Abstract

This article provides an overview of the means and techniques used for subscriber location management in present cellular mobile networks. The overhead due to location management techniques used in the present cellular systems such as GSM, in particular in high traffic spots, has already been a source of concern and is expected to augment dramatically in future wireless systems. Several proposals have been made in the past addressing the problem associated with the cost of location management. The author provides an overview of these contributions; in particular, he classifies the techniques, mainly into non-memory-based and memory-based approaches. Subsequently, a simple evaluation of the current location updating procedure used in GSM is presented in the context of a PCS framework to illustrate the load placed on the radio channels and on the MSC/VLR processing.

Location Management Methods for Third-Generation Mobile Systems

Sami Tabbane, ESPTT

The trends in telecommunications are proceeding with a strong tendency toward increasing need for mobility in the access links within the network. Examples are:

- Residential line access with the proliferation of cordless phones (their penetration rate has passed that of fixed phones in the United States, for instance)
- Business lines with wireless private branch exchange (PBX) access for voice services and wireless LANs for computer-oriented data communications (IEEE 802.11 and HIPERLAN specifications have just completed their standardization)
- Paging systems, which provide a low-cost, always-in-touch service (delivering voice, numeric and alphanumeric messages, and electronic mail) to a still growing part of the population, especially young people
- Cellular systems, which allow telecommunication accesses over wide areas

Furthermore, the growth of mobility aspects in telecommunications networks can be seen at three different levels.

- First, at a *spatial level*, within a few decades users who have been able to roam with a mobile terminal, beginning locally and regionally (with conventional and trunk private mobile radio systems since the '70s), nationally since the early '70s (with cellular systems), internationally since the beginning of the '90s (with international roaming, GSM for instance), and soon globally (satellite systems; global mobile personal communications systems such as IRIDIUM, ICO, or GLOBALSTAR).
- Second, from the *penetration rate* of mobile radio access lines. Indeed, most optimistic surveys have nearly always underestimated the fantastic growth of cellular systems. For instance, Nordic countries like Sweden have penetration rates of more than 30 percent [1].
- Third, the *traffic generated* by each wireless user is constantly growing: on one hand, tetherless (e.g., cellular) subscribers use their mobile terminals more often than they used to (due to changing habits, cheaper call rates, wider mobility, ease of call); on the other hand, the arrival of more capacity-greedy services (Internet accesses, multimedia services, etc.) pushes the need for higher capacities per call (the high-speed circuit-switched data capability defined for GSM Phase 2+ [2] or recent wireless ATM studies [3]). GSM stands for Global System for Mobile Communications.

From all these considerations, it is easy to predict that the generalized mobility features will have serious impacts on future telecommunications networks. Mobility can be categorized into two areas:

- Radio mobility, which mainly consists of the handover process
- Network mobility, which mainly consists of location management (location updating and paging)

Handover processes are essentially based on radio aspects, and the main difficulties in improving handover procedures' performance come from unpredictable and highly fluctuating radio channel behavior [4].

Location management schemes are essentially based on users' mobility and incoming call rate characteristics. The network mobility process has to face strong antagonism between its two basic procedures: *location* and *paging*. The location procedure allows the system to keep the user's location knowledge, more or less accurately, in order to be able to find him, in case of an incoming call, for example. Location registration is also used to bring the user's service profile near its location and allows the network to provide him rapidly with his services (e.g., the visitor location registration, VLR, functions in GSM). The paging process achieved by the system consists of sending paging messages in all cells where the mobile terminal could be located. Therefore, if the location cost is high (and thus the user location knowledge is accurate), the paging cost will be low (paging messages will only be transmitted over a small area). If the location cost is low (and thus the user location knowledge is fuzzy), the paging cost will be high (paging messages will have to be transmitted over a wide area).

0163-6804/97/$10.00 © 1997 IEEE

In first-generation cellular mobile systems, traffic was highly unbalanced. Less than one third of calls were incoming calls; the remaining were outgoing. Therefore, the paging process (required only for incoming calls) was a rather rare event, and thus had rather little impact on the mobility management traffic. The location updating procedure also had little impact due to the large cells. Besides the trends stated earlier, current systems experiment with a balance between incoming and outgoing call rates. The paging process is therefore more important, and thus the location management is also more important. For instance, recent statistics from GSM operators show that, in the Paris dense urban environment, the location updating rate can be 10 times that of the call rate at peak traffic hours.

In the remainder of this article, we introduce the main location procedures used in present systems. The GSM case will be examined in particular. Then, we propose a classification of location management methods defined in the literature for third-generation mobile systems.

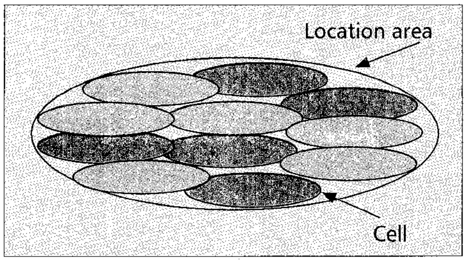


Figure 1. *Location area structure.*

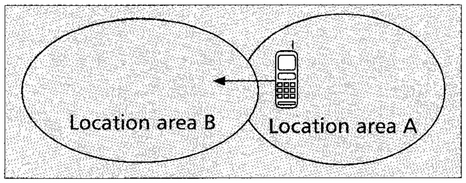


Figure 2. *Location updating on LA crossing.*

Present Location Management Methods

Level 0: No location management

In early wide-area wireless systems (not yet cellular), human operators had to process the calls and the users' location was not managed by the system. A user was able to generate a call through any base station (BS), and paging messages addressed to the called mobiles were transmitted through all BSs. The main characteristics of these systems were very large cells, and low user population and call rates.

Small-capacity cellular systems (with a few tens of BSs serving a few thousand users) may also not use a location management method, even when the standard allows it. If subscriber number and calling rates do not require it, the location management method is not activated; resource consumption for finding users is not so important that its reduction is mandatory.

This *level 0* method is therefore as simple as could be: no location management is realized; the system does not track the mobiles. A search for a called user must therefore be done over the complete radio coverage area and within a limited time. This method is usually referred to as the *flooding* algorithm [5]. It is used in paging systems because of the lack of an uplink channel allowing a mobile to inform the network of its whereabouts. It is also used in small private mobile networks because of their small coverage areas and user populations.

The main advantage of not locating the mobile terminals is obviously simplicity; in particular, there is no need to implement special databases. Unfortunately, it does not fit large networks dealing with high numbers of users and high incoming call rates.

Level 1: Manual Registration

This method requires the user to locate himself by achieving a special procedure if he wishes to receive his incoming calls. From the network side, this method is relatively simple to manage because it just requires the management of an indicator which stores the current location of the user. The mobile is also relatively simple; its task is just limited to scanning the channels to detect paging messages.

This method is currently used in telepoint cordless systems (such as the French Bi-Bop CT2, Cordless Telephone 2). The user has to register itself each time he/she moves to a new island of CT2 beacons. To page a user, the network first transmits messages through the beacon with which he/she registered and, if the mobile does not answer, extends the paging to neighboring beacons.

The main drawback of this method is the constraint for a user to register each time he moves. Nevertheless, this low ergonomy can be balanced by the low equipment and management costs of the network, which allow the operator to offer users attractive fees.

Level 2: Use of Location Areas for Automatic Location Management

Presently, the location method most widely implemented in first- and second-generation cellular systems (NMT, GSM, IS-95, etc.) makes use of location areas (LAs) (Fig. 1). In these wide-area radio networks, location management is done automatically.

Location areas allow the system to track the mobiles during their roaming in the network(s): subscriber location is known if the system knows the LA in which the subscriber is located. When the system must establish a communication with the mobile (to route an incoming call, typically), the paging only occurs in the current user LA. Thus, resource consumption is limited to this LA; paging messages are only transmitted in the cells of this particular LA.

Implementing LA-based methods requires the use of databases. Generally, a home database and several visitor databases are included in the network architecture. We discuss this point later in the GSM example.

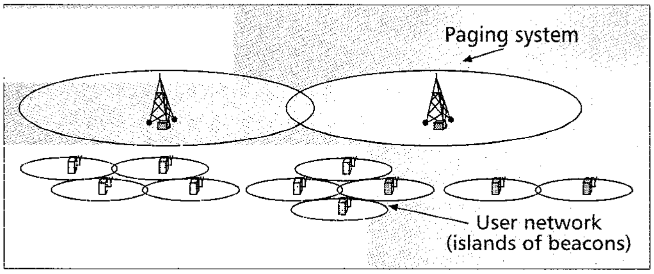
Several location updating methods can be implemented based on LA structuring.

Periodic Location Updating — This method is simplest because it just requires the mobile to periodically transmit its identity to the network. Its drawback is its resource consumption, which is user-ndependent and can be unnecessary if the user does not move from an LA for several hours. Generally, this method is combined with the next one.

Location Updating on LA Crossing — This method (Fig. 2) first requires each BS to periodically broadcast the identity of its LA. Second, the mobile is required to permanently listen to network broadcast information (on the broadcast channel) and to store the current LA identity. If the received LA number differs from the stored one, a location update (LU) procedure is automatically triggered by the mobile.

The advantage of this method is that it only requires LUs when the mobile actually moves. A highly mobile user will generate a lot of LUs; a low mobility user will only trigger a few.

A hybrid method which combines the two previous ones can also be implemented. The mobile generates its LUs each time it detects an LA crossing. Nevertheless, if no communication (related to an LU or a call) has occurred between the



■ **Figure 3.** *Location management with separate paging and transport networks.*

mobile (in idle mode, i.e., powered on but not communicating) and the network for a fixed period (e.g., three hours), the mobile generates an LU (i.e., a periodic LU). This periodic LU typically allows the system to recover user location data in case of a database failure.

Separate User and Paging Networks

The search for the user can be achieved through a network separate from that which carries the calls: a signaling network (*paging network*) is physically distinct from the users' data transport network (*user network*) (Fig. 3). This is the case for some wireless systems offering a telepoint cordless service combined with a paging service. This combination allows users to be reachable even outside beacon islands, which are generally limited to high-density areas.

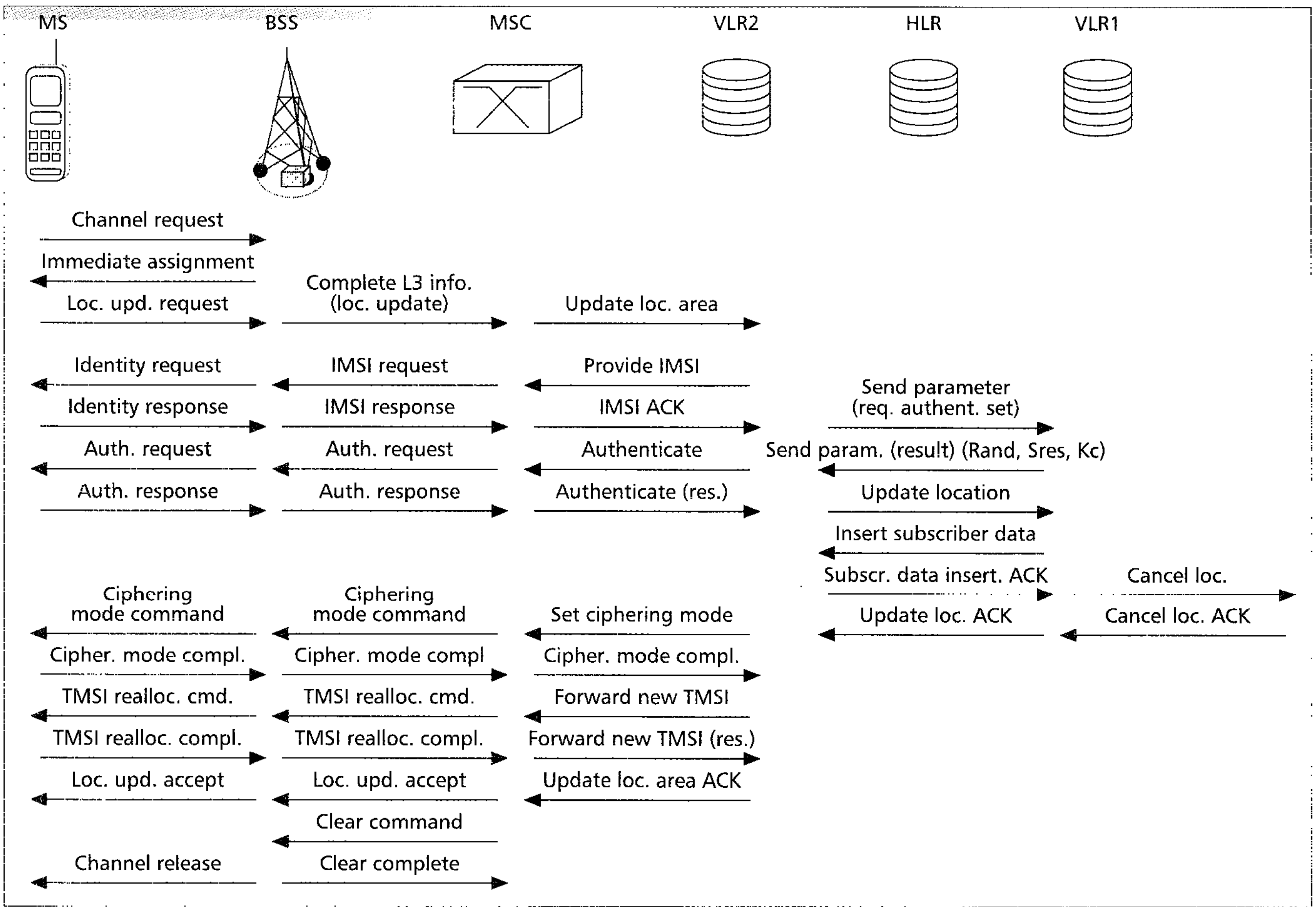
GSM Example

The GSM standard [6] defines a database structure based on:
- An HLR (home location register) where all subscriber related information is stored (access rights, user location, etc.). Security parameters and algorithms are managed by the authentication center (AuC) which is often considered part of the HLR.
- Several VLRs. Each VLR stores part of the data regarding the users located in its related LAs.

The location management method defined in GSM combines the periodic LU method and the LU on LA border crossing. The VLR stores the LA identifier, and the HLR stores the VLR identifier.

This consists of three main types of LU procedures: the intra-VLR LU, the inter-VLR LU using TMSI (temporary mobile subscriber identity), and the inter-VLR LU using IMSI (international mobile subscriber identity). A fourth one, the *IMSI Attach* procedure, is triggered when the mobile is powered on in the LA where it was powered off.

In the following, we present the most complete LU, which is inter-VLR using IMSI. Figure 4 depicts the signaling exchange during the procedure. This procedure mainly consists of the following steps:
- A signaling channel is allocated to the MS, and an LU is requested.
- The MS provides the network with its IMSI, which allows the new VLR (VLR_2) to load the authentication data from the HLR/AuC, mainly the triplets (*Rand, SRES, Kc*) for the authentication and ciphering procedures.
- The VLR is then able to authenticate the MS; if this step succeeds, it updates the location at the HLR. The HLR informs the old HLR (VLR_1) to remove the user's data stored in VLR_1.



■ **Figure 4.** *Inter-VLR LU with IMSI.*

- Ciphering may be required if available.
- A new TMSI is allocated to the MS, and, after acknowledgment of its LU request (first message sent by the MS), the channel is finally released.

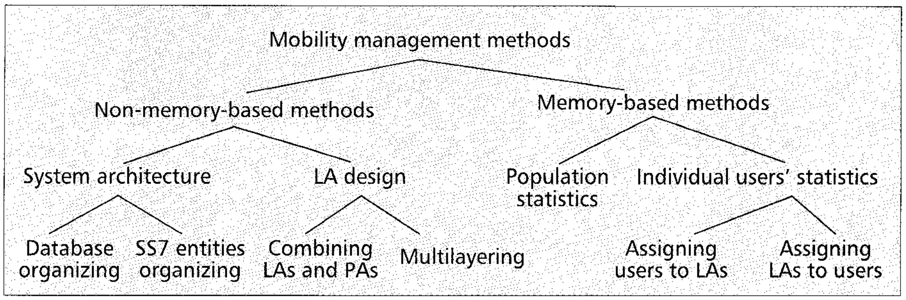


Figure 5. *Mobility management methods classification.*

LA Partitioning

In order to minimize location management cost (LU + paging traffic and processing), the LA must be designed carefully. Two types of methods can be used to design optimum LAs: analytic and heuristic.

Analytic approaches are based on assumptions of homogeneous cell shape, LA structure, and users' movements. One interesting problem is to determine a subscriber mobility model that can capture as much as possible the real subscribers' movements. Common approaches for modeling human movements are mentioned in [7], including:
- The Markovian model, also known as the random-walk model, is a model which describes individual movement behavior.
- The fluid model considers traffic flow as the flow of a fluid, modeling macroscopic movement behavior.

One of the first studies to address the second method has been that of Morales [8]. Let us mention that mobility traces, which record actual movement behavior for certain segments of the population, are certainly more realistic than mobility models.

Generally, authors use a simple model based on fluid flow assumptions. This model is applied, for instance, in [9] in order to determine the size of an LA: a simple method is defined which takes into account R, the cell radius, V, the mean mobile velocity, $LU_{\cos t}$, the cost of LUs (in terms of the number of LU messages required to update the location of a mobile), $PAG_{\cos t}$, the cost of paging (in terms of the number of paging messages required to find a mobile), and N (to be determined), the number of cells per LA. The calculations lead to the following formula, which gives the N optimum value:

$$N_{\text{opt}} = \sqrt{\frac{v.PAG_{\cos t}}{\pi.R.LU_{\cos t}}}$$

In real cases, cell shapes and patterns are not so regular. If less restrictive assumptions are considered, the LA partitioning problem is much more complex and appears as an NP-complete combinatorial problem. Only empirical methods can thus be used to approach the optimal solution.

In [10], for example, the method proposed proceeds in an iterative way by reducing a mobility cost function and starts by a configuration of LAs using an analytical method. Another approach, proposed in [11], makes use of genetic algorithms. Genetic algorithms are used to efficiently group the cells under a mobility cost function constraint. They use several processes, such as elitism, linear normalization of chromosome, and edge-base crossover. Other empirical methods can be used, such as simulated annealing [12] — which is currently used by some GSM operators — single-move heuristic, and steepest descent optimization.

Limits of Present Location Management Methods

The LA-based location management methods are the most adapted and widely used in current cellular (e.g. GSM, IS-54 and IS-95, etc.), in trunk systems such as trans-European trunk radio (TETRA), in cordless systems like Digital European Cordless Telecommunications (DECT), Personal Access Communications System (PACS), Personal Handyphone System (PHS), and so on. Nevertheless, the traffic and processing generated may lead to congestion problems in high-density systems (see the fourth section). One of the main concerns of system designers is therefore to define methods allowing the system to reduce the overhead traffic as much as possible.

In the next section we examine several location management methods proposed within these last years which attempt to reduce the overhead traffic.

Location Management Methods for Third-Generation Systems

We classify the location management methods into two major groups (Fig. 5). In the first, we include all methods based on algorithms and network architecture, mainly on the processing capabilities of the system. The second group gathers the methods based on learning processes which require the collection of statistics on users' mobility behavior, for instance. This second type of method emphasizes the information capabilities of the network.

Memoryless Methods

Database Architecture — LA partitioning, and thus mobility management cost, partly relies on the system architecture (e.g., database locations). Thus, signaling traffic can be reduced by designing an appropriate database organization. In [13–16], various database architectures are proposed with this aim.

Figure 6 presents an architecture where a unique centralized database is used. This is well suited to small and medium networks, typically based on a star topology.

The second one (Fig. 7) is a distributed database architecture which uses several independent databases according to geographical proximity or service providers. It is best suited to large networks including subnetworks managed by different operators and service providers. The GSM worldwide network, defined as the network made up of all interconnected GSM networks in the world, can be such an example of a large network. The main drawbacks of this architecture are clearly the cost of database system acquisition, implementation, and management.

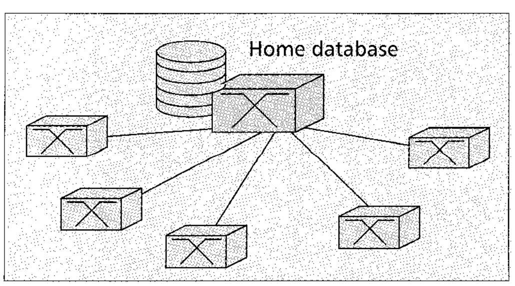
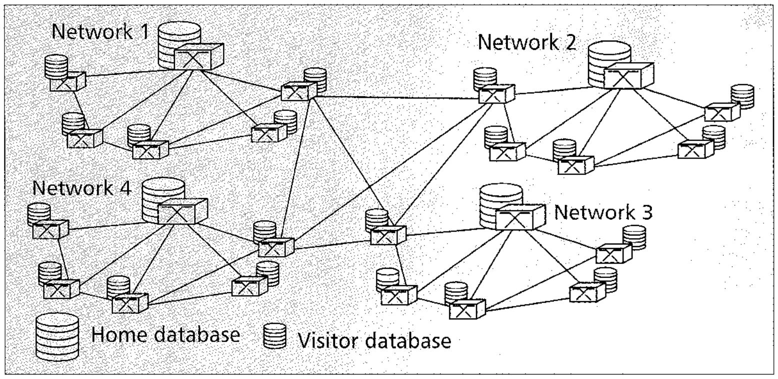


Figure 6. *Centralized database architecture.*

The third case (Fig. 8) is the hybrid database architecture that combines the centralized and distributed database architectures. In this case, a central database (HLR-like) is used to store all user information. Other smaller databases (VLR-like) are distributed all over the network. These VLR databases



■ **Figure 7.** *Distributed database architecture.*

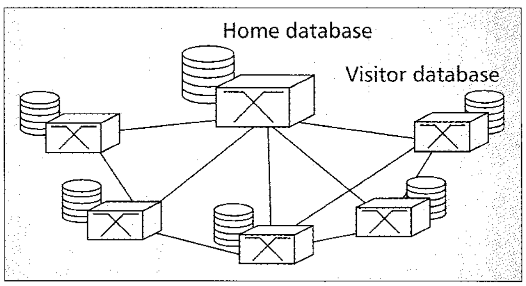
store portions of HLR user records. A single GSM network is an example of such an architecture.

Optimizing Fixed Network Architecture — In second-generation cellular networks and third-generation systems, signaling is managed by the intelligent network (IN) [17]. Appropriately organizing mobility functions and entities can help reduce the signaling burden at the network side. The main advantage of these propositions is that they allow us to reduce the network mobility costs independent of the radio interface and LA organization.

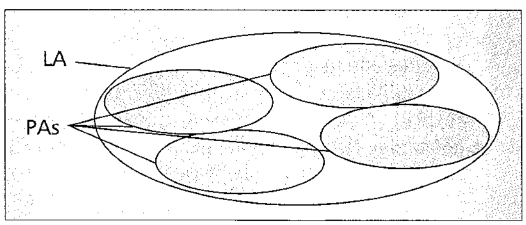
For example, in [18] it is proposed to use different degrees of decentralization of the control functions. Thus, using adapted signaling network nodes, interconnection allows mobility costs to be reduced. The proposition made in [19] is based on a metropolitan area network (MAN) architecture. The higher-layer VLR (VLR gateway) is used with a MAN such that most signaling traffic for LUs in a city area can be handled inside the MAN. Outside the MAN, the traffic burden can consequently be reduced.

Combining Location Areas and Paging Areas (Fig. 9) — In current systems, an LA is defined as both an area in which to locate a user and an area in which to page him. LA size optimization is therefore achieved by taking into account two antagonistic procedures, locating and paging. Based on this observation, several proposals have defined location management procedures which make use of LAs and paging areas (PAs) of different sizes [20]. One method often considered consists of splitting an LA into several PAs.

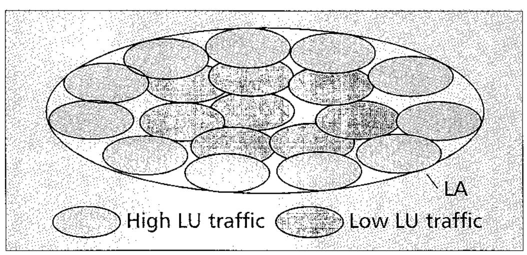
An MS registers only once, that is, when it enters the LA. It does not register when moving between the different PAs of the same LA. For an incoming call, paging messages will be broadcast in the PAs according to a sequence determined by different strategies. For example, the first PA of the sequence can be the one where the MS was last detected by the network. The drawback of this method is the possible delay increase due to large LAs.



■ **Figure 8.** *Hybrid database architecture.*



■ **Figure 9.** *LAs and PAs combined.*



■ **Figure 10.** *Concentrated LU traffic in the LA border cells.*

A method based on the same partitioning principle is investigated in [21], in which each cell of an LA can be considered one individual PA. The interesting result obtained shows that partitioning the LA into several PAs paged sequentially allows first a reduction of the expected number of messages, and second a significant decrease of the paging delay at high loads.

Multilayer LAs — In present location management methods, LU traffic is mainly concentrated in the cells of the LA border (Fig. 10).

Based on this observation and to overcome this problem, Okasaka has introduced the multilayer concept [22]. In his method, each MS is assigned to a given group, and each group is assigned one or several layers of LAs.

According to Fig. 11, it is clear that group 1 and group 2 MSs will not generate LUs in the same cells, thus allowing the LU traffic load to be distributed over all the cells. Nevertheless, this location updating method, although it may help reduce channel congestion, does not help reduce the overall signaling load generated by LUs.

A Procedure for Reducing the Exchanges of Signaling Messages — The reverse virtual call (RVC) setup is a new scheme for delivering mobile terminal calls [23]. It allows, under the constraint that the LA is not smaller than the VLR area, a reduction in the number of signaling messages exchanged between the called and calling databases and switches. The call setup delay is shown to always be reduced by about 50 percent when using the RCV scheme.

Memory-Based Methods

The design of memory-based location management methods has been motivated by the fact that systems do a lot of repetitive actions, which can be avoided if predicted. This is particularly the case for LUs. Indeed, present cellular systems achieve every day, at the same peak hours, almost the same LU processing. Systems act as memoryless processes. Short-term and long-term memory processes can help the system avoid these repetitive actions. Some methods have thus been proposed that are based on user and system behavior observation and statistics.

Short-Term Observation for Dynamic LA and PA Size Assignment/Adjustment — In current systems, the size of LAs is optimized according to mean parameter values, which in practical situations vary over a wide range during the day and from one user to another.

Based on this observation, it is proposed to manage user location by defining multilevel LAs in a hierarchical cellular structure [24]. At each level the LA size is different, and a cell belongs to different LAs of different sizes. According to

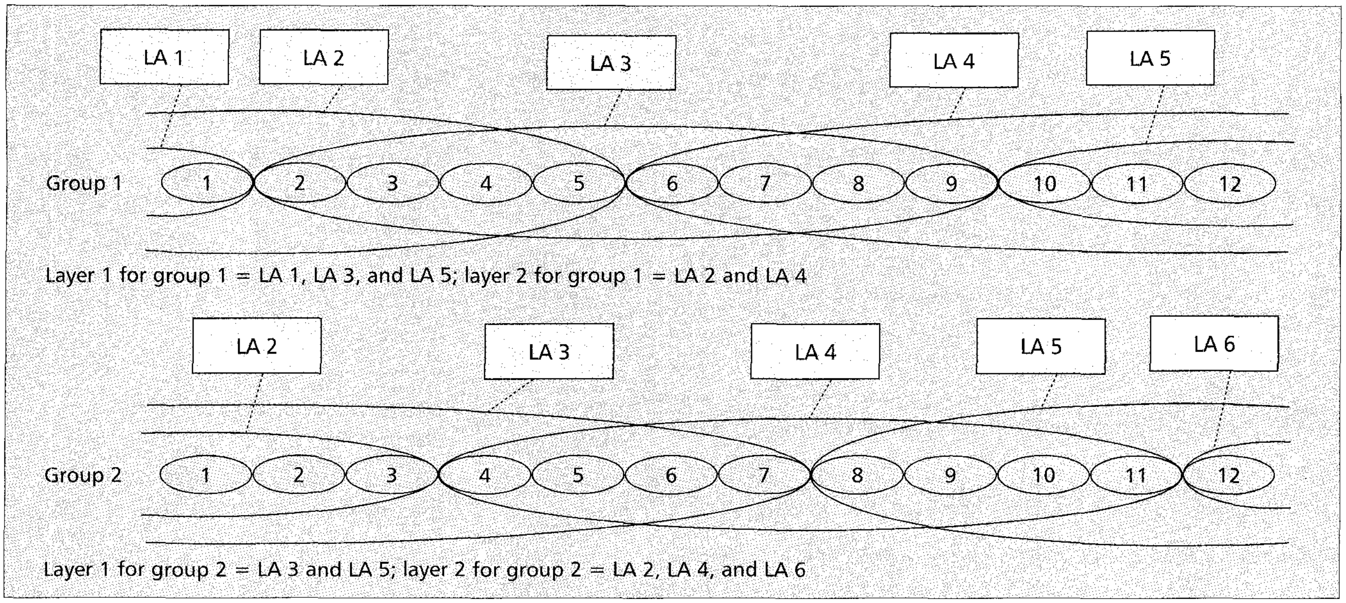


Figure 11. *LA overlapping example with Okasaka's method.*

past and present MS mobility behavior, the scheme dynamically changes the hierarchical level of the LA to which the MS registers. LU savings can thus be obtained.

A variant of this strategy evaluated in [25] consists in requiring from mobiles to register in the cells where they are camped on. Registrations involve a periodic timer which value has to be optimized. Thus, rather than paging a mobile in all the cells of a LA, the mobile will be paged only in the cells visited during the last period: these are the cells where the mobile camped on during its traversal of the LA.

In Fig. 12, high incoming call rate and low-mobility users are directed to small LAs, medium-mobility users are directed to medium-sized LAs, and high-velocity and low incoming call rate users are directed to large LAs.

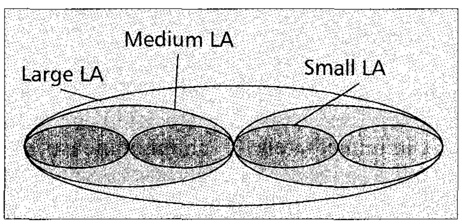


Figure 12. *Multilevel LAs.*

An opposite approach considers that instead of defining LA sizes a priori, these can be adjusted dynamically for every user according to his/her incoming call rate (a) and LU rate (u_k), for instance. In [26–27], a simple method for optimizing LA size (say k cells) individually for each user is described and evaluated. In [26], a mobility cost function denoted $C(k, a, u_k)$ is minimized so that k is permanently adjusted. Each user is therefore related to a unique LA for which size k is adjusted according to his/her particular mobility and incoming call rate characteristics.

Adapting the LA size to each user's parameters values may be difficult to manage in practical situations. This led, in [28], to the definition of a method where the LAs sizes are dynamically adjusted for the whole population, not per user as in the two previous methods. Statistical information about users and mobility in the network is collected in databases and computed. Network characteristics in function of time, place, density, and so on are thus evaluated. Results of this computation allow the network to dynamically (daily, weekly, monthly, yearly, etc.) adjust the LAs' sizes. For instance, during the day, when call rates are high, it is preferable to deal with small LAs. Conversely, at night the call rate is much lower, and therefore larger LAs are better.

Individual User Patterns — Observing that users show repetitive mobility patterns, the *alternative strategy* (AS) is defined [29–30]; its main goal is to reduce the traffic related to mobility management — and thus reduce the LUs — by taking advantage of users' highly predictable patterns. In AS, the system handles a *profile* recording the most probable mobility patterns of each user. The profile of the user can be provided and updated manually by the subscriber himself or determined automatically by monitoring the subscriber's movements over a period of time. For an individual user, each period of time $[t_i, t_j[$ corresponds to a set of location areas, k, in the following way: (a_f, α_f) with 1 ″ f ″ k where a_f is the location area in which the user can be located, αf is the probability that the user is located in a_f, and

$$\alpha_1 > \alpha_2 > \ldots > \alpha_k \left(\sum_{j=1}^{k} \alpha_i < 1 \right).$$

When the user receives a call the system pages him sequentially over the LA a_is until getting an acknowledgment from the mobile. Note that when the delay is important, a parallel paging can be processed.

When the subscriber moves away from the recorded zone $\{a_1, \ldots, a_k\}$ the terminal processes a voluntary registration by pointing out its new LA to the network.

The main savings allowed by this method are due to the non-triggered LUs when the user keeps moving inside his profile LAs. So, the more predictable the user's mobility, the lower the mobility management cost.

A variant of this method, called the Two Location Algorithm (TLA), is proposed and studied in [31]. In this strategy, a mobile stores the two most recently visited LA addresses. The same is done at the HLR level. Obviously, the main advantage of this method relies on the reduction of LUs when a mobile goes back and forth between two LAs.

Predicting Short-Term Movements of the Subscriber — The method proposed in [32] uses a process which predicts the movements of the MS according to its direction of movement, velocity, and so on. Processing and prediction are made at both the MS and the HLR. When actual movements of the MS do not fit with those predicted, a registration is triggered by the mobile to inform the network of its actual location.

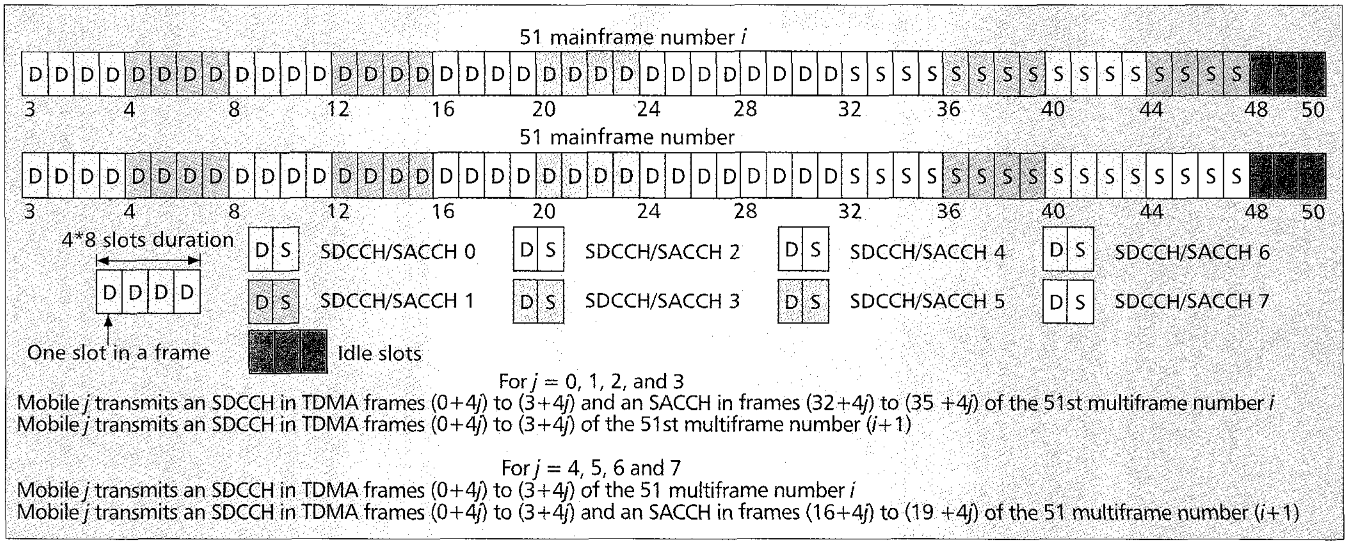


Figure 13. *Channel multiplexing for SDCCH/SACCH transmission.*

Otherwise, no exchange is required, which allows savings in LU processing and signaling.

Mobility Statistics — In [32], a mobility management method similar to AS is defined. It is called *Statistical Paging Area Selection* (SPAS) and is based on location statistics collected by each MS which periodically reports them to the network. These statistics consist of a list of the average duration the MS has been located in each LA. A priority rule is determined to settle the sequence of LAs visited by the mobile. If this sequence is different from the last one reported to the network, the MS transmits it; otherwise, nothing is done. The paging process is achieved in the same way as in AS. When the MS moves to an area that is not on the reported list, it has to process a *temporary location registration* to the network.

In [34], the proposed method provides a means of allowing preconnection and pre-assignment of data or services at the location before the user moves into it, so he can immediately receive service or data. This method clearly applies to location management. Just as are the previous two methods, it is based on users' movement history patterns. Called Mobile Motion Prediction (MMP), it allows the system to predict the future location of the user. Schematically, the MMP combines two movement models: *Movement Circle* (MC), based on a closed-circuit model of user movement behavior, and *Movement Track*, used to predict routine movements. MC is used to predict long-term regular movements.

Let us mention, finally, the method proposed in [35] which makes use of a cache memory for reducing the search cost. The proposal is to store the location of the frequently called mobiles in a local database (i.e., *cache*). This scheme allows the number of queries to the HLR to be reduced, thus reducing the signaling traffic at the fixed network side between the local database and the HLR.

Impact of LUs on Radio Resource Occupancy and the Number of MSC/VLR Transactions

In this section we examine the impact of LUs on RF resource occupancy in a DCS 1800 (or PCS 1900) network (urban environment, small cells, and high user densities). At the network level, we compute the number of transactions processed by the MSC/VLR and due to LUs. The number of transactions is defined here as the messages received or transmitted by the MSC/VLR.

A location updating exchange mainly uses an SDCCH (slow dedicated control channel). The SDCCH allocated to an MS consists of four time slots (for the SDCCH blocks) every 51 time-division multiple access (TDMA) multiframes. Let us consider more generally a channel as a single slot in a TDMA frame, corresponding, for instance, to one TCH/F (traffic channel/full rate). With this definition, a channel can therefore accommodate eight SDCCHs. That is, an LU consumes (during the signaling exchange) 0.125 TCH/F channel (i.e., eight MSs can share alternatively the same time slot on a TDMA 51-multiframe; Fig. 13).

Note — Dedicated channels in GSM (voice, data, and signaling channels) are always accompanied by an SACCH (slow associated control channel). This signaling channel is used for channel measurement transmissions on the uplink and call parameter settings on the downlink (e.g., power control).

Parameters and Hypothesis

In order to quantify the impact of the LUs on RF usage in a cell located at the border of a LA and on transactions processed at the MSC/VLR managing a set of LAs, we introduce the following parameters and hypothesis:

- ρ: density of the MSs in the cell (MSs/km^2),
- S: cell area (km^2); thus, $\rho.S$ gives the number of MSs per cell
- R: cell radius or side of the hexagon (km)
- L: length (km) of the cell exposed perimeter (i.e., perimeter being part of the LA border)
- v: average MS velocity (km/hr)
- $t_{LU}^{(i)}$: average duration of one LU in case i (i = 1: intra-VLR, i = 2: inter-VLR with TMSI, i = 3: inter-VLR with IMSI), equal to the time occupancy of one SDCCH/SACCH
- $TN_{LU}^{(i)}$: number of transactions processed by the MSC/VLR for one LU in case i
- $Pr_{LU}^{(i,j)}$: percentage of LUs in case i in cell number j (each cell is identified by a number)
- $\lambda_{LU}^{(j)}$: number of location updates in cell number j and per hour
- N: number of cells per LA
- N_{LA}: number of LAs managed by an MSC/VLR
- N_p: number of cells located on the perimeter of one LA; its expression is given by [36]:

$$N_p = 6.\sqrt{\frac{N}{3}} - 3 \quad (1)$$

- Cells assumed to be hexagonal
- Maximum blocking probability for the allocation of an SDCCH: 1 percent
- MSs uniformly distributed on the surface area of the cell
- Movements of the MS are decorrelated: the directions of their movements are uniformly distributed on $[0, 2\pi]$.

From [8], the number of location updates in an LA perimeter cell j and per hour is given by

$$\lambda_{LU}^{(j)} = \lambda_{LU} = v.L.\rho/\pi \quad (2)$$

From [36], L is given by the formula

$$L = 6.R.\left(\frac{1}{3} + \frac{1}{2\sqrt{3.N} - 3}\right) \quad (3)$$

Thus, the SDCCH/SACCH resource occupancy in the cell number j, due to MS LUs, is given by the formula

$$Tr_{LU}^{(j)} = \lambda_{LU} \cdot \left[\sum_{i=1}^{3} \Pr_{LU^{(i,j)}} \cdot t_{LU^{(i)}}\right] (\text{Erl}) \quad (4)$$

The number of transactions due to LUs generated in the $N_{LA}.N_p$ LA perimeter cells (which we number from 1 to $N_{LA}.N_p$) and processed per hour by the MSC/VLR is given by the formula

$$TTN_{LU} = \lambda_{LU} \cdot \left[\sum_{j=1}^{N_{LA} \cdot N_p} \left(\sum_{i=1}^{3} \Pr_{LU^{(i,j)}} \cdot TN_{LU^{(i)}}\right)\right] \quad (5)$$

For the numerical results, we consider now third-generation system parameters for high-density areas [37] during the peak traffic hours:

ρ = 10 000 MSs/km^2, R = 500 m, v = 10 km/h, N = 10, N_{LA} = 5.

The number of accesses to the MSC/VLR is

$TN_{LU}^{(1)} = 2$, $TN_{LU}^{(2)} = 14$, $TN_{LU}^{(3)} = 16$.

Furthermore, practical measures done on a GSM network give [38]

$t_{LU}^{(1)}$ = 600 ms, $t_{LU}^{(2)}$ = 3.5 s, $t_{LU}^{(3)}$ = 4.0 s.

Note — ρ can be the sum of ρ_1, ρ_2, and ρ_3, densities for user populations of three different operators.

Radio Resource Occupancy

We will consider two cases: an optimistic case (a cell where only intra-VLR LUs are generated) and a pessimistic case (a cell where only inter-VLR LUs are generated).

First Case — The considered cell (number j) is located at the border of two LAs related to the same VLR. In this case, only intra-VLR LUs will be processed in the cell (we neglect IMSI Attach procedures). Equation 4 gives then $Tr_{LU}^{(j)}$ = 7.30 Erl which, for a 1 percent blocking probability, requires *14/8 = 1.75 channels* (thus nearly one-forth of an RF channel).

Second Case — The considered cell (number j) is located at the border of two LAs related to two different VLRs. In this case, only inter-VLR LUs will be processed in the cell. We assume that we have 80 percent LUs with TMSI and 20 percent LUs with IMSI. Equation 4 then gives $Tr_{LU}^{(j)}$ = 42.46 Erl which, for a 1 percent blocking probability, requires *7 channels* (thus nearly 1 RF channel).

MSC/VLR Transaction Load

To evaluate the load on the MSC/VLR, we have to consider the whole area managed by the same MSC/VLR. It consists of N_{LA} = 5 LAs. We assume that one LA is located in the center of this area and that the remaining four are located on the border of the area. For sake of simplicity, we suppose that in all the perimeter cells belonging to the center LA and in half the perimeter cells of the border LAs, only intra-VLR LUs are generated. For the other half of the perimeter cells of the border LAs only inter-VLR LUs are generated. The number of cells where intra-VLR LUs are generated is given by Eq. 1 and is equal to

$$N_p + 4\frac{N_p}{2} = 18\sqrt{\frac{N}{3}} - 9$$

and the number of cells where inter-VLR LUs are generated is equal to

$$4\frac{N_p}{2} = 12\sqrt{\frac{N}{3}} - 6 .$$

Furthermore, we assume that among the generated inter-VLR LUs, 80 percent are LUs using TMSI and 20 percent are LUs using IMSI.

In this simple scenario, the number of transactions to be processed by the MSC/VLR because of LUs in its LAs is computed using Eq. 5 and is finally equal to

$TTN_{LU} \cong 12 \times 10^6$ transactions at peak hour.

Concluding Remarks

The results obtained in the previous section show that under heavy traffic conditions, the impact of LUs can be important. First, in terms of radio channels used, we showed that between one-fourth to one GSM RF carrier is used for location area boundary crossings. Although this burden cannot directly cause call blockings on the radio interface (in a DCS 1800 network using a 12 reuse cluster, the average number of RF carriers per cell, with three operators, is about 10), it has nevertheless a nonnegligible impact on the traffic channels consumption. Second, in terms of processing at the MSC/VLR side, with a processing power of about 12 x 10^6 transactions/hr, it is clear that blocking can rapidly be reached with our scenario. The MSC/VLR resource dedicated to LUs processing cannot be used for providing services. This is a major concern for operators whose objectives are to provide users with rich and sophisticated services, which require more processing resources at the MSC/VLR side.

The methods presented in the third section are intended to reduce the LUs signaling and processing on either the radio interface or the network equipment (switches and databases mainly). Unfortunately, it is quite difficult to compare these various mobility management schemes. Indeed, there exist no absolute bounds on their relative performances which can be applied to any procedure regardless of assumptions about mobility, network structure and so forth. In particular, there is a real need to obtain realistic performance evaluation models based on actual users' movements taking into account network structure [39]. With this aim, in [40] the authors set a general framework for evaluating location management schemes.

In this article, we have examined the impact of LUs on radio interface consumption and MSC/VLR processing. These evaluations have been made with a future wireless network framework in mind and using a simple fluid model for user movements. They show that the resource usage on the air interface can be important. However, the main impact of LUs is noticed at the MSC/VLR side, where the transaction rates can be very high and may be a cause of either saturation or reduction of the processing power dedicated to providing sophisticated services to subscribers.

REFERENCES

[1] *Mobile Communications International*, May 1997.
[2] "Phase 2+ Work Program," Int'l. doc.,ETSI TC SMG.
[3] "Wireless ATM," *IEEE Pers. Commun.*, Aug. 1996.
[4] G. P. Pollini, "Trends in Handover Design," *IEEE Commun. Mag.*, Mar. 1996, pp. 82–93.
[5] W. C. Y. Lee, *Mobile Cellular Telecommunications Systems*, McGraw-Hill, 1989.
[6] M. Mouly and M. B. Pautet, *The GSM System for Mobile Communications*, 1992.
[7] D. Lam, D. C. Cox, and J. Widom, "Teletraffic Modeling for Personal Communications Services," *IEEE Commun. Mag.*, Feb. 1997, pp. 79–87.
[8] G. Morales-Andres and M. Villen-Altamirano, "An Approach to Modeling Subscriber Mobility in Cellular Radio Networks," *Proc. Telecom Forum '87*, Geneva, Switzerland, Nov. 1987.
[9] R. Thomas, H. Gilbert, and G. Mazziotto, "Influence of the Moving of the Mobile Stations on the Performance of a Radio Mobile Cellular Network," *Proc. 3rd Nordic Sem. Digital Land Mobile Radio Commun.*, Copenhagen, Denmark, Sept.12–15, 1988.
[10] P. Carle and G. Colombo, " Sub-Optimal Solutions for Location and Paging Areas Dimensioning in Cellular Networks," *Proc. IEEE Int'l. Conf. Pers. Commun.*, Tokyo , Japan, Nov. 6–10, 1995.
[11] P. R. L. Gondim, "Genetic Algorithms and the Location Area Partitioning Problem in Cellular Networks," *Proc. IEEE VTC '96*, Atlanta, GA, Apr. 28–May 1, 1996.
[12] S. Behrokh and W. S. Wong, "Optimization Techniques for Location Area Partitioning," *Int'l. Teletraffic Conf. Sem. on UPT*, Italy, 1992.
[13] D. C. C. Wang, "A Survey of Number Mobility Techniques for PCS," *Proc. IEEE Int'l. Conf. Pers. Commun.*, Tokyo, Japan, Nov. 6–10, 1995.
[14] D. C. C. Wang, "A Survey of Number Mobility Techniques for PCS," *Proc. IEEE ICC*, 1994.
[15] N. and S. Tabbane, "Database Architectures and Location Strategies for Mobility Management in Mobile Radio Systems," *Proc. Wksp. Multiaccess, Mobility and Teletraffic for Pers. Commun.*, Paris, France, May 1996.
[16] I. F. Akyildiz and J. S. M. Ho, "On Location Management for Personal Communications Networks," *IEEE Commun. Mag.*, Sept. 1996, pp. 138–45.
[17] B. Jabbari, " Intelligent Network Concepts in Mobile Communications," *IEEE Commun. Mag.*, Feb. 1992.
[18] M. Listanti and S. Salsano, "Impact of Signaling Traffic for Mobility Management in an IN-based PCS Environment," *Proc. IEEE Int'l. Conf. Pers. Commun.*, Tokyo, Japan, Nov. 6–10, 1995.
[19] H. Xie and D. J. Goodman, "Signaling System Architecture Based on Metropolitan Area Networks," Winlab Tech. Rep. TR-39, Aug. 1992.
[20] D. Plassmann, "Location Management for MBS," *Proc. IEEE VTC*, Stockholm, Sweden, June 8–10, 1994, pp. 649–53.
[21] D. Goodman, P. Krishnan, and B. Sugla, "Design and Evaluation of Paging Strategies for Personal Communications," *Wksp. Multiaccess, Mobility and Teletraffic for Pers. Commun.*, Paris, France, May 1996.
[22] S. Okasaka *et al.*, "A New Location Updating Method for Digital Cellular Systems," *Proc. IEEE VTC '91*, Saint Louis, MI, May 1991.
[23] C.-L. I, G. P. Pollini, and R. D. Gitlin, "PCS Mobility Management Using the Reverse Virtual Call Setup Algorithm," *IEEE/ACM Trans. Networking*, vol. 5, no. 1, Feb. 1997, pp. 13–24.
[24] L.-R. Hu and S. S. Rappaport, "An Adaptive Location Management Scheme for Global Personal Communications," *Proc. IEEE Int'l. Conf. Pers. Commun.*, Tokyo, Japan, Nov. 6–10, 1995.
[25] A. Pashtan and I. Arich Cimed, " The CLU Mobility Management Scheme for Digital Cellular Systems," *Proc. IEEE VTC*, Atlanta, GA, Apr. 28–May 1, 1996, pp. 1873–77.
[26] H. Xie, S. Tabbane, and D. J. Goodman, "Dynamic Location Area Management and Performance Analysis," *Proc. IEEE VTC '93*, Secaucus, NJ, May 1993.
[27] S. J. Kim, C. Y. Lee, " Modeling and Analysis of the Dynamic Location Registration and Paging in Microcellular Systems," *IEEE Trans. Vehic. Tech.*, vol. 45, no. 1, Feb. 1996, pp. 82–89.
[28] "Location areas, paging areas and the Network Architecture, "RACE II Deliverable, R2066/PTTNL/MF1/DS/P/001/b1, Apr. 1992.
[29] S. Tabbane, "Comparison between the Alternative Location Strategy (AS) and the Classical Location Strategy (CS)," WINLAB Tech. Rep. 37, Aug. 1992.
[30] S. Tabbane, "An Alternative Strategy for Location Tracking," *IEEE JSAC*, vol. 13, no. 5, June 1995.
[31] Y.-B. Lin, "Reducing Location Update Cost in a PCS Network," *IEEE/ACM Trans. Networking*, vol. 5, no. 1, Feb. 1997, pp. 25–33.
[32] C. H. Rokitansky, " Knowledge-Based Routing Strategies for Large Mobile Networks with Rapidly Changing Topology," *Proc. ICCC '90*, New Delhi, India, Nov. 1990, pp. 541–50.
[33] M. Shirota, Y. Yoshida, and F. Kubota, "Statistical Paging Area Selection Scheme (SPAS) for Cellular Mobile Communication Systems," *Proc. IEEE VTC '94*, Stockholm, Sweden, June 8–10, 1994.
[34] G. Y. Liu and G. Q. Maguire, "A Predictive Mobility Management Algorithm for Wireless Mobile Computing and Communications," *Proc. IEEE Int'l. Conf. Pers. Commun.*, Tokyo, Japan, Nov. 6–10, 1995.
[35] Y.-B. Lin and S.-Y. Hwang, "Comparing the PCS Location Tracking Strategies," *IEEE Trans. Vehic. Tech.*, vol. 45, no. 1, Feb. 1996, pp. 114–21.
[36] E. Alonso, K. Meier Hellstern, and G. Pollini, "Influence of Cell Geometry on Handover and Registration Rates in Cellular and Universal Personal Telecommunications Networks," *Proc. 8th Int'l. Teletraffic Sem.*, Santa Margherita Ligure, Genova, Italy, Oct. 12–14, 1992, pp. 261–70.
[37] D. McFarlane and S. Griffin, " The MPLA Vision of UMTS," CODIT/BT/PM-005/1.0, Feb. 1994.
[38] X. Lagrange, P. Godlewski, and S. Tabbane, "Les Réseaux GSM/DCS," *Hermès*, Nov. 1995, p. 106–7.
[39] B. Jabbari, " Teletraffic Aspects of Evolving and Next-Generation Wireless Communications Networks, " *IEEE Commun. Mag.*, Dec. 1996, pp. 4–9.
[40] C. Rose and R. Yates, "Location Uncertainty in Mobile Networks: A Theoretical Framework," *IEEE Commun. Mag.*, Feb. 1997, pp. 94–101.

BIOGRAPHY

SAMI TABBANE presently teaches and performs research at ESPTT (Ecole Supérieure des Postes et des Télécommunications de Tunis, Cité El Ghazala, Route de Raoued, 2083, Tunisia). His research topics of interest are location management techniques, handover procedures, macrodiversity, and cellular network planning. He has been involved in several Tunisian committees to work on telecommunication network evolution in Tunisia and is currently responsible for cooperation projects with France, Canada, and Portugal in the domain of mobile communications. From 1992 to 1994 he worked with CNET (National Center for Telecommunications Studies of France Télécom) in Issy-les-Moulineaux, France, where he was involved in the European RACE II project. In 1992, he worked at WINLAB (Wireless Information Network Laboratory, Rutgers University, New Jersey) where he started his work on mobility management techniques. He received his Ph.D. from Ecole Nationale Supérieure des Télécommunications of Paris in 1991 and his engineer diploma from Ecole Nationale Supérieure des Mines de Paris in 1988. He is the coauthor of the book *Les Réseaux GSM-DCS* (Hermès, November 1995) and author of a book entitled *Réseaux Mobiles — Concepts de base et techniques opérationnelles* (July 1997).