Adaptive Region-Based Location Management for PCS Systems

Shou-Chih Lo and Arbee L. P. Chen, *Member, IEEE*

***Abstract*—The personal communications services (PCSs) systems can provide ubiquitous and customized services. The key issue, which affects the performance of the whole system, is the location management. In this paper, we propose a region-based location strategy by taking advantage of the user's movement behavior to improve the performance of the conventional systems. Each mobile user is associated with a set of regions, which are derived from the user's movement patterns. The registration processes in the same region can be eliminated such that the cost of location management can be significantly reduced. Several design issues are studied by considering the workload balance and the call-to-mobility ratio for a user. The proposed strategy can be dynamically adjusted based on different system parameters and user behaviors. A performance analysis on the signaling cost and the database access cost is given to justify the benefits of this approach.**

***Index Terms*—Location management, location registration, location tracking, personal communications service (PCS).**

I. Introduction

TELEPHONE services have brought great convenience to our life. The communication quality and service contents have been extensively enhanced with technology advances in telecommunications. Personal communications services (PCSs), with the characteristics of personalized and ubiquitous services, have become an interesting topic [7]. In the PCS environment, the geographic area is divided into adjacent cells, each covered by a base station. A mobile user with a handset can connect to the local base station via a wireless channel to receive or send any message.

Location management is an important task in order to communicate with the user, who may move from one cell to another at anytime. There are two processes for location management: registration and location tracking (or call delivery). Registration corresponds to the process of updating the data about the user's location, while location tracking corresponds to the process of searching the user's current location by the registration data. In general, several cells constitute a *location area* (LA); a registration is required as a user crosses the boundary of the LA. The *mobile switch center* (MSC), which can be viewed as a bridge to connect the wireless network and the wired network, may manage several LAs. To simplify the discussion, we assume there is only one LA within an MSC. The location management defined in IS-41 and the global system for mobile communication (GSM) standards adopts a two-level hierarchical architecture, which is composed of a centralized *home location register* (HLR) and multiple *visitor location registers* (VLRs) [20]. The VLR, which is usually coupled with an MSC, contains location information for the users in its service area, while the HLR contains permanent information for all users and pointers to the current serving VLRs of the users. Any reader interested in the current and proposed protocols for location management in various wireless systems is referred to [2].

In this paper, we consider the problem of how to efficiently locate a user under an environment characterized by the following properties: small cells and high user density. This paper focuses on a special class of users whose movements follow certain patterns. The main idea of our method is to form some regions for each user by grouping the LAs they frequently visit. The registration cost can be significantly reduced by decreasing the registration frequency as the user moves in the same region. We will discuss the suitable condition to put the restraint on the registration and provide a mechanism for the location tracking. Our primary contributions in this paper are as follows: we propose an efficient location strategy by taking advantage of the user's movement behavior; we construct a distributed HLR instead of a centralized HLR to balance the workload; moreover, we consider the ratio of the call arrival rate to the mobility rate [or *call-to-mobility ratio* (CMR)] for reducing the registration cost. The proposed strategy can be dynamically adjusted based on the number of regions, the degree of user mobility, and the system parameters such as the signaling cost between HLR and VLR.

The rest of this paper is organized as follows. In Section II, we present the location management defined in the IS-41 and give a brief introduction on other approaches. Section III presents the design for the region-based strategy in detail. In Section IV, the benefits of this method over IS-41 are indicated by performance analysis. Finally, we present the conclusion and future work in Section V.

II. Related Work

Location management is an important issue in PCS systems for efficiently locating a mobile user. In the IS-41 standard [8], the location information is managed by the HLR and the VLR. The respective flows for the processes of registration and location tracking are depicted in Fig. 1.

The identification number of VLR is periodically broadcast in each LA for assisting the user to judge whether a new LA is entered. If it is the case, a registration message containing the user's *mobile identification number* (MIN) is sent to the new VLR. Then the registration message is forwarded to the HLR. At the HLR, another message is sent to the old VLR to reclaim

Manuscript received November 8, 1999; revised October 23, 2001.

S.-C. Lo is with the Computer and Communication Research Center, National Tsing Hua University, Hsinchu, Taiwan, R.O.C.

A. L. P. Chen is with the Department of Computer Science and Information Engineering, National Dong Hwa University, Hualien, Taiwan, R.O.C.

Publisher Item Identifier S 0018-9545(02)02508-2.

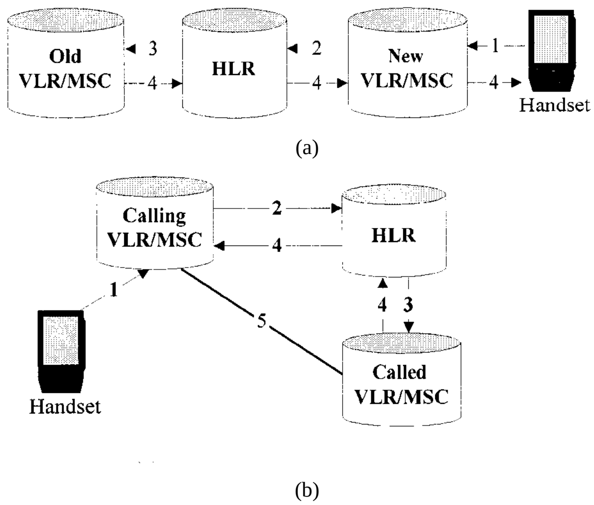


Fig. 1. Location management in the IS-41. (a) Registration process. (b) Location tracking process.

previously occupied resources (referred to as the *deregistration* process). When a call setup is going to be established, a request to query the callee's location is issued from the MSC to the HLR, then forwarded to the VLR where the callee currently visits. A routing number, such as a temporary location directory number (TLDN) in the IS-41 or a mobile station roaming number (MSRN) in the GSM, is returned. Finally, a trunk is established by using the routing number.

The location management specified in the IS-41 standard has several disadvantages. The HLR becomes a bottleneck of the network system. Also, the registration process may incur a big waste for a user frequently crossing LAs while seldom receiving calls. This situation can be stated as the user's having a low CMR. The CMR is defined as the number of calls to the user divided by the number of times the user crosses the LA boundary in a time unit. When the CMR is low, the number of registration processes performed by the user is more than the number of location tracking processes performed by the system to locate the user. An efficient location management strategy should minimize the costs of registration and location tracking. However, the two costs have a tradeoff. The more location information we record (i.e., increase the frequency of registration), the less location tracking cost we need to pay and vice versa. The improvement on the location management can be considered in three aspects: the registration frequency, the registration cost, and the location tracking cost.

A. Techniques for Reducing the Registration Frequency

The number of registrations can be controlled by performing the registration process only when a certain quantity is reached. As illustrated in [1] and [6], the quantity may be measured in time, the number of moves, or the moving distance. These methods have either poor performance or high complexity in implementation. Another possibility is to define the region (group of LAs) and then perform the registration process as the user enters a new region. In [5], the region is predefined, which consists of the neighboring cells of the *reporting center* selected from the base stations. However, an optimal selection of the reporting centers is an nameplate (NP)-complete problem in a hexagonal network configuration. A heuristic algorithm for selecting reporting centers was proposed in [10]. In [9], the region is preallocated by separating the PCS service area into several uniform groups. This approach has worse performance when a user frequently crosses the region boundary. In [26], the size of a region can be dynamically adjusted for a user according to the CMR value. Nevertheless, the region can only be shrunk or expanded in the form of a square shape. In [16], the optimal size and shape of a region is studied under a one-dimensional wireless environment. In [19], the region is composed of the current and the previous LAs that the user has visited. This method is suitable only for users moving back and forth crossing two LAs.

In [4] and [11], the region is generated by partitioning the movement patterns into groups. Hierarchical location servers are built over the regions. The search locating a user is performed either by the top–down or the bottom–up method to traverse the location servers. However, the complete signaling for the location management is not presented. Location management in the case where each user has only one region was addressed in [21] and [24]. The region is defined as a list of LAs where the user is most likely to be. Registration is performed when a user enters an LA not recorded in the list. Since the content of the list may change, retransmission of the list is required. Moreover, the search cost in a large region will be very high. In our approach, multiple regions are generated as in [11]. However, a different search method to locate a user is adopted. Furthermore, we reduce not only the number of registrations but also the cost of a registration by using the CMR value.

B. Techniques for Reducing the Registration Cost

To reduce the registration cost, the technique of *forwarding pointer*, by which the registration message is directed to the last visited VLR, was introduced in [13]. Implicit deregistration was studied in [17], where the obsolete resource (such as TLDN or MSRN) is reused when no free resource is available.

C. Techniques for Reducing the Location Tracking Cost

For reducing the location tracking cost, typical techniques include caching the search result of any outgoing call from an MSC for later use [12], replicating the location data in multiple VLRs [23], and constructing the location databases into a hierarchical architecture [3], [15], [25]. The search to locate a user in a region can be performed by broadcasting, by following the forwarding pointers that direct the moving tracks, or by sequentially searching each area in the order of probabilities [11]. The last method (named the *list method*) is better than others in most cases. The list method is adopted in our strategy for the location search. More complex search mechanisms were studied in [22].

III. Region-Based Location Strategy

In this section, a region-based location strategy is presented. First, we discuss various issues in the strategy design. Then, the architecture and the processing flows of location management for our proposed method are discussed.

A. Design Issues

1) The Regions: Two types of regions are defined based on a user's movement pattern. The *hot region* is a group of LAs that corresponds to a partition of the user's movement pattern [11]. The PCS service area excepting the hot regions constitutes a single *cold region*. Note that the regions are disjointed from each other. We have proposed four strategies to derive the regions from the movement patterns in [14]. To reflect the changes in user behavior, a simple method is to rebuild the regions periodically.

The location data are recorded by the following formats. The data about which region the user is in are recorded at the HLR, while the data about which LA in the region the user is in are recorded at a server associated with the region. When the user crosses the region boundary, both of these data are updated. To reduce the registration cost, we decrease the registration frequency as the user crosses the LA boundary in the same region. In this situation, a search within a region is needed to locate the user. To speed up the search, each LA in the region can be associated with a value that denotes the probability that the user may be there. Hence, the search is performed by sequentially searching each LA in the region in the order of probabilities. The probability information can be either provided by the user or estimated by the user's calling history. In certain conditions, such as when the user is in the cold region or with a high CMR, the search will be expensive. In these conditions, we perform the registration when a user crosses the LA boundary in the cold region or crosses the LA boundary in the hot region when the CMR is high.

2) The Options of the Registration: The location data at the HLR are usually updated when the user crosses the region boundary. This update cost is high in the PCS system. For some location data in the HLR, they may not be used for location tracking when the mobility rate is much higher than the call arrival rate (i.e., CMR is low). Therefore, we can eliminate the registration at the HLR when the CMR of the user is low as a means of reducing the registration cost. That is, the registration message is first sent to the server of the newly entered region, and then the message is optionally sent to the HLR based on the current CMR value. If the location data at the HLR are not up to date, the location tracking should be performed by first sequentially searching each region of the user. Again, each region can be associated with a probability value to speed up the search.

3) The Distributed HLR: A centralized HLR can become a bottleneck of the network system, which causes more call blockings. It is intuitive to build several distributed HLRs with an appropriate data allocation [18]. Although full data replication can balance the workload on querying the HLR for location tracking, the consistency maintenance becomes a problem, especially when a large volume of location updates is issued. Another choice is to partition the HLR. In this case, a function like *global title translation* (GTT) to decide which partition of the HLR stores the location data, given the identifier of a mobile user, should be supported. This model is adopted in our strategy.

B. System Architecture

According to the above design principles, we propose a three-level hierarchical architecture of location management. The logical view of our proposed architecture is shown in Fig. 2.

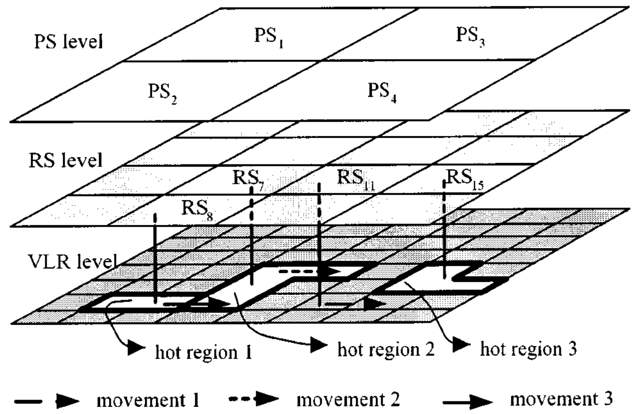


Fig. 2. Three-level hierarchical architecture.

Each square at the bottom level of the architecture represents a VLR (or an LA). The distribution of regions for a user is also shown in the figure. Note that each user has its own set of regions. In the middle level of the architecture, each square represents a *region server* (RS). Each the user's region is managed by the nearest RS, as indicated by the vertical line in the figure. For example, hot regions 1, 2, and 3 are managed by RS_8, RS_7, and RS_{15}, respectively, in the figure, while the cold region is managed by RS_{11}. Generally, an RS may manage several regions of different users. The HLR is partitioned, and each partition is managed by a *profile server* (PS) in the upper level of the architecture. The user's permanent information is stored in one of these PSs. The PS where the user's permanent information is stored is called the user's *home PS*. The placement of the PS depends on the population of mobile users with denser areas having more PSs.

Here, we give a brief description of how to perform the processes of registration and location tracking. Three movement cases in Fig. 2 are considered. When a user enters a new LA, the registration process is performed as follows. It should be decided first whether the user stays in the same region. If the user is still in the cold region (the case of movement 1), a registration message is sent to the RS of the cold region (RS_{11}), which indicates the user's current residing LA. A further notification to the home PS (say, PS_2) is needless. If the user is still in a hot region (the case of movement 2), the registration message is only sent to the RS of the current region (RS_7) when the CMR is high, but no registration is performed when the CMR is low. If the user crosses any region boundary (the case of movement 3), a registration message is sent to the RS of the newly entered region (RS_7). If the current CMR value is higher than the predefined threshold value, the RS then sends a registration message to the home PS (PS_2), which indicates the user's current serving RS. A record cancellation message is also sent from the new RS (RS_7) to the old one (RS_8).

Location tracking proceeds by first querying the home PS of the callee. From the PS, the current serving RS of the callee is found by first searching the recent serving RS (which is recorded in the PS) and then other RSs in the order of probability. The found RS then decides the current residing LA of the callee by searching first the recent residing LA (which is recorded in the RS) and then other LAs in the order of probabilities. The se-

TABLE I
PS TABLE

User ID	Recent RS	Associated RSs
MIN_1	RS_3	$\{(RS_1, \alpha_1), (RS_2, \alpha_2), \ldots\}$
MIN_2	RS_2	$\{(RS_2, \alpha_2), \ldots\}$

TABLE II
RS TABLE

User ID	Recent VLR	Associated VLRs	Home PS	CMR History
MIN_1	VLR_2	$\{(VLR_2, \alpha_2), (VLR_3, \alpha_3), \ldots\}$	PS_2	*Condition*
MIN_2	VLR_3	$\{(VLR_1, \alpha_1), \ldots\}$	PS_1	*Condition*

TABLE III
HANDSET TABLE

User ID	Last Visited VLR	Last Visited RS	CMR History
MIN_1	VLR_2	RS_3	*Condition*

TABLE IV
REGION TABLE

Region Number	Region Server	Associated VLRs
0	RS_4	*NULL*
1	RS_5	$\{VLR_2\ VLR_4 \ldots\}$

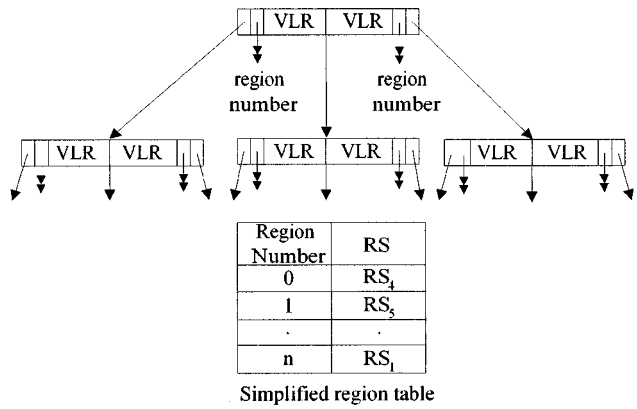


Fig. 3. A B-tree index for the region table.

quential search is due to the optional registration at the PS and the RS.

In Section IV, we will do a performance analysis to show when it is better not to perform registration at the RS and the PS. As we can see, the CMR value of each user is an important basis to adjust the performance of the proposed strategy. To obtain this information, we can implement two counters that record the number of incoming calls and the number of movements for a user.

C. The Location Databases

We introduce the data structure used in our region-based location strategy. Four tables are given as follows.

The *PS table,* as shown in Table I, is stored in the PS. The recent RS field records the RS the user recently registers. The associated RSs field records a list of RSs belonging to the user. Each RS in the list is also associated with a value that denotes the probability that the user may be there. The *RS table,* as shown in Table II, is stored in the RS. The recent VLR field records the VLR in the region the user recently registers. Similarly, the associated VLRs field records a list of VLRs for the LAs in the region with probability values. The home PS field records the user's home PS. The CMR history field describes the condition when the registration at the PS is not needed. The condition is described either by time periods or a value as the threshold.

In the user's handset, the *handset table,* as shown in Table III, is stored. The second and third fields record the last visited VLR and RS, respectively. The CMR history field describes the condition when the registration at the RS is not needed. The *region table,* which records the information about the regions of a user, is also stored in the user's handset. The data structure is shown in Table IV. Each region has a region number, a region server, and a list of VLRs for the LAs in this region. The last field for the cold region (numbered zero) is empty, as it is useless. The region table is used to identify the current residing region of the user, given the user's current serving VLR.

A large region table will incur significant overhead in identifying the region. To reduce this cost, a B-tree index structure can be constructed on the associated VLRs field in Table IV, as demonstrated in Fig. 3. Each VLR entry in the index structure constitutes a search-key value and a pointer for each search key. This pointer points to one of the entries in the simplified region table, where only the region number and the region server are recorded. The search proceeds by using the current VLR identification (ID) as the search key. If the search is successful, one hot region is found. Otherwise, the user is in the cold region.

D. Processing Flows

The processing flows of the registration and location tracking of the proposed strategy are illustrated.

1) Location Registration: When a user enters a new LA, a new VLR ID will be received and recorded in the field of last visited VLR in the handset table. This ID will be compared with the associated VLRs of the regions in the region table for deciding to which region the newly entered LA belongs. If the user is in the same hot region and the recent CMR value is lower than the threshold recorded in the CMR history field of the handset table, no registration is performed. Otherwise, the registration process is performed as shown in Fig. 4. The steps are illustrated as follows.

1) The handset sends the message ⟨MIN, old RS, current RS⟩ to the newly entered VLR. The first two fields in the message are acquired from the handset table, while the last field is the result of the comparison at the region table.
2) The newly entered VLR updates its database, indicating that the user is now staying in its coverage area. Then the message ⟨MIN, old RS, new VLR⟩ is sent to the current serving RS, as indicated by the current RS in the incoming message.

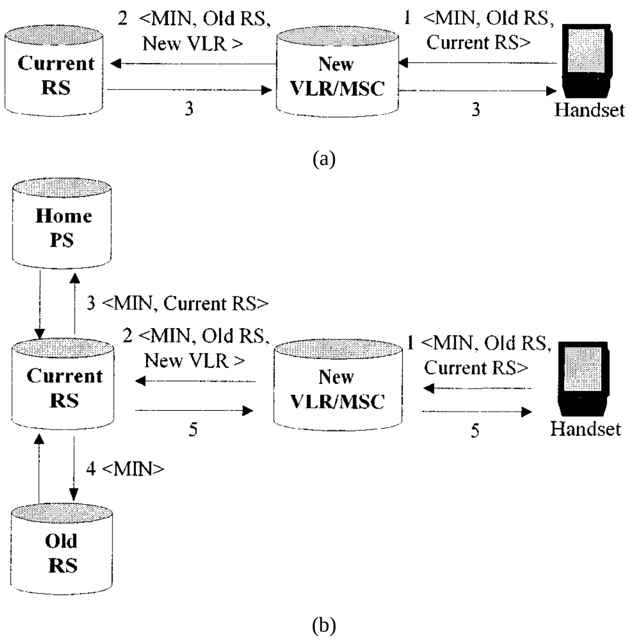


Fig. 4. The flow of registration. (a) Intraregion registration process. (b) Interregion registration process.

3) The current serving RS updates the recent VLR field, indicating the new VLR in the RS table for the user. Also, the current serving RS checks whether the user stays in the same region by referring to the old RS in the incoming message. Case 1 (intraregion registration): the user stays in the same region. Here, the old RS and the current RS in the message are the same. An acknowledgment is transmitted backward to the handset. The registration process is then finished. Case 2 (interregion registration): the user stays in a different region. Go to Step 4).
4) The current serving RS sends the message $\langle$MIN, current RS$\rangle$ to the home PS if the user's CMR is higher than the threshold recorded in the CMR history field of the RS table. The address of the home PS is acquired by the home PS field.
5) The current serving RS sends a resource cancellation message to the last serving RS, as indicated by the old RS in the incoming message.
6) An acknowledgment is transmitted backward to the handset. The handset then updates the last visited RS field to the new value in the handset table.

The resource cancellation message to the last visited VLR is not shown in the above processing flow because we assume that the implicit deregistration process [17] is operated. The implicit deregistration process is really needed for the case where the user enters a new VLR but no registration is performed.

2) Location Tracking: Location tracking is performed as shown in Fig. 5. The steps are illustrated as follows.

1) The handset sends the dial number (DN) of the callee to the current serving MSC.
2) The MSC then transmits the DN to a fixed switch where the global title translation is performed to get the address of the callee's home PS.
3) The request for querying the callee's location is forwarded to the home PS.

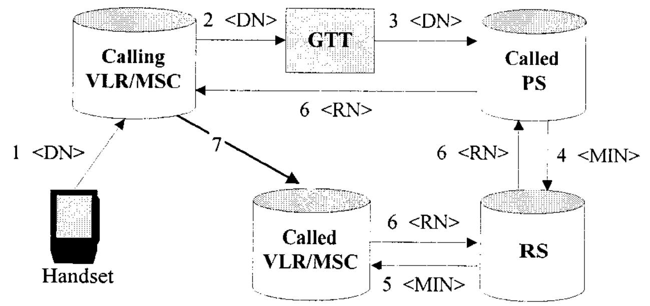


Fig. 5. The flow of location tracking.

4) The home PS searches for the current serving RS of the callee based on the PS table (the recent RS and associated RSs fields).
5) The RS then finds the current residing LA of the callee based on the RS table (the recent VLR and associated VLRs fields). The MSC should issue the paging message to confirm that the callee is in its coverage area.
6) If the callee is reachable in an LA, the routing number (RN) is then returned to the current serving MSC of the caller.
7) Finally, a trunk is established based on the RN.

IV. Performance Analysis

To evaluate the performance of the proposed strategy, we use an analytical model to explore the costs of registration and location tracking. We consider both the cost for transmitting signaling messages and the cost of database access in the analysis. As in [24], the analysis is performed for one user because the effectiveness of this strategy is dominated by personal behaviors. Experiments are performed to compare the cost of the proposed strategy with that of the strategy defined in IS-41.

A. Performance Environment

Assume that all hot regions are of the same size and that there are K LAs per hot region. The number of hot regions is determined by giving the total number of LAs in all hot regions and the region size in units of LAs. To reduce the complexity in the analysis, we assume that all hot regions are separate. The switches functioning as the GTT are colocated at the PSs.

We measure the signaling cost for a specified communication link by its delay value. Assume that these values can be acquired by a table lookup method. In Table V, we denote the delay value for each communication link as a variable. The item *paging* in the table denotes the cost to page a user in an LA. We define q_V, q_R, q_P, and q_H to be the costs for updating or querying the VLR, the RS, the PS, and the centralized HLR, respectively. We assume that $q_H = 1.5q_P$ in our analysis. The cost for performing the GTT is denoted as q_G.

We introduce parameters for modeling the movement behavior of the user in the following:

- N number of LAs in all hot regions;
- K number of LAs in a hot region;
- n number of hot regions (that is, $\lceil N/K \rceil$);
- P_i probability that the user is in region i (note that region 0 denotes the cold region); assume $P_1 \geq P_2 \geq \cdots \geq P_n$;

TABLE V
SIGNALING COST NOTATIONS

Communication Links	Cost	Communication Links	Cost
VLR − RS/GTT	d_1	VLR − PS	d_1+d_3
RS − RS	d_2	VLR − HLR	d_4
PS − RS/GTT	d_3	Page	*paging*

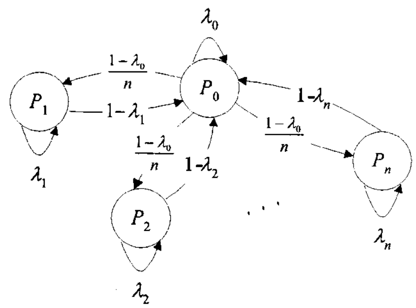


Fig. 6. The user's movement model.

λ_i probability that the user moves in the same region i;
α_i probability that the user is in LA i of a region; assume $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_K$.

The movement behavior can be modeled as a state transition diagram (see Fig. 6). The user may move in the same region or into another region. If the user is in the cold region, we assume that it has the same probability as the user's moving into any hot region. Since hot regions are supposed separate, crossing between hot regions is impossible. Each node in the diagram represents a region, and each edge represents a movement. Symbols in the node and on the edge denote the probability values of a user's being in the corresponding region and the movement, respectively. If the user has the same behavior in a certain time period, we can view the movement transition as a steady state. By the flow balance for each state, we can derive the equations shown in

$$\begin{cases} 1-\lambda_1 = \dfrac{P_0}{nP_1}(1-\lambda_0) \\ 1-\lambda_2 = \dfrac{P_0}{nP_2}(1-\lambda_0) \\ \quad\cdots \\ 1-\lambda_n = \dfrac{P_0}{nP_n}(1-\lambda_0). \end{cases} \tag{1}$$

The performance measurement is based on the total signaling cost and the total database access cost spent on the registration and location tracking. Define λ_c and λ_m to be the average number of incoming calls to the user per unit time and the average number of times the user changes LA per unit time, respectively. CMR can be expressed as λ_c/λ_m. Let C_R and C_L denote the costs for registration and location tracking, respectively. The total cost per time unit can be computed by the expression $(\lambda_c C_L + \lambda_m C_R)$. We separate the cost evaluation for the proposed strategy into four cases according to the option of registration at RS and PS as listed in Table VI. In the following, we list the expressions for evaluating the costs in these cases:

TABLE VI
THE DIFFERENT OPTIONS OF THE REGISTRATION

Strategies	Registration at RS (in the same hot region	Registration at PS
RB1+	*No*	*Yes*
RB1−	*No*	*No*
RB2+	*Yes*	*Yes*
RB2−	*Yes*	*No*

1) RB1+

$$C_R = \sum_{i=0}^{n} P_i \cdot (1-\lambda_i) \cdot (2d_1 + 2d_2 + 2d_3 + q_V + 2q_R + q_P) + P_0 \cdot \lambda_0 \cdot (2d_1 + q_V + q_R)$$

$$C_L = \sum_{i=1}^{n} P_i \cdot \left[2d_1 + 4d_3 + q_G + q_P + q_R + \sum_{j=1}^{K} j \cdot \alpha_j \cdot (2d_1 + paging + q_V) \right] + P_0 \cdot (4d_1 + 4d_3 + paging + q_G + q_P + q_R + q_V)$$

2) RB1−

$$C_R = \sum_{i=0}^{n} P_i \cdot (1-\lambda_i) \cdot (2d_1 + 2d_2 + q_V + 2q_R) + P_0 \cdot \lambda_0 \cdot (2d_1 + q_V + q_R)$$

$$C_L = \sum_{i=1}^{n} P_i \cdot \left[2d_1 + 2d_3 + i \cdot 2d_3 + q_G + q_P + i \cdot q_R + \sum_{j=1}^{K} j \cdot \alpha_j \cdot (2d_1 + paging + q_V) \right] + P_0 \cdot (4d_1 + (n+2) \cdot 2d_3 + paging + q_G + q_P + (n+1) \cdot q_R + q_V)$$

3) RB2+

$$C_R = \sum_{i=0}^{n} P_i \cdot (1-\lambda_i) \cdot (2d_1 + 2d_2 + 2d_3 + q_V + 2q_R + q_P) + \sum_{i=0}^{n} P_i \cdot \lambda_i \cdot (2d_1 + q_V + q_R)$$

$$C_L = \sum_{i=1}^{n} P_i \cdot (4d_1 + 4d_3 + paging + q_G + q_P + q_R + q_V) + P_0 \cdot (4d_1 + 4d_3 + paging + q_G + q_P + q_R + q_V)$$

4) RB2−

$$C_R = \sum_{i=0}^{n} P_i \cdot (1-\lambda_i) \cdot (2d_1 + 2d_2 + q_V + 2q_R) + \sum_{i=0}^{n} P_i \cdot \lambda_i \cdot (2d_1 + q_V + q_R)$$

$$C_L = \sum_{i=1}^{n} P_i \cdot (4d_1 + (i+1) \cdot 2d_3 + paging + q_G + q_P + i \cdot q_R + q_V) + P_0 \cdot (4d_1 + (n+2) \cdot 2d_3 + paging + q_G + q_P + (n+1) \cdot q_R + q_V)$$

5) IS-41

$$C_R = 4d_4 + 2q_V + q_H$$
$$C_L = 4d_4 + paging + q_G + q_H + q_V.$$

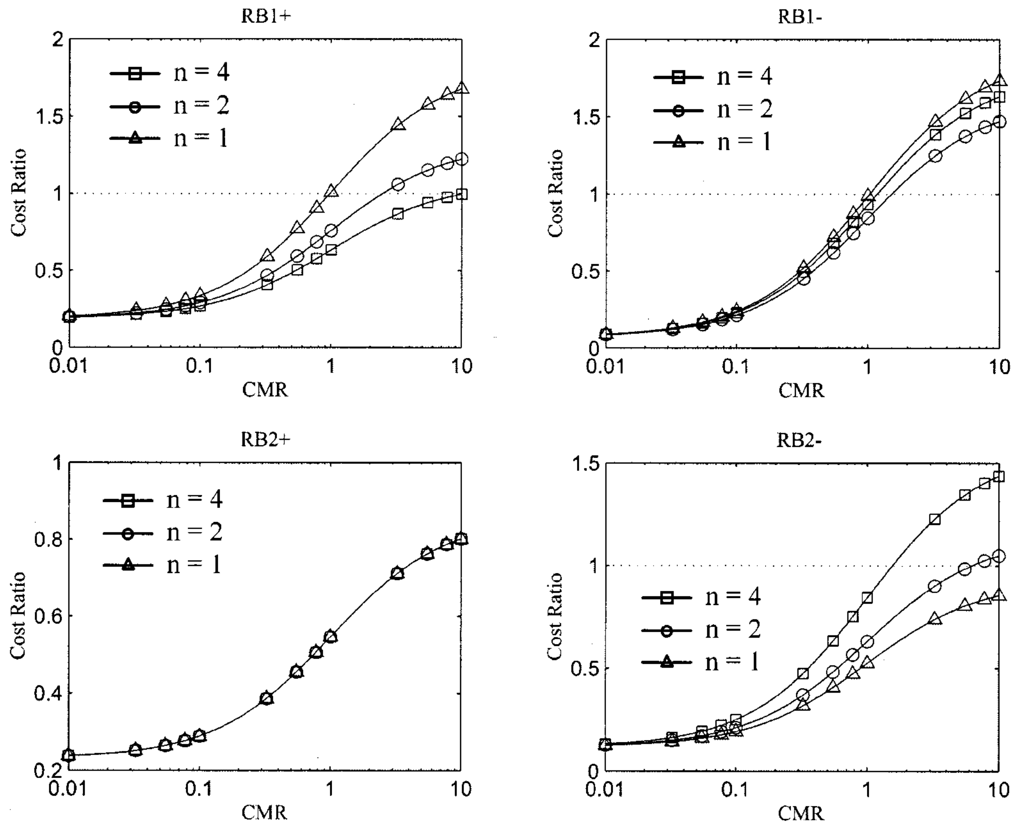


Fig. 7. The effect of the number of hot regions (data set 1).

TABLE VII
SYSTEM PARAMETERS

Parameters	Values
n	1,2, 4
N	32
P_i	$P_i = 0.2, (i = 0)$ $P_i = 0.8/n, (i \neq 0)$
λ_0	0.2

TABLE VIII
SIGNALING COST PARAMETERS

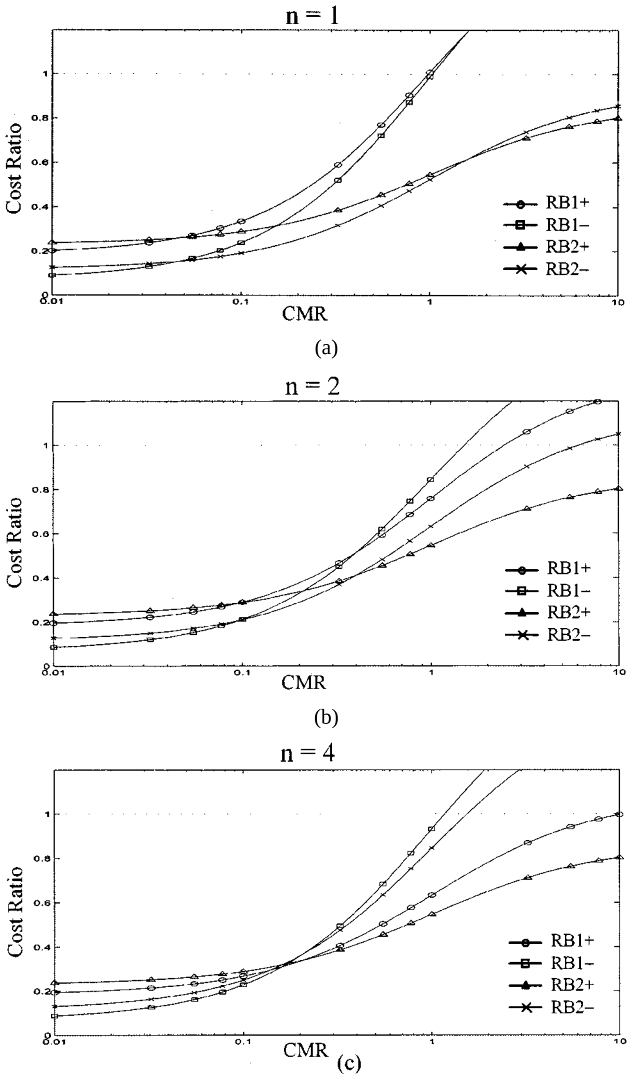
Data Sets	d_1	d_2	d_3	d_4	*Paging*
1	1	2	5	7	0.2
2	1	10	5	7	0.2
3	1	2	10	12	0.2
4	1	10	10	12	0.2

B. Experiments and Results

In this section, we observe the influence of different factors on the performance of the proposed strategy through the experiments. Also, we use the cost ratio of the proposed strategy to that of IS-41 as a comparison of the performance. That is, our proposed strategy outperforms IS-41 when the cost ratio is less than one. The values of signaling cost parameters and database access cost parameters are adopted from [9].

The settings of system parameters in our experiments are listed in Table VII. Here, we assume that the probabilities of appearing in the hot and cold regions have an 80–20 ratio. The probability distributions of a user's appearing in each region and each LA within a region are assumed to both be uniform (i.e., $P_i = 0.8/n$ and $\alpha_i = 1/K$). When the probability distributions are skewed, it means that the user is prone to stay in a certain region or LA. The cost of location tracking by the list method becomes small and our approaches will be even more outstanding. To save space, we omit the experimental results for the cases with skewed probabilities. To simplify the analysis, the cost evaluations on the signaling cost and the database access cost are separate. When one cost is considered, the other cost is negligible.

1) Signaling Cost: We consider four sets of signaling cost parameters as listed in Table VIII, where the values have been normalized based on the value of d_1. At first, we show the effect of the number of hot regions (or the size of a hot region) on the performance. Fig. 7 shows the change of the cost ratios for each strategy as data set 1 is used and n is set to 1, 2, and 4 (the size of hot regions is 32, 16, and 8, respectively). As n increases, the cost of RB1+ becomes smaller, while that of RB2− becomes larger. The reason is that the cost of location tracking dominates in these strategies. The search cost for the RSs is directly proportional to the number of hot regions ($\propto n$), while the search cost for the LAs is directly proportional to the size of a hot region ($\propto 1/n$). Both of these searches are performed in RB1−, so the cost varies as n increases. The current residing LA for a user is recorded



(c)

Fig. 8. The effect of the different options of the registration (data set 1).

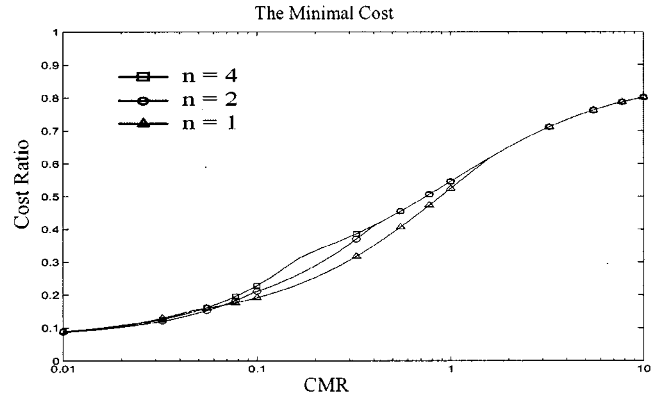


Fig. 9. The best performance for different numbers of regions (data set 1).

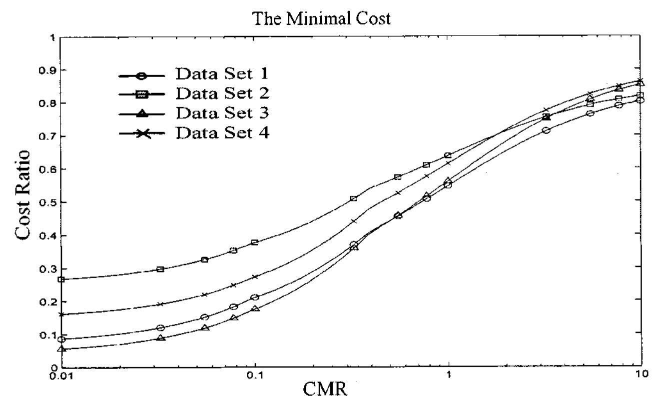


Fig. 10. The effect of varying signaling costs ($n = 2$).

in RB2+ such that no extra search is needed. Hence, the cost of RB2+ is independent of the value of n. When the CMR is high, the location tracking cost becomes dominant and its cost is sensitive to the value of n, so the difference of the cost ratios is large. However, when the CMR is low, the registration cost becomes dominant and its cost is independent of the value of n, so the difference of the cost ratios becomes small.

Second, we observe the effect of the different options of registration listed in Table VI on the performance. The experiment is performed by using data set 1 and various values of n. Fig. 8 shows the change of cost ratios for the strategy with different options, as the value of CMR varies from 0.01 to ten. We found that the strategy with a different option has a different performance as the CMR changes. Also, RB1+, RB1−, and RB2−, which involve the search costs for the RSs and/or the LAs, may perform worse than the IS-41 at high CMRs. This indicates that we should adopt a different option based on the CMR for getting the best performance. That is, the strategy should dynamically follow the options with the lowest cost ratio based on the current CMR. For example, if n is four, the strategy should behave as RB1− when CMR ≤ 0.1 and behave as RB2+ when CMR > 0.1. In both Figs. 7 and 8, we found that the performance improvement at low CMRs is more significant than that at high CMRs. The reason is that our proposed strategy focuses on reducing the registration cost and there is a tradeoff between the registration cost and the location tracking cost.

Fig. 9 shows the performances of the strategy, which always adopts the best option of registration under different values of n. It can be seen that having a number of hot regions ($n = 2$) is the best at low CMRs, while a single hot region is the best at higher CMRs. At low CMRs, the minimal cost comes from RB1−, which performs well as $n = 2$. We arrive at the following conclusion. If a mobile user is prone to have high mobility rates, the location strategy with adequate multiple hot regions should be adopted. Otherwise, the strategy with a single hot region should be adopted.

Fig. 10 shows the effect of signaling costs on the performance of the proposed strategy. The differences between data sets 1 and 2 and between data sets 3 and 4 demonstrate the effect of the signaling cost d_2 between two RSs. The figure shows that the effect of d_2 is more significant at low CMRs. The reason is that d_2 only affects the registration cost, which is dominant at low CMRs. The differences between data sets 1 and 3 and between data sets 2 and 4 demonstrate the effect of the signaling cost d_3 between the PS and the RS. Note that d_3 and d_4 are increased/decreased together in our settings. The effect of d_3 is also significant at low CMRs. The reason is that RB1− and RB2−, which have large cost savings of registration costs, dominate in these cases.

TABLE IX
DATABASE ACCESS COST PARAMETERS

Data Sets	q_R	q_V	q_P	q_G
5	1	3	6	0.5
6	1	3	12	0.5
7	1	6	6	0.5
8	1	6	12	0.5

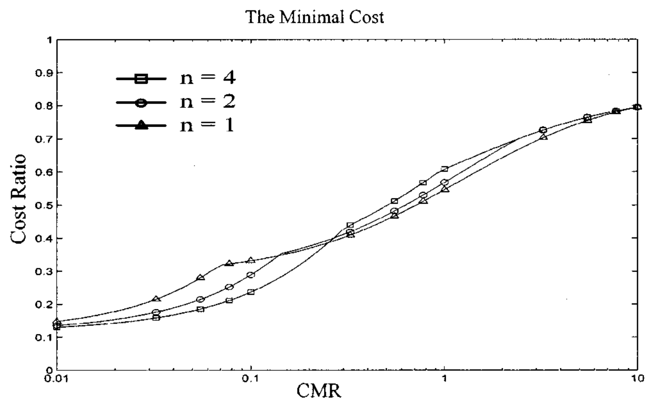


Fig. 11. The best performance for different numbers of regions (data set 5).

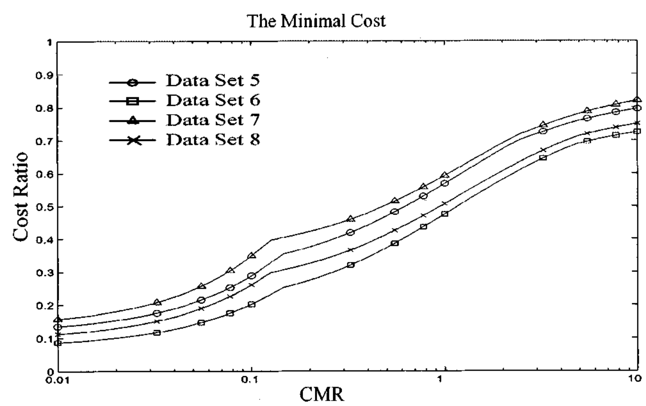


Fig. 12. The effect of varying database access costs ($n = 2$).

2) Database Access Cost: We consider four sets of database access cost parameters as listed in Table IX, where the values have been normalized based on the value of q_R. We perform a similar experiment as done in Fig. 9. The result is shown in Fig. 11. We found that the proposed strategy with the best option outperforms IS-41 even when the database access cost is considered. Moreover, the effect of multiple hot regions is more significant than that shown in Fig. 9. This indicates that no registration in the same hot region can save more database access costs than signaling costs. As the number of hot regions increases, the cumulative cost savings become large.

Fig. 12 shows the effect of database access costs on the performance. The differences between data sets 5 and 6 and between data sets 7 and 8 demonstrate the effect of the database access cost of the PS. The figure shows that more cost savings are obtained as q_P increases (q_H increases too). This reveals the fact that the centralized HLR is a bottleneck of the performance. The differences between data sets 5 and 7 and between data sets 6 and 8 demonstrate the effect of the database access cost of the VLR. It is obvious that the cost ratio increases as q_V increases.

From the experiments, it can be justified that the proposed strategy has a significant improvement over the scheme defined in the IS-41 standard. The proposed strategy will be dynamically adjusted (that is, adopting RB1+, RB1−, RB2+, or RB2−) based on the current CMR for a user. Although we focus on the techniques of reducing the registration cost, other techniques like caching and replication should be involved to reduce the location tracking cost in the system design. Also, the search order among regions or among LAs of a region will affect the location tracking cost. A more effective order can be arranged if the time is also considered in the movement history. Each region can be associated with time periods denoting when the user may be there. Given the time, the system can make a proper decision to locate the user.

V. Summary and Discussion

In this paper, we introduce a location management scheme based on regions. The regions are derived from the user's movement patterns. The proposed approach is effective if the user has a regular movement behavior. Our focus is to reduce the registration cost. We present a three-level hierarchical database architecture that consists of the VLR, the RS, and the profile server (PS). The RS is responsible for the registration within the region. The PS acts as a distributed HLR and records the recent serving RSs of the users. We consider the registration process in two phases: first at the RS and second at the PS. The best performance can be achieved by controlling the registration frequency in each phase. From the experiments, we obtain the following criteria. The registration frequencies in both phases should be reduced at low CMRs (less than 0.1), while only the registration frequency in the second phase should be reduced at middle CMRs (between 0.1 and 1). Our study also indicates that a single hot region is sufficient most of the time, except in the case of low CMRs (less than 0.1).

The movement patterns and call histories of mobile users should be properly managed to derive useful information. We name this work *profile management,* which includes data collection, data mining, and data maintenance. In the first part, the related data are collected and forwarded to the home PS for each user. Mining techniques should be explored in the second part. Five types of patterns can be considered by the following questions. Who is frequently called by a user? Where do most of the incoming calls come from for a user? When does a user receive most of the calls? When and where does a user frequently make a movement? Parts of these patterns have been cleverly used in some papers. In the final part, how to refresh the movement patterns will be a challenge because the moving information has been hidden from the system within a region. One approximation is to refer to the call history. To reduce the cost for managing these patterns, we can group users with similar behaviors to share the same patterns. Profile management is a problem we are currently addressing.

REFERENCES

[1] I. F. Akyildiz and J. S. M. Ho, “On location management for personal communications networks,” *IEEE Commun. Mag.*, pp. 138–145, Sept. 1996.

[2] I. F. Akyildiz, J. McNair, J. S. M. Ho, H. Uzunalioglu, and W. Wenye, “Mobility management in next-generation wireless systems,” *Proc. IEEE*, vol. 87, pp. 1347–1384, Aug. 1999.

[3] B. Awerbuch and D. Peleg, “Concurrent online tracking of mobile users,” in *Proc. ACM SIGCOMM’91*, Sept. 1991, pp. 221–233.

[4] B. R. Badrinath, T. Imielinski, and A. Virmani, “Locating strategies for personal communication networks,” in *IEEE Globecom ’92 Workshop Networking for Personal Communications Applications*, Dec. 1992.

[5] A. Bar-Noy and I. Kessler, “Tracking mobile users in wireless communications networks,” in *IEEE Infocom’93*, vol. 3, 1993, pp. 1232–1239.

[6] ——, “Mobile users: To update or not to update?,” in *INFOCOM’94*, June 1994, pp. 570–576.

[7] D. C. Cox, “Wireless personal communications: What is it?,” *IEEE Personal Commun.*, pp. 20–27, Apr. 1995.

[8] EIA/TIA, “Cellular radio telecommunication intersystem operations,” Tech. Rep. IS-41 (revision C), Nov. 1995.

[9] J. S. M. Ho and I. F. Akyildiz, “Dynamic hierarchical database architecture for location management in PCS networks,” *IEEE/ACM Trans. Networking*, vol. 5, pp. 646–660, Oct. 1997.

[10] A. Hac and X. Zhou, “Locating strategies for personal communication networks: A novel tracking strategy,” *IEEE J. Select. Areas Commun.*, vol. 15, pp. 1425–1436, Oct. 1997.

[11] T. Imielinski and B. R. Badrinath, “Querying in highly mobile distributed environments,” in *Proc. 18th VLDB*, Aug. 1992, pp. 41–52.

[12] R. Jain, Y. B. Lin, C. Lo, and S. Mohan, “A caching strategy to reduce network impacts of PCS,” *IEEE J. Select. Areas Commun.*, vol. 12, pp. 1434–1444, Oct. 1994.

[13] R. Jain and Y. B. Lin, “An auxiliary user location strategy employing forwarding pointers to reduce network impacts of PCS,” in *IEEE Int. Conf. Communications (ICC’95)*, 1995, pp. 740–744.

[14] G. Lee and A. L. P. Chen, “The design of location regions using user movement behaviors in PCS systems,” *Multimedia Tools Applicat.*, vol. 15, no. 2, pp. 187–202, Nov. 2001.

[15] D. Lam, Y. Cui, D. C. Cox, and J. Widom, “A location management technique to support lifelong numbering personal communication services,” *Mobile Comput. Commun. Rev.*, vol. 2, pp. 27–35, Jan. 1998.

[16] Z. Lei and C. Rose, “Wireless subscriber mobility management using adaptive individual location area for PCS systems,” in *IEEE Int. Conf. Communications (ICC’98)*, Atlanta, GA, June 1998.

[17] Y. B. Lin and A. Noerpel, “Implicit deregistration in a PCS network,” *IEEE Trans. Veh. Technol.*, vol. 43, pp. 1006–1010, 1994.

[18] Y. B. Lin, “Location tracking with distributed HLR’s and pointer forwarding,” in *Int. Symp. Communications*, 1995, pp. 31–37.

[19] ——, “Reducing location update cost in a PCS network,” *IEEE/ACM Trans. Networking*, vol. 4, 1996.

[20] S. Mohan and R. Jain, “Two user location strategies for personal communications services,” *IEEE Personal Commun.*, pp. 42–50, 1st quarter 1994.

[21] G. P. Pollini and C. L. I, “A profile-based location strategy and its performance,” *IEEE J. Select. Areas Commun.*, vol. 15, pp. 1415–1424, Oct. 1997.

[22] R. Rezaiifar and A. M. Makowski, “From optimal search theory to sequential paging in cellular networks,” *IEEE J. Select. Areas Commun.*, vol. 15, pp. 1253–1264, Sept. 1997.

[23] N. Shivakumar, J. Jannink, and J. Widom, “Per-user profile replication in mobile environments: Algorithms, analysis, and simulation results,” *MONET*, vol. 2, no. 2, pp. 129–140, 1997.

[24] S. Tabbane, “An alternative strategy for location tracking,” *IEEE J. Select. Areas Commun.*, vol. 13, pp. 880–892, June 1995.

[25] J. Z. Wang, “A fully distributed location registration strategy for universal personal communication systems,” *IEEE J. Select. Areas Commun.*, vol. 11, pp. 850–860, Aug. 1993.

[26] H. Xie, S. Tabbance, and D. J. Goodman, “Dynamic location area management and performance analysis,” *IEEE Trans. Veh. Technol.*, pp. 536–539, May 1993.

Shou-Chih Lo received the B.S. degree in computer science from National Chiao Tung University, Taiwan, R.O.C., in 1993 and the Ph.D. degree in computer science from National Tsing Hua University, Taiwan, in 2000.

He is now with the Computer and Communication Research Center, National Tsing Hua University, Taiwan, as a Postdoctoral Fellow. His current research interests are in the area of mobile and wireless internet with emphasis on mobility management, interworking operation, and MAC protocols with quality of service guarantee. He also works on problems related to index and data allocation on broadcast channels.

Dr. Lo received the Best Thesis Award from the Chinese Institute of Information and Computer Machinery in 2000.

Arbee L. P. Chen (S’80-M’84) received the B.S. degree in computer science from National Chiao-Tung University, Taiwan, R.O.C., in 1977 and the Ph.D. degree in computer engineering from the University of Southern California, Los Angeles, in 1984.

He joined National Tsing Hua University (NTHU), Taiwan, as a National Science Council (NSC) sponsored Visiting Specialist in August 1990 and became a Professor with the Department of Computer Science in 1991. In August 2001, he took a leave from NTHU and became Chairman of the Department of Computer Science and Information Engineering at National Dong Hwa University, Hualien, Taiwan. He was a Member of Technical Staff at Bell Communications Research, NJ, from 1987 to 1990, an Adjunct Associate Professor with the Department of Electrical Engineering and Computer Science, Polytechnic University, Brooklyn, NY, and a Research Scientist at Unisys, California, from 1985 to 1986. He is an editor of several international journals including the *World Wide Web: Internet and Web Information Systems*, (New York: Kluwer). His current research interests include multimedia databases, data mining, and mobile computing.

Dr. Chen was a recipient of the NSC Distinguished Research Award since 1996. He had organized the 1995 IEEE Data Engineering Conference and the 1999 International Conference on Database Systems for Advanced Applications (DASFAA) in Taiwan.