General Modeling and Performance Analysis for Location Management in Wireless Mobile Networks

Yuguang Fang, *Senior Member*, *IEEE*

Abstract—Location management plays a significant role in current and future wireless mobile networks in effectively delivering services to the mobile users on the move. Many schemes have been proposed and investigated extensively in the last decade. However, most performance analyses were carried out either under simplistic assumptions on some time variables or via simulations. In this paper, we present a new analytical approach to investigate the tradeoff (cost) analysis for location management schemes under fairly general assumption. For two specific location management schemes, *Pointer Forwarding Scheme (PFS)* and *Two-Location Algorithm (TLA)*, we present analytical formulae to compute the total costs. These formulae are easy to compute. Numerical results show that the traditional exponential model approximation may lead to a wrong decision in the tradeoff analysis. However, our analytical results can be easily used to find the appropriate parameters in PFS and TLA.

Index Terms—Mobility management, location management, pointer forwarding, two-location algorithm, cost analysis, wireless networks, mobile networks.

1 Introduction

In wireless mobile networks, in order to effectively deliver a service to a mobile user, the location of a called mobile user must be determined within a certain time limit (before the service is blocked). *Location management* is used to track the mobile users that move from place to place in the coverage of a wireless mobile network or in the coverage of multiple communications networks working together to fulfill the grand vision of ubiquitous communications. Thus, location management (also called *mobility management*) is a key component for the effective operations of wireless networks to deliver wireless Internet services (see [5] and references therein).

Wireless networks provide services to their subscribers in the coverage area. Such an area is populated with base stations, each of which is responsible for relaying communications with the mobiles traveling in its coverage called *cells*. A group of cells form a *registration area* (RA), which is managed by a *mobile switching center* (MSC) connecting directly to the *Public Switched Telephone Networks* (PSTN). The wireless networks standards IS-41 ([1]) and GSM MAP ([2]) use two-level strategies for mobility management in that a two-tier system of Home Location Register (HLR) and Visitor Location Register (VLR) databases is deployed (see Fig. 1). The HLR stores the user profiles of its registered mobiles, which contain information such as the mobile's identification number, the type of services subscribed, the quality of service (QoS) requirements, and the current location information. The VLR stores replications of users profiles and the temporary identification number for those mobiles which are currently visiting the associated RA. There are two basic operations in location management: *registration* and *location tracking*. The registration is the process in which a mobile informs the system of its current location, the location tracking is the process in which the system locates its mobile in order to deliver a call service to the mobile. When a user subscribes a service to a wireless network, a record of the mobile user's profile will be created in the system's HLR. Whenever and wherever the mobile user travels in the system's coverage, the mobile's location will be reported to the HLR (registration) according to some strategies. Then, the location in HLR will be used to locate (*find*) the mobile. When the mobile visits a new RA, a temporary record will be created in the VLR of the visited system and the VLR will then send a registration message to the HLR. All the signaling messages are exchanged over the overlaying signaling network using signaling system 7 (SS7) standard.

To deliver a call service to a mobile from an originating switch (MSC), the HLR is queried to find the current VLR of the mobile so that the HLR sends a query to the VLR. The VLR, upon finding the mobile in its charging area, will return a routable address, called the *temporary-location directory number* (TLDN), to the originating switch through the HLR. Based on the TLDN, the originating switch will set up the trunk (e.g., voice circuit with enough bandwidth for voice service) to the mobile. Thus, the call-delivery consists of two parts: *find*, the method to locate the mobile, and trunk setup for the mobile using TLDN. The signaling traffic due to *find* and *registration* can be significant. Various schemes to reduce such traffic have been proposed. In [14], the location cache scheme was proposed and shown to be a

• *The author is with the Department of Electrical and Computer Engineering, University of Florida, 435 Engineering Building, PO Box 116130, Gainesville, FL 32611-6130.*
E-mail: fang@ece.ufl.edu.

Manuscript received 15 July 2001; revised 15 May 2002; accepted 20 May 2002.
For information on obtaining reprints of this article, please send e-mail to: tc@computer.org, and reference IEEECS Log Number 116589.

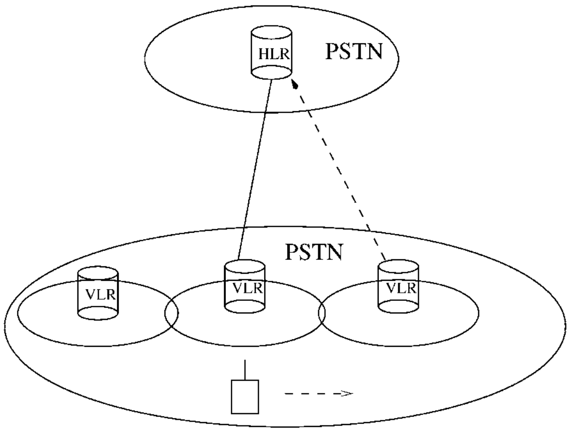


Fig. 1. Two-tier location management architecture.

significant improvement over the IS-41 scheme when the frequency of the incoming calls is high with respect to the mobility. To reduce the signaling traffic from the mobile to HLR, the *pointer forwarding scheme* ([13]) was proposed, based on the observation that it may be better to set up a forwarding pointer from the previous VLR to avoid more expensive registration from the mobile. By storing a location profile, registration traffic can also be reduced; this idea leads to the *alternative location algorithm* (ALA) and *two location algorithm* (TLA) ([19] and references therein). By observing that signaling cost can be significantly affected by the location database distribution, Ho and Akyildiz proposed the *local anchor scheme* to localize the registration traffic ([11]) and the dynamic hierarchical database architecture using *directory registers* ([12]). Recently, Ma and Fang ([22]) proposed a two-level pointer forwarding scheme to combine the advantages of the pointer forwarding scheme and the local anchor scheme. Although most studies focus on second generation personal communications networks (PCN), all location management schemes will be useful in future generation wireless networks ([5]).

Signaling traffic cost relies on many factors such as the terminating call arrivals and the users' mobility. Cost analysis of most location management schemes was carried out under the assumption that most time variables are exponentially distributed. For example, the time between two *served* calls, which we called *interservice time* ([8]), was usually assumed to be exponentially distributed. Although the call arrivals terminating at a mobile, say, $\mathcal{T}$, may be approximated by Poisson process, the interservice time is not identical to the interarrival time due to the *busy-line effect* ([8]), i.e., some call arrivals for the mobile $\mathcal{T}$ may not be connected because $\mathcal{T}$ is serving another call, hence the served calls are, in fact, a "sampled" Poisson process and thus will most likely not be Poisson process. The reason we are interested in the interservice times is that, during the call connection of a mobile, no signaling traffic is necessary from this mobile since the network knows where the mobile is; only when the mobile is idle does it need to make location update. Thus, the location management is affected by the interservice times rather than the interarrival times to the mobile. Moreover, due to the new trend of applications and user habits, even the interarrival time for the terminating calls to a mobile may not be exponentially distributed anymore. Some adaptive or dynamic schemes for choosing some location management parameters depend on the explicit form of cost ([20]); whether such schemes are still effective or not when the interservice times are not exponentially distributed is in question, there is no justification in the literature. In this paper, we will develop a new approach using more general modeling for the time variables involved and present general analytical results for the signaling cost analysis. Our results can be used to investigate the dynamic location management schemes. We will concentrate only on two location management schemes: *Two Location Algorithm* (TLA) and *Pointer Forwarding Scheme* (PFS).

This paper is organized as follows: In the next section, we present the descriptions of two location management schemes we investigate in this paper. We then present the probability distribution of the number of RA crossings in the Section 3, which is crucial for the cost analysis. In Section 4, we present the general analytical results for the signaling costs for TLA and PFS. We will conclude this paper in the Section 5.

2 Location Management Schemes

There are many location management schemes in the literature; [5] presents a very comprehensive survey on all aspects of mobility management. In this paper, we concentrate only on two schemes, the *Two Location Algorithm* (TLA) and *Pointer Forwarding Scheme* (PFS) in order to present our new analytical approach. This approach can, in fact, be applied to the analysis of other schemes.

2.1 IS-41 Scheme

Before we present TLA and PFS, we briefly go over the IS-41 scheme (or GSM MAP). We use the terminology used in [13]. An operation *move* means that a mobile user moves from one RA to another, while an operation *find* is the process to determine the RA a mobile user is currently visiting. The *move* and *find* in second generation location management schemes (such as in IS-41 or GSM MAP) are called *basic move* and *basic find*. In the *basic move* operation, a mobile detects if it is in a new RA. If it is, it will send a registration message to the new VLR; the VLR will send a message to the HLR. The HLR will send a deregistration message to the old VLR, which will, upon receiving the deregistration message, send the cancellation confirmation message to the HLR. The HLR will also send a cancellation confirmation message to the new VLR. In the *basic find*, a call to a mobile $\mathcal{T}$ is detected at a local switch. If the called party is in the same RA, the connection can be set up directly without querying the HLR. Otherwise, the local switch (VLR) queries the HLR for the callee, then HLR will query the callee's VLR. Upon receiving the callee's location, the HLR will forward the location information to the caller's local switch, which will then establish the call connection with the callee.

2.2 Pointer Forwarding Scheme (PFS)

The PFS modifies the *move* and *find* as follows: When a mobile $\mathcal{T}$ moves from one RA to another, it will inform its local switch (and VLR) at the new RA, which will determine whether to invoke the *basic move* or the *forwarding move*. In the *forwarding move*, the new VLR exchanges messages with the old VLR to set up a pointer from the old VLR to the new VLR, but does not involve the HLR. A subsequent call to the mobile $\mathcal{T}$ from some other switches will invoke the *forwarding find* procedure to locate the mobile: queries to the mobile's HLR, as in the *basic find*, and obtains a "potentially outdated" pointer to the old VLR, which will then direct the *find* to the new VLR using the pointer to locate the mobile $\mathcal{T}$. To ensure that the time taken by the *forwarding find* is within the tolerable time limit, the length of the chain of the forwarding pointers must be limited. This can be done by setting up the threshold for chain length to be a number, say, K, i.e., whenever the mobile $\mathcal{T}$ crosses K RA boundaries, it will register itself through the *basic move* (i.e., basic registration with HLR). In this way, the signaling traffic between the mobile and HLR can potentially be curbed.

2.3 Two Location Algorithm (TLA)

In the TLA, a mobile $\mathcal{T}$ has a small built-in memory to store the addresses for the two most recently visited RAs. The record of $\mathcal{T}$ in the HLR also has an extra field to store the two corresponding two locations. The first memory location stores the most recently visited RA. The TLA guarantees that the mobile $\mathcal{T}$ is in one of the two locations. When the mobile $\mathcal{T}$ joins the network, it registers with the network and stores the location in its memory and updates the HLR with its location. When the mobile $\mathcal{T}$ moves to another registration area, it checks whether the RA is in the memory or not. If the new RA is not in the memory, the most recently visited (MRV) RA in the memory is moved to the second memory entry, while the new RA is stored in the MRV position in the memory of $\mathcal{T}$. At the same time, a registration operation is performed to make the same modification in the HLR record. If the address for the new RA is already in the memory, the two locations in the memory of $\mathcal{T}$ are swapped and no registration is needed and no action is taken in HLR record. Thus, in TLA, no registration is performed when a mobile moves back and forth in two locations. The consequence is that the location entries in the mobile and HLR may not be exactly the same: The MRV RA in HLR may not be the MRV RA in reality!

When a call arrives for the mobile $\mathcal{T}$, the two addresses are used to find the actual location of the mobile $\mathcal{T}$. The order of the addresses used to locate $\mathcal{T}$ will affect the performance of the algorithm. If $\mathcal{T}$ is located in the first try (i.e., a *location hit*), then the *find* cost is the same as the one in the IS-41 scheme. Otherwise, the second try (due to the *location miss* in the first try) will find $\mathcal{T}$, which incurs additional cost. This additional cost has to be lower than the saved registration cost in order to make the TLA effective, which will demand a tradeoff analysis.

3 Probability of the Number of RA Boundary Crossings

In order to carry out the performance analysis of such location management schemes, we need to model some of the time variables appropriately and compute the probability distribution of the number of RA boundary crossings. The main results in this section have been presented in [8]. For completeness of the paper, we present the main results which will be used in the subsequent development.

Assuming that the incoming calls to a mobile, say, $\mathcal{T}$, form a Poisson process, the time the mobile stays in a registration area (RA) (also called the *RA residence time*) has a general nonlattice distribution. We will derive the probability $\alpha(K)$ that a mobile moves across K RAs between two served calls arriving to the mobile $\mathcal{T}$. The time between the end of a call served and the start of the following call served by the mobile is called *the interservice time*. It is possible that a call arrives while the previous call served is still in progress ([19]). In this case, the mobile $\mathcal{T}$ cannot accept the new call (the caller senses a busy tone in this case and may hang up). Thus, the interarrival (intercall) times for the calls terminated at the mobile $\mathcal{T}$ are different from the interservice times. This phenomenon is called the *busy line* effect. We emphasize here that, for location management, it is the interservice times, not the interarrival times of calls terminating at the mobile that affects the location update cost because, when a mobile is engaging a call connection, the wireless network knows where the mobile $\mathcal{T}$ is, hence, the mobile does not need to carry out any location update during the call connection. Although the incoming calls form a Poisson process (i.e., the interarrival times are exponentially distributed), the interservice times may not be exponentially distributed. Interservice time bears a similarity to the interdeparture time of a queuing system with blocking, which has been shown to be nonexponential in general. By ignoring the busy line effect and the service time, Lin ([19]) is able to derive an analytical formula for $\alpha(K)$, which has been subsequently used ([4], [3], [19], [26]) for tradeoff analysis for location update and paging. In this section, we assume that the interservice times are generally distributed and derive an analytic expression for $\alpha(K)$.

Let $t_1, t_2, \ldots$ denote the RA residence times and let r_1 denote the residual RA residence time (i.e., the time interval between the time instant the mobile $\mathcal{T}$ registers to the network and the time instant the mobile exits the first RA). Let t_c denote the interservice time between two consecutive served calls to a mobile $\mathcal{T}$. Fig. 2 shows the time diagram for K RA boundary crossings. Suppose that the mobile is in an RA R_1 when the previous call arrives and accepted by $\mathcal{T}$, it then moves K RAs during the interservice time and $\mathcal{T}$ resides in the *j*th RA for a period t_j ($1 \le j \le K+1$). In this paper, we consider a homogeneous wireless mobile network, i.e., all RAs in the network are statistically identical. We also assume that $t_1, t_2, \ldots$ are independent and identically distributed (iid) with a general probability density function $f(t)$. We want to point out that the independence assumption is crucial for our derivation, it is not known yet in the current literature how all performance analysis in the wireless networks can be carried out analytically without this assumption. Let t_c be generally distributed with probability density function $f_c(t)$ and let $f_r(t)$ be the probability density function of r_1. Let $f^*(s)$, $f_c^*(s)$, and $f_r^*(s)$ denote the Laplace-Stieltjes (L-S) transforms (or simply Laplace transforms) of $f(t)$, $f_c(t)$, and $f_r(t)$, respectively. Let $E[t_c] = 1/\lambda_c$ and $E[t_i] = 1/\lambda_m$. From the residual life theorem ([16]), we have

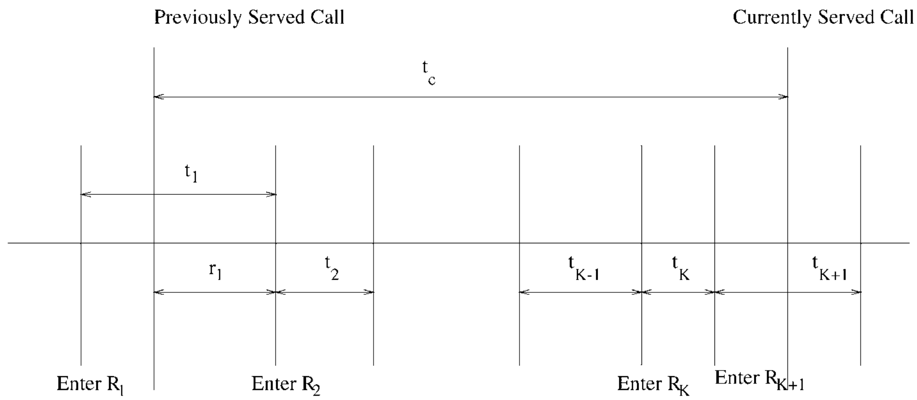


Fig. 2. The time diagram of K RA boundary crossings.

$$
\begin{aligned}
f_r(t) &= \lambda_m \int_t^\infty f(\tau)d\tau = \lambda_m[1-F(t)], \\
f_r^*(s) &= \frac{\lambda_m}{s}[1-f^*(s)],
\end{aligned} \tag{1}
$$

where $F(t)$ is the distribution function of $f(t)$. As a remark, the residual life theorem is valid only when we deal with the steady-state case; we can independently model the RA residence time r_1 in the initiating RA, i.e., we treat r_1 as a random variable which does not have any relationship to the RA residence time t_i. The results we presented in this paper can be applied to this very general case. It is obvious that the probability $\alpha(K)$ is given by

$$\alpha(0) = \Pr[t_c \le r_1], \quad K = 0, \tag{2}$$

$$
\begin{aligned}
\alpha(K) = &\Pr[r_1 + t_2 + \cdots + t_K < t_c \le r_1 + t_2 + \cdots + t_{K+1}], \\
&K \ge 1.
\end{aligned} \tag{3}
$$

Applying the inverse Laplace transform, we can compute $\alpha(0)$ from (2) as follows:

$$
\begin{aligned}
\alpha(0) &= \int_0^\infty \Pr(r_1 \ge t) f_c(t)dt \\
&= \int_0^\infty \frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty} \frac{1-f_r^*(s)}{s} e^{st} ds f_c(t)dt \\
&= \frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty} \frac{1-f_r^*(s)}{s} \int_0^\infty f_c(t)e^{st}dtds \\
&= \frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty} \frac{1-f_r^*(s)}{s} f_c^*(-s)ds,
\end{aligned} \tag{4}
$$

where σ is a sufficiently small positive number which is appropriately chosen for the inverse Laplace transform.

For $K > 0$, $\alpha(K)$ is computed as follows: First, we need to compute $\Pr(r_1 + t_2 + \cdots + t_k \le t_c)$ for any $k > 0$. Let $\xi = r_1 + t_2 + \cdots + t_k$. Let $f_\xi(t)$ and $f_\xi^*(s)$ be the probability density function and the Laplace transform of ξ. From the independence of $r_1, t_2, t_3, \cdots$, we have

$$f_\xi^*(s) = E[e^{-s\xi}] = E[e^{-sr_1}]\prod_{i=2}^{k} E[e^{-st_i}] = f_r^*(s)(f^*(s))^{k-1}.$$

Thus, the probability density function is given by

$$f_\xi(t) = \frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty} f_r^*(s)(f^*(s))^{k-1}e^{st}ds.$$

Noticing that the Laplace transform of $\Pr(\xi \le t)$ is $f_\xi^*(s)/s$, applying the inverse Laplace transform, we have

$$
\begin{aligned}
&\Pr(r_1 + t_2 + \cdots + t_k \le t_c) = \int_0^\infty \Pr(\xi \le t) f_c(t)dt \\
&= \int_0^\infty \frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty} \frac{f_r^*(s)[f^*(s)]^{k-1}}{s} e^{st} ds f_c(t)dt \\
&= \frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty} \frac{f_r^*(s)[f^*(s)]^{k-1}}{s} f_c^*(-s)ds.
\end{aligned}
$$

Taking this into (3), we obtain

$$
\begin{aligned}
\alpha(K) &= \Pr(t_c \ge r_1 + t_2 + \cdots + t_K) \\
&\quad - \Pr(t_c \ge r_1 + t_2 + \cdots + t_{K+1}) \\
&= \frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty} \frac{f_r^*(s)[f^*(s)]^{K-1}[1-f^*(s)]}{s} f_c^*(-s)ds,
\end{aligned} \tag{5}
$$

where σ is a sufficiently small positive number. In summary, we obtain:

Theorem 1. *If the probability density function of the interservice time has only finite possible isolated poles (which is the case when it has a rational Laplace transform), then the probability $\alpha(K)$ that a mobile moves across K RAs during the interservice time is given by*

$$
\begin{aligned}
\alpha(0) &= \frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty} \frac{1-f_r^*(s)}{s} f_c^*(-s)ds \\
\alpha(K) &= \frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty} \frac{f_r^*(s)[1-f^*(s)][f^*(s)]^{K-1}}{s} f_c^*(-s)ds, \\
&\quad K > 0,
\end{aligned} \tag{6}
$$

where σ is a sufficiently small positive number and $f_r^*(s) = \lambda_m(1-f^*(s))/s$.

4 Cost Analysis

Signaling traffic in the location management schemes incurs costs from the operations of *move* and *find*. Most cost analysis for *move* and *find* was carried out under the assumption that some of the time variables are exponentially distributed. The conclusions drawn from such results may not be extended to cases when such an assumption is not valid and the adaptive schemes for choosing certain parameters (such as the threshold for the number of pointers) may not be appropriate, accordingly. There are not many works handling nonexponential situations. In this paper, we relax this assumption so that more general cases can be dealt with analytically. Although we only study three location management schemes, our techniques can, in fact, be extended to analyze other location management schemes. We will postpone such studies to the future.

4.1 Cost Analysis for IS-41

As a baseline comparative study, we present the cost analysis for IS-41 first. We use the same notation as before, t_c denotes the interservice time (the interarrival time for calls terminated at a mobile $\mathcal{T}$ if we neglect the busy-line effect) with average $1/\lambda_c$ and t_i denotes the RA residence time with average $1/\lambda_m$. Let M and F denote the total cost for *basic move* during the interservice time and the total cost for *basic find*, respectively (i.e., the costs incurred in the IS-41 scheme). Since all location management schemes will go through the *move* and *find* whenever a terminating call to a mobile $\mathcal{T}$ arrives, the interservice time forms the fundamental regenerative period for cost analysis; thus, we only need to consider the signaling traffic incurred during this period.

For the IS-41 scheme, whenever the mobile crosses an RA boundary, a registration will be triggered. We assume that the unit cost for a basic registration (i.e., *basic move*) is m. From Theorem 1, M will be equal to the product of m and the average number of registrations incurred during the interservice time, given by

$$\begin{aligned}
M &= m\sum_{K=0}^{\infty} K\alpha(K) \\
&= \frac{m}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty} \frac{f_r^*(s)[1-f^*(s)]}{s}\left(\sum_{K=1}^{\infty} K(f^*(s))^{K-1}\right) f_c^*(-s)ds \\
&= \frac{m}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty} \frac{\lambda_m[1-f^*(-s)]^2}{s^2}\cdot\frac{1}{[1-f^*(-s)]^2} f_c^*(-s)ds \\
&= \frac{m\lambda_m}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty} \frac{1}{s^2} f_c^*(-s)ds \\
&= m\lambda_m\int_0^{\infty} f_c(t)\left(\frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty}\frac{1}{s^2}e^{st}ds\right)dt \\
&= m\lambda_m\int_0^{\infty} f_c(t)tdt = \frac{m\lambda_m}{\lambda_c} = \frac{m}{\rho},
\end{aligned}$$

where $\rho = \lambda_c/\lambda_m$, which is called the *call-to-mobility ratio*. If we use a more general distribution model for r_1, then we should use the following general formula:

$$\begin{aligned}
M &= \frac{m}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty} \frac{f_r^*(s)}{s[1-f^*(s)]} f_c^*(-s)ds \\
&= -m\sum_{p\in\sigma_c} \mathrm{Res}_{s=p}\frac{f_r^*(s)}{s[1-f^*(s)]} f_c^*(-s),
\end{aligned} \tag{7}$$

where σ_c is the set of poles of $f_c^*(-s)$ and $\mathrm{Res}_{s=p}$ indicates the residue of an analytic function at a pole $s = p$. Notice, in the last step, we have used the Residue Theorem from complex analysis ([16]).

The *basic find* operation consists of two parts. The first part includes the interactions between the originating switch and the HLR, while the second part includes the interactions between the HLR, the VLR, the MSC (terminating switch), and the mobile. Without loss of generality, we can assume that *basic find* unit cost and the unit cost of the first part is δ $(0 \le \delta \le 1)$, hence, the second part will be $1-\delta$. We observe that the message exchanges in the first part of *basic find* are almost identical to those in the *basic move*, hence we have $m = \delta$. During one interservice time, we only need one *find* operation. In summary, we obtain the total cost for IS-41 during the interservice time is

$$\begin{aligned}
\mathcal{C}_{IS-41} &= M + F \\
&= -\delta\sum_{p\in\sigma_c} \mathrm{Res}_{s=p}\frac{f_r^*(s)}{s[1-f^*(s)]} f_c^*(-s) + 1 = \frac{\delta}{\rho} + 1.
\end{aligned} \tag{8}$$

From this result, we observe that the total cost for IS-41 only depends on the CMR (i.e., the first moments of RA residence time and interarrival time of terminating calls).

4.2 Cost Analysis for Pointer Forwarding

In this subsection, we develop a more general model to evaluate the performance of the pointer forwarding scheme. Let M' and F' denote the corresponding costs for the pointer forwarding scheme in which every K moves will trigger a new registration. Let S denote the cost of setting up a forwarding pointer between VLRs during a pointer forwarding *move* and let T denote the cost of traversing a forwarding pointer between VLRs during a pointer forwarding *find*. We first derive M' and F'.

Suppose that a mobile $\mathcal{T}$ crosses i RA boundaries, then there are $i - \lfloor i/K \rfloor$ pointer creations (every K moves require $K-1$ pointer creations) and the HLR is updated $\lfloor i/K \rfloor$ times (with pointer forwarding, the mobile $\mathcal{T}$ registers every Kth move). Here, we use $\lfloor x \rfloor$ to denote the floor function, i.e., the largest integer not exceeding x. Thus, we have

$$\begin{aligned}
M' &= \sum_{i=0}^{\infty}\left[\left(i-\left\lfloor\frac{i}{K}\right\rfloor\right)S + \left\lfloor\frac{i}{K}\right\rfloor m\right]\alpha(i) \\
&= S\sum_{i=0}^{\infty} i\alpha(i) + (m-S)\sum_{i=0}^{\infty}\left\lfloor\frac{i}{K}\right\rfloor\alpha(i) \\
&= S\sum_{i=0}^{\infty} i\alpha(i) + (m-S)\sum_{r=0}^{\infty} r\left(\sum_{i=rK}^{(r+1)K-1}\alpha(i)\right).
\end{aligned} \tag{9}$$

Let

$$X(K) = \sum_{i=1}^{\infty} r\left(\sum_{i=rK}^{(r+1)K-1}\alpha(i)\right)$$

$$S(n) = \sum_{k=1}^{n-1}\alpha(k),$$

then, from Theorem 1, we obtain

$$\begin{aligned}S(n) &= \sum_{i=1}^{n-1} \frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty} \frac{f_r^*(s)[1-f^*(s)][f^*(s)]^{i-1}}{s} f_c^*(-s)ds \\ &= \frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty} \frac{f_r^*(s)[1-f^*(s)]}{s}\left(\sum_{i=1}^{n-1}[f^*(s)]^{i-1}\right) f_c^*(-s)ds \\ &= \frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty} \frac{f_r^*(s)[1-(f^*(s))^{n-1}]}{s} f_c^*(-s)ds.\end{aligned} \tag{10}$$

Moreover, we have

$$\begin{aligned}\sum_{i=1}^{N} s(iK) &= \sum_{i=1}^{N} \frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty} \frac{f_r^*(s)[1-(f^*(s))^{iK-1}]}{s} f_c^*(-s)ds \\ &= \frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty} \frac{f_r^*(s)}{s}\left\{\sum_{i=1}^{N}\left[1-(f^*(s))^{iK-1}\right]\right\} f_c^*(-s)ds \\ &= \frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty} \frac{f_r^*(s)}{s}\left\{N-(f^*(s))^{K-1}\sum_{i=1}^{N}\left[(f^*(s))^{K}\right]^{i-1}\right\} \\ &\quad f_c^*(-s)ds \\ &= \frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty} \frac{f_r^*(s)}{s}\left\{N-\frac{(f^*(s))^{K-1}[1-(f^*(s))^{NK}]}{1-(f^*(s))^{K}}\right\} \\ &\quad f_c^*(-s)ds.\end{aligned} \tag{11}$$

Thus, from (10) and (11), we obtain

$$\begin{aligned}X(K) &= \sum_{r=1}^{\infty} r[S((r+1)K)-S(rK)] \\ &= \lim_{N\to\infty}\left\{\sum_{r=1}^{N} r[S((r+1)K)-S(rK)]\right\} \\ &= \lim_{N\to\infty}\left\{NS((N+1)K)-\sum_{r=1}^{N}S(rK)\right\} \\ &= \lim_{N\to\infty}\frac{1}{2\pi j}\left\{\int_{\sigma-j\infty}^{\sigma+j\infty}\frac{f_r^*(s)}{s}\left[N-N(f^*(s))^{(N+1)K-1}\right]f_c^*(-s)ds\right. \\ &\quad \left.\int_{\sigma-j\infty}^{\sigma+j\infty}\frac{f_r^*(s)}{s}\left[N-\frac{(f^*(s))^{K-1}[1-(f^*(s))^{NK}]}{1-(f^*(s))^{K}}\right]f_c^*(-s)ds\right\} \\ &= \lim_{N\to\infty}\frac{1}{2\pi j}\left\{\int_{\sigma-j\infty}^{\sigma+j\infty}\frac{f_r^*(s)}{s}\left[\frac{(f^*(s))^{K-1}}{s[1-(f^*(s))^{K}]}-\frac{f^*(s))^{(N+1)K-1}}{1-(f^*(s))^{K}}\right.\right. \\ &\quad \left.\left.-N(f^*(s))^{(N+1)K-1}\right]f_c^*(-s)ds\right\} \\ &= \frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty}\frac{f_r^*(s)(f^*(s))^{K-1}}{s[1-(f^*(s))^{K}]}f_c^*(-s)ds \\ &\quad -\lim_{N\to\infty}\frac{1}{2\pi j}\left\{\int_{\sigma-j\infty}^{\sigma+j\infty}\frac{f_r^*(s)[N+1-N(f^*(s))^{K}]}{s[1-(f^*(s))^{K}]}\right. \\ &\quad \left.[f^*(s)]^{(N+1)K-1}f_c^*(-s)ds\right\} \\ &= \frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty}\frac{f_r^*(s)(f^*(s))^{K-1}}{s[1-(f^*(s))^{K}]}f_c^*(-s)ds,\end{aligned} \tag{12}$$

where, in the last equation, we have used the following fact: For any positive number $\sigma > 0$, for any complex number s with the real part $Re(s) \geq \sigma$, we always have

$$|f^*(s)| \leq \int_0^{\infty} f(t)|e^{-st}|dt \leq \int_0^{\infty} f(t)e^{-\sigma t}dt = f^*(\sigma) < 1,$$

and it can be shown that, for any analytic function $g(s, N)$, which is a polynomial in N, we always have

$$\lim_{N\to\infty}\int_{\sigma-j\infty}^{\sigma+j\infty} g(s,N)[f^*(s)]^{(N+1)K-1}f_c^*(-s)ds = 0.$$

Noticing that $X(1) = \sum_{i=0}^{\infty} i\alpha(i)$, from (9), we obtain

$$\begin{aligned}M' &= SX(1) + (m-S)X(K) \\ &= \frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty}\left[\frac{Sf_r^*(s)}{s[1-f^*(s)]}+\frac{(m-S)f_r^*(s)(f^*(s))^{K-1}}{s[1-(f^*(s))^{K}]}\right] \\ &\quad f_c^*(-s)ds.\end{aligned} \tag{13}$$

Next, we derive F'. After the last *basic move* operation (if any), the mobile $\mathcal{T}$ crosses $n = i - K\lfloor i/K \rfloor$ RA boundaries. Let $\Theta(n)$ denote the number of pointers to be traced in order to find the mobile $\mathcal{T}$ in the pointer forwarding *find* operation. If the mobile visits an RA more than once (i.e., a "loop" exists among n moves), then $\Theta(n)$ may not need to trace n pointers, thus, $\Theta(n) \leq n$. From this argument and applying Theorem 1, we obtain

$$\begin{aligned}F' &= \sum_{i=0}^{\infty} T\Theta(i-K\lfloor i \rfloor)\alpha(i) + F \\ &= T\sum_{r=0}^{\infty}\sum_{k=0}^{K-1}\Theta(k)\alpha(rK+k) + F \\ &= T\sum_{k=0}^{K-1}\Theta(k)\left(\sum_{r=0}^{\infty}\alpha(rK+k)\right) + F \\ &= T\sum_{k=0}^{K-1}\Theta(k)\sum_{r=0}^{\infty}\frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty}\frac{f_r^*(s)[1-f^*(s)](f^*(s))^{rK+k-1}}{s} \\ &\quad f_c^*(-s)ds + F \\ &= T\sum_{k=0}^{K-1}\Theta(k)\frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty}\frac{f_r^*(s)[1-f^*(s)](f^*(s))^{k-1}}{s} \\ &\quad \left(\sum_{r=0}^{\infty}(f^*(s))^{rK}\right)f_c^*(s)ds + F \\ &= T\sum_{k=0}^{K-1}\Theta(k)\frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty}\frac{f_r^*(s)[1-f^*(s)](f^*(s))^{k-1}}{s[1-(f^*(s))^{K}]} \\ &\quad f_c^*(-s)ds + F \\ &= \frac{T}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty}\frac{f_r^*(s)[1-f^*(s)]}{s[1-(f^*(s))^{K}]}\left(\sum_{k=0}^{K-1}\Theta(k)(f^*(s))^{k-1}\right) \\ &\quad f_c^*(-s)ds + F.\end{aligned}$$

In summary and applying the Residue Theorem ([16]), we finally arrive at:

Theorem 2. *If the interservice time $f_c^*(s)$ has a finite number of poles (such as a proper rational function), then we have*

$$M' = \frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty}\left[\frac{Sf_r^*(s)}{s[1-f^*(s)]} + \frac{(m-S)f_r^*(s)(f^*(s))^{K-1}}{s[1-(f^*(s))^K]}\right] f_c^*(-s)ds = -\sum_{p\in\sigma_c}\mathrm{Res}_{s=p}\left[\frac{Sf_r^*(s)}{s[1-f^*(s)]} + \frac{(m-S)f_r^*(s)(f^*(s))^{K-1}}{s[1-(f^*(s))^K]}\right] f_c^*(-s) \quad (14)$$

$$F' = F + \frac{T}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty}\frac{f_r^*(s)[1-f^*(s)]}{s[1-(f^*(s))^K]}\left(\sum_{k=0}^{K-1}\Theta(k)(f^*(s))^{k-1}\right) f_c^*(-s)ds = F - \sum_{p\in\sigma_c}\mathrm{Res}_{s=p}\frac{f_r^*(s)[1-f^*(s)]}{s[1-(f^*(s))^K]}\left(\sum_{k=0}^{K-1}\Theta(k)(f^*(s))^{k-1}\right) f_c^*(-s), \quad (15)$$

where σ_c denotes the set of poles of $f_c^*(-s)$ and $\mathrm{Res}_{s=p}$ denotes the residue at the pole $s = p$.

If the interservice time is exponentially distributed with parameter λ_c, then we have $f_c^*(-s) = \lambda_c/(-s+\lambda_c)$, from Theorem 2, we obtain

$$M' = \frac{Sf_r^*(\lambda_c)}{1-f^*(\lambda_c)} + \frac{(m-S)f_r^*(\lambda_c)(f^*(\lambda_c))^{K-1}}{1-(f^*(\lambda_c))^K}$$

$$F' = F + \frac{Tf_r^*(\lambda_c)[1-f^*(\lambda_c)]}{1-(f^*(\lambda_c))^K}\left(\sum_{k=0}^{K-1}\Theta(k)(f^*(\lambda_c))^{k-1}\right),$$

which were obtained in [13] with a slight different form.

The worst case for F' would be when all pointers are traced, i.e., when $\Theta(n) = n$. In this case, we have the following result:

$$\begin{aligned} F' &= F + \frac{T}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty}\frac{f_r^*(s)[1-f^*(s)]}{s[1-(f^*(s))^K]}\left(\sum_{k=0}^{K-1}k(f^*(s))^{k-1}\right) f_c^*(-s)ds \\ &= F + \frac{T}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty}\frac{f_r^*(s)[1-f^*(s)]}{s[1-(f^*(s))^K]} \cdot \frac{1-K(f^*(s))^{K-1}+(K-1)(f^*(s))^K}{(1-f^*(s))^2} f_c^*(-s)ds \\ &= F + \frac{T}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty}\frac{f_r^*(s)[1-K(f^*(s))^{K-1}+(K-1)(f^*(s))^K]}{s[1-(f^*(s))^K][1-f^*(s)]} f_c^*(-s)ds \\ &= F - T\sum_{p\in\sigma_c}\mathrm{Res}_{s=p}\frac{f_r^*(s)[1-K(f^*(s))^{K-1}+(K-1)(f^*(s))^K}{s[1-(f^*(s))^K][1-f^*(s)]} f_c^*(-s). \end{aligned}$$

As we mentioned in the previous subsection, we can choose $F = 1$. Thus, the total cost for TLA during the interservice time can be computed as follows:

$$\mathcal{C}_{PFS} = M' + F' = 1 + \frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty} g(s)f_c^*(-s)ds = 1 - \sum_{p\in\sigma_c}\mathrm{Res}_{s=p}g(s)f_c^*(-s), \quad (16)$$

where

$$g(s) = \frac{Sf_r^*(s)}{s[1-f^*(s)]} + \frac{(m-S)f_r^*(s)(f^*(s))^{K-1}}{s[1-(f^*(s))^K]} + \frac{Tf_r^*(s)[1-f^*(s)]}{s[1-(f^*(s))^K]}\left(\sum_{k=0}^{K-1}\Theta(k)\right).$$

4.3 Cost Analysis for TLA

In the Two Location Algorithm (TLA), if HLR has a location miss for a call termination, i.e., the two-location for the called mobile is not the same as the one stored in the HLR, additional signaling traffic will be necessary to setup the call connection. Thus, the probability that the HLR has a location miss for a call setup, say, w, is important to capture the signaling traffic. In [19], this probability is derived under the assumption that the interservice time is exponential. In this subsection, we first derive a more general analytical result to compute this quantity under the assumption that the interservice time is generally distributed, then we present the cost analysis for the TLA.

From the argument in [19], $(1-\omega)$ is the probability that the HLR has the correct view of the latest visited RA when a call arrives (i.e., a location hit occurs and the *find* cost for TLA is the same as that for IS-41). A location hit occurs either when the mobile has not moved since last served call arrival or when the last registration is followed by an even number of moves during the interservice time, or when there are an even number of moves with no registration during the interservice time. Let w_1 denote the probability that there is no move during the interservice time, let w_2 denote the probability that the last served call is followed by an even number of moves without registration during the interservice time, and let w_3 denote the probability that there are an even number of moves with no registration during the interservice time. Then, we will have $1 - w = w_1 + w_2 + w_3$, i.e., $w = 1 - w_1 - w_2 - w_3$. For w_1, we have

$$w_1 = \Pr(r_1 \le t_c) = \alpha(0). \quad (17)$$

Let θ denote the probability of a move without registration. Let $w_2(K)$ denote the probability that the last registration is followed by an even number of moves without registration during the interservice time given that there are K moves during the interservice time. We can easily see that ([19])

$$w_2(K) = (1-\theta)\sum_{i=0}^{\lfloor (K-1)/2\rfloor}\theta^{2i} = \frac{1-\theta^{2\lfloor (K-1)/2\rfloor+2}}{1+\theta}, \quad K>0,$$

where $\lfloor x\rfloor$ indicates the floor function, i.e., the largest integer not exceeding x. Noticing that $w_2(2i+2) = w_2(2i+1)$ for $i \geq 0$, we have

$$\begin{aligned} w_2 &= \sum_{K=1}^{\infty} w_2(K)\alpha(K) = \sum_{i=0}^{\infty} w_2(2i+1)[\alpha(2i+1)+\alpha(2i+2)] \\ &= \frac{1}{1+\theta}\sum_{i=0}^{\infty}(1-\theta^{2(i+1)})[\alpha(2i+1)+\alpha(2i+2)] \\ &= \frac{1}{1+\theta}\sum_{i=1}^{\infty}(1-\theta^{2i})[\alpha(2i-1)+\alpha(2i)]. \end{aligned} \tag{18}$$

The probability w_3 can be computed as follows:

$$w_3 = \sum_{i=1}^{\infty}\theta^{2i}\alpha(2i).$$

Thus, applying Theorem 1, we obtain

$$\begin{aligned} w &= 1 - w_1 - w_2 - w_3 = 1-\alpha(0) \\ &\quad - \frac{1}{1+\theta}\sum_{i=1}^{\infty}(1-\theta^{2i})[\alpha(2i-1)+\alpha(2i)] - \sum_{i=1}^{\infty}\theta^{2i}\alpha(2i) \\ &= 1-\alpha(0) - \frac{1}{1+\theta}\sum_{i=1}^{\infty}[\alpha(2i-1)+\alpha(2i)] + \frac{1}{1+\theta}\sum_{i=1}^{\infty}\theta^{2i} \\ &\quad - \frac{\theta}{1+\theta}\sum_{i=1}^{\infty}\theta^{2i}\alpha(2i) \\ &= 1-\alpha(0) - \frac{1}{1+\theta}(1-\alpha(0)) + \frac{1}{1+\theta}\sum_{i=1}^{\infty}\theta^{2i} \\ &\quad - \frac{\theta}{1+\theta}\sum_{i=1}^{\infty}\theta^{2i}\alpha(2i) \\ &= \frac{\theta}{1+\theta}(1-\alpha(0)) + \frac{1}{1+\theta}\sum_{i=1}^{\infty}\frac{\theta^{2i}}{2\pi j} \\ &\quad \int_{\sigma-j\infty}^{\sigma+j\infty}\frac{f_r^*(s)[f^*(s)]^{2i-2}[1-f^*(s)]}{s}f_c^*(-s)ds \\ &\quad - \frac{\theta}{1+\theta}\sum_{i=1}^{\infty}\frac{\theta^{2i}}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty}\frac{f_r^*(s)[f^*(s)]^{2i-1}[1-f^*(s)]}{s} \\ &\quad f_c^*(-s)ds \\ &= \frac{\theta}{1+\theta}(1-\alpha(0)) + \frac{1}{1+\theta}\frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty}\frac{f_r^*(s)[1-f^*(s)]}{s} \\ &\quad \left[\sum_{i=1}^{\infty}\theta^{2i}[f^*(s)]^{2i-2}\right]f_c^*(-s)ds \\ &\quad - \frac{\theta}{1+\theta}\frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty}\frac{f_r^*(s)[1-f^*(s)]}{s}\left[\sum_{i=1}^{\infty}\theta^{2i}[f^*(s)]^{2i-1}\right] \\ &\quad f_c^*(-s)ds \\ &= \frac{\theta}{1+\theta} - \frac{\theta}{(1+\theta)2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty}\frac{1-f_r^*(s)}{s}f_c^*(-s)ds \\ &\quad + \frac{\theta^2}{(1+\theta)2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty}\frac{f_r^*(s)[1-f^*(s)]}{s[1-\theta^2 f^{*2}(s)]}f_c^*(-s)ds \\ &\quad - \frac{\theta^3}{(1+\theta)2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty}\frac{f_r^*(s)[1-f^*(s)]f^*(s)}{s[1-\theta^2 f^{*2}(s)]}f_c^*(-s)ds \\ &= \frac{\theta}{1+\theta} - \frac{\theta}{(1+\theta)2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty}\frac{1-f_r^*(s)}{s}f_c^*(-s)ds \\ &\quad + \frac{\theta^2}{(1+\theta)2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty}\frac{f_r^*(s)[1-f^*(s)](1-\theta f^*(s))}{s[1-\theta^2 f^{*2}(s)]} \\ &\quad f_c^*(-s)ds \\ &= \frac{\theta}{1+\theta} + \frac{\theta}{(1+\theta)2\pi j} \\ &\quad \int_{\sigma-j\infty}^{\sigma+j\infty}\left[-\frac{1-f_r^*(s)}{s} + \frac{\theta f_r^*(s)[1-f^*(s)]}{s[1+\theta f^*(s)]}\right]f_c^*(-s)ds \\ &= \frac{\theta}{1+\theta} + \frac{\theta}{(1+\theta)2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty}\left[-\frac{1}{s} + \frac{f_r^*(s)(1+\theta)}{s[1+\theta f^*(s)]}\right]f_c^*(-s)ds \\ &= \frac{\theta}{1+\theta}\left(1 - \frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty}\frac{1}{s}f_c^*(-s)ds\right) \\ &\quad + \frac{\theta}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty}\frac{f_r^*(s)}{s[1+\theta f^*(s)]}f_c^*(-s)ds \\ &= \frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty}\frac{\theta f_r^*(s)}{s[1+\theta f^*(s)]}f_c^*(-s)ds, \end{aligned} \tag{19}$$

where we have used the following result:

$$\begin{aligned} &\frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty}\frac{1}{s}f_c^*(-s)ds \\ &= \int_0^{\infty} f_c(t)\left(\frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty}\frac{1}{s}e^{st}dsdt\right) \\ &= \int_0^{\infty} f_c(t)dt = 1. \end{aligned}$$

Applying the Residue Theorem ([16]), we obtain:

Theorem 3. *If the interservice time is distributed with rational Laplace transform, the probability of a location miss in TLA is given by*

$$w = \frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty}\frac{\theta f_r^*(s)}{s[1+\theta f^*(s)]}f_c^*(-s)ds, \tag{20}$$

where σ is a sufficiently small positive number. If σ_c denotes the set of poles of $f_c^*(-s)$, then we have

$$w = -\sum_{p\in\sigma_c}\mathrm{Res}_{s=p}\frac{\theta f_r^*(s)}{s[1+\theta f^*(s)]}f_c^*(-s), \tag{21}$$

where $\mathrm{Res}_{s=p}$ denotes the residue at the pole $s=p$.

If the interservice time is exponentially distributed, which is the case studied in [19], then we have $f_c^*(s) = \lambda_c/(s+\lambda_c)$, hence, from Theorem 3, we obtain

$$w = -\mathrm{Res}_{s=\lambda_c}\frac{\theta f_r^*(s)}{s[1+\theta f^*(s)]}\cdot\frac{\lambda_c}{-s+\lambda_c} = \frac{\theta f_c^*(\lambda_c)}{1+\theta f^*(\lambda_c)},$$

which is exactly the same as in [19] after some simplification of the result in [19].

Now, we are ready to carry out the cost analysis for PFS. As for the IS-41 analysis, we assume that the unit cost for *basic find* is $m = 1$ with the cost δ for the first part of operation in *basic find* and $1 - \delta$ for the second part. Suppose that the mobile moves across K RAs during the interservice time. The conditional probability $\Pr[I = i|K]$ that i *registration* operations are performed among the K moves has a Bernoulli distribution:

$$\Pr[I = i|K] = \binom{K}{i}\theta^{K-i}(1-\theta)^i,$$

where θ is the probability that, when a mobile $\mathcal{T}$ moves, the new RA address is in the mobile's memory. Then, the average number of registrations during the interservice time for TLA is given by

$$\begin{aligned}
n_{TLA} &= \sum_{K=0}^{\infty}\sum_{i=0}^{K} i\Pr[I = i|K]\alpha(K) \\
&= \sum_{K=0}^{\infty}\left(\sum_{i=0}^{K} i\binom{K}{i}\theta^{K-i}(1-\theta)^i\right)\alpha(K) \\
&= \sum_{K=0}^{\infty} K(1-\theta)\alpha(K) = (1-\theta)X(1) \\
&= \frac{1-\theta}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty}\frac{f_r^*(s)}{s[1-f^*(s)]}f_c^*(-s)ds \\
&= -(1-\theta)\sum_{p\in\sigma_c}\mathrm{Res}_{s=p}\frac{f_r^*(s)}{s[1-f^*(s)]}f_c^*(-s),
\end{aligned}$$

where we have used (12). Thus, the total cost for registration during interservice time is $c_1 = \delta n_{TLA}$.

For the operations of the second part in the *basic find* for TLA, if we have a location hit (i.e., the location entries in the memories of both HLR and the mobile are identical), the *find* cost will be the same as in *basic find*; if there is a location miss, extra cost from HLR to the VLR (second part of the *basic find* operation) will be incurred, thus the total cost for the *find operation* in TLA will be

$$c_2 = (1-\omega)\cdot 1 + \omega[1 + (1-\delta)] = 1 + (1-\delta)\omega.$$

In summary, the total signaling cost during the interservice time for TLA is given by

$$\begin{aligned}
\mathcal{C}_{TLA} &= c_1 + c_2 = \delta n_1 + 1 + (1-\delta)\omega \\
&= 1 + \frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty}\left[\frac{(1-\theta)\delta f_r^*(s)}{s[1-f^*(s)]} + \frac{(1-\delta)\theta f_r^*(s)}{s[1+\theta f^*(s)]}\right] \\
&\quad f_c^*(-s)ds \\
&= 1 - \sum_{p\in\sigma_c}\mathrm{Res}_{s=p}\left[\frac{(1-\theta)\delta f_r^*(s)}{s[1-f^*(s)]} + \frac{(1-\delta)\theta f_r^*(s)}{s[1+\theta f^*(s)]}\right] \\
&\quad f_c^*(-s),
\end{aligned} \tag{23}$$

where σ_c is the set of poles of $f_c^*(-s)$. If we use (1), we can obtain the following:

$$\begin{aligned}
\mathcal{C}_{TLA} = 1 + \frac{(1-\theta)\delta}{\rho} - (1-\delta)\theta\lambda_m \\
\sum_{p\in\sigma_c}\mathrm{Res}_{s=p}\frac{1-f^*(s)}{s^2[1+\theta f^*(s)]}f_c^*(-s).
\end{aligned}$$

5 Interesting Special Cases

We notice that all general results presented so far can be easily applied when the interservice time is distributed with the rational Laplace transform. Since distributions with rational Laplace transforms are dense in the space of probability distributions, we can always use a distribution model with Laplace transform to approximate any given distribution to any desired accuracy. Recently, a few distribution models have been proposed to model various time variables in the wireless cellular networks; the hyper-Erlang model and the SOHYP (Sum of Hyper-Exponential) model are two of the important models (see [9] and references therein). Due to the simplicity and generality of the hyper-Erlang model, we will concentrate on this model in this section.

The *hyper-Erlang* distribution has the following probability density function and Laplace transform:

$$\begin{aligned}
f_{he}(t) &= \sum_{i=1}^{M} p_i\frac{(m_i\eta_i)^{m_i}t^{m_i-1}}{(m_i-1)!}e^{-m_i\eta_i t} \quad (t \ge 0), \\
f_{he}^*(s) &= \sum_{i=1}^{M} p_i\left(\frac{m_i\eta_i}{s+m_i\eta_i}\right)^{m_i},
\end{aligned} \tag{24}$$

where

$$p_i \ge 0, \quad \sum_{i=1}^{M} p_i = 1,$$

and $M, m_1, m_2, \ldots, m_M$ are nonnegative integers and $\eta_1, \eta_2, \ldots, \eta_M$ are positive numbers.

It has been demonstrated ([9], [15]) that the hyper-Erlang distribution can be used to approximate the distribution of any nonnegative random variable. Many distribution models, such as the exponential model, the Erlang model, and the hyper-exponential model, are special cases of the hyper-Erlang distribution. More importantly, the moments of a hyper-Erlang distribution can be easily obtained. If ξ is hyper-Erlang distributed, as in (24), then its kth moment is given by

$$E[\xi^k] = (-1)^k f_{he}^{*(k)}(0) = \sum_{i=1}^{M} p_i\frac{(m_i+k-1)!}{(m_i-1)!}(m_i\eta_i)^{-k}.$$

The parameters p_i, m_i, and η_i $(i = 1, 2, \ldots, M)$ can be found by fitting a number of moments from field data in practice.

Assuming now that the interservice time is hyper-Erlang distributed with distribution given in (24) with

$$\lambda_c = \left(\sum_{i=1}^{M} p_i/\eta_i\right)^{-1},$$

then, applying the Residue Theorem, we can easily obtain

$$\mathcal{C}_{PFS} = 1 - \sum_{i=1}^{M} p_i \frac{(-1)^{m_i}(m_i\eta_i)^{m_i}}{(m_i-1)!} g^{(m_i-1)}(m_i\eta_i)$$

$$\mathcal{C}_{TLA} = 1 + \frac{(1-\theta)\delta}{\rho} - (1-\delta)\theta\lambda_m \sum_{i=1}^{M} p_i \frac{(-1)^{m_i}(m_i\eta_i)^{m_i}}{(m_i-1)!} h^{(m_i-1)}(m_i\eta_i),$$

where

$$g(s) = \frac{\lambda_m S}{s^2} + \frac{\lambda_m(\delta - S)[1-f^*(s)][f^*(s)]^{K-1}}{s^2[1-(f^*(s))^K]} + \frac{\lambda_m T[1-f^*(s)]^2}{s^2[1-(f^*(s))^K]}\left(\sum_{k=0}^{K-1}\Theta(k)\right)$$

$$h(s) = \frac{1-f^*(s)}{s^2[1+\theta f^*(s)]}$$

and $x^{(i)}(s)$ denotes the ith derivative of the function $x(s)$ at point s.

If the interservice time is hyper-exponentially distributed with Laplace transform

$$f_c^*(s) = \sum_{i=1}^{M} p_i \frac{\eta_i}{s+\eta_i},$$

then we have

$$\mathcal{C}_{PFS} = 1 + \sum_{i=1}^{M} p_i\eta_i g(\eta_i)$$

$$\mathcal{C}_{TLA} = 1 + \frac{(1-\theta)\delta}{\rho} + (1-\delta)\theta\lambda_m \sum_{i=1}^{M} p_i\eta_i h(\eta_i).$$

If the interservice time is exponentially distributed with Laplace transform

$$f_c^*(s) = \frac{\lambda_c}{s+\lambda_c},$$

then we have

$$\mathcal{C}_{PFS} = 1 + \lambda_c g(\lambda_c)$$

$$\mathcal{C}_{TLA} = 1 + \frac{(1-\theta)\delta}{\rho} + (1-\delta)\theta\lambda_m\lambda_c h(\lambda_c).$$

6 Numerical Results

In this section, we present a few numerical results to show how distribution models affect the outcome of cost analysis and how approximation using the exponential model is inappropriate. As we mentioned earlier, all new proposed schemes for location management have to be compared against IS-41, a standard for second generation wireless systems and one which has been incorporated into the third generation wireless systems.

For illustration purpose, we choose the hyper-exponential distribution model for the interservice time and choose the Gamma distribution model for the RA residence time. We also use the distribution model with the same average interservice time to approximately model the interservice time and study whether the exponential model is enough for the cost analysis.

Fig. 3 shows the results for PFS. In this figure, the interservice time is modeled by the hyper-exponential distribution with the following parameters: $M = 2$, $p_1 = p_2 = 0.5$, $\eta_1 = 1$, and $\eta_2 = 0.5$, and with its average $1/\lambda_c = p_1/\eta_1 + p_2/\eta_2 = 1.5$. The RA cell residence time is characterized by the Gamma distribution with the parameters: $\gamma = 0.1$, λ_m varies. Other parameters are given as follows: $K = 10$, $\delta = 0.8$, $S = 0.1$, and $T = 0.05$. We also consider the worst case scenario for PFS: We assume $\Theta(n) = n$, i.e., there are no loops in the pointer chain, which gives the highest pointer tracing cost. From Fig. 3, we observe that there are significant differences between the total cost of PFS with the actual distribution for interservice time and the total cost of PFS with the exponential approximation. Particularly, we observe that the crossing points of cost curves with the cost curve for IS-41 are different, which implies that the wrong decision in the tradeoff analysis can be drawn if the exponential distribution model is used. For example, when $\rho = 2$, the cost curves show that IS-41 works better; however, if we use the exponential approximation, we conclude that PFS is still better than the IS-41; a wrong conclusion is drawn! One may suggest the following adaptive location management scheme: Between IS-41 and PFS, use the one with lower cost. If this is the case, then the switching decision point will depend on the probability distribution, hence the approximation by the exponential distribution model may not be appropriate. Therefore, since the distribution model for the interservice time does affect the tradeoff decision, we have to use the general formula we present in this paper.

Fig. 4 shows the result for TLA. In this example, the interservice distribution is modeled by the hyper-Erlang distribution with the following set of parameters: $p_1 = p_2 = 0.5$, $m_1 = 1$, $m_2 = 5$, $\eta_1 = 1$, $\eta_2 = 0.01$, $\delta = 0.2$, $\theta = 0.5$, and $\lambda_c = (p_1/\eta_1 + p_2/\eta_2)^{-1} = 0.0198$, while the RA cell residence time is exponentially distributed with the average varying. From this figure, we observe the following phenomena: When the call-to-mobility factor is low, the TLA scheme is better. The cost under exponential approximation deviates from the actual cost. If we use the exponential model, we may switch the location management scheme to the IS-41 when $\rho > 1$, although, in reality, the TLA still works better until $\rho > 2$. This implies that the probability distribution model of the inte-service time does play a significant role in the decision making process in the cost analysis.

7 Conclusions

In this paper, we have developed a new analytical approach to carrying out the cost analysis for location management in wireless mobile networks. We focus on two well-known location management schemes: the pointer forwarding scheme and two location algorithm (a special case of alternative location algorithm) and present analytical results for signaling traffic analysis under very general assumptions. These results can be used to investigate how to choose the design parameters in location management.

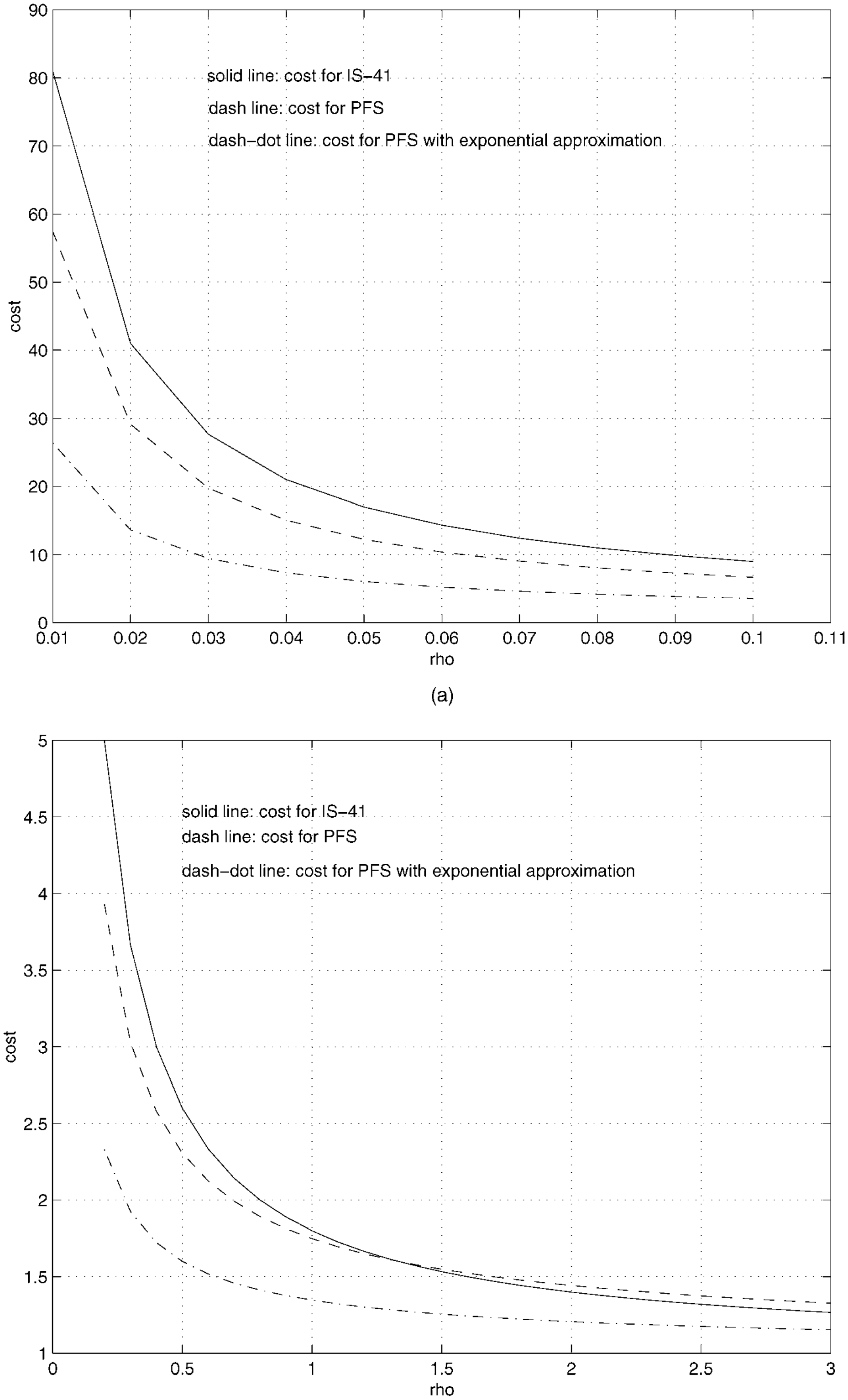


Fig. 3. Comparison results for PFS and IS-41: Solid line is for IS-41, dashed line is for PFS, dash-dot line is for PFS using exponential approximation for the interservice time.

ACKNOWLEDGMENTS

This work was supported in part by US National Science Foundation Faculty Early Career Development Award under grant ANI-0093241 and the US Office of Naval Research Young Investigator Award under grant N000140210464.

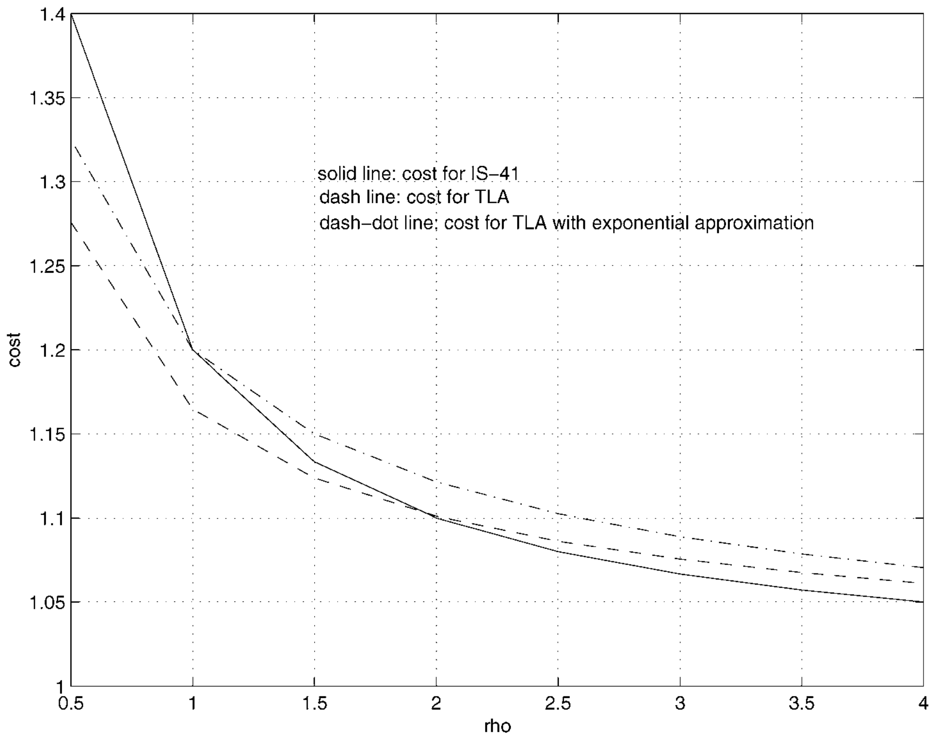


Fig. 4. Comparison results for TLA and IS-41: The solid line is for IS-41, the dashed line is for TLA, the dash-dot line is for TLA using exponential approximation for the interservice time.

REFERENCES

[1] EIA/TIA, "Cellular Radio-Telecommunications Intersystem Operations," EIA/TIA Technical Report IS-41 Revision B, 1991.

[2] ETSI, *Digital Cellular Telecommunications System (Phase 2+): Mobile Application Part (MAP) Specification (GSM 09.02 Version 7.51 Release)*, 1998.

[3] I.F. Akyildiz, J.S.M. Ho, and Y.-B. Lin, "Movement-Based Location Update and Selective Paging for PCS Networks," *IEEE/ACM Trans. Networking,* vol. 4, no. 4, pp. 629-638, 1996.

[4] I.F. Akyildiz and W. Wang, "A Dynamic Location Management Scheme for Next Generation Multi-Tier PCS Systems," *IEEE J. Selected Areas in Comm.,* to appear.

[5] I.F. Akyildiz, J. McNair, J.S.M. Ho, H. Uzunalioglu, and W. Wang, "Mobility Management in Next-Generation Wireless Systems," *Proc. IEEE,* vol. 87, no. 8, pp. 1347-1384, Aug. 1999.

[6] A. Bar-Noy, I. Kessler, and M. Sidi, "Mobile Users: To Update or Not Update?" *Wireless Networks,* vol. 1, no. 2, pp. 175-186, July 1994.

[7] Y. Fang, "Movement-Based Location Management and Tradeoff Analysis for Wireless Mobile Networks," *IEEE Trans. Computers,* to appear.

[8] Y. Fang, I. Chlamtac, and Y.B. Lin, "Portable Movement Modeling for PCS Networks" *IEEE Trans. Vehicular Technology,* vol. 49, no. 4, pp. 1356-1363, July 2000.

[9] Y. Fang and I. Chlamtac, "Teletraffic Analysis and Mobility Modeling for PCS Networks," *IEEE Trans. Comm.,* vol. 47, no. 7, pp. 1062-1072, July 1999.

[10] Y. Fang, I. Chlamtac, and Y.B. Lin, "Channel Occupancy Times and Handoff Rate for Mobile Computing and PCS Networks," *IEEE Trans. Computers,* vol. 47, no. 6, pp. 679-692, June 1998.

[11] J.S.M. Ho and I.F. Akyildiz, "Local Anchor Scheme for Reducing Signaling Costs in Personal Communications Networks," *IEEE/ACM Trans. Networks,* vol. 4, no. 5, pp. 709-725, Oct. 1996.

[12] J.S.M. Ho and I.F. Akyildiz, "Dynamic Hierarchical Database Architecture for Location Management in PCS Networks," *IEEE/ACM Trans. Networks,* vol. 5, no. 5, pp. 646-660, Oct. 1997.

[13] R. Jain and Y.B. Lin, "An Auxiliary User Location Strategy Employing Forwarding Pointers to Reduce Network Impacts of PCS," *ACM-Baltzer Wireless Networks,* vol. 1, no. 2, pp. 197-210, 1995.

[14] R. Jain, Y.B. Lin, C.N. Lo, and S. Mohan, "A Caching Strategy to Reduce Network Impacts of PCS," *IEEE J. Selected Areas in Comm.,* vol. 12, no. 8, pp. 1434-1445, 1994.

[15] F.P. Kelly, *Reversibility and Stochastic Networks.* New York: John Wiley & Sons, 1979.

[16] L. Kleinrock, *Queueing Systems: Theory, Volume I.* New York: John Wiley & Sons, 1975.

[17] J. Li, H. Kameda, and K. Li, "Optimal Dynamic Mobility Management for PCS Networks," *IEEE/ACM Trans. Networking,* vol. 8, no. 3, pp. 319-327, 2000.

[18] W. Li and A.S. Alfa, "A PCS Network with Correlated Arrival Process and Splitted-Rating Channels," *IEEE J. Selected Areas in Comm.,* vol. 17, no. 7, pp. 1318-1325, 1999.

[19] Y.B. Lin, "Reducing Location Update Cost in a PCS Network," *IEEE/ACM Trans. Networking,* vol. 5, no. 1, pp. 25-33, 1997.

[20] Y.B. Lin, "Determining the User Locations for Personal Communications Networks," *IEEE Trans. Vehicular Technology,* vol. 43, pp. 466-473, 1994.

[21] Y.B. Lin, S. Mohan, and A. Noerpel, "Queueing Priority Channel Assignment Strategies for Handoff and Initial Access for a PCS Network," *IEEE Trans. Vehicular Technology,* vol. 43, no. 3, pp. 704-712, 1994.

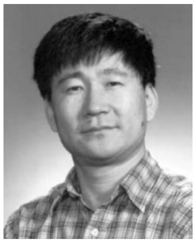
[22] W. Ma and Y. Fang, "Two-Level Pointer Forwarding Strategy for Location Management in PCS Networks," *IEEE Trans. Mobile Computing,* vol. 1, no. 1, pp. 32-45, Jan.-Mar. 2002.

[23] S. Mohan and R. Jain, "Two User Location Strategies for Personal Communications Services," *IEEE Personal Comm.,* pp. 42-50, 1994.

[24] C. Rose and R. Yates, "Minimizing the Average Cost of Paging under Delay Constraints," *Wireless Networks,* vol. 1, no. 2, pp. 211-219, 1995.

[25] S. Tabbane, "An Alternative Strategy for Location Tracking," *IEEE J. Selected Areas in Comm.,* vol. 13, no. 5, pp. 880-892, 1995.

[26] W. Wang and I.F. Akyildiz, "Intersystem Location Update and Paging Schemes for Multitier Wireless Networks," *Proc. ACM MobiCom 2000,* Aug. 2000.



Yuguang Fang (S'92-M'94-S'96-M'97-SM'99) received the BS and MS degrees in mathematics from Qufu Normal University, Qufu, Shandong, China, in 1984 and 1987, respectively, the PhD degree from the Department of Systems, Control, and Industrial Engineering at Case Western Reserve University, Cleveland, Ohio, in January 1994, and the PhD degree from the Department of Electrical and Computer Engineering at Boston University, Massachusetts, in May 1997. From 1987 to 1988, he held research and teaching positions in both the Department of Mathematics and the Institute of Automation at Qufu Normal University. He held a postdoctoral position in the Department of Electrical and Computer Engineering at Boston University from June 1994 to August 1995. From June 1997 to July 1998, he was a visiting assistant professor in the Department of Electrical Engineering at the University of Texas at Dallas. From July 1998 to May 2000, he was an assistant professor in the Department of Electrical and Computer Engineering at the New Jersey Institute of Technology, Newark, New Jersey. Since May 2000, he has been an assistant professor in the Department of Electrical and Computer Engineering at the University of Florida, Gainesville. His research interests span many areas, including wireless networks, mobile computing, mobile communications, automatic control, and neural networks. He has published more than 60 papers in refereed professional journals and conferences. He received the US National Science Foundation Faculty Early Career Development Award in 2001 and the US Office of Naval Research Young Investigator Award in 2002. He is listed in Marquis *Who's Who in Science and Engineering*, *Who's Who in America*, and *Who's Who in World*. Dr. Fang has actively engaged in many professional activities. He is a senior member of the IEEE and a member of the ACM. He is an editor for the *IEEE Transactions on Communications*, an editor for the *IEEE Transactions on Wireless Communications*, an editor for *ACM Wireless Networks*, an area editor for the *ACM Mobile Computing and Communications Review*, an associate editor for the *Wiley International Journal on Wireless Communications and Mobile Computing*, and feature editor for Scanning the Literature in *IEEE Personal Communications*. He was an editor for the *IEEE Journal on Selected Areas in Communications: Wireless Communications Series*. He has also been actively involved with many professional conferences such as ACM MobiCom '02, ACM MobiCom '01, IEEE INFOCOM '00, INFOCOM '98, IEEE WCNC '02, WCNC '00 (technical program vice-chair), WCNC '99, and the International Conference on Computer Communications and Networking (IC3N '98) (technical program vice-chair).