North American and European standards for mobile communications

Two User Location Strategies for Personal Communications Services

SESHADRI MOHAN, RAVI JAIN

The vision of nomadic personal communications is the ubiquitous availability of services to facilitate the exchange of information (voice, data, video, image, etc.) between nomadic end users independent of time, location, on access arrangements. To realize this vision, it is necessary to locate users who move from place to place. The strategies commonly proposed are two-level hierarchical strategies, which maintain a system of home and visited databases — home location register (HLR) and visitor location register (VLR) — to keep track of user locations. Two standards exist for carrying out two-level hierarchical strategies using HLRs and VLRs. The standard commonly used in North America is the Electronics Industry Association/Telecommunications Industry Association (EIA/TIA) Interim Standard 41 (IS-41) [1], and in Europe the Global System for Mobile Communications (GSM) [2, 3].

In this article we introduce these two strategies for locating users and provide a tutorial on their usage. In the following section, different forms of mobility in the context of PCS and a reference model for a PCS architecture are discussed. The user location strategies specified in the IS-41 and GSM standards, respectively, are described, and then, using a simple example, a simplified analysis of the database loads generated by each strategy is presented. Also briefly discussed are possible modifications to these protocols that are likely to result in significant benefits by reducing query and update rate to databases and/or reducing the signaling traffic.

The choice of platforms on which to realize the two location strategies (IS-41 and GSM) may vary from one service provider to another. A possible realization of these protocols based on the advanced intelligent network (AIN) architecture [4, 5] and Signaling System Number 7 (SS7) is described herein. It is also worthwhile to point out that several strategies have been proposed in the literature for locating users, many of which attempt to reduce the signalling traffic and database loads imposed by the need to locate users in PCS.

An Overview of PCS

This section explains different aspects of mobility in PCS using an example of two nomadic users who wish to communicate with each other, and describes a reference model for PCS.

Aspects of Mobility: An Example

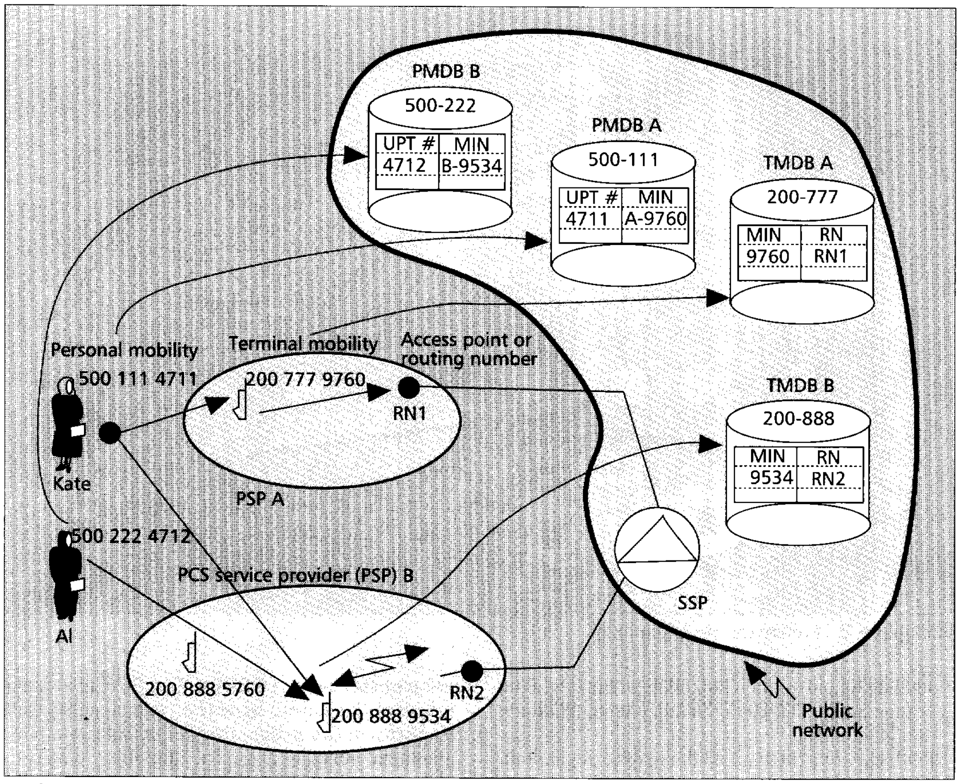
PCS can involve two possible types of mobility: terminal mobility and personal mobility.

Terminal mobility allows a terminal to be identified by a unique terminal identifier independent of the point of attachment to the network. Calls intended for that terminal can therefore be delivered to that terminal regardless of its network point of attachment. To facilitate terminal mobility, a network must provide several functions, which include those that locate, identify, and validate a terminal and provide services (e.g., deliver calls) to the terminal based on the location information. This implies that the network must store and maintain the location information of the terminal based on a unique identifier assigned to that terminal. An example of a terminal identifier is the IS-41 EIA/TIA cellular industry term "mobile identification number" (MIN), which is a North American numbering plan (NANP) number; it is stored in the terminal at the time of manufacture and cannot be changed. A similar notion exists in GSM (to be discussed).

Personal mobility allows a PCS user to make and receive calls independently of both the network point of attachment and a specific PCS terminal. This implies that the services to which a user has subscribed (stored in that user's service profile) are available to the user even if the user moves or changes terminal equipment. Functions needed to provide personal mobility include those that identify (authenticate) the end user and provide services to an end user independent of both the terminal and the location of the user. An example of a functionality needed to provide personal mobility for voice calls is the need to maintain a user's location information based on a unique number, the universal personal telecommunications (UPT) number, assigned to that user. UPT numbers are also NANP numbers. Another example would allow end users to define and manage their service profiles to enable users to tailor services to suit their needs. The way in which GSM caters to personal mobility via smart cards is described in a following section.

For the purposes of the example that follows, the terminal identifiers (TIDs) and UPT numbers are NANP numbers; the distinction is that TIDs address terminal mobility and UPT numbers address personal mobility. Though we have assigned two different numbers to address personal and terminal mobility concerns, the same effect could be achieved by a single identifier assigned to the terminal, which varies depending on

1070-9916/94/$04.00 © 1994 IEEE



■ **Figure 1.** *Illustrating terminal and personal mobility.*

the user currently utilizing the terminal. For simplicity, we assume that two different numbers are assigned.

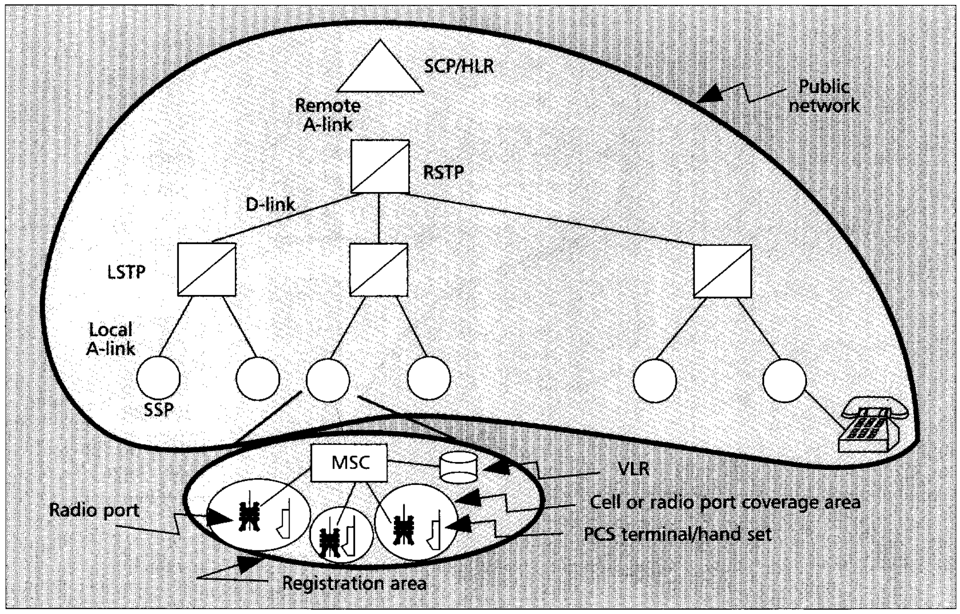
Figure 1 illustrates the terminal and personal mobility aspects of PCS, which will be explained via an example. Let us assume that users Kate and Al have subscribed to PCS services from PCS service provider (PSP) A and PSP B, respectively. Kate receives her UPT number, say 500 111 4711, from PSP A. She also owns a PCS terminal with TID 200 777 9760. Al receives his UPT number, 500 222 4712, from PSP B, and he owns a PCS terminal with TID 200 888 5760. Each has been provided a personal identification number (PIN) by her/his respective PSP when subscription began. We assume that the two PSPs have subscribed to PCS access services from a certain network provider such as a local exchange carrier (LEC). (Depending on the capabilities of the PSPs, the access services provided may vary. Examples of access services include translation of a UPT number to a routing number (RN), terminal and personal registration, and call delivery [6, 7].)

When Kate plugs her terminal into the network or activates it, the terminal registers itself with the network by providing its TID to the network. The network creates an entry for the terminal in an appropriate database, which in this example is entered in terminal mobility database (TMDB) A. The entry provides a mapping of her terminal's TID, 200 777 9760, to an RN, RN1. All of these activities happen without Kate being aware of them. After activating her terminal, Kate registers herself at that terminal by entering her UPT number (500 111 4711) to inform the network that all calls to her UPT number are to be delivered to her at her terminal. For security reasons the network may want to authenticate her, and she may be prompted to enter her PIN number into her terminal. (Alternatively, if the terminal is equipped with a smart card reader, she may enter her smart card into the reader. Other techniques, such as voice recognition, may be employed.) Assuming that she is authenticated, Kate has now registered herself. As a result of personal registration by Kate, the network creates an entry for her in personal mobility database (PMDB) A, which maps her UPT number to the TID of the terminal at which she registered. Similarly, when Al activates his terminal and registers himself, appropriate entries are created in TMDB B and PMDB B.

Functions needed to provide personal mobility include those that identify (authenticate) the end user and provide services to an end user independent of both the terminal and the location of the user.

Now Al wishes to call Kate, so he dials Kate's UPT number (500 111 4711). The network carries out the following tasks:

- The switch analyzes the dialed digits and recognizes the need for AIN service, determines that the dialed UPT number needs to be translated to an RN by querying PMDB A, and hence queries PMDB A.
- PMDB A searches its database and determines that the person with UPT number 500 111 4711 is currently registered at the terminal with TID 200 777 9760.



■ **Figure 2.** *A reference model for PCS.*

- PMDB A then queries TMDB A for the RN of the terminal with TID 200 777 9760. TMDB A returns the RN (RN1).
- PMDB A returns RN1 to the originating switch.
- The originating switch directs the call to switch RN1, which then alerts Kate's terminal. The call is completed when Kate picks up her terminal.

Kate may take her terminal wherever she goes and perform registration at her new location. From then on, the network will deliver all calls for her UPT number to her terminal at the new location. In fact, she may actually register on someone else's terminal too. For example, suppose that Kate and Al agree to meet at Al's place to discuss a school project they are working on together. Kate may register herself on Al's terminal (TID 200 888 9534). The network will now modify the entry corresponding to 4711 in PMDB A to point to B-9534. Subsequent calls to Kate will be delivered to Al's terminal.

It is quite possible that one or more registration areas may be served by a single PSP. The PSP may have one or more HLR(s) for serving its service area.

The scenario given here is used only to illustrate the key aspects of terminal and personal mobility; an actual deployment of these services may be implemented in ways different from those suggested here. Personal registration will not be discussed further in this article. The analyses that follow consider only terminal mobility, but may easily be modified to include personal mobility.

A Model for PCS

Figure 2 illustrates the reference model used for the comparative analysis. The model assumes that the HLR resides in a service control point (SCP) connected to a regional signal transfer point (RSTP). The SCP is a storehouse of the service logic (i.e., functionality used to perform the processing required to provide advanced services, e.g., speed calling, outgoing call screening, etc.) in the AIN architecture [4, 5]. The RSTP and the local STP (LSTP) are packet switches, connected together by various links such as A-links or D-links, which perform the signaling functions of the SS7 network. Such functions include, for example, global title translation for routing messages between the AIN switching system, which is also referred to as the service switching point (SSP), and SCP and IS-41 messages [1]. Several SSPs may be connected to an LSTP.

The reference model in Fig. 2 introduces several terms that are explained below. We have tried to keep the terms and discussions fairly general. However, wherever possible equivalent cellular terms from IS-41 or GSM are indicated.

For our purposes, the geographical area served by a PCS system is partitioned into a number of radio port coverage areas (or cells in cellular terms) each of which is served by a radio port (or, equivalently, base station), which communicates with PCS terminals in that cell. A registration area (also known in the cellular world as "location area") is composed of a number of cells. The base stations of all cells in a registration area are connected by wireline links to a mobile switching center (MSC). It is assumed that each registration area is served by a single VLR. The MSC of a registration area is responsible for maintaining and accessing the VLR, and switching between radio ports. The VLR associated with a registration area is responsible for maintaining a subset of the user information contained in the HLR.

The terminal registration process is initiated by terminals whenever they move into a new registration area. The base stations of a registration area periodically broadcast an identifier associated with that area. The terminals periodically compare an identifier they have stored with the identifier of the registration area being broadcast. If the two identifiers differ, the terminal recognizes that it has moved from one registration area to another and will therefore generate a registration message. It also replaces the previous registration area identifier with that of the new area. Movement of a terminal within the same registration area will not generate registration messages. Registration messages may also be generated when terminals are switched on. Similarly, messages are generated to deregister terminals when they are switched off.

PCS services may be provided by different types of commercial service vendors — PSPs. Three different types of PSPs and the different access services that a public network may provide to them are described in [6, 7]. For example, a PSP may have full network capabilities with its own switching, radio management, and radio port capabilities. Certain others may not have switching capabilities, and others may have only radio port capabilities. The model in Fig. 2 assumes full PSP capabilities. The analysis later in this article is based on this model, and modifications may be necessary for other types of PSPs.

It is also quite possible that one or more registration areas may be served by a single PSP. The PSP may have one or more HLR(s) for serving its service area. In such a situation, users who move within the PSP's serving area may generate traffic to the PSP's HLR (not shown in Fig. 2) but not to the network's HLR (shown in Fig. 2). In the interest of keeping the discussions simple, it is assumed that there is one-to-one correspondence between SSPs and MSCs, and also between MSCs, registration areas, and VLRs. One impact of locating the SSP, MSC, and VLR in separate physical sites connected by SS7 signaling links would be to increase the required signaling message volume on the SS7 network. This model assumes that the messages between the SSP and the associated MSC and VLR do not add to signaling load on the public network. Other configurations and assumptions could be studied for which the analysis may need to be suitably modified. However, the underlying analysis techniques will not differ significantly.

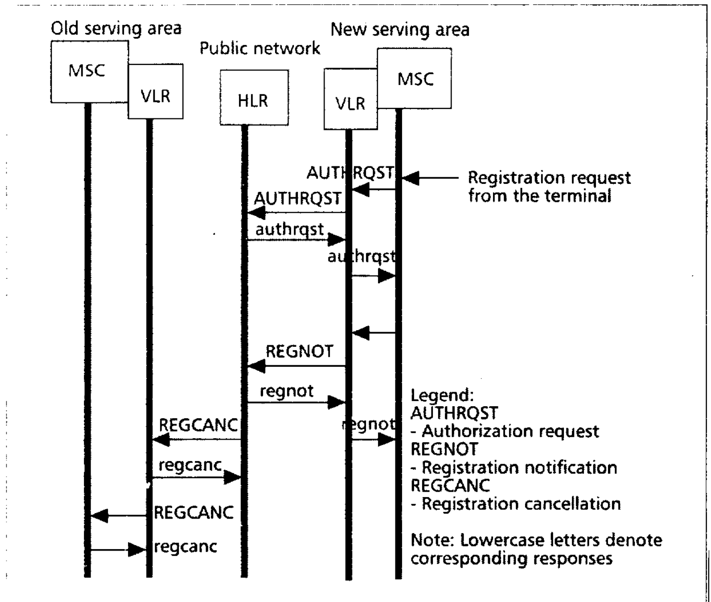
Preliminaries

The message flow for call origination, call delivery, and terminal registration, also sometimes called location registration, based on the IS-41 protocol, will now be described. This protocol is described in detail in [1]; only an outline is provided here.

Terminal/Location Registration

During IS-41 registration, signaling is performed between the following pairs of network elements: the new serving MSC and the associated database (or VLR); the new database (VLR) in the visited area and the HLR in the public network; and the HLR and the VLR in former visited registration area or the old MSC serving area.

Figure 3 shows the signaling message flow diagram for IS-41 registration activity, focusing only on the essential elements of the message flow relating to registration; for details of variations from the basic registration procedure, see [7].



■ **Figure 3.** *Signal flow diagram for registration in IS-41.*

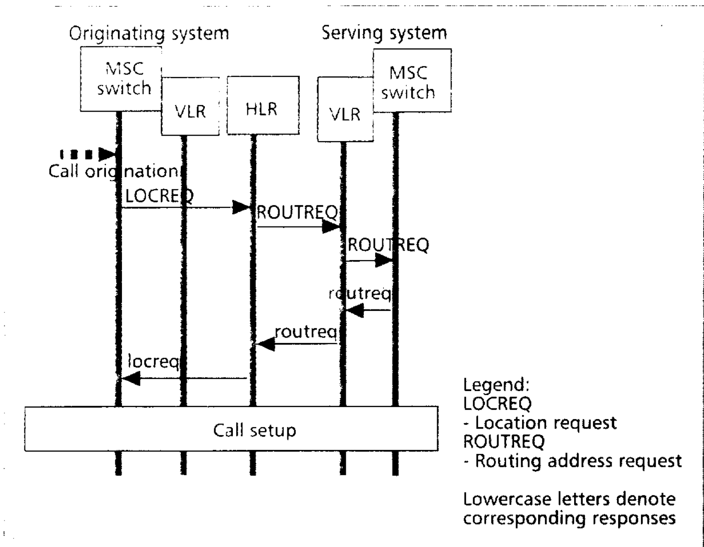
The following steps describe the activities that take place during registration.

- Step 1: Once a terminal enters a new registration area, the terminal sends a registration request to the MSC of that area.
- Step 2: The MSC sends a AUTHRQST (authentication request) message to its VLR to authenticate the terminal, which in turn sends the request to the HLR. The HLR sends its response in the authrqst message.
- Step 3: If the terminal is authenticated, the MSC sends a REGNOT (registration notification) message to its VLR.
- Step 4: The VLR in turn sends a REGNOT message to the HLR serving the terminal. The HLR updates the location entry corresponding to the terminal to point to the new serving MSC/ VLR. The HLR sends a response back to the VLR, which may contain relevant parts of the user's service profile. The VLR stores the service profile in its database and also responds to the serving MSC.
- Step 5: If the user/terminal was registered previously in a different registration area, the HLR sends a REGCANC (Registration Cancellation) message to the previously visited VLR. On receiving this message, the VLR erases all entries for the terminal from the record and sends a REGCANC message to the previously visited MSC, which then erases all entries for the terminal from its memory.

The protocol shows authentication request and registration notification as separate messages. If the two messages can be packaged into one message, the rate of queries to HLR may be cut in half. This does not necessarily mean that the total number of messages is cut in half.

Call Delivery

The signaling message flow diagram for IS-41 call delivery is shown in Fig. 4. The following steps describe the activities that take place during call delivery.



■ **Figure 4.** *Signaling flow diagram for call delivery in IS-41.*

- Step 1: A call origination is detected and the number of the called terminal (e.g., MIN) received by the serving MSC. Observe that the call could have originated from within the public network from a wireline telephone or from a wireless terminal in an MSC/VLR serving area. (If the call originated within the public network, the AIN SSP analyzes the dialed digits and sends a query to the SCP.)

There has been increased interest in GSM in North America, since it is possible that early deployment of PCS will be facilitated by using the communication equipment already available from European manufacturers who use the GSM standard.

- Step 2: The MSC determines the associated HLR serving the called terminal and sends a LOCREQ (location request) message to the HLR.
- Step 3: The HLR determines the serving VLR for that called terminal and sends a ROUTEREQ to the VLR, which forwards it to the MSC currently serving the terminal.
- Step 4: Assuming that the terminal is idle, the serving MSC allocates a temporary identifier — called a temporary local directory number (TLDN) — to the terminal and returns a response to the HLR containing this information. The HLR forwards this information to the originating SSP/MSC in response to its LOCREQ message.
- Step 5: The originating SSP requests call setup to the serving MSC of the called terminal via the SS7 signaling network using the usual call setup protocols.

Similar to the considerations for reducing signaling traffic for location registration, the VLR and HLR functions could be united in a single logical database for a given serving area and collocated; furthermore, the database and switch can be integrated into the same piece of physical equipment or collocated. In this manner, a significant portion of the messages exchanged between the switch, HLR, and VLR, as shown in Fig. 4, will not contribute to signaling traffic.

The Global System for Mobile Communications

In this section the user location strategy proposed in the European GSM standard and its offshoot, DCS1800, is described. Recently, there has been increased interest in GSM in North America, since it is possible that early deployment of PCS will be facilitated by using the communication equipment already available from European manufacturers who use the GSM standard. Since the GSM standard is relatively unfamiliar to North American readers, we first give some background and introduce the various abbreviations. The reader will find additional details in [2]. For an overview on GSM, refer to [3].

The abbreviation GSM originally stood for "Groupe Special Mobile," a committee created within the pan-European standardization body Conference Europėenne des Posts et Telecommunications (CEPT) in 1982. There were numerous national cellular communication systems and standards in Europe at the time, and the aim of GSM was to specify a uniform standard around the newly reserved 900-MHz frequency band with a bandwidth of twice 25 MHz. The phase 1 specifications of this standard were frozen in 1990. In addition, in 1990, at the request of the United Kingdom, the specification of a version of GSM adapted to the 1800-MHz frequency, with a bandwidth of twice 75 MHz, was begun. This variant is referred to as Digital Cellular System 1800 (DCS1800); the abbreviation GSM900 is sometimes used to distinguish between the two variations, with the abbreviation GSM being used to encompass both GSM900 and DCS1800. The motivation for DCS1800 is to provide higher capacities in densely populated urban areas, particularly for PCS. The DCS1800 specifications were frozen in 1991, and by 1992 all major GSM900 European operators had begun operation.

At the end of 1991, activities concerning the post-GSM generation of mobile communications were begun by the standardization committee, using the term universal mobile telecommunications system (UMTS) for this effort. In 1992, the name of the standardization committee was changed from GSM to Special Mobile Group (SMG) to distinguish it from the 900-MHz system itself, and GSM was chosen as the commercial trademark of the European 900-MHz system, where GSM now stands for Global System for Mobile Communications.

The GSM standard has now been widely adopted in Europe and is under consideration in several other non-European regions, including the United Arab Emirates, Hong Kong, and New Zealand. In 1992, Australian operators officially adopted GSM.

Architecture

In this section we describe the GSM architecture, focusing on those aspects which differ from the architecture assumed in the IS-41 standard.

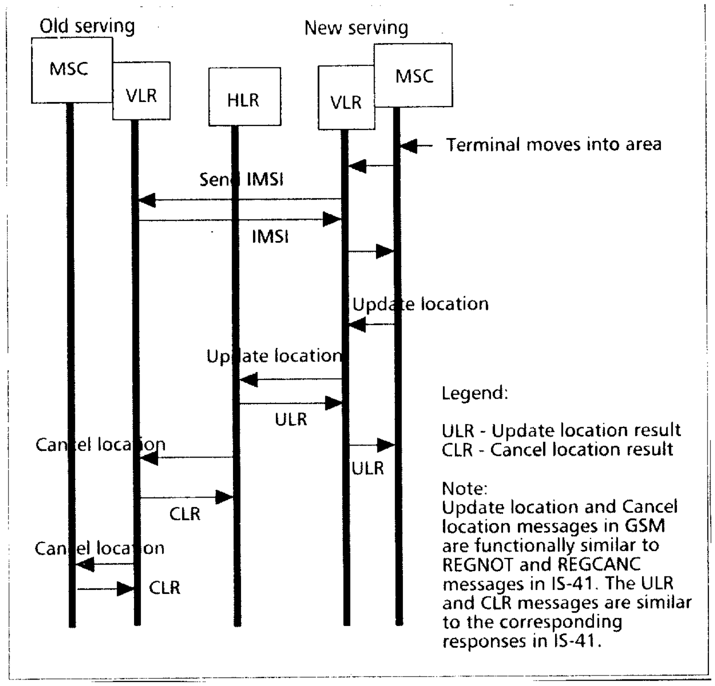
A major goal of the GSM standard was to

enable users to move across national boundaries and still be able to communicate. However, it was considered desirable that the operational networks within each country be operated independently. Each of the operational networks is called a public land mobile network (PLMN), and its commercial coverage area is confined to the borders of one country (although some radio coverage overlap at national boundaries may occur), and each country may have several competing PLMNs.

A GSM customer subscribes to a single PLMN, called the home PLMN, and subscription information includes the services to which the customer subscribes. During normal operation, a user may elect to choose other PLMNs as their service becomes available (as either the user moves or new operators enter the marketplace). The user's terminal, which in GSM is called a mobile station (MS), assists the user in choosing a PLMN in this case, either presenting a list of possible PLMNs to the user using explicit names (e.g., DK Sonofon for the Danish PLMN) or choosing automatically based upon a list of preferred PLMNs stored in the terminal's memory. This PLMN selection process allows users to choose between the services and tariffs of several competing PLMNs. Note that the PLMN selection process differs from the cell selection and handoff process, which a terminal carries out automatically without any user intervention, typically based upon received radio signal strengths; thus, additional intelligence and functionality are required in the terminal.

The geographical area covered by a PLMN is partitioned into MSC serving areas, and a registration area is constrained to be a subset of a single MSC serving area. The PLMN operator has complete freedom to allocate cells to registration areas. Each PLMN has, logically speaking, a single HLR, although this may be implemented as several physically distributed databases, as for IS-41. Each MSC also has a VLR, and a VLR may serve one or several MSCs. As for IS-41, it is interesting to consider how the VLR should be viewed in this context. The VLR can be viewed as simply a database off-loading the query and signaling load on the HLR, and hence logically tightly coupled to the HLR, or as an ancillary processor to the MSC. This distinction is not academic; in the first view, it would be natural to implement a VLR as serving several MSCs, while in the second each VLR would serve one MSC and be physically closely coupled to it. For GSM, the MSC implements most of the signaling protocols, and at present all switch manufacturers implement a combined MSC and VLR, with one VLR per MSC [2].

A GSM MS is split into two parts, one containing the hardware and software for the radio interface, and the other containing subscriber-specific and location information, called the subscriber identity module (SIM). The SIM can be removed from the terminal and is the size of a credit card or smaller. The SIM is assigned a unique identity within the GSM system, called the international mobile subscriber identity (IMSI), which is used by the user location strategy as described below. The SIM also stores authentication information, services lists, PLMN selection lists, and so on, and can itself be protected by a password or PIN.



■ **Figure 5.** *Flow diagram for registration in GSM.*

The SIM can be used to implement a form of large-scale mobility called "SIM-roaming." The GSM specifications standardize the interface between the SIM and the terminal so that a user carrying his or her SIM can move between different terminals and use the SIM to personalize the terminal. This capability is particularly useful for users who move between PLMNs that have different radio interfaces. The user can use the appropriate terminal for each PLMN coverage area while obtaining the personalized facilities specified in his or her SIM. Thus, SIMs address personal mobility. In the European context, the usage of two closely related standards at different frequencies, namely GSM900 and DCS1800, makes this capability an especially important one and facilitates interworking between the two systems.

User Location Strategy

We present a synopsis of the user location strategy in GSM using call flow diagrams similar to those used to describe the strategy in IS-41. Figure 5 illustrates the flow of GSM registration.

In order to describe the registration procedure, it is first useful to clarify the different identifiers used. The SIM of the terminal is assigned a unique identity, called the IMSI, as mentioned above. However, to increase confidentiality and make more efficient use of the radio bandwidth, the IMSI is not normally transmitted over the radio link. Instead, the terminal is assigned a temporary mobile subscriber identity (TMSI) by the VLR when it enters a new registration area. The TMSI is valid only within a given registration area and is shorter than the IMSI. The IMSI and TMSI are identifiers that are internal to the system and assigned to a terminal or SIM, and should not be confused with the user's number that would be dialed by a calling party; the latter is a separate number called the mobile subscriber ISDN number (MSISDN),

Activity	HLR Updates	VLR Updates	HLR Queries	VLR Queries
Mobility-related activities at registration	749	5.85	749	5.85
Mobility-related activities at deregistration		5.85		
Call Origination				8.7
Call Delivery			1116	8.7
Total (per Registration Area)	5.85	11.7	14.57	23.25
Total (Network)	749	1497.6	1865	2976

■ **Table 1.** *IS-41 query and update rates (per second) to HLR and VLR.*

and is similar to the usual telephone number in a fixed network.

We now describe the procedure during registration. The terminal can detect when it has moved into the cell of a new registration area from the system information broadcast by the base station in the new cell. The terminal initiates a registration update request to the new base station; this request includes the identity of the old registration area and the TMSI of the terminal in the old area. The request is forwarded to the MSC, which in turn forwards it to the new VLR. Since the new VLR cannot translate the TMSI to the IMSI of the terminal, it sends a request to the old VLR to send the IMSI of the terminal corresponding to that TMSI. In its response, the old VLR also provides the required authentication information. The new VLR then initiates procedures to authenticate the terminal. If the authentication succeeds, the VLR uses the IMSI to determine the address of the terminal's HLR.

The procedure for delivering calls to mobile users in GSM is very similar to that in IS-41. The sequence of messages between the caller's and called party's MSC/VLRs and the HLR is identical to that shown in the call flow diagrams for IS-41.

The ensuing protocol is then very similar to that in IS-41, except for the following differences. When the new VLR receives the registration affirmation (similar to "regnot" in IS-41) from the HLR, it assigns a new TMSI to the terminal for the new registration area. The HLR also provides the new VLR with all relevant subscriber profile information required for call handling (e.g., call screening lists, etc.) as part of the affirmation message. Thus, in contrast with IS-41, authentication and subscriber profile information are obtained from both the HLR and the old VLR, not just the HLR.

The procedure for delivering calls to mobile users in GSM is very similar to that in IS-41. The sequence of messages between the caller's and called party's MSC/VLRs and the HLR is identical to that shown in the call flow diagrams for IS-41, although the names, contents, and lengths of messages may be different; hence, the details are left out. The interested reader is referred to [2, 3] for further details.

Analysis of Database Traffic Rate for IS-41 and GSM

In the two following sections, the common set of assumptions on which to base comparison of the two strategies is outlined.

The Mobility Model for PCS Users

In the analysis that follows, we assume a simple mobility model for PCS users. The model, described in [8], assumes that PCS users carrying terminals are moving at an average velocity of v, and their direction of movement is uniformly distributed over $[0, 2\pi]$. Assuming that PCS users are uniformly populated with a density of ρ and the registration area boundary is of length L, it has been shown that the rate of registration area crossing, R, is given by

$$R = \frac{\rho v L}{\pi}. \qquad (1)$$

Using equation (1), we can calculate the signaling traffic due to registration, call origination, and delivery.

We now need a set of assumptions so that we may proceed to derive the traffic rate to the databases using the model in Fig. 2.

Additional Assumptions

The following assumptions are made in performing the analysis:

- 128 total registration areas.
- Square registration area size $(7.575\ \text{km})^2$ = 57.4 sq. km, with border length L = 30.3km.
- Average call origination rate = average call termination (delivery) rate = 1.4/hr/terminal.
- Mean density of mobile terminals = ρ = 390/sq. km.
- Total number of mobile terminals = 128 x 57.4 x 390 = 2.87 million.
- Average call origination rate = average call termination (delivery) rate = 1.4/hr/terminal.
- Average speed of a mobile, v = 5.6 km/hr.
- Fluid flow mobility model.

The assumptions regarding the total number of terminals may also be obtained by assuming that a certain public network provider serves 19.15 million users and that 15 percent (or 2.87 million) of the users also subscribe to PCS services from various PSPs.

Note that we have adopted a simplified model that ignores situations in which PCS users may turn their handsets on and off, which will generate additional registration and deregistration traffic. The model also ignores wireline registrations. These activities will increase the total number of queries and updates to HLRs and VLRs.

Analysis of IS-41

Using equation (1) and the parameter values assumed above, we can compute the traffic due to registration. The registration traffic is generated by mobile terminals moving into a new registration area, and this must equal the mobile terminals moving out of the registration area, which is, according to equation (1)

$$R_{\text{reg, VLR}} = \frac{390 \times 30.3 \times 5.6}{3600\pi} = 5.85\,/\,\text{s}$$

This must also be equal to the number of deregistrations (registration cancellations):

$$R_{\text{Dereg,VLR}} = 5.85/\text{s}.$$

The total number of registration messages per second arriving at the HLR will be

$$R_{\text{reg,HLR}} = R_{\text{reg,VLR}} \times \text{total no. of registration areas} = 749/\text{s}.$$

The HLR should therefore be able to handle roughly 750 updates per second. We observe from Fig. 3 that authenticating terminals generates as many queries to VLR and HLR as the respective number of updates generated due to registration notification messages.

The number of queries that the HLR must handle during call origination and delivery can be similarly calculated. Queries to HLR are generated when a call is made to a PCS user. The SSP, that receives the request for a call, generates a location request (LOCREQ) query to the SCP controlling the HLR. The rate of such queries must be equal to the rate of calls made to PCS users. This is calculated as

$$R_{\text{CallDeliv,HLR}} = \text{Call Rate per User} \times \text{Total Number of Users} = \frac{1.4 \times 2.87 \times 10^5}{3600} = 1116/\text{s}.$$

For calls originated from a mobile terminal by PCS users, the switch authenticates the terminal by querying the VLR. The rate of such queries is determined by the rate of calls originating in an SSP serving area, which is also a registration area. This is given by

$$R_{\text{CallOrig,VLR}} = \frac{1116}{128} = 8.7\,/\,\text{s}.$$

This is also the number of queries needed to authenticate terminals of PCS users to which calls are delivered:

$$R_{\text{CallDeliv,VLR}} = 8.7/\text{s}.$$

Table 1 summarizes the calculations.

Analysis of GSM

Calculations for query and update rates for GSM may be performed in the same manner as for IS-41, and they are summarized in Table 2. The difference between this table and Table 1 is that in GSM, the new serving VLR does not query the HLR separately in order to authenticate the terminal during registration; hence, there are no HLR queries during registration. Instead, the entry (749 queries) under HLR queries in Table 1, corresponding to mobility-related authentication activity at registration, gets equally divided between the 128 VLRs.

Observe that with either protocol the total database traffic rates are conserved, where the total database traffic for the entire network is given by the sum of all the entries in the last row, Total (network):

HLR updates + VLR updates + HLR queries + VLR queries

From Table 1 and Table 2, we see that this quantity equals 7087.

This conclusion is independent of variations in the assumptions mentioned previously. For example, if the PCS penetration (the percentage of the total users subscribing to PCS services) were to increase from 15 to 30 percent, all the entries in the two tables will double; hence, the total database traffic generated by the two protocols will still be equal.

Activity	HLR Updates	VLR Updates	HLR Queries	VLR Queries
Mobility-related activities at registration	749	5.85		11.7
Mobility-related activities at deregistration		5.85		
Call Origination				8.7
Call Delivery			1116	8.7
Total (per Registration Area)	5.85	11.7	8.72	29.1
Total (Network)	749	1497.6	1116	3724.8

■ **Table 2.** *GSM query and update rates (per second) to HLR and VLR.*

Previous studies indicate that the signaling traffic and database queries associated with PCS due to user mobility are likely to grow to levels well in excess of that associated with the conventional call.

Reducing Signaling During Call Delivery

In the above two subsections, we provided a simplified analysis of some scenarios associated with user location strategies and the associated database queries and updates required. Previous studies indicate that the signaling traffic and database queries associated with PCS due to user mobility are likely to grow to levels well in excess of that associated with the conventional call [9, 10]. It is therefore desirable to study modifications to the two protocols that would result in reduced signaling and database traffic. We now provide some suggestions.

For both GSM and IS-41, the delivery of a call to a mobile user involves four messages: from the caller's VLR to the called party's HLR, from the HLR to the called party's VLR, from the called party's VLR to the HLR, and from the HLR to the caller's VLR. The last two involve the HLR, which relays the routing information provided by the called party's VLR to the caller's VLR. An obvious modification to the protocol would be to have the called VLR directly send the routing information to the calling VLR. This would substantially reduce the total load on the HLR and on signaling network links. Such a modification to the protocol may not be easy, of course, due to administrative, billing, legal, or security concerns. This would also violate the query/response model adopted in IS-41, requiring further analysis.

A related question is whether the routing information obtained from the called party's VLR could instead be stored in the HLR. This routing information could be provided to the HLR, for example, whenever a terminal registers in a new registration area. If this were possible, two of the four messages involved in call delivery could be eliminated. This point was discussed at length by the GSM standards body, who arrived at the present strategy. The reason for this decision was to reduce the number of temporary routing numbers allocated by VLRs to terminals in their registration areas. If a temporary routing number (TLDN in IS-41 or MSRN in GSM) is allocated to a terminal for the whole duration of its stay in a registration area, the quantity of numbers required is much greater than if a number is assigned on a per-call basis.

If a temporary routing number is allocated to a terminal for the whole duration of its stay in a registration area, the quantity of numbers required is much greater than if a number is assigned on a per-call basis.

Other strategies may be employed to reduce signaling and database traffic; we discuss these in [12, 13]. One method is to use intelligent paging by exploiting users' mobility behavior stored in user profiles [11]. A discussion of these techniques is beyond the scope of this article.

Conclusions

In this article we have provided an overview of data management for user location in PCS, such as aspects of personal and terminal mobility, registration, deregistration, and call delivery. A tutorial is provided on the two most common strategies for locating users in PCS: the North American Interim Standard, IS-41, and the Pan-European Standard, GSM. A simplified analysis is provided to show the reader the extent of database and signaling traffic likely to be generated by PCS services. Finally, some suggestions that are likely to result in reduced traffic were provided. Details of more efficient location strategies appear in [12], and a survey of several approaches to user location strategies is presented in [13].

Acknowledgments

Thanks are due to N. N. Crystal, D. F. Daly, G. E. Feldkamp, S. Levenson, P. J. Louis, D. R. Lukacs, J. F. Rizzo, and R. S. Wolff for their helpful comments that improved the quality and readability of this article. We also wish to thank D. Lamerdin of Pacific Bell and M. Horrer of the TINA Consortium for their help in obtaining information about GSM.

References

[1] EIA/TIA IS-41.3 (Revision B), "Cellular Radiotelecommunications Intersystem Operations: Automatic Roaming," July 1991.
[2] M. Mouly and M.-B. Pautet, "The GSM System for Mobile Communications," Palaiseau, France, 1992.
[3] E. Lycksel, "GSM System Overview," Swedish Telecom. Admin., Jan. 1991.
[4] "Advanced Intelligent Network Release 1 Network and Operations Plan," issue 1, Bellcore Spec. Rep. SR-NPL-001623, June 1990.
[5] R. K. Berman and J. H. Brewster, "Perspectives on the AIN Architecture," *IEEE Commun. Mag.*, vol. 1, no. 2, Feb. 1992, pp. 27-3.
[6] "Personal Communications Services (PCS) Network Access Service Alternatives," issue 1, Bellcore Spec. Rep. SR-INS-002245, Apr. 1992.
[7] "Network and Operations Plan for Access Services to Personal Communications Services (PCS) Providers," Bellcore Spec. Rep. SR-TSV-002402, Aug. 1992.
[8] R. Thomas, H. Gilbert, and G. Mazziotto, "Influence of the Mobile Station on the Performance of a Radio Mobile Cellular Network," *Proc. 3rd Nordic Sem.*, paper 9.4, Copenhagen, Denmark, Sept. 1988.
[9] K. Meiller-Hellstern and E. Alonso, "The Use of SS7 and GSM to Support High Density Personal Communications," *Proc. ICC '92*.
[10] C. N. Lo, R. S. Wolff, and R. C. Bernhardt, "Expected Network Database Transaction Volume to Support Personal Communications Services," 1st International. Conf. Universal Personal Commun., Dallas, Texas, Sept. 1992.
[11] S. Tabbane, "Evaluation of an Alternative Location Strategy for Future High Density Wireless Communications Systems," Tech. Rep. WINALAB-TR-51, Rutgers University, Jan. 1993.
[12] R. Jain, J. Lin, C.N. Lo, and S. Mohan, "A Caching Strategy to Reduce Network Impacts of PCS," submitted for publication, Sept. 1993.
[13] R. Jain, "A Classification and Survey of Strategies for Locating Users in Personal Communications Services Systems," submitted for publication, Aug. 1993.

Biographies

SESHADRI MOHAN [M '79] received a Ph.D. in electrical and computer engineering from McMaster University, Ontario,in 1980. Since then, he has worked for Bell Laboratories, Holmdel, New Jersey, and has taught at Wayne State and Clarkson Universities, where he conducted research in the area of coding algorithms, and computer communications and multiple access protocols. He joined Bellcore, Morristown, New Jersey, in 1990, where he is currently a member of the Network Architecture and Analysis Research Laboratory. His current research inte rests include designing efficient data management and location strategies for the support of personal communications applications and investigating the applicability of distributed computing and communications architectures for nomadic personal communications and information networking. He has coauthored the text, *Source and Channel Coding: An Algorithmic Approach* (Kluwer Academic Publishers). He is a member of the ACM.

RAVI JAIN [M '89] received a Ph.D in computer science from the University of Texas at Austin in 1992, having previously worked at Syntrex Inc., SES Inc., and the Schlumberger Laboratory for Computer Science on developing communications and systems software, performance modeling, and parallel programming. In 1992 he joined Bellcore, Morristown, New Jersey, where his research interests include design and analysis of algorithms and techniques for efficient resource management, focusing on reducing network impacts of supporting PCS, and Intelligent Vehicle-Highway Systems (IVHS) applications. He has authored several publications in the area of communications and parallel computing and has co-chaired conference sessions, workshops, and panels. He is a member of ACM, ORSA, and CPSR.