JOURNAL OF INFORMATION SCIENCE AND ENGINEERING 18, 1011-1025 (2002)

Short Paper

A Forwarding Pointer-Based Cache Scheme for Reducing Location Management Cost in PCS Networks

KI-SIK KONG, JOON-MIN GIL*, YOUN-HEE HAN+, UI-SUNG SONG
AND CHONG-SUN HWANG
Distributed Systems Lab.
Department of Computer Science and Engineering
**Institute of Basic Science*
Korea University
SungBuk-Gu, Seoul 136-701, Republic of Korea
+*Samsung Advanced Institute of Technology*
Republic of Korea

This paper proposes a forwarding pointer-based cache scheme (PB-Cache scheme) that can reduce the signaling cost for location management in PCS networks. In the existing cache scheme, the use of cache information can effectively reduce the signaling traffic for locating frequently called mobile users. However, when the cache information is obsolete, it results in much more signaling traffic than that of the IS-41. In order to solve this problem, we propose a new location cache scheme called the PB-Cache scheme, which exploits a user's movement locality as well as call locality. Even if the cached information is not up-to-date, the called user can be found by tracing forwarding pointers starting from that VLR pointed in the cache instead of querying the HLR. Thus, the PB-Cache scheme can effectively reduce the frequent access to the HLR and the signaling traffic for location management. Moreover, it distributes the signaling and database access load on the HLR to the VLR's. Analytical results indicate that the PB-Cache scheme significantly outperforms the other schemes when a user's call-to-mobility ratio is high or the signaling traffic to the HLR is heavy.

Keywords: location management, forwarding pointer, cache, location registration, call delivery, personal communications service (PCS)

1. INTRODUCTION

Personal Communications Service (PCS) networks provide wireless services to subscribers that are free to travel, and the network access point of a mobile terminal (MT's) changes as it moves around the network coverage area. A location management scheme, therefore, is necessary to effectively keep track of the MT's and locate a called MT when a call is initiated [1]. There are two commonly used standards for location management: IS-41 [2-6] and GSM [3, 7]. Both are based on a two-level database hierarchy, which consists of Home Location Register (HLR) and Visitor Location Registers (VLR's).

The whole network coverage area is divided into cells. There is a Base Station (BS) installed in each cell and these cells are grouped together to form a larger area called a Registration Area (RA). All BS's belonging to one RA are wired to a Mobile Switching

Received September 28, 2001; accepted April 15, 2002.
Communicated by Jang-Ping Sheu, Makato Takizawa and Myongsoon Park.

Center (MSC). The MSC/VLR is connected to the Local Signal Transfer Point (LSTP) through the local A-link, while the LSTP is connected to the Regional Signal Transfer Point (RSTP) through the D-link. The RSTP is, in turn, connected to the HLR through the remote A-link [8, 9].

Location management is one of the most important issues in PCS networks. As the number of MT's increases, location management scheme under the IS-41 has gone through many problems such as increasing traffic in network, bottleneck to the HLR, and so on. To overcome these problems under the IS-41, a number of works have been reported. A cache scheme [10] was proposed to reduce the signaling cost for call delivery by reusing the cached information about the called MT's location from the previous call. When a call arrives, the location of the called MT is first identified in the cache instead of sending query messages to the VLR. When a cache hit occurs, the cache scheme can save one query to the HLR and traffic along some of the signaling links as compared to the IS-41. This is especially significant when the call-to-mobility ratio (CMR) of the MT is high. However, a penalty has to be paid when there is "location miss" since the cache information is not always up-to-date. In this paper, we propose an enhanced cache scheme called PB-Cache scheme. By exploiting a user's movement locality as well as call locality at the same time, this scheme can reduce the access to the HLR and the signaling traffic for location management throughout the networks.

This paper is organized as follows: Section 2 introduces a new location cache scheme employing forwarding pointers. An analytical model for comparing location management cost is given in section 3. Section 4 shows an analysis of the location management cost under the PB-Cache scheme. Finally, conclusions are given in section 5.

2. A FORWARDING POINTER-BASED CACHE SCHEME

Fig. 1 shows an illustrative example of locating the MT under the PB-Cache scheme. The VLR_0 represents the calling VLR associated with the RA the caller resides. We assume that the cache information exists in the MSC_0 and the cache entry for the called MT (MT_1) currently points to the VLR_1. Let's consider that the MT_1 has moved from the RA associated with VLR_1 to the RA associated with VLR_2 after the last call arrived. Then, the current location of the MT_1 is the RA associated with the VLR_2. When the next call arrives, the MSC_0 first queries the pointed VLR, that is, VLR_1. In this case, since the existing cache scheme has to perform the call delivery procedure of the IS-41 after an unsuccessful query for the cache, it results in the waste of the signaling traffic as compared to the IS-41. However, under the PB-Cache scheme, even if the cache information is obsolete, it traces the pointer chain without querying the HLR until the current location of the called MT is found within the maximum pointer chain length of *K* (see Table 1). So, the saving of one query to the HLR and traffic along some of the signaling links can be obtained. Note that the pointer chain length has to be limited due to the maximum pointer setup delay requirement. In Fig. 1, we assume that the maximum pointer chain length, denoted by *K*, is preconfigured to be one. Therefore, unless the MT_1 moves into the RA associated with the VLR_3, it can be found through both cache and forwarding pointer information without querying the HLR. In the following, more detailed procedures for the PB-Cache scheme are described.

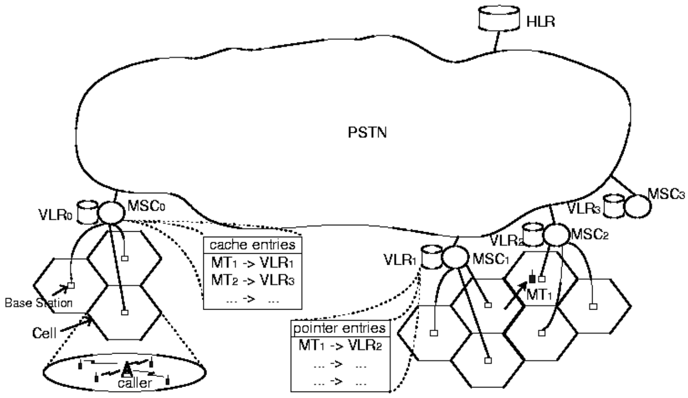


Fig. 1. An example of locating an MT under the PB-Cache scheme.

2.1 Location Registration

Compared with the IS-41, most procedures for location registration under the PB-Cache scheme are exactly the same as those of the IS-41 except that the forwarding pointer is additionally set up between the two VLR's.

2.2 Call Delivery

Most procedures for call delivery under the PB-Cache scheme are almost the same as those of the cache scheme except that the called MT is traced through the pointer chain length of *K*. When the cache information is obsolete, the pointer chain is traced to find a called MT starting from the VLR pointed in the cache. If the called MT is located within the pointer chain length of *K* from that VLR, it can be found without querying the HLR.

Fig. 2 describes the procedures for a cache hit under the PB-Cache scheme. The cache hit under the PB-Cache scheme contains two situations: One is the situation that the cache information is correct. Thus, the called MT is found after the only one query to the pointed VLR. The other is the situation that the cache information is not correct. In this case, after querying the pointed VLR, the called MT is found by tracing through the pointer chain length of *K* (See step 3 in Fig. 2). The cache miss under the PB-Cache scheme occurs when the called MT is not found even if the forwarding pointer chain has been traced until the length of *K*. After this, the same call delivery procedure as that of the IS-41 is performed. In this case, the current location of the called MT is transmitted from the HLR to the calling VLR together with the cache update request message.

3. PERFORMANCE MODELING

3.1 Basic Assumptions and Parameters

The basic assumptions for the performance modeling are as follows:

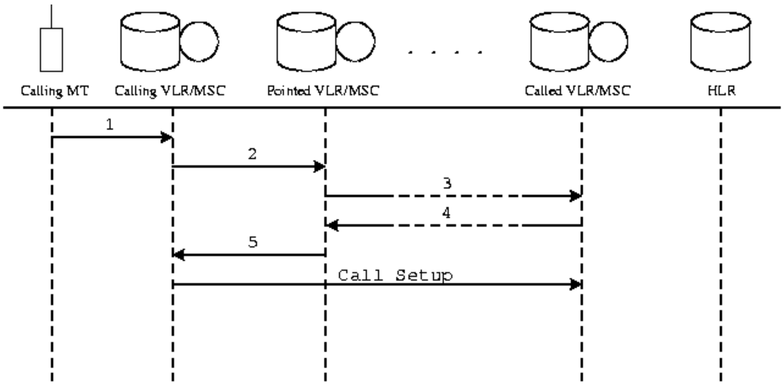


Fig. 2. Call delivery under the PB-Cache scheme (cache hit).

- The call arrivals to an MT follows a Poisson distribution with mean λ_c.
- The RA residence time of an MT follows a Gamma distribution with mean $\frac{1}{\lambda_m}$, where λ_m is the movement rate.
- We use the extended meaning of the cache hit ratio under the PB-Cache scheme. In other words, if the MT can be found within the pointer chain length of K even if the cache information is obsolete, we include this case with a cache hit. Therefore, $p = \overline{p} + \hat{p}$ (see Table 1).

Table 1 shows the costs and parameters used for the performance modeling.

Table 1. Costs and parameters.

Parameter	Description
C_{la}	Cost for sending a signaling message through the local A-link
C_d	Cost for sending a signaling message through the D-link
C_{ra}	Cost for sending a signaling message through the remote A-link
C_v	Cost for a query or an update of the VLR
C_h	Cost for a query or an update of the HLR
t	Probability (VLR pointed in the calling MSC's cache is the VLR within the same LSTP area)
m	Probability (Any movement or searching by a user is within its current LSTP area)
j	Length of pointer chain traced until the MT is found ($0 < j \leq K$)
K	Maximum allowable length of pointer chain which can be traced to locate the MT
q	MT's CMR (call-to-mobility ratio: CMR)
p	Cache hit ratio under the PB-Cache scheme
$\overline{p}$	Cache hit ratio under the cache scheme
$\hat{p}$	Probability (MT is found by tracing forwarding pointers when the cache information is not correct).

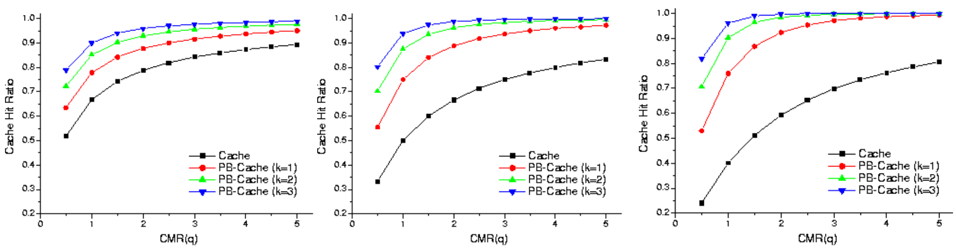
3.2 Comparison of the Cache Hit Ratios

The cache hit ratio is determined by a called MT's CMR. The CMR is defined as the expected number of calls to a called MT from a given originating MSC during the period that the called MT visits an RA. Thus, the CMR can be denoted by $\frac{\lambda_c}{\lambda_m}$. We assume that the RA residence time follows a Gamma distribution. The Gamma distribution is selected for its flexibility and generality. By setting appropriate parameters, a Gamma distribution can be also used to represent the distribution for a set of measured data. The Laplace transform, $f_m^*(s)$, of a Gamma distribution [11, 12] with mean $\frac{1}{\lambda_m}$ and variance V is

$$f_m^*(s) = \left(\frac{\lambda_m \gamma}{s + \lambda_m \gamma}\right), \qquad \gamma = \frac{1}{V{\lambda_m}^2} \tag{1}$$

In order to simplify the analysis, we will first set $\gamma = 1$ such that the RA residence time of an MT follows an exponential distribution (see Fig. 3(b)). Thus, it can be expressed as $f_m^*(\lambda_c) = \frac{\lambda_m}{\lambda_c + \lambda_m}$. Then, we will investigate the effect of the variance of the RA residence time on the performance of the PB-Cache scheme by setting different γ values.

The cache hit ratio under the PB-Cache scheme can be obtained by the sum of the probabilities that the called MT resides in each RA between two consecutive phone calls. In the following, $\alpha(K)$ is the probability that the called MT moves across K RA's between two consecutive phone calls [13-15]. Thus, we can derive the cache hit ratios under both the PB-Cache scheme and the cache scheme from $\alpha(K)$.



(a) $5/{\lambda_m}^2$ (b) $1/{\lambda_m}^2$ (c) $1/5{\lambda_m}^2$

Fig. 3. Cache hit ratios under the different RA residence time distributions.

$$\alpha(K) = \begin{cases} 1 - \frac{1}{q}\left[1 - f_m^*(\lambda_c)\right] & (K = 0) \\ \frac{1}{q}\left[1 - f_m^*(\lambda_c)\right]^2 \left[f_m^*(\lambda_c)\right]^{K-1} & (K > 0) \end{cases} \tag{2}$$

Using (2), $p, \overline{p}$, and $\hat{p}$ can be shown to be

$$\overline{p} = \alpha(0) \tag{3}$$

$$\hat{p} = \alpha(1) + \alpha(2) + \cdots + \alpha(K) = \sum_{i=1}^{K} \alpha(i) \tag{4}$$

$$p = \overline{p} + \hat{p} = \alpha(0) + \alpha(1) + \alpha(2) \cdots + \alpha(K) = \sum_{i=0}^{K} \alpha(i) \tag{5}$$

Equation (3) indicates the cache hit ratio of the cache scheme, and (5) indicates the cache hit ratio of the PB-Cache scheme. As mentioned above, (4) means the probability that the MT is found by tracing forwarding pointers when the cache information is not up-to-date. Fig. 3 shows the cache hit ratios under the different RA residence time distributions. This indicates that the cache hit ratios of the PB-Cache scheme are much higher than that of the cache scheme. Note that the cache hit ratios of the PB-Cache scheme have little effect on RA residence time variance, while the cache hit ratio of the cache scheme has much effect on it. Using the higher cache hit ratio, the PB-Cache scheme can reduce the access to the HLR and the signaling traffic for location management throughout the networks.

3.3 Analytical Model for Location Management Cost

In PCS networks, the delays for location registration and call delivery are considered to be some of the most important performance criteria. In our analytical model, the signaling and database access costs are used to measure the performance of both the PB-Cache scheme and the cache scheme. These costs are associated with the delays needed to perform the signal transmission and database update/query. For the analytical model, we also assume that there are q call arrivals for an MT between two RA crossings. Based on these assumptions, we derive location management cost for the IS-41, the cache scheme, and the PB-Cache scheme, respectively.

3.3.1 Location management cost under the IS-41

The signaling cost for location registration (UNC_{IS}) and the signaling cost for call delivery (SNC_{IS}) are $4(C_{la} + C_d + C_{ra})$, respectively. Then, the total signaling cost (NC_{IS}) is

$$NC_{IS} = UNC_{IS} + qSNC_{IS} = 4(1 + q)(C_{la} + C_d + C_{ra}) \tag{6}$$

The database access cost for location registration (UDC_{IS}) and the database access cost for call delivery (SDC_{IS}) are $2C_v + C_h$, respectively. Then, the total database access cost (DC_{IS}) is

$$DC_{IS} = UDC_{IS} + qSDC_{IS} = 2C_v + C_h + q(2\,C_v + C_h) \tag{7}$$

From (6) and (7), the total cost (TC_{IS}) is expressed as

$$TC_{IS} = NC_{IS} + DC_{IS} = 4(1 + q)(C_{la} + C_d + C_{ra}) + 2C_v + C_h + q(2C_v + C_h) \tag{8}$$

3.3.2 Location management cost under the cache scheme

The signaling cost for location registration under the cache scheme is the same as that of the IS-41. Therefore, the signaling cost for location registration (UNC_{Ca}) is

$$UNC_{Ca} = UNC_{IS} = 4(C_{la} + C_d + C_{ra}) \tag{9}$$

The signaling cost for call delivery (SNC_{Ca}) is expressed as

$$SNC_{Ca} = \overline{p}SNC_{CaH} + (1-\overline{p})SNC_{CaM} \tag{10}$$

where SNC_{CaH} and SNC_{CaM} are the signaling costs for call delivery when a cache hit and a cache miss occur, respectively.

$$SNC_{CaH} = 4C_{la} + 4(1 - t)\, C_d = 4[C_{la} + (1 - t)C_d] \tag{11}$$

$$SNC_{CaM} = 4[C_{la} + (1 - t)\, C_d] + 4(C_{la} + C_d + C_{ra}) \tag{12}$$

Replacing SNC_{CaH} and SNC_{CaM} by (11) and (12), (10) is re-written as

$$\begin{aligned} SNC_{Ca} &= \overline{p}SNC_{CaH} + (1-\overline{p})SNC_{CaM} \\ &= 4[C_{la} + (1-t)C_d] + 4(C_{la} + C_d + C_{ra})(1-\overline{p}) \end{aligned} \tag{13}$$

Then, total signaling cost (NC_{Ca}) is

$$\begin{aligned} NC_{Ca} &= UNC_{Ca} + qSNC_{Ca} \\ &= 4(C_{la} + C_d + C_{ra}) + q\{4[C_{la} + (1-t)C_d] + 4(C_{la} + C_d + C_{ra})(1-\overline{p})\} \end{aligned} \tag{14}$$

The database access cost for location registration (UDC_{Ca}) is

$$UDC_{Ca} = UDC_{IS} = 2C_v + C_h \tag{15}$$

The database access cost for call delivery (SDC_{Ca}) is expressed as

$$SDC_{Ca} = \overline{p}SDC_{CaH} + (1-\overline{p})SDC_{CaM} \tag{16}$$

where SDC_{CaH} and SDC_{CaM} are the database access costs for call delivery when a cache hit and a cache miss occur, respectively.

$$SDC_{CaH} = 2C_v \tag{17}$$

$$SDC_{CaM} = 3C_v + C_h \tag{18}$$

Replacing SDC_{CaH} and SDC_{CaM} by (17) and (18), (16) is now given by

$$SDC_{Ca} = \overline{p}SDC_{CaH} + (1-\overline{p})SDC_{CaM} = 3C_v + C_h - \overline{p}(C_v + C_h) \tag{19}$$

Then, the total database access cost (DC_{Ca}) is

$$DC_{Ca} = UDC_{Ca} + qSDC_{Ca} = 2C_v + C_h + q[3C_v + C_h - \overline{p}(C_v + C_h)] \tag{20}$$

From (14) and (20), the total cost (TC_{Ca}) is expressed as

$$\begin{aligned} TC_{Ca} &= NC_{Ca} + DC_{Ca} \\ &= 4(C_{la} + C_d + C_{ra}) + q\{4[C_{la} + (1-t)C_d] + 4(C_{la} + C_d + C_{ra})(1-\overline{p})\} \\ &\quad + 2C_v + C_h + q[3C_v + C_h - \overline{p}(C_v + C_h)] \end{aligned} \tag{21}$$

3.3.3 Location management cost under the PB-cache scheme

The signaling cost for location registration (UNC_{PB}) is

$$UNC_{PB} = UNC_{IS} + S_N = 4(C_{la} + C_d + C_{ra}) + 4[C_{la} + (1-m)C_d] \tag{22}$$

where S_N is the additional signaling cost for setting up forwarding pointer between the VLR's. The signaling cost for call delivery (SNC_{PB}) is expressed as

$$SNC_{PB} = pSNC_{PBH} + (1-p)SNC_{PBM} \tag{23}$$

where SNC_{PBH} and SNC_{PBM} are the signaling costs for call delivery when a cache hit and a cache miss occur, respectively. To quantify j, we assume that on average $j = \frac{K}{2}$. Therefore, the signaling cost for call delivery when a cache hit occurs is

$$SNC_{PBH} = 4[C_{la} + (1-t)C_d] + 2K[C_{la} + (1-m)C_d] \tag{24}$$

The signaling cost for call delivery when a cache miss occurs is

$$SNC_{PBM} = 4[C_{la} + (1-t)C_d] + 4K[C_{la} + (1-m)C_d] + 4(C_{la} + C_d + C_{ra}) \tag{25}$$

Replacing SNC_{PBH} and SNC_{PBM} by (24) and (25), (23) now becomes

$$\begin{aligned} SNC_{PB} &= pSNC_{PBH} + (1-p)SNC_{PBM} \\ &= 4(C_{la} + C_d + C_{ra})(1-p) + 2K(2-p)[C_{la} + (1-m)C_d] \\ &\quad + 4[C_{la} + (1-t)C_d] \end{aligned} \tag{26}$$

Then, the total signaling cost (NC_{PB}) is

$$\begin{aligned} NC_{PB} &= UNC_{PB} + qSNC_{PB} \\ &= 4(C_{la} + C_d + C_{ra}) + 4[C_{la} + (1-m)C_d] + q\{4(C_{la} + C_d + C_{ra})(1-p) \\ &\quad + 2K(2-p)[C_{la} + (1-m)C_d] + 4[C_{la} + (1-t)C_d]\} \end{aligned} \tag{27}$$

The database access cost for location registration (UDC_{PB}) is

$$UDC_{PB} = UDC_{IS} + S_D = 4C_v + C_h \tag{28}$$

where S_D is the additional database access cost for setting up forwarding pointer between the VLR's.

The database access cost for call delivery (SDC_{PB}) is expressed as

$$SDC_{PB} = pSDC_{PBH} + (1 - p)SDC_{PBM} \tag{29}$$

where SDC_{PBH} and SDC_{PBM} are the database access costs for call delivery when a cache hit and a cache miss occur, respectively. Let $j = \frac{K}{2}$, the database access cost for call delivery when a cache hit occurs is

$$SDC_{PBH} = (\frac{K}{2} + 2)C_v \tag{30}$$

The database access cost for call delivery when a cache miss occurs is

$$SDC_{PBM} = (K + 2)C_v + C_h + C_v = (K + 3)C_v + C_h \tag{31}$$

Replacing SDC_{PBH} and SDC_{PBM} by (30) and (31), (29) now becomes

$$\begin{aligned} SDC_{PB} &= pSDC_{PBH} + (1-p)SDC_{PBM} \\ &= (K+3)C_v + C_h - p[(\frac{K}{2}+1)C_v + C_h]\} \end{aligned} \tag{32}$$

Then, the total database access cost (DC_{PB}) is

$$\begin{aligned} DC_{PB} &= UDC_{PB} + qSDC_{PB} \\ &= 4C_v + C_h + q\{(K+3)C_v + C_h - p[(\frac{K}{2}+1)C_v + C_h]\} \end{aligned} \tag{33}$$

From (27) and (33), the total cost (TC_{PB}) is expressed as

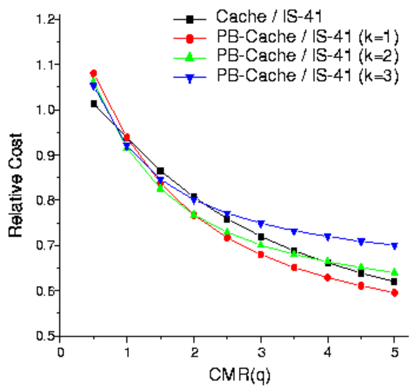
$$\begin{aligned} TC_{PB} &= NC_{PB} + DC_{PB} \\ &= 4(C_{la} + C_d + C_{ra}) + 4[C_{la} + (1-m)C_d] + q\{4(C_{la} + C_d + C_{ra})(1-p) \\ &\quad + 2K(2-p)[C_{la} + (1-m)C_d] + 4[C_{la} + (1-t)C_d]\} + 4C_v + C_h \\ &\quad + q\{(K+3)C_v + C_h - p[(\frac{K}{2}+1)C_v + C_h]\} \end{aligned} \tag{34}$$

4. PERFORMANCE ANALYSIS

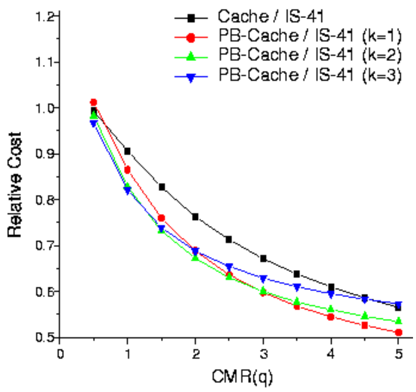
We evaluate the performance of the PB-Cache scheme by comparing it with the IS-41 and the cache scheme based on the analytical model described in section 3. Similar to [4, 6], we define the relative cost of the PB-Cache scheme as the ratio of the total cost for the PB-Cache scheme to that of the IS-41. A relative cost of 1 means that the costs under both schemes are exactly the same. In order to estimate *m*, we use the model of the called MT's mobility across LSTP areas (see [16] for more details). Then, the probability of intra-LSTP movement can be estimated to be approximately 0.87, and *t* is assumed to be 0.2.

4.1 Signaling Cost

We first evaluate the case of a dominating signaling cost by setting the database access cost parameters, C_v and C_h to 0. Parameter sets 1 and 2 show the cases when the cost for sending a message to the HLR is relatively low. Parameter sets 3 and 4 show the cases when the cost for sending a message to the HLR is relatively high. As the number of the mobile users keeps increasing rapidly, parameter set 4 is especially expected to be the common case. Fig. 4 shows the relative signaling cost for both the cache scheme and the PB-Cache scheme when the parameter sets 2 and 4, as given in Table 2(a), are used. We can see that the PB-Cache scheme for $K = 1$, on the whole, results in the lowest signaling cost as compared with other schemes.



(a) set 2



(b) set 4

Fig. 4. Signaling cost.

4.2 Database Access Cost

In the following, we evaluate the case when the database access cost dominates by setting the signaling cost parameters, C_{la}, C_d, and C_{ra} to 0. Fig. 5 shows the relative database access cost for both the cache scheme and the PB-Cache scheme when the parameter sets 6 and 7, as given in Table 2(b), are used. Note that the situation in Fig. 5(b) offsets the additional database access cost incurred by tracing more VLR queries against

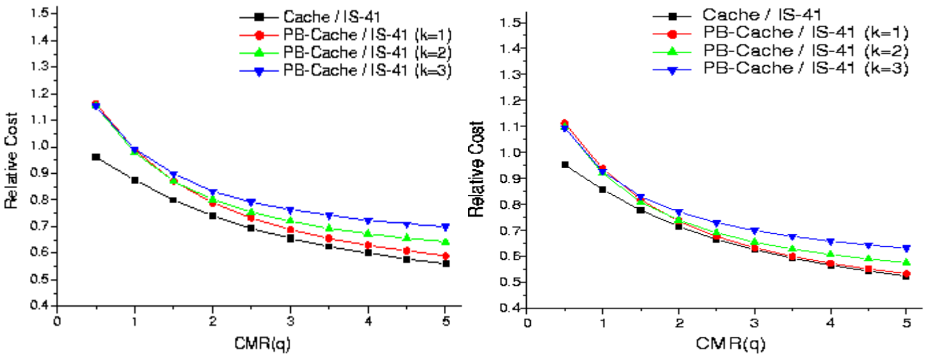
Table 2. Cost parameter sets.

set	C_{la}	C_d	C_{ra}
1	1	3	3
2	1	3	5
3	1	5	7
4	1	5	10

(a) Signaling cost

set	C_v	C_h
5	1	3
6	1	4
7	1	5

(b) Database access cost



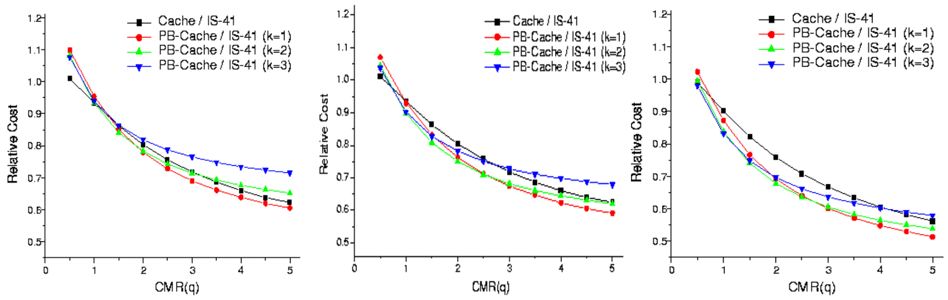
(a) set 6 (b) set 7

Fig. 5. Database access cost.

fewer HLR queries, and this will be more apparent as the cost for accessing the HLR becomes higher.

4.3 Total Cost

We compare the total cost of the PB-Cache scheme with that of the IS-41 and the cache scheme. Fig. 6 demonstrates the relative total cost of the cache scheme and the PB-Cache scheme. We can see that the PB-Cache scheme for $K = 1$, on the whole, results in the lowest total cost compared with other schemes. As we can see in Fig. 6, for high CMR, the reduction in the total cost of the PB-Cache scheme is very significant when the cost for accessing to the HLR is relatively high (See Fig. 6(c)). However, when C_{ra} and C_h are relatively low (See Fig. 6(a)), the reduction in the total cost of the PB-Cache scheme for $K = 1$ is almost the same as that of the cache scheme. These results may be expected because the PB-Cache scheme tries to reduce the number of messages going to the HLR by searching the MT based on the VLR pointed in the cache. Thus, the decision of the appropriate maximum pointer chain length is essential to improve the performance of the PB-Cache scheme. Note that the pointer chain length, K can be configured and changeable according to the user's CMR and the signaling traffic condition.



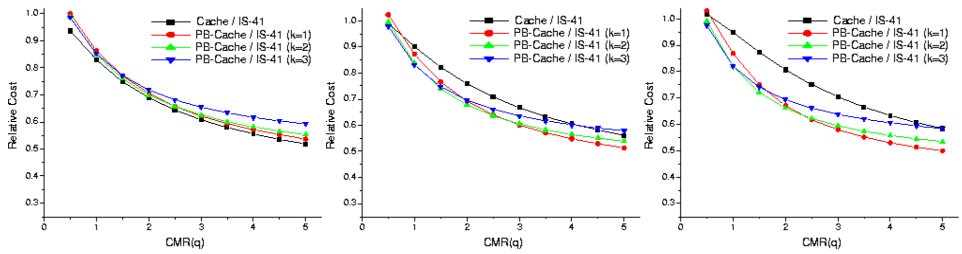
(a) set 2 and set 5 (b) set 3 and set 6 (c) set 4 and set 7

Fig. 6. Total cost.

4.4 Effect of RA Residence Time Variance on the Total Cost

We now investigate the effect of RA residence time variance on the total cost of the PB-Cache scheme. The Laplace transform of a Gamma distribution is given by (1). Since the variance of a Gamma distribution is $\frac{1}{\gamma \lambda_m^2}$, we adjust the variance of the RA residence time by changing the value of γ. A small value of γ results in high variance and the RA residence time may deviate from its average value significantly. In contrast, a large value of γ results in low variance and the RA residence time stays close to its mean value most of time. Fig. 7 shows the effect of RA residence time variance on the total cost of the PB-Cache scheme for parameter sets 4 and 7 as given in Table 2.

The results shown in Fig. 7 indicate that the PB-Cache scheme has little effect on RA residence time variance, while the cache scheme has much effect on it. This is due to the effect of the cache hit ratios under the different RA residence time distributions (See Fig. 3). Note that the cache hit ratios of the PB-Cache scheme have little effect on it, while the cache hit ratio of the cache scheme has much effect on it. In other words, when the RA residence time variance is low, the cache hit ratio under the cache scheme decreases, and vice versa. In Fig. 7 (c), the total cost of the PB-Cache scheme, therefore, becomes relatively lower than that of the cache scheme when the RA residence time variance is low.



(a) $5/\lambda_m^2$ (b) $1/\lambda_m^2$ (c) $1/5\lambda_m^2$

Fig. 7. Total cost of the PB-Cache scheme under gamma RA residence time distributions with different variance values.

5. CONCLUSIONS

In this paper, a location management scheme called the PB-Cache scheme was proposed. The PB-Cache scheme is an enhanced variation of the alternative location algorithm proposed in [10]. In short, the primary idea is to exploit a user's movement locality as well as call locality at call delivery. In the PB-Cache scheme, the information of the VLR pointed in the cache is used as a hint of the called MT's location information, and even if it is not correct, the called MT can ultimately be found by tracing forwarding pointers starting from that VLR pointed in the cache instead of querying the HLR. By reducing the access to the HLR, the PB-Cache scheme, in most cases, results in a significant reduction in the total location management cost compared with other schemes, and distributes the signaling traffic and database access load on the HLR to the VLR's, i.e, the bottleneck in the HLR can be effectively reduced. The analytical results have indicated that the PB-Cache scheme significantly outperforms IS-41 and the existing cache scheme when the CMR is high or the signaling traffic to the HLR is heavy.

REFERENCES

1. F. Akyildiz et al., "Mobility management in next-generation wireless systems," in *Proceedings of IEEE*, Vol. 87, 1999, pp. 1347-1384.
2. EIA/TIA, "Cellular radio-telecommunications intersystem operations," Technical Report IS-41 Revision B, EIA/TIA, 1991.
3. S. Mohan and R. Jain, "Two user location strategies for personal communications services," *IEEE Personal Communications*, 1994, pp. 42-50.
4. R. Jain and Y. -B. Lin, "Local anchor scheme for reducing signaling cost in personal communication networks," *IEEE/ACM Transactions on Networking*, Vol. 4, 1996, pp. 709-726.
5. I. R. Chen, T. M. Chen, and C. Lee, "Agent-based forwarding strategies for reducing location management cost in mobile networks," *ACM/Baltzer Journal on Mobile Networks and Applications* (*MONET*), Vol. 6, 2001, pp. 103-113.
6. J. S. M. Ho and I. F. Akyildiz, "Dynamic hierarchical database architecture for location management in PCS networks," *IEEE/ACM Transactions on Networking*, Vol. 5, 1997, pp. 646-661.
7. M. Mouly and M. B. Pautet, *The GSM System for Mobile Communications*, 1992.
8. D. R. Wilson, "Signaling system no.7, IS-41 and cellular telephony networking," in *Proceedings of IEEE*, Vol. 80, 1992, pp. 664-652.
9. G. P. Pollini, K. S. Meier-Hellstern, and D. J. Goodman, "Signaling traffic volume generated by mobile and personal communications," *IEEE Communications Magazine*, 1995, pp. 60-65.
10. R. Jain, Y. -B. Lin, C. Lo, and S. Mohan, "A caching strategy to reduce network impacts of PCS," *IEEE Journal on Selected Areas in Communications*, 1994, pp. 1434-1445.
11. Y. Fang, I. Chlamtac, and Y. -B. Lin, "Portable movement modeling for PCS networks," *IEEE Transactions on Vehicular Technology*, Vol. 49, 2000, pp. 1356-1363.
12. Y. -B. Lin, "Performance analysis for dual band PCS networks," *IEEE Transactions*

on Computers, Vol. 49, 2000, pp. 148-159.
13. Y. -B. Lin, "Modeling techniques for large-scale PCS networks," *IEEE Communications Magazine*, 1997, pp. 102-107.
14. Y. -B. Lin, "Reducing location update cost in a PCS network," *IEEE/ACM Transactions on Networking*, Vol. 5, 1997, pp. 25-33.
15. J. Li, H. Kameda, and K. Li, "Optimal dynamic mobility management for PCS networks," *IEEE/ACM Transactions on Networking*, Vol. 8, 2000, pp. 319-327.
16. R. Jain and Y. -B. Lin, "An auxiliary user location strategy employing forwarding pointers to reduce network impacts of PCS," *Wireless Networks*, Vol. 1, 1995, pp. 197-210.

Ki-Sik Kong received his B.S. and M.S. degrees in Computer Science and Engineering from Korea University, Seoul, Korea in 1999 and 2001, respectively. Currently, he is a Ph.D. candidate in Computer Science and Engineering at the same university, and is also a researcher in the Research Institute of Computer Science and Engineering Technology at Korea University. His research interests include distributed and mobile computing, mobile IP, and wireless and mobile networks.

Joon-Min Gil received his M.S. degree in Computer Science from Korea University, Seoul, Korea in 1996, and his Ph.D. degree in Computer Science and Engineering at the same university in 2000. From 2001 to 2002, he was a Visiting Research Associate in the Department of Computer Science at the University of Illinois at Chicago, U.S.A. Since 1998, he has worked for the Institute of Basic Science at Korea University, Seoul Korea as a Research Associate. His recent research interests include distributed and mobile computing, wireless networks, moving object databases, and soft computing.

Youn-Hee Han received his Ph.D. degree in Computer Science and Engineering from Korea University, Seoul, Korea in 2002. Currently, he is a technical engineer of i-Networking Laboratory of Samsung Advanced Institute of Technology, where he has been leading and conducting research on IP-based mobility management and IP paging. Prior to his current position, he was a researcher in the Institute of Basic Science at Korea University, where he was responsible for researching the location management in mobile computing, from 1998 to 2002.

Ui-Sung Song received his B.S. and M.S. degrees in Computer Science and Engineering from Korea University, Seoul, Korea in 1997 and 1999, respectively. He is currently a Ph.D. candidate in Computer Science and Engineering from Korea University. Also, he is currently a researcher in the Research Institute of Computer Science and Engineering Technology at Korea University. His research interests include mobile IP, personal communications service networks, and ad-hoc networks.

Chong-Sun Hwang received his M.S. degree in Mathematics from Korea University, Seoul, Korea in 1970, and his Ph.D. degree in Statistics and Computer Science from the University of Georgia in 1978. From 1978 to 1980, he was an Assistant Professor at South Carolina Lander State University. He is currently a Full Professor in the Department of Computer Science and Engineering at Korea University, Seoul, Korea. Since 1995, he has been a Dean in the Graduate School of Computer Science and Technology at Korea University. His recent research interests include distributed systems, distributed algorithms, and mobile computing systems.