Dynamic Hierarchical Database Architecture for Location Management in PCS Networks

Joseph S. M. Ho, *Member, IEEE*, and Ian F. Akyildiz, *Fellow, IEEE*

***Abstract*—This paper introduces a dynamic hierarchical database architecture for location management in personal communications service (PCS) networks. The proposed scheme allows the dynamic adjustments of user location information distribution based on the mobility and calling patterns of the mobile terminals (MT's). A unique distribution strategy is determined for each MT, and location pointers are set up at selected remote locations which indicate the current location of the MT's. This method effectively reduces the signaling and database access overhead for location registration and call delivery. Besides, the required processing is handled by a distributed network of *directory registers* and centralized coordination is not necessary. The functions of the other network elements, such as the home location register (HLR) and the visitor location registers (VLR's), remain primarily unchanged. This greatly facilitates the deployment of this scheme in current PCS networks.**

***Index Terms*—Call delivery, directory register, home location register (HLR), location registration, visitor location register (VLR).**

I. INTRODUCTION

PERSONAL communications service (PCS) provides seamless and uninterrupted communication to mobile subscribers. PCS users carrying mobile terminals (MT's) can communicate with a remote terminal (mobile or static) regardless of its current location and mobility pattern. However, unlike static networks, such as the Internet, where routing information is embedded in the address of each node on the network, the current location of an MT cannot be obtained from its identification number. A location management scheme, therefore, is necessary to effectively keep track of the MT's and to locate a called MT when a call is initiated.

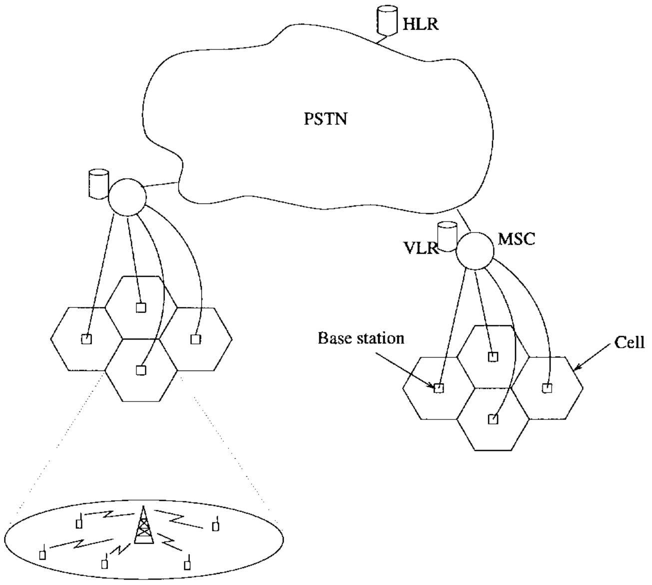


Fig. 1. PCN signaling network architecture.

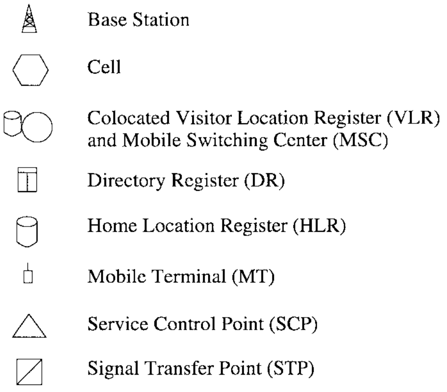


Fig. 2. Glossary of symbols.

There are two commonly used standards for mobility management: IS-41 [1], [11] and GSM [11], [12]. IS-41 is generally used in North America, while GSM is common in Europe. Under these standards, the location information of each MT is stored in databases which are updated as the MT moves to another location and are queried when a call is initiated. Fig. 1 shows an overview of the current PCS network architecture. For clarity, a glossary of symbols used throughout this paper is given in Fig. 2. The PCS coverage area is divided into *cells*. All MT's within a cell communicate with a *base station* through wireless links. The base station is, in turn, connected to the wireline network through a *mobile switching center* (MSC). The wireline network carries user information and signaling messages among the MSC's and the location databases. In current PCS systems, the *public switched telephone network* (PSTN) is used as the backbone wireline network.

Both the IS-41 and GSM standards employ a two-level database architecture consisting of the *home location register* (HLR) and the *visitor location registers* (VLR's). The HLR

Manuscript received June 19, 1996; revised January 31, 1997; approved by IEEE/ACM TRANSACTIONS ON NETWORKING Editor C. E. Perkins.

J. S. M. Ho was at the Gorgia Institute of Technology, Atlanta, GA. He is now with Nortel Wireless Networks, Richardson, TX 75083 USA (e-mail: joeho@nortel.com).

I. F. Akyildiz is with the Broadband and Wireless Networking Lab, School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332 USA (e-mail: ian@ee.gatech.edu).

Publisher Item Identifier S 1063-6692(97)07052-0.

is a centralized database containing the *user profiles* of its assigned subscribers. These user profiles record information such as the type of services subscribed, the quality of service (QoS) requirements, the billing information, and the current locations of the MT's. The VLR's are distributed throughout the PCS network, and each one stores the information of the MT's currently residing in its associated area. Depending on the network configuration, the user profile of an MT may be replicated at its current serving VLR. There are two possible implementations of the database structure [11]. In the first implementation, the VLR serves a number of MSC's and the major function of the VLR is to offload the query and signaling load at the HLR. In the second implementation, each VLR serves a single MSC and the VLR acts as an auxiliary processor to the MSC. The second implementation is more widely used in today's PCS networks. As described in [11], [12], all GSM switch manufacturers implement a combined MSC and VLR with one VLR per MSC.

We assume throughout this paper that the VLR is colocated with the MSC. When an MT moves to another MSC, signaling messages are exchanged among the HLR and the new and the old VLR's in order to record the new location information of the MT in the databases. We call this database update process *location registration*. Similarly, when a call is initiated, signaling messages are exchanged among the VLR of the calling MT and the HLR and the VLR of the called MT in order to locate the called MT (the "callee"). We call this callee location process *call delivery*. Depending on the relative position of the HLR and the VLR's, location management may result in significant signaling load to the network. As the number of PCS subscribers keeps increasing, it is demonstrated in [9] and [10] that the volume of signaling and database access traffic will increase beyond the capacity of the current network design. Methods for reducing the signaling and database access traffic are, therefore, necessary before PCS use can become widespread.

In this paper, we introduce a dynamic hierarchical database architecture for PCS location management. Under this approach, an additional level of databases called the *directory registers* (DR's) is introduced. Each DR determines the *location information distribution strategies* for its associated MT's. Location *pointers* are then set up at selected remote DR's indicating the current location of the MT. Using this scheme, location information is distributed more effectively and the signaling and database access overhead for location registration and call delivery is greatly reduced. Moreover, the functions of the HLR, the VLR's, and the MSC's remain primarily unchanged. This facilitates deployment of the proposed scheme in current PCS networks. We will present the proposed database architecture and its associated location management scheme in more detail in Section IV.

This paper is organized as follows. In Section II, we give a brief introduction to recent research in location management. Section III describes the PCS signaling network architecture and the IS-41 location management standard. Section IV introduces the proposed hierarchical database architecture and the location registration and call delivery procedures associated with this new database architecture. An analysis of the signaling and database access costs for the proposed scheme is also given in this section. Section V presents an enhancement for the proposed location management scheme. Conclusions are given in Section VI.

II. RECENT RESEARCH

A *per-user location caching* strategy is introduced in [6] which reduces the signaling cost for call delivery by reusing the cached information about a called MT's location from a previous call. The performance of this scheme depends on the probability that the cached information is valid. In [7], the author introduces a threshold scheme that dynamically determines the time when a cache record becomes obsolete.

In [13], the authors propose a *user profile replication* scheme. Based on this scheme, user profiles are replicated at selected locations such that the call delivery delay can be reduced. The replication decision is made by a centralized system which must collect the mobility and calling parameters of the whole user population from time to time. This may not be feasible in current PCS networks because of the large number of PCS network providers involved. Besides, generating and distributing the replication decision for a large user population[1] is a computationally intensive and time-consuming process which may incur significant amount of network bandwidth.

In [5], a *location forwarding* strategy is proposed to reduce the signaling cost for location registration. Whenever an MT moves to another VLR, a pointer is set up from the old VLR to the new VLR. When an incoming call arrives, the HLR determines the serving VLR of the MT by transversing the pointer chain. This method reduces the location tracking cost when the call-to-mobility ratio is low. However, a number of VLR queries have to be performed in order to locate the MT. This introduces additional overhead in the call delivery process.

In [14], the author proposes a fully distributed database architecture for location management. Under this scheme, location information of the MT's is distributed among a large number of location databases, organized as a tree structure. Each database contains location information of the MT's residing in its subtree. Under this scheme, database accesses are localized. However, a large number of location databases may have to be accessed during location registration and call delivery. In [4], a partitioning scheme is introduced for the fully distributed location database architecture. Partitions are generated by grouping the location databases among which the MT's move frequently; location registration is performed only when an MT enters a partition. This scheme minimizes the number of location registrations in areas where the mobility rate of the MT's is high. In [3], the authors suggest that it is not necessary to place a location database at every node of the tree structure. A method for determining the near optimal database placement is introduced such that the number of database queries and updates is minimized.

---

[1] The mobile user population in the United States is expected to exceed 60 million by the year 2005.

The location management schemes described above can be classified into two categories. For the first category, the IS-41 or GSM centralized location database architecture is used and auxiliary strategies [5]–[7], [13] are introduced which reduce the signaling or database access overhead for either location registration or call delivery. For the second category, the fully distributed location database architecture [3], [4], [14] is used such that location information for the MT's are distributed among a number of databases. The performance of the schemes in both categories depends on the mobility and calling parameters of the MT's. Neither the centralized nor the fully distributed database architecture is superior under all circumstances [2].

Most of the recent research efforts intend to modify the distribution of location information in order to achieve lower signaling overhead. For example, in [5], location forwarding reduces the overhead for location registration by distributing the location information of the MT in a number of recently visited VLR's. In [3], however, a reduction in the database access overhead is achieved by limiting the distribution of the location information.

We believe the optimal distribution of location information is user dependent and the distribution strategy for one MT may not be appropriate for another MT. In this paper, we introduce a database architecture which facilitates the dynamic adjustment of the location information distribution. Based on this scheme, the distribution strategy for each MT is dynamically determined based on its mobility and call arrival parameters. The required computation is handled by a network of DR's and the distribution strategy for each MT is determined independently without centralized coordination. As a result, this scheme is easily scalable. Finally, the proposed scheme reduces the overhead for both location registration and call delivery. This results in significant overall reduction in signaling and database access costs, compared to the IS-41 scheme.

III. SYSTEM DESCRIPTION

Signaling messages in current PCS systems are carried by the *Signaling System No. 7* (SS7) network. Fig. 3 shows a simplified SS7 network architecture. The HLR and VLR's store the user profiles for their associated MT's. The functions of the HLR and the VLR's were described in Section I. The MSC's provide switching functions for MT's in their associated area. In addition, they are responsible for location registration and paging procedures, and handoffs. The *Signal Transfer Point* (STP) is a switch on the SS7 network which is responsible for the routing of signaling messages based on their destination addresses. The *Service Control Point* (SCP) contains the database, such as the HLR, and the associated logic which handles database query and update requests initiated by the MSC's. These network elements are interconnected by wireline links. For reliability reasons, the STP's are installed in pairs and there exists a number of links between two network elements.

Procedures for location registration and call delivery are specified in the IS-41 standard. According to IS-41, the HLR knows the exact MSC's where its associated MT's reside. The area covered by an MSC is called a *Registration Area* (RA) and an MT performs a location registration when it enters an RA. Fig. 4 shows the signaling diagram for location registration under IS-41. In Fig. 4 (as well as in Figs. 5 and 8–10), the "( )" indicates the message number, the "[ ]" indicates the cost for the particular signaling exchange, and the "{ }" at the bottom of the figure indicates the cost for accessing the particular database. These signaling and database access costs will be discussed in more detail in Section IV-C. The procedure for location registration under IS-41 is described as follows.

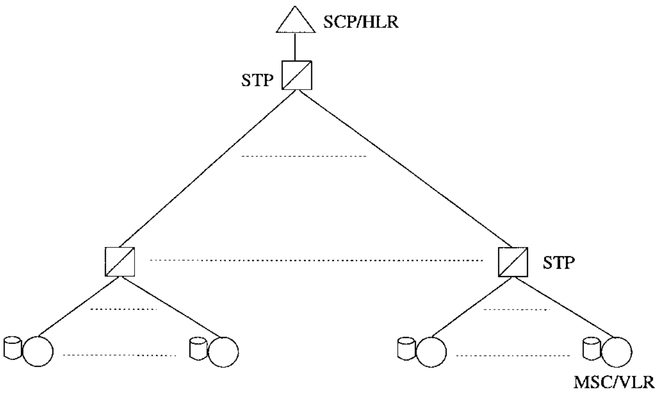


Fig. 3. Signaling network.

1) The MT detects that it has entered a new RA and sends a location update message to the MSC through the base station.
2) The MSC updates its associated VLR indicating that the MT is residing in its area and sends a location registration message to the HLR.
3) The message is routed to an STP which determines the HLR of the called MT from its *Mobile Identification Number* (MIN) by a table lookup procedure called *Global Title Translation* (GTT). The location registration message is then forwarded to the HLR.
4) The HLR updates its record indicating the current serving MSC of the MT and sends a registration acknowledgment message to the new MSC.
5) The HLR sends a registration cancellation message to the old MSC.
6) The old MSC deletes the record of the MT in its associated VLR and sends a cancellation acknowledgment message to the HLR.

The IS-41 call delivery procedure is described as follows (see Fig. 5).

1) A call is initiated by an MT and the base station forwards the call initiation signal to the MSC.
2) The MSC sends a location request message to the HLR through an STP where GTT is performed.
3) The location request is forwarded to the HLR of the called MT.
4) The HLR sends a location request message to the MSC serving the called MT.
5) The MSC determines the cell location of the called MT and assigns it a *Temporary Location Directory Number*

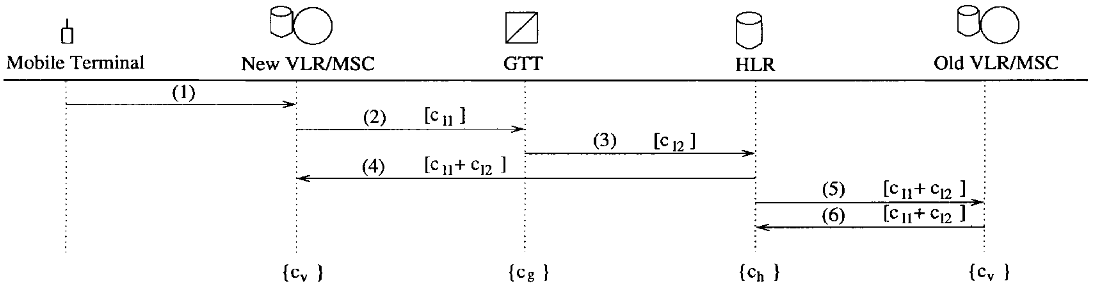


Fig. 4. Location registration under IS-41.

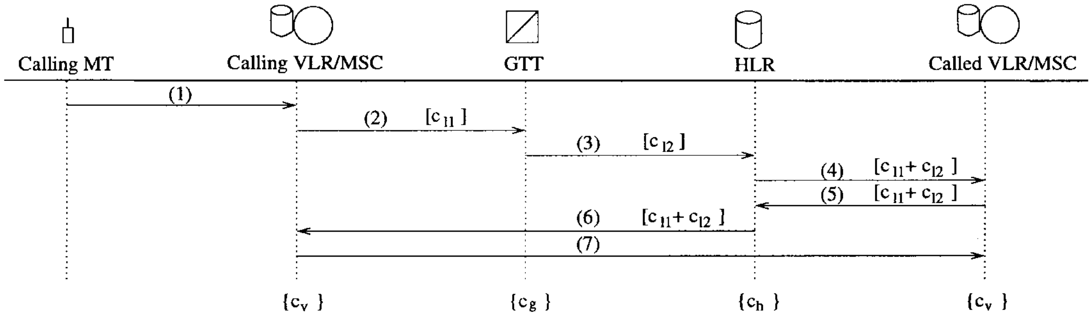


Fig. 5. Call delivery under IS-41.

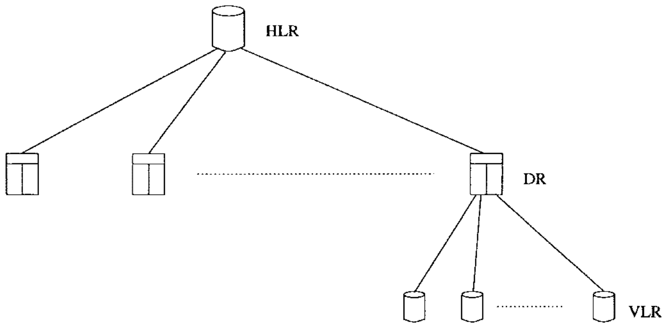


Fig. 6. Hierarchical database architecture.

(TLDN). The MSC then sends this TLDN to the HLR.
6) The HLR forwards the TLDN to the calling MSC.
7) The calling MSC sets up a connection to the called MSC using this TLDN.

IV. DYNAMIC HIERARCHICAL DATABASE ARCHITECTURE

In this section, we introduce a dynamic hierarchical database architecture for location management in PCS networks. The proposed database architecture is based on that of the IS-41 standard with the addition of a new level of databases called *directory registers* (DR's). The placements and the operations of the HLR, the VLR's, and the MSC's remain mostly unchanged, while additional processing is handled by the DR's. Fig. 6 shows the proposed database architecture which contains the HLR, the DR's, and the VLR's. We assume that the DR's are installed in SCP's on the SS7 network and each DR serves a number of MSC's. The area covered by a DR is called the *directory area* (DA).

Throughout this paper, we assume a DR consists of the database and the computer for providing the required location management functions. The responsibilities of the DR's are as follows:

1) periodically compute and store the location information distribution strategy for its associated MT's;
2) store location pointers which indicate the current locations of local and selected remote MT's;
3) forward location registration and call initiation requests from it associated MT's to the appropriate network elements as indicated by the location pointers.

Three types of location pointers are defined for each MT.

- *Local Pointer:* A local pointer for an MT is stored at its serving DR which indicates the current serving MSC of the MT. Based on the proposed scheme, the DR contains a local pointer for each of its associated MT's. Local pointers improve location registration performance.
- *Direct Remote Pointer:* A direct remote pointer for an MT is stored at a remote DR which indicates the current serving MSC of the MT. Direct remote pointers improve call delivery performance.
- *Indirect Remote Pointer:* An indirect remote pointer for an MT is stored at a remote DR which indicates the current serving DR of the MT. Indirect remote pointers improve both location registration and call delivery performance.

These pointers are used to distribute the location information for the MT's. For example, suppose that the HLR of a given MT is located in New York and it is currently roaming in San Francisco. If a significant number of the incoming calls for the MT are originated from Los Angeles, a remote pointer can be set up for the MT in the DR at the Los Angeles area. When the next call is initiated for this MT from Los Angeles, the search can be forwarded directly to San Francisco without requiring a query at the HLR which is located in New York. This reduces

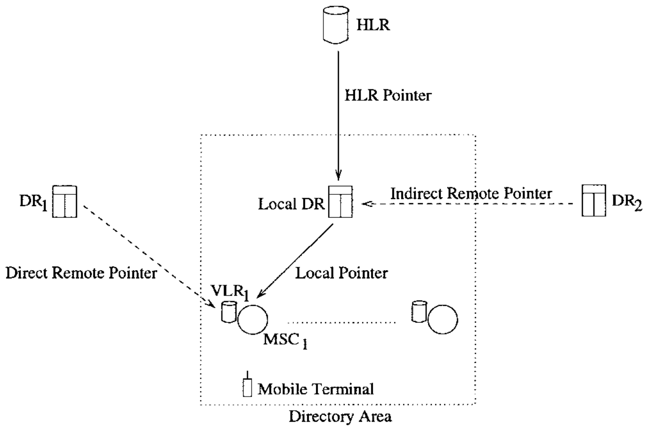


Fig. 7. An example location information distribution strategy.

the signaling overhead for call delivery. On the other hand, the HLR can be set up to record the ID of the serving DR (instead of the serving MSC) of the MT. When the MT moves to another RA within the same DA in San Francisco area, only the local pointer at the DR has to be updated. Again, it is not necessary to access the HLR in New York. This reduces the signaling overhead for location registration. Unlike other schemes [5], [6], [13] as described in Section II, an advantage of this scheme is that it can reduce the overhead for both location registration and call delivery.

As the mobility and call arrival patterns vary from user to user, the pointer configuration for each MT must be uniquely determined. Based on our proposed scheme, the DR periodically computes a location information distribution strategy for each of its associated MT's which specifies:

1) the set of remote DR's where direct remote pointers for the MT should be set up;
2) the set of remote DR's where indirect remote pointers for the MT should be set up;
3) whether the HLR should record the ID of the serving MSC or the serving DR of the MT.

Fig. 7 shows the pointer configuration for a given MT. It can be seen in Fig. 7 that this MT is currently residing under MSC_1 and it has a direct remote pointer and an indirect remote pointer from DR_1 and DR_2, respectively. The HLR records the ID of the current serving DR of the MT. When the MT moves to another MSC within the same DA, the local pointer and the direct remote pointers become invalid and the DR updates these pointers to indicate the correct location of the MT. When the MT moves to another DA, all pointers associated with the MT become invalid. The DR informs the HLR of the location change and updates all pointers associated with the MT.

The DR recomputes the pointer configuration periodically at constant time intervals or at specific time instants of the day. If a new remote pointer is to be added, a message is sent to the corresponding remote DR to set up the new pointer. If an existing remote pointer is no longer necessary, it is deleted, instead of updated, when the remote pointer becomes invalid. Note that a direct remote pointer becomes invalid when the MT moves to another MSC and a indirect remote pointer becomes invalid when the MT moves to another DA.

The location pointers are implemented as entries in a directory table stored at the DR. When the DR receives a location registration or a call delivery request, a table lookup operation is performed and the request is forwarded to the appropriate network elements. The DR is not involved in the processing required for location registration and call delivery (such as the recording of billing information). This results in small processing overhead at the DR as compared to that of the HLR and the VLR's.

Similar to the STP's, the DR's may be installed in pairs to enhance fault tolerance. One of the DR in the pair is *active* and the other DR is *inactive*. Both DR's, active and inactive, contain exactly the same information. However, only the active DR is responsible for the required processing. When the active DR becomes unavailable due to failure or maintenance, the inactive DR takes over the normal processing and becomes the active DR.

For security and privacy reasons, the service provider of a particular MT may not want the location information of this MT to be stored at remote DR's. This can be achieved by setting up a *DR overriding* flag in the user profile of the MT. The DR's will be denied access to the information of the MT if this flag is set. Under this situation, the IS-41 location management scheme is used for this MT and the HLR always contains the ID of the serving MSC of this MT.

The procedures for location registration and call delivery under this new database architecture are given in Section IV-A. The method for determining the location information distribution strategy for an MT is described in Section IV-B. It is demonstrated in Section IV-C that the proposed scheme can significantly reduce both the signaling and database access overhead for location management.

A. Location Registration and Call Delivery

When an MT moves to another MSC, a location registration is performed which i) updates or creates the local pointer for the MT at the serving DR, ii) deregisters the MT at the previous MSC and DR (if an inter-DA movement is performed), iii) updates the location information of the MT at the HLR if necessary, and iv) updates the remote pointers of the MT if necessary. Figs. 8 and 9 show the location registration procedure. The steps shown in Figs. 8 and 9 are described as follows.

1) The MT moves to a new MSC and sends a location update message to this new MSC.
2) The new MSC updates its VLR indicating that the MT is now residing in its associated area and sends a location registration message to the DR.
3) If an intra-DA movement has occurred, then we have the following.
    a) The DR updates the local pointer for the MT indicating the new MSC and sends a registration cancellation message to the old MSC.
    b) The old MSC sends the user profile of the MT to the DR.

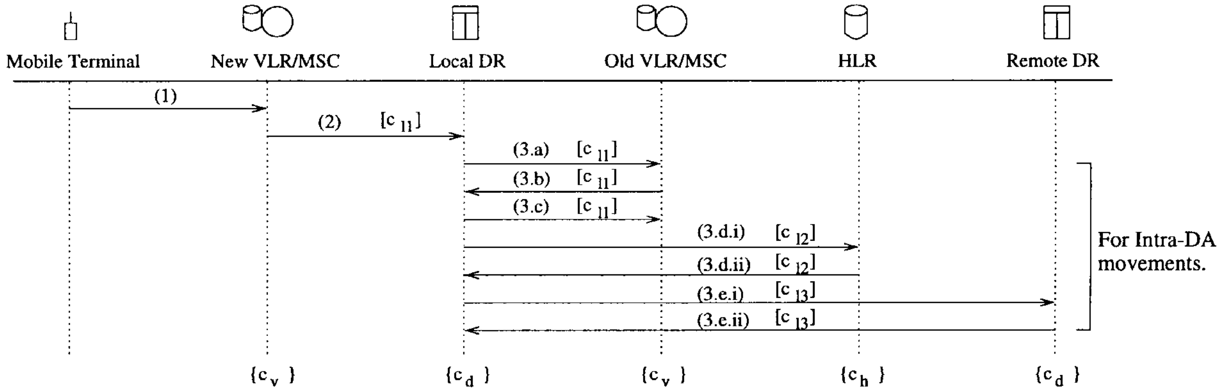


Fig. 8. Location registration under new database architecture.

- c) The DR sends an acknowledgment message to the old MSC which then deletes the record for the MT in its associated VLR.
- d) If the HLR records the ID of the serving MSC of the MT, then:
  - i) the DR reports the location change to the HLR;
  - ii) the HLR sends an acknowledgment message to the DR.
- e) If direct remote pointers exist in one or more remote DR's, then:
  - i) the serving DR sends a pointer update message to all the remote DR's that contain a direct remote pointer for the MT;
  - ii) The remote DR's update the direct remote pointers and send acknowledgment messages to the serving DR.

  Otherwise, if an inter-DA movement has occurred, we have the following.
- f) The DR creates a directory entry for the MT and records the ID of the serving MSC. The DR then sends a location registration message to the HLR.
- g) The HLR sends a registration cancellation message to the old DR.
- h) The old DR sends a registration cancellation message to the old MSC.
- i) The old MSC sends the user profile of the MT to the old DR.
- j) The old DR sends an acknowledgment to the old MSC which then deletes the record for the MT in its associated VLR.
- k) The old DR sends the directory information of the MT to the HLR.
- l) The HLR sends an acknowledgment to the old DR which then deletes its record for the MT.
- m) The HLR sends the directory information and the user profile of the MT to the new DR.
- n) The new DR sends an acknowledgment to the HLR.
- o) If direct or indirect remote pointers exist in one or more remote DR's, then:
  - i) the serving DR sends a pointer update message to all the DR's that contain a remote pointer for the MT;
  - ii) the remote DR's update the pointers and send acknowledgment messages to the serving DR.

4) The new DR sends the user profile of the MT to the new MSC.
5) The new MSC sends an acknowledgment to the new DR.

Call delivery involves the determination of the current serving MSC of a called MT. Fig. 10 shows the procedure for call delivery under the proposed database architecture. The steps as shown in Fig. 10 are described as follows.

1) A call is initiated by an MT and the base station forwards the call initiation signal to the MSC.
2) The MSC sends a location request message to the local DR.
3) If a local pointer for the called MT exists at the DR, then we have the following.
   - a) The DR sends a location request message to the local MSC which serves the called MT.
   - b) The MSC assigns a TLDN to the called MT and sends this TLDN to the DR.

   Otherwise, if a direct remote pointer for the called MT exists at the DR, then we have the following.
   - c) The DR sends a location request message to the remote MSC which serves the called MT.
   - d) The MSC assigns a TLDN to the called MT and sends this TLDN to the calling DR.

   Otherwise, if an indirect remote pointer for the called MT exists at the DR, then we have the following.
   - e) The DR sends a location request message to the DR of the called MT.
   - f) The called DR forwards the location request message to the MSC of the called MT.
   - g) The MSC assigns a TLDN to the called MT and sends this TLDN to the calling DR.

   Otherwise, if no pointer exists for the called MT, then we have the following.
   - h) The DR sends a location request to the HLR of the MT.
   - i) If the ID of the serving MSC of the called MT is available at the HLR, then:
     - i) the HLR sends a location request message to the MSC of the called MT;

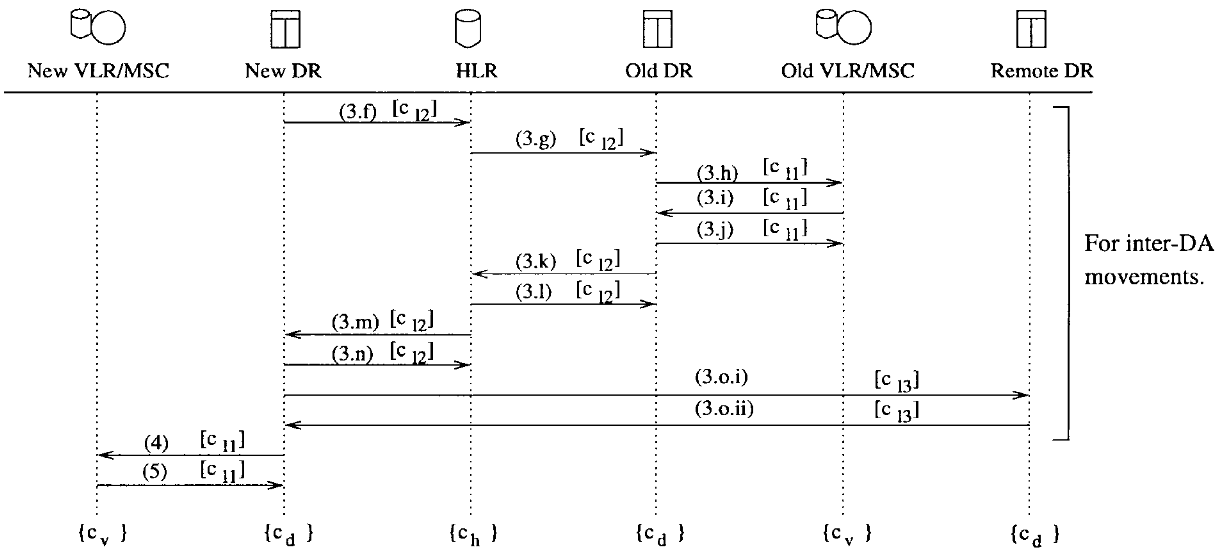


Fig. 9. Location registration under new database architecture (continued).

- ii) the MSC assigns a TLDN to the called MT and sends this TLDN to the HLR.

otherwise:

- iii) the HLR sends a location request message to the serving DR of the called MT;
- iv) the called DR forwards the location request to the serving MSC of the called MT;
- v) the MSC assigns a TLDN to the called MT and sends this TLDN to the HLR.

j) The HLR forwards the TLDN to the calling DR.

4) The calling DR forwards the TLDN to the calling MSC.

5) The calling MSC sets up a connection to the called MSC using this TLDN.

The signaling and database access overhead for call delivery is reduced if a remote pointer exists at the DR of the calling MT. However, additional processing is required after each movement to update the remote pointers. Similarly, the signaling and database access overhead for location registration can be reduced if location change is reported to the HLR only after inter-DA movements. However, this will increase the database access overhead for call delivery as an additional DR access is necessary to locate a called MT. There are tradeoffs among different pointer configurations and we will introduce a method for determining the location information distribution strategy for a particular MT in Section IV-B.

B. Per-User Location Information Distribution Strategy

The location information distribution strategy for each MT specifies the pointer configuration for the MT at the remote DR's and the location information available at the HLR. The DR periodically computes the distribution strategy for each of its associated MT's based on their mobility and call arrival parameters. This subsection describes a method for determining the per-user distribution strategy which significantly reduces the signaling and database access costs for location management.

For a target MT, we define λ_c, $1/\lambda_m$, and q to be the incoming call arrival rate, the average RA residence time, and the inter-DA movement probability (this is the probability that a movement to another RA will result in departure from the current residing DA), respectively. The values of $1/\lambda_m$ and q describe the frequency and the locality of the movements, respectively. For example, a vehicle in an interstate highway usually travels at a relatively high speed in a constant direction and thus the values for both λ_m and q are large. On the other hand, a pedestrian travels slowly and the movements are generally confined to a specific area. In this case, the values for λ_m and q are relatively small.

In order to determine the location information distribution strategy for a particular MT, a scheme for deriving the values of λ_c, λ_m, and q for the MT is needed. One possible method to derive these parameters is to implement three counters in the MT's user profile which record the number of call arrivals to the MT, the number of inter-RA movements, and the number of inter-DA movements, respectively. A counter is increased by one when the corresponding event occurs (e.g., the call arrival counter is increased by one when an incoming call arrives) and it is reset to 0 at specific time intervals (e.g., once per hour). We call this the measurement intervals. The counter values obtained at the end of a measurement interval are used to predict the parameters for the next measurement interval. For example, assuming that the call arrival counter recorded a value of 5 at the end of a one-hour measurement interval. An incoming call arrival rate, λ_c, of 5 calls/h will be assumed for the next measurement interval. The measurement interval should coincide with the time interval for updating the pointer configurations of the MT's such that the pointers are generated based on the most up-to-date parameters. The advantage of this method is that it is simple to implement and it does not significantly increase the processing load at the MSC. The disadvantage of this method is that the behavior of the MT may change significantly and this method cannot capture the change until the end of the current measurement interval.

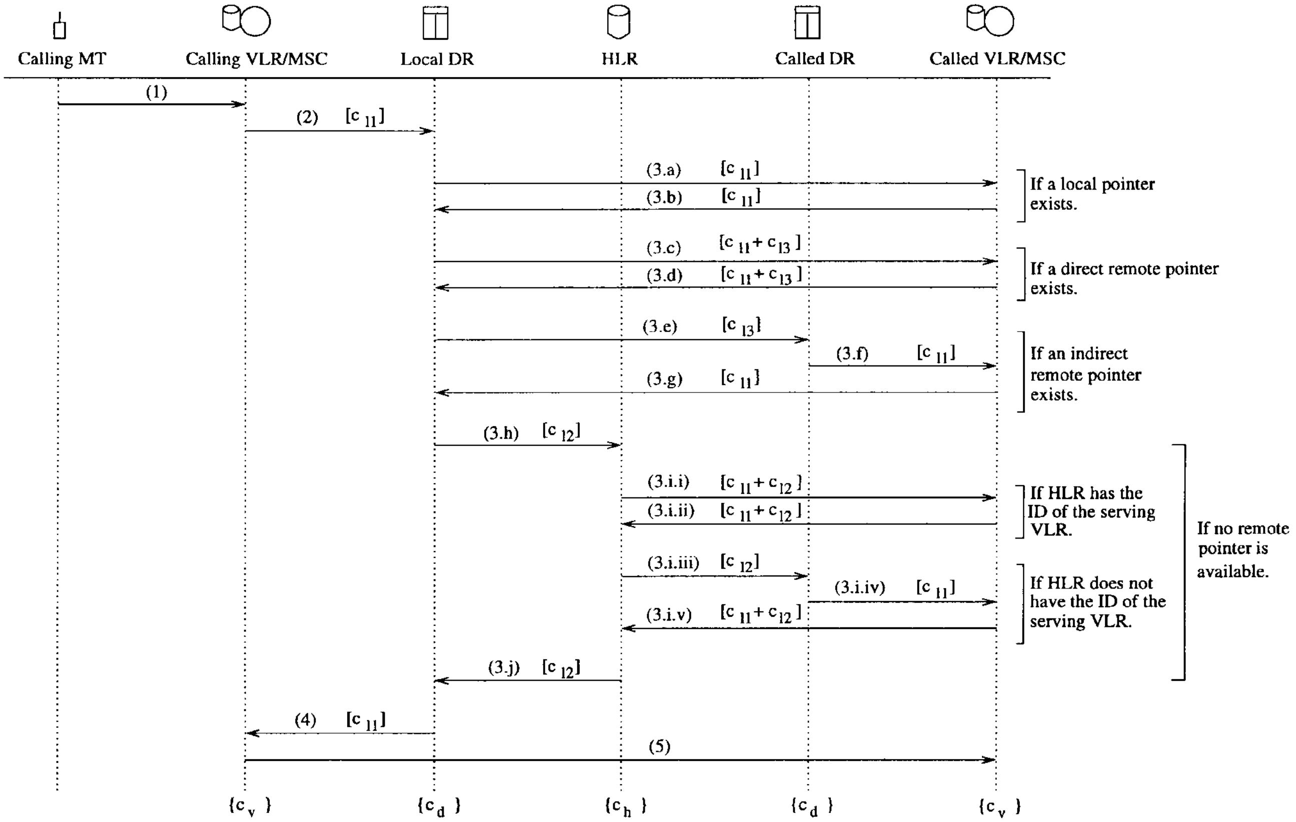


Fig. 10. Call delivery under new database architecture.

An alternative method for determining the values of λ_c, λ_m, and q is based on the historical data of the particular MT. In general, the mobility and calling patterns of an MT do not vary significantly from day to day. As a result, the mobility and calling data obtained during a given time interval in the previous day can be used to predict the behavior of the MT during the same time interval in the current day. This method is accurate for the majority of the MT's as most subscribers exhibit similar mobility and calling patterns from day to day. A disadvantage of this method is that the network must record the previous day's data for each MT. This will increase the disk space requirement at the HLR.

In this paper, we will present and analyze the proposed location management scheme assuming that the values for the incoming call arrival rate, λ_c, the average RA residence time, $1/\lambda_m$, and the inter-DA movement probability, q, are readily available.

As described in Section IV-A, location management involves the exchange of signaling messages among the network elements. The costs for location registration and call delivery depend on the relative locations of the network elements involved as well as the overhead for accessing the databases. Here we define a number of parameters which represent the costs for location registration and call delivery under various situations.

α_1	The cost for call delivery if a direct remote pointer for the called MT is available at the calling DR.
α_2	The cost for call delivery if an indirect remote pointer for the called MT is available at the calling DR.
α_3	The cost for call delivery if no remote pointer is available and the HLR records the ID of the serving DR of the MT.
α_4	The cost for call delivery if no remote pointer is available and the HLR records the ID of the serving MSC of the MT.
α_5	The cost for call delivery if the called MT is residing in the same DA.
β_1	The cost for location registration after an intra-DA movement if reporting of location change to the HLR is required.
β_2	The cost for location registration after an intra-DA movement if reporting of location change to the HLR is not required.
β_3	The cost for location registration after an inter-DA movement.
γ	The cost for updating a remote pointer.

Note that the location registration costs β_i $(i = 1, 2, 3)$ do not include the cost for updating remote pointers because the number of remote pointers is not known at this time. The cost for updating a particular remote pointer is given by γ. For simplicity, we assume that γ does not depend on the location of the remote DR. The above parameters are used to determine the location information distribution strategy for each MT. The derivation of these parameters are described in more detail in Section IV-C.

Three pointer configurations are possible for a given MT at a remote DR:

P1) a direct remote pointer for the MT is set up at the remote DR;

P2) an indirect remote pointer for the MT is set up at the remote DR;

P3) no remote pointer for the MT is set up at the remote DR.

The call delivery cost is lower when a remote pointer for the called MT is available at the DR associated with the calling MT. However, the pointer must be updated when the called MT moves to another location. This results in additional signaling and database access cost. For each remote DR, we select the pointer configuration for the MT that results in the lowest total cost for call delivery and pointer update. Let θ_i^k ($i = 1, 2, 3$) be the cost per unit time for delivering calls from DA k to the MT if pointer configuration Pi is used

$$\theta_i^k = \lambda_c^k \alpha_i \qquad \text{for } i = 1, 2, 3 \tag{1}$$

where λ_c^k is the arrival rate of calls originating from DR k and destined for the MT. Let ϕ_i ($i = 1, 2, 3$) be the cost per unit time for updating the MT's remote pointer at a DR if pointer configuration Pi is used at this DR

$$\phi_i = \begin{cases} \lambda_m \gamma & \text{if } i = 1 \\ q\lambda_m \gamma & \text{if } i = 2 \\ 0 & \text{if } i = 3. \end{cases} \tag{2}$$

Let μ_i^k ($i = 1, 2, 3$) be the total cost per unit time for call delivery and remote pointer update for the MT with respect to remote DR k when pointer configuration Pi is used. The expression for μ_i^k is

$$\mu_i^k = \theta_i^k + \phi_i \qquad \text{for } i = 1, 2, 3. \tag{3}$$

To minimize the location management cost for the MT with respect to DR k, we select the pointer configuration at DR k which results in the smallest total cost, μ_i^k. For example, if μ_1^k is the smallest among μ_1^k, μ_2^k, and μ_3^k, then pointer configuration P1 is used, and so on.

Assume S is the set of remote DR's that contain direct or indirect remote pointers for the MT as determined by the above procedure. Let λ_l be the arrival rate of calls from the residing DA of the MT and λ_h be the arrival rate of calls from remote DAs which do not contain remote pointers for the MT. The expression for λ_h is

$$\lambda_h = \lambda_c - \lambda_l - \sum_{y \in S} \lambda_c^y. \tag{4}$$

Delivering calls from DAs where remote pointers are not available requires HLR queries. Therefore, λ_h can also be considered as the rate of HLR query due to call delivery.

As described in Section IV, two HLR configurations are possible:

H1) the HLR records the ID of the serving MSC of the MT;

H2) the HLR records the ID of the serving DR of the MT.

If configuration H1 is selected, the cost for call delivery is lower as one less DR query is required to locate the MT. However, the DR must report each location change of the MT to the HLR. This results in higher location registration cost. Let π_1 and π_2 be the location registration costs per unit time for configurations H1 and H2, respectively

$$\pi_1 = \lambda_m[(1-q)\beta_1 + q\beta_3] \tag{5}$$

$$\pi_2 = \lambda_m[(1-q)\beta_2 + q\beta_3]. \tag{6}$$

For calls originating from DAs where pointers for the MT are not available, let ϵ_1 and ϵ_2 be the call delivery costs per unit time under configurations H1 and H2, respectively

$$\epsilon_1 = \lambda_h \alpha_4 \tag{7}$$

$$\epsilon_2 = \lambda_h \alpha_3. \tag{8}$$

The total cost per unit time for location registration and call delivery (for calls that requires HLR queries) under configurations H1 and H2 are

$$\rho_1 = \pi_1 + \epsilon_1 \tag{9}$$

and

$$\rho_2 = \pi_2 + \epsilon_2 \tag{10}$$

respectively. To minimize the overall cost for location registration and call delivery, configuration H1 is selected if $\rho_1 < \rho_2$. Otherwise, configuration H2 is used.

Here is a summary of the steps for determining the location information distribution strategy for an MT.

1) For each remote DR, compute the average total call delivery and pointer update cost for the three pointer configurations, P1, P2, and P3, using (3). Select the pointer configuration that results in the lowest average total call delivery and pointer update cost.
2) Determine the rate of HLR query, λ_h, using (4). Then, compute the average total location registration and call delivery cost (for calls that require HLR query) under HLR configurations H1 and H2 using (9) and (10), respectively.
3) Select the HLR configuration that results in lower location registration and call delivery cost.

C. Performance Analysis

For the analysis given in this paper, we assume the cost for database access and the cost for transmitting signaling messages are available. For PCS networks, the delays for connection establishment, location registration, and call delivery are considered to be some of the most important performance criteria. A user's perception on the quality of the network is greatly affected by connection set up delay. In this analysis, we assume the database access and signaling costs are measured by the delays required to complete the database update/query and signal transmission, respectively. These delay values can be obtained by on-line measurements, if such measurement functions are available in the particular network, or by a table lookup process. Given the particular time of the day, the table provides the access cost for a specified database or the signaling cost for a specified communication link. These values are then used for determining the location information distribution strategy for the MT's. The table is located at each DR and it should be updated frequently to reflect the status of the network.

TABLE I
EXPRESSIONS FOR LOCATION REGISTRATION AND CALL DELIVERY COSTS

Cost Parameter	Expression
α_1	$2c_v + c_d + 4c_{l1} + 2c_{l3}$
α_2	$2c_v + 2c_d + 4c_{l1} + 2c_{l3}$
α_3	$2c_v + 2c_d + c_h + 4(c_{l1} + c_{l2})$
α_4	$2c_v + c_d + c_h + 4(c_{l1} + c_{l2})$
α_5	$2c_v + c_d + 4c_{l1}$
β_1	$2c_v + c_d + c_h + 6c_{l1} + 2c_{l2}$
β_2	$2c_v + c_d + 6c_{l1}$
β_3	$2c_v + 2c_d + c_h + 6(c_{l1} + c_{l2})$
γ	$c_d + 2c_{l3}$

To simplify the analysis, we assume the costs for a database update is equal to the cost for a database query. We define c_v, c_h, and c_d to be the costs for updating or querying the VLR, the HLR, and the DR, respectively. We also define c_{l1}, c_{l2}, and c_{l3} to be the costs for sending a signaling message between a DR and one of its associated MSC's, between an HLR and a DR and between two DR's, respectively. These signaling costs include the switching cost at the sender and at the intermediate STP's, and the cost for transmitting the signaling message through the communication links.

Based on the procedures for location registration and call delivery as described in Section IV-A, the expressions for the call delivery costs, α_i ($i = 1, \cdots, 5$), the location registration costs, β_j ($j = 1, 2, 3$), and the remote pointer update cost, γ, as described in Section IV-B are given in Table I. In Table I, we assume that when a user profile is retrieved from the database, it is temporarily stored at a cache memory for the duration of the location update or call delivery. As a result, no database access is necessary for subsequent queries to the same user profile. In addition, updates to the database are recorded at the cache memory and they are written to the database only at the end of the location update or call delivery procedure. Consequently, a maximum of one update or query is performed at each database involved in a location update or call delivery. Furthermore, in Table I, we assume that signaling messages between the HLR and an MSC and between a remote DR and an MSC are routed through the DR serving this particular MSC. The cost for sending these signaling messages are, therefore, $c_{l1} + c_{l2}$ and $c_{l1} + c_{l3}$.

For example, delivering a call to an MT given that a direct remote pointer is available at the local DR requires one query at the calling VLR (c_v), one query at the DR (c_d), and one query at the called VLR (c_v). The total database access cost is $2c_v + c_d$. In addition, delivering this call requires two message exchanges between a DR and a local MSC ($2c_{l1}$), and two message exchanges between a DR and a remote MSC ($2c_{l1}+2c_{l3}$). The total signaling cost is $4c_{l1}+2c_{l3}$. As a result, the expression for α_1 given in Table I is $2c_v+c_d+4c_{l1}+2c_{l3}$. In another example, performing a location update after an intra-DA movement where reporting to the HLR is necessary requires one update at each of the following databases: the new VLR (c_v), the DR (c_d), the HLR (c_h), and the old VLR (c_v). The total database access cost is $2c_v + c_d + c_h$. In addition, this type of location update requires three message exchanges between the DR and the new VLR ($3c_{l1}$), three message exchanges between the DR and the old VLR ($3c_{l1}$), and two message exchanges between the DR and the HLR ($2c_{l2}$). The total signaling cost is $6c_{l1} + 2c_{l2}$. As a result, the expression for β_1 given in Table I is $2c_v+c_d+c_h+6c_{l1}+2c_{l2}$.

According to the per-user location information distribution strategy, given that a remote pointer exists at DR k, the cost per unit time for delivering calls from DA k and for updating the remote pointer at DR k is

$$\mu^k = \min\left(\mu_1^k, \mu_2^k\right) \tag{11}$$

where μ_1^k and μ_2^k can be computed by (3). The cost per unit time for location registration and for delivering calls that require HLR queries is

$$\rho = \min(\rho_1, \rho_2) \tag{12}$$

where ρ_1 and ρ_2 are given by (9) and (10), respectively. The cost per unit time for delivering calls originating from the local DA is $\lambda_l\alpha_5$. As a result, the total cost per unit time for location registration, remote pointer update, and call delivery is

$$C = \rho + \lambda_l\alpha_5 + \sum_{k\in S} \mu^k \tag{13}$$

where S is the set of remote DR's that contain remote pointers for the MT.

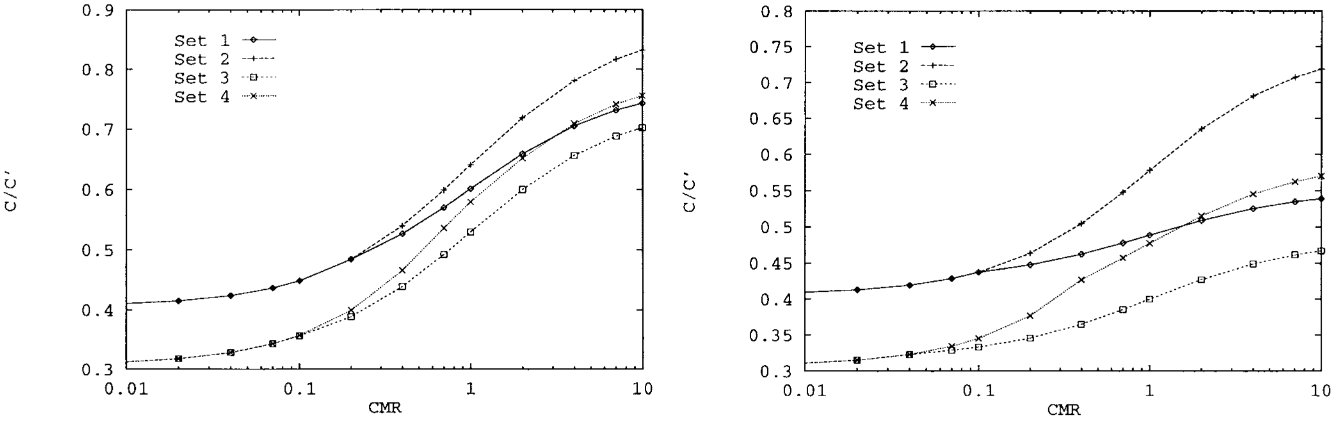
In the following, we evaluate the performance of the proposed location management scheme by comparing with the IS-41 standard under various mobility, call arrival, and cost parameters. Based on the IS-41 location registration and call delivery procedures as described in Section III, the expression for the total cost per unit time for location registration and call delivery is

$$C' = (\lambda_m + \lambda_c)[4(c_{l1} + c_{l2}) + 2c_v + c_h + c_g] \tag{14}$$

where c_g is the cost for performing global title translation (GTT). In (14), we assume that the cost of a signaling message exchange between an MSC and an STP is equal to the cost of a signaling message exchange between an MSC and its DR, c_{l1}. Similarly, the cost of a signaling message exchange between an STP and the HLR is equal to the cost of a signaling message exchange between a DR and the HLR, c_{l2}.

We define the relative cost of the proposed location management scheme as the ratio of the total cost per unit time for the proposed scheme to that of the IS-41 standard, C/C'. A relative cost of 1 means that the costs under both schemes are the same. We also define the call-to-mobility ratio (CMR) as the ratio of the call arrival rate to the mobility rate; in other words $CMR = \lambda_c/\lambda_m$.

In the following analysis, we assume that the inter-DA movement probability, q, is 0.125. This value is selected based on the assumption that there are 8×8 RA's arranged in a square in each DA and the random walk mobility model is used (see Section V for more details). To simplify the analysis, we carry out the evaluation assuming that either the signaling



(a) (b)

Fig. 11. Signaling cost for (a) $v_l = v_r = 0.15$ and (b) $v_l = v_r = 0.3$.

TABLE II
SIGNALING COST PARAMETERS

Set	c_{l1}	c_{l2}	c_{l3}
1	1	5	2
2	1	5	10
3	1	10	2
4	1	10	10

or the database access cost dominates. When the signaling cost dominates, the database access cost is assumed to be negligible, and vice versa.

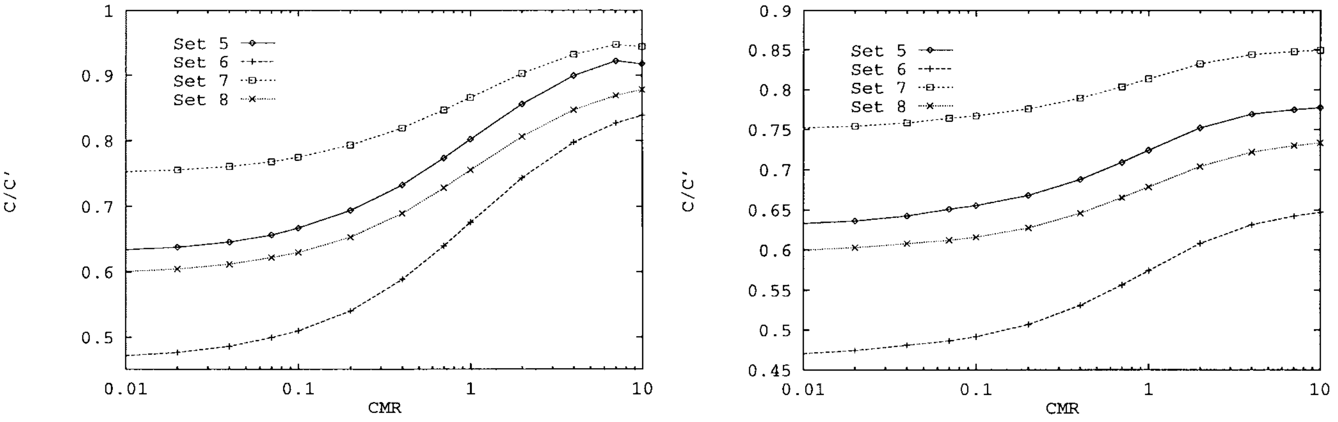
We first evaluate the case when the signaling cost dominates by setting the database access cost parameters, c_v, c_h, and c_d, to 0. We consider the four sets of signaling cost parameters given in Table II. Since the DR is located close to its associated MSC's, signaling messages between a DR and one of its MSC's can be exchanged without going through intermediate STP's. As a result, the signaling cost between the DR and the local MSC, c_{l1}, is expected to have the smallest value. In this study, all signaling cost parameters are normalized to c_{l1} such that $c_{l1} = 1$. The selected data sets allow us to study the effect of varying the cost parameters c_{l2} and c_{l3} on the performance of the location management scheme. Let v_l and v_r be the fractions of the incoming calls for the target MT that are originated from the local DA and from remote DA k, respectively. We assume that the origination of the remaining calls are evenly distributed over all other DAs such that the call arrival rate from each of these DAs is small. As a result, no remote pointers are set up in these remaining DAs based on the proposed location management scheme.

Fig. 11(a) and (b) shows the relative signaling cost for $v_l = v_r = 0.15$ and $v_l = v_r = 0.3$, respectively, as the value of CMR varies from 0.01–10. In general, the relative cost increases with CMR. When the CMR is low, the mobility rate is high compared to the call arrival rate and the cost for location registration dominates. Significant cost saving can be obtained by reporting location change to the HLR only after an inter-DA movement. When the CMR is high, the mobility rate is low relative to the call arrival rate and the cost saving from location registration diminishes. However, when CMR is high, the call delivery cost dominates and the saving in call delivery cost becomes significant. This saving is attributed to the lower cost for delivering calls from the local DA and from DA k where a remote pointer is set up. This cost saving increases and thus lowers the relative cost as the fractions v_l and v_r increase.

Comparisons between data sets 1 and 2 and between data sets 3 and 4 demonstrate that the effect of varying the signaling cost between two DR's, c_{l3}, is significant only when CMR is high. As described before, at low CMR, the location registration cost dominates. As no signaling messages are sent directly between two DR's during location registration, the effect of varying c_{l3} is limited. At high CMR, a smaller c_{l3} values results in lower cost for delivering a call from DA k where a remote pointer is available. As a result, the relative signaling cost is lower when c_{l3} is smaller.

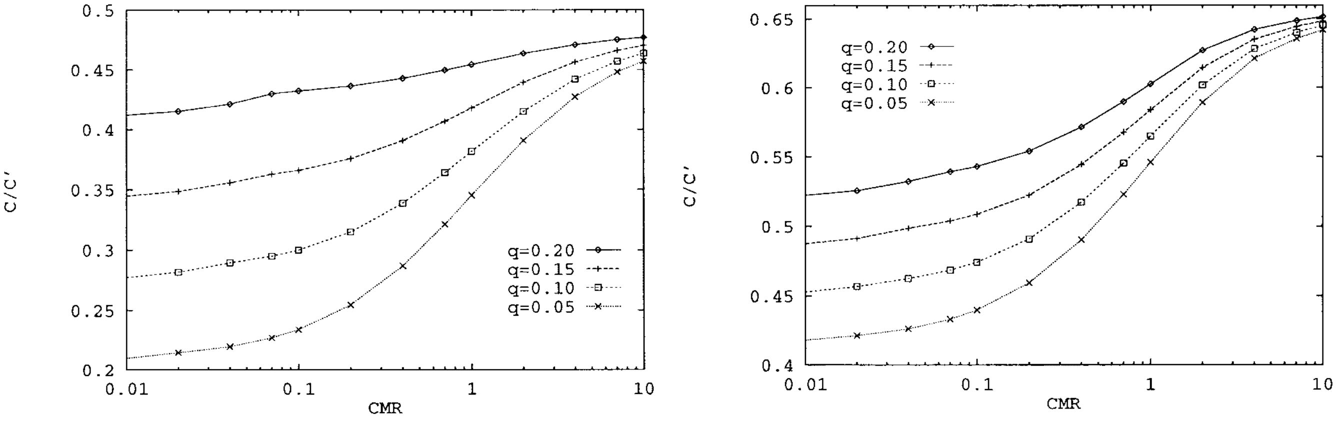
Comparisons between data sets 1 and 3 and between data sets 2 and 4 demonstrate that an increase in the signaling cost between the DR and HLR, c_{l2}, results in lower relative cost under all CMR values. Under the proposed location registration and call delivery schemes, cost saving is achieved by reducing the number of relatively expensive HLR accesses. As a result, increasing the signaling cost between the DR and the HLR increases the cost saving when the proposed scheme is used and thus lowers the relative cost.

In the following analysis, we consider the situations when the database access cost dominates by setting the signaling cost parameters, c_{l1}, c_{l2}, and c_{l3}, to 0. We consider the four sets of signaling cost parameters given in Table III. Since a DR access involves a simple table lookup or update procedure, the DR access cost, c_d, is expected to have the lowest value. Therefore, all cost parameters are normalized to c_d such that $c_d = 1$. The selected data sets allow us to study the effect of varying the cost parameters c_v and c_h on the performance of the location management scheme. In this analysis, we assume the cost for GTT, c_g, is 50% of the DR access cost such that $c_g = 0.5c_d$. The is reasonable as GTT is a simple mapping procedure which generates the address of the HLR given the mobile identification number, the cost for GTT is expected to be low compared to the cost for accessing the DR.



(a) (b)

Fig. 12. Database access cost for (a) $v_l = v_r = 0.15$ and (b) $v_l = v_r = 0.3$.



(a) (b)

Fig. 13. (a) Signaling and (b) database access costs for data sets 3 and 6 at different inter-DA movement probability, q.

TABLE III
DATABASE ACCESS COST PARAMETERS

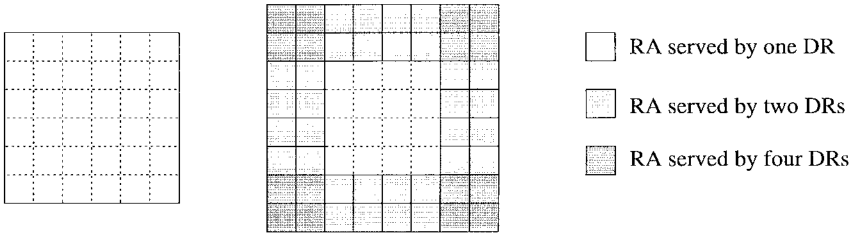
Set	c_v	c_h	c_d
5	3	6	1
6	3	12	1
7	6	6	1
8	6	12	1

Fig. 12(a) and (b) shows the relative database access cost for $v_l = v_r = 0.15$ and $v_l = v_r = 0.3$, respectively, as the value of CMR varies from 0.01–10. The relative cost increases as the CMR increases and the relative cost decreases as the fractions v_l and v_r increase. These can be explained using the same reasoning as in the case when the signaling cost dominates. Comparisons between data sets 5 and 6 and between data sets 7 and 8 demonstrate that an increase in the HLR access cost, c_h, results in lower relative cost under all CMR values. As described before, the proposed location registration and call delivery schemes achieve cost saving by reducing the number of relatively expensive HLR accesses. As a result, increasing the HLR access cost increases the cost saving and thus lowers the relative cost.

Comparisons between data sets 5 and 7 and between data sets 6 and 8 demonstrate that an increase in the VLR access cost, c_v, results in higher relative cost under all CMR values. As the number of VLR accesses is the same under both the proposed scheme and the IS-41 standard, a higher VLR access cost diminishes the overall cost saving. As a result, the relative cost becomes higher as the VLR access cost increases.

Fig. 13 shows the relative signaling and database access costs for data sets 3 and 6 (as given in Tables II and III, respectively) for $v_l = v_r = 0.3$ when different inter-DA movement probabilities, q, are used. These two data sets are selected because they result in the lowest signaling and database access costs, respectively (see Figs. 11 and 12). We investigate the effect of varying the value of q on the relative signaling and database access costs.

A smaller inter-DA movement probability results in a smaller number of HLR and indirect remote pointer updates and thus lower relative costs. It can be seen in Fig. 13 that as q decreases, both the relative signaling cost and the relative database access cost decrease. For high CMR values, the inter-DA movement rate is low even for large q values and the cost saving obtained by reducing q is not significant. However, for low CMR values, the inter-DA movement rate is relatively high. Lowering the value of q results in significant reduction



(a) (b) (c)

Fig. 14. (a) Directory area with $k = 6$. (b) Overlapped directory area with $k = 6$ and $n = 2$.

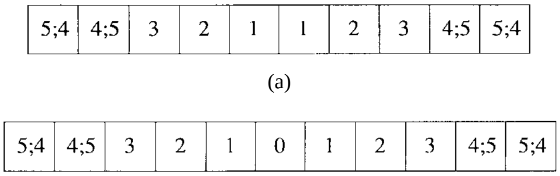
in both the relative signaling and the relative database access costs.

The performance analysis given in this section demonstrated that the proposed location management scheme results in significant saving in both the signaling and database access costs for location registration and call delivery. Based on the cost parameters considered in this analysis, the signaling and database access cost can be reduced by as much as 70 and 50%, respectively, compared to the IS-41 standard. These cost savings depend on the mobility and call arrival parameters of the target MT. However, as long as the cost for accessing the HLR is relatively high, which is usually the case, the cost saving is significant.

V. Directory Area Overlapping

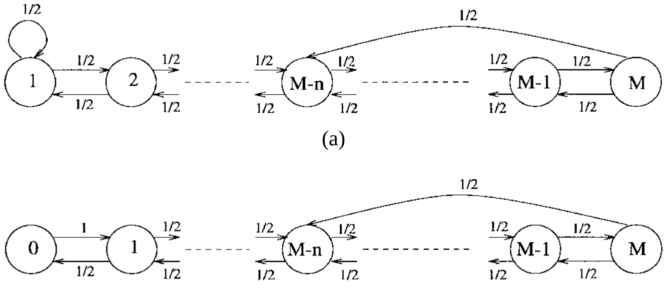
It is demonstrated in Section IV-C that the overhead of the proposed location management scheme depends on the inter-DA movement probability q. A small q value results in low signaling and database access overhead and vice versa. One method for lowering the probability q, and thus the location management cost, is to increase the size of each DA. However, this increases the number of subscribers served by the DR and, in turn, increases processing load and the memory requirement at the DR. Here we introduce a method called *directory area overlapping*. Given a uniform distribution of MT's throughout the PCS service area, this method allows increased DA size while keeping the average number of subscribers per DR unchanged. Assume that RA's are square shaped and there are $k \times k$ RA's per DA in the original system where DA overlapping is not applied. Fig. 14(a) shows an example DA with $k = 6$. If DA-overlapping is used, we extend the dimension of each DA by n such that each DA contains $(k+n) \times (k+n)$ RA's and the outermost n layers of RA's overlap with that of the neighboring DAs. Fig. 14(b) shows an example of DA-overlapping with $k = 6$ and $n = 2$. It can be seen that each DA consists of 8 × 8 RA's. Under this arrangement, 16 RA's are served solely by the original DR, 32 RA's are served by two DR's and 16 RA's are served by four DR's.

Using this method, the inter-DA movement probability, q, decreases because the size of each DA is increased. Since each MT is served by one DR at a time, for a uniform distribution of mobile subscribers, the average number of MT's served by each DR is the same as the case when directory area overlapping is not used. In addition, an MT is located at an inner portion of a new DA as soon as it leaves the previous DA.



(a) (b)

Fig. 15. One-dimensional mobility model for $n = 2$: (a) $k = 8$ and (b) $k = 9$.



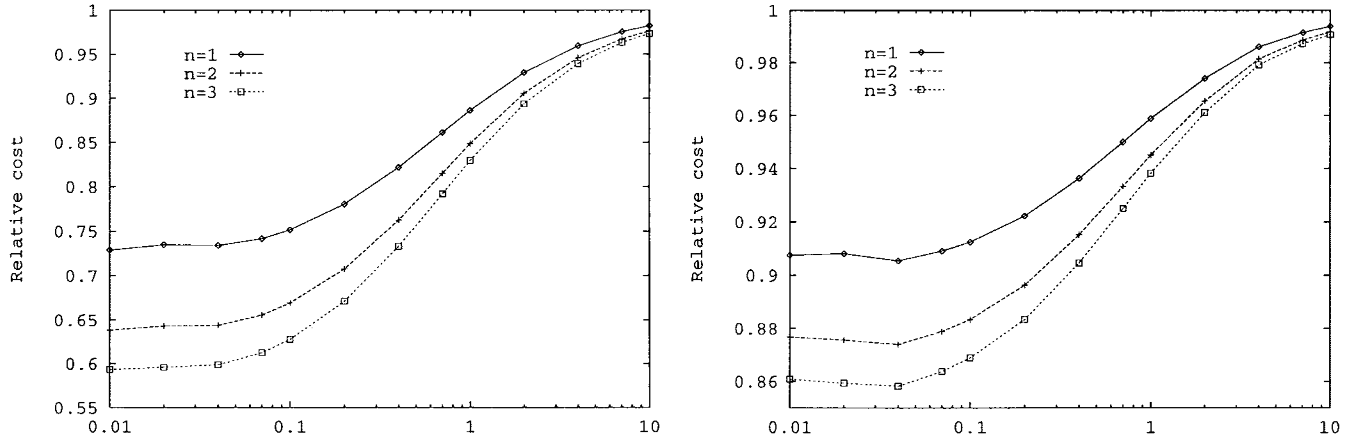
(a) (b)

Fig. 16. State transition diagram for (a) even and (b) odd DA dimensions.

This significantly reduces the number of inter-DA movements, especially for MT's located close to the DA boundaries making frequent movements between the two DAs.

We evaluate the performance of DA overlapping assuming that the MT's travel according to a two-dimensional random walk model with square shaped RA's. Based on this assumption, when an MT moves from one RA to another RA, there is an equal probability, i.e., 1/4, that any one of the four neighboring RA's is selected as the destination.

To determine the inter-DA movement probability q, we first consider a one-dimensional random walk model and then generalize the result for the two-dimensional model. For the one-dimension model, each RA has two neighbors and the probability that the MT will move to one of these neighbors during its next movement is 1/2. For $n = 2$, Fig. 15(a) and (b) shows the one-dimensional service areas for $k = 8$ and $k = 9$, respectively. If $k+n$ is even [see Fig. 15(a)], we label the center two RA's by 1. Otherwise, if $k+n$ is odd [see Fig. 15(b)], we labeled the center RA by 0. The numbering increases toward the edge of the DA. Since n RA's at each edge of the DA belongs to two DAs, they are assigned two numbers each with respect to one of its associated DAs. We then set up the state transition diagrams for even and odd $k+n$ values as shown in Fig. 16. Here, the state is defined as the



(a) (b)

Fig. 17. Relative (a) signaling cost and (b) database access cost for $k = 8$ and $n = 1, 2, 3$.

label of the current RA with respect to the current serving DR of the MT. Let p_i be the steady state probability of state i and let M be the highest state of the state diagrams such that $M = \lfloor (k+n)/2 \rfloor$. The balance equations of the state transition diagrams for $n < M - 1$ and $M \geq 2$ are[2]

$$p_i = \tfrac{1}{2} p_{i-1} + \tfrac{1}{2} p_{i+1} \qquad \text{for } 1 < i < M \text{ and } i \neq M - n \tag{15}$$

$$p_M = \begin{cases} p_{M-1} & \text{for } n = 0 \\ \tfrac{1}{2} p_{M-1} & \text{for } n > 0 \end{cases} \tag{16}$$

$$p_{M-n} = \tfrac{1}{2} p_M + \tfrac{1}{2} p_{M-n-1} + \tfrac{1}{2} p_{M-n+1} \tag{17}$$

$$p_1 = \begin{cases} p_2 & \text{for } k+n \text{ is even} \\ p_0 + \tfrac{1}{2} p_2 & \text{for } k+n \text{ is odd} \end{cases} \tag{18}$$

$$p_0 = \begin{cases} 0 & \text{for } k+n \text{ is even} \\ \tfrac{1}{2} p_1 & \text{for } k+n \text{ is odd.} \end{cases} \tag{19}$$

The steady state probabilities, p_i $(i = 0 \cdots M)$, can be solved numerically using the above balance equations. The inter-DA movement probability for the one-dimensional model, denoted by q', can then be obtained as

$$q' = \tfrac{1}{2} p_M. \tag{20}$$

To compute the inter-DA movement probability for the two-dimensional model q, we break down the movements into two categories: left–right movements and up–down movements. Given that a left–right movement is performed by the MT, the probability that the movement results in a DA boundary crossing can be obtained using the one-dimensional model as $q_{lr} = q'$. Similarly, given that an up–down movement is performed by the MT, the probability that the movement results in a DA boundary crossing is $q_{ud} = q'$. Based on the two-dimensional random walk model, half of the movements are left–right movements and the other half up–down movements. As a result, the inter-DA movement probability for the two-dimensional model, denoted by q, is

$$q = \tfrac{1}{2} q_{lr} + \tfrac{1}{2} q_{ud} = q'. \tag{21}$$

TABLE IV
INTER-DA MOVEMENT PROBABILITY, q, FOR $k = 8, 15$ AND $n = 0, \cdots, 3$

k	$n = 0$	$n = 1$	$n = 2$	$n = 3$
8	0.1250	0.0625	0.0417	0.0313
15	0.0667	0.0333	0.0222	0.0167

This demonstrates that the inter-DA movement probability for the two-dimensional model is the same as that for the one-dimensional model.

Table IV shows the value of q for $k = 8$ and $k = 15$ while the value of n varies from 0–3. It can be seen that q decreases significantly when DA overlapping is used. The value of q decreases by 50% when n is increased from 0 to 1. Therefore, a small degree of overlapping achieves a significant reduction in the inter-DA movement probability.

Fig. 17(a) shows the signaling cost for $n = 1, 2, 3$ relative to the signaling cost when $n = 0$. Similarly, Fig. 17(b) shows the database access cost for $n = 1, 2, 3$ relative to the database access cost when $n = 0$. The signaling and database access costs at a given n value is obtained using (13). A relative cost value of 1 means that there is no cost advantage using DA-overlapping. In Fig. 17, we assume $k = 8$ and $v_l = v_r = 0.3$. Data set 3 from Table II is used in Fig. 17(a) and data set 6 from Table II is used in Fig. 17(b). It can be seen in Fig. 17 that the relative signaling cost and the database access cost can be as low as 60% and 85%, respectively, compared to the costs when DA-overlapping is not used. The cost reduction using DA-overlapping is due to the lowering of the inter-DA movement probability such that the number of relatively expensive HLR accesses is reduced. When CMR is high, the mobility and the inter-DA movement rates are relatively low. The cost associated with inter-DA movements do not account for a significant fraction of the total cost, so the saving due to DA-overlapping is limited. When CMR is low, the mobility and the inter-DA movement rates are high and the saving due to DA-overlapping is larger. The most significant cost reduction resulted from incrementing n by 1 occurs when n is increased from 0 to 1. This suggests that a small degree of overlapping is sufficient to benefit from DA-overlapping.

---

[2] The balance equations for the boundary cases can be set up accordingly and are not presented here because of space limitations.

VI. CONCLUSIONS

In this paper, we introduced a location management scheme based on a three-level database architecture. This new database architecture consists of the HLR and the VLR's in the original IS-41 standard as well as a new type of database called the directory register (DR). The DR's determine the location information distribution strategy for each of their associated MT's and store location information of local and selected remote mobile terminals (MT's). The distribution strategy for each MT is unique and is determined based on its mobility and call arrival parameters. We introduced an algorithm for determining the per-user location information distribution strategy which significantly reduces the signaling and database access costs for location management. The cost reduction is most significant when the call-to-mobility ratio is low and the cost for accessing the HLR is high. It is also demonstrated that the cost for location management is smaller when the inter-directory area (DA) movement is low. An enhancement to the proposed location management scheme called directory area overlapping is then introduced. This scheme lowers the inter-DA movement probability and thus results in additional reduction in the signaling and database access costs.

The proposed location management scheme relies on the newly introduced directory registers to distribute the location information of the MT's. All additional processing, such as the computation of the location information distribution strategies, are handled by the DR's. Besides, the proposed scheme is completely distributed. The distribution strategy for each MT can be computed independently at the local DR using information available at the serving VLR. As a result, no centralized control is necessary.

REFERENCES

[1] EIA/TIA, "Cellular radio-telecommunications intersystem operations," Tech. Rep. IS-41 Revision B, EIA/TIA, 1991.

[2] I. F. Akyildiz and J. S. M. Ho, "On location management for personal communications networks," *IEEE Commun. Mag.*, vol. 34, pp. 138–145, Sept. 1996.

[3] V. Anantharam, M. L. Honig, U. Madhow, and V. K. Wei, "Optimization of a database hierarchy for mobility tracking in a personal communications network," *Performance Evaluation*, vol. 20, no. 1–3, pp. 287–300, May 1994.

[4] B. R. Badrinath, T. Imielinski, and A Virmani, "Locating strategies for personal communication network," in *Proc. Workshop Networking Personal Commun. Appl.*, Dec. 1992.

[5] R. Jain and Y. B. Lin, "An auxiliary user location strategy employing forwarding pointers to reduce network impact of PCS," *ACM-Baltzer J. Wireless Networks*, vol. 1, no. 2, pp. 197–210, July 1995.

[6] R. Jain, Y. B. Lin, and S. Mohan, "A caching strategy to reduce network impacts of PCS," *IEEE J. Select. Areas Commun.*, vol. 12, Oct. 1994.

[7] Y. B. Lin, "Determining the user locations for personal communications services networks," *IEEE Trans. Veh. Technol.*, vol. 43, pp. 466–473, Aug. 1994.

[8] Y. B. Lin and S. K. DeVries, "PCS network signaling using SS7," *IEEE Personal Commun. Mag.*, vol. 2, pp. 44–55, June 1995.

[9] K. Meier-Hellstern and E. Alonso, "The use of SS7 and GSM to support high density personal communications," in *Proc. IEEE ICC*, 1992, pp. 1698–1720.

[10] C. N. Lo, R. S. Wolff, and R. C. Bernhardt, "Expected network database transaction volume to support personal communications services," in *Proc. 1st Int. Conf. Universal Personal Commun.*, Sept. 1992, pp. 09.03/1–6.

[11] S. Mohan and R. Jain, "Two user location strategies for personal communications services," *IEEE Personal Commun. Mag.*, vol. 1, pp. 42–50, First Quarter 1994.

[12] M. Mouly and M. B. Pautet, "The GSM system for mobile communications," M. Mouly, 49 rue Louise Bruneau, Palaiseau, France, 1992.

[13] N. Shivakumar and J. Widom, "User profile replication for faster location lookup in mobile environments," in *Proc. ACM MOBICOM'95*, Nov. 1995, pp. 161–169.

[14] 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.

Joseph S. M. Ho (S'93–M'96) received the B.S.E.E. and M.S.E.E. degrees from the University of Washington at Seattle in 1987 and 1989, respectively, and the Ph.D. degree in electrical and computer engineering from the Georgia Institute of Technology in 1996.

He is currently a Member of the Scientific Staff at Northern Telecom. His current research interests include design and analysis of PCS networks, wireless ATM networks, satellite based PCS networks, mobile computing, and performance evaluation. He is a member of Tau Beta Pi.

Ian F. Akyildiz (F'96) received the B.S., M.S., and Ph.D. degrees in computer engineering from the University of Erlangen-Nuernberg, Germany, in 1978, 1981, and 1984, respectively.

Currently, he is a Full Professor with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta. His current research interests are in ATM networks, wireless networks, and satellite communication.

Dr. Akyildiz is an Editor for IEEE/ACM TRANSACTIONS ON NETWORKING, *Computer Networks and ISDN Systems Journal*, *ACM-Baltzer Journal of Wireless Networks*, *ACM-Springer Journal of Multimedia Systems*, and, in 1992–1996, for IEEE TRANSACTIONS ON COMPUTERS. He served as the Program Chair for the ACM/IEEE Mobicom'96 conference and is the Program Chair for the IEEE INFOCOM'98 conference. He is an ACM Fellow and is a National Lecturer for ACM. He received the "Don Federico Santa Maria Medal" for his services to the Universidad de Federico Santa Maria in Chile. He is listed in *Who's Who in the World* (Platinum Edition). He received the ACM Outstanding Distinguished Lecturer Award for 1994 and the IEEE Leonard G. Abraham Price Award for best paper published in IEEE JOURNAL OF SELECTED AREAS IN COMUNICATIONS in 1996.