Mobility and Security Management in the GSM System and Some Proposed Future Improvements

ASHA MEHROTRA, MEMBER, IEEE, AND LEONARD S. GOLDING, LIFE FELLOW, IEEE

Important aspects of mobility and security in the Global System for Mobile communications (GSM) system are discussed in this paper. Mobility management functions are broadly categorized into three groups: a) mobile turned on, b) mobile turned off, and c) mobile in conversation. The paper first outlines the mobile synchronization sequence followed by its mobility functions: mobile identification, authentication, international mobile station identity attach/detach, and its location update. The important role of security in the GSM system is fully explored, including authentication, encryption, and positive identification of mobile equipment before the user is provided with the service. The future of mobility management, with respect to subscriber identification module roaming, intersystem roaming, advancement in mobile service, and its impact on data base requirements, is covered in subsequent paragraphs.

***Keywords*—** *Encryption, GSM signaling protocols, intersystem roaming, mobile temporary ID, mobility management, security management, SIM roaming, subscriber, subscriber identity module, TMSI attach and detach, white, black, and gray equipment list.*

I. INTRODUCTION

Mobility management (MM) entails the Global System for Mobile communications (GSM) system's keeping track of the mobile while it is on the move. Basically, we have two different situations: mobile idle and mobile busy. These two cases lead to all the relevant cases we need to consider: a) mobile station (MS) is turned off, b) MS is turned on but is in the idle state, and c) MS is in the conversational mode.

a) *MS turned off:* In this case, the mobile cannot be reached by the network because it does not answer a paging message. It does not inform the system about possible changes of location area (LA), as it is simply inoperative. In this case, the MS is simply considered detached from the system [international mobile station identity (IMSI) detached].

Manuscript received June 13, 1997.

The authors are with Hughes Network Systems, Germantown, MD 20876 USA (e-mail: Amehrotra@hns.com; Lgolding@hns.com).

Publisher Item Identifier S 0018-9219(98)04245-5.

b) *MS turned on, idle state:* Here, the system can page the MS successfully. This is the situation in which the MS is considered attached (IMSI attached). While on the move, the MS has to check that it is always connected to the best received broadcast control channel (BCCH). This procedure is called roaming. While on the move, the mobile must also inform the system about changes of LA, which is called location updating.

c) *MS busy:* The radio network has traffic channels allocated for the data flow to/from the MS. While moving, the MS must also be able to change to a new traffic channel as the signal on the traffic channel drops below an unacceptable level, which is called handover. In order to decide whether to hand over, the mobile switching center (MSC) base station controller (BSC in some cases) interprets information received from the MS and base transceiver station (BTS) known as locating.

In view of the above, we shall discuss the complete mobility aspect of the system, which revolves around these three states.

Another important aspect of the GSM system is security. At an early stage in the development of the Pan-European mobile radio system GSM, it was apparent that the weakest part of the system was the radio path, as this could be easily eavesdropped upon with radio equipment. There was also a need to authenticate users of the system so that the resources were not misused by nonsubscribers. It is easy to see that the public land mobile network (PLMN) needs a higher level of protection than traditional telecommunications networks. Therefore, to protect the system against the two cases mentioned above, the GSM system has been reinforced by the following four security techniques [1]–[3]:

- anonymity: mobile temporary identification (ID) (TMSI);
- authentication;

0018–9219/98$10.00 © 1998 IEEE

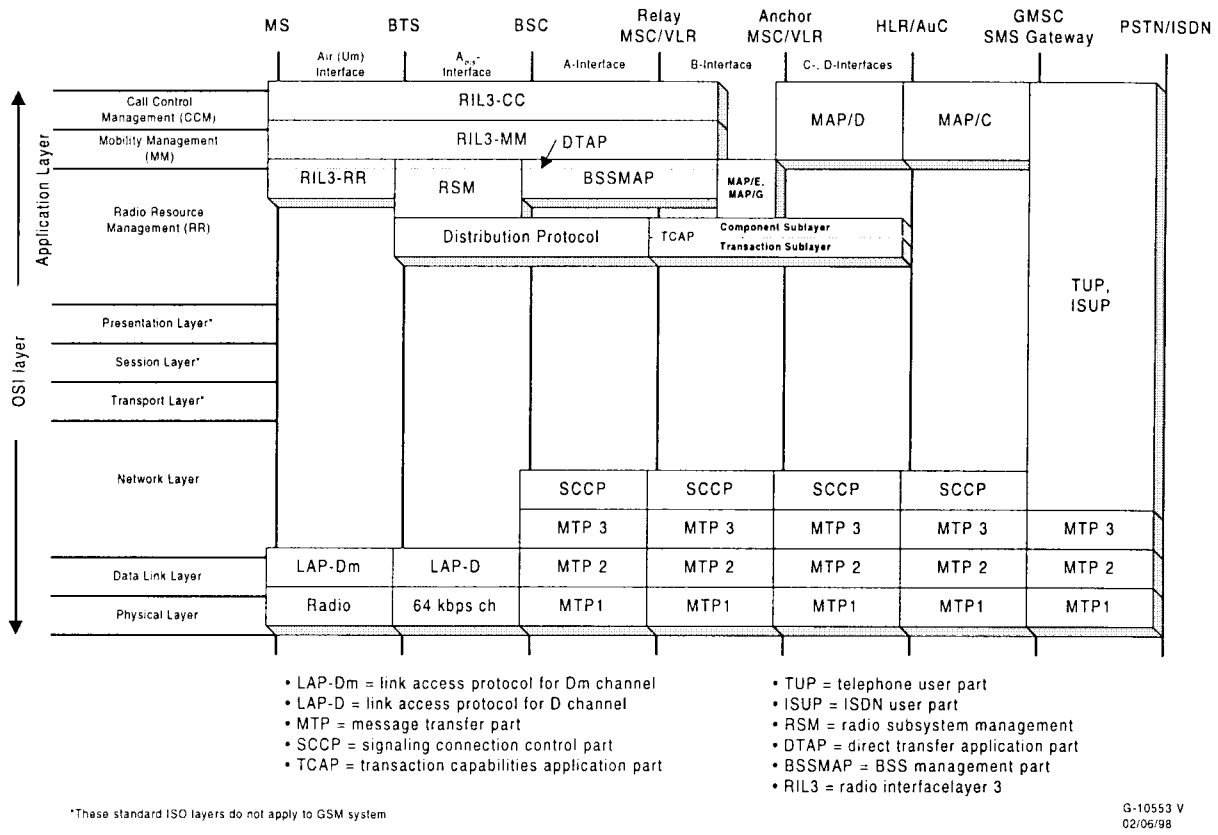


Fig. 1. GSM signaling protocols [3, p. 266].

- encryption;
- user's subscriber identity module (SIM) module and mobile equipment ID.

Among these, we shall elaborate on three important techniques: authentication, encryption, and mobile equipment ID.

To make the discussion complete, the last section of the paper will deal with future improvements one should pursue in these areas. The following are candidates for discussion: improvement in SIM roaming, intersystem roaming, future service, and data base requirements. Before we discuss the mobility aspect of the user, let us briefly introduce the GSM protocols and how the mobile synchronizes within the GSM system. Both factors are intimately tied to the mobility aspect of GSM users.

II. GSM Signaling Protocols

GSM uses the International Consultative Committee on Telegraphy and Telephony Signaling System 7 (SS7) protocols for its signaling [1], [4]–[5]. Fig. 1 shows key elements of a GSM system as vertical bars, with the spaces between them representing interfaces among different subsystems. The protocol is based on open system interconnection (OSI) layers, which are shown at the left.

At the physical layer, the air interface uses radio frequency (RF) radio transmission. The A_{bis}-Interface between BTS and BSC uses 64 kb/s channels on wire cable, optical fiber, or microwave links based on the best availability, designer's choice, and ease of installation. All other interfaces use the SS7 message transfer part (MTP) level 1 with channel rates of 64 kb/s, where the medium can be coaxial cable, balanced wire pairs, or fiber-optic cable.

The data link layer uses integrated services digital network (ISDN) link access protocol D (LAP-D) on the A_{bis}-Interface and a version called LAP-D_m on radio links to MS's. LAP-D_m uses the GSM RF interface framing and synchronization scheme rather than that of ISDN. All other interfaces use the SS7 MTP level 2 protocol for the data link layer. The air and A_{bis}-Interfaces have no network layer. All other interfaces within GSM use both MTP level 3 and the SS7 signaling connection control part (SCCP). MTP 3 provides routing of general connectionless messages to the various network nodes, while SCCP routes connection-oriented-messages, specific to a particular user transaction, all the way to the destination terminals. No GSM interfaces have need for use of the SS7 transport, session, and presentation layers. The rest of the interfaces are at the application layer 3. Interfaces to the fixed networks [public switched telephone network, ISDN] use their own standard telephone user part and ISDN user part protocols. Once a radio resource (RR) connection has been established, there exists a physical point-to-point bidirectional connection between two RR entities. At one instant of time, there will not be more than one RR connection available between a mobile station and the MSC. One RR connection may be used to support several MM connections.

The SS7 application layer has several sublayers specific to the GSM system. The radio interface layer 3 (RIL3)

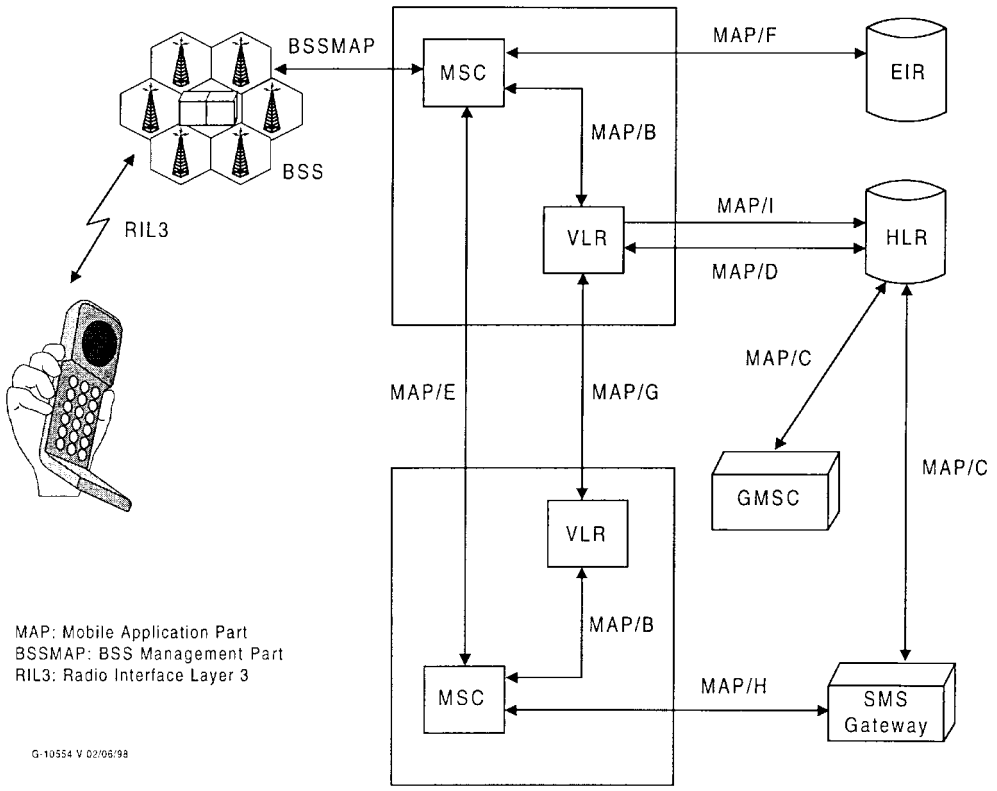


Fig. 2. MAP protocol connections [3, p. 297].

RR management protocols establish and release connections between an MS and various BSC's for the duration of a call despite user movement. The radio subsystem management protocol provides RR functions between the BTS and BSC. The direct transfer application part (DTAP) protocols provide RR messages between the MS and MSC. The base station subsystem management part (BSSMAP) protocols provide RR messages between the BSC and MSC. The distinction between DTAP and BSSMAP is provided by a small "distribution" protocol below them.

The RIL3 MM protocols deal with MS location management (cell assignment, location updating, paging for incoming calls, etc.) and security aspects of the system (authentication, user identity protection, etc.). This protocol is the main interest of this paper. The RIL3 call control management protocols deal with call control (CC), supplementary services, and short messages. All mobility management and call control functions reside in the MSC. Thus, all CC and MM messages are interpreted neither by the BSC nor by the BTS. They are transparently conveyed to its corresponding entity within the MSC. Before exchanging messages between MM entities of the mobile station and the network (MSC), an RR connection needs to be established.

The transaction capabilities application part provides correlation between individual operations and structured exchanges in building up a complete transaction. The transaction sublayer manages transactions on an end-to-end basis. The component sublayer correlates commands and responses within a dialog. Noncall-related signaling between network parts is handled by the various mobile application part (MAP) protocols, details of which are shown in Fig. 2 [2], [5]–[7]. These protocols are designated as MAP/B through MAP/H. For example, signaling between MSC and equipment ID register (EIR) is through MAP/F. Although these messages are not directly related for completing the actual voice or data transmission path, they are essential for the proper establishment of the traffic channel, as we shall see in the mobility discussion.

III. Mobile Initialization

Prior to establishing any communication links to other parties, the MS must first acquire synchronization with the GSM system. This process begins after the MS is turned on in a PLMN. The first step of the process is for the MS to search for and acquire a frequency control channel (FCCH) burst on some common control frequency channel. The mobile will scan all or part of 124 RF channels and obtain the average signal strength of each channel. During the scanning process, several readings of the RF level have to be taken so that the mobile gets an accurate estimate of the channel power. Thus, the scanning may take several seconds.

For each of the 124 channels, starting with the one of highest signal strength level, the mobile searches for the FCCH. This is the first step of the process known as frequency synchronization. The frequency correction burst is unique and easily recognizable. The FCCH burst is a long sine wave that is offset by 67.7 kHz from the carrier frequency. The cell transmits all zeros for the frequency

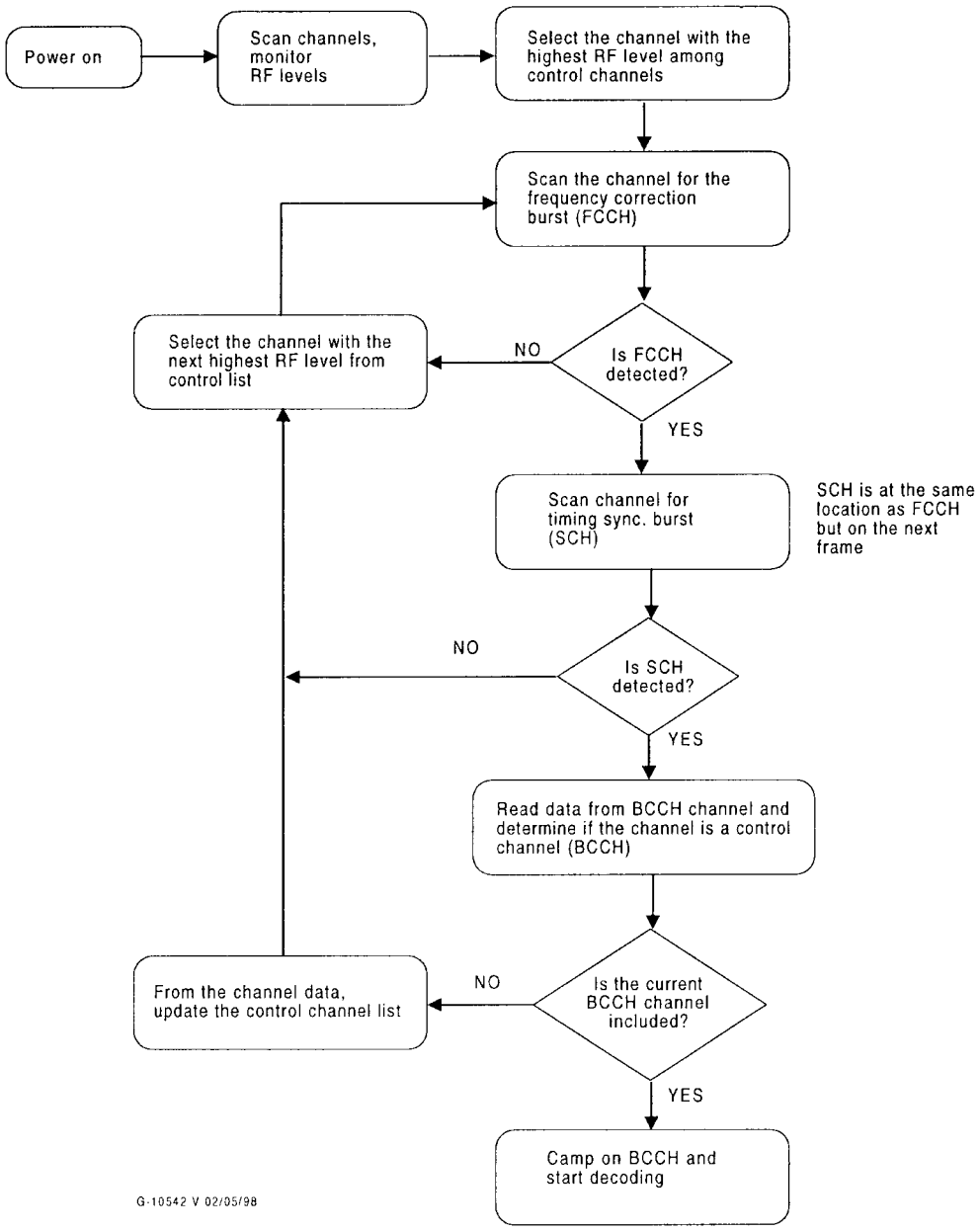


Fig. 3. Initial mobile acquisition [3, p. 113].

correction signal. The mobile has to take out this offset before an estimate of the carrier frequency can be made. This process of frequency synchronization is shown as the first step in Fig. 3 [3]. If no frequency burst is detected, then the mobile can go to a channel with the next highest signal strength level.

After the frequency correction burst is detected, the MS will try to synchronize with the time synchronization burst synchronization channel (SCH). The SCH always occurs in the next frame in the same time slot as the FCCH. This is eight burst periods later than the FCCH. The SCH contains precise timing information on the timeslot boundaries to permit refining the received slot timing. The SCH message also contains the current frame number to which the MS synchronizes. This time synchronization is generally carried out in two steps: coarse and fine. Here, the internally stored synchronizing pattern is correlated, and at the peak of correlation, the channel is considered to be synchronized. If synchronization does not occur, the process of frequency synchronization with the next highest channel in the list may start. If the synchronization is successful, the mobile will read the time division multiple access (TDMA) frame number and the base station identity code. Assuming that the mobile is in sync and decodes the information on a BCCH, the BCCH information will contain such items as adjacent cell list, BCCH location of adjacent cells, minimum received signal strength, and location area identity. The BCCH information also provides beacon frequencies of surrounding BTS cells, etc. All BCCH transmissions are at a standard power level, which permits the MS to compare received power from its own BTS as well as from adjoining BTS's. Therefore, when the

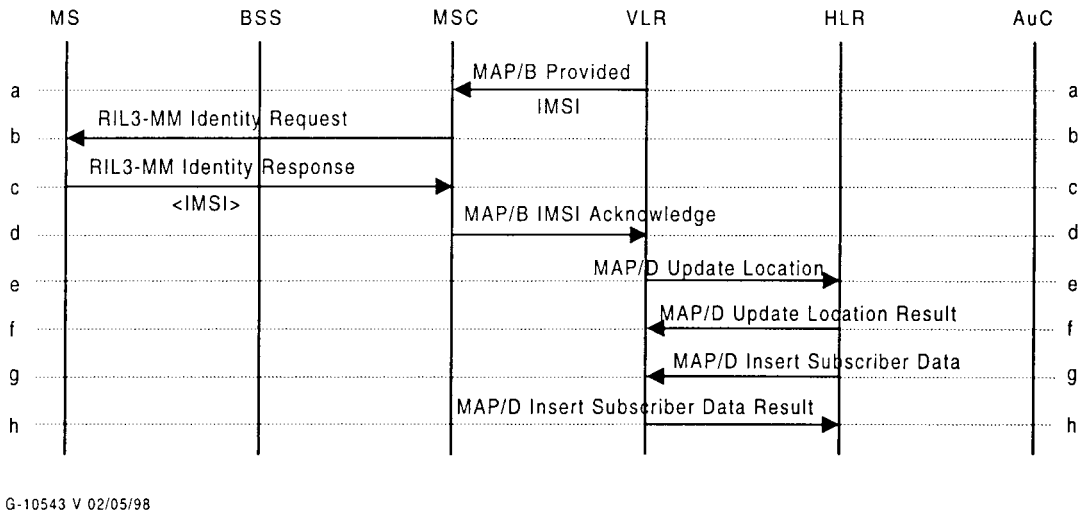


Fig. 4. Mobile identification process.

BCCH information is correctly decoded, the mobile follows one of the two paths discussed below.

a) If the BCCH information includes the present BCCH channel, then the mobile will simply stay on the channel.

b) If the current channel is not in BCCH information list, or the received signal strength level is below the desired level, the mobile will continue searching for the next control channel.

After the mobile has successfully synchronized to a valid BCCH, the mobile is now ready to register, receive paging, or originate an outgoing call.

IV. Mobility Functions

Among all functions that the mobile is at liberty to perform, we shall consider only those connected with the MM layer or the mobility aspect of the GSM system. Those functions are [2]–[3], [5]–[8]:

- mobile identification;
- authentication;
- IMSI attach and detach;
- location update.

A. Identification Procedure

The identification procedure is used to identify the MS/SIM by its IMSI if the visitor location register (VLR) does not recognize the TMSI sent by the MS. This lack of recognition can be the result of the mobile user's changing the MSC/VLR area from the last time he accessed the system or can be due to some other reason. If identification is required, the VLR first sends a MAP/B Provide IMSI message to the MSC, as shown in Fig. 4 [3]. As a result of this message, MSC sends an RIL3-MM Identity Request message to the MS. The MS responds by returning an RIL3-MM Identity Response message containing its IMSI to the MSC. The MSC then sends the MAP/B IMSI acknowledge to the VLR. If the IMSI is currently not in the VLR, then the VLR must get its file from the home location register (HLR) identified in the IMSI. To do this, the VLR sends the HLR a MAP/D Update Location message. Assuming that the IMSI is in fact registered in the HLR, the HLR responds with a MAP/D Update Location Result message, followed by a MAP/D Insert Subscriber Data message containing other pertinent data needed by the VLR. The VLR acknowledges the data transfer with a MAP/D Insert Subscriber Data result message to the HLR.

B. Authentication

The authentication process may be run at each and every location update and at the initiation of every new service request. The process starts at VLR. If the VLR determines that authentication is required, it sends a MAP/D Send Parameters message to the HLR, which relays this message to the authentication center (AuC). The AuC then draws a value for the random challenge random number (RAND) and applies algorithms A3 and A8 to generate the signed response (SRES) and the cipher key. The complete process of authentication is discussed in section 4.0. The AuC then returns the triplet (RAND, signed response, K_c) value to the VLR in a MAP/D Send Parameters Result message. Actually, the AuC normally calculates and sends a few such triplets at a time for each requesting MS, so the VLR only has to request parameters from the AuC if it has no stored unused triplets for the particular MS [10]–[12].

The VLR then sends a MAP/B Authenticate message to the MSC, which in turn sends an RIL3-MM Authentication Request message containing RAND to the MS over the air. The MS calculates the required signed response challenge (SRESc) using the algorithm A3 and authentication key K_i stored in the SIM. The SRESc is returned to the MSC in a RIL3-MM Authentication Response message. The MSC compares the SRESc with the signed response, and if they agree, it sends the MS an RIL3-MM Service Accept message. The MSC also sends the VLR a MAP/B Authentication Complete message. The protocol for the authentication process is shown in Fig. 5.

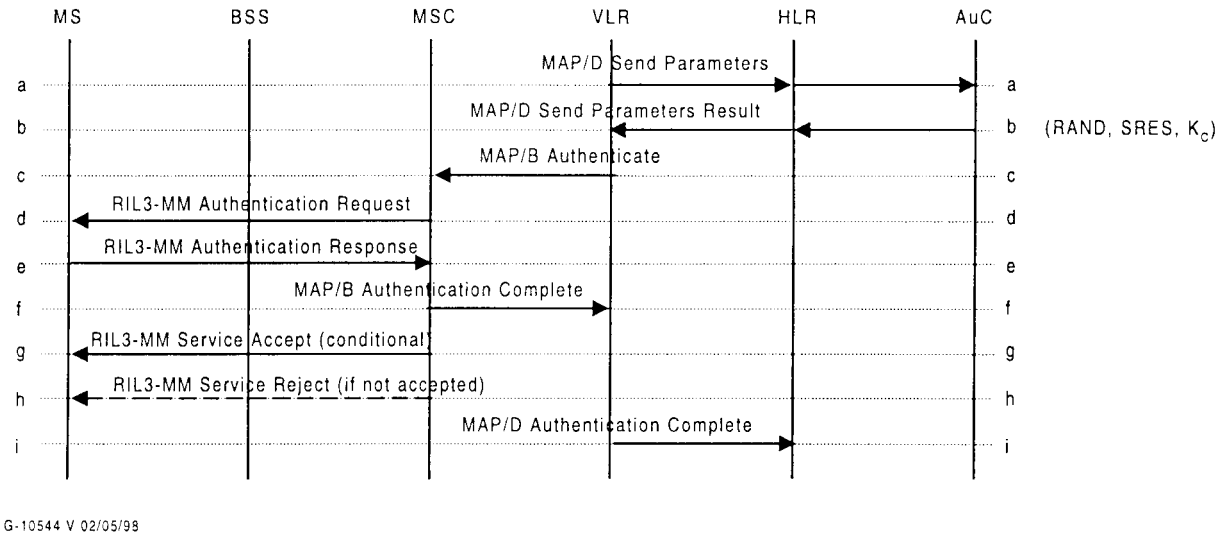


Fig. 5. MS authentication process [3, p. 119].

C. IMSI Attach and Detach

The IMSI attach and detach procedures register and deregister the mobile to the system. If the mobile user is attached, he will be paged in the location area of the user's presence. If the mobile user is detached, the system will not waste its resources in paging for an incoming call.

1) IMSI Attach: The IMSI attach procedure is used by the MS to indicate that it has reentered the active state (power on). IMSI attach is invoked if the attach/detach procedures are required by the network and an IMSI is activated in an MS (i.e., activation of an MS with plug-in SIM or the insertion of a card in a SIM card-operated MS, etc.) within the coverage area of the network or if an MS with an IMSI activated outside the coverage area enters the coverage area. The IMSI attached is marked in the MSC/VLR with an "attached" flag. The following sequence of events describes the IMSI attach procedure, shown in Fig. 6 [2]–[3]. Upon turning on the power, the MS sends an RIL3-RR Channel Request message to the BSS on the random access channel (RACH). The network assigns the channel, and the BSS sends an RIL3-RR IMM Assignment message to the MS over the access grant channel (AGCH). This message assigns the stand-alone dedicated control channel (SDCCH) channel to the mobile. After the channel is assigned, the MS sends an RIL3-MM IMSI Attach message over the SDCCH channel to the BSS, which is forwarded first to the MSC and then to the VLR as a MAP/B protocol message. The VLR sends an acknowledgment to the MSC, "IMSI Attach Acknowledge," as a MAP/B protocol, which is forwarded to the BSS and then to the MS. The MSC also sends "Clear Command" for the channel release to the BSS as the BSSMAP protocol, which is then forwarded to the MS. Upon receiving the RIL3-RR Disconnect signal from the MS, a Clear Complete message is sent to the MSC as a BSSMAP protocol.

2) IMSI Detach Procedure: Similar to the IMSI attach procedure, the IMSI detach procedure may be invoked by an MS if the MS is deactivated or if the SIM is detached from the MS. A flag (ATTN) broadcasted in the System Information message on the BCCH is used by the network to indicate to the MS whether the detach procedure is required. The procedure causes the MS to be declared inactive in the network. Once the IMSI detach procedure is active, the MS can neither transmit nor receive. The system will also not page the MS.

The IMSI detach procedure starts with the MS's sending an RIL3-RR Channel Request message on RACH to BSS. The BSS assigns an SDCCH channel and notifies the channel assignment to the MS over the AGCH. The MS then sends an RIL3-MM IMSI Detach Indication message to the BSS. The message identifies the MS (indicated here as ⟨TMSI⟩) and contains an 8-bit code indicating IMSI detach. After receiving an IMSI Detach Indication message from MS, the BSS forward this message in a BSSMAP complete layer 3 information message to the MSC. The MSC in turn updates the state of the MS in the VLR with a MAP/B Detach IMSI message. At this stage, all terminating calls to the MS are rejected, and the system does not page the mobile anymore. The VLR forwards this message to the HLR as a MAP/D Deregister Mobile Subscriber, and the HLR marks the MS as deregistered. The HLR forwards a MAP/B Deregistration Accepted message to the VLR, which, in turn, sends a MAP/B Acknowledge IMSI Detachment message to the MSC. No response is returned to the MS, as shown in Fig. 7 [2]–[3], [8]–[10]. This is correct because the mobile would be switched off before the return message is sent from the BSS to the MS. The MSC sends a BSSMAP Clear Command to the BSS to clear the SDCCH channel assigned to the MS. The BSS acknowledges with a BSSMAP Clear Complete message to the MSC.

3) Location Update: The MS location updating is performed to tell the system where to search for the MS during paging for an incoming call. If the location is known to a definite subregion of a particular PLMN, this will reduce

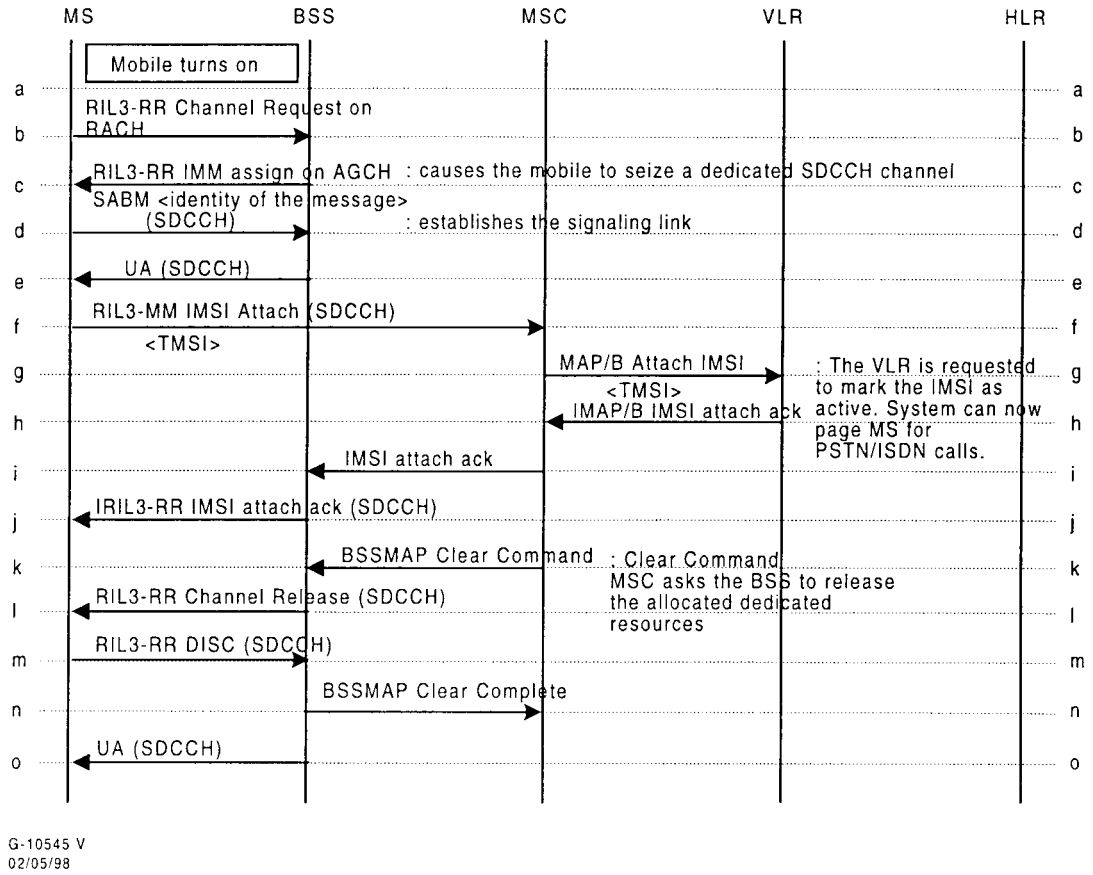


Fig. 6. IMSI attach.

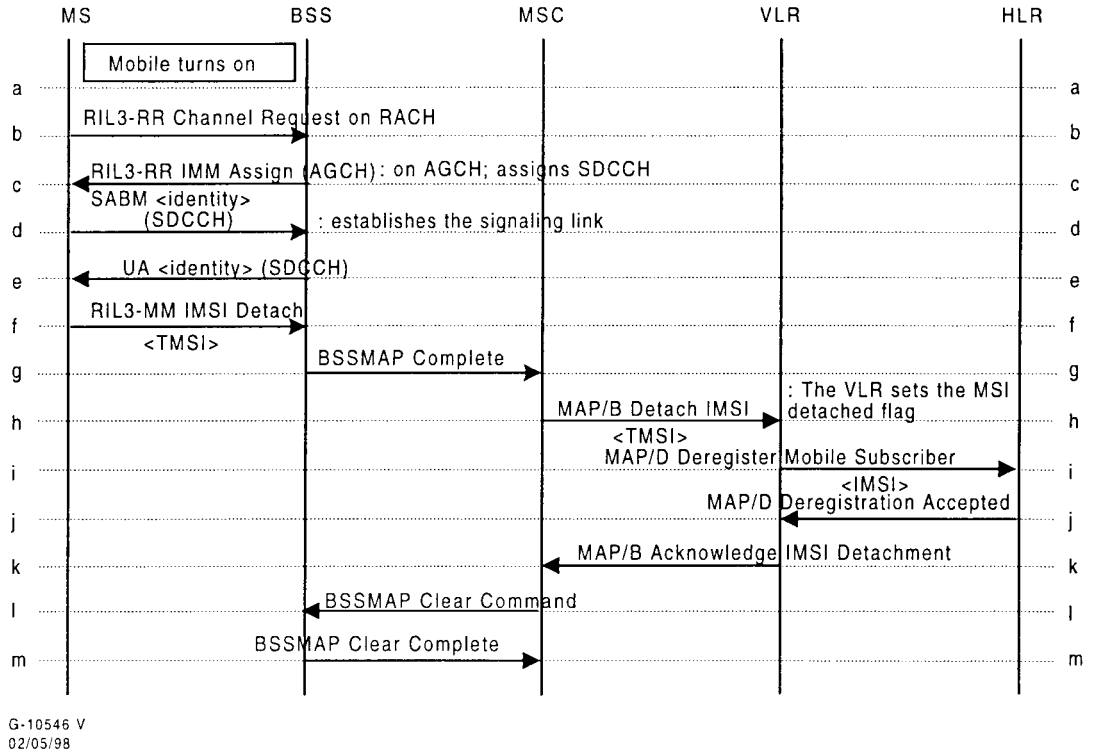


Fig. 7. IMSI detach.

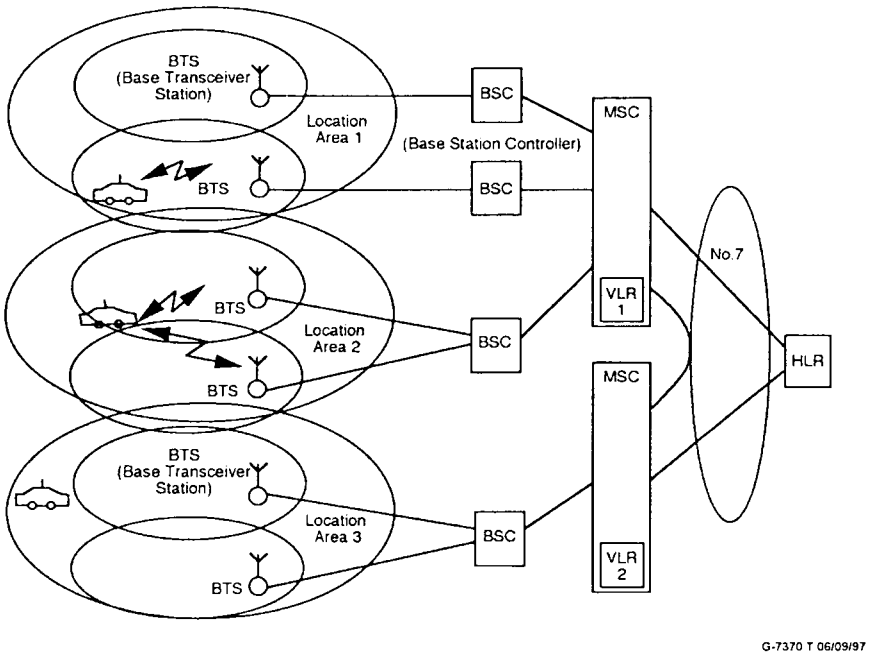


Fig. 8. Location updating [10, p. 303].

the number of cells where the mobile has to be paged, thereby reducing the load on the system.

The MS location is determined from the cell identification of the strongest BCCH signal received by the MS. The MS regularly measures the received signal strengths of the BCCH's for all surrounding cells at least once every six seconds (superframe cycle). It stores at least the six strongest BCCH measurements and their identifications in the SIM, which can subsequently be used for handover decisions. The MS also transmits the location area of the strongest cell to the MSC during location updating. The location area may be a single cell or a contiguous group of cells under the control of one BSC, as shown in Fig. 8 [3]–[4].

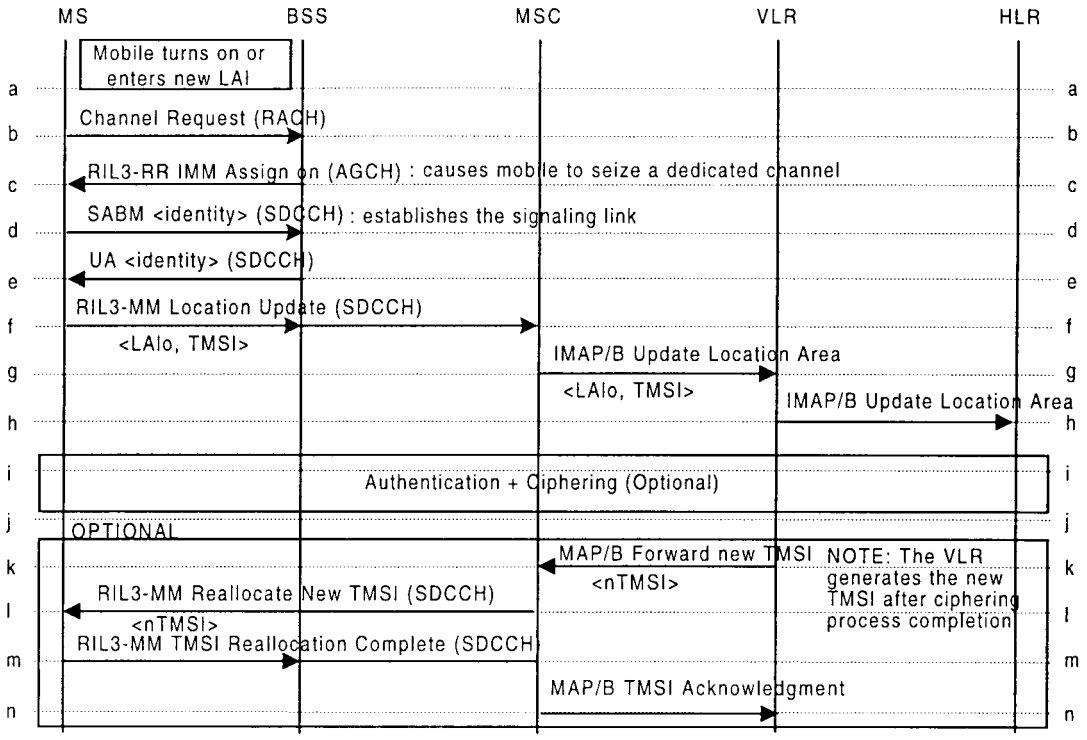
A cellular system requires that the user location of all active mobile units be known at all times as they roam. As seen in Fig. 8, each cell is served by one BTS. Each location area is divided into many cells, which may be served by one or more BSC's. The VLR may serve one or more location areas. An inactive mobile is ignored by the system. As soon as the mobile switches its power, it retrieves its stored location-area identity and compares it with the one being broadcast within its present cell. If they match, the mobile does not have to do anything, as the subscriber is already correctly located; however, if it does not match, the mobile identifies itself by transmitting its IMSI together with the identities of the previous and present location areas. The BSS transmits this information to the associated VLR.

Each time an MS moves into a new location area, the corresponding VLR is informed. If both the present and previous areas are served by the same VLR, the mobile station is given a new TMSI, and its location is updated in the VLR memory. On the other hand, if the mobile enters a new VLR area, its HLR, the old VLR, and the new VLR are informed. The old VLR erases the data for the mobile, and the new VLR records relevant parameters needed to process calls.

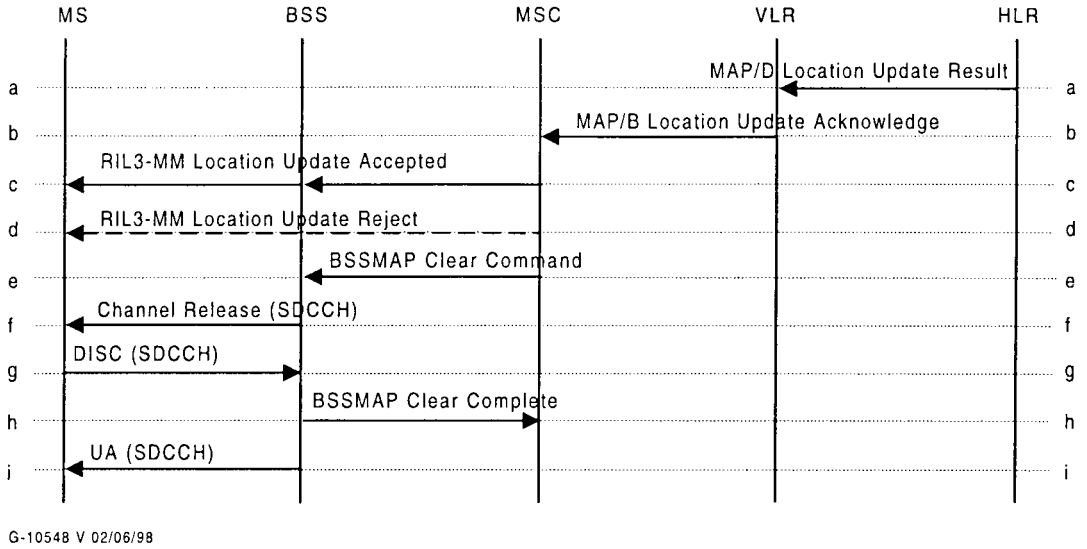
The message sequence is shown in Fig. 9. The MS is switched on in a location area different from the previous one, or it moves across boundaries of a location area in the idle state RIL3-Location Updating Request message, which is sent from the MS to the BSS and is relayed to the MSC. The MSC in turn alerts the VLR by a MAP/B Update Location Area message. The message contains the old location area that the mobile had in its storage along with its TMSI (designated here as $\langle \mathrm{LAI}_o, \mathrm{TMSI} \rangle$). The process of authentication, ciphering, and TMSI reallocation can now start. After completion of the ciphering process, the message is sent from the VLR to the MSC for reallocation of the TMSI if desired (Forward New TMSI). A TMSI Reallocation Complete message is sent from the MS to the BSS after reallocation of new TMSI. The HLR sends a MAP/D Location Update Result message to the VLR, which in turn sends a MAP/B Location Update Acknowledge message to the MSC. This message is subsequently forwarded to the MS as a RIL3-RR Location Update Accepted message. In the event that the HLR rejects the request, the VLR RIL3-MM Location Update Reject message is sent from the MSC to the MS (shown dotted). Either the accept or the reject message is initiated from MSC. Location area updating may not be accepted due to the following reasons:

- unknown subscriber;
- unknown location area;
- roaming not allowed;
- system failure.

After the location update accept or reject message, the MSC asks the BSS to release the allocated dedicated resource



(a)



(b)

Fig. 9. Location update process.

by sending a BSSMAP Clear Command message to the BSS, which then forwards it to the MS. A BSSMAP Clear Complete message follows from the BSS to the MSC, which completes the location updating process.

In the next section, we shall discuss the most important security aspects of the GSM system.

V. Security Aspects of the GSM System

At an early stage in the development of the Pan European mobile radio system GSM, it was realized that security was an important issue that needed to be addressed. It was apparent that the weakest part of the system was the radio path, as this could be easily eavesdropped upon with radio equipment. There was also a need to authenticate users of the system so that the resources are not misused by nonsubscribers [3], [10]–[12].

Therefore, the objective of this section is to outline clearly the most important security features adapted in GSM, including: a) authentication; b) ciphering; and c) an

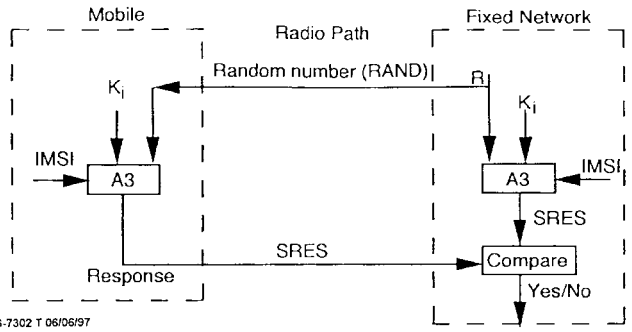


Fig. 10. Generic authentication process.

equipment ID, which ensures that no stolen or unauthorized mobile equipment is used in the system.

A. Authentication

The authentication feature ensures that, to a very high level of probability, the user is the one he claims to be. The purpose of the authentication is to protect the network against unauthorized use. It also enables the protection of the GSM PLMN. Subscriber authentication is performed at each registration, at each call setup attempt (mobile originating or terminated), and before performing some supplementary services, such as activation or deactivation of the mobile (IMSI attach, IMSI detach). Authentication is not mandatory prior to IMSI attach and detach procedures. The frequency with which a particular PLMN applies the authentication procedure to its own subscribers is their responsibility. However, a PLMN shall apply the authentication procedure to visiting subscribers as often as this feature is applied to those subscribers in their home PLMN.

GSM uses a sophisticated technique for authentication that consists of asking a question that only the right subscriber equipment (in this case, the SIM) can answer. The heart of this method is that a large number of such questions exist, and it is unlikely that the question can be answered correctly by the wrong MS. The generic process of authentication is shown in Fig. 10 [3], [10]. The authentication algorithm (called A3 in the GSM specifications) computes from a RAND, both at the MS and at the AuC, a signed response, using an individual secret key K_i attached to the mobile subscriber. The number RAND, whose value is drawn randomly between 0 and $2^{128} - 1$, is used to generate the response by the mobile as well as by the fixed part of the network. It should be noted that the authentication process is carried out both at the mobile and at the MSC simultaneously. The BSS remains transparent to this process. It should also be noted that the mobile only receives the random number over the radio path and in turn returns the signed response to the network. Thus, an air interface mobile designation is not disclosed. At subscription time, the subscriber authentication key K_i is allocated to the subscriber together with its IMSI. The key K_i is stored in the AuC and used to generate a triplet (K_c, signed response, RAND) within the GSM system. As stated above, the same K_i is also stored at the mobile in the subscriber ID (SIM). In the AuC, the following steps are carried out in order to produce one triplet. A nonpredictable RAND is produced. RAND and K_i are used to calculate the signed response and the ciphering key (K_c) using two different algorithms (A3, A8). This triplet (RAND, signed response, and K_c) is for each and every user and is then delivered to the HLR. This procedure is shown in Fig. 11 [3], [9]–[11].

The AuC begins the authentication and cipher key generation procedures after receiving an identification of the subscriber from the MSC/VLR. The AuC first queries the HLR for the subscriber's authentication key K_i. It then generates a 128-bit RAND for use as a challenge, to be sent to the MS for verification of the MS's authenticity. RAND is also used by the AuC, with K_i in the algorithm A3 for authentication, to calculate the expected correct signed response from the MS. RAND and K_i are also used in the AuC to calculate the cipher key K_c with algorithm A8. The signed response is a 32-bit number, and K_c is a 64-bit number.

The values of RAND, signed response, and K_c are transmitted to the MSC/VLR for interaction with the MS. Algorithms A3 and A8 are not fully standardized by GSM and may be specified at the direction of PLMN operators. Different PLMN's may use different and proprietary versions of these algorithms. Also, to protect the secrecy of the user, the authentication key K_i is not sent to the MSC/VLR. Based on the discretion of the PLMN operator, K_i can be of any format and length. The MSC/VLR forwards the value of the RAND to the MS, which also has the correct K_i and algorithm A3, which is stored in its SIM. The SIM then uses RAND and K_c in these algorithms to calculate the authentication SRESc and the cipher key, K_c. The MS sends the calculated response, SRESc back to the MSC/VLR, which compares it with the value signed response received from the HLR/AuC. If the SRESc and the signed response agree, the subscriber access to the system is granted, and the cipher key K_c is transferred to the BTS for use in encrypting and decrypting messages to and from the MS. If the SRESc (computed signed response at the mobile) and the signed response disagree, the subscriber access to the system is denied. In summary, the VLR initiates authentication toward the MS and checks the authentication result. The complete process is shown in Fig. 12.

B. Encryption

Obtaining good protection against unauthorized listening is not an easy matter with analog transmission, but digital transmission brings an excellent level of protection by using digital cryptographic methods. The confidentiality feature on physical connections (physical radio channels) means that the user information and signaling exchanged between the BTS and the MS are not made available or disclosed to unauthorized individuals, entities, or processes. The purpose of this feature is to ensure the privacy of the user information (voice and nonvoice) as well as the user-related signaling elements.

All speech and data are ciphered, and all associated signaling information is protected. The ciphering algorithm is synchronized with the TDMA clock and adds very little

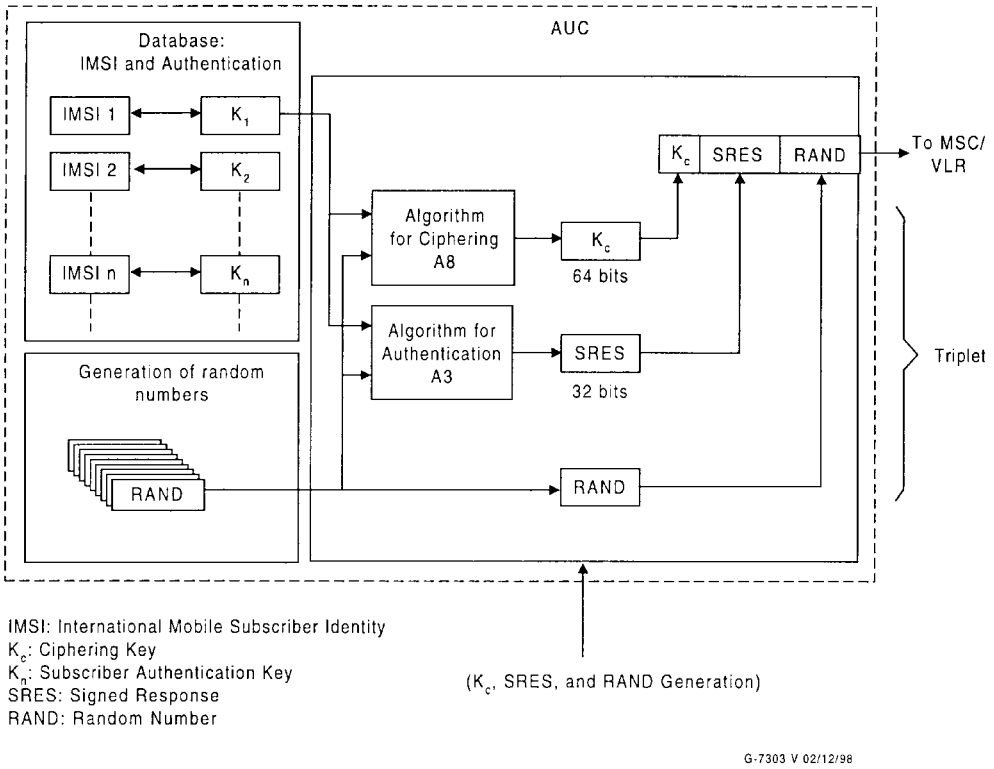


Fig. 11. Generation of K_c, signed response, and RAND at the AuC [3, p. 157].

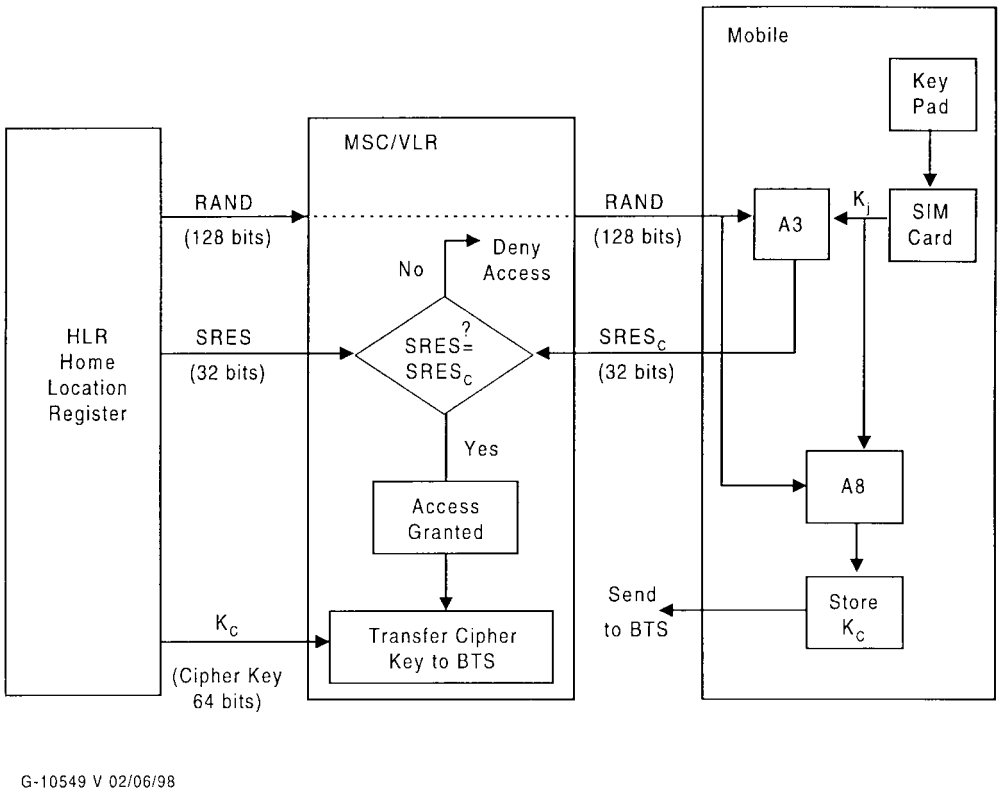


Fig. 12. Authentication process in GSM system [3, p. 158].

complexity to the MS. The cipher key is obtained as a side product of the authentication procedure and differs from call to call. The GSM is designed so that a single encryption algorithm is used for protection of all transmitted data in dedicated mode, whether it is user information (speech or data), user-related signaling (e.g., the messages carrying

the called phone numbers, etc.), or even system-related signaling (e.g., the messages carrying radio measurement results to prepare for handover). Encryption is a process in which a series of bits is transformed by mathematical or logical functions into another series of bits. The number of transformations is determined by the key so that, for an exhaustive search, all possible keys must be tried. The confidentiality of the information elements carried on the radio path (signaling and user data) is ensured by systematic encryption. The ciphering/deciphering algorithm (called A5) uses a cipher key K_c, which is allocated to each mobile subscriber during the authentication procedures. The key K_c is computed from the RAND by an algorithm (called A8) driven by the mobile subscriber authentication key K_i. Algorithm A8 is common to all GSM's. Fig. 11 has shown the process of generating K_c. For the authentication procedure, when a signed response is being calculated at the mobile, the ciphering key (K_c) is also calculated using another algorithm (A8), as shown in Fig. 12. This key setting takes place in the fixed system and in the MS. At the ciphering start command (from VLR to BSS), K_c is used by the MS and the BTS in order to cipher and decipher the bitstream that is sent over the radio path. In addition to the authentication procedures, a key setting may be initiated by the network as often as the network operator wishes. The command to use the encryption key is sent over the logical channel and dedicated control channel, as soon as the identity of the mobile subscriber is known by the network [2]–[7], [11]–[12].

The key K_c must be agreed upon by the mobile station and the network prior to the start of encryption. The choice in GSM is to compute the key K_c independently from the effective start of encryption during the authentication process. K_c is then stored in a nonvolatile memory inside the SIM so as to be remembered even after a switched-off phase. This key is also stored in the visited MSC/VLR on the network side and is ready to be used for the start of encryption. It should be noted that the actual encryption/decryption of user data (e.g., speech) takes place within the mobile station and the BSS. For this purpose, the encryption key is downloaded from the MSC to the BTS via the BSC. After authentication, the transmission is ciphered, and K_c is used for ciphering/deciphering. This process is shown in Fig. 13.

Data flow on the radio path is obtained by a bit-per-bit binary addition of the user data flow and ciphering bitstream generated by the GSM algorithm A5 using a ciphering key (K_c). This exact process of encryption/decryption at the mobile and at BTS is shown in Fig. 14. Code words S_1 and S_2 for downlinks and uplinks are changed at every frame. When modulo 2 is added with plain text, S_1 outputs cipher text. On the other side, the cipher text, when modulo 2 is added with S_1, outputs the plain text. The ciphering/deciphering function is placed on the transmission chain between the interleave and the modulator. Since A3 and A8 are always running together, these two are implemented as a single algorithm in most cases. The algorithm A3 is standardized in the whole of GSM.

C. Equipment Identification Definition

The administrative use of the international mobile equipment identity (IMEI) enables the operator to check the mobile equipment identity at call setup [1]–[10]. The purpose of this feature is to make sure that no stolen or unauthorized mobile equipment is used in the system, which is another security aspect of the system.

The equipment identification procedure consists of the MSC/VLR's requesting the IMEI from the MS and sending it to a stand-alone entity called EIR.

On reception of the IMEI at the AuC, the EIR makes use of three possible defined lists.

- A white list containing all numbered series of all equipment identities that have been allocated in the different participating GSM countries.
- A black list containing all equipment identities that are barred. This listing may be the result of stolen equipment.
- A gray list containing (at the operator's decision) faulty or nonapproved mobile equipment. This equipment is under observation but not barred for service.

Although the GSM specification recommends using the equipment ID at each and every call, the frequency of identification really lies with the individual operator. The system operator can make decisions in this regard. The equipment identification process starts with the MSC/VLR's requesting MS for its IMEI. In response, the MS sends its ID that, if positively checked by the equipment ID register, allows the mobile to proceed further with the call. Mobile is not allowed to continue with the call if the equipment ID does not match the stored value of the ID in the register. The complete equipment identification process is shown in Fig. 15 [3]. As shown in the figure, an IMEI request is initiated by the MSC/VLR combination as a result of the MS's requesting a call setup. Upon receiving the IMEI request, the MS sends the equipment identification to the MSC/VLR, which is subsequently checked against the stored values in the EIR.

In the next section, we shall discuss future enhancement in those areas where there are presently limitations.

VI. Solution Approach for Future Systems

In this section, we shall discuss approaches to some solutions for the future systems based on the present limitations of the GSM system. The areas discussed are:

- SIM roaming;
- intersystem roaming;
- future mobile service;
- data base requirements.

A. SIM Roaming

One solution of roaming can be solved by having a SIM card that works on multiple standards. Such a card will be highly desirable for future use and will allow the subscriber

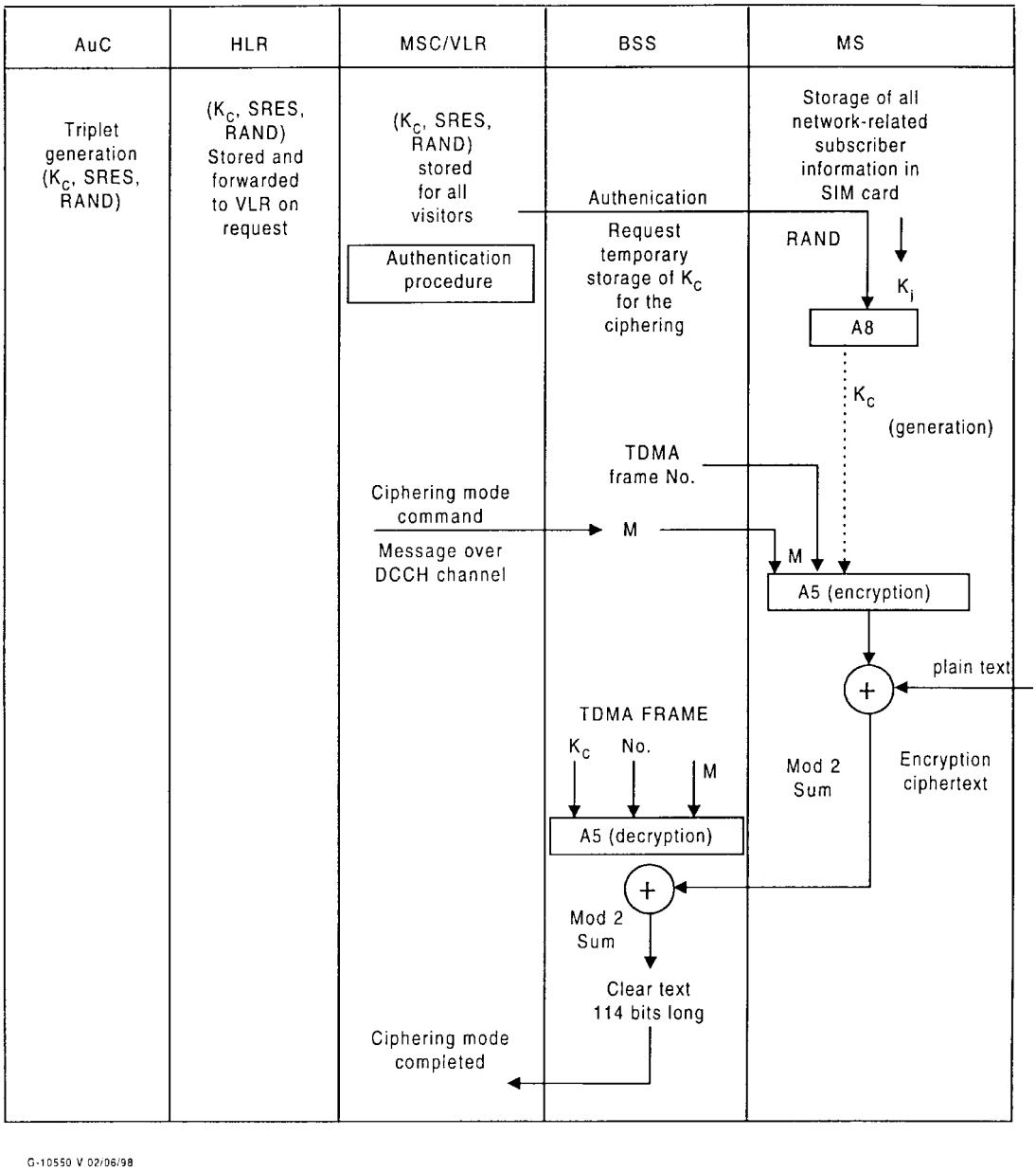


Fig. 13. Sequential steps for encryption and decryption process [3, p. 160].

not to carry a privately owned handset from place to place. Wherever the user goes, he can rent a cellular telephone, and it will work with the user's SIM card.

In this case, the SIM card will have data according to several standards. For example, such a dual-mode SIM card can operate on the GSM system when inserted into a GSM terminal, i.e., a GSM terminal would access the GSM directory on the SIM. Similarly, a packet data terminal would access the packet data directory on the SIM when the SIM is inserted into a packet data terminal. In the future, this concept can be expanded, and one SIM card can work on many standards.

This solution has many advantages. First, it allows for the use of already existing terminals (thus not requiring new terminals to be developed) for end users to roam among different standards. Second, the use of such a SIM card can enable the mobile subscriber to have the necessary data for all cellular standards on one SIM card, enabling a mobile subscriber to choose different systems based on personal choice and the availability of the system. This multisystem SIM card has many advantages.

- It allows for use of already existing terminals.
  - — No new terminal development is required for end users roaming among different standards.
- The use of such a SIM card would enable the mobile subscriber to have the necessary data for all cellular standards on one SIM card.
  - — This enables a mobile subscriber to choose different systems based on personal choice and the availability of the system.

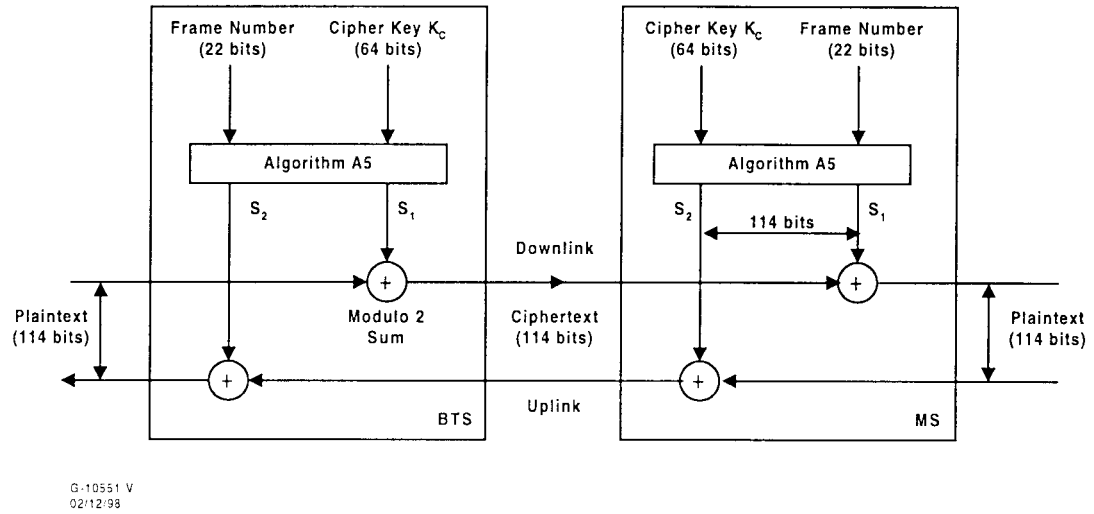


Fig. 14. Encryption/decryption process [3, p. 161].

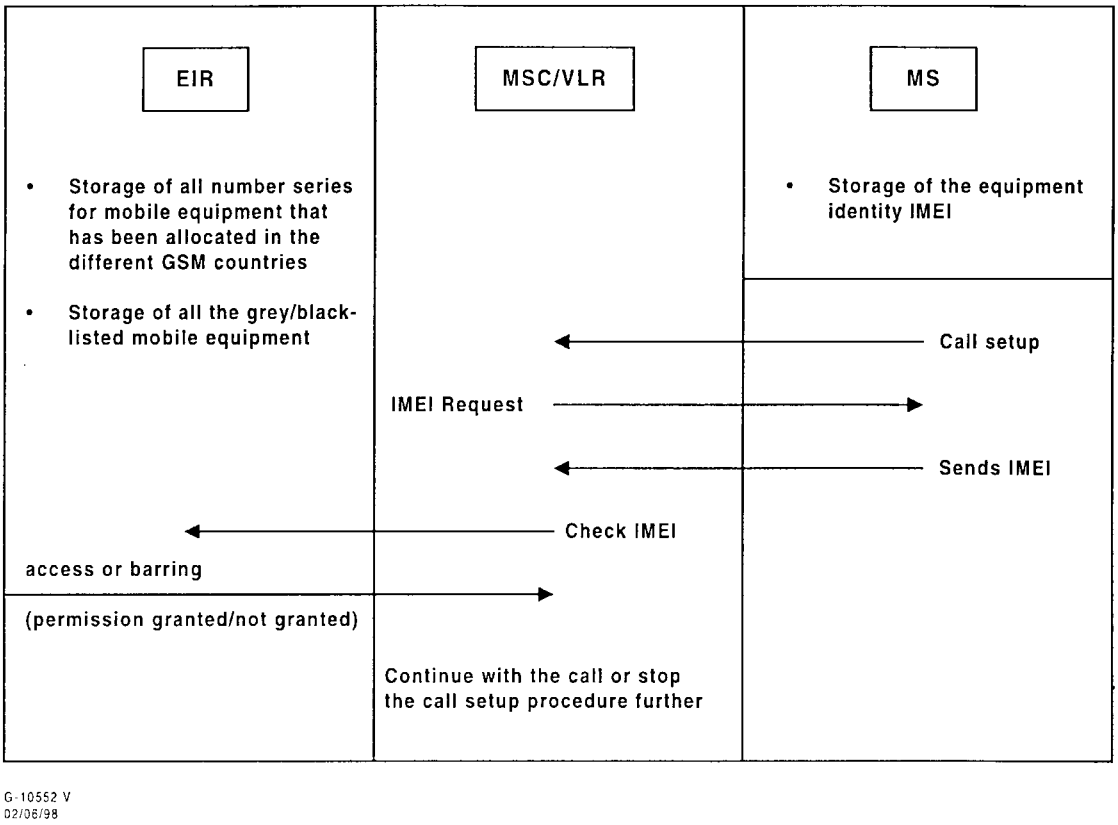


Fig. 15. Equipment identification process.

Fig. 16 shows such a SIM card working on systems A and B [3], [15]–[18].

B. Intersystem Roaming

Currently, a number of cellular standards exist, each of them supporting seamless roaming and call delivery within each standard. These standards include air interface procedures, whereby an MS can access a visited network in the same way as it would access its home network. Network interfaces allow networks run by different operators to be interconnected using commonly agreed upon signaling procedures. One standard is GSM, which supports international roaming. Other examples include the Advanced Mobile Phone System (AMPS) standard, which supports roaming within North America today, and packet data terminals, which support roaming within Japan. In most of the above-mentioned cellular standards, the MAP is the signaling protocol that handles the roaming signaling in the respective network. For a standard to provide roaming, it is equally important that the standard include a fully

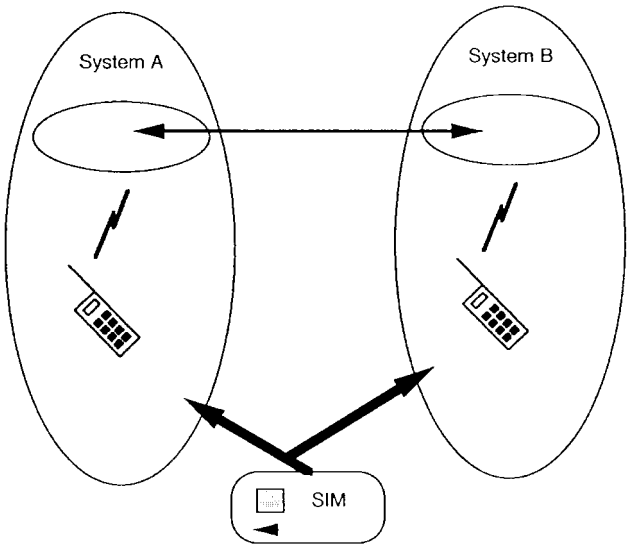


Fig. 16. SIM roaming.

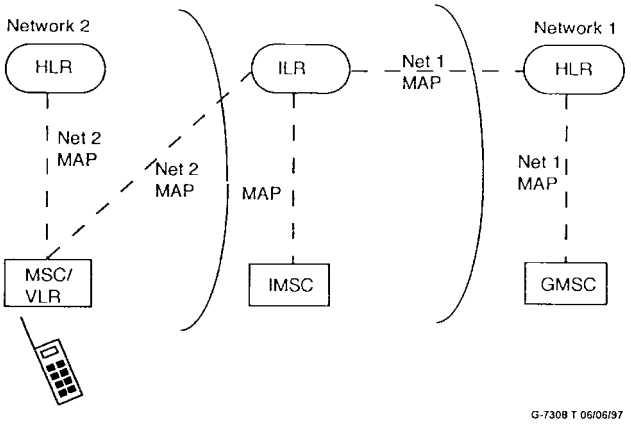


Fig. 17. General network architecture.

specified air interface as well as the necessary network interface(s). However, all the standards mentioned here contain their own specific air interface specifications as well as their specific MAP protocols, which do not allow roaming between two different systems. Thus, it is not possible to roam between GSM and AMPS, nor is it possible to roam between GSM and packet data terminals. At a future date, the scheme presented here will allow roaming between two different systems, each signaling within its own system and in its own MAP protocol.

The scheme shown in Fig. 17 introduces new nodes that have the purpose of providing interworking between two different systems [3], [19]–[24]. The nodes that perform interworking between two different cellular standards are the interworking location register (ILR) and the interworking mobile switching center (IMSC). The ILR works as follows. In the view of the HLR in the home network (Network 1 in the figure), the ILR is seen as a VLR. In view of the MSC/VLR in the visited network, the ILR is seen as an HLR. This circumstance implies that the ILR can ensure that the two interconnected standards can cooperate. An

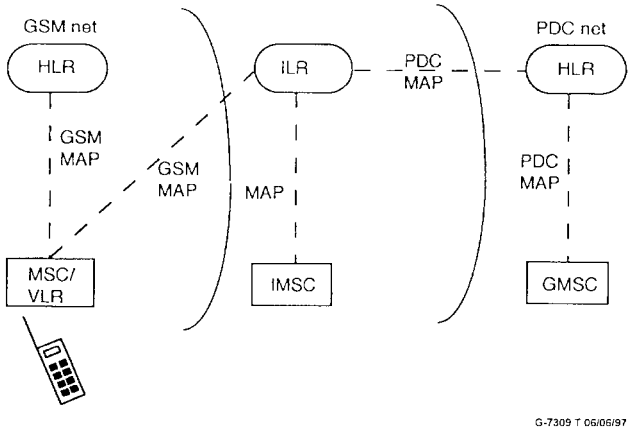


Fig. 18. Network architecture for roaming from packet data to GSM [3, p. 386].

example below provides the basic scheme for international roaming.

1) Example: Let us consider an example of two independent systems having an ILR/IMSC interface between them. Let us also assume that a subscriber belonging to the GSM system switches on the power. The location updating procedure from the GSM terminal will be handled by the MSC/VLR in the GSM network. The GSM MSC/VLR will communicate with the ILR using the normal GSM MAP protocol, assuming the ILR to be the GSM HLR described previously. When receiving the GSM MAP location update procedure from the GSM MSC/VLR, the ILR will perform a packet data MAP location registration procedure toward the packet data HLR to update the location data for the subscriber in the packet data HLR. When the GSM MAP procedure is completed between the GSM MSC/VLR and the ILR, the GSM location is stored in the GSM MSC/VLR and the ILR in accordance with the GSM standard. Now a call from a fixed subscriber to a packet data subscriber who is roaming in GSM will work as follows. The gateway mobile switching center (GMSC) in the packet data network will, when receiving the call from the fixed network, ask the packet data HLR to terminate call-routing information. The packet data HLR will thus provide the packet data GMSC with routing information, which enables the packet data GMSC to route the call to the IMSC. The ILR will, when asked for GSM routing information, provide a GSM roaming number to the IMSC. The IMSC will now forward the call to the GSM MSC/VLR using the received GSM roaming number, and the call is completed in the GSM MSC/VLR as a normal GSM mobile terminated call. The main reason for having the IMSC is that the functional interface between the GMSC and the MSC/VLR differs in the GSM and packet data systems. To enable routing of a call originally set up for a packet data subscriber when the subscriber is roaming in a GSM network, the IMSC has the role of interworking between packet data and GSM, specifically handling routing of calls to roaming subscribers. This scheme is shown in Fig. 18.

The approach described above will allow subscribers to roam internationally by making use of the multimode

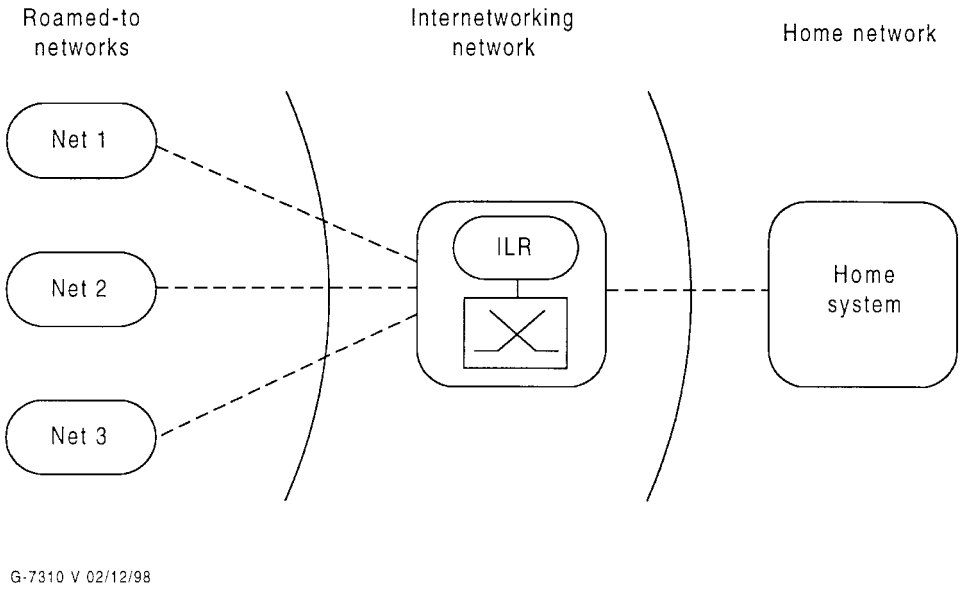


Fig. 19. Multimode terminal [3, p. 386].

terminals (multimode terminals will allow a subscriber to access different systems) as shown in Fig. 19 [18]–[20]. This terminal will allow a mobile to have worldwide service just by having a single multimode terminal.

C. Future Mobile Service and Data Base Requirements

As discussed above, future universal mobile telecommunications service (UMTS) will consist of a wide variety of interconnected networks and facilities. For mobile originated transactions, let us assume that a terminal can access the base station, which responds to a request for service and can then proceed to negotiate conditions and price.

Because payment can be by electronic cash over the air, a user does not have to subscribe to or belong to the network being used. The base station or the network to which it is attached will have sufficient intelligence to carry out an electronic cash or credit transaction. This is likely to be a commonplace function in the near future for many types of remote shopping. Because cash is transferred electronically, the subscriber does not have to belong to the specific network from which he wants the service. We then have the problem of locating the mobile terminal and establishing the connection via one or more competing networks. The following example is provided to show how the mobile can be located potentially in a UMTS/future public land mobile telephone system.

Consider the following example. There is an urgent need to contact you: I call your personal number. It is routed to your home because, after 6 p.m., you are most often at home. The phone in your home rings. After five seconds with no response, I am asked if I wish to pay for a search or to be connected to an answering machine or voice mail service. Because I need to speak to you urgently, I ask for a "grade I search" (assume grade I is a worldwide search, no expense spared, costs up to my credit limit; lesser grades possible can be activated also with a limitation of an area search, such as one country or a few countries). The telecom operator to whom I am connected may now contact a location agency (LA). The LA will search a data base to see what it knows about you. Based on knowledge of your habits and any recent calls to or from you, the search may then proceed by one or more of the following ways. You are working late: try your office. Ask all personal communications network and cellular operators within the country, "Can you connect to X, and if so, what is the charge?" Ask the local paging operators to page you, but suppose the LA data base knows its last contact with you was an hour before from XYZ hotel in New York. A call is made to your personal number via the hotel switchboard. No response is obtained from either your room extension or the hotel's cordless telephone system. Therefore, a New York radio-paging operator is asked to page you. A response comes back from your pocket phone.

The LA computer will then calculate the cost of a call using its knowledge of networks and organizations to determine the lowest cost options and then call me back and offer to connect me to you at a cost of $2 per minute for voice or $5 per megabyte of data transmission. The associated data base requirements, based on the complex problem we have just shown, will now be discussed.

D. Data Base Requirements

To locate an individual, a comprehensive data base must contain all relevant information for the user. The data base might also keep success statistics for some or all of the "usual places" for an individual. For instance, it could generate and keep hour-by-hour probabilities. Other information in the data base might include which services the user subscribes to or authorizes to use, etc. The usual information about the subscriber, which has to be stored in the data base, is shown in Table 1 [20]–[23]. It should be noted that possibly there will be many LA's maintaining the data base. Each individual agency will only have to

Table 1 Location Agency Data Base Entries

Possible Entries for the Subscriber in the Database		
Subscriber Name		
Home Number		
Office Number		
Personal Number		
Carphone Number		
Usual location: Home		
except:		
Mon to Fri	0800 to 1700	Office
Sat	1000 to 1200	Gaithersburg Shopping Center, MD
Sun	0900 to 1300	Golf Club at Montgomery Village, Gaithersburg
Location put in by user as a form of "follow me." Last known location (e.g., 04/30/95 from 12.00 hrs to 16.00 hrs. (San Francisco))		

maintain a fraction of the world's capacity. Apparently the activity (load) factor has to be estimated. Data base information sources can be from telephone companies, mobile network operators, national numbering authorities, and directly from the users for whom the data base has been created. We now provide an example of a rough estimate for the total data base storage requirement and the number of location requests/second.

1) Example: Assuming that the world population is ≈6 billion and that only 20% will use the service, let us also assume that, on the average, there are four transactions per day per person and 10% of those will require locating the destination terminal. Then, the number of location processes per second is

$$6 \times 10^9 \times 0.2 \times 4 \times 0.1 = 480 \times 10^6 \text{ per day} = 5500 \text{ queries/second.}$$

Since this will vary during the time of the day, we can assume a safe factor of two, which will bring us to ≈11 000 queries/second. If each user requires about 300 bytes on the average, the total storage requirement for 6×10^9 people is $300 \times 0.2 \times 6 \times 10^9 = 3600$ Gbytes.

Total data base size, update loads, and access times are calculated assuming once again that only 20% of the world population has the service and that each entry is updated on the average four times a day. The total number of transactions per day $= 0.2 \times 6 \times 10^9 \times 4 = 2.4 \times 10^9$. Assuming that the peak load is double the average, the total peak load becomes 4.8×10^9.

VII. Summary and Conclusions

Protocols using the MM layer have been discussed. Important aspects of authentication, encryption, and the positive identification of mobile equipment before providing the user with service have been fully explored. Authentication ensures that the network is accessed by the legitimate subscribers. The radio path is protected due to ciphering. Equipment ID ensures that the mobile is using the correct brand of transceivers. True mobility for the user can be achieved only by multiple entries on the SIM card, by design of multimode terminals, and by the availability of fast and large data bases.

References

[1] European Telecommunication Standard Institute/Global System for Mobility, "Mobile radio interface layer 3 specification," ETSI/GSM section 4.08, Apr. 1993.

[2] M. Mouly and M.-B. Pautet, "The GSM system for mobile communications," M. Mouly and M.-B. Pautet, Eds., section 2.3, 1992.

[3] A. Mehrotra, *GSM System Engineering*. Norwood, MA: Artech House, 1997.

[4] A. Maloberti *et al.*, "Radio subsystem functions and elements," presented at the GSM Seminar-Budapest, Hungary, session 3.1, Oct. 1990.

[5] U. J. Brune, "The mobile application part protocol," presented at the GSM Seminar-Budapest, Hungary, session 4.2, Oct. 1990.

[6] M.-B. Pautet and M. Mouly, "GSM protocol architecture: Radio sub-system signaling," in *Proc. IEEE VT Conf.*, 1991, pp. 326–332.

[7] Global System for Mobility, "Mobile application part (MAP) specification," GSM Recommendation 09.02.

[8] M. Mouly and M.-B. Pautet, "Security Management," in *The GSM System for Mobile Communications*, M. Mouly and M.-B. Pautet, Eds., section 7.2, pp. 477–492, 1992.

[9] S. M. Redl, M. K. Weber, and M. W. Oliphant, "Security parameter," in *An Introduction to GSM*. Norwood, MA: Artech House, 1995, section 3.8, pp. 44–48.

[10] A. Mehrotra, *Cellular Radio Analog and Digital Systems*. Norwood, MA: Artech House, 1994, section 7.5.2.4, pp. 305–309.

[11] European Telecommunication Standard Institute/Global System for Mobility, ETSI/GSM specification vol. 2.17, section 3, Jan. 1993.

[12] European Telecommunication Standard Institute/Global System for Mobility, ETSI/GSM specification vol. 3.20, section 3, Jan. 1993.

[13] A. Mehrotra, *Cellular Radio Performance Engineerings*. Norwood, MA: Artech House, 1994.

[14] O. Spaniol, A. Fasbender, S. Hoff, J. Kaltwasser, and J. Kassubek, "Impact of mobility on telecommunication and data communication networks," *IEEE Personal Commun. Mag.*, pp. 20–33, Oct. 1995.

[15] B. Gabelgaard, "Converging GSM and IN—Deploying PCS services in GSM," in *Proc. Telecom 1995*, session 5.1, pp. 21–25.

[16] J. S. da Silva, "Mobile and personal communications—The European R&D perspective," in *Proc. Telecom 1995*, session 1.3, pp. 51–55.

[17] R. Prasad, "European radio propagation and subsystems research for the evolution of mobile communications," *IEEE Commun. Mag.*, Feb. 1996.

[18] M. Grant, "Personal mobility vs. terminal mobility: Development of personal numbering services in Europe's personal communications industry," in *Proc. Telecom 1995*, session 1.3, pp. 41–43.

[19] E. Damosso and G. de Brito, "Cost 231 achievements as a support to the development of UMTS: A look into the future," *IEEE Commun. Mag.*, pp. 90–96, Feb. 1996.

[20] M. Mouly and M.-B. Pauted, "Current evolution of the GSM systems," *IEEE Personal Commun. Mag.*, pp. 9–19, Oct. 1995.

[21] R. W. Gibson, "Location Agencies" for universal mobile telecommunications services," in *Proc. Telecom 1995*, session 1.3, pp. 45–49.

[22] P. Ramsdale, "The path to third generation personal communications systems," in *Proc. Telecom 1995*, session 5.4, pp. 174–180.

[23] D. MacFarlane, "Third generation mobile—UMTS and FPLMTS," in *Proc. Telecom 1995*, session 5.4, pp. 185–188.

[24] H. Sherry, "A united states perspective PCS standards," in *Proc. Telecom 1995*, session 6.2, pp. 351–355.

Asha Mehrotra (Member, IEEE) received the Ph.D. degree in electrical engineering from the Polytechnic Institute of New York, Brooklyn.

He is an Advisory Engineer with Hughes Network Systems, a Hughes Electronics Corporation company, where he is working on problems of mobility management, handover, diversity, and encryption on the ICO project. Prior to joining Hughes Network Systems, he was with TASC as a Senior Principal Engineer, where he worked on various classified and unclassified projects, including a consulting role with the Systems Engineering Group on the IRIDIUM project. He has been an Adjunct Professorial Lecturer at George Washington University, Washington, DC, for nearly 20 years. He has published three books on cellular communication, including his latest publication on GSM systems from Artech House.

Leonard S. Golding (Life Fellow, IEEE) received the A.B. and B.Sc. degrees from Columbia University, New York, and the doctorate degree from Yale University, New Haven, CT.

He currently is a Vice President with Hughes Network Systems (HNS), a Hughes Electronics Corporation company. He is in charge of systems engineering at HNS and has been involved in directing and working on projects in direct broadcast satellite, mobile satellite system, and and new multimedia fixed satellite system programs. He also has been involved in terrestrial wireless communication projects, including personal communications systems such as perceptual audio coder and broad-band wireless local loop systems in the millimeter-wave bands. He also oversees HNS's standards activities and is involved in new business development activities. He has served on numerous professional and industry committees and has more than 40 publications. He has received ten patents.