Chapter 16

# MOBILITY MANAGEMENT FOR WIRELESS NETWORKS: MODELING AND ANALYSIS

Yuguang Fang and Wenchao Ma
*Department of Electrical and Computer Engineering*
*University of Florida*
*435 Engineering Building, P.O.Box 116130*
*Gainesville, FL 32611*
*Tel: (352) 846-3043, Fax: (352) 392-0044*
fang@ece.ufl.edu

**Abstract** Mobility management plays a significant role in the current and the 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 past. However, most performance analyses were carried out either under simplistic assumptions on some time variables or via simulations. Recently, we have developed a new analytical approach to investigating the modeling and performance analysis for mobility management schemes under fairly general assumption. In this chapter, we present the techniques we have developed for this approach lately and summarize the major results we have obtained for a few mobility management schemes such as *Movement-based Mobility Management*, *Pointer Forwarding Scheme (PFS)*, *Two-level location management scheme*, and *Two-Location Algorithm (TLA)*.

**Keywords:** Mobility management, location management, registration, paging, cellular, wireless 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). When the call connection is in progress while the mobile is moving, the network has to follow the mobile users and allocate enough resource to provide seamless continuing service without user awareness that in fact the network facility (such as base station) is changing. *Mobility 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, *mobility management* is a key component for the effective operations of wireless networks to deliver wireless Internet services (see [6] and references therein).

Wireless networks provide services to their subscribers in the coverage area. Such area is populated with base stations, each of which is responsible for relaying communication services for the mobiles traveling in its coverage called *cells*. A group of cells form a *registration area (RA)*, which is managed by *mobile switching center* (MSC) connecting directly to the *Public Switched Telephone Networks (PSTN)*. In this chapter, we will use *a mobile, a mobile user, a mobile subscriber* and *a mobile terminal*, interchangeably. The netshell communications architecture is shown in Figure 16.1.

When a mobile communicates with another user (either wired user or wireless user), the mobile will connect to the nearest base station (or the base station which provides the strongest receiving signal strength), which then establishes the communication with the other user over the existing communication infrastructure (such as PSTN). When a mobile user engaging a communication and moving from one cell to another, the base station in the new cell will allocate a channel to continue to provide the service to the mobile user without interruption (if possible). The switch from one channel to another (or from one base to another) is called *handoff*. One of important aspect of the mobility management is the handoff management: how to achieve a smooth handoff without degrading the service currently in progress over the air. When users do not engage any communications and move around, the system has to track them in order to deliver possible services to them. This requires mobiles to inform the network their whereabouts when they move, thus the system can locate them based on the previously reported information. This process is called *location management* (sometimes is also called *mobility management*). In this chapter, we focus on the location management.

Location management is a unique feature for wireless networking, and needs to be addressed carefully. In second generation wireless systems, the wireless networks standards IS-41 ([15]) and GSM MAP ([16]) use two-level strategies for mobility management in that a two-tier system consisting of Home Location Register (HLR) and Visitor Location Register (VLR) databases is deployed (see Figure 16.2). Although there are some modifications on the mobility management for the third gen-

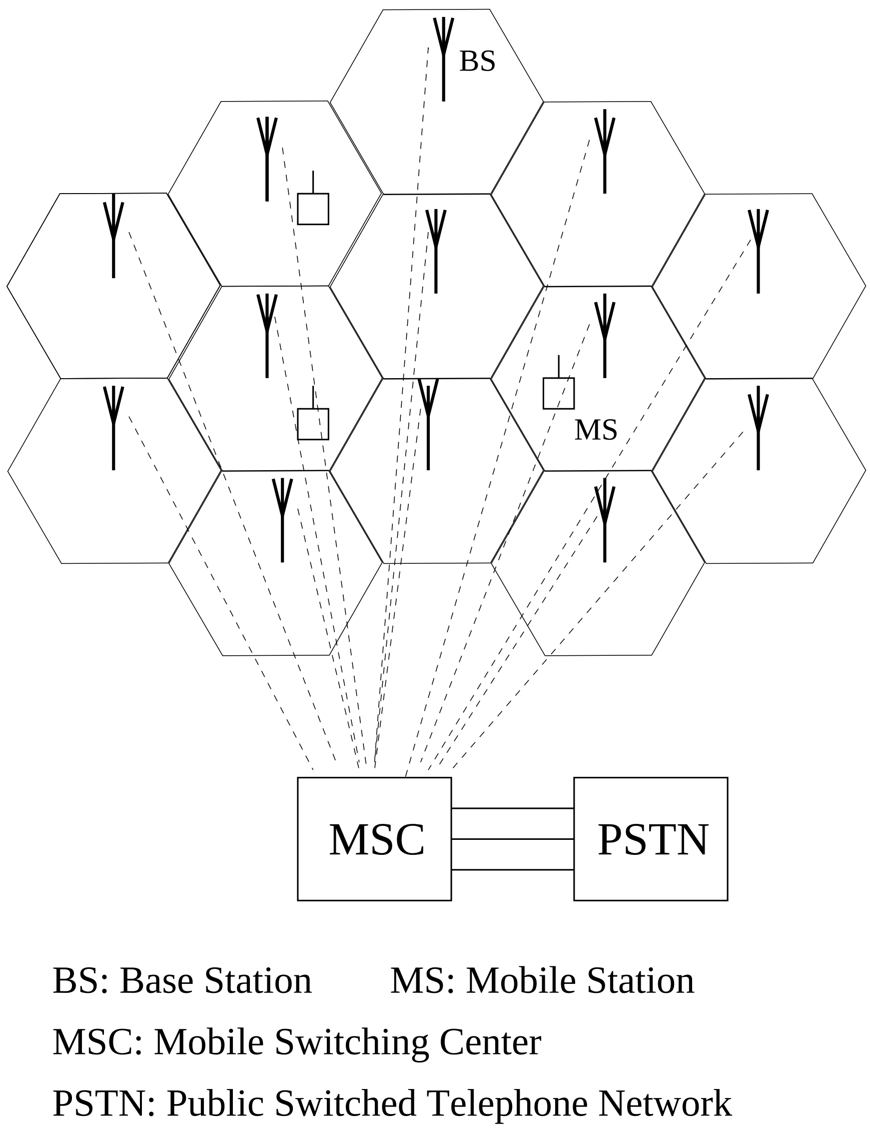


*Figure 16.1.* The communication architecture for wireless cellular systems

eration wireless systems (such as the introduction of *gateway location register* or *GLR*), the fundamental operations in mobility management remains more or less the same. Hence, we will use the IS-41 standard as the baseline study here. In two-tier mobility management architecture shown in Figure 16.2, the HLR stores the user profiles of its registered mobiles, which contain the user profile 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 the replications of user 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 (also called *location update* in many situations) is the process that a mobile informs the system of its current location, the location tracking (also called *call delivery*) is the process that 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 area (called *registration area (RA)*), a temporary record will be created in the VLR of the visited system, and the VLR will then sends 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 calling mobile contacts its MSC, which then initiates a query to the caller's HLR to find where the callee is last reported (i.e., the VLR the callee was sighted). When the HLR receives this query, it will send 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 (MSC) through the HLR. Based on the TLDN, the originating switch (MSC) will set up the trunk (e.g., voice circuit with enough bandwidth for voice service) to the mobile. Thus, the call delivery process consists of two parts: *find*, the method to locate the mobile, and the trunk setup for the mobile using TLDN. The signaling traffic due to *find* and *registration ( location update)* can be significant. Various schemes to reduce such traffic have been proposed. A very excellent survey on this topic is presented in [6]. There are mainly two approaches to reduce the signaling traffic. One is to manage the mobility databases physically or logically, the other is to control the location update frequency more intelligently. The funda-

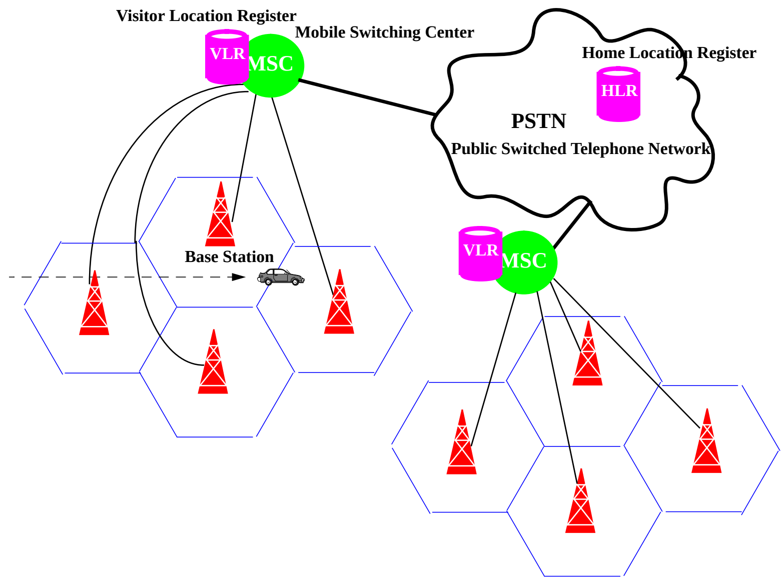


*Figure 16.2.* Two-tier location management architecture ([44])

mental idea in both approaches is to localize the signaling traffic. Here are a few representative schemes proposed in the past. In [35], the *location cache scheme* was proposed and shown significant improvement over 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 the HLR, the *pointer forwarding scheme* ([34]) was proposed based on the observation that it may be better to setup a forwarding pointer from the previous VLR to avoid more expensive registration from the mobile. By storing location profile, registration traffic can also be reduced, this idea leads to the *alternative location algorithm (ALA)* and *two location algorithm (TLA)* ([42] 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 ([30]) and the dynamic hierarchical database architecture using *directory registers* ([29]). Recently, Ma and Fang ( [47]) proposed a two-level pointer forwarding scheme to combine the advantages of pointer forwarding scheme and local anchor scheme. To maintain the freshness of the location information, active location update (also known as *autonomous registration*) was also suggested in the standard as an option ([16]). How often the active location update should be done is still an open question. Three schemes were investigated in the past ([7]): *timer-based, movement-based* and *distance-based*. It has been shown that the most effective scheme among the three is distance-based: whenever a mobile moves away from the previously reported spot in certain distance, the mobile will carry out the location update. However, the implementation is complicated as most mobiles do not have GPS receivers equipped. The time-based scheme is simple to use, which has been used in most mobiles, however, it is not efficient in terms of signaling traffic, particularly for the mobiles with low mobility. It turns out that the movement-based (MB) scheme strikes the balance between the other two in both efficiency and implementation, which will be the one we investigate in this chapter. The problem is how to choose the movement threshold. The large the the threshold, the larger the uncertainty region of the mobile; while the smaller the threshold, the smaller the uncertainty region, but the more the signaling traffic. All above schemes are for location management. Another major part of signaling traffic comes from the the last segment during the call delivery process: the paging. When the callee's VLR receives a query for the callee, a paging process is initiated. Depending on the information available and resource consideration, sequential paging, parallel paging or selective paging can be chosen ([56]).

To effectively evaluate a mobility management scheme, we have to deal with signaling traffic caused by both location update and call delivery. How to quantify the signaling traffic and the cost analysis become important. Signaling traffic cost relies on many factors such as the terminating call arrivals and the users' mobility. In the past, cost analysis of most mobility management schemes were carried out under the assumption that most time variables are exponentially distributed. For example, the time between two *served* calls, which we called *inter-service time* ( [20]), was usually assumed to be exponentially distributed. Although the call arrivals terminating at a mobile, say, $\mathcal{T}$, may be approximately modeled by Poisson process, the inter-service time is not identical to the inter-arrival time due to the *busy-line effect* ([20]), i.e., some call arrivals for the mobile $\mathcal{T}$ may not be connected because $\mathcal{T}$ is serving another call, this is particularly true for data connections that tend to connect to the wireless systems longer than voice connection on average. Hence the served calls are in fact a "sampled" Poisson process, thus will be most likely not Poisson process. The reason we are interested in the inter-service 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 to inform the system. Thus, the location management is affected by the inter-service times rather than the inter-arrival times to the mobile. Moreover, due to the new trend of applications and user habits, even the inter-arrival 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 ([43]), whether such schemes are still effective or not when the inter-service times are not exponentially distributed is a question, there is no justification in the literature.

Recently, we have developed a new analytical approach to systematically analyze the performance of a few mobility management schemes under more realistic assumptions ([17, 18, 47]). In this chapter, we will present this analytical approach for the modeling and performance analysis for mobility management. We focus on the general analytical results for the signaling cost analysis for a few mobility management schemes we have studied in the past to illustrate our approach. Obviously, our results can be used to design the dynamic location management schemes.

This chapter is organized as follows. In the next section, we present the descriptions of a few known location management schemes we will study in this chapter. We then present a general framework for the evaluation of mobility management. The crucial analytical result on the probability distribution of the number of area boundary crossings are given in the fourth section. In the fifth section, we present the general analytical results for the signaling costs for the mobility management schemes we are interested in. We will conclude this chapter in the last section.

## 2. Mobility Management Schemes

There are many location management schemes in the literature, [6] presents a very comprehensive survey on all aspects of mobility management. In this chapter, we only concentrate on the schemes for which we have analytical results under fairly general assumptions. For some interesting schemes such as per-user caching scheme ([35]), location anchoring scheme ([30]), and location profile based scheme ([52]), we do not have the analytical results as yet, we will not include them in this section. More details can be found in [6].

### 2.1 IS-41 Scheme

To set the baseline comparison study, we first briefly go over the IS-41 scheme (or GSM MAP), the de facto standards for second generation wireless systems. We use the terminology used in [34]. An operation *move* means that a mobile user moves from one Registration Area (RA) (also called Location Area (LA)) 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 de-registration message to the old VLR, which will, upon receiving the de-registration 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*, 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 setup 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 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. Periodic location update is optional for improving the efficiency of IS-41 and GSM MAP ([15, 16]).

## 2.2 Movement-based Mobility Management

There are many ways to improve the performance of the mobility management detailed in the IS-41. To minimize the signaling traffic due to the location update while keeping the location information fresh, we need to determine when we would need location update. In the current literature, three location update schemes were proposed and studied ( [3, 7]): *distance-based location update*, *movement-based location update* and *time-based location update*. In distance-based locate update scheme, location update will be performed when a mobile terminal moves $d$ cells away from the cell in which the previous location update was performed, where $d$ is a distance threshold. In the movement-based location update scheme, a mobile terminal will carry out a location update whenever the mobile terminal completes $d$ movements between cells (whenever the mobile moves from one cell to another, we count it as one move), where $d$ is the movement threshold. In the time-based location update scheme, the mobile terminal will update its location every $d$ time units, where $d$ is the time threshold. It has been shown ([7]) that the distance-based location update scheme gives the best result in terms of signaling traffic, however, it may not be practical because a mobile terminal has to know its own position information in the network topology. The time-based location update scheme is the simplest to implement, however, unnecessary signaling traffic may result (imagine a terminal stationary for a long period may not need to do any update before it moves). The movement-based location update scheme seems to be the best choice in terms of signaling traffic and implementation complexity. We will analyze the movement-based mobility management scheme here.

## 2.3 Pointer Forwarding Scheme (PFS)

The PFS modifies the *move* and *find* used in the IS-41 in the following fashion ([34]). 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 then 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 setup 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 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 be curbed potentially.

## 2.4 Two-Level Pointer Forwarding Scheme

Two-level pointer forwarding scheme (TLPDS) is a generalization of PFS, which attempts to localize the location update signaling traffic and is consistent with the current hierarchical architecture of wireless system design ([47]). Instead of using one-level pointers in PFS, we activate a new layer of management entities, called *Mobility Agent (MA)*, the VLR which is in charge of multiple RAs, reflecting the regional activities of mobile users. The Two-Level Pointer Forwarding Scheme modifies the basic procedures used in IS-41 as follows. When a mobile moves from one RA to another, it informs the switch (and the VLR) at the new RA about the old RA. It also informs the new RA about the previous MA it was registered at. The switch at the new RA determines whether to invoke the *BasicMOVE* (update to the HLR) or the *TwoLevelFwdMOVE* (update to either the previous RA or the previous MA). In *TwoLevelFwdMOVE*, the new VLR exchanges messages with the old VLR or the old MA to set up a forwarding pointer from the old VLR to the new VLR. If a pointer is set up from the previous MA, the new VLR is selected as the current MA. The *TwoLevelFwdMOVE* procedures do not involve

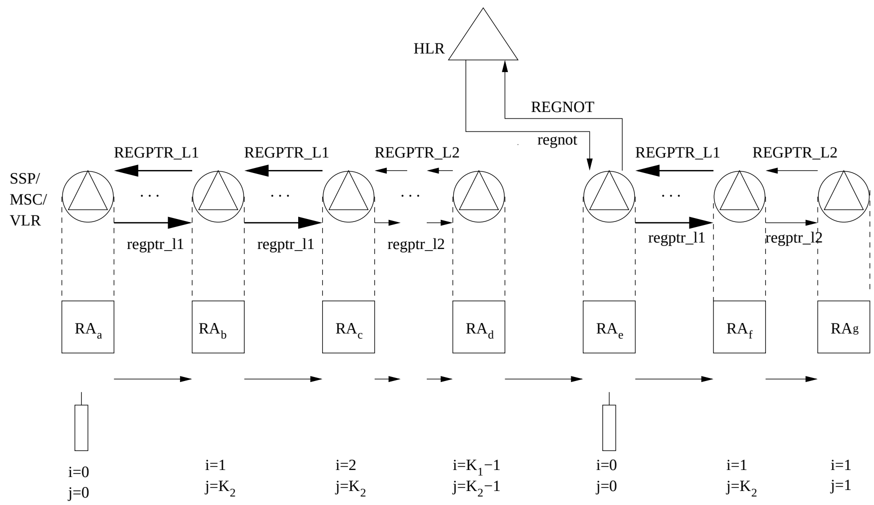


*Figure 16.3.* TLPDS location management architecture ([47])

the user's HLR. For example, Figure 16.3 shows a *Two-Level Forward MOVE* procedures with level_1 pointers chain threshold limited to 3. Assume that a user moves from RAa to RAg (these RAs are not necessary to be adjacent) and RAa is the user's MA. When the user leaves RAa but before enters RAb, the user informs the new VLRs and the level_2 pointers are built from the old VLR to the new VLR. When the user enters RAb, the chain threshold for level_2 pointer is reached, so RAb is selected as the user's new MA and a level_1 pointer is set up from the old MA to the new MA. At the same time, level_2 pointer chain is reset. The similar procedures are used at RAc. A level_1 pointer is set up from RAb to RAc, and the VLR in RAc is the new user's MA. As the user keeps moving. In RAe, the threshold of level_2 pointer chain is reached again, while this time the threshold of the level_1 pointer chain is reached too. Instead of exchanging information with the previous MA, the *BasicMOVE* is invoked. The HLR is updated with the user current location. The messages REGPTR_L1 and REGPTR_L2 are messages from the new VLR to the old VLR specifying that a level_1 or level_2 forwarding pointer is to be set up; messages regptr_l1 and regptr_l2 are the confirmations from the old VLR (or MA). In this figure, the VLRs in RAa, RAb, RAe and RAf are selected as the user's MAs.

To locate a mobile user, the *Find* procedure, called *TwolevelFwdFIND* procedure, is invoked for the subsequent calls to the mobile user from some other switches. The user's HLR is queried first as in the basic strategy, and a pointer to the user's potentially outdated MA is obtained.

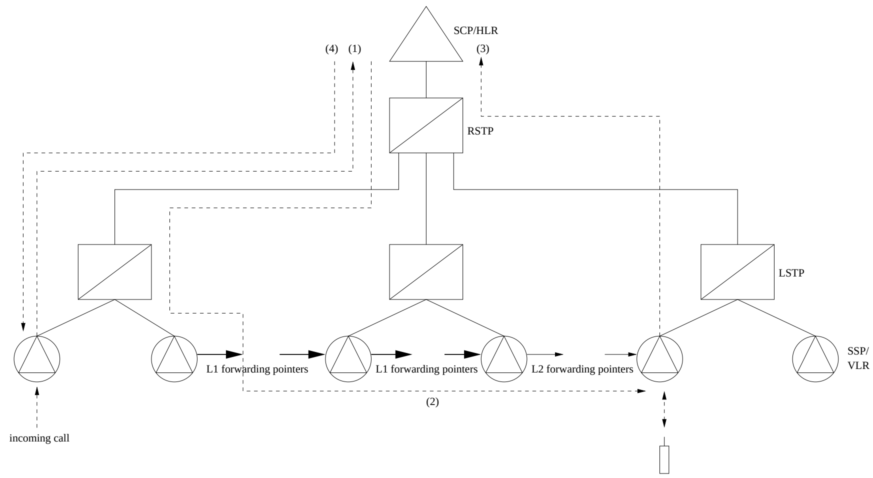


*Figure 16.4.* TLPDS Find procedure ([47])

The pointer chain is followed to find out the user's current location (see Figure 16.4). As we can see, in the two-level forwarding scheme, the chain length can be longer than that in the basic pointer forwarding scheme without increasing the *Find* penalty significantly. The previous study [34] shows that more saving can be obtained with longer chain. However, the pointer chain length is limited by the delay restriction requirement. By appropriately tuning the two thresholds in our schemes, we can mitigate the signaling cost without too much increase in the call setup delay.

## 2.5 Two Location Algorithm (TLA)

In the TLA ([42]), a mobile $\mathcal{T}$ has a small built-in memory to store the addresses for the two most recently visited RA's. 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, swaps the two locations in the memory of $\mathcal{T}$ 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 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. A General Framework for Performance Evaluation

Consider a wireless mobile network with cells of the same size. A mobile terminal visits a cell for a time interval which is generally distributed, then moves to the neighboring cell with equal probability (we are interested in this paper the homogeneous wireless networks in which all cells are statistically identical and all registration areas (RAs) are statistically identical). When a mobile is engaging a service, the network knows the location of the mobile, hence the basic idea to evaluate the performance of a mobility management scheme is to study the overall signaling traffic caused by location update and call delivery (*Find* or *Paging*) when the mobile is NOT engaging communication through the network, i.e., we only need to study the overall signaling traffic during the inter-service time. To carry tradeoff analysis, the signaling traffic has to be quantified, hence certain cost mapping may be necessary. Based on the cost structure for signaling traffic, performance comparison can be made possible. Moreover, some mobility management schemes have some tunable parameters, which need to be chosen to optimize the performance. By quantifying the signaling traffic, optimization problems can be formulated and solved analytically.

We take the movement-based location update scheme as one example to illustrate the framework for the cost analysis. Let $d$ denote the movement threshold, i.e., a mobile terminal will perform a location update whenever the mobile terminal makes $d$ movements (equal to the number of serving cell switching) after the last location update. When an incoming call to a mobile terminal arrives, the network initiates the paging process to locate the called mobile terminal. Thus, both location update and terminal paging will incur signaling traffic, the location updates consume the uplink bandwidth and mobile terminal power, while the terminal paging mainly utilizes the downlink resource, hence the cost factors for both processes are different. When uplink signaling traffic is high or power consumption is a serious consideration, it may be better to use terminal paging instead. Different users or terminals may have different quality of service (QoS) requirement, location update and paging may be designed to treat them differently. All these factors should be considered and a more general cost function, which reflects these considerations, is desirable. A reasonable cost should consider two factors: location update signaling and paging signaling, which can be captured by the following general cost function:

$$\mathcal{C}(d) = \mathcal{C}_u(N_u(d), \lambda_u, q_s) + \mathcal{C}_p(N_p(d), \lambda_p, q_s) \qquad (16.1)$$

where $N_u(d)$ and $N_p(d)$ are the average number of location updates and the average number of paging messages under the movement-based location update and a paging scheme with movement threshold $d$, respectively, during a typical period of inter-service time, $\lambda_u$ is the signaling rate for location updates at a mobile switching center (MSC) and $\lambda_p$ is the signaling rate for paging at an MSC, $q_s$ indicates the QoS factor, $\mathcal{C}_u$ and $\mathcal{C}_p$ are two functions, reflecting the costs for location updates and paging. Depending on the choice of the two functions, we can obtain different mobility management schemes, particularly, we can obtain the optimal movement threshold $d$ to minimize the total cost $\mathcal{C}(d)$, which gives us the best tradeoff scheme.

In the current literature, the total cost is chosen to be the linear combination of the location update and the paging signaling traffic, i.e., $\mathcal{C}_c(N_u(d), \lambda_u, q_s) = UN_u(d)$, $\mathcal{C}_p(N_p(d), \lambda_p, q_s) = PN_p(d)$, where $U$ is the cost factor for location update and $P$ is the cost factor for terminal paging ([4, 40, 42]). When the signaling traffic for location updates in an area is too high, some mobile terminals may be advised to lower their location updates, this can be done by choosing a function $\mathcal{C}_u$ to contain a factor $1/(\lambda_{\max} - \lambda_u)$ where $\lambda_{\max}$ is the maximum allowable signaling rate for location update. For example, if we choose $\mathcal{C}_u = U/(\lambda_{\max} - \lambda_u)$, when $\lambda_u$ is approaching $\lambda_{\max}$, the update cost will become huge, the optimization will prefer to use paging more, leading to fewer location updates. How to choose the appropriate cost functions is a challenging problem to be addressed in the future.

For other schemes, the total cost will consist of two parts as well. The first part reflects the cost resulting from the location update, while the second part is contributed by the call delivery cost (*Find* operation). In the movement-based mobility management scheme we discussed above, since the call delivery cost mainly comes from the terminal paging, hence we can use the terminal paging cost. Thus, the remaining task is to find the quantities we need in the cost function for each scheme.

## 4. Probability of the Number of Area Boundary Crossings

In order to carry out the performance analysis of mobility management schemes, we need to model some of the time variables appropriately and compute the probability distribution of the number of RA boundary crossings or cell boundary crossings. For example, in the mobility-based mobility management, we need to find the average number of location updates during the inter-service time, hence need to determine the distribution of cell boundary crossings. In other schemes (such as the PFS or TLPDS), we need to find the probability distribution of RA crossings to determine the cost for location update and call delivery. The results in this subsection have been presented in [20]. Although the results there are for RA crossings, they are valid for the cell crossings too, the only change to be made is to replace the RA residence time by the cell residence time. As we will show that this distribution plays a significant role in our analytical approach. For the simplicity of presentation, in this section, we use *an area* to denote either a registration area (RA) or a cell so that we can use the results to determine the probability of the number of RA boundary crossings and the probability of the number of cell boundary crossings as long as we use the corresponding residence time.

Assume that the incoming calls to a mobile, say, $\mathcal{T}$, form a Poisson process. The time the mobile stays in an *area* is called the *area residence time (ART)*. The time the mobile stays in a *registration area (RA)* is called the *RA residence time (RRT)*. The time the mobile stays in a cell is called the *cell residence time (CRT)*. Thus, ART can be either RRT or CRT for later applications. We assume that the ART is generally distributed with a non-lattice distribution (i.e., the probability distribution does not have discrete component). The time between the end of a call served and the start of the following call served by the mobile is called *the inter-service time*. It is possible that a call arrives while the previous call served is still in progress ([42]). In this case, the mobile $\mathcal{T}$ cannot accept the new call (the caller senses busy tone in this case and may hang up). Thus, the inter-arrival (inter-call) times for the calls terminated at the mobile $\mathcal{T}$ are different from the inter-service times. This phenomenon is called the *busy line* effect. We emphasize here that

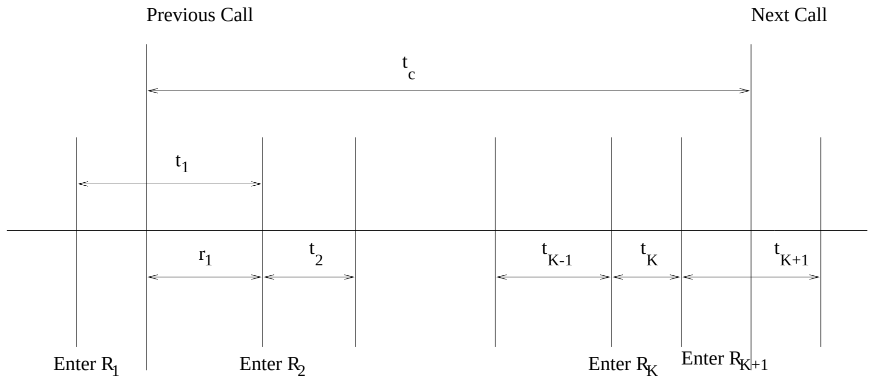


*Figure 16.5.* The time diagram for $K$ area boundary crossings ([20])

for mobility management, it is the inter-service times, not the inter-arrival 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 inter-service times may not be exponentially distributed. Inter-service time bears the similarity to the inter-departure time of a queueing system with blocking, which has been shown to be non-exponential in general. In this chapter, we assume that the service time is negligibly shorter than the inter-service time (i.e., ignoring the service time) and derive the probability $\alpha(K)$ that a mobile moves across $K$ *areas* between two served calls arriving to the mobile $\mathcal{T}$. In what follows, we will use the *area residence time (ART)* to denote either the RRT or the CRT, depending on the mobility management schemes we evaluate. By ignoring the busy line effect and the service time, assuming that the inter-service time is exponentially distributed, Lin ([42]) is able to derive an analytical formula for $\alpha(K)$, which has been subsequently used ([5, 4, 42, 55]) for tradeoff analysis for location update and paging. In this section, we assume that the inter-service times are generally distributed and derive an analytic expression for $\alpha(K)$.

Let $t_1, t_2, \ldots$ denote the area residence times and let $r_1$ denote the residual area 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 *area*). Let $t_c$ denote the inter-service time between two consecutive served calls to a mobile $\mathcal{T}$. Figure 16.5 shows the time diagram for $K$ area boundary crossings. Suppose that the mobile is in an *area* $R_1$ when the previous call arrives and accepted by $\mathcal{T}$, it then moves $K$ *areas* during the inter-service time, and $\mathcal{T}$ resides in the $j$th *area* for a period $t_j$ $(1 \leq j \leq K+1)$. In this chapter, we consider a homogeneous wireless mobile network, i.e., all *areas* (either registration areas or cells) in the network are statistically identical, hence ARTs $t_1, t_2, \ldots$ 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 ([38]), we have

$$\begin{aligned} f_r(t) &= \lambda_m \int_t^\infty f(\tau)d\tau = \lambda_m \left[1 - F(t)\right], \\ f_r^*(s) &= \frac{\lambda_m}{s}[1 - f^*(s)], \end{aligned} \tag{16.2}$$

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 area residence time $r_1$ in the initiating *area*, i.e., we treat $r_1$ as a random variable which does not have any relationship to the ART $t_i$. The results we present here can be applied to this very general case. It is obvious that the probability $\alpha(K)$ is given by

$$\alpha(0) = \Pr[t_c \leq r_1], \quad K = 0, \tag{16.3}$$

$$\alpha(K) = \Pr[r_1 + t_2 + \cdots + t_K < t_c \leq r_1 + t_2 + \cdots + t_{K+1}], \quad K \geq 1 \tag{16.4}$$

Applying the inverse Laplace transform, we can compute $\alpha(0)$ from (16.3) as follows:

$$\begin{aligned} \alpha(0) &= \int_0^\infty \Pr(r_1 \geq t) f_c(t)dt = \int_0^\infty \frac{1}{2\pi j} \int_{\sigma - j\infty}^{\sigma + j\infty} \frac{1 - f_r^*(s)}{s} e^{st} ds f_c(t) dt \\ &= \frac{1}{2\pi j} \int_{\sigma - j\infty}^{\sigma + j\infty} \frac{1 - f_r^*(s)}{s} \int_0^\infty f_c(t) e^{st} dt ds \\ &= \frac{1}{2\pi j} \int_{\sigma - j\infty}^{\sigma + j\infty} \frac{1 - f_r^*(s)}{s} f_c^*(-s) ds \end{aligned} \tag{16.5}$$

where $\sigma$ 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 $\Pr(r_1 + t_2 + \cdots + t_k \leq t_c)$ 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 $\xi$. ¿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 $\Pr(\xi \leq t)$ is $f_\xi^*(s)/s$, applying the inverse Laplace transform, we have

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

Taking this into (16.4), we obtain

$$\begin{aligned} \alpha(K) &= \Pr(t_c \geq r_1 + t_2 + \cdots + t_K) - \Pr(t_c \geq r_1 + t_2 + \cdots + t_{K+1}) \\ &= \frac{1}{2\pi j} \int_{\sigma-j\infty}^{\sigma+j\infty} \frac{f_r^*(s)[f^*(s)]^{K-1}[1 - f^*(s)]}{s} f_c^*(-s) ds \end{aligned} \tag{16.6}$$

where $\sigma$ is a sufficiently small positive number. In summary, we obtain

THEOREM 16.1 *If the probability density function of the inter-service 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 $K$ areas (either RAs or cells) during the inter-service time is given by*

$$\begin{aligned} \alpha(0) &= \frac{1}{2\pi j} \int_{\sigma-j\infty}^{\sigma+j\infty} \frac{1 - f_r^*(s)}{s} f_c^*(-s) ds \\ \alpha(K) &= \frac{1}{2\pi j} \int_{\sigma-j\infty}^{\sigma+j\infty} \frac{f_r^*(s)[1 - f^*(s)][f^*(s)]^{K-1}}{s} f_c^*(-s) ds, \quad K > 0 \end{aligned} \tag{16.7}$$

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

If the inter-service time is exponentially distributed, then we can apply the Residue Theorem ([39]) to obtain the following simple result:

$$\begin{aligned} \alpha(0) &= 1 - \frac{1 - f^*(\lambda_c)}{\rho}, \\ \alpha(K) &= \frac{1}{\rho}[1 - f^*(\lambda_c)]^2[f^*(\lambda_c)]^{K-1}, \quad K > 0, \end{aligned}$$

where $\rho = \lambda_c/\lambda_m$ is the *call-to-mobility ratio*. This result has been obtained in [42] using a different approach.

## 5. Performance Evaluation

Signaling traffic in mobility management schemes incur from the operations of *move* (may cause location update) and *find* (may need message exchanges and terminal paging). Most performance evaluation for *move* and *find* were 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 the 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 or the threshold in movement-based mobility management scheme) may not be appropriate accordingly. There are not much works handling the non-exponential situations. Recently, we have developed a new approach which handles the non-exponential situations. In this section, we present our analytical results for signaling cost.

### 5.1 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 inter-service time (the inter-arrival 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 moves* during the inter-service 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 inter-service 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 a RA boundary, a registration is triggered. We assume that the unit cost for a basic registration (i.e., *basic move*) is $m$. From Theorem 16.1, $M$ will be equal to the product of $m$ and the average number of registrations incurred during the inter-service time, given by ($f(t)$ and $f^*(s)$ are for the RA residence time)

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

where $\rho = \lambda_c / \lambda_m$, which is called the *call-to-mobility ratio.*

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. Thus, $F \geq m$. The *basic find* will be the paging cost in one RA. Thus, we obtain the total cost for IS-41 during the inter-service time is

$$\mathcal{C}_{IS-41} = M + F = \frac{m}{\rho} + F. \tag{16.8}$$

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

## 5.2 Cost Analysis for Movement-based Mobility Management (MB)

The periodic location update (autonomous registration) is only an option in the IS-41 or the GSM MAP. The operation for such location update is simple, however, the scheme may not be efficient. The movement-based mobility management (MB) is a better choice in both the efficiency and simplicity. As we mentioned, the location update cost relies on the average number of location updates during the inter-service time. In this section, we present the analytical results for this scheme.

**5.2.1 Average Number of Location Updates.** Let $d$ be the threshold for the movement-based location update scheme, i.e., a mobile will make a location update every $d$ crossings of cell boundary when the mobile does not engage in service. Obviously, the average number of location updates during the inter-service time will determine the locate update cost. So, we first want to find this quantity. We assume that $f(t)$ denote the probability density for the cell residence time (CRT) in this section, all other notation will remain the same as before. Then, the average number of location update during an inter-service time interval under movement-based location update scheme can be expressed as ([4])

$$N_u(d) = \sum_{i=1}^{\infty} i \sum_{k=id}^{(i+1)d-1} \alpha(k) \tag{16.9}$$

where $\alpha(k)$ in this subsection denotes the probability that a mobile crosses $k$ cells during the inter-service time. In what follows in this subsection, we present the computation for $N_u(d)$. Let

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

then, from Theorem 16.1, we obtain

$$\begin{aligned} S(n) &= \sum_{k=1}^{n-1} \frac{1}{2\pi j} \int_{\sigma-j\infty}^{\sigma+j\infty} \frac{f_r^*(s)[1-f^*(s)][f^*(s)]^{k-1}}{s} f_c^*(-s)ds \\ &= \frac{1}{2\pi j} \int_{\sigma-j\infty}^{\sigma+j\infty} \frac{f_r^*(s)[1-f^*(s)]}{s} \left( \sum_{k=1}^{n-1} [f^*(s)]^{k-1} \right) f_c^*(-s)ds \\ &= \frac{1}{2\pi j} \int_{\sigma-j\infty}^{\sigma+j\infty} \frac{f_r^*(s)[1-(f^*(s))^{n-1}]}{s} f_c^*(-s)ds \end{aligned} \tag{16.10}$$

Moreover, we have

$$\begin{aligned} \sum_{k=1}^{N} S(kd) &= \sum_{k=1}^{N} \frac{1}{2\pi j} \int_{\sigma-j\infty}^{\sigma+j\infty} \frac{f_r^*(s)[1-(f^*(s))^{kd-1}]}{s} f_c^*(-s)ds \\ &= \frac{1}{2\pi j} \int_{\sigma-j\infty}^{\sigma+j\infty} \frac{f_r^*(s)}{s} \left\{ \sum_{k=1}^{N} \left[1-(f^*(s))^{kd-1}\right] \right\} f_c^*(-s)ds \\ &= \frac{1}{2\pi j} \int_{\sigma-j\infty}^{\sigma+j\infty} \frac{f_r^*(s)}{s} \left\{ N - \frac{(f^*(s))^{d-1}[1-(f^*(s))^{Nd}]}{1-(f^*(s))^d} \right\} f_c^*(-s)ds \end{aligned} \tag{16.11}$$

Thus, from (16.9), (16.10) and (16.11), after some mathematical manipulation (details can be found in [17]), we obtain

$$\begin{aligned} N_u(d) &= \sum_{i=1}^{\infty} i\,[S((i+1)d) - S(id)] \\ &= \lim_{N\to\infty} \left\{ \sum_{i=1}^{N} i\,[S((i+1)d) - S(id)] \right\} \\ &= \lim_{N\to\infty} \left\{ NS((N+1)d) - \sum_{i=1}^{N} S(id) \right\} \\ &= \frac{1}{2\pi j} \int_{\sigma-j\infty}^{\sigma+j\infty} \frac{f_r^*(s)(f^*(s))^{d-1}}{s[1-(f^*(s))^d]} f_c^*(-s)ds \end{aligned}$$

where $\sigma > 0$ is a sufficiently small positive number.

In summary, we obtain

THEOREM 16.2 *If the inter-service time $f_c^*(s)$ has a finite number of poles (such as a proper rational function) and the cell residence time (CRT) is generally distributed with probability density function $f(t)$ and L-S transform $f^*(s)$, then the average number of location updates $N_u(d)$ is given by*

$$N_u(d) = \frac{1}{2\pi j} \int_{\sigma-j\infty}^{\sigma+j\infty} \frac{f_r^*(s)(f^*(s))^{d-1}}{s[1-(f^*(s))^d]} f_c^*(-s)ds \tag{16.12}$$

*where $f_r^*(s)$ is the L-S transform of the probability density of the residual cell residence time. If $f_c^*(s)$ has a finite number of poles, which is the case when it is a rational function, then we have*

$$N_u(d) = -\sum_{p\in\sigma_c} \operatorname*{Res}_{s=p} \frac{f_r^*(s)(f^*(s))^{d-1}}{s[1-(f^*(s))^d]} f_c^*(-s), \tag{16.13}$$

*where $\sigma_c$ is the set of poles of $f_c^*(-s)$.*

**5.2.2 Tradeoff Cost Analysis: A Case Study.** In this subsection, we present some results for the tradeoff analysis under movement-based location update scheme and some paging scheme with the linear cost functional for illustration purpose.

*Location Update Cost*:

Let $m$ denotes the unit cost for location update (i.e., the cost for a *basic move* as in IS-41), then total location update cost is given by

$$\mathcal{C}_u(d) = -m \sum_{p \in \sigma_c} \operatorname*{Res}_{s=p} \frac{f_r^*(s)(f^*(s))^{d-1}}{s[1-(f^*(s))^d]} f_c^*(-s), \tag{16.14}$$

If the inter-service time $t_c$ is exponentially distributed with parameter $\lambda_c$, then $f_c^*(s) = \lambda_c/(s+\lambda_c)$, from (16.14), we can easily obtain

$$\mathcal{C}_u(d) = \frac{m f_r^*(\lambda_c)[f^*(\lambda_c)]^{d-1}}{1-[f^*(\lambda_c)]^d}. \tag{16.15}$$

This result was obtained in [40] (noticing that $f_r^*(s) = \lambda_m(1-f^*(s))/s$) via a different approach.

*Paging Cost:*

We consider the paging strategy used in [40] for our case study. Consider the hexagonal layout for the wireless network, all cells are statistically identical. According to the movement-based location update scheme, a mobile terminal moves at most $d$ cells away from the previous position where it performs the last location update. Thus, a mobile terminal will surely be located in a cell which is less than $d$ cell away from the previously reported position. If we page in the circular area with $d$ cells as radius and with the previously reported position as the center, then we can definitely find the mobile terminal. Thus, this paging scheme is the most conservative among all paging. If we let $P$ denote the unit cost for each paging in a cell, the maximum paging cost for this paging scheme is given by ([40])

$$\mathcal{C}_p(d) = P(1+3d(d-1)), \tag{16.16}$$

*Total Cost:*

The unit costs (cost factors) $m$ and $P$ can be chosen to reflect the significance of the signaling (which may be significantly different from each other because they use different network resources). Given $m$, $P$ and the movement threshold $d$, the total cost for location update and paging will be given by

$$\begin{aligned}
\mathcal{C}_{MB}(d) &= \mathcal{C}_u(d) + \mathcal{C}_p(d) \\
&= -m \sum_{p \in \sigma_c} \operatorname*{Res}_{s=p} \frac{f_r^*(s)(f^*(s))^{d-1}}{s[1-(f^*(s))^d]} f_c^*(-s) + P(1+3d(d-1)).
\end{aligned}$$

(16.17)

We observe that as long as we find the probability distributions of cell residence time and inter-service time, we can find the total cost using 16.17. Many numerical results ([17]) have shown that the cost function $\mathcal{C}_{MB}(d)$ is a convex function of $d$. If this is true, then we can find the unique optimal threshold $d^*$ to minimize the cost. Unfortunately, we have not proved the convexity of $\mathcal{C}(d)$ yet.

## 5.3 Cost Analysis for Pointer Forwarding

In this subsection, we evaluate the performance of pointer forwarding scheme (PFS). We need to quantify the signaling traffic incurring in this scheme. Let $M'$ and $F'$ denote the corresponding costs for the pointer forwarding scheme during the inter-service time, in which every $K$ moves will trigger a new registration, where $K$ is the maximum pointer chain length. In this subsection, a move means the crossing from one RA to another. 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 during the inter-service time, 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, because the mobile $\mathcal{T}$ registers every $K$th move). Here, we use $\lfloor x \rfloor$ to denote the floor function, i.e., the largest integer not exceeding $x$. Let $f(t)$ and $f^*(s)$ denote the probability density function and its corresponding L-S transform of the RA residence time (RRT). Let $\alpha(k)$ denote the probability that the mobile crosses $k$ RAs during the inter-service time. Then, we obtain

$$\begin{aligned} M' &= \sum_{i=0}^{\infty} \left[ \left( i - \left\lfloor \frac{i}{K} \right\rfloor \right) S + \left\lfloor \frac{i}{K} \right\rfloor m \right] \alpha(i) \\ &= S \sum_{i=0}^{\infty} i\alpha(i) + (m - S) \sum_{i=0}^{\infty} \left\lfloor \frac{i}{K} \right\rfloor \alpha(i) \\ &= S \sum_{i=0}^{\infty} i\alpha(i) + (m - S) \sum_{r=0}^{\infty} r \left( \sum_{i=rK}^{(r+1)K-1} \alpha(i) \right) \qquad (16.18) \end{aligned}$$

Let

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

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

then, from Theorem 16.1 and following the same procedure we used in the last subsection, we obtain

$$S(n) = \frac{1}{2\pi j} \int_{\sigma - j\infty}^{\sigma + j\infty} \frac{f_r^*(s)[1 - (f^*(s))^{n-1}]}{s} f_c^*(-s) ds, \tag{16.19}$$

$$\sum_{i=1}^{N} s(iK) = \frac{1}{2\pi j} \int_{\sigma - j\infty}^{\sigma + j\infty} \frac{f_r^*(s)}{s} \left\{ N - \frac{(f^*(s))^{K-1}[1 - (f^*(s))^{NK}]}{1 - (f^*(s))^K} \right\} f_c^*(-s) ds \tag{16.20}$$

and

$$\begin{aligned} X(K) &= \sum_{r=1}^{\infty} r \left[ S((r+1)K) - S(rK) \right] \\ &= \frac{1}{2\pi j} \int_{\sigma - j\infty}^{\sigma + j\infty} \frac{f_r^*(s)(f^*(s))^{K-1}}{s[1 - (f^*(s))^K]} f_c^*(-s) ds \end{aligned} \tag{16.21}$$

Noticing that $X(1) = \sum_{i=0}^{\infty} i\alpha(i)$, from (16.18), we obtain

$$\begin{aligned} M' &= SX(1) + (m - S)X(K) \\ &= \frac{1}{2\pi j} \int_{\sigma - j\infty}^{\sigma + j\infty} \left[ \frac{S f_r^*(s)}{s[1 - f^*(s)]} + \frac{(m - S) f_r^*(s)(f^*(s))^{K-1}}{s[1 - (f^*(s))^K]} \right] f_c^*(-s) ds \end{aligned} \tag{16.22}$$

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 tracked in order to find the mobile $\mathcal{T}$ in the pointer forwarding *find* operation. If the mobile visits a 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 16.1, we obtain

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

$$= \frac{T}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty} \frac{f_r^*(s)[1-f^*(s)]}{s[1-(f^*(s))^K]}\left(\sum_{k=0}^{K-1}\Theta(k)(f^*(s))^{k-1}\right) f_c^*(-s)ds + F$$

In summary and applying the Residue Theorem ([39]), we finally arrive at

THEOREM 16.3 *If the inter-service time $f_c^*(s)$ has a finite number of poles (such as a proper rational function) and the RA residence time is generally distributed with probability density function $f(t)$ and L-S transform $f^*(s)$, then we have*

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

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

If the inter-service time is exponentially distributed with parameter $\lambda_c$, then we have $f_c^*(-s) = \lambda_c/(-s+\lambda_c)$, from Theorem 16.3, we obtain

$$\begin{aligned} M' &= \frac{Sf_r^*(\lambda_c)}{1-f^*(\lambda_c)}+\frac{(m-S)f_r^*(\lambda_c)(f^*(\lambda_c))^{K-1}}{1-(f^*(\lambda_c))^K} \\ F' &= F+\frac{Tf_r^*(\lambda_c)[1-f^*(\lambda_c)]}{1-(f^*(\lambda_c))^K}\left(\sum_{k=0}^{K-1}\Theta(k)(f^*(\lambda_c))^{k-1}\right) \end{aligned}$$

which were obtained in [34] 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:

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

Thus, the total cost for PFS during the inter-service time can be computed as follows:

$$\mathcal{C}_{PFS} = M' + F' = F + \frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty} g(s) f_c^*(-s) ds = F - \sum_{p\in\sigma_c} \operatorname*{Res}_{s=p} g(s) f_c^*(-s) \tag{16.23}$$

where

$$\begin{aligned} g(s) &= \frac{S f_r^*(s)}{s[1-f^*(s)]} + \frac{(m-S) f_r^*(s)(f^*(s))^{K-1}}{s[1-(f^*(s))^K]} \\ &\quad + \frac{T f_r^*(s)[1-f^*(s)]}{s[1-(f^*(s))^K]}\left(\sum_{k=0}^{K-1}\Theta(k)\right). \end{aligned}$$

## 5.4 Cost Analysis for Two-Level Pointer Forwarding

Two-Level Pointer Forwarding (TLPFS) is a generalization of PFS in the sense that a new level of mobility agents (MA) is added between the home systems and visiting systems to regionalize (or localize) the signaling traffic. In [47], we derive the signaling cost under exponentially distributed inter-service time. Although we could obtain more general analytical results for this scheme under more general assumption in a similar way used for PFS, for simplicity, we only present the results in [47] and leave the derivation of more general results to the readers.

We use the same notation as before, we let $t_c$ denote the inter-service time with average $1/\lambda_c$ and let $f(t)$ denote the probability density function of the RA residence time with average $1/\lambda_m$. The call-to-mobility ratio (CMR) is then given by $\rho = \lambda_c/\lambda_m$. We assume that the inter-service time is exponentially distributed. We define $\mathcal{C}_{IS-41}$ and $\mathcal{C}_{TLPFS}$ as the total costs of updating the location information (location update) and locating the user (location tracking) during the inter-service time in IS-41 and in the TLPFS, respectively. The following parameters used in the TLPFS:

$m =$ the cost of a single invocation of *BasicMOVE*.

$M =$ the total cost of all the *BasicMOVESs* during the inter-service time.

$F =$ the cost of a single *BasicFIND*.

$M' =$ the expected cost of all *TwoLevelFwdMOVEs* during the inter-service time.

$F' =$ the average cost of the *TwoLevelFwdFIND*.

$S_1 =$ the cost of setting up a forwarding pointer (level_1 pointer) between MAs during a *Two-LevelFwdMOVE*.

$S_2 =$ the cost of setting up a forwarding pointer (level_2 pointer) between VLRs during a *Two-LevelFwdMOVE*.

$T_1 =$ the cost of traversing a forwarding pointer (level_1 pointer) between MAs during a *Two-LevelFwdFIND*.

$T_2 =$ the cost of traversing a forwarding pointer (level_2 pointer) between VLRs during a *Two-LevelFwdFIND*.

$K_1 =$ the threshold of level_1 pointer chain.

$K_2 =$ the threshold of level_2 pointer chain.

$\alpha(k) =$ the probability that there are $k$ RA crossings during the inter-service time.

Then, we have

$$\begin{aligned} \mathcal{C}_{IS-41} &= M + F = m/\rho + F, \\ \mathcal{C}_{TLPFS} &= M' + F'. \end{aligned}$$

Thus, we need to derive formulas for $M'$ and $F'$.

In the TLPFS, we observe that the HLR is updated only every $K_1 \cdot K_2$ moves ($K_1$ and $K_2$ are the level_1 and level_2 pointer chain length thresholds, respectively), while forwarding pointers are set up for all other moves. If a mobile user crosses $i$ RA boundaries during the inter-service time, then the HLR is updated $\lfloor \frac{i}{K_1 K_2} \rfloor$ times, there are also $\lfloor \frac{i}{K_2} \rfloor - \lfloor \frac{i}{K_1 K_2} \rfloor$ level_1 pointer creations (every $K_2$ moves may require a level_1 pointer creation but sometimes the HLR is updated and level_1 pointer is not set up), and there are the level_2 pointers created for all the rest $i - \lfloor \frac{i}{K_2} \rfloor$ moves. Thus, we obtain

$$\begin{aligned} M' &= \sum_{i=0}^{\infty} \left\{ \left\lfloor \frac{i}{K_1 K_2} \right\rfloor m + \left( \left\lfloor \frac{i}{K_2} \right\rfloor - \left\lfloor \frac{i}{K_1 K_2} \right\rfloor \right) S_1 \right. \\ &\quad \left. + \left( i - \left\lfloor \frac{i}{K_2} \right\rfloor \right) S_2 \right\} \alpha(i) \\ &= \sum_{i=0}^{\infty} i S_2 \alpha(i) + \sum_{i=0}^{\infty} \left\lfloor \frac{i}{K_2} \right\rfloor (S_1 - S_2) \alpha(i) \\ &\quad + \sum_{i=0}^{\infty} \left\lfloor \frac{i}{K_1 K_2} \right\rfloor (m - S_1) \alpha(i). \qquad (16.24) \end{aligned}$$

The cost $F'$ is derived as follows. After the last *BasicMove* operations (if any), the mobile user traverses $\left\lfloor \frac{i-\left\lfloor \frac{i}{K_1K_2}\right\rfloor K_1K_2}{K_2}\right\rfloor$ level_1 pointers and $i-\left\lfloor \frac{i}{K_1K_2}\right\rfloor K_1K_2-\left\lfloor \frac{i-\left\lfloor \frac{i}{K_1K_2}\right\rfloor K_1K_2}{K_2}\right\rfloor K_2$ level_2 pointers. Thus, we obtain

$$
\begin{aligned}
F' &= F+\sum_{i=0}^{\infty}\left\{\left\lfloor \frac{i-\left\lfloor \frac{i}{K_1K_2}\right\rfloor K_1K_2}{K_2}\right\rfloor T_1\right.\\
&\quad \left.+\left(i-\left\lfloor \frac{i}{K_1K_2}\right\rfloor K_1K_2-\left\lfloor \frac{i-\left\lfloor \frac{i}{K_1K_2}\right\rfloor K_1K_2}{K_2}\right\rfloor K_2\right)T_2\right\}\alpha(i)\\
&= F+(T_1-K_2T_2)\sum_{i=0}^{\infty}\left\lfloor \frac{i-\left\lfloor \frac{i}{K_1K_2}\right\rfloor K_1K_2}{K_2}\right\rfloor \alpha(i)\\
&\quad +T_2\sum_{i=0}^{\infty}\left(i-\left\lfloor \frac{i}{K_1K_2}\right\rfloor K_1K_2\right)\alpha(i).
\end{aligned}
\tag{16.25}
$$

For simplicity, we let $g=f_m^*(\lambda_c)$. We observe that all summations in both equations (16.24) and (16.25) bear close similarities to those in subsection 16.5.3, hence applying a similar procedure, we can find the analytical results for $M'$ and $F'$. The details can be found in [47]. In summary, we obtain

THEOREM 16.4 *If the inter-service time is exponentially distributed and the RA residence time is generally distributed with probability density function $f(t)$ and L-S transform $f^*(s)$, then the total signaling cost for TLPFS is given by*

$$C_{TLPFS}=M'+F'$$

*where $g=f_m^*(\lambda_c)$ and*

$$
\begin{aligned}
M' &= \frac{S_2}{\rho}+\frac{(1-g)g^{K_2-1}(S_1-S_2)}{\rho(1-g^{K_2})}+\frac{(1-g)g^{K_1K_2-1}(m-S_1)}{\rho(1-g^{K_1K_2})}\\
F' &= F+\frac{[1-K_1K_2g^{K_1K_2-1}+(K_1K_2-1)g^{K_1K_2}]T_2}{\rho(1-g^{K_1K_2})}\\
&\quad +\frac{(T_1-K_2T_2)(1-g)[g^{K_2}-K_1g^{K_1K_2}+(K_1-1)g^{(K_1+1)K_2}]}{\rho g(1-g^{K_1K_2})(1-g^{K_2})}.
\end{aligned}
$$

## 5.5 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 are 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 [42], this probability is derived under the assumption that the inter-service time is exponential. In this subsection, we first derive a more general analytical result to compute this quantity under the assumption that the inter-service time is generally distributed, then we present the cost analysis for the TLA.

¿From the argument in [42], $(1-\omega)$ 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 location update is followed by an even number of moves during the inter-service time, or when there are an even number of moves with no location update during the inter-service time. Here, in this subsection, a move means a boundary crossing from one RA to another. Let $w_1$ denote the probability that there is no move during the inter-service time, let $w_2$ denote the probability that the last served call is followed by an even number of moves without location update during the inter-service time, and let $w_3$ denote the probability that there are an even number of moves without location update during the inter-service time. Then, we have $1-w = w_1 + w_2 + w_3$, i.e., $w = 1 - w_1 - w_2 - w_3$. Let $f(t)$ denote the probability density function for the RA residence time (RRT) and other related notations remain the same as before. Let $\alpha(k)$ denote again the probability that a mobile user crosses $k$ RA boundaries during the inter-service time. For $w_1$, we have

$$w_1 = \Pr(r_1 \leq t_c) = \alpha(0) \tag{16.26}$$

Let $\theta$ denote the probability of a move without location update. Let $w_2(K)$ denote the probability that the last registration followed by an even number of moves without location update during the inter-service time given that there are $K$ moves (RA crossings) during the inter-service time. We can easily obtain that ([42])

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

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

$$w_2 \;=\; \sum_{K=1}^{\infty} w_2(K)\alpha(K) = \sum_{i=0}^{\infty} w_2(2i+1)[\alpha(2i+1) + \alpha(2i+2)]$$

$$= \frac{1}{1+\theta}\sum_{i=1}^{\infty}(1-\theta^{2i})[\alpha(2i-1)+\alpha(2i)] \qquad (16.27)$$

The probability $w_3$ can be computed as follows:

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

Thus, applying Theorem 16.1 and after some mathematical manipulations similar to the techniques we used for PFS and TLPFS (the details can be found in [18]), we obtain

$$\begin{aligned} w &= 1 - w_1 - w_2 - w_3 \\ &= 1-\alpha(0) - \frac{1}{1+\theta}(1-\alpha(0)) + \frac{1}{1+\theta}\sum_{i=1}^{\infty}\theta^{2i} - \frac{\theta}{1+\theta}\sum_{i=1}^{\infty}\theta^{2i}\alpha(2i) \\ &= \frac{\theta}{1+\theta} - \frac{\theta}{(1+\theta)2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty}\frac{1-f_r^*(s)}{s}f_c^*(-s)ds \\ &\quad + \frac{\theta^2}{(1+\theta)2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty}\frac{f_r^*(s)[1-f^*(s)]}{s[1-\theta^2 f^{*2}(s)]}f_c^*(-s)ds \\ &\quad - \frac{\theta^3}{(1+\theta)2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty}\frac{f_r^*(s)[1-f^*(s)]f^*(s)}{s[1-\theta^2 f^{*2}(s)]}f_c^*(-s)ds \\ &= \frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty}\frac{\theta f_r^*(s)}{s[1+\theta f^*(s)]}f_c^*(-s)ds \qquad (16.28) \end{aligned}$$

Applying the Residue Theorem ([39]), we obtain

THEOREM 16.5 *If the inter-service time is distributed with rational Laplace transform $f_c^*(s)$ and the RA residence time (RRT) is generally distributed with probability density function $f(t)$ and with the L-S transform $f^*(s)$, the probability of a location miss in TLA is given by*

$$w = \frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty}\frac{\theta f_r^*(s)}{s[1+\theta f^*(s)]}f_c^*(-s)ds \qquad (16.29)$$

*where $\sigma$ is a sufficiently small positive number and $f_r^*(s)$ is the L-S transform of the residual RRT. If $\sigma_c$ denotes the set of poles of $f_c^*(-s)$, then we have*

$$w = -\sum_{p\in\sigma_c}\operatorname{Res}_{s=p}\frac{\theta f_r^*(s)}{s[1+\theta f^*(s)]}f_c^*(-s). \qquad (16.30)$$

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

If the inter-service time is exponentially distributed, which is the case studied in [42], then we have $f_c^*(s) = \lambda_c/(s+\lambda_c)$, hence from Theorem 16.5, we obtain

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

which is exactly the same as in [42] after some simplification of the result in [42].

Now, we are ready to carry out the cost analysis for TLA. We still assume that the cost for *basic move* is $m$ and the cost for the *basic find* is $F$. The *basic find* consists of two parts: the first part is the message exchange from a mobile to the HLR, which is more or less the cost for the *basic move*; the second part is the message forwarding from the HLR to the callee plus the possible terminal paging. Hence, usually we have $F \geq m$ and the the cost for the second part in the *basic find* is $F - m$. Suppose that the mobile moves across $K$ RAs during the inter-service time. The conditional probability $\Pr[I = i|K]$ that $i$ *location update* 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 $\theta$ 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 location update during the inter-service time for TLA is given by

$$\begin{aligned}
n_{TLA} &= \sum_{K=0}^{\infty}\sum_{i=0}^{K} i \Pr[I=i|K]\alpha(K) = \sum_{K=0}^{\infty}\left(\sum_{i=0}^{K} i\binom{K}{i}\theta^{K-i}\theta^i\right)\alpha(K) \\
&= \sum_{K=0}^{\infty} K(1-\theta)\alpha(K) = (1-\theta)X(1) \\
&= \frac{1-\theta}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty} \frac{f_r^*(s)}{s[1-f^*(s)]} f_c^*(-s)ds \\
&= -(1-\theta)\sum_{p\in\sigma_c}\operatorname*{Res}_{s=p} \frac{f_r^*(s)}{s[1-f^*(s)]} f_c^*(-s) \qquad (16.31)
\end{aligned}$$

where we have used (16.21). Thus, the total cost for registration during inter-service time is $c_1 = mn_{TLA}$.

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 incur, thus the total cost for the *find operation* in TLA will be

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

In summary, the total signaling cost during the inter-service time for TLA is given by

$$\begin{aligned}
\mathcal{C}_{TLA} &= c_1 + c_2 = \delta n_{TLA} + F + (F-m)\omega \\
&= F + \frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty}\left[\frac{(1-\theta)m}{s[1-f^*(s)]} + \frac{(F-m)\theta}{s[1+\theta f^*(s)]}\right] f_r^*(s) f_c^*(-s)ds \\
&= F - \sum_{p\in\sigma_c}\operatorname*{Res}_{s=p}\left[\frac{(1-\theta)m}{s[1-f^*(s)]} + \frac{(F-m)\theta}{s[1+\theta f^*(s)]}\right] f_r^*(s) f_c^*(-s)
\end{aligned} \tag{16.32}$$

where $\sigma_c$ is the set of poles of $f_c^*(-s)$.

## 6. Some Interesting Special Cases

We notice that all general results presented so far can be easily applied when the inter-service time is distributed with 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 rational Laplace transform to approximate any given distribution to any desired accuracy. Some well-known distribution models used in queueing theory, such as Coxian model and phase-type distribution model, can be used to model both inter-service time and the area residence time (RRT or CRT), from which we can easily obtain the analytical results for the performance evaluation of mobility management schemes we investigated in this chapter. 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 [23] and references therein). They all belong to the phase-type distribution model, however, they are simple enough for analytical analysis and general enough to capture the statistics of the random time variables of interest. Due to the simplicity and generality of the hyper-Erlang model, we will concentrate on this model in this subsection.

The *hyper-Erlang* distribution (a more appropriate term may be *mixed-Erlang distribution*) has the following probability density function and Laplace transform:

$$f_{he}(t) = \sum_{i=1}^{M} p_i \frac{(m_i\eta_i)^{m_i} t^{m_i-1}}{(m_i-1)!} e^{-m_i\eta_i t} \quad (t \geq 0),$$

$$f_{he}^*(s) = \sum_{i=1}^{M} p_i \left( \frac{m_i \eta_i}{s + m_i \eta_i} \right)^{m_i}, \tag{16.33}$$

where

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

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

It has been demonstrated ([23, 37]) 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 $\xi$ is hyper-Erlang distributed as in equation (16.33), then its $k$th 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 $\eta_i$ ($i = 1, 2, \ldots, M$) can be found by fitting a number of moments from field data in practice.

Assuming now that the inter-service time is hyper-Erlang distributed with distribution given in equation (16.33) with

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

then, applying the Residue Theorem ([39]), we can easily obtain

$$\begin{aligned}
\mathcal{C}_{MB}(d) &= m \sum_{i=1}^{M} p_i \frac{(-1)^{m_i - 1} (m_i \eta_i)^{m_i}}{(m_i - 1)!} d^{(m_i - 1)}(m_i \eta_i) \\
&\quad + P(1 + 3d(d-1)) \\
\mathcal{C}_{PFS} &= F + \sum_{i=1}^{M} p_i \frac{(-1)^{m_i - 1} (m_i \eta_i)^{m_i}}{(m_i - 1)!} g^{(m_i - 1)}(m_i \eta_i) \\
\mathcal{C}_{TLA} &= F + \sum_{i=1}^{M} p_i \frac{(-1)^{m_i - 1} (m_i \eta_i)^{m_i}}{(m_i - 1)!} h^{(m_i - 1)}(m_i \eta_i)
\end{aligned}$$

where

$$d(s) = \frac{f_r^*(s) (f^*(s))^{d-1}}{s[1 - (f^*(s))^d]},$$

$$\begin{aligned} g(s) &= \frac{Sf_r^*(s)}{s[1-f^*(s)]} + \frac{(m-S)f_r^*(s)(f^*(s))^{K-1}}{s[1-(f^*(s))^K]} \\ &\quad + \frac{Tf_r^*(s)[1-f^*(s)]}{s[1-(f^*(s))^K]}\left(\sum_{k=0}^{K-1}\Theta(k)\right), \\ h(s) &= \left[\frac{(1-\theta)m}{s[1-f^*(s)]} + \frac{(F-m)\theta}{s[1+\theta f^*(s)]}\right] f_r^*(s) \end{aligned}$$

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

If the inter-service 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

$$\begin{aligned} \mathcal{C}_{MB}(d) &= m\sum_{i=1}^{M} p_i\eta_i d(\eta_i) + P(1+3d(d-1)) \\ \mathcal{C}_{PFS} &= F + \sum_{i=1}^{M} p_i\eta_i g(\eta_i) \\ \mathcal{C}_{TLA} &= F + \sum_{i=1}^{M} p_i\eta_i h(\eta_i) \end{aligned}$$

If the inter-service time is exponentially distributed with Laplace transform

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

then we have

$$\begin{aligned} \mathcal{C}_{MB}(d) &= m\lambda_c d(\lambda_c) + P(1+3d(d-1)) \\ \mathcal{C}_{PFS} &= F + \lambda_c g(\lambda_c) \\ \mathcal{C}_{TLA} &= F + \lambda_c h(\lambda_c) \end{aligned}$$

## 7. Conclusions

In this chapter, we have presented a new analytical approach to carrying out the performance evaluation for mobility management in wireless mobile networks. We focus on a few mobility management schemes we have investigated in the past few years and present the analytical results for the signaling traffic under very general assumption. These analytical results can be used to choose the design parameters in the mobility management. Moreover, from these results, possible adaptive mobility management schemes can be developed based on the analytical results we present in this chapter. It is our hope that this chapter will inspire more applications of this novel approach to solve other problems in wireless networking.

# References

[1] A. Abutaleb and V. O. K. Li, "Paging strategy optimization in personal communication systems," *ACM Wireless Networks*, vol.3, pp.195-204, 1997.

[2] A. Abutaleb and V. O. K. Li, "Location update optimization in personal communication systems," *ACM Wireless Networks*, vol.3, pp.205-216, 1997.

[3] I.F. Akyildiz and J.S.M. Ho, "Dynamic mobile user location update for wireless PCS networks," *ACM Wireless Networks*, vol.1, pp.187-196, 1995.

[4] 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.

[5] I.F. Akyildiz and W. Wang, "A dynamic location management scheme for next generation multi-tier PCS systems," To appear in *IEEE Journal on Selected Areas in Communications.*

[6] I.F. Akyildiz, J. McNair, J.S.M. Ho, H. Uzunalioglu and W. Wang, "Mobility management in next-generation wireless systems," *Proc. of the IEEE*, vol.87, no.8, pp.1347-1384, August 1999.

[7] A. Bar-Noy, I. Kessler and M. Sidi, "Mobile users: to update or not update?" *ACM Wireless Networks*, vol.1, no.2, pp.175-186, July 1994.

[8] R. Caceres and V. N. Padmanabhan, "Fast and scalable handoffs for wireless internetworks", *Proc. ACM Mobicom'96*, pp. 56-66, 1996.

[9] A. T. Campbell, J. Gomez, S. Kim, C.-Y. Wan, Z. R. Turanyi and A. G. Valko, "Comparison of IP micromobility protocols," *IEEE Wireless Communications*, February 2002.

[10] A. T. Campbell, J. Gomez, A. G. Valko, "An overview of cellular IP," *IEEE Wireless Communications and Networking Conference (WCNC'99)*, New Orleans, September 1999.

[11] F.M. Chiussi, D. A. Khotimsky and Santosh Krishnan, "Mobility management in third-generation all-IP networks", *IEEE Communications Magazine*, pp. 124-135, September 2002.

[12] D. C. Cox, "Wireless personal communications: what is it?" *IEEE Personal Comm. Mag.*, pp.20-35, April 1995.

[13] D. R. Cox, *Renewal Theory.* John Wiley & Sons, New York, 1962.

[14] S. Das, A. Misra, P. Agrawal and S. K. Das, "TeleMIP: Telecommunications-enhanced mobile IP architecture for fast intradomain mobility," *IEEE Personal Communications*, pp. 50-58, August 2000.

[15] EIA/TIA, "Cellular radio-telecommunications intersystem operations," *EIA/TIA Technical Report IS-41 Revision B*, 1991.

[16] ETSI, *Digital cellular telecommunications system (phase 2+): mobile application part (MAP) specification (GSM 09.02 version 7.51 Release)*, 1998.

[17] Y. Fang, "Movement-based location management and tradeoff analysis for wireless mobile networks," Accepted for publication in *IEEE Transactions on Computers.*

[18] Y. Fang, "General modeling and performance analysis for location management in wireless mobile networks," *IEEE Transactions on Computers*, **51**(10), 1169-1181, October 2002.

[19] Y. Fang, I. Chlamtac and H. Fei, "Analytical results for optimal choice of location update interval for mobility database failure restoration in PCS networks," *IEEE Transactions on Parallel and Distributed Systems*, vol.11, no.6, pp.615-624, June 2000.

[20] Y. Fang, I. Chlamtac and Y. B. Lin, "Portable movement modeling for PCS networks" *IEEE Trans. Veh. Tech.*, **49**(4), 1356-1363, July 2000.

[21] Y. Fang, I. Chlamtac and H. Fei, "Analytical results for optimal choice of location update interval for mobility database failure restoration in PCS networks," *IEEE Transactions on Parallel and Distributed Systems*, Vol.11, No.6, pp.615-624, June 2000.

[22] Y. Fang, I. Chlamtac and H. Fei, "Failure recovery of HLR mobility databases and parameter optimization for PCS networks," *Journal of Parallel and Distributed Computing*, Vol. 60, No. 4, pp. 431-450, April 2000.

[23] Y. Fang and I. Chlamtac, "Teletraffic analysis and mobility modeling for PCS networks," *IEEE Transactions on Communications*, Vol. 47, No. 7, pp. 1062-1072, July 1999.

[24] Y. Fang, I. Chlamtac and Y. B. Lin, "Billing strategies and performance analysis for PCS networks," *IEEE Transactions on Vehicular Technology*, Vol. 48, No. 2, pp. 638-651, 1999.

[25] Y. Fang, I. Chlamtac and Y. B. Lin, "Channel occupancy times and handoff rate for mobile computing and PCS networks," *IEEE Transactions on Computers*, Vol. 47, No. 6, pp. 679-692, 1998.

[26] Y. Fang, I. Chlamtac and Y. B. Lin, "Modeling PCS networks under general call holding times and cell residence time distributions," *IEEE/ACM Transactions on Networking.* Vol. 5, No. 6, pp. 893-906, December, 1997.

[27] Y. Fang, I. Chlamtac and Y. B. Lin, "Call performance for a PCS networks," *IEEE Journal on Selected Areas in Communications*, Vol. 15, No. 8, pp. 1568-1581, October, 1997.

[28] E. Gustafsson, A. Jonsson and C. Perkins , "Mobile IP regional registration," *Internet draft, draft-ietf-mobileip-reg-tunnel-04.txt*, work in progress, March 2001.

[29] J.S.M. Ho and I.F. Akyildiz, "Dynamic hierarchical database architecture for location management in PCS networks," *IEEE/ACM Transactions on Networks*, **5**(5), 646-660, October 1997.

[30] J.S.M. Ho and I.F. Akyildiz, "Local anchor scheme for reducing signaling costs in personal communications networks," *IEEE/ACM Transactions on Networks*, **4**(5), 709-725, October 1996.

[31] J. S. M. Ho and I. F. Akyildiz, "Mobile user location update and paging under delay constraints," *ACM Wireless Networks*, vol. 1, no. 4, pp. 413-425, December 1995.

[32] D. Hong and S. S. Rappaport, "Traffic model and performance analysis for cellular mobile radio telephone systems with prioritized and nonprioritized handoff procedures," *IEEE Trans. Veh. Tech.*, vol.35, no.3, pp.77-92, 198

[33] B. Jabbari, "Teletraffic aspects of evolving and next-generation wireless communication networks," *IEEE Comm. Mag.*, pp.4-9, December 1994.

[34] R. Jain and Y.B. Lin, "An auxiliary user location strategy employing forwarding pointers to reduce network impacts of PCS," *ACM Wireless Networks*, **1**(2), 197-210, 1995.

[35] R. Jain, Y.B. Lin, C.N. Lo and S. Mohan, "A caching strategy to reduce network impacts of PCS," *IEEE Journal on Selected Areas in Communications*, **12**(8), 1434-1445, 1994.

[36] I. Katzela and M. Naghshineh, "Channel assignment schemes for cellular mobile telecommunication systems: a comprehensive survey," *IEEE Personal Communications*. vol.3, no.3, pp.10-31, June 1996.

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

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

[39] W. R. LePage, *Complex Variables and the Laplace Transform for Engineers*. Dover Publications, Inc., New York, 1980.

[40] J. Li, H. Kameda and K. Li, "Optimal dynamic mobility management for PCS networks," *IEEE/ACM Transactions on Networking*, vol.8, no.3, pp.319-327, 2000.

[41] W. Li and A. S. Alfa, "A PCS network with correlated arrival process and splitted-rating channels," *IEEE Journal on Selected Areas in Communications*, **17**(7), 1318-1325, 1999.

[42] Y. B. Lin, "Reducing location update cost in a PCS network," *IEEE/ACM Transactions on Networking*, vol.5, no.1, pp.25-33, 1997.

[43] Y.B. Lin, "Determining the user locations for personal communications networks," *IEEE Transactions on Vehicular Technology*, **43**, 466-473, 1994.

[44] Y.B. Lin and I. Chlamtac, *Wireless and Mobile Network Architectures*, John Wiley and Sons, New York, 2001.

[45] Y.B. Lin, Y.R. Haung, Y.K. Chen and I. Chlamtac, "Mobility management: from GPRS to UMTS,", *Wireless Communications and Mobile Computing (WCMC)*, vol.1, no.4, 339-359, 2001.

[46] Y. B. Lin, S. Mohan and A. Noerpel "Queueing priority channel assignment strategies for handoff and initial access for a PCS network," *IEEE Trans. Veh. Technol.*, vol.43, no.3, pp.704-712, 1994.

[47] W. Ma and Y. Fang, "Two-level pointer forwarding strategy for location management in PCS networks," *IEEE Transactions on Mobile Computing*, Vol.1, No.1, 32-45, Jan.-Mar. 2002.

[48] S. Mohan and R. Jain, "Two user location strategies for personal communications services," *IEEE Personal Communications*, pp.42-50, First quarter, 1994.

[49] C. E. Perkins, "Mobile IP," *IEEE Communications Magazine*, pp. 84-99, May 1997.

[50] R. Ramjee, K. Varadhan, L. Salgarelli, S. R. Thuel, S.-Y. Wang and T. La Porta, "HAWAII: a domain-based approach for supporting mobility in wide-area wireless networks," *IEEE/ACM Trans. of Networking*, vol. 10, no. 5, June 2002.

[51] C. Rose and R. Yates, "Minimizing the average cost of paging under delay constraints," *ACM Wireless Networks*, vol.1, no.2, pp.211–219, 1995.

[52] S. Tabbane, "An alternative strategy for location tracking," *IEEE J. on Selected Areas in Comm.* vol.13, no.5, pp.880-892, 1995.

[53] W. Wang and I.F. Akyildiz, "A dynamic location management scheme for next generation multi-tier PCS systems," *IEEE Transactions on Wireless Communications*, Vol.1., No.1 , pp. 178-190, January 2002.

[54] W. Wang and I.F. Akyildiz, "A new signaling protocol for intersystem roaming in next-generation wireless systems," *IEEE Journal on Selected Areas in Communications*, Vol. 19, No. 10, pp. 2040-2052, October 2001.

[55] W. Wang and I.F. Akyildiz, "Intersystem location update and paging schemes for multitier wireless networks," *Proc. of the ACM MobiCom'2000*, Boston, August 2000.

[56] W. Wang, I.F. Akyildiz and G. Stuber, "Effective paging schemes with delay bounds as QoS constraints in wireless systems," *ACM Wireless Networks*, No. 5, pp. 455-466, September 2001.

[57] J. Xie and I. F. Akyildiz, "A distributed dynamic regional location management scheme for mobile IP," *Proc. IEEE INFOCOM 2002*, pp. 1069-1078, 2002.