Cost Analysis of Movement-Based Location Management in PCS Networks: An Embedded Markov Chain Approach

Xian Wang, *Member, IEEE*, Xianfu Lei, *Member, IEEE*, Pingzhi Fan, *Senior Member, IEEE*, Rose Qingyang Hu, *Senior Member, IEEE*, and Shi-Jinn Horng

***Abstract*—In this paper, we develop an approach of embedded Markov chain to analyze the signaling cost of a movement-based location management (MBLM) scheme. This approach distinguishes itself from those developed in the literature in the following aspects. 1) It considers the location area (LA) architecture used by personal communication service (PCS) networks for location management. 2) It considers two different call handling models that determine after a call whether a location update should be performed. 3) It considers the effect of the call holding time on the call handling models. 4) It proposes to use a fluid flow model to describe the dependence between the cell and the LA residence time. We derive closed-form analytical formulas for the signaling cost, whose accuracy is manifested by a simulation. Based on the analytical formulas, we conduct a numerical study to evaluate the influence of various parameters on the signaling cost. The formulas can contribute to the implementation of the MBLM scheme in PCS networks including Fourth-Generation (4G) Long-Term Evolution. The modeling approach developed in this paper can be exploited to model other location management schemes.**

***Index Terms*—Embedded Markov chain, fluid flow model, location area (LA), modeling, movement-based location management (MBLM).**

Manuscript received September 22, 2012; revised September 8, 2013; accepted September 21, 2013. Date of publication October 9, 2013; date of current version May 8, 2014. This work was supported by the Innovative Intelligence Base Project (111 Project) under Grant 111-2-14. The work of P. Fan was supported in part by the 973 project under Grant 2012CB316100, by the Key Grant Project of the Chinese Ministry of Education under Grant 311031100, and by the Young Innovative Research Team Project of Sichuan Province under Grant 2011JTD0007. The work of X. Wang was supported in part by the National Science Foundation of China under Grant 61171096 and in part by the Fundamental Research Funds for the Central Universities under Grant SWJTU12CX093. The work of S.-J. Horng was supported in part by the One Hundred Talents Program 2012, Sichuan Providence, and by the National Science Council under Contract NSC-99-2916-I-011-002-A1. The review of this paper was coordinated by Prof. J.-M. Chung.

X. Wang and P. Fan are with the School of Information Science and Technology, Southwest Jiaotong University, Chengdu 610031, China (e-mail: drwangxian@gmail.com; p.fan@ieee.org).

X. Lei and R. Q. Hu are with the Department of Electrical and Computer Engineering, Utah State University, Logan, UT 84322-4100 USA (e-mail: xflei81@yahoo.com.cn; rose.hu@usu.edu).

S.-J. Horng is with the School of Information Science and Technology, Southwest Jiaotong University, Chengdu 610031, China, and also with the Department of Computer Science and Information Engineering, National Taiwan University of Science and Technology, Taipei 106, Taiwan (e-mail: horngsj@yahoo.com.tw).

Digital Object Identifier 10.1109/TVT.2013.2285118

I. Introduction

A. Motivation

In a personal communication service (PCS) network, the location of a user equipment (UE) must be known prior to the successful delivery of an incoming call to the UE. The PCS network keeps track of the location of the UE through a mechanism called *location management* (LM). LM comprises two essential operations, namely, *location update* (LU), through which a UE dynamically updates its registration of location in the network, and *paging*, through which upon the arrival of an incoming call, the network pinpoints the accurate area where the UE is currently residing to route the incoming call. The network pages the UE by broadcasting a polling signal in the cells close to the cell where the last LU before the arrival of the incoming call occurs. Both LU and paging incur significant cost due to the usage of wireless and wired bandwidth, the battery power of the UE, and the computing capacities of the UE and the network. Therefore, the cost of LU and paging should be minimized. Intuitively, the more frequently LUs are performed, the more accurate the location information will be, and thus, the less the paging cost will be. However, frequent LUs generate high LU cost. On the other hand, sparse LUs cause the increase in the paging cost. Therefore, a tradeoff exists between the LU and paging costs.

For the purpose of LM, the coverage of a PCS network is partitioned into registration areas, each of which consists of an aggregation of cells. In the circuit-switched (CS) and packet-switched (PS) service domains of Second-Generation (2G) Global System for Mobile Communication (GSM) and Third-Generation (3G) Universal Mobile Telecommunications System (UMTS), these registration areas are called location areas (LAs) and routing areas (RAs), respectively, whereas in Fourth-Generation (4G) Long-Term Evolution (LTE), these registration areas are called tracking areas (TAs). The existing PCS networks use a two-tier system composed of home and visited databases for LM. The home database stores the profiles of the UEs whose primary subscriptions are within the LAs/RAs/TAs that the home database serves. The profile of a UE stored at the home database contains information about the UE's identification, services subscribed, authentication parameters, current location, and so forth. The visited database stores the replications of the profiles and the temporary identities of the UEs that are currently visiting the LAs/RAs/TAs that the

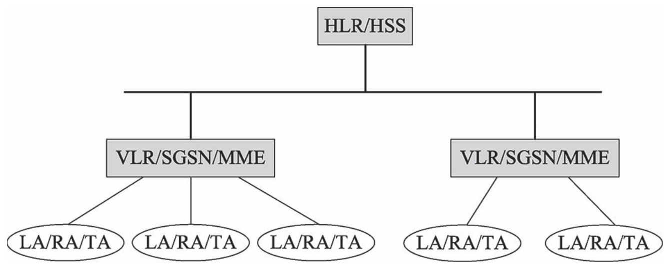

Fig. 1. Architecture used for location management in GSM, UMTS, and LTE.

visited database serves. The home and visited databases are called, respectively, home location register (HLR) and visitor location register (VLR) in the CS domain of GSM and UMTS, HLR and serving GPRS support node (SGSN) in the PS domain of GSM and UMTS, and home subscriber server (HSS) and mobility management entity (MME) in LTE. Fig. 1 shows the architecture used for LM in GSM, UMTS, and LTE. cdmaOne and CDMA2000 use a similar architecture for LM as GSM and UMTS [1]. The architecture shown in Fig. 1 is a reference architecture of the two-tier system used by PCS networks for LM. This architecture has been pervasively used in the literature. Hereafter, we concentrate on the CS domain of GSM and UMTS, and when coming across LM-related terms, unless necessary, we will not bother to mention their counterparts in the PS domain or in other PCS networks. Despite the fact that a VLR can serve more than one LA, in this paper, to simplify the analysis, we assume that a VLR serves only one LA. This assumption was widely used in the literature [2]–[10]. The modeling approach developed in this paper can be easily extended to deal with the situation that a VLR serves more than one LA.

The LM schemes employed in existing 2G and 3G PCS networks, i.e., IS-41 [11] and GSM MAP [12], are intrinsically a static LM scheme, where an LU is triggered by the crossing of an LA boundary. Under the static scheme, when the UE moves from an old LA into a new LA, the UE initiates an LU toward the new VLR that serves the new LA, which, in turn, reports its new location to the HLR. After receiving the report, the HLR instructs the old VLR that serves the old LA to delete the record of the UE and simultaneously transfers the UE's subscriber data to the new VLR. Under the static scheme, all the UEs have paging areas of the same size, i.e., the size of an LA. The static scheme is not cost effective, because it neglects the mobility characteristics of individual UEs. Intuitively, for UEs with low mobility (i.e., frequent call arrivals or, equivalently, large cell residence time), frequent LUs are necessary to reduce the paging cost that, in this case, dominates the signaling cost, whereas for UE devices with high mobility, infrequent LUs are necessary to reduce the LU cost that dominates the signaling cost. Imagine a situation where a UE never has an incoming call. In such a situation, no LU is needed. Therefore, a cost-effective LM scheme should be dynamic and can adapt to different mobility patterns.

To cope with the drawback of the static LM scheme, a plethora of dynamic schemes have been proposed in the literature. The existing dynamic schemes can be classified into three categories. The first category of LM schemes reduce the LU cost by localizing the LU toward the HLR when the UE moves into a new LA, such as the pointer forwarding scheme and its variant proposed in [6], [7], and [13] and the local anchor scheme developed in [8]. The second category of LM schemes exploit the fairly regular moving pattern of the UE to decrease the LU cost, such as the alternative location algorithm proposed in [14]–[16] and its special case, the two location algorithm studied in [9] and [17], and the lookahead strategy proposed in [18]. The third category of LM schemes make a tradeoff between the LU and paging costs to minimize the overall cost. The third category comprises the distance-based LM (DBLM), movement-based LM (MBLM), and time-based LM (TDLM) schemes proposed in [19]. The first and second categories can hardly be realized in existing PCS networks, because the first category requires the introduction of new network entities, and the second category necessitates the moving history that a PCS network may have difficulty to obtain. Under the three schemes of the third category, LUs are performed based on the distance in terms of cells traveled, the number of cells crossed, and the elapsed time since the last LU. Among the three schemes, the DBLM scheme has the lowest signaling cost. However, to realize the DBLM scheme, the network must provide the UE with information about the distance between different cells, which is infeasible in existing PCS networks. The TBLM scheme is the simplest to be implemented. However, it may produce unnecessary signaling traffic (imagine that a stationary UE for a long time does not need to perform an LU before it moves). The MBLM scheme seems to be the most practical because of its effectiveness and easy-to-implement nature [20], [21]. The operation complexity of the MBLM scheme at the UE side is a movement counter used to tally the number of cells crossed. In this paper, we focus on the modeling and cost analysis of the MBLM scheme.

B. Existing Work

Before surveying the literature, first, we introduce two different call handling models that determine after a call whether an LU should be performed. The two models, namely, the *call without LU* (CWLU) model and the *call plus LU* (CPLU) model, were proposed by Li in [22]. Under the CWLU model, no LU is performed after the completion of a call unless during the call, the movement threshold has been reached or an LA boundary has been crossed. Under the CPLU model, an LU is always performed after a call. Note that during the call, no LU is performed [1], because during this period, the network knows the specific cell where the UE is camping. The details of the two models will be given in Section II-D.

Based on whether the LA (i.e., HLR/VLR) architecture and the CWLU and CPLU models are taken into account, the existing researches on the modeling and cost analysis of the MBLM scheme can be assorted into four groups.

- The first group comprises the research that ignored the LA architecture. This group includes [18]–[29]. The emphases of this group are on the following aspects: 1) adopting different probability distributions for the call holding time and the cell residence time [19], [21], [25]; 2) calculating the optimal movement threshold [23], [24]; 3) deriving another performance metric rather than the signaling cost [27]; 4) designing optimal sequential paging schemes [20], [26]; 5) exploring UE moving pattern (history) to enhance the MBLM scheme [18], [28]; 6) developing new analyzing approaches [21], [22], [24]; and 7) applying the MBLM scheme in group LM [29].

- The second group is composed of the researches that took into account the LA architecture. The studies carried out in [2]–[5], [10], and [30] are in this group. When the LA architecture is considered, an LU occurs either when the UE moves into a new LA or when the number of cells crossed by the UE reaches a predefined threshold, which is called a movement threshold. In this case, the challenge in cost analysis for the MBLM scheme is a mathematical model that can describe the LUs due to LA boundary crossings or the reaching of the movement threshold. Li *et al.* in [2] conducted the pioneering work in developing such a model. Xiao *et al.* in [10] and [30] extended the work of Li *et al.* [2] to calculate the optimal fractional movement threshold and design a dynamic MBLM scheme for 3G cellular networks, where HLRs, gateway location registers, and VLRs form a three-tier hierarchical structure of location databases. Based on the model developed in [2], Rodriguez-Dagnino and Takagi, in [3] and [4], developed a renewal theory approach to analyze the signaling cost of the MBLM scheme. The renewal theory approach can release the strict restrictions imposed in [2] and [5] on the statistical distributions of the call interarrival time, as well as the cell and the LA residence time, and thus can deal with more general cases. The model developed in [2] can be enhanced from the following two aspects. First, the model overlooked the VLR LUs due to LA boundary crossings. The cost of these overlooked VLR LUs can be counted toward the signaling cost by increasing the cost of an HLR LU to cover the costs of an HLR and a VLR LU. Second, the model assumes that the cell and the LA boundary are independent, resulting in ambiguity during LU and call delivery processes. In [5], Wang *et al.* advanced the model developed in [2] from the given two aspects. The models used in [2]–[5] lack in a mathematical approach to describe the dependence between the cell and the LA residence time, despite that some forms of dependence between the two variables were considered during the calculation of some conditional probabilities. The dependence between the two variables always exists, just because an LA is made up of cells.
- The third group consists of researches that considered only the CPLU but not the CWLU call handling model. All the existing researches except that done by Li in [22] belong to this group. The fourth group considered both the CWLU and the CPLU model. Hitherto, only [22] belongs to this group. All the existing researches in the third and fourth groups ignored the impact of the call holding time on the CWLU or/and the CPLU model by assuming that the call holding time is nil. The results obtained in this paper suggest that the call holding time has an impact on the signaling costs of both the CWLU and the CPLU model. In addition, [22] also ignored the LA architecture.

As a summary, the existing studies on the modeling and cost analysis of the MBLM scheme leave space for improvement for the following reasons: 1) They failed to develop an approach that can describe the dependence between the cell and the LA residence time and, at the same time, is convenient for mathematical derivations. 2) They failed to develop a mathematical model to assess the impact of the call holding time on the signaling cost of the MBLM scheme under the CWLU and the CPLU model.

C. Our Contributions

Cost analysis is necessary for optimizing the MBLM scheme. In the literature, the characteristics of the LM in PCS networks are not fully respected so as to limit the complexity of mathematical modeling and analysis, such that the results obtained may be unavailable in PCS networks. In this paper, we develop a comprehensive mathematical model to analyze the signaling cost of the MBLM scheme. The main contributions of this paper are fourfold. 1) We take into account the LA architecture. 2) We consider both the CWLU and CPLU call handling models and emphasize the impact of the call holing time. 3) We propose to use a fluid flow model to characterize the dependence between the cell and the LA residence time. 4) We develop an embedded Markov chain to describe the LUs occurring during the call interarrival time and after a call and obtain closed-form analytical results for the signaling cost, whose accuracy is tested by simulation. The results obtained in this paper are instrumental in implementing the MBLM scheme in PCS networks, including 4G LTE. The modeling approach developed in this paper can be harnessed to model other LM schemes, such as the DBLM scheme and the TA list-based LM scheme used in LTE, in PCS networks having the LA or similar architecture.

The thread followed in this paper is as follows: Section II introduces an LA architecture, as well as mobility, traffic, and call handling modes. Sections III and IV analyze the signaling cost of the MBLM scheme under the CWLU and the CPLU model, respectively. Section V tests the accuracy of the analytical results by simulation and carries out a numerical study to investigate the impact of various parameters on the signaling cost. Finally, Section VI concludes this paper.

II. LA Architecture and Mobility, Traffic, and Call Handling Models

A. LA Architecture

A specific LA architecture is not required by the mathematical model developed in this paper, because any LA architecture can be mapped into the model by determining parameter ξ as per (5) given the expected values of the cell and the LA residence time. However, it is required by the simulation that will be carried out in Section V-B to test the accuracy of the modeling approach. Assume that cells are regular hexagons of identical size and that an LA is constructed by arranging cells in concentric cycles around a center cell. Each cell has six neighbor cells. Refer to the set of cells that are r, $r = 1, 2, \ldots$, cells away from the center cell as *ring* r. Particularly, refer to

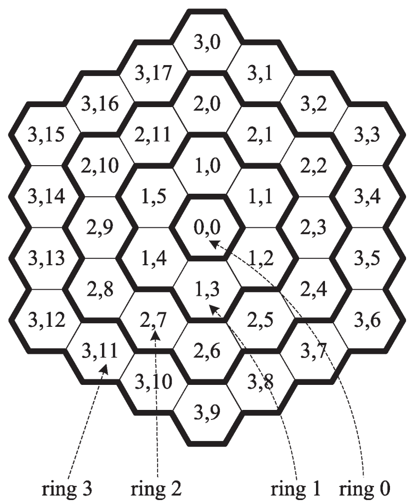

Fig. 2. LA of radius 4.

the center cell as ring 0. Denote by n the radius of an LA in terms of cells. Fig. 2 shows an LA of radius 4. An LA of radius n consists of n rings, ranging from ring 0 to ring $n-1$. Ring 0 has only one cell, and ring r, $r = 1, 2, \ldots, n-1$, has $6r$ cells. The size, i.e., the number of cells, of an LA of radius n, denoted by $\mathcal{O}(n)$, is

$$\mathcal{O}(n) = 3n^2 - 3n + 1, \qquad n = 1, 2, \ldots. \tag{1}$$

B. Mobility Model

Mobility model concerns the cell and the LA residence time, as well as the UE movements among cells. Assume that the cell residence time follows an exponential distribution of rate η. In the literature, a random walk model has been extensively used to describe the UE movements among cells. This model is suitable for simulation but hinders mathematical analysis. To tackle this problem, for mathematical analysis, we use a fluid flow model to characterize the UE movements among cells, whereas for simulation, we use the random walk model. Next, we expatiate on the two models.

1) Random Walk Model: Suppose that LAs are configured as per the ring LA architecture given in Section II-A. Under the random walk model, after leaving its current cell, the UE moves to one of the six neighbor cells of the current cell with equal probability, i.e., 1/6. The random walk model can be accurately described by a 2-D Markov chain with space size $\mathcal{O}(n) + 6(n-1)$ [31]. When n is large, mathematical operations of the 2-D Markov chain will incur high computation complexity. A method that can alleviate this problem is to approximately describe the random walk model by a Markov chain with a smaller space size at the cost of compromising the accuracy. In [31], Akyildiz *et al.* proposed a new approach to simplify the 2-D Markov chain. While retaining high accuracy, the proposed approach reduces the space size of the 2-D Markov chain by about 5/6. To further reduce the space size, 1-D Markov chains can be used to describe the random walk model [20], [26], [32], [33]. A drawback with the Markov chains for the random walk model, whether 2-D or 1-D, is that in most situations, these chains only allow us to do numerical calculations, such that it is arduous to obtain closed-form analytical results. An exception is that when it comes to the expected number of steps required to move out of an LA after entering the LA through ring $n-1$, which is denoted by $S_{n,n}$, the 1-D Markov chain proposed in [20] can produce a very simple and accurate result. When the UE movements between cells are described by the 1-D Markov chain proposed in [20], $S_{n,n}$ can be expressed as [33]

$$S_{n,n} = \frac{3n^2 - 3n + 1}{2n - 1}, \qquad n = 1, 2, \ldots. \tag{2}$$

There is no perceivable discrepancy between the $S_{n,n}$'s calculated as per (2) and the accurate 2-D Markov chain for the random walk model.

2) Fluid Flow Model: The fluid flow model was proposed in [16], [34], and [35] to calculate the rate with which an individual UE or the UE ensemble in a closed plane cross the border of the plane. This rate is referred to as the border crossing rate. In the literature, the fluid flow model has been extensively used to describe the residence times in areas of diverse scales, ranging from a cell, a paging area, an LA, a subnet, and an access network, to a PCS network [36]–[41]. To use the fluid flow model to characterize the UE movements among cells, the following assumptions are necessary. 1) UEs are uniformly distributed throughout an LA. 2) The direction of the UE movement relative to the cell and LA borders is uniformly distributed over $[0, 2\pi)$. 3) The UE speeds at different locations are independent and identically distributed (i.i.d.). 4) All the cells in an LA and all the LAs in the PCS network are homogeneous. That is, the residence times in different cells/LAs are i.i.d. Denote by $E(V)$, L, and A the average speed of a UE and the perimeter and area of the closed plane, respectively. Under the fluid flow model, the border crossing rate of a UE is

$$\frac{E(V)L}{\pi A}. \tag{3}$$

Denote by t_c and t_{LA} with mean $1/\tau$ the cell and the LA residence time, respectively. Remember that t_c has been assumed to follow an exponential distribution of mean $1/\eta$. Denote by L_c/L_{LA} and A_c/A_{LA} the perimeter and the area of a cell/LA, respectively. It follows from (3) that

$$\frac{\tau}{\eta} = \frac{L_{\mathrm{LA}}}{L_c}\frac{A_c}{A_{\mathrm{LA}}}. \tag{4}$$

Denote by ξ the probability that when the UE moves into a new cell from an old cell, the new and old cells belong to the same LA. Since $\tau = (1-\xi)\eta$, it follows that

$$\xi = 1 - \frac{\tau}{\eta}. \tag{5}$$

After leaving a cell, with probabilities ξ and $1-\xi$, a UE remains in and moves out of its current LA, respectively. This trick of determining whether a UE leaves its current LA after leaving a cell was also applied in [37] and underpins the mathematical modeling of the MBLM scheme in this paper. The suitability of this trick will be manifested in Section V-B by a simulation based on the random walk model. Denote by

$f_{\mathrm{LA}}(t)$ the probability density function of t_{LA}. Let $f^*(s)$ be the Laplace transform of function $f(t)$. It follows that

$$\begin{aligned} f^*_{\mathrm{LA}}(s) &= \sum_{k=1}^{+\infty} \xi^{k-1}(1-\xi)\left(\frac{\eta}{s+\eta}\right)^k \\ &= \frac{(1-\xi)\eta}{s+(1-\xi)\eta} = \frac{\tau}{s+\tau}. \end{aligned}$$

The last equation reveals that under the fluid flow model, when the cell residence time is exponentially distributed, the LA residence time must do so. Therefore, in this paper, we assume that t_{LA} follows an exponential distribution of rate τ.

Next, we calculate the given parameters when the LA has the ring architecture introduced in Section II-A. Let R denote the side length of a hexagonal cell. It follows that

$$\begin{aligned} L_{\mathrm{LA}} &= 6 \times 3R + 6(n-2) \times 2R = 6(2n-1)R \\ L_c &= 6R \\ A_{\mathrm{LA}} &= \mathcal{O}(n)A_c \\ \frac{\tau}{\eta} &= \frac{2n-1}{3n^2-3n+1} \qquad (6) \\ \xi &= \frac{3n^2-5n+2}{3n^2-3n+1} \qquad (7) \end{aligned}$$

$n = 1, 2, \ldots$. Equation (6) suggests that the $S_{n,n}$ defined in Section II-A can be expressed as

$$S_{n,n} = \frac{1/\tau}{1/\eta} = \frac{3n^2-3n+1}{2n-1}, \qquad n = 1, 2, \ldots$$

which is identical with (2). Therefore, when it comes to $S_{n,n}$, the fluid flow model does not compromise accuracy.

C. Traffic Model

Denote by t_a and t_h the call interarrival time and the call holding time, respectively. Assume that t_a and t_h follow two exponential distributions of rates λ and μ, respectively. That is, the call arrival process to the UE is Poisson.

D. Call Handling Models

Let case I represent the case that during t_h, the LA of the UE has changed; let case II represent the case that during t_h, the LA of the UE has not changed, but the movement threshold, which is denoted by M, has been reached; and let case III represent the case that during t_h, neither the LA of the UE has changed nor has the movement threshold been reached. Under the CWLU model, after completing a call if case I occurs, then the UE performs an LU toward both the VLR and the HLR; if case II occurs, then the UE performs an LU toward only the VLR; otherwise, if case III occurs, no LU is performed. Under the CPLU model, after completing a call if case I occurs, then the UE acts as under the CWLU model; otherwise, if case II or case III occurs, the UE performs an LU toward only the VLR. Note that no LU is performed during t_h [1]. Under the MBLM scheme implemented in a network with the LA architecture, when case I or case II takes place, after the call, an LU must be performed; otherwise, the network may fail in paging the UE to deliver the next incoming call, because the paging operation is carried out only in cells meeting the following two criteria. First, the distance between these cells and the cell where the last LU before the call arrival occurs is less than M cells. Second, these cells are within the LA where the last LU before the call arrival occurs. If we neglect the influence of t_h by assuming that t_h is nil, then under the CWLU model, no LU occurs after a call, whereas under the CPLU model, an LU always occurs after a call. This is the scenario considered in [22].

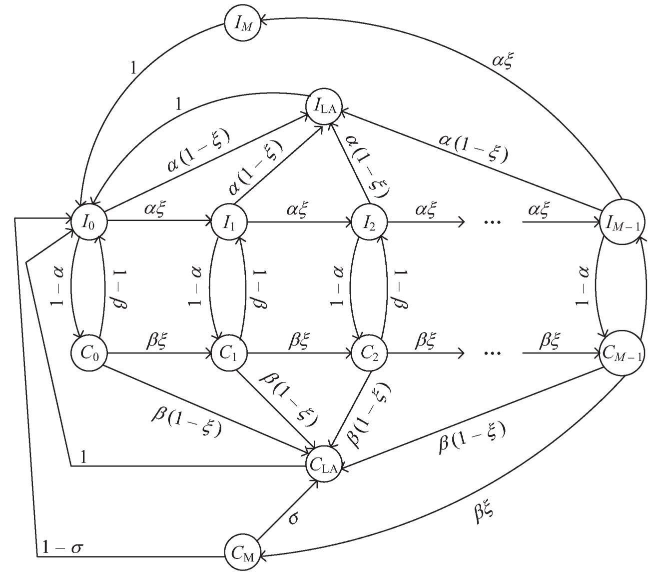

Fig. 3. Embedded Markov chain for the MBLM scheme under the CWLU call handling model.

III. Signaling Cost of the Movement-Based Location Management Scheme Under the Call Without LU Model

A. Embedded Markov Chain Model

Fig. 3 shows an embedded Markov chain for the MBLM scheme under the CWLU call handling model. The states in the chain can be classified into two types. The first type of state, which is denoted by I_{sub}, where sub represents some subscript, comprises the states visited during call interarrival time t_a. The second type of state, which is denoted by C_{sub}, consists of the states visited during call holding time t_h. Specifically, states I_j and C_j, $j = 0, 1, \ldots, M$, represent the states that the movement counter reaches j during t_a and t_h, respectively; state I_{LA} represents the state that during t_a, the UE has just moved into a new LA; and state C_{LA} represents the state that during t_h, the UE has crossed at least one LA.

First, we derive the one-step state transition probabilities of the embedded Markov chain. Denote by $\mathrm{E}_{\mathrm{c}\to\mathrm{c}}$ the event that when the UE moves into a new cell, the new and the old cell belong to the same LA. Denote by $\mathrm{E}_{t_a<t_c}/\mathrm{E}_{t_h<t_c}$ the event that the remaining call interarrival/holding time (which is the time from when the UE enters a state until the arrival/completion of the next/current call) is smaller than the cell residence time. Denote, respectively, by $\mathrm{E}_{\mathrm{c}\to\mathrm{LA}}$, $\mathrm{E}_{t_a>t_c}$, and $\mathrm{E}_{t_h>t_c}$ the events that complement events $\mathrm{E}_{\mathrm{c}\to\mathrm{c}}$, $\mathrm{E}_{t_a<t_c}$, and $\mathrm{E}_{t_h<t_c}$. Owing to the memoryless property of exponential distributions,

the remaining call interarrival/holding time follows the same distribution as the call interarrival/holding time. When in state I_j/C_j, $j = 0, 1, \ldots, M-1$, in the next step, the UE transits to state I_{j+1}/C_{j+1} if events $\mathrm{E}_{t_a>t_c}/\mathrm{E}_{t_h>t_c}$ and $\mathrm{E}_{\mathrm{c}\to\mathrm{c}}$ occur, or transits to state $I_{\mathrm{LA}}/C_{\mathrm{LA}}$ if events $\mathrm{E}_{t_a>t_c}/\mathrm{E}_{t_h>t_c}$ and $\mathrm{E}_{\mathrm{c}\to\mathrm{LA}}$ take place, or transits to state C_j/I_j if event $\mathrm{E}_{t_a<t_c}/\mathrm{E}_{t_h<t_c}$ happens. After entering state I_M or I_{LA}, the UE initiates an LU toward only the VLR or toward both the VLR and the HLR and then transits to state I_0. We assume that the time required in the UE and the network to process an LU is nil. Therefore, states I_M and I_{LA} are transient states with respect to their state residence time. When staying in state C_M, if the remaining call holding time is larger than the remaining LA residence time (i.e., if the UE traverses at least one LA during the remaining call holding time), in the next step, the UE transits to state C_{LA}; otherwise, the UE initiates an LU toward only the VLR and then transits to state I_0. After entering state C_{LA}, in the next step (i.e., after completing the current call), the UE performs an LU toward both the VLR and HLR and then transits to state I_0. Following the given narration, we can write out the set of one-step state transition probabilities of the embedded Markov chain, i.e.,

$$\begin{cases} \Pr(I_j \to I_{j+1}) = \Pr(\mathrm{E}_{t_a>t_c}, \mathrm{E}_{\mathrm{c}\to\mathrm{c}}) = \alpha\xi \\ \Pr(I_j \to I_{\mathrm{LA}}) = \Pr(\mathrm{E}_{t_a>t_c}, \mathrm{E}_{\mathrm{c}\to\mathrm{LA}}) = \alpha(1-\xi) \\ \Pr(I_j \to C_j) = \Pr(\mathrm{E}_{t_a<t_c}) = 1-\alpha \\ \Pr(I_M \to I_0) = \Pr(I_{\mathrm{LA}} \to I_0) = \Pr(C_{\mathrm{LA}} \to I_0) = 1 \\ \Pr(C_j \to C_{j+1}) = \Pr(\mathrm{E}_{t_h>t_c}, \mathrm{E}_{\mathrm{c}\to\mathrm{c}}) = \beta\xi \\ \Pr(C_j \to C_{\mathrm{LA}}) = \Pr(\mathrm{E}_{t_h>t_c}, \mathrm{E}_{\mathrm{c}\to\mathrm{LA}}) = \beta(1-\xi) \\ \Pr(C_j \to I_j) = \Pr(\mathrm{E}_{t_h<t_c}) = 1-\beta \\ \Pr(C_M \to C_{\mathrm{LA}}) = \Pr(t_h > t_{\mathrm{LA}}) = \sigma \\ \Pr(C_M \to I_0) = \Pr(t_h < t_{\mathrm{LA}}) = 1-\sigma \end{cases}$$

$j = 0, 1, \ldots, M-1$, where ξ is given in (5), and

$$\alpha \triangleq \frac{\eta}{\eta+\lambda}$$
$$\beta \triangleq \frac{\eta}{\eta+\mu}$$
$$\sigma \triangleq \frac{\eta(1-\xi)}{\eta(1-\xi)+\mu}.$$

Next, we derive the equilibrium probabilities of the embedded Markov chain. Denote by Θ the state space of the chain and by Π_θ, $\theta \in \Theta$, the equilibrium probability of state θ. In the embedded Markov chain, Π_θ is directly proportional to the number of times rather than the duration time state θ has been visited, which explains that although states I_M and I_{LA} are transient states, they have nonzero equilibrium probabilities [cf. (8)]. It follows from Fig. 3 that the set of balance equations of the chain is given by

$$\begin{cases} \Pi_{I_0} = \Pi_{I_M} + \Pi_{I_{\mathrm{LA}}} + \Pi_{C_M}(1-\sigma) + \Pi_{C_{\mathrm{LA}}} + \Pi_{C_0}(1-\beta) \\ \Pi_{I_j} = \Pi_{I_{j-1}}\alpha\xi + \Pi_{C_j}(1-\beta) \\ \Pi_{I_M} = \Pi_{I_{M-1}}\alpha\xi \\ \Pi_{I_{\mathrm{LA}}} = (\Pi_{I_0} + \Pi_{I_1} + \cdots + \Pi_{I_{M-1}})\alpha(1-\xi) \\ \Pi_{C_0} = \Pi_{I_0}(1-\alpha) \\ \Pi_{C_j} = \Pi_{C_{j-1}}\beta\xi + \Pi_{I_j}(1-\alpha) \\ \Pi_{C_M} = \Pi_{C_{M-1}}\beta\xi \\ \Pi_{C_{\mathrm{LA}}} = (\Pi_{C_0} + \Pi_{C_1} + \cdots + \Pi_{C_{M-1}})\beta(1-\xi) + \Pi_{C_M}\sigma \end{cases}$$

$j = 1, \ldots, M-1$. After some mathematical manipulations, we have

$$\Pi_{I_j} = \frac{\xi^j}{\alpha+\beta-2\alpha\beta}\left\{\left[\beta(1-\alpha)\Delta^j + \alpha(1-\beta)\right]\Pi_{I_0} + \beta(1-\beta)(1-\Delta^j)\Pi_{C_0}\right\}$$

$$\Pi_{C_j} = \frac{\xi^j}{\alpha+\beta-2\alpha\beta}\left\{\left[\alpha(1-\beta)\Delta^j + \beta(1-\alpha)\right]\Pi_{C_0} + \alpha(1-\alpha)(1-\Delta^j)\Pi_{I_0}\right\}$$

$j = 0, 1, \ldots, M-1$, where

$$\Delta \triangleq \frac{\alpha\beta}{\alpha+\beta-\alpha\beta}.$$

Therefore, we obtain the set of equilibrium probabilities expressed in terms of Π_{I_0}, i.e.,

$$\begin{cases} \Pi_{I_j} = \frac{\Pi_{I_0}}{\alpha+\beta-2\alpha\beta}\begin{bmatrix} \beta^2(1-\alpha)(\Delta\xi)^j \\ +(1-\beta)(\alpha+\beta-\alpha\beta)\xi^j \end{bmatrix} \\ \Pi_{C_j} = \frac{\Pi_{I_0}}{\alpha+\beta-2\alpha\beta}\begin{bmatrix} -\alpha\beta(1-\alpha)(\Delta\xi)^j \\ +(1-\alpha)(\alpha+\beta-\alpha\beta)\xi^j \end{bmatrix} \\ \Pi_{I_M} = \alpha\xi\Pi_{I_{M-1}} \\ \Pi_{C_M} = \beta\xi\Pi_{C_{M-1}} \\ \Pi_{I_{\mathrm{LA}}} = \frac{\Pi_{I_0}\alpha(1-\xi)}{\alpha+\beta-2\alpha\beta}\begin{bmatrix} \beta^2(1-\alpha)S(M,\Delta\xi) \\ +(1-\beta)(\alpha+\beta-\alpha\beta)S(M,\xi) \end{bmatrix} \\ \Pi_{C_{\mathrm{LA}}} = \frac{\Pi_{I_0}\beta(1-\xi)}{\alpha+\beta-2\alpha\beta}\begin{bmatrix} -\alpha\beta(1-\alpha)S(M,\Delta\xi) \\ +(1-\alpha)(\alpha+\beta-\alpha\beta)S(M,\xi) \end{bmatrix} \\ \qquad + \beta\xi\sigma\Pi_{C_{M-1}} \end{cases} \tag{8}$$

$j = 0, 1, \ldots, M-1$, where

$$S(m,x) \triangleq \sum_{k=0}^{m-1} x^k = \frac{1-x^m}{1-x}, \qquad m \geq 1.$$

Finally, we deal with the state residence time, i.e., the time from when the UE enters a state until the UE leaves the state. Denote by r_θ, $\theta \in \Theta$, the residence time in state θ. It follows that

$$\begin{cases} r_{I_j} = \min(t_a, t_c) \\ r_{I_M} = r_{I_{\mathrm{LA}}} = 0 \\ r_{C_j} = \min(t_h, t_c) \\ r_{C_M} = \min(t_h, t_{\mathrm{LA}}) \\ r_{C_{\mathrm{LA}}} = t_h \end{cases} \tag{9}$$

$j = 0, 1, \ldots, M-1$. Denote by r the time between two consecutive state transitions. It follows that

$$r = \bigcup_{\theta\in\Theta} \Pi_\theta\, r_\theta$$

$$\overline{r} \triangleq E(r) = \sum_{\theta\in\Theta} \Pi_\theta E(r_\theta) = \Pi_{I_{\mathrm{LA}}}\frac{1-\alpha}{\alpha(1-\xi)}\left(\frac{1}{\lambda}+\frac{1}{\mu}\right). \tag{10}$$

$$\begin{cases} n \geq M \begin{cases} \text{Case I: } n \leq 2M-2 & (12) \\ \text{Case II: } n \geq 2M-1 & (13) \end{cases} \\ n+1 \leq M \leq 2n-2 \begin{cases} \text{Case III: } 2M \leq 3n-2 \quad (14) \\ 2M \geq 3n-1 \begin{cases} \text{Case IV: } 2M = 3n-1 & (15) \\ \text{Case V: } 2M \geq 3n & (16) \end{cases} \end{cases} \\ \text{Case VI: } M \geq 2n-1 \quad (17) \end{cases}$$

Fig. 4. Different cases considered in the calculation of the PA size.

B. Paging Scheme

Refer to the area where the UE can be located upon the arrival of an incoming call as the *paging area* (PA). Paging schemes can be classified into two categories, namely, the *parallel scheme* that simultaneously pages all the cells in a PA and *sequential schemes* that partition a PA into disjoint subareas and page these subareas one after another until the UE is located. Compared with the parallel scheme, sequential schemes have less signaling cost but longer delay.

Denote by $P(j)$, $j = 0, 1, \ldots, M-1$, the probability that upon the arrival of an incoming call, the UE has crossed j cells since the last LU prior to the call arrival. It follows that

$$P(j) = \frac{\Pi_{I_j} E(r_{I_j})}{\sum_{k=0}^{M-1} \Pi_{I_k} E(r_{I_k})}, \qquad j = 0, 1, \ldots, M-1.$$

It follows from (8) and (9) that

$$P(j) = \frac{\beta^2(1-\alpha)(\Delta\xi)^j + (1-\beta)(\alpha+\beta-\alpha\beta)\xi^j}{\beta^2(1-\alpha)S(M,\Delta\xi) + (1-\beta)(\alpha+\beta-\alpha\beta)S(M,\xi)}. \tag{11}$$

Denote by $P_{\text{d}}(j)$, $j = 0, 1, \ldots, M-1$, the probability that upon a call arrival, the UE is j cells away from the cell where the last LU occurs. The distribution of $P_{\text{d}}(j)$ is a prerequisite for designing optimal sequential paging schemes that can meet a bound on the paging delay and, at the same time, minimize the paging cost. For the case of no LA architecture, given the $P(j)$ distribution, the combinatorial method developed in [20] can be used to calculate $P_{\text{d}}(j)$. However, when the LA architecture exists, such a method does not exist and deserves a separate research. For this reason, in this paper, we focus on the parallel paging scheme.

Define PA size as the number of cells in a PA. The PA size is independent of the call handling model. That is, the CWLU and CPLU models have the same PA size. Suppose that LAs are configured as per the ring architecture given in Section II-A and that the UE movements among cells follow the random walk model given in Section II-B1. Due to space limitations, the calculation of the PA size will be detailed in a separate paper. Here, we present only the final results for the PA size. Denote by N_{PA} the PA size. As shown in Fig. 4, N_{PA} is calculated in six different cases. In Fig. 4, after a case, the number of the equation for the case is given. Equations (12)–(17) are as follows:

$$N_{\text{PA}} = \mathcal{O}(M) + p_1 + p_2 + p_3, \qquad M \leq n \leq 2M-2 \tag{12}$$

where $\mathcal{O}(\cdot)$ is defined in (1) and

$$\begin{aligned} k &\triangleq 2M-2-n, \\ p_1 \times \mathcal{O}(n) &\triangleq -M(M-1)(2M-1)(3n-M-1) \\ p_2 \times \mathcal{O}(n) &\triangleq \frac{1}{4}(M-k-2)^2(M-k-3)(9M+3k+1) \\ &\quad - \frac{1}{2}(M-k-2)(M-k-3) \\ p_3 \times \mathcal{O}(n) &= 2(k+2)(2M+k)(M-1-k)(M-2-k) \\ &\quad + \frac{3}{2}(k+2)(k-1)(M-2-k)(M+1+k) \\ &\quad + \frac{1}{4}k(k+1)(k+2)(k+3). \end{aligned}$$

$$\begin{aligned} N_{\text{PA}} &= \mathcal{O}(M) - \frac{1}{\mathcal{O}(n)}\frac{1}{4}M(M-1) \\ &\quad \times (31M^2 + 16Mk - 39M - 8k + 14), \\ &\quad n \geq 2M-1 \end{aligned} \tag{13}$$

where $k \triangleq n - 2M + 1$.

$$N_{\text{PA}} = \mathcal{O}(n) + p_1 + p_2 + p_3, \quad M \geq n+1, 2M \leq 3n-2 \tag{14}$$

where

$$\begin{aligned} k &\triangleq 3n-2M-2 \\ p_1 \times \mathcal{O}(n) &\triangleq -(2n-1)(2n-M-1)(2n-M)(n+M-1) \\ p_2 \times \mathcal{O}(n) &\triangleq \frac{1}{4}(M-n-1)(M-n) \\ &\quad \times(-3M^2 - 15n^2 + 18Mn - 5M + 5n - 2) \\ p_3 \times \mathcal{O}(n) &\triangleq 2(M-n)(M-n+1)(M+n+k)(k+2) \\ &\quad + \frac{3}{2}(M-n)(M+k+1)(k-1)(k+2) \\ &\quad + \frac{1}{4}k(k+1)(k+2)(k+3). \end{aligned}$$

$$\begin{aligned} N_{\text{PA}} &= \mathcal{O}(n) + p_1 + p_2 + p_3, \\ &\quad 2M = 3n-1, \quad M \leq 2n-2 \end{aligned} \tag{15}$$

where

$$\begin{aligned} p_1 \times \mathcal{O}(n) &\triangleq -(2n-1)(2n-M)(2n-M-1)(n+M-1) \\ p_2 \times \mathcal{O}(n) &\triangleq \frac{3}{64}(n-1)(n-3)(n-5)(7n-1) \\ p_3 \times \mathcal{O}(n) &\triangleq \frac{1}{4}(n-1)(10n^2-23n+3). \end{aligned}$$

$$\begin{aligned} N_{\text{PA}} &= \mathcal{O}(n) + \frac{1}{4}\frac{1}{\mathcal{O}(n)}(2n-M)(2n-M-1) \\ &\quad \times (5M^2 - 8n^2 - 12Mn + M + 14n - 6), \\ &\quad 2M \geq 3n, \quad M \leq 2n-2. \end{aligned} \tag{16}$$

$$N_{\text{PA}} = \mathcal{O}(n), \quad M \geq 2n-1. \tag{17}$$

The correctness of (12)–(17) will be manifested in Section V-B through a simulation.

TABLE I
COMPARISON BETWEEN N'_{PA} AND N_{PA} WHEN $n = 15$

M	1	5	10	15	20	25	29
N'_{PA}	1	61	271	631	1141	1801	2437
N_{PA}	1	52.9	195.5	368.4	571.8	607.7	631
rel. diff. (%)	0	15.4	38.6	71.3	120.4	196.4	286.2

In the literature, whether or not the LA architecture is considered, the PA size, which is denoted by N'_{PA}, was unanimously expressed as

$$N'_{\mathrm{PA}} = \mathcal{O}(M) \tag{18}$$

which apparently has no relation with the LA architecture and is an upper bound on the PA size. Table I compares the PA size calculated as per (18) and that calculated as per (12)–(17), when $n = 15$, and M ranges from 1 to 29. In Table I, the entry *rel. diff.* represents the relative difference of N'_{PA} to N_{PA}.

C. Signaling Cost

In PCS networks with the LA architecture, there are four types of LU, denoted, respectively, by a VLR LU, a VLR′, an HLR LU, and an HLR′ LU. A VLR/HLR LU is an LU performed to update the UE registration of location in the VLR/HLR when no radio resource (RR) connection exists before performing the LU. A VLR′/HLR′ LU is an LU performed to update the UE registration of location in the VLR/HLR right after the completion of a call. Since an RR connection has been already established at the beginning of a call, there is no need to establish another RR connection for performing a VLR′/HLR′ LU. Thus, compared with a VLR/HLR LU, a VLR′/HLR′ LU has less signaling cost. Denote by $N_{\mathrm{VLR,st}}/N_{\mathrm{VLR',st}}/N_{\mathrm{HLR,st}}/N_{\mathrm{HLR',st}}$ and $N_{\mathrm{paging,st}}$ the expected numbers of VLR/VLR′/HLR/HLR′ LUs performed and cells paged due to a state transition (cf. Fig. 3), respectively. Referring to Fig. 3, the transition $I_M \to I_0$ results in a VLR LU; the transition $I_{\mathrm{LA}} \to I_0$ generates a VLR and an HLR LU; the transition $C_M \to I_0$ incurs a VLR′ LU; the transition $C_{\mathrm{LA}} \to I_0$ triggers a VLR′ and an HLR′ LU; and the transition $I_j \to C_j,\ j = 0, 1, \ldots, M-1$, leads to the paging of N_{PA} cells. It follows that

$$\begin{cases} N_{\mathrm{VLR,st}} = \Pi_{I_M} + \Pi_{I_{\mathrm{LA}}} \\ N_{\mathrm{VLR',st}} = \Pi_{C_M}(1-\sigma) + \Pi_{C_{\mathrm{LA}}} \\ N_{\mathrm{HLR,st}} = \Pi_{I_{\mathrm{LA}}} \\ N_{\mathrm{HLR',st}} = \Pi_{C_{\mathrm{LA}}} \\ N_{\mathrm{paging,st}} = \sum\limits_{j=0}^{M-1} \Pi_{I_j}(1-\alpha)N_{\mathrm{PA}}. \end{cases}$$

Denote by $\delta_{\mathrm{VLR}}/\delta_{\mathrm{VLR'}}/\delta_{\mathrm{HLR}}/\delta_{\mathrm{HLR'}}$ and δ_{paging} the costs of performing a VLR/VLR′/HLR/HLR′ LU and a paging in a cell, respectively. Denote by $\mathrm{TC_{st}}$ the expected signaling cost per state transition, which is the cost due to the LUs and paging triggered by a state transition. It follows that

$$\begin{aligned} \mathrm{TC_{st}} = {} & N_{\mathrm{VLR,st}}\delta_{\mathrm{VLR}} + N_{\mathrm{VLR',st}}\delta_{\mathrm{VLR'}} + N_{\mathrm{HLR,st}}\delta_{\mathrm{HLR}} \\ & + N_{\mathrm{HLR',st}}\delta_{\mathrm{HLR'}} + N_{\mathrm{paging,st}}\delta_{\mathrm{paging}}. \end{aligned}$$

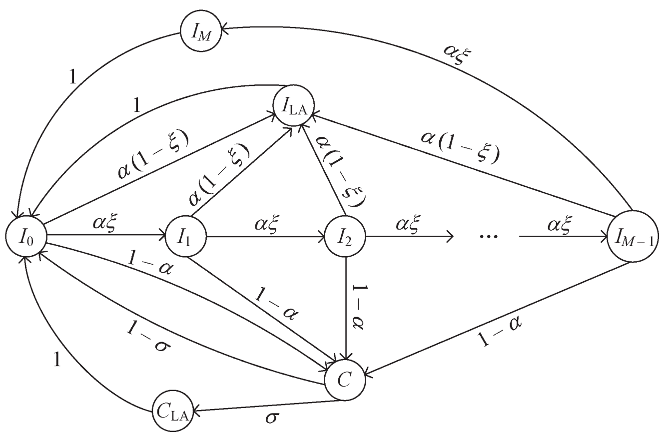

Fig. 5. Embedded Markov chain for the MBLM scheme under the CPLU call handling model.

Since $\mathrm{TC_{st}}$ is the cost generated during the time between two state transitions r, the expected signaling cost per unit time, which is denoted by $\mathrm{TC_{ut}}$, is

$$\mathrm{TC_{ut}} = \mathrm{TC_{st}}/\overline{r}.$$

In the literature, the de facto performance metrics regarding the signaling cost of an LM scheme include the expected number of LUs performed during the call interarrival time and the expected number of cells paged to locate the UE when an incoming call arrives. That is, the recurrent period over which the signaling cost due to LUs and paging is evaluated is the call interarrival time. We refer to this type of signaling cost as the *signaling cost per call arrival*. When the call holding time is considered, the recurrent period for evaluating the signaling cost per call arrival is the sum of the call interarrival time and the call holding time. Therefore, the expected signaling cost per call arrival, which is denoted by TC, is

$$\mathrm{TC} = \mathrm{TC_{ut}}\left(\frac{1}{\lambda} + \frac{1}{\mu}\right) = \frac{\mathrm{TC_{st}}}{\overline{r}}\left(\frac{1}{\lambda} + \frac{1}{\mu}\right). \tag{19}$$

Denote by $N_{\mathrm{VLR}}/N_{\mathrm{VLR'}}/N_{\mathrm{HLR}}/N_{\mathrm{HLR'}}$ and N_{paging} the expected numbers of VLR/VLR′/HLR/HLR′ LUs performed and cells paged per call arrival, respectively. It follows from (19) that

$$\begin{cases} N_{\mathrm{VLR}} = \frac{\eta}{\lambda}\left[(1-\xi) + \xi P(M-1)\right] \\ N_{\mathrm{VLR'}} = \sigma + \frac{(1-\sigma)\eta\xi^M(1-\Delta^M)}{\lambda\Delta S(M,\Delta\xi) + \mu S(M,\xi)} \\ N_{\mathrm{HLR}} = \frac{\eta}{\lambda}(1-\xi) \\ N_{\mathrm{HLR'}} = \sigma \\ N_{\mathrm{paging}} = N_{\mathrm{PA}}. \end{cases} \tag{20}$$

Appendix A proves (20).

IV. SIGNALING COST OF THE MOVEMENT-BASED LOCATION MANAGEMENT SCHEME UNDER THE CALL PLUS LU MODEL

Fig. 5 shows an embedded Markov chain for the MBLM scheme under the CPLU model, where state C represents the state that an incoming call arrives. It follows from Fig. 5 that

$$\begin{aligned} \Pi_{I_j} &= \Pi_{I_{j-1}}\alpha\xi, \qquad j = 1, \ldots, M-1 \\ \Pi_{I_j} &= (\alpha\xi)^j \Pi_{I_0}, \qquad j = 0, 1, \ldots, M-1. \end{aligned}$$

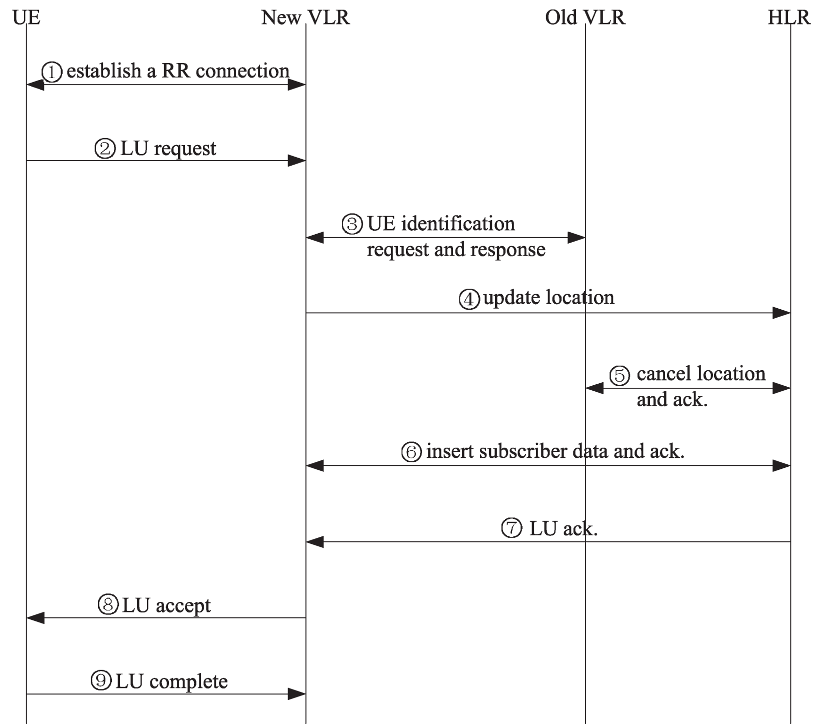

Fig. 6. Simplified procedure of a combined VLR and HLR LU.

The residence time in state I_j is

$$r_{I_j} = \min(t_a, t_c), \qquad j = 0, 1, \ldots, M-1.$$

The probability that upon the arrival of an incoming call the UE is in state I_j is

$$P(j) = \frac{\Pi_{I_j} E(r_{I_j})}{\sum_{k=0}^{M-1} \Pi_{I_k} E(r_{I_k})} = \frac{\Pi_{I_j}}{\sum_{k=0}^{M-1} \Pi_{I_k}} = \frac{(\alpha\xi)^j}{S(M, \alpha\xi)} \qquad (21)$$

$j = 0, 1, \ldots, M-1$. Equation (21) suggests that the $P(j)$ distribution of the CPLU model is independent of the call holding time, because under the CPLU model, after a call, an LU always occurs irrespective of the call holing time. In [26], without considering the LA architecture, Zhu and Leung derived the $P(j)$ distribution of the MBLM scheme under the CPLU model. The $P(j)$ derived in [26, eq. (20)] is

$$P(j) = \frac{\alpha^j}{S(M, \alpha)}, \qquad j = 0, 1, \ldots, M-1. \qquad (22)$$

When there is no LA architecture, the LA boundary crossing rate $\tau = 0$. It follows from (5) that, in this case, $\xi = 1$. When $\xi = 1$, (21) reduces to (22). Following the principle of the CPLU model and the logic used in Appendix A to prove (20), without going through the standard steps followed in Section III to reach (20), we can directly reach

$$\begin{cases} N_{\mathrm{VLR}} = \frac{\eta}{\lambda}\left[(1-\xi) + \xi P(M-1)\right] \\ N_{\mathrm{VLR}'} = 1 \\ N_{\mathrm{HLR}} = \frac{\eta}{\lambda}(1-\xi) \\ N_{\mathrm{HLR}'} = \sigma \\ N_{\mathrm{paging}} = N_{\mathrm{PA}}. \end{cases} \qquad (23)$$

Appendix B discusses the N_{VLR} given in (23). Therefore, the expected signaling cost per call arrival of the MBLM scheme under the CPLU model is

$$\mathrm{TC} = N_{\mathrm{VLR}}\delta_{\mathrm{VLR}} + N_{\mathrm{VLR}'}\delta_{\mathrm{VLR}'} + N_{\mathrm{HLR}}\delta_{\mathrm{HLR}} + N_{\mathrm{HLR}'}\delta_{\mathrm{HLR}'} + N_{\mathrm{paging}}\delta_{\mathrm{paging}}.$$

V. Performance Evaluation

A. Preparation

Before performance evaluation, first, we need to determine the values of parameters δ_{paging}, δ_{VLR}, $\delta_{\mathrm{VLR}'}$, δ_{HLR}, and $\delta_{\mathrm{HLR}'}$. We assume that the cost of a paging/LU is due to the usage of wired and wireless bandwidth and neglect other factors that may contribute to the cost. Fig. 6 shows a simplified procedure of a combined VLR and HLR LU [42], [43]. The procedures of the four types of LU can be tailored from Fig. 6. A VLR LU consists of steps ①, ②, and ⑧. A VLR′ LU is composed of steps ② and ⑧. Both an HLR and an HLR′ LU are made up of steps ③–⑦, and ⑨. Referring to Fig. 6, the links involved to perform an LU can be classified into six types, namely, a local/long-haul wired link that connects two adjacent/distant network entities using a wired link, and a common/dedicated wireless uplink/downlink over which information is transmitted from the UE/VLR to the VLR/UE using a common/dedicated wireless link. Considering the asymmetrical capacities of wireless uplink and downlink and the abundant capacity of wired links, in Table II, we set the costs of the six types of link. In Fig. 6, step ① uses one common wireless uplink and one common wireless downlink. Note that the common wireless uplink used in step ① to transmit the UE's access request is a random access channel (RACH). The RACH is a contention-based channel and, each time, can convey only one UE's access request if there is no collision that occurs when more than one UE simultaneously select the same RACH to transmit their

TABLE II
COSTS OF VARIOUS TYPES OF LINK

link type	cost
local wired link	1
long-haul wired link	5
common wireless downlink	5
common wireless uplink	25
dedicated wireless downlink	25
dedicated wireless uplink	125

access requests. Therefore, when counting the cost, the RACH used in step ① should be treated as a dedicated wireless uplink. Moreover, the UE may have contended several times for the RACH before successfully gaining the RACH to transmit its access request. The equivalent effect is that the UE has used the RACH a number of times, which is denoted by N_{RACH}, to transmit its access request. N_{RACH} depends on such factors as the number of RACHs allocated and the traffic intensity in a cell. In this paper, we randomly assume that $N_{\mathrm{RACH}} = 1.5$. Both steps ② and ⑨ use one dedicated wireless uplink; step ③ uses two local wired links; both steps ④ and ⑦ use one long-haul wired link; both steps ⑤ and ⑥ use two long-haul wired links; step ⑧ uses one dedicated wireless downlink. When performing a paging in a cell, only one common wireless downlink is involved. Therefore, we have

$$\begin{cases} \delta_{\mathrm{VLR}} = N_{\mathrm{RACH}} \times 125 + 5 + 125 + 25 = 342.5 \\ \delta_{\mathrm{VLR}'} = 125 + 25 = 150 \\ \delta_{\mathrm{HLR}} = \delta_{\mathrm{HLR}'} = 2 \times 1 + 5 + 2 \times 5 + 2 \times 5 + 5 + 125 \\ \qquad = 157 \\ \delta_{\mathrm{paging}} = 5. \end{cases}$$

B. Comparison Between Simulation and Analytical Results

To test the accuracy of the modeling approach developed in this paper, here, we compare the signaling cost of the MBLM scheme obtained through simulation with those calculated as per the analytical formulas derived in this paper and [2]–[5]. In the simulation, LAs are configured as per the ring architecture introduced in Section II-A, and the UE movements among cells follow the random walk model introduced in Section II-B1. As a convention, only after the simulation has entered a steady state can we begin to collect simulation data required to calculate a performance metric. Usually, the realization of the steady state necessitates that the simulation has been running for a long enough time before we begin to collect the simulation data. This conventional method of realizing the steady state is time consuming. Moreover, it is difficult to determine whether the simulation has entered the steady state. It can be proved that the equilibrium probability for the UE to reside in any of the cells within an LA is $1/\mathcal{O}(n)$. That is, in the steady state, the UE has equal probability to reside in any cell. This property can be exploited to design a simpler method to realize the steady state without the warm-up process required by the conventional method. As shown in Fig. 2, identify a cell in ring x by a coordinate (x, y), $x = y = 0$, or $1 \le x \le n-1$ and $0 \le y \le 6x-1$. Let $\mathrm{PM}(x, y)$ be the performance metric obtained by simulating a certain number of incoming call arrivals, which is denoted by N_{call}, when the location of the UE at the beginning of the simulation is (x, y), $x = y = 0$, or $1 \le x \le n-1$ and $0 \le y \le 6x-1$. From a statistical perspective, the performance metric, which is denoted by PM, can be expressed as

$$\mathrm{PM} = \left[\mathrm{PM}(0,0) + \sum_{x=1}^{n-1}\sum_{y=0}^{6x-1} \mathrm{PM}(x,y)\right] \frac{1}{\mathcal{O}(n)}. \tag{24}$$

In [31], Akyildiz *et al.* observed that for two different cells belonging to ring x, which are denoted by (x, y_1) and (x, y_2), $1 \le x \le n-1$ and $0 \le y_1, y_2 \le 6x-1$, if $y1 \equiv y2 \bmod x$, then (x, y_1) and (x, y_2) are of the same type. See [31] for the definition of the term *type* used here. Thus, there are x types of cell in ring x, $1 \le x \le n-1$; each type has six cells. When $y_1 \equiv y_2 \bmod x$, $\mathrm{PM}(x, y_1) = \mathrm{PM}(x, y_2)$, $1 \le x \le n-1$ and $0 \le y_1, y_2 \le 6x-1$. For all the cells in ring x, $1 \le x \le n-1$, we only need to obtain $\mathrm{PM}(x,0), \mathrm{PM}(x,1), \ldots, \mathrm{PM}(x, x-1)$ instead of all $\mathrm{PM}(x,y)$'s, $0 \le y \le 6x-1$. Therefore, (24) can be simplified as

$$\mathrm{PM} = \left[\mathrm{PM}(0,0) + 6 \times \sum_{x=1}^{n-1}\sum_{y=0}^{x-1} \mathrm{PM}(x,y)\right] \frac{1}{\mathcal{O}(n)}. \tag{25}$$

Compared with the simulation based on (24), that based on (25) can reduce the simulation time by about 5/6. Equation (25) is the rationale behind the simulation conducted in this paper.

In the literature, when modeling the MBLM scheme, references [2]–[5] considered the LA architecture. Although not explicitly stated, the call handling model used in [2]–[5] is the CPLU model, and the influence of call holding time t_h on the call handling model was neglected by assuming that $t_h = 0$. When $t_h = 0$, under the CPLU model, after a call, a VLR′ LU is performed. However, [2]–[5] overlooked this VLR′ LU. Moreover, the N_{VLR}'s derived in [2]–[4] overlooked the VLR LUs due to LA boundary crossings. To make the ensuing comparison fair, the N_{VLR}'s derived in [2]–[4] are amended in this paper to incorporate the VLR LUs due to LA boundary crossings, and the $N_{\mathrm{VLR}'}$ for [2]–[5] is set to 1. The N_{VLR} derived in [2] by Li *et al.* is as follows:

$$N_{\mathrm{VLR}} = \frac{\eta}{\lambda}(1-\xi)\frac{1}{(2-\xi)^M - 1} + \frac{\eta^M}{(\eta+\lambda)^M - \eta^M} + \boxed{\frac{\eta}{\lambda}(1-\xi)}. \tag{26}$$

The N_{VLR} derived in [3] and [4] by Rodriguez-Dagnino and Takagi is as follows:

$$\begin{aligned} N_{\mathrm{VLR}} =& \frac{\lambda}{\eta(1-\xi)+\lambda}\frac{1}{(2-\xi+\lambda/\eta)^M - 1} \\ &+ \frac{\eta(1-\xi)}{\eta(1-\xi)+\lambda}\frac{1}{(2-\xi)^M - 1}\left[2 + \frac{\eta}{\lambda}(1-\xi)\right] \\ &+ \boxed{\frac{\eta}{\lambda}(1-\xi)}. \end{aligned} \tag{27}$$

The component $\boxed{\eta/\lambda(1-\xi)}$ in (26) and (27) represents the number of VLR LUs due to LA boundary crossings that were overlooked in [2]–[4]. The N_{VLR} derived in [5] by Wang *et al.* is as follows:

$$N_{\mathrm{VLR}} = \frac{\frac{\eta}{\lambda}(1-\xi)(2-\xi+\lambda/\eta)^M + 1}{(2-\xi+\lambda/\eta)^M - 1}. \tag{28}$$

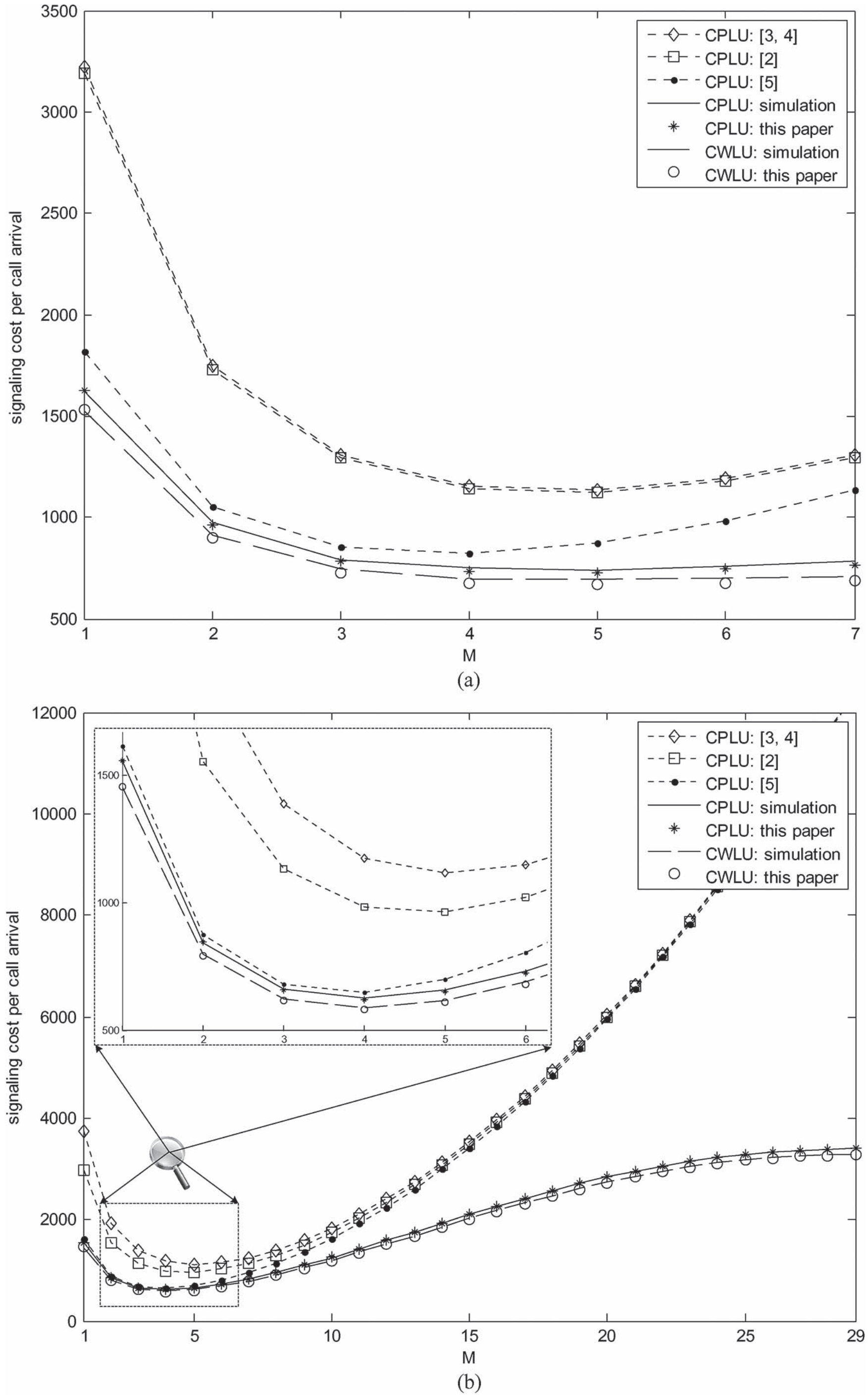

Fig. 7. Comparison between the simulation and the analytical result of the signaling cost. $\lambda = 1$, $\mu = 8$, and $\eta = 4$. (a) $n = 5$ and $N_{\mathrm{call}} = 5000$. (b) $n = 15$ and $N_{\mathrm{call}} = 1000$.

The other four quantities derived in [2]–[5] are as follows:

$$\begin{cases} N_{\mathrm{VLR}'} = 1 \\ N_{\mathrm{HLR}} = \frac{\eta}{\lambda}(1-\xi) \\ N_{\mathrm{HLR}'} = 0 \\ N_{\mathrm{paging}} = \mathcal{O}(M). \end{cases}$$

The N_{HLR} derived in [2]–[5] is accurate.

Fig. 7 compares the signaling cost per call arrival of the MBLM scheme obtained through simulation with those calculated as per the analytical formulas derived in [2]–[5] and this paper. In Fig. 7, we observe the following: 1) In all the situations, the analytical result of this paper is close to the simulation result. 2) In some situations, there is a pronounced discrepancy between the analytical results obtained in [2]–[5] and the simulation result. 3) With the increase of M, the difference between the signaling costs obtained in [2]–[5] decreases. 4) Referring to Fig. 7(b), when M is relatively large compared with n, as M increases, the signaling costs obtained through simulation and in this paper slightly change, whereas those obtained in [2]–[5] quadratically increase with respect to M. The explanation for the current observation is as follows. When

TABLE III
COMPARISON BETWEEN THE ANALYTICAL AND SIMULATION RESULTS OF THE PA SIZE WHEN $n = 15$ AND $N_{\text{call}} = 1000$

M		1	5	10	15	20	25	29
analytical		1	52.9	195.6	368.4	517.8	607.7	631
simulation	CWLU	1	53.2	196.3	368.6	516.5	607.5	631
	CPLU	1	53.0	195.9	368.8	520.1	607.4	631
relative error (%)		0	0.36	0.27	0.08	0.10	0.03	0

the LA architecture is considered, the PA size is a function of M and n [cf. (12)–(17)]. When M is relatively large, the PA size depends more on the number of cells in an LA $\mathcal{O}(n)$ than on M, such that the PA size does not quadratically increase with respect to M. On the contrary, without respect to the configuration of an LA, the PA size obtained in [2]–[5] solely depends on M and is $\mathcal{O}(M)$, which is a quadratic function of M. As mentioned in [2], this kind of approximation for the PA size is acceptable only when $M \ll n$. When M is relatively large, the LU cost is small, and the paging cost dominates the signaling cost, such that we have the current observation. 5) The call handing model has an impact on the signaling cost.

To demonstrate the accuracy of the PA size obtained in this paper, i.e., (12)–(17), Table III compares the analytical and simulation results of the PA size. In Table III, parameters λ, η, and μ are assigned the same values as those used in Fig. 7. In Table III, the entry *relative error* represents the relative error of the analytical result to the simulation result, which is the average of the simulation results of the CWLU and CPLU models. It follows from Table III that 1) as mentioned in Section III-B, the PA sizes of the CWLU and CPLU models are almost identical and that 2) the analytical and simulation results are very close.

C. Impact of Various Parameters on the Signaling Cost

Here, based on numerical results calculated as per the analytical formulas derived in this paper, we evaluate the impact of various parameters on the signaling cost per call arrival of the MBLM scheme. Figs. 8–10 show the signaling cost versus the movement threshold when n, λ/η, and λ/μ take different values. In Figs. 8–10, we have the following observations.

1) The signaling cost is a downward convex function of M, suggesting that there is an optimal M, which is denoted by M_{opt}, that can minimize the signaling cost. In the case of very small λ/η, e.g., $\lambda/\eta = 0.01$, $M_{\text{opt}} = \infty$. That is, in this case, the static LM scheme has the least signaling cost. The explanation is as follows. When λ/η is very small, it follows from (29) that a large number of cells will be crossed during the call interarrival time, such that the cost due to VLR and HLR LUs is very large. Thus, large M is desirable to curb this cost. By setting $M = \infty$, this cost reaches the minimum level. On the contrary, in the case of very large λ/η, e.g., $\lambda/\eta = 100$, $M_{\text{opt}} = 1$. In this case, the cost due to VLR and HLR LUs is negligible compared with the paging cost, such that small M is desirable to curb the paging cost. By setting $M = 1$, the paging cost reaches the minimum level.
2) Denote by TC_{CWLU} and TC_{CPLU} the signaling costs per call arrival of the CWLU and CPLU models, respectively. Define
$$\begin{cases} \text{diff} \triangleq \text{TC}_{\text{CPLU}} - \text{TC}_{\text{CWLU}} \\ \text{diff}_{\min} \triangleq \min\limits_{M \le 2n+1} \{\text{diff}\} \\ \text{diff}_{\max} \triangleq \max\limits_{M \le 2n+1} \{\text{diff}\}. \end{cases}$$
In Figs. 8–10, the relation $\text{diff}_{\min} > 0$ holds, except in the case of $n = 8$ or 15, $\lambda/\eta = 0.01$, and $\lambda/\mu = 1$ or 100, where $-1.36 \le \text{diff}_{\min} < 0$. In Figs. 8–10, $0 < \text{diff}_{\max} \le 150$. The current observation reveals that in most cases, the CWLU model is superior to the CPLU model, whereas in a few cases, the former is inferior to the latter, but the inferiority is insignificant compared with their signaling costs. Overall, the CWLU model is better than the CPLU model. In the following, we prove that $\text{diff}_{\max}$ is upper bounded by the cost of performing a VLR′ LU, i.e., 150. The two models have identical N_{HLR}'s, $N_{\text{HLR}'}$'s, and N_{paging}'s but different N_{VLR}'s and $N_{\text{VLR}'}$'s. Denote by $N_{\text{VLR,CPLU}}$ and $N_{\text{VLR}',\text{CPLU}}$ the N_{VLR} and $N_{\text{VLR}'}$ of the CPLU model, respectively. In the same way, we define $N_{\text{VLR,CWLU}}$ and $N_{\text{VLR}',\text{CWLU}}$. Note that $N_{\text{VLR}',\text{CPLU}} = 1$. First, we consider a case that under the CWLU model, after a call, no VLR′ LU occurs. In this case, we have $N_{\text{VLR}',\text{CWLU}} = 0$ and $N_{\text{VLR,CWLU}} \ge N_{\text{VLR,CPLU}}$. In the CWLU model, the cost due to VLR and VLR′ LUs is larger than or equal to $\delta_{\text{VLR}} N_{\text{VLR,CPLU}}$, whereas the same cost of the CPLU model is $\delta_{\text{VLR}'} + \delta_{\text{VLR}} N_{\text{VLR,CPLU}}$. It follows that $\text{TC}_{\text{CPLU}} - \text{TC}_{\text{CWLU}} \le \delta_{\text{VLR}'}$. Next, we consider the other case that under the CWLU model, after a call, a VLR′ LU occurs. In this case, the two modes have identical N_{VLR}'s and $N_{\text{VLR}'}$'s and, accordingly, identical signaling costs. Therefore, combining these two cases, it follows that $\text{diff}_{\max}$ is upper bounded by $\delta_{\text{VLR}'}$.
3) In the case of very small λ/η (e.g., $\lambda/\eta = 0.01$) or very large n (e.g., $n = 15$), the difference between the signaling costs of the CPLU and CWLU models is indiscernible. In the case of very small λ/η, the number of cells crossed during the call interarrival time is very large such that the cost due to VLR and HLR LUs is large. In the case of very large n, if M is not too small, then the PA size is large such that the paging cost is large. In the given two cases, the singling cost is large such that the difference between the signaling costs of the two models is negligible compared with their signaling costs.
4) The call holding time t_h exerts influence on the signaling cost. The influence of t_h on the signaling cost of the CPLU model is bounded by the cost of performing an HLR′ LU (i.e., 157), whereas the influence on the signaling cost of the CWLU model is bounded by the cost of performing a VLR′ and an HLR′ LU (i.e., $150 + 157 = 307$). Under the CPLU model, when the expected value of t_h $1/\mu \to 0$, we have $N_{\text{HLR}'} = 0$, and when $1/\mu \to \infty$, we have $N_{\text{HLR}'} = 1$. Therefore, the influence of t_h on the signaling cost of the CPLU model does not exceed the cost of an HLR′ LU. In the same way, the explanation for the CWLU model is straightforward.

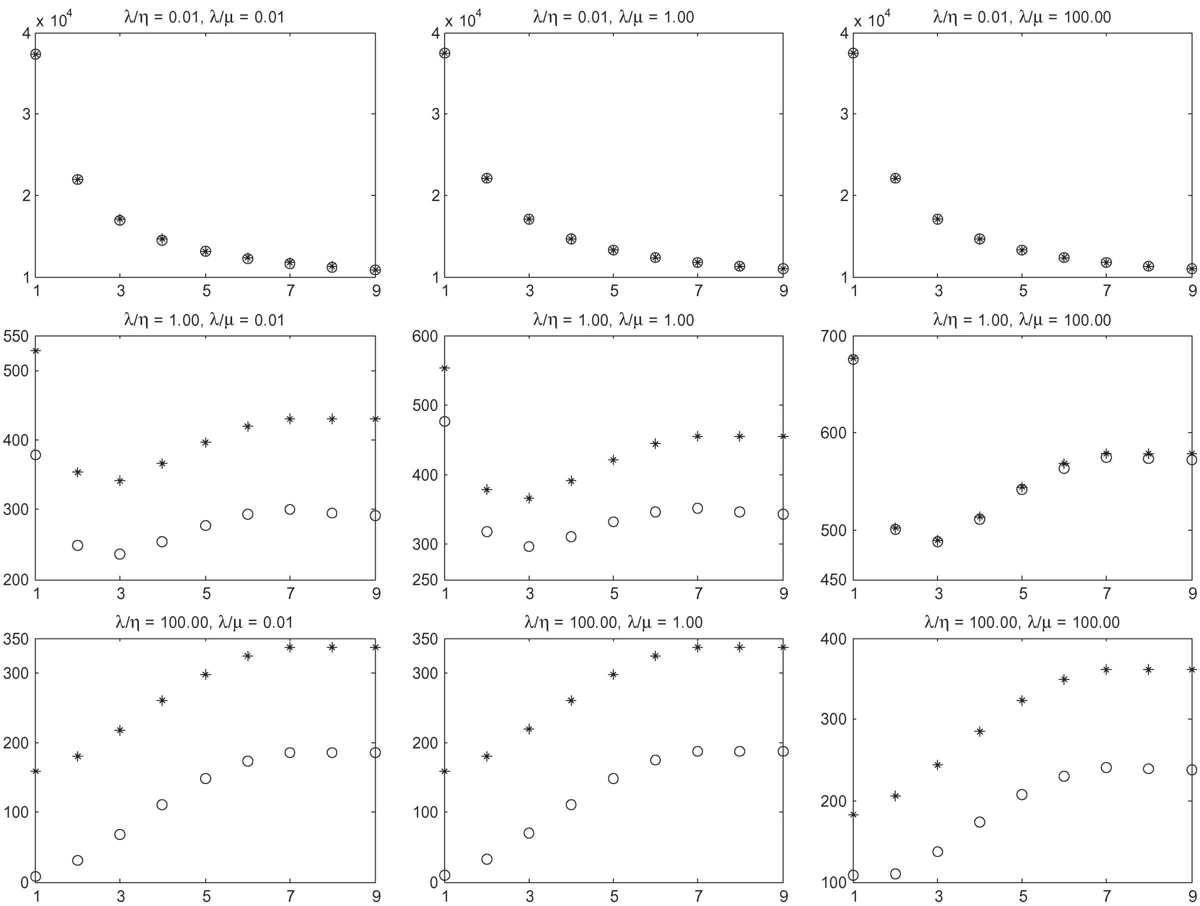

Fig. 8. Signaling cost versus movement threshold. The horizontal and the vertical axis represent the movement threshold and the signaling cost, respectively. (○) CWLU model. (∗) CPLU model. $n = 4$ and $\lambda = 1$.

VI. Conclusion

In this paper, we have developed an embedded Markov chain model to analyze the signaling cost of the MBLM scheme. Compared with the models used in the literature, the developed model is the most comprehensive model that takes into account the LA architecture as well as the CWLU and CPLU call handling models. Moreover, the impact of the call holding time on the CWLU and CPLU models was also taken into account. We employed the fluid flow model to describe the dependence between the cell and the LA residence time. We derived closed-form analytical formulas for the signaling cost, whose accuracy was tested by simulation. Based on the analytical formulas, we conducted a numerical study to evaluate the impact of various parameters on the signaling cost. The observations we observed from the simulation and the numerical study are summarized as follows. 1) There is a discrepancy between the signaling cost obtained through simulation and those calculated as per the analytical formulas derived in [2]–[5]. 2) The signaling cost calculated as per the analytical formulas derived in this paper closely matches the simulation result, manifesting the feasibility of the modeling approach developed in this paper. 3) As observed in the literature, the signaling cost is a downward convex function of the movement threshold. However, when the movement threshold reaches a large value relative to the radius of an LA and continues to increase, the signaling cost changes very slowly. This phenomenon was not observed in the literature. 4) The difference between the signaling costs of the CWLU and CPLU models is bounded by the cost of performing a VLR′ LU. 5) The call holding time exerts influence on the signaling costs of the CWLU and CPLU models, with the influence on the CWLU model being bounded by the cost of performing a VLR′ and an HLR′ LU and the influence on the CPLU model being bounded by the cost of performing an HLR′ LU.

Next, we discuss the operation complexity, signaling cost, and paging delay factors of the MBLM scheme under the CWLU and CPLU call handling models. As to the operation complexity of the MBLM scheme, at the UE side, a movement counter is required to tally the number of crossed cells, and at the network side, the VLR must have the capability to determine the PA for the UE. As mentioned in Section II-D, the cells constituting the PA need to satisfy two criteria. First, they are less than M cells away from the cell where the last LU occurs. Second, they and the cell where the last LU occurs belong to the same LA. Therefore, to determine the PA, the VLR must have the topology information of the cells within the LA it serves. The value of M can be transferred from the VLR to the UE when the UE performs an LU or using LM procedures whenever needed. In the aspect of the signaling cost, overall, the CWLU model is superior to the CPLU model, and the two have the same paging delay. In fact, the CWLU model is the call handling model used by GSM and UMTS [1].

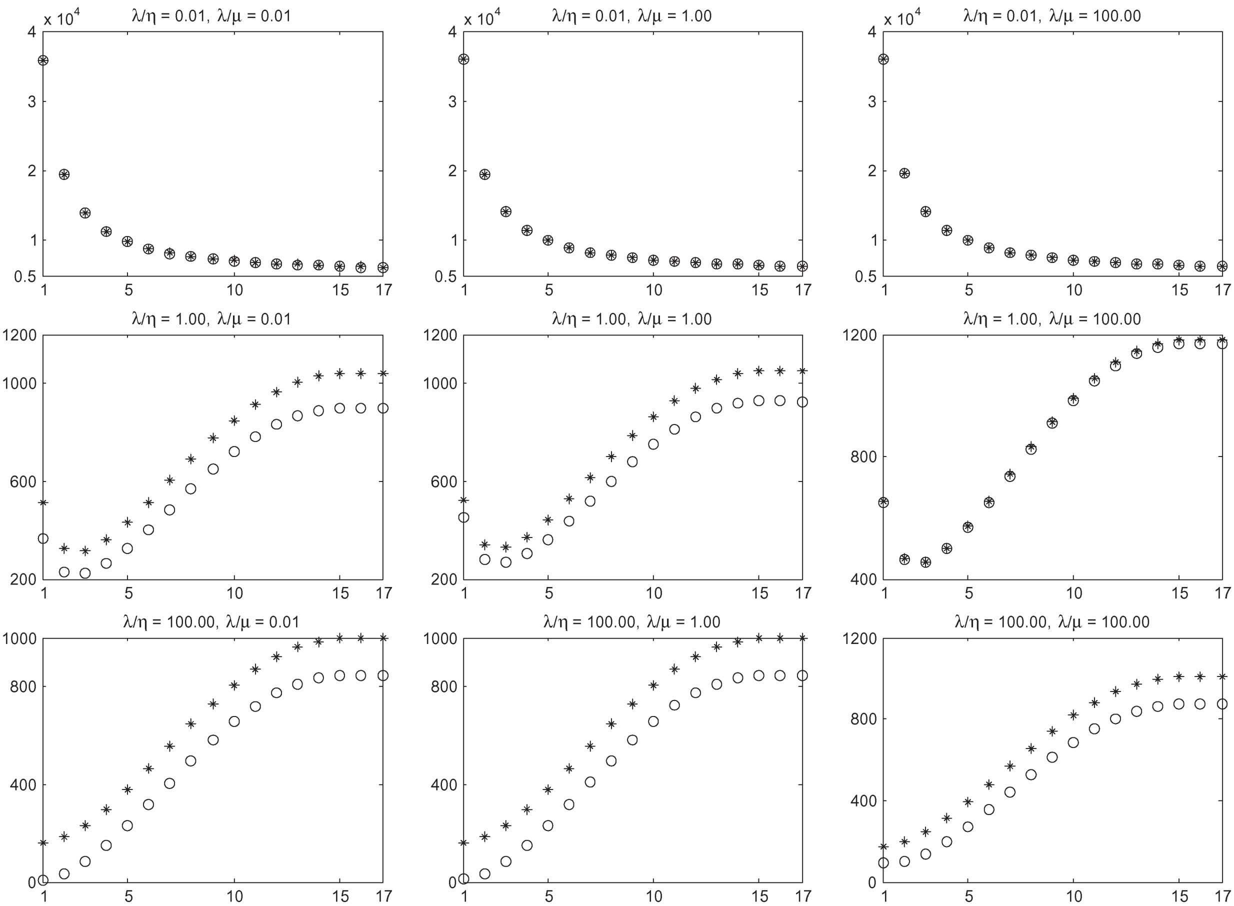

Fig. 9. Signaling cost versus movement threshold. The horizontal and the vertical axis represent the movement threshold and the signaling cost, respectively. (○) CWLU model. (∗) CPLU model. $n = 8$ and $\lambda = 1$.

Finally, we briefly discuss the potential application of the MBLM scheme in LTE. When being applied in existing 2G and 3G PCS networks, the static LM scheme has the following problems: the ping-pong LU effect and uneven distribution of LU signaling. UEs at the boundary of neighboring LAs usually move back and forth between these LAs, resulting in the so-called ping-pong LU effect. Under the static scheme, all the LUs occur in the boundary cells of an LA, such that LU signaling is unevenly distributed among the cells within an LA. If a large number of UEs simultaneously move into a boundary cell of an LA, e.g., all the UE devices on a train moving into a new LA, there is a risk of LU signaling storm in the boundary cell due to a large number of LUs launched in a short period. This is an undesirable case, because it could degrade the quality of service in the boundary cell and even cause signaling congestion. To tackle the problems of the static scheme, LTE adopts a TA list (TAL)-based LM scheme. Under the TAL-based scheme, the UE is assigned a group of TAs, which is referred to as a TAL. Only when the UE moves out of the TAL does it need to perform an LU. Different UEs can be assigned different TALs to evenly distribute LU signaling. Despite the advancement possessed by the TAL-based scheme, it remains an arduous task to design optimal TALs for UEs having diverse mobility and traffic characteristics, because the design process necessitates the information regarding the UE movements among TAs, and the network will pay a high cost to collect this information [44]. The MBLM scheme can help simplify this task. First, try to include more TAs that the UE haunts into the TAL for the UE. This way, we can maximally reduce the LU cost at the price of increasing the PA size, causing unbalance between the LU and the paging cost. Second, implement the MBLM scheme to achieve a balance (i.e., trade-off) between the LU and the paging cost. Treating a TAL as an LA, it is straightforward to implement the MBLM scheme in LTE. Under the MBLM scheme implemented in LTE, a UE performs an LU either when moving out of its current TAL or when reaching the movement threshold. Similarly, the CWLU and CPLU call handling models can be adapted to the TAL-based scheme. Therefore, despite that nearly two decades have elapsed since it was first proposed in 1995 in [19], in LTE, the MBLM scheme still has vitality. In addition to the TAL, the MBLM scheme endows LTE another level of freedom in designing LU strategies for UEs having diverse mobility and traffic characteristics. The analytical formulas derived in this paper are instrumental in implementing the MBLM scheme in PCS networks including LTE. Moreover, the modeling approach developed in this paper can be applied to model other LM schemes, such as the DBLM and TAL-based schemes, in PCS networks having the LA or similar architecture.

APPENDIX A
PROOF OF (20)

The correctness of N_{paging} is indubitable. First, we prove N_{VLR}. A VLR LU is due to the crossing of an LA boundary or

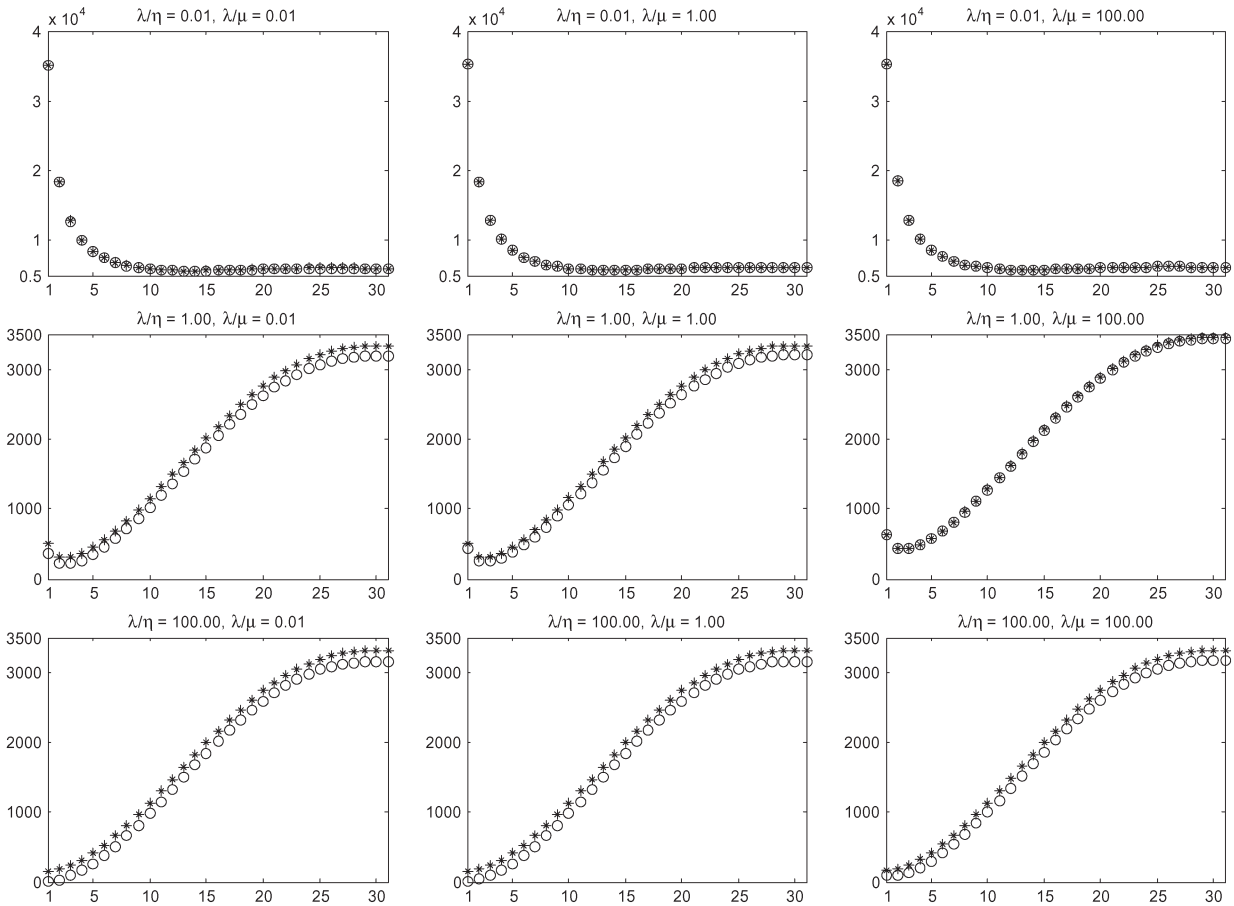

Fig. 10. Signaling cost versus movement threshold. The horizontal and the vertical axis represent the movement threshold and the signaling cost, respectively. (○) CWLU model. (∗) CPLU model. $n = 15$, and $\lambda = 1$.

the reaching of the movement threshold during call interarrival time t_a. Denote by $N(t_c, t_a)$ the number of cells crossed during t_a and by P_{VLR} the probability that a cell crossing leads to a VLR LU. The number of VLR LUs occurring during t_a follows a binomial distribution with parameters $N(t_c, t_a)$ and P_{VLR}. Since t_c is exponential, $N(t_c, t_a)$ can be described by an equilibrium renewal process [45]. It follows that

$$E\left[N(t_c, t_a)\right] = \frac{E(t_a)}{E(t_c)} = \frac{\eta}{\lambda}. \tag{29}$$

If event $\mathrm{E}_{\mathrm{c}\to\mathrm{LA}}$ takes place or if the UE is in state I_{M-1} and event $\mathrm{E}_{\mathrm{c}\to\mathrm{c}}$ takes place, then the UE needs to perform a VLR LU. Since during t_a, the possible states the UE can camp in include states I_j's, $j = 0, 1, \ldots, M$, and state I_{LA}. The probability that during t_a the UE is in state I_{M-1} is

$$\frac{\Pi_{I_{M-1}} E(r_{I_{M-1}})}{\sum_{j=0}^{M} \Pi_{I_j} E(r_{I_j}) + \Pi_{I_{\mathrm{LA}}} E(r_{I_{\mathrm{LA}}})} = \frac{\Pi_{I_{M-1}}}{\sum_{j=0}^{M-1} \Pi_{I_j}} = P(M-1).$$

It follows that

$$\begin{aligned} P_{\mathrm{VLR}} &= \Pr(\mathrm{E}_{\mathrm{c}\to\mathrm{LA}}) + \Pr(\mathrm{E}_{\mathrm{c}\to\mathrm{c}}) P(M-1) \\ &= (1-\xi) + \xi P(M-1). \end{aligned}$$

Therefore, we reach

$$N_{\mathrm{VLR}} = E\left[N(t_c, t_a)\right] P_{\mathrm{VLR}} = \frac{\eta}{\lambda}\left[(1-\xi) + \xi P(M-1)\right]$$

which proves N_{VLR}. In a similar way, we can prove N_{HLR}.

Next, we prove $N_{\mathrm{VLR}'}$. $N_{\mathrm{VLR}'}$ is actually the probability that a VLR$'$ LU occurs after a call. As per the CWLU model, after a call, no VLR$'$ LU occurs if during the call holding time no LA has been crossed and the movement threshold has not been reached. Denote by $t_{c,k}$ the residence time in the kth cell. It follows that

$$\begin{aligned} 1 - N_{\mathrm{VLR}'} &= \sum_{j=0}^{M-1} P(j) \sum_{k=0}^{M-1-j} \Pr\begin{pmatrix} t_h > t_{c,1} + \cdots + t_{c,k}, \\ t_h < t_{c,1} + \cdots + t_{c,k+1} \end{pmatrix} \xi^k \\ &= \sum_{j=0}^{M-1} P(j) \sum_{k=0}^{M-1-j} \beta^k (1-\beta) \xi^k. \end{aligned}$$

It follows from the last equation and (11) that

$$N_{\mathrm{VLR}'} = \sigma + \frac{(1-\sigma)\eta\xi^M(1-\Delta^M)}{\lambda\Delta S(M, \Delta\xi) + \mu S(M, \xi)}$$

which proves $N_{\mathrm{VLR}'}$. After a call, an HLR$'$ LU occurs if at least one LA has been crossed during the call holding time. It follows that

$$N_{\mathrm{HLR}'} = \Pr(t_h > t_{\mathrm{LA}}) = \sigma$$

which proves $N_{\mathrm{HLR}'}$.

APPENDIX B
DISCUSSIONS ON THE N_{VLR} GIVEN IN (23)

- When $\xi = 1$, i.e., when there is no LA architecture, the N_{VLR} given in (23) reduces to

$$N_{\mathrm{VLR}} = \frac{\alpha^M}{1-\alpha^M}$$

which is the result obtained in [20, eqs. (2) and (9)], [21, eqs. (12) and (14)], and [22, Th. 6], [23], [24], and [27].

- As was done in this paper, the studies in [2]–[5] considered the LA architecture. Here, we use a special case where an LA contains only one cell to show that in some situations, the results derived in [2]–[5] may differ from the accurate result. In this case, during the call interarrival time whenever moving out of a cell, the UE performs a VLR and an HLR LU. It follows from (29) that in this case

$$N_{\mathrm{VLR}} = \frac{\eta}{\lambda}. \tag{30}$$

In the current case, $\tau = \eta$, and thus, it follows from (5) that $\xi = 0$. It follows from (26) that in this case, the N_{VLR} derived in [2] by Li *et al.* becomes

$$N_{\mathrm{VLR}} = \frac{1}{2^M-1}\frac{\eta}{\lambda} + \frac{\alpha^M}{1-\alpha^M} + \frac{\eta}{\lambda}. \tag{31}$$

It follows from (27) that in this case, the N_{VLR} derived in [3] and [4] by Rodriguez-Dagnino and Takagi becomes

$$N_{\mathrm{VLR}} = (1-\alpha)\frac{1}{(2+\lambda/\eta)^M-1} + \alpha\frac{1}{2^M-1}\left(2+\frac{\eta}{\lambda}\right) + \frac{\eta}{\lambda}. \tag{32}$$

It follows from (28) that in this case, the N_{VLR} derived in [5] by Wang *et al.* becomes

$$N_{\mathrm{VLR}} = \frac{1}{(2+\lambda/\eta)^M-1}\left(1+\frac{\eta}{\lambda}\right) + \frac{\eta}{\lambda}. \tag{33}$$

Equations (31)–(33) all contradict (30).

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers, whose comments substantially enhanced the quality of this paper.

REFERENCES

[1] Y.-B. Lin and A.-C. Pang, *Wireless and Mobile All-IP Networks*. New York, NY, USA: Wiley, 2005.

[2] J. Li, Y. Pan, and X. Jia, "Analysis of dynamic location management for PCS networks," *IEEE Trans. Veh. Technol.*, vol. 51, no. 5, pp. 1109–1119, Sep. 2002.

[3] R. M. Rodriguez-Dagnino and H. Takagi, "Movement-based location management for general cell residence times in wireless networks," *IEEE Trans. Veh. Technol.*, vol. 56, no. 5, pp. 2713–2722, Sep. 2007.

[4] R. M. Rodriguez-Dagnino and H. Takagi, "Application of renewal theory to call handover counting and dynamic location management in cellular mobile networks," *Eur. J. Oper. Res.*, vol. 204, no. 1, pp. 1–13, Jul. 2010.

[5] X. Wang, P. Fan, J. Li, and Y. Pan, "Modeling and cost analysis of movement-based location management for PCS networks with HLR/VLR architecture and general location area and cell residence time distributions," *IEEE Trans. Veh. Technol.*, vol. 57, no. 6, pp. 3815–3831, Nov. 2008.

[6] R. Jain and Y.-B. Lin, "An auxiliary user location strategy employing forwarding pointers to reduce network impacts of PCS," *Wireless Netw.*, vol. 1, no. 2, pp. 197–210, 1995.

[7] W. Ma and Y. Fang, "Two-level pointer forwarding strategy for location management in PCS networks," *IEEE Trans. Mobile Comput.*, vol. 1, no. 1, pp. 32–45, Jan.–Mar. 2002.

[8] J. S. M. Ho and I. F. Akyildiz, "Local anchor scheme for reducing signaling costs in personal communications networks," *IEEE/ACM Trans. Netw.*, vol. 4, no. 5, pp. 709–725, Oct. 1996.

[9] Y.-B. Lin, "Reducing location update cost in a PCS network," *IEEE/ACM Trans. Netw.*, vol. 5, no. 1, pp. 25–33, Feb. 1997.

[10] Y. Xiao, Y. Pan, and J. Li, "Design and analysis of location management for 3G cellular networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 15, no. 4, pp. 339–349, Apr. 2004.

[11] "Cellular radio-telecommunications intersystem operations," EIA/TIA, Dallas, TX, USA, EIA/TIA Tech. Rep. IS-41, Rev. B, 1991.

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

[13] T. Imielinski and B. R. Badrinath, "Querying in highly mobile distributed environments," in *Proc. 18th VLDB Conf.*, Vancouver, BC, Canada, 1992, pp. 41–52.

[14] S. Tabbane, "Comparison between the alternate location strategy and the classic location strategy," WINLAB, Rutgers Univ., Piscataway, NJ, USA, Tech. Rep. WINLAB-TR-37, 1992.

[15] G. P. Pollini and S. Tabbane, "The intelligent network signaling and switching costs of an alternate location strategy using memory," *IEEE Trans. Veh. Technol.*, vol. 4, no. 5, pp. 709–725, Oct. 1996.

[16] H. Xie, S. Tabbane, and J. D. Goodman, "Dynamic location area management and performance analysis," in *Proc. 43rd IEEE Veh. Technol. Conf.*, Secaucus, NJ, USA, 1993, pp. 536–539.

[17] Y. Fang, "General modeling and performance analysis for location management in wireless mobile networks," *IEEE Trans. Comput.*, vol. 51, no. 10, pp. 1169–1181, Oct. 2002.

[18] V. C. Giner, "On lookahead strategy for movement-based location update: A general formulation," in *Proc. PERFORM*, 2011, pp. 153–166.

[19] A. Bar-Noy, I. Kessler, and M. Sidi, "Mobile users: To update or not to update?" *Wireless Netw.*, vol. 1, no. 2, pp. 175–185, 1995.

[20] I. F. Akyildiz, J. S. M. Ho, and Y.-B. Lin, "Movement-based location update and selective paging for PCS networks," *IEEE/ACM Trans. Netw.*, vol. 4, no. 4, pp. 629–638, Aug. 1996.

[21] Y. Fang, "Movement-based mobility management and trade off analysis for wireless mobile networks," *IEEE Trans. Comput.*, vol. 52, no. 6, pp. 791–803, Jun. 2003.

[22] K. Li, "Cost analysis and minimization of movement-based location management schemes in wireless communication networks: A renewal process approach," *Wireless Netw.*, vol. 17, no. 4, pp. 1031–1053, May 2011.

[23] J. Li, H. Kameda, and K. Li, "Optimal dynamic mobility management for PCS networks," *IEEE/ACM Trans. Netw.*, vol. 8, no. 3, pp. 319–327, Jun. 2000.

[24] R. M. Rodriguez-Dagnino, J. J. Ruiz-Cedillo, and H. Takagi, "Dynamic mobility management for cellular networks: A delayed renewal process approach," *IEICE Trans. Commun.*, vol. E85-B, no. 6, pp. 1069–1074, Jun. 2002.

[25] Y. Zhang, J. Zheng, L. Zhang, Y. Chen, and M. Ma, "Modeling location management in wireless networks with generally distributed parameters," *Comput. Commun.*, vol. 29, no. 12, pp. 2386–2395, Aug. 2006.

[26] Y. Zhu and V. C. M. Leung, "Optimization of sequential paging in movement-based location management based on movement statistics," *IEEE Trans. Veh. Technol.*, vol. 56, no. 2, pp. 955–964, Mar. 2007.

[27] Y. Zhu and V. C. M. Leung, "Joint distribution of numbers of location updates and cell boundary crossings in movement-based location management schemes," *IEEE Commun. Lett.*, vol. 11, no. 12, pp. 943–945, Dec. 2007.

[28] J. H. Baek, J. Y. Seo, and D. C. Sicker, "Performance analysis and optimization of an improved dynamic movement-based location update scheme in mobile cellular networks," in *Proc. ICCSA*, 2005, vol. 4, pp. 528–537.

[29] Y. W. Chung, "Performance analysis of movement-based group location management for mobile stations riding on transportation systems in cellular systems," in *Proc. SAINT*, 2011, pp. 333–337.

[30] Y. Xiao, "Optimal fractional movement-based scheme for PCS location management," *IEEE Commun. Lett.*, vol. 7, no. 2, pp. 67–69, Feb. 2003.

[31] I. F. Akyildiz, Y.-B. Lin, W.-R. Lai, and R.-J. Chen, "A new random walk model for PCS networks," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 7, pp. 1254–1260, Jul. 2000.

[32] K. Li, "Analysis of distance-based location management in wireless communication networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 24, no. 2, pp. 225–238, Feb. 2013.

[33] V. C. Giner, V. Pla, and P. Escalle-Garcia, "Mobility models for mobility management," in *Network Performance Engineering*, D. Kouvatsos, Ed. Berlin, Germany: Springer-Verlag, 2011, ser. Next Generation Internet, pp. 716–745.

[34] G. Morales-Andres and M. Villen-Altamirano, “An approach to modeling subscribers mobility in cellular radio networks,” in *Proc. Forum Telecom*, Geneva, Switzerland, Dec. 1987, pp. 185–189.

[35] R. Thomas, H. Gilbert, and G. Mazziotto, “Influence of the movement of the mobile station on the performance of the radio cellular network,” in *Proc. 3rd Nordic Sem. Digital Land Mobile Radio Commun.*, Copenhagen, Denmark, Sep. 1988, pp. 9–14.

[36] C. Makaya and S. Pierre, “An analytical framework for performance evaluation of IPv6-based mobility management protocols,” *IEEE Trans. Wireless Commun.*, vol. 7, no. 3, pp. 972–983, Mar. 2008.

[37] F. V. Baumann and I. G. Niemegeers, “An evaluation of location management procedures,” in *Proc. ICUPC*, San Diego, CA, USA, 1994, pp. 359–364.

[38] K. S. Meier-Hellstern, E. Alonso, and D. R. O’Neil, “The use of SS7 and GSM to support high density personal communications,” in *Proc. ICC*, Chicago, IL, USA, Jun. 1992, pp. 1698–1702.

[39] S. Mohan and R. Jain, “Two user location strategies for personal communications services,” *IEEE Pers. Commun.*, vol. 1, no. 1, pp. 42–50, Apr.–Jun. 1994.

[40] C. N. Lo, R. S. Wolff, and R. C. Bernhardt, “An estimate of network database transaction volume to support universal personal communications services,” in *Proc. 1st. Int. Conf. Universal Pers. Commun.*, Dallas, TX, USA, Sep. 1992, pp. 09.03-1–09.03-6.

[41] I. F. Akyildiz and W. Wang, “A dynamic location management scheme for next-generation multitier PCS systems,” *IEEE Trans. Wireless Commun.*, vol. 1, no. 1, pp. 178–189, Jan. 2002.

[42] Y.-B. Lin and I. Chlamtac, *Wireless and Mobile Network Architectures*. New York, NY, USA: Wiley, 2000.

[43] [Online]. Available: http://www.eventhelix.com/realtimemantra/telecom/GSM_Location_Update_Sequence_Diagram.pdf

[44] S. M. Razavi, “Tracking area planning in cellular networks,” Ph.D. dissertation, Linkoping University, Linkoping, Sweden, 2011.

[45] R. M. Rodriguez-Dagnino and H. Takagi, “Counting handovers in a cellular mobile communication network: Equilibrium renewal process approach,” *Performance Eval.*, vol. 52, no. 2/3, pp. 153–174, Apr. 2003.

Xian Wang (M’09) received the B.Eng. and Ph.D. degrees in communication and information systems from Southwest Jiaotong University (SWJTU), Chengdu, China, in 2002 and 2008, respectively.

He is currently an Associate Professor with the School of Information Science and Technology, SWJTU. His research interests include mobility management and performance modeling for wireless mobile networks.

Xianfu Lei (M’13) was born in Lian Yun Gang, China, in December 1981. He received the Ph.D. degree in communication and information systems from Southwest Jiaotong University (SWJTU), Chengdu, China.

Since December 2012, he has been a Research Fellow with the Department of Electrical and Computer Engineering, Utah State University, Logan, UT, USA. He was with the Key Laboratory of Information Coding and Transmission, SWJTU. He has published over 35 technical papers. His current research interests include cooperative communications, cognitive radio, physical-layer security, and heterogeneous networks.

Dr. Lei is the Technical Program Committee (TPC) Cochair of the 10th International Wireless Communications and Mobile Computing Conference (IWCMC 2014) Workshop on Green Communications and Networking and the Tutorial Chair for the International Conference on Computing, Management, and Telecommunications (ComManTel’14). He has been a TPC member of the IEEE Global Communications Conference (GLOBECOM) from 2011 to 2013; the IEEE International Conference on Communications (ICC) from 2011 to 2014; the IEEE Wireless Communications and Networking Conference (WCNC) from 2011 to 2014; the IEEE Vehicular Technology Conference (VTC) from 2011 to 2013; and the IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC) from 2011 to 2013. He is an active reviewer for various IEEE journals and major conferences. He is a member of the IEEE Communications Society and the IEEE Green ICT Community.

Pingzhi Fan (S’94–M’95–SM’99) received the M.S. degree in computer science from Southwest Jiaotong University (SWJTU), Chengdu, China, in 1987 and the Ph.D. degree in electronic engineering from Hull University, Yorkshire, U.K., in 1994.

He is currently a Professor and the Director of the Institute of Mobile Communications with SWJTU. He is also the Chief Scientist of the China 973 program and has been a Guest Professor with Leeds University, Leeds, U.K., since 1997 and a Guest Professor with SWJTU since 1999. He is the inventor of 22 patents and the author of over 350 research papers and eight books, including six books published by John Wiley & Sons Ltd., RSP (1996), IEEE Press (2003 and 2006), Springer (2004), and Nova Science (2007). His research interests include high-mobility wireless communications, spread-spectrum and code-division multiple-access techniques, and information theory and coding.

Dr. Fan received the U.K. Overseas Research Students Award in 1992 and the National Natural Science Foundation of China Outstanding Young Scientist Award in 1998.

Rose Qingyang Hu (S’95–M’98–SM’06) received the B.S. degree in electrical engineering from the University of Science and Technology of China, Hefei, China; the M.S. degree in mechanical engineering from the Polytechnic Institute of New York University, Brooklyn, NY, USA; and the Ph.D. degree in electrical engineering from the University of Kansas, Lawrence, KS, USA.

From January 2002 to June 2004, she was an Assistant Professor with the Department of Electrical and Computer Engineering, Mississippi State University, Mississippi State, MS, USA. She has more than ten years of R&D experience with Nortel, RIM, and Intel as a Technical Manager, a Senior Wireless System Architect, and a Senior Research Scientist. She is currently an Associate Professor with the Department of Electrical and Computer Engineering, Utah State University, Logan, UT, USA. She has extensively published and holds numerous patents in her research areas. Her current research interests include next-generation wireless communications, wireless network design and optimization, green radios, multimedia quality of service/quality of experience, communication and information security, wireless system modeling, and performance analysis.

Dr. Hu is currently serving on the editorial board of the IEEE WIRELESS COMMUNICATIONS MAGAZINE, the *Security and Communication Networks Journal*, *Wireless Communications and Mobile Computing*, and the *KSII Transactions on Internet and Information Systems Journals*. She is a five-time Guest Editor for the IEEE COMMUNICATIONS MAGAZINE, the IEEE WIRELESS COMMUNICATIONS MAGAZINE, and the IEEE NETWORK MAGAZINE. She served as the Technical Program Committee (TPC) Cochair for the International Conference on Computing, Networking, and Communication (ICNC 2014); the TPC Vice Chair for the IEEE International Conference on Green Computing and Communications (GreenCom 2013) and the IEEE/IFIP International Conference on Embedded and Ubiquitous Computing; the Workshop Cochair for the International Conference on Communications and Networking in China (Chinacom 2013); and the Symposium Cochair for the IEEE International Conference on Communications (ICC 2012/2014), the IEEE Wireless Communications and Mobile Computing (WCNC 2013), the Ninth International Wireless Communications and Mobile Computing Conference (IWCMC 2013), the International Conference on Computing, Networking, and Communications (ICNC 2013), and the IEEE International Conference on Smart Grid Communications (SmartGridComm 2012). She is a member of the Phi Kappa Phi and Epsilon Pi Epsilon Honor Societies. She received the Best Paper Award from the IEEE Global Communications Conference (GLOBECOM 2012). She has been also listed multiple times in the *World Who’s Who, Marquis Who’s Who*, and *Cambridge Who’s Who*.

Shi-Jinn Horng received the B.S. degree in electronics engineering from the National Taiwan Institute of Technology, Taipei, Taiwan, in 1980; the M.S. degree in information engineering from National Central University, Zhongli, Taiwan, in 1984; and the Ph.D. degree in computer science from National Tsing Hua University, Hsinchu, Taiwan, in 1989.

He is currently a Professor and the Director of the Department of Computer Science and Information Engineering with National Taiwan University of Science and Technology, Taipei. His research interests include very-large-scale integration design, wireless communication, security, and parallel algorithms.