDING ASSURANCES TO GSM SECURITY

In a 1993 paper presented at the IEEE Global Telecommunications Conference, Hung-Yu Lin and Lein Harn propose authentication mechanisms that would provide substantial improvements in the level of security offered by the protocol. The key enhancements to standard GSM authentication proposed by Lin and Harn are:

1. The mobile station authenticates the base station, as well as the GSM network authenticating the mobile station;
2. As an option, users employ a personal code (PIN) to initiate service;
3. A time stamp is used to counter replay attacks;
4. Clock synchronization between mobile station and the network is more sophisticated than in standard GSM;
5. Different temporary identifiers are used for each individual call; and
6. The GSM data encryption function (A5) is supplemented with a one-way hash function.

The basic subscriber authentication mechanism proposed by Lin and Harn proceeds as follows (I will emphasize the aspects in which this sequence differs from GSM authentication).

1. Each base station continually broadcasts the current time, t, to support time synchronization with mobile stations.
2. When a mobile station wants to initiate a communication session it sends its unique identifier (UID), the service domain identifier (SDID) of its home network, and the concatenated value of the time stamp, t, and a random number R1 selected by the mobile station. The sequence <t, R1> is encrypted by applying an encryption function, f, using the mobile station's secret key, k.
As in standard GSM, the secret key is known only to the mobile station and to the home service domain (HSD). It is not transmitted over the network.

3. The mobile station also proceeds to calculate a session key, k1, for this communication session. k1 is generated by applying the hash function, h, to the string <k, R1>.

4. The visited service domain (VSD) forwards the data elements to the subscriber’s home service domain (HSD), using the SDID to determine where the HSD is located.

5. The HSD, which knows the secret key k for the subscriber, decrypts t and r1. It examines the value of the time stamp t to insure that this request to establish a communication session is a fresh one and not a replay (standard GSM does not provide this check).

6. The HSD calculates k1, by applying the hash function to k and r1. The HSD then sends r1 and k1 back to the VSD, together with an account number AN and an expiration code EXPIRE for the communication session (EXPIRE can either represent a span of time or a limit on the number of calls that can be placed during the current communication session). The mobile station and VSD now share the same session key k1.

7. The VSD generates a temporary identifier, I1, and, using the encryption function f with the session key k1, sends <I1, R1> back to the mobile station.

8. The mobile station confirms the value of R1 that it has received back from the VSD (Note that when the mobile station first sent R1 to the VSD, it was encrypted with the mobile station’s secret key, and thus was not accessible to the VSD).

9. The communication can now proceed with the use of the session key k1 for data encryption.

If, while in communication with the same VSD, the subscriber wants to place another call, it is not necessary to contact the HSD again. Rather, the mobile station sends I1 and R2 to the VSD, where R2 is a new random number that the mobile station has generated. Note that when the VSD originally sent I1 to the mobile station, the value was encrypted, so by now sending the correct value back to the VSD, the mobile station has in effect
authenticated itself. The VSD then uses the hash function, $h$, with the prior session key $k_1$ and $R_2$, to generate a new session key, $k_2$ (the mobile station can, of course, generate the same session key $k_2$ on its side). The VSD sends back the string $<I_2, R_2>$, encrypted with the encryption function $f$ in the presence of the new session key $k_2$. Now, the mobile station can initiate the next call using session key $k_2$. This approach can be employed for further calls until the limit imposed by EXPIRE is encountered.

See Figure 4.6 for a schematic definition of the basic Lin-Harn authentication protocol. This protocol differs from standard GSM authentication in several regards. It is the mobile station, for instance, rather than the network, which generates random numbers. By requiring that the visited service domain transmit a random number in encrypted form to the HSD, and then forward the corresponding decrypted value back to the mobile station, there is more provision for authentication of the network to the mobile station here than is present in standard GSM. Further, the use of a time-stamp in an encrypted string guards against replay attacks by eavesdroppers. However, weaknesses remain. The original communication from the mobile station to the VSD, for instance, includes the subscriber’s unique identifier in plain-text. This exposure grows in significance if the subscriber is frequently roaming between service domains. In addition, the VSD, having once received a session key from the HSD, is in a position to place fraudulent calls that originate from others than the real subscriber.
Figure 4.6: Schematic illustration of the exchange of messages in the Lin-Harn authentication protocol.

Lin and Harn propose additional enhancements to their protocol designed to address these shortcomings. These include:

**User ID Confidentiality:** Lin and Harn propose the addition of a public-key cryptographic protocol, possibly RSA, for the limited role of concealing the subscriber's user identification (UID) when it is first submitted to the network. Under this scheme, each service domain would choose two large, secret prime numbers, p and q. Subscribers would then receive N, the product of p x q, and a public key e, when they register for service. When submitting its initial request for service to the VSD, the mobile station would use the public-key encryption algorithm with e to encrypt its UID, as well as the results of the application of the private-key encryption process to <t, R1>. With this addition, it is only possible for the home service domain to decrypt and read the subscriber's UID.

**Controlled Random-Number Generation:** Lin and Harn also propose the use of a second one-way hash function, l, which would support the generation of the values R1, R2, R3 in a sequence that is uniquely
determined by the seed value $R_1$ and the secret key $k$. That is to say, $R_i = l(k, R_{i-1})$. Now, both the mobile station and the home service domain know what the next random value should be, and consequently what the next session key $k_i$ should be. This eliminates the prospect that a rogue VSD will sponsor illegitimate calls.

**Simplified Roaming Support:** With the implementation of the controlled random number generation advocated above, it also becomes easier to support hand-offs between service domains as the subscriber roams. Because the expected session-key sequence can now be monitored by the mobile station and the HSD, in a roaming scenario the old service domain can pass the current temporary ID and current session key along to the new service domain.

One major merit Lin and Ham claim for this set of enhanced protocols is that they can be implemented through changes in the interaction between the mobile station and the home service domain. The behavior of the VSD does not need to change. Thus, different home service domains can offer different sets of security services to their subscribers without requiring that the entire infrastructure be upgraded.