A Pointer Forwarding Based Local Anchoring (POFLA) Scheme for Wireless Networks

Wenchao Ma, *Member, IEEE,* and Yuguang Fang, *Senior Member, IEEE*

***Abstract*—For a Personal Communication System (PCS) network to effectively deliver services to the mobile users, it must have an efficient way to keep track of mobile users. Location management fulfills this task through location registration and call delivery procedures. To reduce the signaling traffic, many schemes such as local anchoring, pointer forwarding, and two-level pointer forwarding schemes have been proposed in the past few years. In this paper, we present a novel location management scheme, *pointer forwarding based local anchoring* scheme, that intends to mitigate the signaling traffic while keeping the tracking delay in check. In this scheme, one visitor location register (VLR) traversed by a user is selected as the mobility agent (MA) for the user at a time, which forms another level of management in order to make some registration traffic localized. The idea is as follows: instead of always carrying out location update to the home location register (HLR), which would become the bottleneck otherwise, many location updates are accomplished by the mobility agent. Moreover, pointers can be set up between VLRs under the coverage of the same MA. When the distance between the MA and the current VLR exceeds a certain threshold, the current VLR will become the new MA, which also updates the HLR (where a user's location information is the identification of the reported MA). The numerical results show that this scheme can significantly reduce the network signaling traffic load without increasing much of the call setup delay.**

***Index Terms*—Location management, mobility management, Personal Communication System (PCS) networks, user mobility pattern.**

I. Introduction

IN wireless cellular networks such as Personal Communication Systems (PCSs), the mobile users communicate with each other via the fixed infrastructure using mobile telephones or portable computers. The user locations can change from time to time in the network coverage area. To deliver services efficiently to the mobile users, the network needs some mechanism to track the users under certain time delay restraint. Location management fulfills this task through the location registration and call delivery procedures. Two standards currently exist for PCS location management: *IS-41* [1] and *GSM MAP* [2]. IS-41 is commonly used in North America, and GSM MAP is popular in Europe and Asia.

Manuscript received May 12, 2004; revised October 29, 2004. This work was supported in part by National Science Foundation under Faculty Early Career Development Award ANI-0093241. The review of this paper was coordinated by Prof. B. Li.

W. Ma is with Microsoft Research Asia, Beijing 100080, P.R. China (e-mail: t-wenma@microsoft.com).

Y. Fang is with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611 USA (e-mail: fang@ece.ufl.edu).

Digital Object Identifier 10.1109/TVT.2005.844651

In third-generation (3G) wireless communication networks, a series of new technologies has been developed to enable the service providers to offer high-quality broadband multimedia services to mobile users in a cost-effective way. These technologies include IMT-2000, UMTS, GPRS, EDGE, CDMA-2000, and W-CDMA. Although some new functional entities are introduced in these new technologies, the basic scheme for location management has not changed significantly. One of the changes is that the registration areas (RAs) in the 3G systems become smaller and smaller. Although the reduction of RA size can improve the system performance in terms of packet loss, handoff, and user localization, it may also degrade the performance in terms of signaling traffic load. When a call request for a mobile user arrives, the user's current location must be determined before the services can be delivered to the user. Therefore some mobility databases must be designed to store the users' up-to-date location information that can be retrieved for the future service delivery. The databases store all the users' location information and need to be updated in a certain manner because of user mobility. With the increase of the mobile user number and the reduction of RA size, more and more update and retrieval operations have to be performed. Currently, both IS-41 and the GSM MAP share similar characteristics: they use a two-tier hierarchical system of *home location register* (HLR) and *visitor location register* (VLR) databases. Based on these strategies, a series of update (registration) operations will be initiated by the user terminal every time the user crosses the RA boundaries. If many mobile users are far away from their HLRs, heavy signaling traffic over the network will incur. This problem becomes more serious with the increase of the mobile users' population and the reduction of the RA size.

Many research works have been carried out to minimize the mobility management signaling traffic load. The *local anchoring scheme* is proposed in [3]. In this scheme, a VLR close to a user is selected as the local anchor for his mobile terminal. Whenever the user moves from one RA to another, the mobile terminal will perform location update to the local anchor. The local anchor for the mobile terminal will not change until a call request to the mobile user arrives. Whenever a new call arrives, the VLR serving the user becomes the new local anchor and a location update will be carried out to the HLR. The local anchoring scheme avoids frequent updates to HLR at the expense of increasing the local signaling traffic. The drawback of this scheme is that when the user keeps moving constantly without receiving any call, the updates to the local anchor may become costly, a similar bottleneck as the HLR. The authors [3] also attempted to improve the local anchoring scheme by dynamically selecting local anchors based on the user's expected future location registration and call delivery costs, leading to the *dynamic local anchoring scheme*. Jain and Lin [4] proposed another scheme called *per-user pointer forwarding scheme*. In this scheme, when a mobile user moves from one RA to another, a pointer is set up from the previous VLR to the current one, then the user can be traced through the pointer chain during the call delivery procedure. Since the pointer setup cost is usually less than the update cost to HLR, this scheme works well for users with high mobility. However, the penalty of this scheme is the longer time delay for locating the user. The longer the pointer chain, the less the signaling traffic, however, the longer the delay for finding the user. To avoid long connection setup latency, a threshold of the pointer chain length is set. The user needs to perform registration to the HLR whenever the pointer chain threshold is reached.

By combining the above two schemes, we recently proposed the *two-level pointer forwarding* strategy in [12]. In this scheme, we use two types of pointers. Some VLRs are selected as the *mobility agents* (MAs), which are responsible for location management in a larger area comparing to the RAs and can be geographically distributed. The pointers between MAs are level_1 pointers, and those between VLRs in the same charging domain of MAs are level_2 pointers. When a user crosses the boundaries of RAs, the level_2 pointers are set up. If the level_2 pointer chain threshold is reached, the current RA is selected as an MA for this user and a level_1 pointer is set up from the previous MA to the new one. Calls to a given user will query the HLR first and follow the level_1 pointer chain to the current MA, then reach the user's current VLR by tracking the level_2 pointer chain. The user does not need to update the HLR until the level_1 pointer chain threshold is reached. The chain threshold in the two-level pointer forwarding strategy can be much longer than that in the pointer forwarding scheme, and it also can have shorter call setup delay because of the level_1 pointer chain.

In this paper, we propose another new location management scheme called *pointer forwarding based local anchoring* (POFLA) scheme. It is similar to the two-level pointer forwarding scheme in the sense that some VLRs are selected as MAs, and there are two types of pointers in both schemes. However, the major difference is that there is only one MA in the pointer chain for POFLA at any given time while multiple MAs may exist concurrently for two-level pointer scheme. In the POFLA, when a user crosses a RA boundary and enters a new RA under the control of a different VLR, a pointer is set up from the old VLR to the new one without updating the MA. We call this type of pointer *low-level pointers* (or L-pointers). This procedure is similar to the two-level pointer or per-user forwarding scheme. If the threshold for L-pointer chain is reached, a higher level pointer is set up from the MA to the current VLR, and we call this type of pointer *high-level pointer* (or H-pointer). In this manner, the POFLA operates as the local anchoring scheme, and the difference is that the L-pointers instead of the H-pointer will be built again when the user keeps moving away. In this paper, we carry out the performance evaluation of the above four schemes under different *call-to-mobility ratios* (CMRs) using an analytical approach. We show that, under certain conditions, the POFLA performs better than the per-user forwarding and the local anchoring schemes. Although the two-level pointer forwarding scheme may show similar performance as our new strategy, the POFLA is simpler to implement in the practical systems.

This paper is organized as follows. In Section II, we describe the general system architecture of wireless cellular systems. The details of the new scheme are given in Section III. In Section IV, we derive the cost functions for the POFLA scheme and compare the performances of all the strategies under different conditions. The conclusions are presented in Section V.

II. System Description

In wireless networks, the user's current location must be determined first before any service can be delivered to the corresponding wireless access point. In a PCS network, the service area covered by the network is divided into cells. Each cell is primarily served by one base station, although a base station may serve one or more cells. Base stations are the users' radio port. An RA consists of an aggregation of a number of cells, forming a contiguous geographical region. The location management protocols for PCS networks consist of two major parts: location registration (or location update) and call delivery. In the location registration procedure, a mobile terminal updates its current location information to some network databases, and the information can be retrieved for the future call delivery procedure.

Currently, both the IS-41 and the GSM MAP share the same characteristics. They use a two-tier system consisting of HLR and VLR databases. The signaling network used to set up calls is distinct from that used to actually transport the information contents of the calls. It is the signaling network, connecting all the databases, that accomplishes the location management task for the PCS network. In real systems, a VLR or serving GPRS support node (SGSN) can be in charge of one or multiple RAs according to the user density. Without loss of generality, we assume there is one VLR or SGSN for every RA in our analysis model and use VLR as the example to illustrate our scheme. A VLR stores the profiles of the users who are currently residing in its charge area. The HLR stores the users' profiles and the IDs of current serving VLRs in addition to some information such as security and billing information. According to the current mobility management strategy, a mobile terminal keeps monitoring the broadcast signaling for the ID information of cells and RAs. If the mobile terminal finds that the RA changes, it sends a location update message to the new VLR via the current serving base station. Upon receiving the update message, the new VLR forwards it to the user's HLR to request the user's profile and finish the authentication, authorization, and accounting (AAA) procedures at the same time. The HLR updates the user's serving VLR ID in its database and sends a deregistration message to the user's old VLR. The old VLR deletes the user's entry in its database and responds with an acknowledgment message to the HLR. In the call delivery procedure, when a call for this user is initiated by some other caller, the location request message is sent to the user's HLR to find out the user's current location. The HLR forwards the message to the user's current serving VLR. The current VLR can determine the user's current serving cell by paging all the cells in its charging area. After finding out the

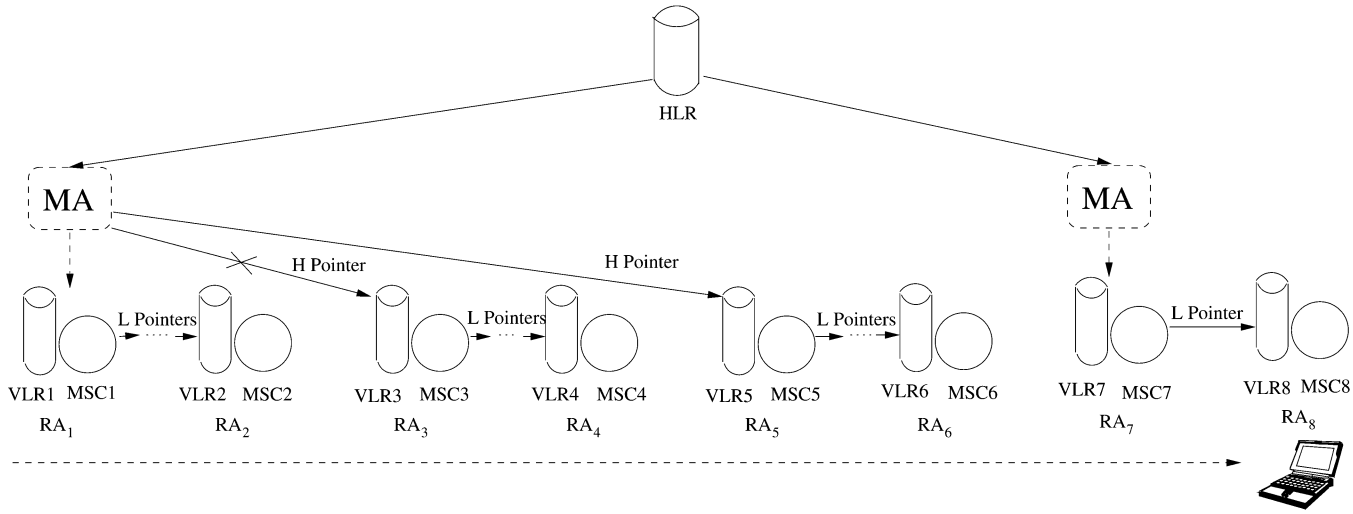


Fig. 1. POFLA strategy procedures.

user's current access point (base station), the VLR sends a responding message back to the HLR with a temporary number allocated to the called user for routing purpose. Then, a traffic channel can be assigned between the caller and the callee after the HLR receives the temporary number and forwards it to the caller's serving switch. It is obvious that, in this scheme, the location update traffic will increase dramatically with the increase of the user population and the reduction of the RA size, especially for the HLR, which may become bottleneck for the PCS and future wireless networks.

III. POFLA Scheme

In the POFLA scheme, the basic location update and call delivery procedure are modified to achieve better performance. A VLR is selected as the MA for a user and may change during the user's movement. Since the selection of MAs for the users is based on their locations, the signaling traffic can be distributed evenly among the network. This can avoid the bottleneck effect normally experienced by the HLR.

The basic location update procedure is modified as follows: every time a user enters a new RA served by a different VLR, the mobile terminal registers to the new VLR and informs the new VLR about the old VLR and MA. The old VLR and MA may be the same VLR the user visited. The VLR in the new RA determines what to do based on the mobile location update information. The new VLR has three options: it can request the old VLR to set up a pointer to itself, which is called L-pointer in our scheme; it can update the MA and request to set up a pointer from the MA to itself, called H-pointer; or it can decide to update the user's new location to the HLR directly and it itself becomes the new MA. The user's current VLR becomes new MA, too, whenever a connection setup is completed.

Fig. 1 shows the location update and call delivery procedures in the POFLA scheme with the H-pointer number limit setting three. Assume that a mobile user moves from RA_1 to RA_8 (these RAs are not necessarily adjacent) and VLR_1 is the user's current MA. At the beginning, the user is in RA_1 and VLR_1 is the user's current serving VLR too. VLR_1 is selected as the user's current MA because either the user just receives an incoming call in RA_1 or VLR_1 just updates the user's new location to the HLR. When the user leaves RA_1, but before enters RA_3, the mobile terminal informs the new VLR and a pointer chain consisting of L-pointers is set up just as in the per-user forwarding scheme ([4]). When the user enters RA_3, the chain threshold for L-pointers is reached. In this situation, VLR_3 will update the user's new location to the current MA, i.e., VLR_1. An H-pointer is set up from MA to VLR_3. At the same time, the L-pointer chain is reset. The same procedure is repeated in VLR_5 and the previous H-pointer is deleted when a new H-pointer is built from MA to VLR_5. If the user keeps moving, in RA_7, the threshold for L-pointer chain is reached again. At this time, the limit of the H-pointer number is reached too. Instead of exchanging information with the previous MA and setting up a new H-pointer, VLR_7 will update the user's location to the HLR directly and VLR_7 itself is selected as the new MA for that the user. The reason for updating the HLR instead of the MA is that the cost of setting up and traversing the pointer chain between MA and the current serving VLR may be costly when the user is far away from the MA, and the connection setup delay for an incoming call may become intolerable. If an incoming call arrives before the mobile user changes his MA, the current serving VLR is selected as the user's current MA because the HLR has the knowledge of the user's current location after the connection setup and it is not necessary to go through the pointer chain again to locate the user for future service deliveries.

The call delivery procedure in the POFLA scheme is straightforward. When the subsequent calls are initiated from some other switches to the user, the user's HLR is queried first and a pointer to the user's current MA is obtained. Then the pointer chain is followed to find the user's current location. For example, in Fig. 1, if the user is in RA_6 when a call arrives, the user can be reached by following the H-pointer to VLR_5 and L-pointer chain to VLR_6. After the connection setup, the current serving VLR becomes the user's new MA.

Because of the H-pointer, the chain length in the POFLA scheme can be longer than that in the basic pointer forwarding scheme without generating long connection setup delay. In Section IV, we can see that the delay for the POFLA scheme is much less than that for the per-user forwarding scheme. Our study also shows, under some assumptions, the POFLA scheme performs better than the static and dynamic local anchoring schemes. Although in some cases, the two-level pointer forwarding scheme ([12]) can generate similar results as those of the POFLA scheme, the new scheme is simpler to implement in practical systems.

IV. Performance Analysis

In this section, we develop an analytic model to derive the cost functions and compare the performance of the POFLA scheme with those of the other three schemes.

A. The Signaling Cost Functions

The mobile users in a PCS can be characterized by their CMRs. All the schemes are evaluated under different CMRs in this paper. The CMR of a user is defined as the expected number of received calls for a user during the period that the user visits an RA. If calls are *received* by the user at an average rate λ and the time the user resides in a given RA has average value $1/\mu$, then the CMR, denoted as ρ, is given by

$$\rho = \frac{\lambda}{\mu}. \tag{1}$$

In order to compare the signaling cost with that of the basic IS-41 location management scheme, we need to consider the mobility management procedures used in IS-41 and our new scheme. We observe that a mobile terminal needs to update its location only when the terminal does not engage in communications with the fixed communication infrastructure (i.e., the network), hence we only need to compare the signaling traffic in the time interval between call services (i.e., the interval between the end of the current call and the beginning of the next call, which is called *interservice time* in [5] and [6]). Assume that a mobile terminal crosses a number of RAs during the interservice time: ignoring the busy-line effect, the interservice time can be approximated by the interarrival time of calls to the terminal. If the basic user location update scheme (IS-41) is used, the user's HLR will be updated every time the user moves to a new RA. In the POFLA scheme, the HLR is updated after every $K_1 \cdot K_2$ moves (K_1 and K_2 are the L-pointer chain threshold and H-pointer number limit, respectively), and pointers are set up for all other moves. We define C and C' to be the total costs of updating the location information (location update) and tracking the user (call delivery) during the interservice time for the basic IS-41 and the POFLA strategies, respectively. For convenience, we list all the notations used in our analysis as follows:

K_1	threshold for the L-pointer chain;
K_2	number limit of H-pointer (i.e., every K_2 updates to an MA will result in the change of a new MA);
m	average cost of location update to the HLR;
M	total location update cost during the interservice time in the IS-41 scheme;
F	total cost of call delivery in the IS-41 scheme;
M'	total location update cost in the POFLA scheme during the interservice time;
F'	total call delivery cost in the POFLA scheme;
S_1	pointer setup cost of an L-pointer;
$S_{2,j}$	pointer setup cost of the jth H-pointer;
T_1	cost of traversing an L-pointer between two adjacent VLRs;
$T_{2,j}$	cost of traversing the jth H-pointer;
$\alpha(i)$	probability that there are i RA crossings during the interservice time;
P	processing cost of setting up a pointer;
G	signaling cost of setting up an L-pointer;
β	coefficient of signaling cost for an H-pointer ($\beta \geq 1$).

Then, we can express the total costs during the interservice time for the two location management schemes as follows:

$$C = M + F = \frac{m}{\rho} + F \tag{2}$$

$$C' = M' + F'. \tag{3}$$

Since pointers are set up in the POFLA scheme, we need to define the pointer setup and traversing costs for further analysis. Every time a pointer is built, the signaling messages will be transmitted back and forth. For pointer traversing, the signaling message is transmitted in one direction only. So we define the costs of pointer setup and traversing for the L-pointer as S_1 and T_1, respectively

$$S_1 = G + P \tag{4}$$

$$T_1 = \frac{1}{2}G + P. \tag{5}$$

The costs of H-pointers are not fixed values because the length of the H-pointers changes with the user's location. In this paper, we express the costs for H-pointers as follows:

$$S_{2,j} = \begin{cases} 0, & \text{if } j = 0 \\ j\beta G + P, & \text{Otherwise} \end{cases} \tag{6}$$

$$T_{2,j} = \begin{cases} 0, & \text{if} j = 0 \\ \frac{1}{2}j\beta G + P, & \text{Otherwise} \end{cases} \tag{7}$$

where the subscript j means the setup cost or the traversing cost for the jth H-pointer.

Now, we can derive the formula of M' and F' as follows: suppose that a user crosses i RA boundaries during the interservice time. The HLR is updated $\lfloor i/K_1K_2 \rfloor$ times. If we call the summation from the 0th to the $(K_2 - 1)$th H-pointer setup cost, $\sum_{j=0}^{K_2-1} S_{2,j}$, the MA update cost, then there are $\lfloor i/K_1K_2 \rfloor$ times such MA update costs that would incur during the interservice time with i RA crossings. In addition, $\lfloor i - \lfloor i/K_1K_2 \rfloor K_1K_2/K_1 \rfloor$ H-pointer setups and $i - \lfloor i/K_1 \rfloor$ L-pointer setups occurred in the remaining i RA crossings. Thus, we can obtain

$$M' = \sum_{i=0}^{\infty} \left\{ \left\lfloor \frac{i}{K_1K_2} \right\rfloor m + \left(i - \left\lfloor \frac{i}{K_1} \right\rfloor \right) S_1 + \left\lfloor \frac{i}{K_1K_2} \right\rfloor \times \left(\sum_{j=0}^{K_2-1} S_{2,j} \right) + \sum_{j=0}^{\lfloor i - \lfloor i/K_1K_2 \rfloor K_1K_2/K_1 \rfloor} S_{2,j} \right\} \alpha(i). \tag{8}$$

The cost F' can be derived in a straightforward way. In order to reach the user's current location, the signaling message is sent to the MA and then delivered through one H-pointer and

$i - \lfloor i/K_1 \rfloor K_1$ L-pointers before reaching the current location. So, we have

$$F' = F + \sum_{i=0}^{\infty} \left\{ T_{2,\zeta} + (i - \left\lfloor \frac{i}{K_1} \right\rfloor K_1) T_1 \right\} \alpha(i) \quad (9)$$

where $\zeta = \lfloor i - \lfloor i/K_1 K_2 \rfloor K_1 K_2 / K_1 \rfloor$. Notice that we can easily obtain (10) and (11) as shown at the bottom of the page. Substituting (10) and (11) into (8), we can obtain (12) as shown at the bottom of the page. M_1 can be simplified according to the definition of $\alpha(i)$

$$M_1 = \frac{S_1}{\rho} = \frac{G+P}{\rho}. \quad (13)$$

In order to evaluate $\alpha(i)$, we invoke the following assumptions.

1) The interservice time is exponentially distributed with average $1/\lambda$, i.e., the served calls to the mobile terminal form Poisson process with rate λ.
2) The residence time of the mobile user at a registration area is a random variable with a general probability density function $f_m(t)$ and the Laplace transform

$$f_m^*(s) = \int_{t=0}^{\infty} f_m(t) e^{-st} dt$$

and with average RA residence time $1/\mu$.

For simplicity, we denote $g = f_m^*(\lambda)$. Based on the above assumptions, we obtain the probability $\alpha(i)$ (see [7] for the detailed derivation)

$$\alpha(i) = \frac{(1-g)^2 g^{i-1}}{\rho}. \quad (14)$$

Applying variable substitution $i = jK_1 + k$, then we obtain

$$\alpha(jK_1 + k) = \frac{(1-g)^2}{\rho g} (g^{K_1})^j g^k = y z^j x^k \quad (15)$$

where

$$y = \frac{(1-g)^2}{\rho g}, \quad z = g^{K_1}, \quad x = g.$$

Thus, we have

$$\begin{aligned} M_2 &= y S_1 \sum_{j=0}^{\infty} \sum_{k=0}^{K_1 - 1} j z^j x^k \\ &= \frac{y S_1 (1 - x^{K_1})}{1 - x} \left(\sum_{j=0}^{\infty} j z^j \right) \\ &= \frac{(1-g)(G+P) g^{K_1 - 1}}{\rho (1 - g^{K_1})}. \end{aligned} \quad (16)$$

Similarly, we can obtain M_3 by using substitution $i = jK_1K_2 + k$; then

$$\alpha(jK_1K_2 + k) = \frac{(1-g)^2}{\rho g} (g^{K_1 K_2})^j g^k = y z^j x^k \quad (17)$$

$$\sum_{j=0}^{K_2 - 1} S_{2,j} = \frac{K_2 (K_2 - 1) \beta G + 2 (K_2 - 1) P}{2} \quad (10)$$

$$\sum_{j=0}^{\lfloor i - \lfloor i/K_1K_2 \rfloor K_1K_2/K_1 \rfloor} S_{2,j} = \frac{\left(\left\lfloor \frac{i - \lfloor \frac{i}{K_1K_2} \rfloor K_1K_2}{K_1} \right\rfloor + 1 \right) \left\lfloor \frac{i - \lfloor \frac{i}{K_1K_2} \rfloor K_1K_2}{K_1} \right\rfloor \beta G + 2 \left\lfloor \frac{i - \lfloor \frac{i}{K_1K_2} \rfloor K_1K_2}{K_1} \right\rfloor P}{2} \quad (11)$$

$$\begin{aligned} M' = & \underbrace{S_1 \sum_{i=0}^{\infty} i \alpha(i)}_{M_1} - \underbrace{S_1 \sum_{i=0}^{\infty} \left\lfloor \frac{i}{K_1} \right\rfloor \alpha(i)}_{M_2} \\ & + \underbrace{\frac{(K_2 - 1) K_2 \beta G + 2 (K_2 - 1) P + 2m}{2} \sum_{i=0}^{\infty} \left\lfloor \frac{i}{K_1 K_2} \right\rfloor \alpha(i)}_{M_3} \\ & + \underbrace{\frac{\beta G}{2} \sum_{i=0}^{\infty} \left\lfloor \frac{i - \lfloor \frac{i}{K_1K_2} \rfloor K_1 K_2}{K_1} \right\rfloor^2 \alpha(i)}_{M_4} \\ & + \underbrace{\frac{\beta G + 2P}{2} \sum_{i=0}^{\infty} \left\lfloor \frac{i - \lfloor \frac{i}{K_1K_2} \rfloor K_1 K_2}{K_1} \right\rfloor \alpha(i)}_{M_5} \end{aligned} \quad (12)$$

where

$$y = \frac{(1-g)^2}{\rho g}, \quad z = g^{K_1K_2}, \quad x = g.$$

$$\begin{aligned} M_3 &= \frac{(K_2-1)K_2\beta G + 2(K_2-1)P + 2m}{2} \\ &\quad \times \sum_{i=0}^{\infty} \left\lfloor \frac{i}{K_1K_2} \right\rfloor \alpha(i) \\ &= \frac{(K_2-1)K_2\beta G + 2(K_2-1)P + 2m}{2} \\ &\quad \cdot \frac{(1-g)g^{K_1K_2-1}}{\rho(1-g^{K_1K_2})}. \end{aligned} \tag{18}$$

In a similar manner, we can also obtain M_4 and M_5 as follows:

$$\begin{aligned} M_4 &= \frac{\beta G}{2} \sum_{j=0}^{\infty} \sum_{k=0}^{K_1K_2-1} \left\lfloor \frac{k}{K_1} \right\rfloor^2 \alpha(jK_1K_2+k) \\ &= \frac{\beta G}{2} \sum_{j=0}^{\infty} \sum_{k=0}^{K_1K_2-1} \left\lfloor \frac{k}{K_1} \right\rfloor^2 yz^jx^k \\ &= \frac{\beta G}{2} y \sum_{j=0}^{\infty} z^j \left(\sum_{n=0}^{K_2-1} \sum_{m=0}^{K_1-1} n^2 x^{nK_1+m} \right) \\ &= \frac{\beta G(1-g)}{2\rho g(1-g^{K_1K_2})(1-g^{K_1})^2} \\ &\quad \cdot \Big\{ g^{K_1} + g^{2K_1} - K_2^2 g^{K_1K_2} \\ &\qquad + (2K_2^2 - 2K_2 - 1)g^{K_1(K_2+1)} \\ &\qquad - (K_2-1)^2 g^{K_1(K_2+2)} \Big\} \end{aligned} \tag{19}$$

and (20) as shown at the bottom of the page.

Finally, we find the expression $M' = M_1 - M_2 + M_3 + M_4 + M_5$.

If $T_{2,\zeta} \neq 0$. We can obtain F' in the following fashion:

$$\begin{aligned} F' &= F + \sum_{i=0}^{\infty} \Bigg\{ \frac{1}{2} \left\lfloor \frac{i - \lfloor \frac{i}{K_1K_2} \rfloor K_1K_2}{K_1} \right\rfloor \beta G \\ &\qquad + P + \left(i - \left\lfloor \frac{i}{K_1} \right\rfloor K_1 \right) \left(\frac{1}{2}G + P \right) \Bigg\} \alpha(i) \\ &= F + \underbrace{\left(\frac{1}{2}G + P \right) \sum_{i=0}^{\infty} i\alpha(i)}_{F_1} \\ &\quad - \underbrace{\left(\frac{1}{2}G + P \right) K_1 \sum_{i=0}^{\infty} \left\lfloor \frac{i}{K_1} \right\rfloor \alpha(i)}_{F_2} \\ &\quad + \underbrace{\left(\frac{1}{2}\beta G \sum_{i=0}^{\infty} \left\lfloor \frac{i - \lfloor \frac{i}{K_1K_2} \rfloor K_1K_2}{K_1} \right\rfloor + P \right) \alpha(i)}_{F_3}. \end{aligned} \tag{21}$$

We can compute F' in a similar fashion. Notice that if we use the substitution $i = jK_1K_2 + k$, when $k = 0, 1, \ldots, K_1 - 1$, $T_{2,\zeta} = 0$. So we can obtain F' as follows:

$$\begin{aligned} F' &= F + \frac{G+2P}{2\rho} \left[1 - \frac{K_1(1-g)g^{K_1-1}}{1-g^{K_1}} \right] \\ &\quad + \frac{P(1-g)(g^{K_1-1} - g^{K_1K_2-1})}{\rho(1-g^{K_1K_2})} \\ &\quad + \frac{\beta G(1-g)[(K_2-1)g^{K_1(K_2+1)} - K_2 g^{K_1K_2} + g^{K_1}]}{2\rho g(1-g^{K_1K_2})(1-g^{K_1})}. \end{aligned} \tag{22}$$

In summary, we have the following theorem.

Theorem: If the interservice time is exponentially distributed and the RA residence time is generally distributed with Laplace transform $f_m^*(s)$, then the total location update cost and total cost for call delivery are given as shown in (23) at the bottom of the next page, where $g = f_m^*(\lambda)$.

For demonstration purposes, we assume that the RA residence time is Gamma distributed with mean $1/\mu$. The reason that Gamma distribution is selected is its flexibility in setting

$$\begin{aligned} M_5 &= \frac{\beta G + 2P}{2} \sum_{j=0}^{\infty} \sum_{k=0}^{K_1K_2-1} \left\lfloor \frac{k}{K_1} \right\rfloor \alpha(jK_1K_2+k) \\ &= \frac{\beta G + 2P}{2} \sum_{j=0}^{\infty} \sum_{k=0}^{K_1K_2-1} \left\lfloor \frac{k}{K_1} \right\rfloor yz^jx^k \\ &= \frac{\beta G + 2P}{2} y \sum_{j=0}^{\infty} z^j \sum_{n=0}^{K_2-1} n x^{nK_1} \sum_{m=0}^{K_1-1} x^m \\ &= \frac{(\beta G + 2P)(1-g)[(K_2-1)g^{K_1(K_2+1)} - K_2 g^{K_1K_2} + g^{K_1}]}{2\rho g(1-g^{K_1K_2})(1-g^{K_1})} \end{aligned} \tag{20}$$

various parameters and can be used to fit the first two moments of the field data. The Laplace transform of a Gamma distribution is

$$f_m^*(s) = \left(\frac{\gamma\mu}{s+\gamma\mu}\right)^{\gamma}.$$

Thus, we have

$$g = f_m^*(\lambda) = \left(\frac{\gamma\mu}{\lambda+\gamma\mu}\right)^{\gamma} = \left(\frac{\gamma}{\rho+\gamma}\right)^{\gamma}. \tag{24}$$

In particular, when $\gamma = 1$, we have an exponential distribution for the RA residence time.

B. Performance Analysis for Exponential RA Residence Time

We first consider the situation that the RA residence time is exponentially distributed. By setting $\gamma = 1$, (24) becomes

$$g = \frac{1}{1+\rho}.$$

In our analysis, we do not address issues regarding the content of messages and other information transfer which may occur during a call connection setup. For simplicity, we assume that the message sizes are equal for all signaling transactions. Since we only compare the relative performance of the aforementioned schemes with the POFLA scheme under various CMRs, the conclusions will not be affected by the simplification.

Notice that, in the simplified IS-41 or the GSM MAP procedures, the location update and call delivery involve the same number of messages between HLR and VLR databases, so we choose $m = F$. Without loss of generality, we normalize $m = F = 1$. G is the signaling transmission cost and P is the processing cost. P usually includes the database transaction costs. The values of P and G should be much less than m or F. In this paper, we do not assign any practical meaning to these parameters; they can be explained as the signaling message traffic exchanged during the location management procedures or the time delay experienced in the real systems.

In Fig. 2, we plot the relative location update, call delivery, and net costs of three schemes as functions of CMR. Here for the POFLA scheme and the two-level pointer forwarding scheme, we assume $K_1 = K_2 = 3$; for the per-user forwarding scheme, the threshold is nine $(K_1 \times K_2)$. In this figure, we also assume $P = 0.05$, $G = 0.1$, and $\beta = 1.5$. As we can see in Fig. 2(a), the POFLA scheme generates higher values than the per-user forwarding scheme. It is obvious because, in the latter scheme, only the L-pointers are set up, while in the POFLA scheme, a new VLR would set up an H-pointer to the MA when the threshold for L-pointers is reached, which costs more. For the two-level pointer forwarding scheme, the level_2 pointer is the L-pointer and the level_1 pointer is usually shorter than the H-pointer [12]. So the location update cost for the two-level pointer forwarding scheme is in the middle of them. Although, with the same length of the pointer threshold, the per-user forwarding scheme can generate less location update cost, it has the largest call delivery cost among the three strategies [see Fig. 2(b)]. For some users with small CMR, which means that the users have higher mobility relative to call arrival rate, the call delivery cost for the per-user forwarding scheme is much higher than those for the other two schemes. In practical systems, it could be embodied as the delay the users have to wait before a connection can be set up. In the POFLA scheme, usually fewer pointers have to be traversed than the two-level pointer forwarding scheme before a user can be located, so the POFLA scheme has the least call delivery cost. In Fig. 2(b), the POFLA scheme does not outperform the two-level pointer forwarding scheme too much; however, it is easier to implement in practical systems because only one VLR is selected as MA at any time. This reduces the VLR numbers involved in the chain. Whenever a connection is set up, the called user can be located by following one H-pointer and multiple L-pointers if any. Although the three schemes perform differently in location update and call delivery, the total net costs for the three schemes are similar for high CMRs [Fig. 2(c)].

We increase the signaling transfer coefficient β from 1.5 to 3 in Fig. 3, which means that the H-pointer setup cost is higher. Under these conditions, both the location update and call delivery costs in the POFLA and two-level pointer forwarding schemes increase. Since the pointer setup and traversing costs for L-pointer keep unchanged, the performance of the per-user forwarding scheme does not change either. Even when the costs of H-pointer increase, the connection setup costs of the POFLA and the two-level pointer forwarding schemes are still less than

$$\begin{aligned}
M' =& \frac{G+P}{\rho} - \frac{(1-g)(G+P)g^{K_1-1}}{\rho(1-g^{K_1})} + \frac{(K_2-1)K_2\beta G + 2(K_2-1)P + 2m}{2} \\
&\cdot \frac{(1-g)g^{K_1K_2-1}}{\rho(1-g^{K_1K_2})} + \frac{\beta G(1-g)}{2\rho g(1-g^{K_1K_2})(1-g^{K_1})^2} \\
&\cdot \left\{ g^{K_1} + g^{2K_1} - K_2^2 g^{K_1K_2} + (2K_2^2 - 2K_2 - 1)g^{K_1(K_2+1)} - (K_2-1)^2 g^{K_1(K_2+2)} \right\} \\
&+ \frac{(\beta G + 2P)(1-g)[(K_2-1)g^{K_1(K_2+1)} - K_2 g^{K_1K_2} + g^{K_1}]}{2\rho g(1-g^{K_1K_2})(1-g^{K_1})}, \\
F' =& F + \frac{G+2P}{2\rho}\left[1 - \frac{K_1(1-g)g^{K_1-1}}{1-g^{K_1}}\right] + \frac{P(1-g)(g^{K_1-1} - g^{K_1K_2-1})}{\rho(1-g^{K_1K_2})} \\
&+ \frac{\beta G(1-g)[(K_2-1)g^{K_1(K_2+1)} - K_2 g^{K_1K_2} + g^{K_1}]}{2\rho g(1-g^{K_1K_2})(1-g^{K_1})}
\end{aligned} \tag{23}$$

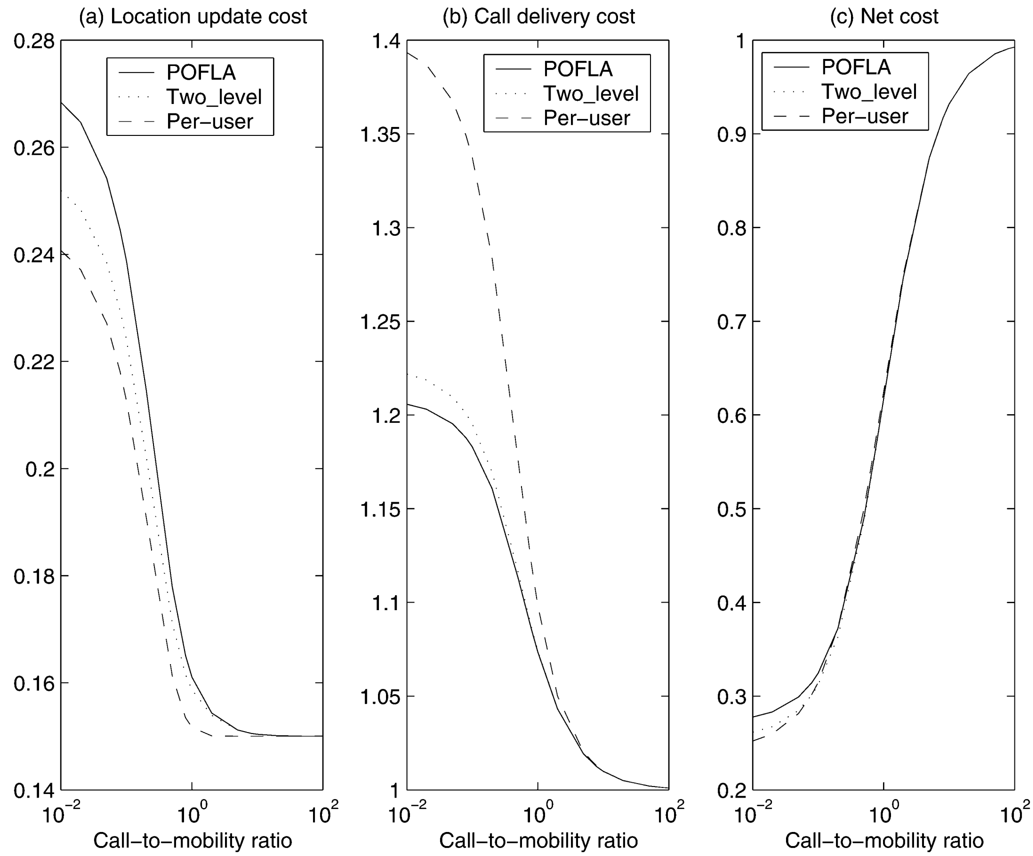


Fig. 2. The relative costs for the three schemes with $P = 0.05$, $G = 0.1$, and $\beta = 1.5$.

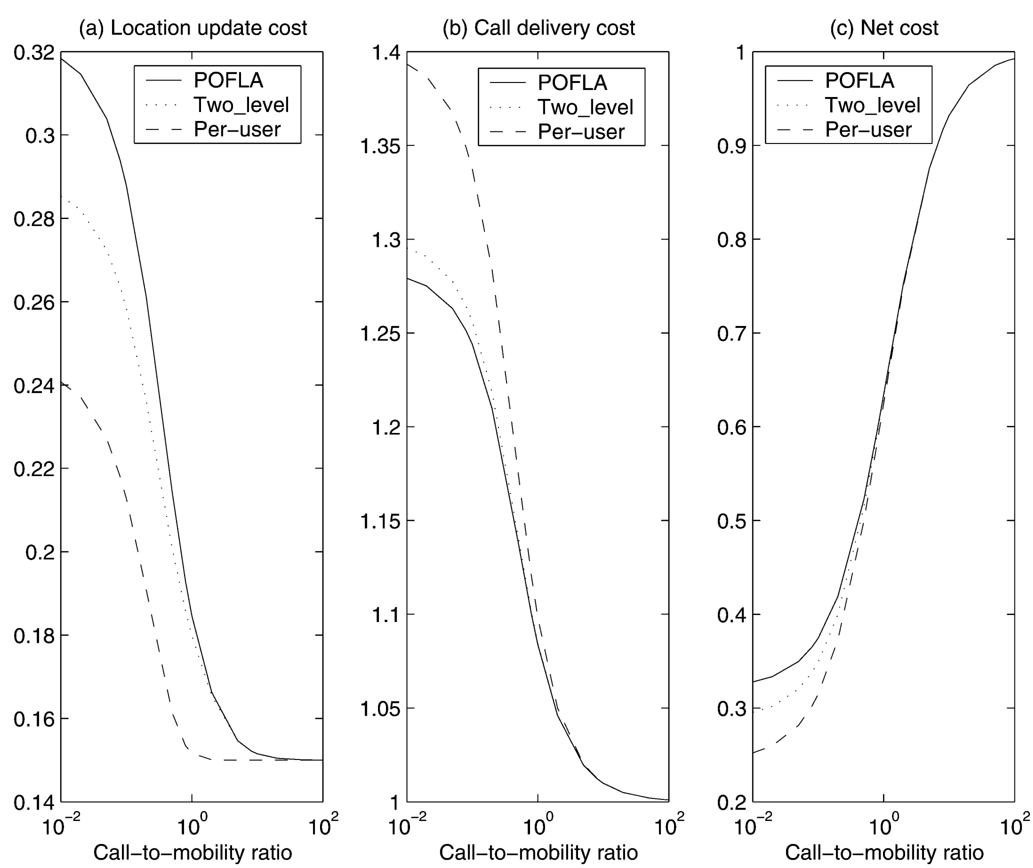


Fig. 3. The relative costs for the three schemes with $P = 0.05$, $G = 0.1$, and $\beta = 3$.

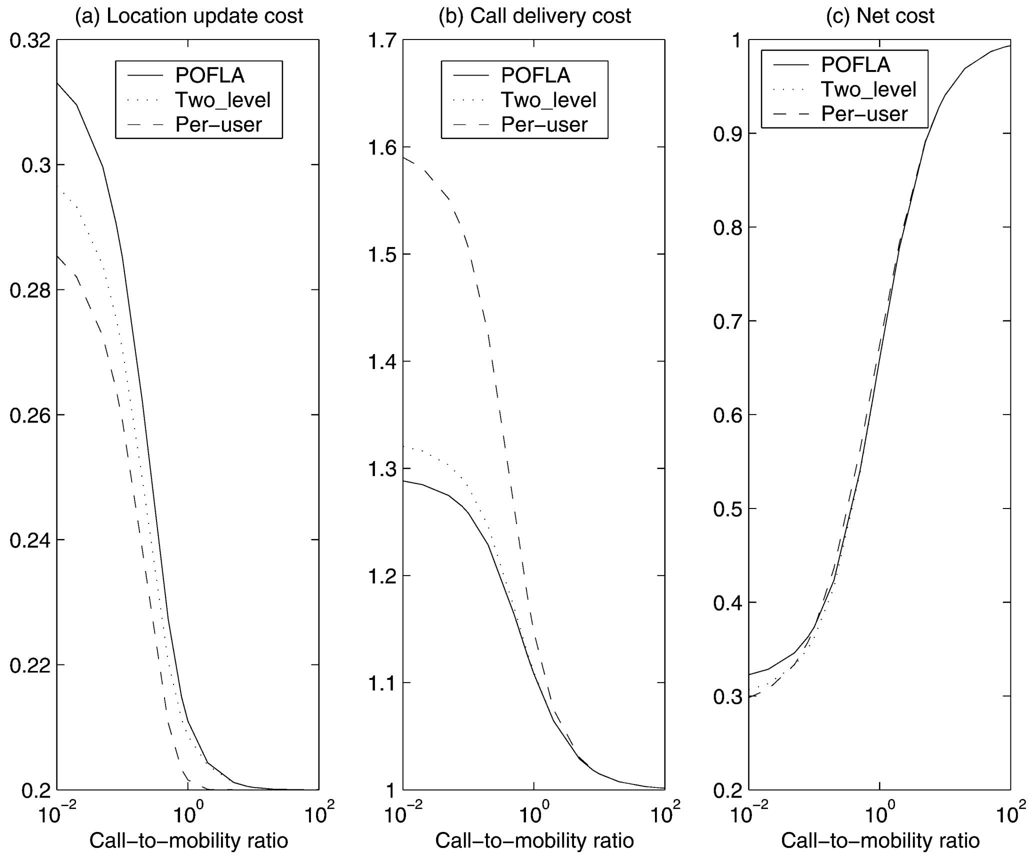


Fig. 4. The relative costs for the three schemes with $P = 0.1$, $G = 0.1$, and $\beta = 1.5$.

that for the per-user forwarding scheme [Fig. 3(b)]. In Fig. 3(c), when the CMR is very low, the per-user forwarding scheme can perform better; when the CMR is larger than one, the net cost for the three schemes is similar.

Fig. 4 shows the performance of the three strategies when the pointer processing cost P is increased to 0.1. With the increase of user population and the mobility, the processing cost for pointer management would increase too. The processing cost includes database transactions and may generate extra delay with larger number of operational requests. As we can observe from Fig. 4, the costs for all three schemes increase. The POFLA scheme will generate the least connection delay, and the total net performance is very close.

In this paper, we also compare the POFLA scheme with that of the local anchoring scheme proposed in [3]. The authors has suggested two variants of the local anchoring scheme—static and dynamic. The static local anchoring scheme is easier to implement. In this scheme, a VLR served a user is selected as the local anchor for that user and will not change until the next call arrives. In the dynamic local anchoring scheme, the user's current local anchor might change to the current serving one according to the user's expected future events. The dynamic local anchoring scheme is more difficult to implement than the static one; however, the results in [3] show that the dynamic scheme can guarantee that the net cost is less than the basic IS-41 or the GSM MAP strategy, and the static scheme might generate higher cost than the basic scheme under certain conditions. In fact, the local anchoring scheme is a special case of the POFLA scheme. If we let $K_1 = 1$, then the POFLA scheme reduces to the dynamic local anchoring scheme. The performance comparisons of the two schemes are shown in Fig. 5. In order to make the comparison fair, the effective pointer chain length $(K_1 \times K_2)$ in the POFLA scheme is the same as the dynamic local anchoring scheme. In [3], the decision of the local anchor change is made based on the user's next event, which is derived according to the user's mobility pattern. In our analysis, we assume the local anchor changes to make sure the net cost will not exceed the basic IS-41 scheme cost. In Fig. 5, we assume $P = 0.05$, $G = 0.1$, $\beta = 1.5$, and $m = F = 1$. Based on these assumptions, we obtain that the effective pointer chain length is four, so we set $K_1 = K_2 = 2$. It can be seen that in the local update, call delivery, and total net costs, the POFLA has better performance than the dynamic local anchoring scheme. In Fig. 6, we also compare the total net cost of the POFLA scheme with that of the static local anchoring scheme. In this figure, we assume $K_1 = K_2 = 3$. We can see that when the CMR is low, the static local anchoring scheme involves higher traffic load.

C. Sensitivity to the Variance of the RA Residence Time

We now investigate the POFLA scheme performance sensitivity to the variance of the RA residence time. We assume the RA residence time has a Gamma distribution. For a Gamma distribution, the variance is $V = \mu^2/\gamma$, i.e., a large γ implies a small variance.

Fig. 7 shows the effect of γ on M'/M, F'/F, and C'/C, respectively. In these figures, we can observe that the increase

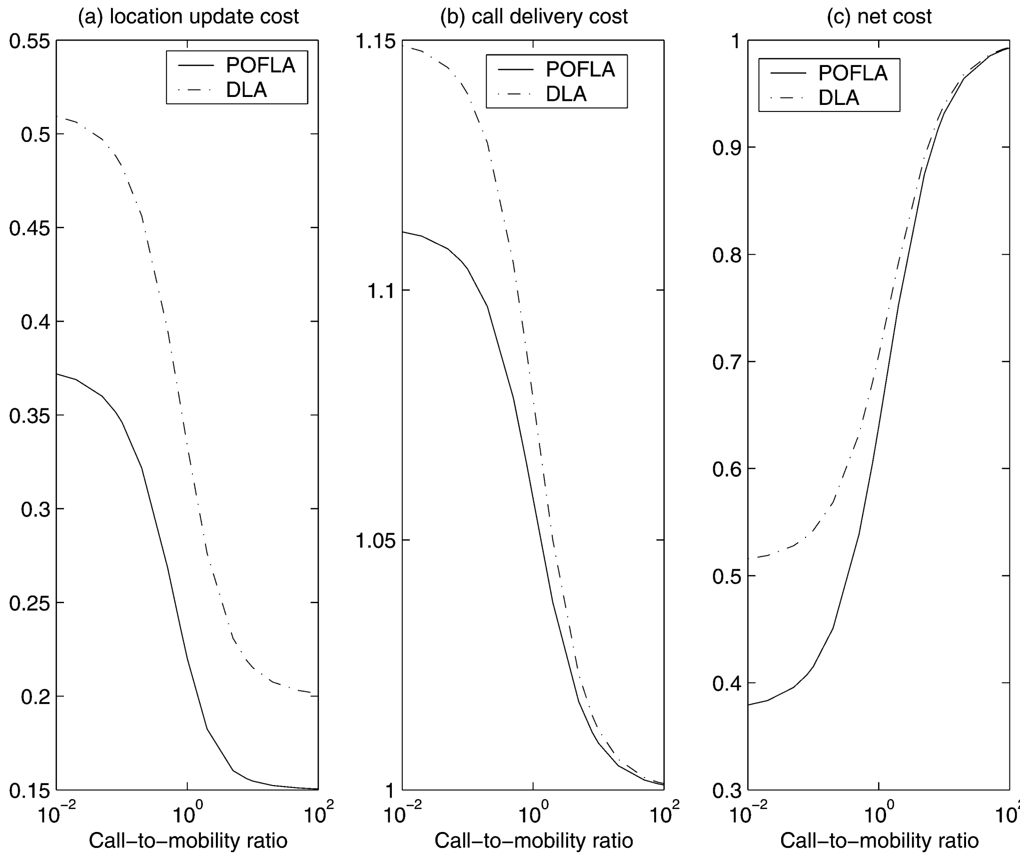


Fig. 5. The relative costs for the POFLA and dynamic local anchoring schemes with $P = 0.05$, $G = 0.1$, and $\beta = 1.5$.

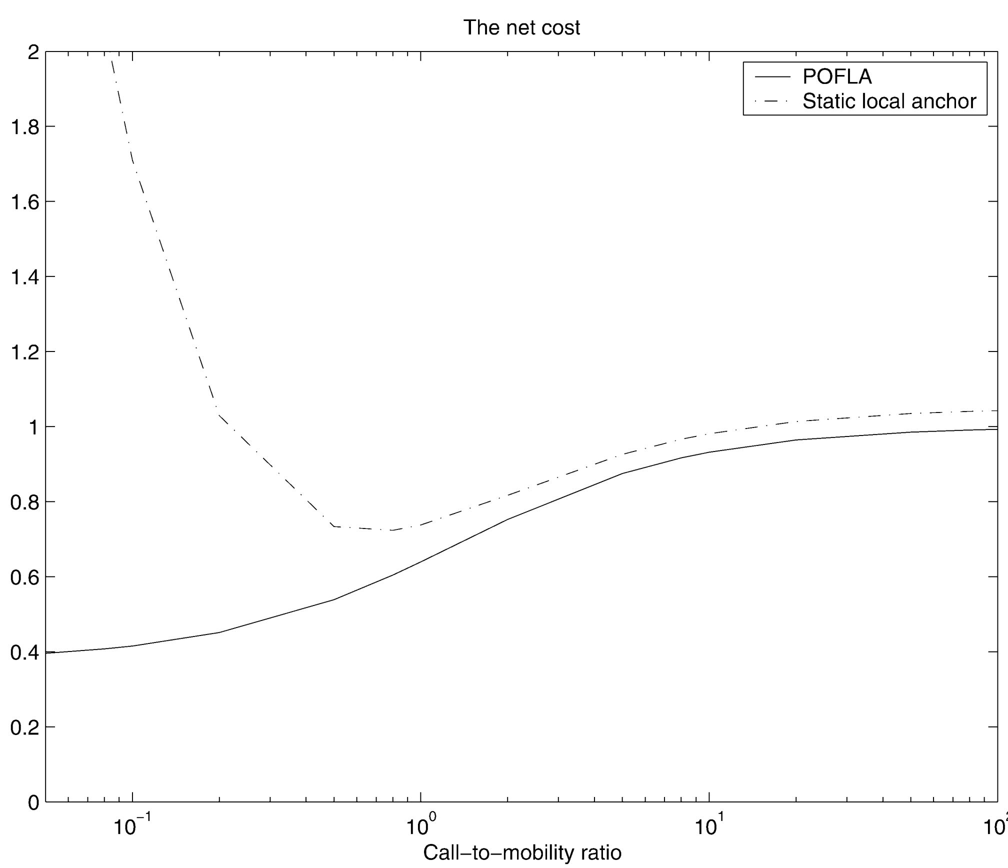


Fig. 6. The relative net costs for the POFLA and static local anchoring schemes with $P = 0.05$, $G = 0.1$, and $\beta = 1.5$.

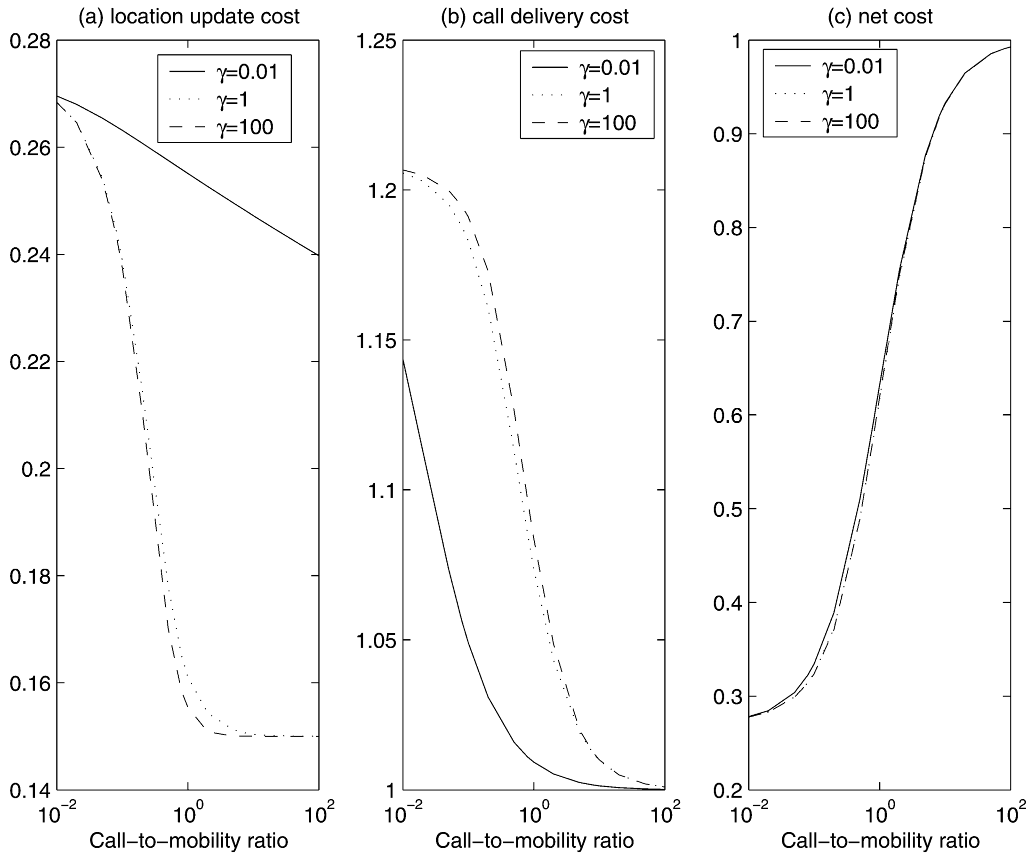


Fig. 7. The effect of variance of residence time (γ) with $P = 0.05$, $G = 0.1$, and $\beta = 1.5$.

of the variance of RA residence time (smaller γ) causes the increase of M'/M and the decrease of F'/F; the net effect on C_F/C_B is not significant. The large variance implies that the number of RA boundaries the user crosses during the interservice time varies greatly. If the user crosses many RAs, a longer H-pointer chain will be created. When the limit of H-pointer K_2 is reached, the HLR will be updated, resulting in an increase of M'. On the other hand, if fewer boundaries are crossed, only shorter pointer chains are set up, and the pointer creation/tracing cost will be saved. The net effect is the increase of M'. In Fig. 7(b), when the variance is high, if the crossed RA boundary number is small, the call delivery cost is reduced; if the number is large, the pointer chain could be shortened by H-pointer or the update to HLR. The net effect is a significant improvement in F'/F. The total net effect of the residence time variance on total cost ratio C'/C is not significant for all CMRs [see Fig. 7(c)].

V. Conclusions

In this paper, we proposed a new location management scheme—pointer forwarding based local anchoring (POFLA). In this scheme, one mobility agent and two types of pointers are introduced. The location update to the HLR can be mitigated by setting up a pointer from the mobility agent or the previous VLR to the current one. The scheme can reduce the long distance signaling traffic at the expense of certain increase in the local signaling traffic. It is beneficial when the cost of communicating with HLR is much higher than that for local signaling exchange. The advantage of the POFLA is that it can keep the call delivery latency low and reduce the total system cost at the same time. In this paper, we also undertake the performance comparison of the new scheme with the per-user forwarding, the local anchoring, and the two-level pointer forwarding schemes using an analytical approach. The results show the significant advantages of the POFLA scheme.

References

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

[2] "Digital Cellular Telecommunications System (Phase 2+): Mobile application part (MAP) specification," ETSI, GSM 09.02 version 7.51 Release, 1998.

[3] J. Ho and F. Akyildiz, "Local anchor scheme for reducing signaling costs in personal communications networks," *IEEE/ACM Trans. Netw.*, vol. 4, pp. 709–725, Oct. 1996.

[4] R. Jain and Y. B. Lin, "An auxiliary user location strategy employing forwarding pointers to reduce network impacts of PCS," *Wireless Networks*, vol. 1, pp. 197–210, 1995.

[5] Y. Fang, "Movement-based location management and tradeoff analysis for wireless mobile networks," *IEEE Trans. Comput.*, vol. 52, pp. 791–803, Jun. 2003.

[6] ——, "General modeling and performance analysis for location management in wireless mobile networks," *IEEE Trans. Comput.*, vol. 51, pp. 1169–1181, Oct. 2002.

[7] Y. Fang, I. Chlamtac, and Y. B. Lin, "Portable movement modeling for PCS networks," *IEEE Trans. Veh. Technol.*, vol. 87, pp. 1347–1384, Aug. 1999.

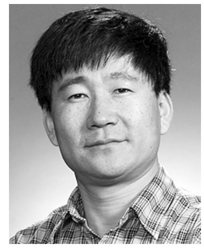
[8] I. F. Akyildiz, J. McNair, J. S. M. Ho, H. Uzunalioglu, and W. Wang, "Mobility management in next-generation wireless systems," *Proc. IEEE*, vol. 4, Oct. 1996.

[9] Y. B. Lin and I. Chlamtac, *Wireless and Mobile Network Architectures*. New York: Wiley, 2001.

[10] R. Jain, Y. B. Lin, C. Lo, and S. Mohan, "A caching strategy to reduce network impacts of PCS," *IEEE J. Sel. Areas Commun.*, vol. 12, pp. 1434–1444, Oct. 1994.
[11] S. Tabbane, "Location management methods for third-generation mobile system," *IEEE Commun. Mag.*, pp. 72–84, Aug. 1997.
[12] W. Ma and Y. Fang, "Two-level pointer forwarding strategy for location management in PCS networks," *IEEE Trans. Mobile Comput.*, vol. 1, pp. 32–45, Jan.–Mar. 2002.

Wenchao Ma (S'99–M'03) received the B.S. and M.S. degrees from Beijing University of Posts and Telecommunications, Beijing, China, in 1995 and 1998, respectively, and the Ph.D. degree from the University of Florida, Gainesville, in 2003.

From 1998 to 1999, he was an Engineer with China Telecom. He is currently with Microsoft Research Asia as an Associate Researcher. His research interests include wireless multimedia networks, mobility management, mobile computing, mobile IP, and broadband wireless access.



Yuguang Fang (S'92–M'94–SM'99) received the Ph.D. degree in systems and control engineering from Case Western Reserve University, Cleveland, OH, in 1994 and the Ph.D. degree in electrical engineering from Boston University, Boston, MA, in 1997.

From 1997 to 1998, he was a Visiting Assistant Professor in the Department of Electrical Engineering, University of Texas at Dallas. From 1998 to 2000, he was an Assistant Professor in the Department of Electrical and Computer Engineering, New Jersey Institute of Technology, Newark. In 2000, he joined the Department of Electrical and Computer Engineering, University of Florida, Gainesville, where he has been an Associate Professor since 2003. He has published more than 100 papers in refereed professional journals and conferences. He is an Editor of *ACM Wireless Networks.*

Prof. Fang is currently an Editor for many journals, including IEEE TRANSACTIONS ON COMMUNICATIONS, IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, and IEEE TRANSACTIONS ON MOBILE COMPUTING. He is also an active participant in conference organization such as the Program Vice-Chair for IEEE INFOCOM'2005, Program Cochair for the Global Internet and Next Generation Networks Symposium in IEEE Globecom'2004, and Program Vice Chair for the 2000 IEEE Wireless Communications and Networking Conference (WCNC'2000). He received the National Science Foundation Faculty Early Career Award in 2001 and the Office of Naval Research Young Investigator Award in 2002.