Predictive Distance-Based Mobility Management for Multi-Dimensional PCS Networks

Accepted for publication in the *ACM/IEEE Transactions on Networking*

Ben Liang, *member*, IEEE, and Zygmunt J. Haas, *Senior Member*, IEEE

Abstract— **This paper presents a mobile tracking scheme that exploits the predictability of user mobility patterns in wireless PCS networks. In this scheme, a mobile's future location is predicted by the network, based on the information gathered from the mobile's recent report of location and velocity. When a call is made, the network pages the destination mobile around the predicted location. A mobile makes the same location prediction as the network does; it inspects its own location periodically and reports the new location when the distance between the predicted and the actual locations exceeds a threshold. Due to the use of prediction, the mobility management cost of our scheme is significantly reduced, as compared with the non-predictive schemes. To more realistically represent the various degrees of velocity correlation in time, a Gauss-Markov mobility model is used. For practical systems where the mobility pattern varies over time, we propose a dynamic Gauss-Markov parameter estimator that provides the mobility parameters to the prediction algorithm.**

Based on the Gauss-Markov model, we describe an analytical framework to evaluate the cost of mobility management for the proposed scheme. We also present an approximation method that reduces the computational complexity of the cost evaluation for multi-dimensional systems. We then compare the cost of predictive mobility management against that of the regular, non-predictive distance-based scheme, for both the case with ideal Gauss-Markov mobility pattern and the case with time-varying mobility pattern.

The performance advantage of the proposed scheme is demonstrated under various mobility patterns, call patterns, location inspection cost, location updating cost, mobile paging cost, and frequencies of mobile location inspections. As a point of reference, prediction can reduce the mobility management cost by more than 50% for all systems, where a the mobile users have moderate mean velocity and where performing a single location update is as least as expensive as paging a mobile in one cell.

I. Introduction

In the operation of wireless personal communication service (PCS) networks, mobility management deals with the tracking, storage, maintenance, and retrieval of mobile location information. Two commonly used standards, the EIA/TIA Interim Standard 41 in North America [1] and the Global System for Mobile Communications in Europe [22], partition their coverage areas into a number of location areas(LA), each consisting of a group of cells. When a mobile enters an LA, it reports to the network the information about its current new location (*location update*). When an incoming call arrives, the network simultaneously pages the mobile (*terminal paging*) in all cells within the LA, where the mobile currently resides. In these standards, the LA coverage is fixed for all users and cannot adapt to different and differing user traffic and mobility patterns.

Dynamic LA management schemes have been proposed in [15][18][32], where an LA is defined per mobile, as to adapt to its traffic and mobility patterns. However, when a call arrives, the system still needs to page the mobile in all cells within the LA, operation that is costly in most cases.

Dynamic mobility management schemes [2][3][5][8][14][19][21] discard the notion of LA borders altogether. A mobile in these schemes updates its location[1] based on either elapsed time, number of crossed cell borders, or traveled distance. All these parameters can be dynamically adapted to each mobile's traffic and mobility patterns, hence providing better cost-effectiveness than the LA scheme. In schemes that do not rely on LAs, upon call arrival, the network pages the destination mobile using a selective paging scheme [24], starting from the cell location where the mobile last updated and outwards, in a shortest-distance-first order. With the assumption of a random-walk mobility model, this paging scheme is equivalent to the highest-probability-first algorithm, which, indeed, incurs the minimum paging cost. In particular, in the distance-based scheme, a mobile performs location update whenever it is some threshold distance away from the location where it last updated. For a system with the memoryless random-walk mobility pattern, the distance-based scheme has been proven to result in less location updating and paging costs than schemes based on time or number of cell boundary crossings [5].

However, in systems where a user's future velocity is correlated with its past and current velocity (i.e., non-memoryless), the highest-probability location of a mobile is generally not the cell where the mobile last reported. Thus, predicting the mobile's current location when a call arrives would result in less paging traffic. Consequently, a mobility management scheme that takes advantage of the predictability of the mobiles' location can lower the mobility management cost.

In our proposed predictive distance-based mobility management scheme, the future location of a mobile is predicted based on the probability density function of the mobile's location, which is given by a Gauss-Markov mobility model based on its location and velocity at the time of the recent location updates. The prediction information is made available to both, the network and the mobiles. Therefore, a mobile is aware of the network's prediction of its location. The mobile checks its position periodically (*location inspection*) and performs location update whenever it reaches some threshold distance (*updating distance*) away from the predicted location. To locate a mobile,

B. Liang is with the Department of Electrical and Computer Engineering, University of Toronto, Ontario, Canada. Z. J. Haas is with the School of Electrical and Computer Engineering, Cornell University, New York, USA. This work was supported in part by the NSF grant number NCR-9704404.

[1] We do not address in this paper the exact mechanism by which a mobile monitors its location and velocity. A mobile may determine its location through a variety of methods, including the Global Positioning System, signal triangulation, base-station self-identifying beacons, or a combination of the above. It may then average position displacement over time to find its velocity. Other methods and related references on mobile location and velocity determination can be found in [13].

the network pages the mobile starting from the predicted location and outwards, in a shortest-distance-first order, and until the mobile is found.

Besides the application of prediction, another feature that sets our work apart from the original investigation of distance-based mobility management is the notion of sporadic location inspections. In a distance-based updating scheme, a mobile needs to monitor its location and/or velocity. In practical systems, this cannot be done continuously due to bandwidth and computation power limitations. We assume that a mobile measures its location and velocity periodically, at the location inspection instants. We further study how the optimal location inspection period is affected by the mean, variance, and memory level of a mobile's velocity. Previous work on distance-based mobility management schemes [2][5][14][21] do not incorporate this assumption.

Summaries of many aforementioned mobility management schemes can be found in the survey papers [4] and [30]. A preliminary version of the predictive distance-based mobility management scheme, based on a one-dimensional system and without considering the cost of location inspections, is presented in [20] and summarized in [30]. Other related work include the combination of timer, movement, and distance approaches into a general state-based framework in [26], and an information-theoretic approach in [6]. Most recently, a cell-shape independent stochastic model has been proposed in [31], which provides an iterative algorithm to compute the optimal updating threshold for distance-based mobility management.

The rest of this paper is organized as follows. In Section II, we describe the Gauss-Markov mobility model and the prediction algorithm. We demonstrate the broad applicability of the Gauss-Markov model, derive the optimal location prediction, and introduce an algorithm to dynamically estimate the Gauss-Markov parameters when the mobile movement pattern is time-varying. Section III presents the analytical framework for evaluating the cost of the predictive distance-based scheme. This cost is defined as the sum of a mobile's location inspection cost, location updating cost, and the cost incurred in paging the mobile. The numerical results are presented in Section IV. We study how the optimal location inspection period and updating distance are affected by the traffic pattern, the mobility pattern, and the relative costs of location inspection, location updating, and paging. To evaluate the performance gain of the predictive scheme, the cost of the non-predictive distance-based scheme is measured through simulations, and is compared with the predictive scheme. Our results indicate that, in general, prediction can significantly lower the mobility management cost. We show that the proposed scheme performs well, even when used in conjunction with a suboptimal parameter estimation algorithm, in systems with time-varying mobility pattern. Furthermore, we explain why the performance gain of the predictive scheme does not necessarily increase as the memory level of a mobile's movement increases. Finally, the concluding remarks are given in Section V.

II. System Description

A. *The Gauss-Markov Mobility Model*

A mobile user usually travels with a destination in mind. Furthermore, the change of mobile's velocity within a short time is limited due to physical restrictions. Therefore, a mobile user's future location and velocity are likely to be correlated with its past and current location and velocity. Thus, the memoryless nature of the random-walk model makes it unsuitable to represent such behavior. Another widely used mobility model in cellular network analysis is the fluid-flow model [32][33]. The fluid-flow model is suitable for vehicle traffic on highways, but not pedestrian movements with frequent "stop-and-go" behavior. A discrete Markovian model is reported in [5]. However, in this model, the velocity of the mobiles is overly simplified and characterized by three states only. A memoryless Brownian motion with drift model is used in [25] and [8]. It is an improvement to the random-walk model, but still cannot represent the time correlation in mobile's velocity. In this work, we introduce a mobility model based on the Gauss-Markov process [11][23][29]. The Gauss-Markov process has appeared in diverse fields, such as the theory and applications of signal estimation and economic forecast. We postulate that it can be used to model the mobile movement in a PCS network, since it captures the essence of the correlation of a mobile's velocity in time. Furthermore, the Gauss-Markov model represents a wide range of user mobility patterns, including, as its two extreme cases, the random-walk and the constant velocity fluid-flow models.

A.1 The One-Dimensional Case

For simplicity, we first illustrate the Gauss-Markov mobility model in a one-dimensional system. In this model, a mobile's velocity is assumed to be correlated in time and modeled by a Gauss-Markov random process. In continuous-time, a stationary Gaussian process $v(t)$ is a Gauss-Markov process if it has the autocorrelation function [11][23][29]

$$R_v(\tau) = \mathrm{E}[v(t)v(t+\tau)] = \sigma^2 e^{-\beta|\tau|} + \mu^2 , \tag{1}$$

where σ^2 is the variance of $v(t)$, μ is its mean, and $\beta \geq 0$ determines the degree of memory. When $\mu = 0$, equation (1) is also sometimes called the Ornstein-Uhlenbeck solution of the Brownian motion with zero restoring force [23].

We define a discrete version of the mobile velocity with $v_n = v(n\Delta t)$, and the memory level $\alpha = e^{-\beta\Delta t}$, where Δt is normalized to unity throughout this paper.[2] Then the discrete Gauss-Markov process described by (1) can be represented by the following recursive realization:

$$v_n = \alpha v_{n-1} + (1-\alpha)\mu + \sigma\sqrt{1-\alpha^2}\,w_{n-1} , \tag{2}$$

where $\{w_n\}$ is an uncorrelated Gaussian process with zero mean and unit variance and is independent of $\{v_n\}$.

If the initial velocity v_0 is Gaussian with mean μ and variance σ^2, from equation (2), we clearly have $R_v(k) = \sigma^2 e^{-\beta k} + \mu^2$.

[2] In a practical system design, the value of Δt should match the parameters of each individual system. It should not be too small, such that the model identification and system optimization become intractably complex. It also should not be too large, such that the details of a user's movement is lost. However, an investigation into the optimal selection of Δt is outside the scope of this paper.

Therefore, equation (2) satisfies the definition (1). However, since there is no guarantee that v_0 has mean μ and variance σ^2, or even that v_0 is Gaussian, the process described by (2) is generally not stationary. In this case, μ is the asymptotic mean of v_n, and σ is the asymptotic standard deviation of v_n, when n approaches infinity.

Clearly, as α approaches zero, or β approaches infinity, (2) represents a drifting random-walk mobility pattern with mean μ and standard deviation σ, and that as α approaches one, or β approaches zero, (2) represents a constant velocity fluid-flow mobility pattern with $v_n = v_0$ for all n. Therefore, not only does the Gauss-Markov model represents a wide spectrum of mobility patterns with various degree of memory, it also includes the random-walk and fluid-flow models as its two extreme cases.

A.2 The Multi-dimensional Case

It is straight forward to extend the Gauss-Markov mobility model to the two or three-dimensional systems by using vectors of Gauss-Markov processes. In what follows, we use a two-dimensional system to illustrate how this is done.

In the 2-D case, the location and velocity of a mobile is represented by the random vectors $\mathbf{s}_n = [s_n^x, s_n^y]^T$ and $\mathbf{v}_n = [v_n^x, v_n^y]^T$, respectively, where the superscripts denote the dimensions. Similarly, we define $\bar{\alpha} = [\alpha^x, \alpha^y]^T$, $\bar{\mu} = [\mu^x, \mu^y]^T$, and $\bar{\sigma} = [\sigma^x, \sigma^y]^T$. Then, the 2-D velocity process can be expressed as follows:

$$\mathbf{v}_n = \bar{\alpha} \odot \mathbf{v}_{n-1} + (1 - \bar{\alpha}) \odot \bar{\mu} + \bar{\sigma} \odot \sqrt{1 - \bar{\alpha}^2} \odot \mathbf{w}_{n-1} \ , \tag{3}$$

where $\odot$ denotes element-by-element multiplication, and $\{\mathbf{w}_n\} = [\{w_n^x\}, \{w_n^y\}]^T$ is a 2-D uncorrelated Gaussian process with zero mean and unit variance and is independent of $\{\mathbf{v}_n\}$. In (3), we have made an important assumption that the velocity processes in both dimensions are uncorrelated. It is easy to rewrite (3) to represent the more general case of correlated dimensions. However, as will become obvious in Section III, the cost analysis with even the simpler, uncorrelated model can be highly computationally intensive. The additional dimension of conditional probabilities in the correlated model would render the model practically intractable.

For simplicity of presentation, one may further assume that the velocity has the same memory level, α, the same asymptotic mean, μ, and the same asymptotic standard deviation, σ, in both dimensions. In this isotropic case, equation (3) becomes

$$\mathbf{v}_n = \alpha \mathbf{v}_{n-1} + (1 - \alpha)\mu + \sigma\sqrt{1 - \alpha^2}\mathbf{w}_{n-1} \ . \tag{4}$$

A.3 Duplicating Other Mobility Models

To our knowledge, no real data are available in the public domain that describe the user mobility pattern in the desired granularity to validate a mobility model. For example, mobile traces reported in [27] and [28] only give coarse movement patterns between large areas, in hourly to daily time-frames, and aggregated among many users. Therefore, none of the commonly used mobility models in literature can be validated against real user data.

Nevertheless, the wide applicability of the proposed Gauss-Markov model can still be demonstrated. We illustrate how the parameter set (α, μ, σ) can be tuned such that the Gauss-Markov model can duplicate the mobile movements described by three of the most popular mobility models: the random-walk model [25][8], the fluid-flow model [32], and the random way-point model [7][9].[3] By showing this, we demonstrate that the Gauss-Markov model is, indeed, a mobility model that can represent a variety of mobility models.

1) The random-walk model (with drift) has no memory. It can be represented by

$$\mathbf{v}_n^{(RW)} = \bar{\mu}^{(RW)} + \bar{\sigma}^{(RW)} \odot \mathbf{w}_n^{(RW)} \ , \tag{5}$$

where $\bar{\mu}^{(RW)}$ is a vector of drift velocities, $\bar{\sigma}^{(RW)}$ is a vector of velocity standard deviations, and $\{\mathbf{w}_n^{(RW)}\}$ is a vector of uncorrelated Gaussian processes with zero mean and unit variance. Comparing this with (3), the corresponding sub-family of the Gauss-Markov model has the parameters $\{\bar{\alpha} = \mathbf{0}, \bar{\mu} = \bar{\mu}^{(RW)}, \bar{\sigma} = \bar{\sigma}^{(RW)}\}$.

2) The fluid-flow model has constant velocity in all dimensions. It can be represented by

$$\mathbf{v}_n^{(FF)} = \mathbf{c}^{(FF)} \ , \tag{6}$$

where $\mathbf{c}^{(FF)}$ is a vector of constant velocities. Comparing this with (3), there are several sub-families of the Gauss-Markov model that generate movement patterns the same as those described by the fluid-flow model. They are $\{\bar{\alpha} = \mathbf{1}, \mathbf{v}_0 = \mathbf{c}^{(FF)}\}$, $\{\bar{\alpha} = \mathbf{0}, \bar{\mu} = \mathbf{c}^{(FF)}, \bar{\sigma} = \mathbf{0}\}$, and $\{\bar{\mu} = \mathbf{c}^{(FF)}, \bar{\sigma} = \mathbf{0}, \mathbf{v}_0 = \mathbf{c}^{(FF)}\}$.

3) The random way-point model is the most prevalent mobility model in mobile ad hoc network (MANET) simulations. In this model, a mobile cycles through two phases: constant-velocity and motionless. In each cycle, it randomly selects a destination, moves toward the destination at a randomly selected velocity until it reaches the destination, stays at the destination for a randomly chosen amount of time, and starts a new cycle by randomly selecting a new destination. Therefore, the movement of a mobile under this model consists of segments of fluid-flow with various level of velocity (including zero velocity).

Since the stochastic process of mobile velocity under the random way-point model is non-stationary, it cannot be obtained by *direct* application of the Gauss-Markov model. However, with the Gauss-Markov model, one only needs to employ an additional random number generator to generate movement patterns that follow the random way-point model description. This random number generator determines the duration of each segment of movement. In each segment, the Gauss-Markov model, as shown in *2)*, generates fluid-flow-like movement patterns. Thus, the combined mobility model consists the following cycle:

i. Generate random number t;
ii. Follow the fluid-flow model for t seconds;
iii. Generate random number t;
iv. Stop for t seconds;
v. Repeat.

[3] A comparative study of various mobility models, including random-walk, Gauss-Markov, and random way-point, in the context of mobile ad hoc network routing, can be found in [16].

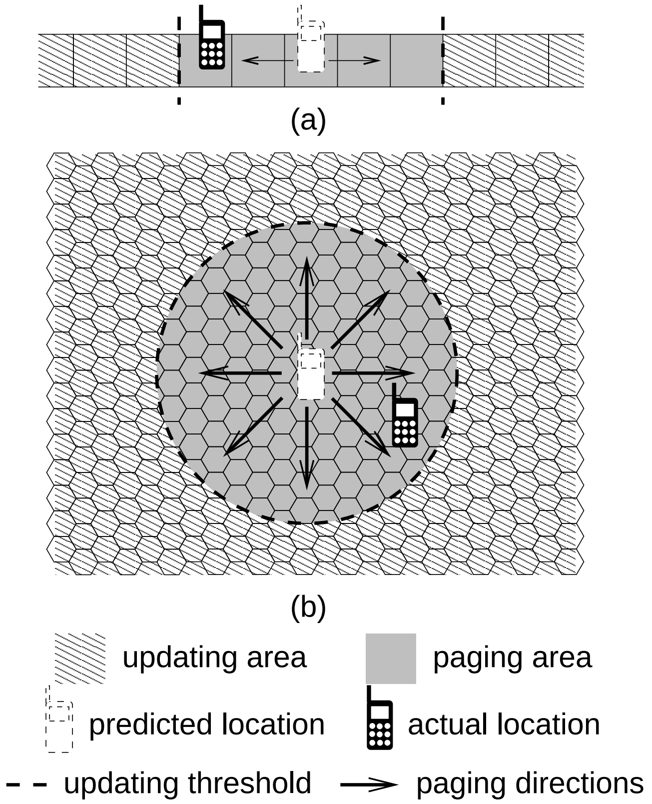

Fig. 1. Predictive Location Updating and Selective Paging: (a) One-Dimensional System (b) Two-Dimensional System

B. Predictive Distance-Based Mobility Management

In what follows, we describe the predictive mobility management scheme in the context of 1-D and 2-D systems. This can easily be extended to systems of three dimensions, such as those in an office building, for example.

B.1 Predictive Location Updating and Selective Paging

As explained in Section I, a mobile cannot monitor its location continuously. Here, we assume that location inspections are performed by the mobiles periodically, at every m units of time. Then, the regular, *non-predictive* distance-based mobility management scheme has the following components:

- *updating*: At the k^{th} location inspection instant, a mobile observes its current location $\mathbf{s}_{km}$. It transmits a location update to the PCS network, if $\mathbf{s}_{km}$ is greater than the distance threshold away from the location where it last updated.
- *paging*: When a call is made to a mobile, the system pages the mobile in cells at and around the mobile's last reported location, in a shortest-distance-first order, and until the mobile is found[4] .

The proposed *predictive* distance-based scheme has the following components:

- *prediction*: The PCS network makes predictions of the mobile's location $\mathbf{p}_n$, at time instants of mobile location inspections and call arrivals to the mobile, based on the history of the mobile's reported location and velocity. The mobile is made aware of the network's prediction, either by information transfer from the network during the last location updating or by local, concurrent computation using the same prediction algorithm as the network.
- *updating*: At the k^{th} location inspection instant, a mobile observes its current location $\mathbf{s}_{km}$. It transmits a location update to the PCS network if $|\mathbf{s}_{km} - \mathbf{p}_{km}|$ is greater than the distance threshold.
- *paging*: When a call is made to a mobile at time n, the system pages the mobile in cells at and around the predicted location $\mathbf{p}_n$, in a shortest-distance-first order, and until the mobile is found[5] .

Figures 1 illustrates the predictive updating and paging schemes for the 1-D and 2-D cases, respectively.

In paging a mobile, there exists a trade-off between the paging traffic bandwidth usage and the time delay in locating the mobile (e.g., a full system paging scheme incurs maximum bandwidth usage and minimum delay). There is a spectrum of means to trade off between bandwidth usage and paging delay. In the rest of this paper, for both the predictive and the non-predictive cases, we have assumed a sequential, ring-by-ring paging scheme, where all cells of equal distance to the predicted location or the last reported location, respectively, are paged at the same time. For systems with different bandwidth and delay constraint requirements, the proposed predictive scheme can be easily extended using the methods similar to those reported in [14], [3], and [24].

The updated mobile location and velocity are assumed to be contained in location databases residing in the wireline network (e.g., Home Location Registers). The reliability issues of location databases are not addressed in this paper. For reliability considerations, methods similar to those described in [12] could be employed.

B.2 Location Prediction

Next, we derive a formula for the predicted location of a mobile as a function of time, when its last reported location and velocity are given. Let $n = 0$ be the initial time when a mobile last updates its location and velocity. We can recursively expand (4) to express $\mathbf{v}_n$ explicitly in terms of the initial velocity $\mathbf{v}_0$,

$$\mathbf{v}_n = \alpha^n \mathbf{v}_0 + (1-\alpha^n)\mu + \sigma\sqrt{1-\alpha^2}\sum_{i=0}^{n-1}\alpha^{n-i-1}\mathbf{w}_i \ . \quad (7)$$

Define $\mathbf{s}_0 = 0$, and $\mathbf{s}_n = \mathbf{s}_{n-1} + \mathbf{v}_{n-1}$, we have

$$\begin{aligned}\mathbf{s}_n = \sum_{i=0}^{n-1}\mathbf{v}_i = \frac{1-\alpha^n}{1-\alpha}\mathbf{v}_0 + \left(1 - \frac{1-\alpha^n}{1-\alpha}\right)\mu \\ + \sigma\sqrt{1-\alpha^2}\sum_{i=0}^{n-1}\sum_{j=0}^{i-1}\alpha^{i-j-1}\mathbf{w}_j \ .\end{aligned} \quad (8)$$

It's easy to show that any linear combination of $\mathbf{s}_0$, $\mathbf{s}_1$, ..., and $\mathbf{s}_n$ is Gaussian. Hence, $\mathbf{s} \triangleq \{\mathbf{s}_n\}$ is a Gaussian process.

Let $\mathbf{z}_n$ be the displacement of a mobile at time n from its last updated location, given that no location updating is performed

[4] A time or distance threshold may be imposed, beyond which the paging terminates, in case the sought mobile has detached from the network.

[5] See the previous footnote.

up to time $n-1$, i.e.,

$$|\mathbf{s}_{km} - \mathbf{p}_{km}| \leq N \ , \tag{9}$$

for $k = 1, 2, \ldots, \lceil n/m \rceil - 1$, where N is the updating distance threshold. The joint PDF of $\mathbf{z} \triangleq \{\mathbf{z}_n\}$ is

$$f_{\mathbf{z}}(\xi) = \begin{cases} \dfrac{f_{\mathbf{s}}(\xi)}{\displaystyle\int_{|\xi_{km} - \mathbf{p}_{km}| \leq N} f_{\mathbf{z}}(\xi) d\xi} \ , \\ \qquad |\xi_{km} - \mathbf{p}_{km}| \leq N \ , k = 1, 2, \ldots, \lceil n/m \rceil - 1 \ ; \\ 0 \ , \qquad \text{otherwise} \ , \end{cases} \tag{10}$$

where $\xi \triangleq \{\xi_n\}$, and $f_{\mathbf{s}}(\xi)$ is the joint PDF of $\mathbf{s}$.

Since $\{\mathbf{s}_n\}$ is a Gaussian process, the maximum-likelihood estimation of $\mathbf{s}_n$ is $\mathrm{E}[\mathbf{s}_n]$. Due to symmetry in the definition of $\{\mathbf{z}_n\}$, as shown in (10), the maximum-likelihood estimation of $\mathbf{z}_n$ is

$$\begin{aligned} \mathbf{p}_n &= \mathrm{E}[\mathbf{z}_n] = \mathrm{E}[\mathbf{s}_n] \\ &= \frac{1-\alpha^n}{1-\alpha}\mathbf{v}_0 + \left(1 - \frac{1-\alpha^n}{1-\alpha}\right)\mu \ . \end{aligned} \tag{11}$$

Furthermore, since the PDF of $\mathbf{z}_n$ is concave and symmetrical, the proposed shortest-distance-first paging scheme conforms with the highest-probability-first rule for optimal selective paging. Therefore, we can expect the predictive scheme to incur lower mobility management cost than the non-predictive distance-based scheme.

C. *Gauss-Markov Model Parameter Estimation*

In a practical system, the mobility pattern can be approximated by the Gauss-Markov model with varying mean, variance, and memory level. Applying the algorithm in Section II-B, the movement of a mobile is predicted based on the estimated parameters $\hat{\mu}$, $\hat{\sigma}$, and $\hat{\alpha}$, obtained at the previous location updates. Therefore, the accurate location prediction depends upon the accurate dynamic estimation of the Gauss-Markov parameters, as well as the degree of variations in these parameters between location updates.

The topic of dynamic parameter estimation has seen many contributions within the signal-processing community. In this work, we propose a simple estimator that computes the mean, variance, and memory level of a mobile, *separately*, in each dimension. We define a parameter estimation window w_{est}. In practice, w_{est} is chosen such that it is small enough for the mobile movement to remain approximately stationary but large enough to provide sufficient amount of data samples. Given the previous w_{est} samples of the mobile velocity $\{v_i\}$, the Gauss-Markov parameters are estimated by

$$\hat{\mu} = \frac{1}{w_{est}} \sum_{i=1}^{w_{est}} v_i \tag{12}$$

$$\hat{\sigma}^2 = \frac{1}{w_{est}-1} \sum_{i=1}^{w_{est}} (v_i - \hat{\mu})^2 \tag{13}$$

$$\hat{\sigma}_1^2 = \frac{1}{w_{est}-1} \sum_{i=1}^{w_{est}-1} (v_i - \hat{\mu})(v_{i+1} - \hat{\mu}) \tag{14}$$

$$\hat{\alpha} = \begin{cases} 1, & \text{if } \hat{\sigma} \approx 0 \\ \max\{0, \frac{\hat{\sigma}_1^2}{\hat{\sigma}^2}\}, & \text{otherwise} \ . \end{cases} \tag{15}$$

In the above estimation, both $\hat{\sigma}^2$ and $\hat{\sigma}_1^2$ are biased because of the correlation between elements in $\{v_i\}$. Despite this drawback, the reasons that we have chosen this algorithm are twofold. First, this estimator is simple to implement in practice. Second, a sub-optimal parameter estimator does *not* invalidate the correctness of the proposed predictive location updating and paging scheme.

With correlated data and limited estimation window size, an optimal estimator has not been found. However, it is stated in [17] that the above estimator can out-perform an unbiased estimator.

Thus, the Gauss-Markov parameter estimates of each mobile are recomputed at each of its location updating events. These estimates are used by the mobile and the system to predict the future location of the mobile. Section IV-C demonstrates the effectiveness of combining the above estimator with distance-based predictive mobility management.

III. Cost Evaluation of the Predictive Distance-Based Mobility Management Scheme

In this section, we describe an analytical framework to evaluate the mobility management cost of the predictive distance-based scheme, assuming ideal Gauss-Markov mobility pattern. The performance of the proposed scheme in the non-ideal, time-varying case is studied in Section IV.

The total cost of mobility management *per call arrival* is defined as

$$C_{total} = C_{inspect} + C_{update} + C_{page} \ , \tag{16}$$

where $C_{inspect}$, C_{update}, and C_{page} are the average location inspection cost per call arrival, the average updating cost per call arrival, and the average paging cost per call arrival, respectively.

Clearly, there exists a tradeoff between location updating and mobile paging. The more often a mobile terminal updates its location, the more precise is the network's knowledge of its location when a call arrives, which could lead to lower paging cost. However, location updating drains the limited energy supply of a mobile terminal, takes away part of the precious wireless channel capacity, and induces access to potentially far-away and heavily-loaded location databases. Therefore, the rate of location updating, which depends on the location updating threshold in this case, should be optimized to reduce the total cost.

Furthermore, the frequency of location inspection is another design parameter that needs to be optimized. On the one hand, location inspection takes up the limited processor power and energy supply of a mobile terminal. On the other hand, it improves the mobile's awareness of its location for timely location updating, which is particularly important when the frequency of call arrivals is high.

In order to determine the optimal location inspection frequency and the optimal location updating threshold, we propose the following framework for evaluating the cost of mobility management. We first describe the cost evaluation framework of a 1-D system. We then extend that to 2-D and 3-D systems, and follow that with an approximation scheme, which reduces the computational complexity of multi-dimensional cost evaluation to that of a 1-D system.

The following notations are used throughout this section:

1-D Cost Evaluation

- m: location inspection period
- N: location updating threshold
- C_i: cost of a single location inspection
- C_u: cost of a single autonomous location update
- C_p: cost of paging a single cell, normalized to 1
- $f_{call}(n)$: PDF of call inter-arrival time
- λ: call arrival rate (if call arrivals are Poisson)
- $f_{upd}(n)$: PDF of time between a location update and the next autonomous update
- η_n: velocity at time n given that no location update is performed up to time $\lfloor n/m \rfloor m$
- μ_{η_n}: mean of η_n
- u_n: mean-shifted η_n, i.e., $\eta_n - \mu_{\eta_n}$
- z_n: displacement at time n, given that no location update is performed up to time $\lfloor n/m \rfloor m$
- μ_{z_n}: mean of z_n
- r_n: mean-shifted z_n, i.e., $z_n - \mu_{z_n}$
- u'_{km}: $k = \lfloor n/m \rfloor$, mean-shifted velocity at the k^{th} location inspection, given that no location update is performed up to time $(k-1)m$.
- r'_{km}: $k = \lfloor n/m \rfloor$, mean-shifted displacement at the k^{th} location inspection, given that no location update is performed up to time $(k-1)m$.
- y_k: $y_k = \sum_{i=km}^{(k+1)m-1} u_i$
- $\delta_{k,i}$: $\delta_{k,i} = \sum_{j=km}^{km+i-1} u_{km+j}$
- $P_N(r)$: $P_N(r) = \begin{cases} 1, & -N \leq r \leq N \\ 0, & \text{otherwise} \end{cases}$
- $R_{upd}(k)$: probability that there is no update up to and including the k^{th} location inspection
- $U(n)$: number of location updates between two successive call arrivals of n time units apart
- $M(n)$: expected value of $U(n)$
- $F^{(i)}(n)$: distribution of time to the i^{th} location update
- $K_p(r)$: number of cells in which the network needs to page a mobile when the mobile is at a displacement r from its predicted location
- S: area of a cell
- $g(i)$: expected number of cells paged, when a call arrives at time i after the last location update
- $P(n)$: expected number of cells paged, when a call arrives at time n from the last call arrival

2-D Cost Evaluation

- superscripts $(\cdot)^x$, $(\cdot)^y$, and $(\cdot)^{(2d)}$: denote the dimension of a quantity or function in the 2-D cost evaluation
- $P_N^{(2d)}(r^x, r^y)$: location cut-off function
- $P_L^{(2d)}(r^x, r^y)$: approximate location cut-off function

A. 1-D Cost Evaluation

Since the position of a mobile (e.g., the residence cell identity) is closely monitored by the network within a phone call duration, a call arrival has the same effect as a mobile location update. To distinguish the two types of location updating, we refer to a location update based on distance as *autonomous update*. Furthermore, we can view a call duration as a point in time, where a location update is performed, and a new cycle of location inspection, location updating, and mobile paging starts.

Suppose calls arrive with the inter-arrival time PDF $f_{call}(n)$. The average location inspection cost per call is

$$C_{inspect} = C_i \sum_{n=0}^{\infty} f_{call}(n) \lfloor n/m \rfloor \ , \tag{17}$$

where C_i denotes the cost of a single location inspection. If we assume that the call arrival times are exponentially distributed with arrival rate λ, we can simplify (17) to

$$C_{inspect} = C_i \frac{(1-\lambda)^m}{1-(1-\lambda)^m} \ . \tag{18}$$

The detailed derivation is presented in Appendix A.

To determine C_{update}, we first consider the time interval between two consecutive autonomous location updates without the interruption of phone calls. We compute the PDF of time between a location update (autonomous update or call-arrival update) and the next autonomous update, denoted $f_{upd}(n)$.

Let η_n and z_n be the velocity and displacement of a mobile at time n, respectively, given that no location updating is performed up to the location inspection at time $km = \lfloor n/m \rfloor m$. Shifting the center of the PDFs of η_n and z_n to the origin, we define

$$u_n = \eta_n - \mu_{\eta_n} \tag{19}$$
$$r_n = z_n - \mu_{z_n} \ . \tag{20}$$

Then, we have $u_0 = r_0 = 0$, and for $n \neq km$,

$$u_n = \alpha u_{n-1} + \sigma \sqrt{1-\alpha^2} w_{n-1} \ , \tag{21}$$

and

$$r_n = r_{n-1} + u_{n-1} \ . \tag{22}$$

We further define u'_{km} and r'_{km} as the mean-shifted velocity and displacement of a mobile at the k^{th} location inspection, respectively, given that no location updating is performed up to the $(k-1)^{th}$ location inspection (i.e., r'_{km} could be greater than N). Obviously, u_{km} is u'_{km} conditioned on $|r'_{km}| \leq N$, and r_{km} is r'_{km} conditioned on $|r'_{km}| \leq N$.

In order to compute the probability of an update at the k^{th} location inspection, we need to compute the PDF of r_{km}. It's easy to see that the statistics of (r_{km}, u_{km}) is completely determined by $(r_{(k-1)m}, u_{(k-1)m})$. Therefore, we can compute the joint PDF $f_{r_{km},u_{km}}(r,u)$, through a recursion on k as follows.

Let

$$y_k = \sum_{i=km}^{(k+1)m-1} u_i \ . \tag{23}$$

Then,

$$r'_{(k+1)m} = r_{km} + y_k \ . \tag{24}$$

By the Markovian property of u_n, we have

$$\begin{aligned} & f_{y_k|(u_{km},u'_{(k+1)m})}(y,\xi_1,\xi_2) \\ =& f_{y_0|(u_0,u'_m)}(y,\xi_1,\xi_2) = \frac{1}{\sigma_{y_0|u_0}\sqrt{2\pi(1-\gamma^2)}} \\ & \cdot \exp\left\{ -\frac{\left[(y-\mu_{y_0|u_0}) - \gamma\frac{\sigma_{y_0|u_0}}{\sigma_{u'_m|u_0}}(\xi_2 - \mu_{u'_m|u_0})\right]^2}{2\sigma^2_{y_0|u_0}(1-\gamma^2)} \right\} \ , \end{aligned} \tag{25}$$

where

$$\mu_{u'_m|u_0} = \alpha^m \xi_1 \ , \tag{26}$$

$$\mu_{y_0|u_0} = \frac{1-\alpha^m}{1-\alpha}\xi_1 \ , \tag{27}$$

$$\sigma^2_{u'_m|u_0} = (1-\alpha^{2m})\sigma^2 \ , \tag{28}$$

$$\sigma^2_{y_0|u_0} = \frac{-(1-\alpha^m)^2 + m(1-\alpha^2) - 2\alpha(1-\alpha^m)}{(1-\alpha)^2}\sigma^2 \ , \tag{29}$$

$$\gamma = \frac{C_{(u'_m,y_0)|u_0}}{\sigma_{u'_m|u_0}\sigma_{y_0|u_0}} \ , \tag{30}$$

and

$$C_{(u'_m,y_0)|u_0} = \frac{(1-\alpha^m)(\alpha-\alpha^m)}{1-\alpha}\sigma^2 \ . \tag{31}$$

The derivation of $f_{y_0|(u_0,u'_m)}(y,\xi_1,\xi_2)$ is presented in Appendix B.

Since r_{km} and y_k are independent given u_{km}, we have, from (24),

$$\begin{aligned} & f_{r'_{(k+1)m}|(u_{km},u'_{(k+1)m})}(r,\xi_1,\xi_2) \\ =& f_{r_{km}|u_{km}}(r,\xi_1) * f_{y_{km}|(u_{km},u'_{(k+1)m})}(r,\xi_1,\xi_2) \\ =& f_{r_{km}|u_{km}}(r,\xi_1) * f_{y_0|(u_0,u'_m)}(r,\xi_1,\xi_2) \ , \end{aligned} \tag{32}$$

where $*$ denotes linear 1-D convolution over the variable r.

Then, by Bayes rule and taking the marginal density, we have

$$\begin{aligned} & f_{r'_{(k+1)m},u'_{(k+1)m}}(r,\xi_2) \\ =& \int_{-\infty}^{\infty} f_{r'_{(k+1)m}|(u_{km},u'_{(k+1)m})}(r,\xi_1,\xi_2) f_{u_{km},u'_{(k+1)m}}(\xi_1,\xi_2)d\xi_1 \\ =& \int_{-\infty}^{\infty} f_{r'_{(k+1)m}|(u_{km},u'_{(k+1)m})}(r,\xi_1,\xi_2) f_{u'_{(k+1)m}|u_{km}}(\xi_2,\xi_1) \\ & \quad \cdot \int_{-\infty}^{\infty} f_{r_{km},u_{km}}(\zeta,\xi_2)d\zeta d\xi_1 \ , \end{aligned} \tag{33}$$

where

$$\begin{aligned} & f_{u'_{(k+1)m}|u_{km}}(\xi_2,\xi_1) = f_{u'_m|u_0}(\xi_2,\xi_1) \\ =& \frac{1}{\sigma_{u'_m|u_0}\sqrt{2\pi}} \exp\left[-\frac{(\xi_2 - \mu_{u'_m|u_0})^2}{2\sigma^2_{u'_m|u_0}} \right] \ . \end{aligned} \tag{34}$$

Because of the distance-based updating, we obtain

$$\begin{aligned} & f_{r_{(k+1)m},u_{(k+1)m}}(r,u) \\ =& \frac{f_{r'_{(k+1)m},u'_{(k+1)m}}(r,u)P_N(r)}{\int_{-\infty}^{\infty}\int_{-N}^{N} f_{r'_{(k+1)m},u'_{(k+1)m}}(r,u)drdu} \ , \end{aligned} \tag{35}$$

where

$$P_N(r) = \begin{cases} 1, & -N \le r \le N \\ 0, & \text{otherwise} \ . \end{cases} \tag{36}$$

Let $R_{upd}(k)$ be the probability that there is no update up to and including the k^{th} location inspection instant. We have $R_{upd}(0) = 1$, and

$$\begin{aligned} & R_{upd}(k+1) \\ =& R_{upd}(k)\int_{-\infty}^{\infty}\int_{-N}^{N} f_{r'_{(k+1)m},u'_{(k+1)m}}(r,u)drdu \ . \end{aligned} \tag{37}$$

The above iteration to compute $f_{r_{km}}(r)$ can be terminated either when $R_{upd}(k)$ is sufficiently small, or when the series of PDFs $f_{r_{km},u_{km}}(r,u)$ converges, e.g., in terms of normalized squared error,

$$\frac{\int_{-\infty}^{\infty}\int_{-N}^{N}\left[f_{r_{km},u_{km}}(r,u) - f_{r_{(k-1)m},u_{(k-1)m}}(r,u)\right]^2 drdv}{\int_{-\infty}^{\infty}\int_{-N}^{N}\left[f_{r_{km},u_{km}}(r,u)\right]^2 drdv} < \epsilon \ ,$$

where ϵ is a pre-defined threshold. Suppose the iteration is terminated at the q^{th} step. Since $f_{r_{km},u_{km}}(r,u)$ uniquely determines $f_{r'_{(k+1)m},u'_{(k+1)m}}(r,u)$, for $k \ge q$, $R_{upd}(k+1)/R_{upd}(k)$ is essentially constant and

$$R_{upd}(k+1) \approx \frac{R_{upd}(q)}{R_{upd}(q-1)}R_{upd}(k) \ . \tag{38}$$

Thus, we can compute $R_{upd}(k)$ successively for all $k \ge q$. We carry out this computation until $R_{upd}(k)$ is small enough.

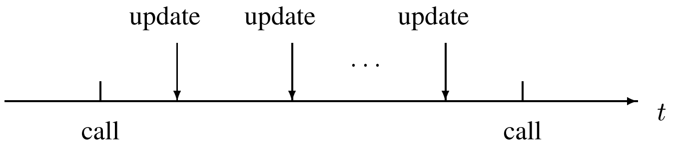

Fig. 2. Renewal Process between Two Successive Call Arrivals

Thus, we obtain the PDF of the time between a location update to the next autonomous update by

$$f_{upd}(km) = R_{upd}(k-1) - R_{upd}(k) \qquad k = 1, 2, \dots , \tag{39}$$

and $f_{upd}(n) = 0$ for n not a integer multiple of m. Since $f_{upd}(km)$ is not dependent on the history of a mobile's location or velocity before time $k = 0$, the time between consecutive autonomous updates is independent and identically distributed.

Next, we consider the location updates between two successive call arrivals. Since a call arrival has the same effect as a location update, the time between the first call arrival and the first autonomous update has PDF $f_{upd}(n)$. As shown in Figure 2, the i.i.d. location update time intervals comprise a renewal process with the probability density function $f_{upd}(n)$.

Let $U(n)$ denote the number of location updates within the time interval of length n between two successive call arrivals. Then,

$$\Pr[U(n) = i] = F^{(i)}(n) - F^{(i+1)}(n) , \tag{40}$$

where

$$F^{(i)}(n) \triangleq \sum_{l=0}^{n} f^{(i)}(l) , \tag{41}$$

and $f^{(i)}(n)$ is the PDF of time between a location update and the ith autonomous update in the future, as given by the following recursive relation

$$f^{(1)}(n) = f_{upd}(n) \tag{42}$$

$$f^{(i)}(n) = f^{(i-1)}(n) * f_{upd}(n) , \tag{43}$$

where $*$ denotes the discrete convolution.

Let $M(n)$ be the expected value of $U(n)$. Then,

$$\begin{aligned} M(n) &= \sum_{i=1}^{\infty} i(F^{(i)}(n) - F^{(i+1)}(n)) \\ &= \sum_{i=1}^{\infty} iF^{(i)}(n) - \sum_{i=2}^{\infty}(i-1)F^{(i)}(n) \\ &= \sum_{i=1}^{\infty} F^{(i)}(n) . \end{aligned} \tag{44}$$

Using (44), the average updating cost per call arrival can be obtained by

$$C_{update} = C_u \sum_{n=0}^{\infty} f_{call}(n) M(n) , \tag{45}$$

where C_u is the cost of a single autonomous location update.

To find C_{page}, we first need to determine the distribution of r_n, for $n = km+1, km+2, \dots , km+m-1$, given $f_{r_{km},u_{km}}(r, u)$. Let

$$\delta_{k,i} = \sum_{j=km}^{km+i-1} u_{km+j} , \quad i = 1, 2, \dots , m . \tag{46}$$

Similar to the mean and variance of $y_0|v_0$ derived in Appendix B, we can show that $\delta_{k,i}|u_{km}$ is Gaussian with mean $\frac{1-\alpha^i}{1-\alpha} u_{km}$ and variance $\frac{-(1-\alpha^i)^2 + i(1-\alpha^2) - 2\alpha(1-\alpha^i)}{(1-\alpha)^2}\sigma^2$. Then, since

$$r_{km+i} = r_{km} + \delta_{k,i} , \tag{47}$$

and r_{km} and $\delta_{k,i}$ are mutually independent given u_{km}, we have

$$f_{r_{km+i}}(r) = \int_{-\infty}^{\infty} f_{r_{km},u_{km}}(r, u) * f_{\delta_{k,i}|u_{km}}(r, u) du , \tag{48}$$

where $*$ denotes convolution in terms of r.

Let $K_p(r)$ be the number of cells in which the network needs to page a mobile when the mobile is at a displacement of r from its predicted location. In a 1-D predictive system with a shortest-distance-first paging rule,

$$K_p(r) = 2\lfloor r/S + 0.5 \rfloor + 1 , \tag{49}$$

where S is the cell size. Then, the expected number of cells paged when a call arrives at time i from the last location updating is

$$g(i) = \int_{-\infty}^{\infty} K_p(r) f_{r_i}(r) dr . \tag{50}$$

Let $P(n)$ be the expected number of cells paged when a call arrives at time n from the last call arrival. We can compute $P(n)$ with the following recursive equation:

$$\begin{aligned} P(n) = &R_{upd}\left(\lfloor n/m \rfloor\right) g(n) + f_{upd}(m) P(n-m) \\ &+ f_{upd}(2m) P(n-2m) + \cdots \\ &+ f_{upd}\left(\lfloor n/m \rfloor m\right) P\left(n - \lfloor n/m \rfloor m\right) . \end{aligned} \tag{51}$$

Then, the average paging cost per call arrival is given by

$$C_{page} = C_p \sum_{n=0}^{\infty} f_{call}(n) P(n) , \tag{52}$$

where C_p is the cost of paging a single cell.

B. 2-D Cost Evaluation

Section III-A describes the analytical framework for evaluating the predictive mobility management scheme in a 1-D system. In a 2-D system, similarly to equations (19) and (20), we can define mean-shifted velocity, $\mathbf{u}_n = (u_n^x, u_n^y)^T$, and mean-shifted displacement, $\mathbf{r}_n = (r_n^x, r_n^y)^T$, given that no location updating is performed up to the $\lfloor n/m \rfloor^{th}$ location inspection. Then, similarly to equations (19)-(39), we can compute $f_{upd}^{(2d)}(n)$

and $f_{r^x_{km}, r^y_{km}, u^x_{km}, u^y_{km}}(r^x, r^y, u^x, u^y)$ through an iterative procedure over $f_{\mathbf{r}_{km}, \mathbf{u}_{(k-1)m}, \mathbf{u}_{km}}(\mathbf{r}, \mathbf{u}_1, \mathbf{u}_2)$. Here, the location cut-off function, as shown in equation (36), becomes

$$P_N^{(2d)}(r^x, r^y) = \begin{cases} 1, & \sqrt{(r^x)^2 + (r^y)^2} \le N \\ 0, & \text{otherwise} , \end{cases} \tag{53}$$

where N is the threshold radius for location updating.

Then, the computation of $C_{inspect}$ remains unchanged, and the computation of C_{update} can be carried out exactly as in equations (40)-(45). The computation of C_{page} is similar to that described in equations (46)-(52), except that $K_p(\mathbf{r})$ in (49) needs to be redefined.

The number of cells paged by the network to find a mobile is largely dependent upon the cell shape and lattice arrangement. In order to obtain analytical results that are compatible with all cell configurations, we assume that the number of cells paged is proportional to the size of the area paged by the network. Since the optimal, shortest-distance-first and ring-by-ring paging pattern is used, the area paged when the mobile is at (r^x, r^y) is a circle centered at the predicted location and with radius $\sqrt{(r^x)^2 + (r^y)^2}$. Thus, here the 2-D $K_p^{(2d)}(\mathbf{r})$ is defined as

$$K_p^{(2d)}(r^x, r^y) = \left\lfloor \pi \left((r^x)^2 + (r^y)^2 \right) / S^{(2d)} \right\rfloor + 1 , \tag{54}$$

where $S^{(2d)}$ is the average area of a cell.

As hinted by equations (32)-(34), the direct extension of the 1-D algorithm to the 2-D case can potentially be very computationally intensive, since it requires convolution and other manipulation of $f_{\mathbf{r}_{km}, \mathbf{u}_{(k-1)m}, \mathbf{u}_{km}}(\mathbf{r}, \mathbf{u}_1, \mathbf{u}_2)$ with six free variables. In what follows, we introduce an approximating algorithm for 2-D cost evaluation, which requires only as much computation as the 1-D algorithm.

C. 2-D Cost Approximation

As shown in Section III-B, the difficulty there is in computing the joint PDF of location and velocity in both dimensions. This problem can be alleviated by an approximation based on the assumption that a mobile performs location updating separately in each of the two dimensions. In this case, we define a new location cut-off function

$$P_L^{(2d)}(r^x, r^y) = \begin{cases} 1, & -L \le r^x \le L \text{ and } -L \le r^y \le L \\ 0, & \text{otherwise} , \end{cases} \tag{55}$$

where L is the threshold distance for location updating in each dimension, and

$$L = \sqrt{\pi N^2} , \tag{56}$$

which maintains the same size of the range of possible mobile location immediately after a location inspection.

With this location updating assumption, the mobile location and update time statistics can be computed independently in each dimension, and then combined to give the 2-D statistics. The 2-D autonomous update time is the minimum of the 1-D autonomous update time in each dimension. In particular, since we have assumed that the mobiles have the same and independent mobility pattern in both dimensions, we have

$$R_{upd}^{(2d)}(k) = R_{upd}^2(k) , \tag{57}$$

and hence,

$$f_{upd}^{(2d)}(km) = 2 f_{upd}(km) R_{upd}(km) + f_{upd}^2(km) . \tag{58}$$

The 2-D mobile location PDF can be obtained by

$$f_{r^x_n, r^y_n}(r^x, r^y) = f_{r_n}(r^x) f_{r_n}(r^y) . \tag{59}$$

Then, we can use the same procedures described in Section III-B to compute the updating and paging costs. To make the approximation results more precise, we can modify $K_p^{(2d)}(r^x, r^y)$ to reflect the mobile location constraint introduced by equation (55). For simplicity, in this work, we only consider this constraint at time $n = km$. In this case, the paging area is the intersection of a circle and a square centered at the predicted location, with radius $N = \sqrt{(r^x)^2 + (r^y)^2}$ and side-length L, respectively. Thus, we redefine

$$K_p^{(2d)}(r^x, r^y) = \begin{cases} \left\lfloor \frac{\pi N^2}{S^{(2d)}} \right\rfloor + 1, & N \le L \\ \left\lfloor \frac{\pi N^2}{S^{(2d)}} - 4(N^2 \tan^{-1} \frac{\sqrt{N^2 - L^2}}{L} \right. & \\ \left. \quad - \frac{L\sqrt{N^2 - L^2})}{S^{(2d)}} \right\rfloor + 1, & N > L \end{cases} \tag{60}$$

Since (55) represents a sub-optimal updating scheme, C_{total} obtained here is an upper-bound of the actual cost that can be obtained by the direct 2-D computation in Section III-B. In all experiments that we have conducted, this upper-bound is within 2% of the actual cost. Thus, we conclude that the approximation is quite reasonable, allowing us to significantly reduce the computation complexity.

IV. Numerical Results and Comparisons

In this section, we evaluate the cost of a two-dimensional system. When analyzing the cost of the predictive scheme in the case of ideal Gauss-Markov mobility pattern, we use the upper-bound approximation described in Section III-C to expedite the computations. In the case of time-varying mobility pattern, the performance of the predictive scheme, in conjunction with the proposed parameter estimator, is demonstrated through simulation.

In order to reduce the redundancy of numerical representation, we scale both the spatial parameters and the cost parameters. We normalize all distance and velocity values according to the cell size, such that the area of a cell is assumed to be one. Furthermore, since only the *relative* costs of location inspection, location updating, and mobile paging affect the performance optimization, we also normalize all costs to have the units of C_p, the cost of a single paging. Thus, C_u now represents the ratio of the cost of one location update to the cost of a single paging, and C_i now represents the ratio of the cost of one location inspection to the cost of a single paging.

The costs C_i, C_u, and C_p depend on multiple factors such as the mobile battery and processor capacity, the system wireless bandwidth, and the location database architecture. It is impossible to quantify all these factors using a single metric. In fact, it is difficult to properly quantify any one of these three costs with a single metric. Therefore, the values of C_i and C_u in our analysis merely represent the relative importance of location inspection, location updating, and mobile paging in evaluating the performance of a system.

Clearly, the relative importance of these different costs is determined by the system design details on a case-by-case basis. Furthermore, these costs may change over time, depending on the time varying characteristics of the battery power, wireless channel usage, database loading, network congestion, and many other factors that a system designer need to consider. In this work, we consider the broad applicability of predictive mobility management in different networks. Therefore, instead of giving the numerical results of one specific type of network, we study the optimization and performance of the proposed scheme with a wide range of C_i and C_u values.

For the case of ideal Gauss-Markov mobility pattern, we are interested in understanding how C_u, C_i, and the other system variables, namely, the memory factor exponent β, the average velocity μ, the standard deviation σ, the location inspection period m, and the updating distance N, affect the performance of the predictive scheme. In particular, we are interested in the joint optimization of m and N, namely,

$$(m_{opt}, N_{opt}) = \arg\min_{m,N} C_{total}(m, N, \beta, \mu, \sigma, \lambda, C_u, C_i) \ , \tag{61}$$

which gives the minimum cost of mobility management $C_{total}(m_{opt}, N_{opt})$.

We are also interested in the performance gain that the predictive scheme can achieve over the non-predictive distance-based scheme. In particular, we are interested in how the cost ratio is affected by the aforementioned system parameters.

Although our analytical framework accommodates a general call-arrival distribution, in the following numerical computations, we assume that the inter-arrival time of calls is exponentially distributed with arrival rate λ [calls/unit time].

A. Joint Optimization of m and N with Ideal Gauss-Markov Mobility Pattern

Assuming ideal prediction, μ has no effect on the cost of predictive mobility management. Other than that, however, it is not easy to solve equation (61) using general numerical optimization tools. Therefore, in what follows, given a set of system parameters, we first compute C_{total} for a range of values of m and N and then pick the pair that gives the minimum cost.

In Figures 3-8, we study the effect of β, σ, λ, C_i, and C_u on the minimization of C_{total}.

Figures 3 and 4 present the plots of the minimum C_{total} and the corresponding (m_{opt}, N_{opt}), respectively, as β increases along the values $\{10^{-2}, 10^{-1.5}, 10^{-1}, 10^{-0.5}, 10^0, 10^{0.5}\}$ (i.e., α decreases along the values $\{0.99, 0.97, 0.90, 0.73, 0.37, 0.04\}$). For each curve, σ takes a value from $\{10^{-1}, 10^{-0.5}, 10^0, 10^{0.5}\}$. In both figures, $\lambda = 0.01$, $C_u = 10$, and $C_i = 1$.

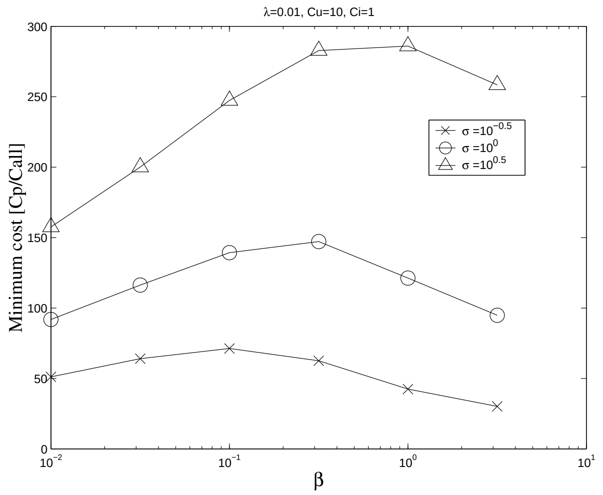

Fig. 3. Minimum cost vs. β and σ for $\lambda = 0.01$, $C_u = 10$, and $C_i = 1$

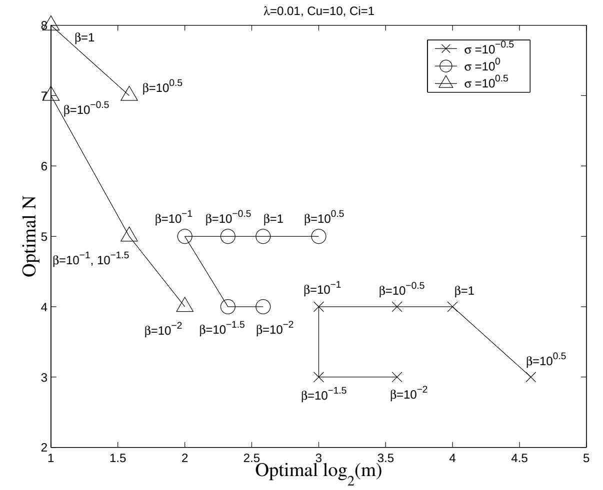

Fig. 4. Joint optimization of m and N vs. λ and C_i for $\lambda = 0.01$, $C_u = 10$, and $C_i = 1$

Figure 3 illustrates a seemingly counter-intuitive phenomenon. It shows that the minimal total cost is not always decreasing with the memory level ($\alpha = e^{-\beta}$) of a mobile's velocity. On the one hand, for fixed σ and small β , as β increases, a mobile's velocity becomes less correlated, so the total cost of using prediction increases. On the other hand, when β is large, the variance of $\sum_{i=0}^{n} v_i$ approaches a local minimum $n\sigma^2$, so a mobile's position becomes more predictable, and therefore the total cost decreases.

In addition, Figure 3 also shows that the minimal total cost is an increasing function of the variance of a mobile's velocity. This matches the intuition that the cost of predictive mobility management increases as a mobile's velocity becomes more chaotic.

Figure 4 shows that, corresponding to the results in Figure 3, the optimal m and the optimal N are respectively convex and concave functions of β. However, the optimal m is always a decreasing function of σ, and the optimal N is always an increasing function of σ. Therefore, as a mobile's movement becomes more unpredictable, more frequent location inspections

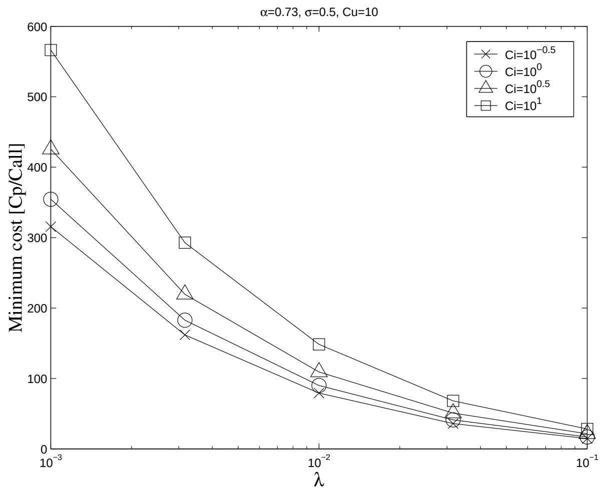

Fig. 5. Minimum cost vs. λ and C_u for $\beta = 10^{-0.5}$, $\sigma = 0.5$, and $C_u = 10$

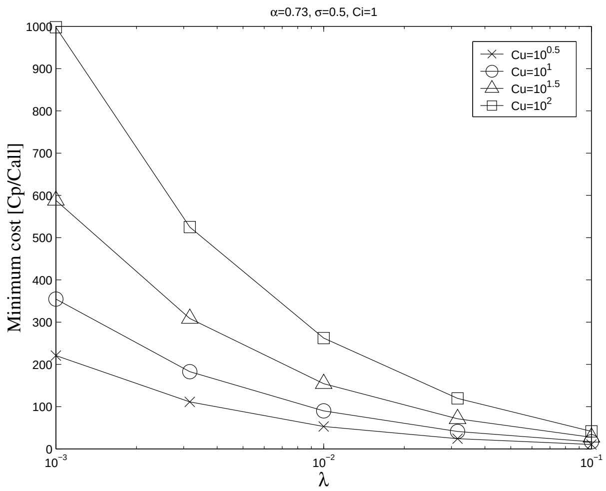

Fig. 7. Minimum cost vs. λ and C_u for $\beta = 10^{-0.5}$, $\sigma = 0.5$, and $C_i = 1$

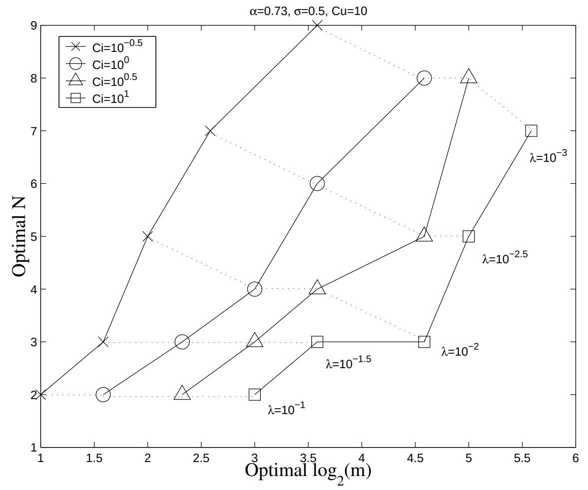

Fig. 6. Joint optimization of m and N vs. λ and C_i for $\beta = 10^{-0.5}$, $\sigma = 0.5$, and $C_u = 10$

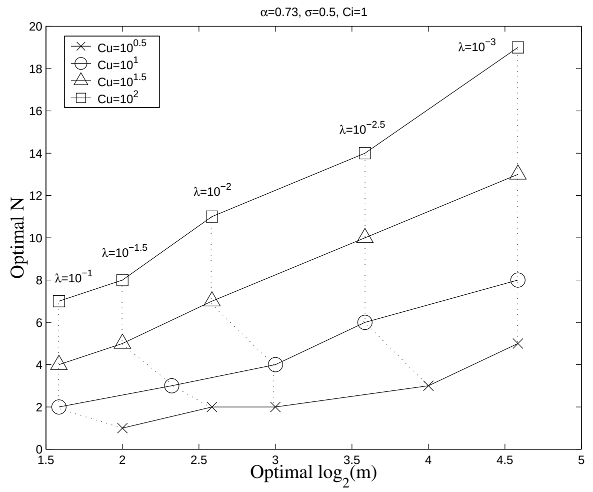

Fig. 8. Joint optimization of m and N vs. λ and C_u for $\beta = 10^{-0.5}$, $\sigma = 0.5$, and $C_i = 1$

and longer update threshold distance should be used.

Figures 5 and 6 present the plots of the minimum C_{total} and the corresponding (m_{opt}, N_{opt}), respectively, as λ increases along the values $\{10^{-3}, 10^{-2.5}, 10^{-2}, 10^{-1.5}, 10^{-1}\}$ [arrivals/time unit]. Here, C_u is fixed at 10, and for the various curves, C_i takes a value from $\{10^{-0.5}, 10^0, 10^{0.5}, 10^1\}$. Figures 7 and 8 present the plots of the minimum C_{total} and the corresponding (m_{opt}, N_{opt}), respectively, as λ increases along the values $\{10^{-3}, 10^{-2.5}, 10^{-2}, 10^{-1.5}, 10^{-1}\}$ [arrivals/time unit]. Here, C_i is fixed at 1, and for the various curves, C_u takes a value from $\{10^{0.5}, 10^1, 10^{1.5}, 10^2\}$. In all of these figures, $\beta = 10^{-0.5}$ (i.e., $\alpha = 0.73$), and $\sigma = 0.5$.

The plots in Figures 5 and 7 demonstrate that the minimum cost of mobility management per call arrival is approximately an exponentially decreasing function of the call arrival probability. Since a mobile's location is tracked during a call, as the inter-arrival time decreases, there is less unpredictable mobile random movement between calls, which leads to lower mobility management cost per call. As λ approaches 1 [arrivals/time unit], the total cost should approach $1[C_p]$. In this case no location inspection or location updating is necessary, and the only cost is incurred when a call arrives and the system pages once to verify that the mobile is there.

Figures 6 and 8 demonstrate that both the optimal location inspection period and the optimal updating distance decrease as the call arrival probability increases. This is intuitive since, as call arrives more often, a mobile is better off to check and update its location more often to reduce the cost of the frequent paging operation. From these figures, we also see that m_{opt} is more sensitive to C_i than N_{opt} is, and on the other hand, N_{opt} is more sensitive to C_u than m_{opt} is. For example, when $\lambda = 0.01$ [arrivals/time unit], as shown in Figure 6, as C_i goes from 1 to $10^{0.5}$, m_{opt} increases from 8 to 12, while N_{opt} remains unchanged at 4. However as shown in Figure 8, as C_u goes from $10^{1.5}$ to 10^2, N_{opt} increases from 7 to 11, while m_{opt} remains unchanged at 4. The relative insensitiveness of the optimal m and N to some of the system parameters is a welcomed property, especially when the proposed scheme is applied to a system where the cost factors and are time-varying and need to be estimated.

B. Comparison with the Non-predictive Distance-Based Scheme

For the non-predictive distance-based scheme, we use computer simulations to determine its cost when the system parameters take various combinations of values. In these simulations, we assume an infinite two-dimensional space that is divided into cells of size 1, where a mobile travels according to the two-dimensional isotropic Gauss-Markov process defined by the mobility parameters β, μ, and σ. The simulations are time-driven. At the time of initiation, the mobile is assumed to have just experienced a call arrival. Thus, it starts from the origin ($\mathbf{s}_0 = \mathbf{0}$), and has initial velocity with the Gaussian distribution defined by μ and σ. The time of the next call arrival is randomly generated following the exponential distribution with arrival rate λ. Until a call arrives, the mobile inspects its position every m clock ticks. If the mobile is N or more unit of distance away from the origin, a location update is performed, and the origin is shifted to the current location. When a call arrives, the paging cost is computed based on the mobile's distance from the origin, using equation (54). The updating cost is computed based on counting the total number of location updates performed since time initiation. The above experiment is repeated 10^5 times, and the average is taken.

To find the joint optimization of m and N through simulations in the non-predictive case would be too daunting a task. Instead, to facilitate the comparison between the two schemes, for both schemes, we treat m as a non-design parameter, and only optimize N. Thus,

$$N^p_{opt} = \arg\min_N C^p_{total}(N, \beta, \mu, \sigma, \lambda, C_u, C_i, m) \quad (62)$$

$$N^{np}_{opt} = \arg\min_N C^{np}_{total}(N, \beta, \mu, \sigma, \lambda, C_u, C_i, m)\ . \quad (63)$$

Furthermore, since both the predictive and non-predictive schemes have the same $C_{inspect}$, independent of N, we drop this term and define the performance gain

$$G(\beta, \mu, \sigma, \lambda, C_u, m) = \frac{C^{np}_{update}(N^{np}_{opt}) + C^{np}_{page}(N^{np}_{opt})}{C^{p}_{update}(N^{p}_{opt}) + C^{p}_{page}(N^{p}_{opt})}\ . \quad (64)$$

Thus, the predictive performance gain is a function of six independent variables. Instead of attempting to plot the performance gain in the six-dimensional space, we divide the variables into two groups, the mobility-related parameters (β, μ, σ) and the traffic and mobility-management parameters (λ, C_u, m). For each group of variables, we study the effect of these variable in detail, while the variables in the other group are fixed.

In Figures 9-11, we study the effect of the mobility pattern, namely, (β, μ, σ), on the performance gain. The other parameters are set to $(\lambda, C_u, m) = (0.01, 10, 10)$.

Figure 9 presents the plots of the performance gain versus β, for various values of μ, and with fixed $\sigma = 0.5$. Here β takes the values $\{10^{-2}, 10^{-1.5}, 10^{-1}, 10^{-0.5}, 10^0, 10^{0.5}\}$. This corresponds to the memory factor $\alpha = e^{-\beta}$ taking the values $\{0.99, 0.96, 0.90, 0.73, 0.37, 0.042\}$. For the various curves, μ takes a value from $\{0, 10^{-1}, 10^{-0.5}, 10^0\}$. These plots demonstrate the important fact that the performance gain is a convex function of β. On the one hand, when β is small, the user mobility has a high memory level, which favors the predictive scheme. On the other hand, at $\alpha = 0$, the variance of $\sum_{i=0}^{n} v_i$ reaches

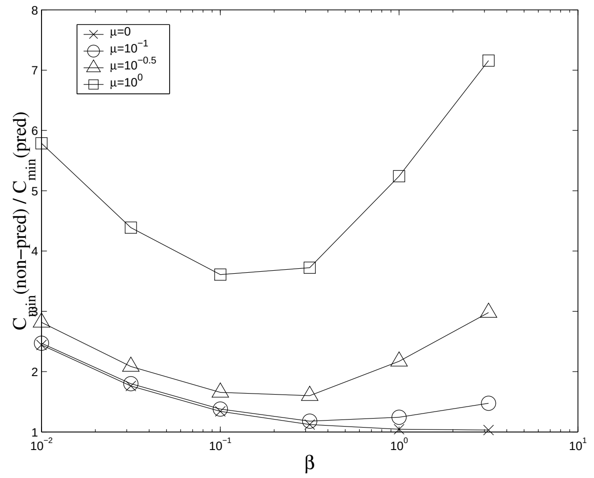

Fig. 9. Performance gain vs. β and μ for $\lambda = 0.01$, $C_u = 10$, $m = 10$, and $\sigma = 0.5$

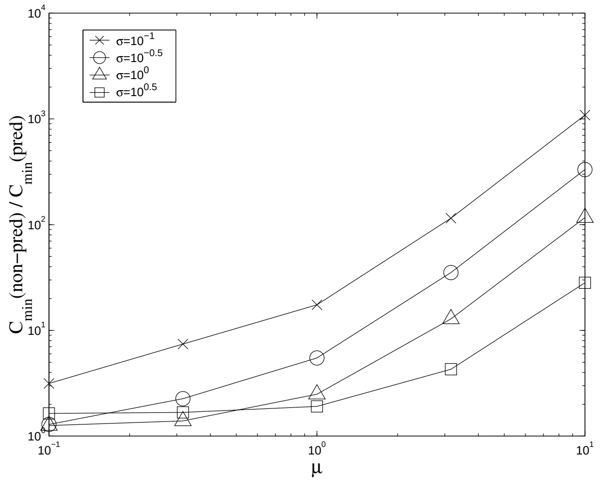

Fig. 10. Performance gain vs. μ and σ for $\lambda = 0.01$, $C_u = 10$, $m = 10$, and $\beta = 10^{-0.5}$ ($\alpha = 0.73$)

a local minimum, $n\sigma^2$. Thus, when β is large (i.e., α is small) and μ is not too small (larger than 0.1 in this case), the disadvantage of the non-predictive scheme is mainly determined by a mobile's average velocity as explained below. Therefore, in this case, since μ does not affect the cost of the predictive scheme, the predictive performance gain is larger for larger β. In other words, a smaller variance in the mobile velocity, as observed over the location inspection periods, reflects a more fluid-flow-liked movement pattern, which favors the predictive scheme.

When $\mu = 0$, the performance gain decreases from about 2.5 down to unity, as the memory factor of the system decreases from 0.99 to 0.042. In particular, for $\mu = 0$ and $\alpha \approx 0$, the mobility of the mobile has the pattern of random-walk. In this case, the predictive scheme does not have any advantage over the non-predictive one. However, in all other cases, the predictive scheme results in substantial savings.

Figure 10 presents the plots of the performance gain versus μ, for various values of σ, and with fixed $\beta = 10^{-0.5}$ ($\alpha = 0.73$). Here μ takes the values $\{10^{-1}, 10^{-0.5}, 10^0, 10^{0.5}, 10^1\}$. For the

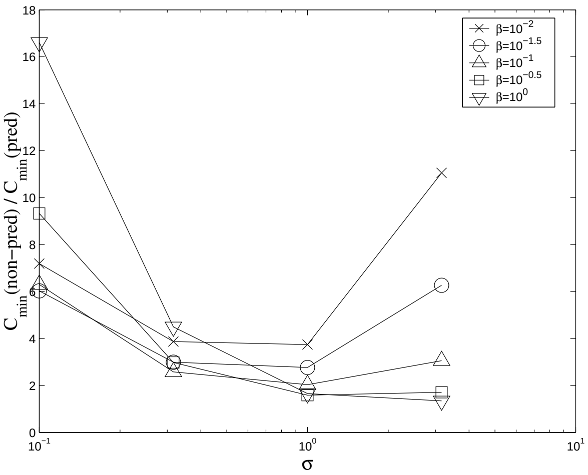

Fig. 11. Performance gain vs. σ and β for $\lambda = 0.01$, $C_u = 10$, $m = 10$, and $\mu = 0.5$

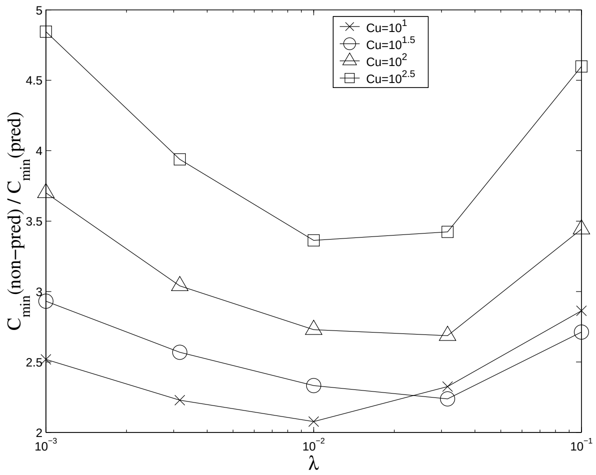

Fig. 12. Performance gain vs. λ and C_u for $\mu = 0.5$, $\sigma = 0.5$, $\beta = 10^{-0.5}$, and $m = 10$

various curves, σ takes a value from $\{10^{-1}, 10^{-0.5}, 10^{0}, 10^{0.5}\}$. These plots demonstrate that the performance gain is a faster-than-linearly increasing function of μ. Maximum savings are achieved when $\mu >> \sigma$, since, in this case, the mobile mobility pattern is close to the fluid-flow model, where a mobile's velocity and location are easily predictable. In this case, the mobile never needs to update its location, and therefore, the only cost incurred using the predictive scheme is the cost of paging once in the cell of the predicted location, when a call arrives, to verify that the mobile is indeed there. The performance gain can grow without bound as μ approaches infinity. For example, when $\sigma = 0.1$ and $\mu = 10$, a performance gain greater than 1000 can be achieved. However, when $\mu = 0.1$ and $\sigma > 0.1$, the performance gain is close to unity.

Figure 11 shows the plots of the performance gain versus σ, for various values of β, and with fixed $\mu = 0.5$. Here σ takes the values $\{10^{-1}, 10^{-0.5}, 10^{0}, 10^{0.5}\}$. For the various curves, β takes a value from $\{10^{-2}, 10^{-1.5}, 10^{-1}, 10^{-0.5}, 10^{0}\}$. These plots demonstrate that, when the memory level is high enough (e.g., $\alpha > 0.73$), the performance gain is a convex, non-monotonic function of σ. On the one hand, for small values of σ, as shown in the last figure, the mobility pattern become more like the fluid-flow model, and hence the predictive performance gain increases. On the other hand, for large σ, the advantage of using prediction becomes more prevalent, as long as the memory level is high. When the memory level is relatively low (e.g., $\alpha \leq 0.73$), the performance gain is approximately an exponentially decreasing function of σ.

In Figures 12-14, we study how the parameters (λ, C_u, m) affect the performance gain. We set the mobility parameters $(\beta, \mu, \sigma) = (10^{-0.5}, 0.5, 0.5)$.

Figure 12 presents the plots of the performance gain versus λ, for various values of C_u, and with fixed $m = 10$. Here λ takes the values $\{10^{-3}, 10^{-2.5}, 10^{-2}, 10^{-1.5}, 10^{-1}\}$ [arrivals/time unit]. For the various curves, C_u takes a value from $\{10^{1}, 10^{1.5}, 10^{2}, 10^{2.5}\}$. These plots demonstrate that the predictive performance gain is a convex function of λ. On the one hand, when the call arrival rate is low, the inter-arrival time becomes longer, and a mobile drifts farther away from the location where the last call arrived. This leads to higher updating and paging costs for the non-predictive scheme. On the other hand, when the call arrival rate is high, as shown in Section IV-A, the cost of the predictive scheme approaches one. In this case, the disadvantage of the non-predictive scheme, due to the non-zero mean velocity, becomes more dominant as λ becomes larger. In particular, the transition point is around $\lambda = 0.01$ for the given set of parameters.

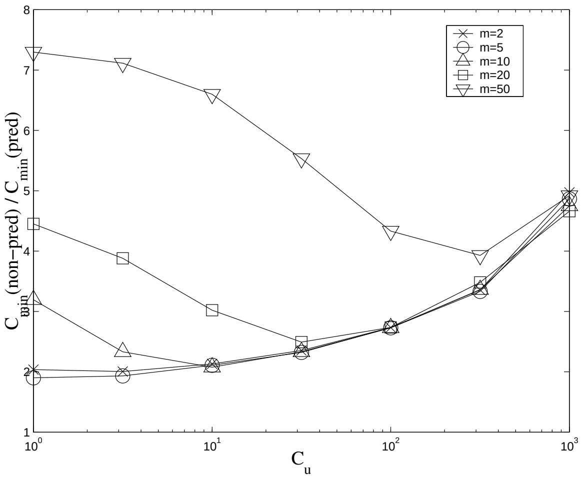

Fig. 13. Performance gain vs. C_u and m for $\mu = 0.5$, $\sigma = 0.5$, $\beta = 10^{-0.5}$, and $\lambda = 0.01$

Figure 13 shows the plots of the performance gain versus C_u, for the various values of m, and with fixed $\lambda = 0.01$. Here C_u takes the values $\{10^{0}, 10^{0.5}, 10^{1}, 10^{1.5}, 10^{2}, 10^{2.5}, 10^{3}\}$. For the various curves, m takes a value from $\{2, 5, 10, 20, 50\}$. These plots demonstrate that, when m is small (e.g., $m \leq 5$), the performance gain is an approximately linearly increasing function of the updating cost to paging cost ratio. However, when m is large (e.g., $m \geq 10$), the performance gain is a convex function of C_u (note that the C_u axis is in the logarithmic scale). As a call arrives in-between location inspections, a mobile may have

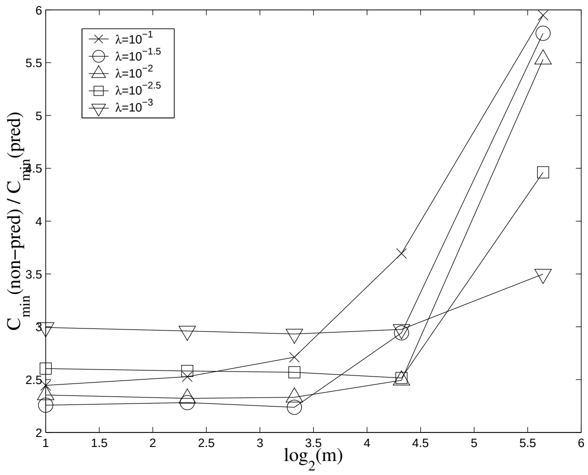

Fig. 14. Performance gain vs. m and λ for $\mu = 0.5$, $\sigma = 0.5$, $\beta = 10^{-0.5}$, and $C_u = 10^{1.5}$

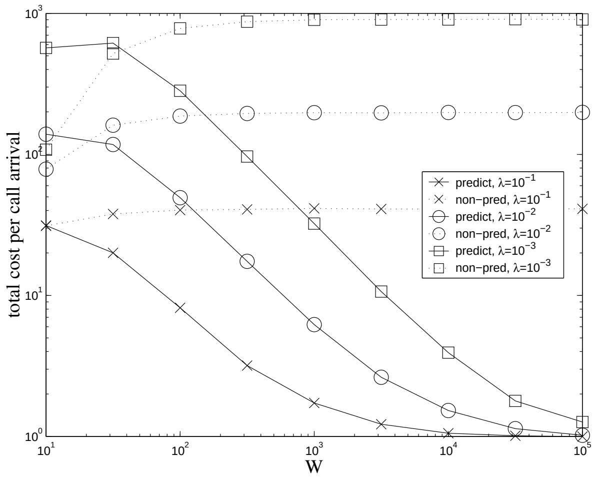

Fig. 15. Comparison of minimum (over N) total cost per call arrival, for $C_u = 10$, with the random way-point mobility pattern

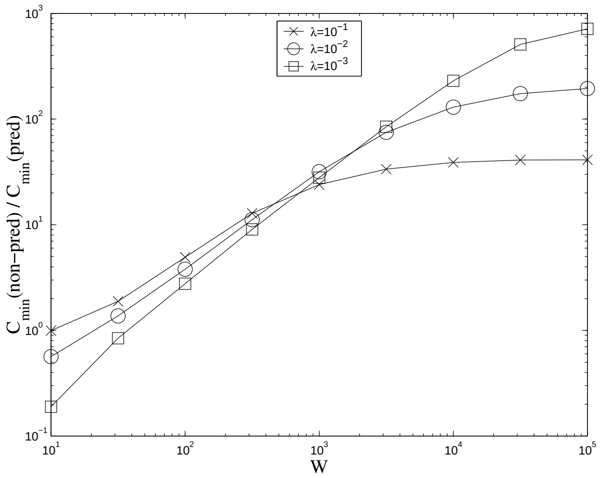

Fig. 16. Performance gain vs. W, for $C_u = 10$, with the random way-point mobility pattern

traveled beyond the updating distance. In the non-predictive scheme, the system needs to page in more cells than it does in the predictive scheme. This problem of the non-predictive scheme becomes more prominent when the paging cost to updating cost becomes larger, i.e., when C_u becomes smaller. Furthermore, the disadvantage of the non-predictive scheme is more apparent as m increases. This is studied in more detail in the next figure.

Figure 14 presents the plots of the performance gain versus m, for the various values of λ, and with fixed $C_u = 10^{1.5}$. Here m takes the values $\{2, 5, 10, 20, 50\}$. For the various curves, λ takes a value from $\{10^{-3}, 10^{-2.5}, 10^{-2}, 10^{-1.5}, 10^{-1}\}$ [arrivals/time unit]. These plots demonstrate that the performance gain is an increasing function of the location inspection period. This shows that the predictive scheme improves the performance of a system where the location inspection cost is relatively high.

C. Dynamic Gauss-Markov Parameter Estimation

We use simulation to study the performance of the proposed scheme in systems where the mobility pattern cannot be ideally represented by the Gauss-Markov model. We simulated the two-dimensional, random way-point mobility model as described in Section II-A.3. The time-varying Gauss-Markov parameters are estimated as described in Section II-C and applied to the predictive scheme as described in Section II-B.

Figures 15 and 16 illustrates the effectiveness of combining dynamic parameter estimation and predictive mobility management. In both figures, the parameter estimation window is $w_{est} = 10$. The call arrival rate λ takes the values $\{10^{-1}, 10^{-2}, 10^{-3}\}$. The cost per location update is assumed to be $C_u = 10[C_p]$.[6] We further assume that $m = 1$. In the random way-point model, we assume that the mobile velocity has unit magnitude with zero pause time (i.e., when a mobile reaches a destination point, it immediately moves toward a new destination). The system coverage area is assumed to be a square with side length W. Clearly, the larger the W, the more predictable is the mobile movement. In particular, when W approaches infinity, a mobile appears to move with constant velocity with a pre-selected random direction.

Figure 15 plots the minimum total cost (obtained with the optimal location threshold) of location updating and mobile paging, per call arrival, for both the predictive scheme and the non-predictive scheme. This figure shows that, as W increases, the mobile movement approaches the constant-velocity fluid-flow model, and the cost of the non-predictive scheme approaches its maximum value. This maximum value can be determined analytically, since when a mobile moves with unit constant velocity,

$$C_{total}^{np}(N) = \int_0^\infty \lambda e^{-\lambda t} \left(C_u \frac{t}{N} + \frac{1}{N} \int_0^N \pi s^2 ds \right) dt \ . \quad (65)$$

It is easy to optimize N in the above and to obtain the minimum cost of the non-predictive scheme $C_{total}^{np}(N_{opt}^{np}) = 1.92 \left(\frac{C_u}{\lambda}\right)^{\frac{2}{3}}$. The simulation model can be verified by substituting the values of C_u and λ into the above equation and comparing the results with the simulation results.

[6] Simulation results with other C_u values yield similar results and are omitted here.

In contrast, as W increases, the cost of the predictive scheme decreases. As expected, this cost approaches $1[C_p]$ when W is very large. This is the case where the mobile location is perfectly predictable, and, hence, the best scheme is not to update at all, and each time a call arrives, the system pages the mobile once.

Figure 16 plots the performance gain of the predictive scheme as defined in (64). This figure shows that when W is not too large, the performance gain is approximately linearly increasing with W. When W is very large, as expected from the results shown in the previous figure, the performance gain levels off and approaches its global maximum. As a point of reference, for the moderate case where $C_u = 10$ and $\lambda = 0.01$, the predictive scheme outperforms the non-predictive scheme by a maximal factor of 200!

It is worth noting that, in the case of dynamic Gauss-Markov parameter estimation, the predictive performance gain can degrade to less than 1 when W is very small. This is in contrast with the ideal case shown in IV-B, where the predictive scheme performs as well as the non-predictive scheme even when the mobile movement is memoryless. This performance degradation is a direct result of our sub-optimal parameter estimation algorithm. For example, if we had *a priori* knowledge (e.g., through long-term observation) that the mobile movement follows the random way-point pattern, a better parameter estimation using $w_{est} = 1$ could be employed. For systems with general mobility patterns, improving the parameter estimation algorithm would further improve the performance gain of the predictive scheme beyond what is shown in Figure 16.

V. Conclusions

Mobile users in PCS networks move with wide variety of mobility patterns, especially in networks with multi-layer, macrocellular and micro-cellular infrastructures [10]. The fixed location-area-based mobility management schemes are not flexible enough to adapt to various and varying user traffic and mobility patterns. Furthermore, most of the existing dynamic mobility management schemes assume either memoryless random-walk or constant velocity fluid-flow as the user mobility model. However, neither of these mobility models can practically represent the random and correlated mobile movement in PCS networks. In this paper, we have considered the time-correlation in a mobile user's movement. We have presented a novel predictive distance-based mobility management scheme, which takes full advantage of the correlation between a mobile's current velocity and location and its future velocity and location. We have introduced a mobility model based on the Gauss-Markov random process, which captures the various degrees of velocity correlation in time. It even includes both the random-walk and fluid-flow models as its two extreme cases. An analytical framework is introduced to evaluate the performance of the proposed scheme. The analytical framework enables us to study the effects of various parameters on the mobility management cost, to optimize the location inspection frequency and location updating threshold through numerical analysis, and to determine the relative performance of the proposed scheme over the non-predictive scheme.

The numerical results have demonstrated the important fact that the performance gain of the predictive scheme is not an increasing function of the memory level of a mobile's movement. This validates our observation that the *memory level* does not directly indicate the *predictability* of a mobile's movement. Overall, however, the numerical analysis confirms our intuition that the cost of predictive mobility management is inversely proportional to the predictability of a mobile's mobility pattern. In particular, our numerical results suggest that this cost is a monotonic function of the velocity variance and the call arrival rate, that it is a concave function of the memory level, and that it is independent of the mean velocity. Correspondingly, the optimal location inspection period and the optimal updating distance are monotonic functions of the velocity variance and the call arrival rate, but they are, respectively, convex and concave functions of the memory level. In the span of parameter values under consideration, the performance improvement of the predictive scheme ranges from unity, in the random-walk case, to a factor of a few orders of magnitude, in the constant velocity case. As a point of reference, as shown in Figure 10, in the near fluid-flow case, where $\sigma = 0.1$, the performance gain quickly increases from 8, when $\mu = 0.3$, to 100, when $\mu = 3$. In more general terms, the mobility management cost can be reduced by more than 50% for all systems where the mobiles have moderate mean velocity (e.g., $\mu \geq 0.5$), and where performing a location update by a mobile is as least as expensive as paging a mobile in a cell (e.g., $C_u \geq 1$).

Our analysis using ideal Gauss-Markov mobility pattern has demonstrated the potentially enormous advantage of using prediction in mobility management. For practical systems with time-varying mobility patterns, we have introduced a dynamic parameter estimation algorithm to be employed in conjunction with the proposed predictive scheme. Simulation results of systems with the random way-point mobility model have demonstrated that the proposed predictive scheme can largely outperform the non-predictive scheme, even with a sub-optimal parameter estimator.

The vast performance gain of predictive mobility management, as shown by both analysis and simulation, offsets the cost of enabling prediction in the mobile terminal and the wireline backbone. This justifies the investment of such software and hardware devices in the next generation wireless personal communication networks.

APPENDIX A: DERIVATION OF $C_{inspect}$ WITH EXPONENTIALLY DISTRIBUTED CALL INTER-ARRIVAL TIME

In this case, the PDF of the call inter-arrival time is $f_{call}(n) = \lambda(1-\lambda)^n$. From equation (17), we have

$$\begin{aligned} C_{inspect} &= C_i \sum_{n=0}^{\infty} \lambda(1-\lambda)^n \lfloor n/m \rfloor \\ &= C_i \sum_{i=0}^{\infty} \sum_{j=im}^{(i+1)m-1} \lambda(1-\lambda)^j \lfloor j/m \rfloor \\ &= C_i \sum_{i=0}^{\infty} i \sum_{j=im}^{(i+1)m-1} \lambda(1-\lambda)^j \\ &= C_i \left(1-(1-\lambda)^m\right) \sum_{i=0}^{\infty} i(1-\lambda)^{im} \\ &= C_i \frac{(1-\lambda)^m}{1-(1-\lambda)^m} . \end{aligned} \tag{66}$$

APPENDIX B: DERIVATION OF $f_{y_0|(u_0,u'_m)}(y,\xi_1,\xi_2)$

Since $u_0, u_1, \dots, u_{m-1}, u'_m$ are jointly Gaussian, $y_0 = \sum_{i=0}^{m-1} u_i$ is jointly Gaussian along with $u_0, u_1, \dots, u_{m-1}, u'_m$. With slight abuse of notation, we have

$$\begin{aligned} f_{y_0|(u_0,u'_m)}(y,\xi_1,\xi_2) &= \frac{f_{(y_0,u'_m)|u_0}(y,\xi_1,\xi_2)}{f_{u'_m|u_0}(\xi_2,\xi_1)} \\ = f_{(y_0|u_0)|(u'_m|u_0)}(y,\xi_1,\xi_2) &= \frac{1}{\sigma_{y_0|u_0}\sqrt{2\pi(1-\gamma^2)}} \\ \cdot \exp &\left\{ - \frac{\left[(y-\mu_{y_0|u_0}) - \gamma \frac{\sigma_{y_0|u_0}}{\sigma_{u'_m|u_0}}(\xi_2 - \mu_{u'_m|u_0})\right]^2}{2\sigma^2_{y_0|u_0}(1-\gamma^2)} \right\} , \end{aligned} \tag{67}$$

where $\mu_{u'_m|u_0}$ and $\mu_{y_0|u_0}$ are the means of $u'_m|u_0$ and $y_0|u_0$, respectively, $\sigma_{u'_m|u_0}$ and $\sigma_{y_0|u_0}$ are the standard deviations of $u'_m|u_0$ and $y_0|u_0$, respectively, and γ is the correlation factor of $y_0|u_0$ and $u'_m|u_0$.

From the definition of u'_m, we have

$$\mu_{u'_m|u_0} = \alpha^m u_0 , \tag{68}$$

and

$$\begin{aligned} \sigma^2_{u'_m|u_0} &= \sigma^2(1-\alpha^2) \sum_{i=0}^{m-1} \alpha^{2(m-i-1)} \\ &= (1-\alpha^{2m})\sigma^2 . \end{aligned} \tag{69}$$

The mean of y_0 given u_0 is

$$\mu_{y_0|u_0} = \sum_{i=0}^{m-1} \mu_{u_i|u_0} = \sum_{i=0}^{m-1} \alpha^i u_0 = \frac{1-\alpha^m}{1-\alpha} u_0 . \tag{70}$$

To find a formula for $\sigma_{y_0|u_0}$, we first consider the covariance between $u_i|u_0$ and $u_j|u_0$. For any i and j satisfying $0 \le i < j \le m-1$, we have

$$\begin{aligned} &\mathrm{E}[(u_i|u_0 - \mu_{u_i|u_0})(u_j|u_0 - \mu_{u_i|u_0})] \\ =&\mathrm{E}\left[\sigma^2(1-\alpha^2)\sum_{l=0}^{i-1}\alpha^{i-l-1}w_l \sum_{l=0}^{j-1}\alpha^{j-l-1}w_l\right] \\ =&\sigma^2(1-\alpha^2)\sum_{l=0}^{i-1}\alpha^{i-l-1}\alpha^{j-l-1} \\ =&\sigma^2(1-\alpha^{2i})\alpha^{j-i} . \end{aligned} \tag{71}$$

Then,

$$\begin{aligned} \sigma^2_{y_0|u_0} &= \mathrm{E}\left[\left(\sum_{i=0}^{m-1} u_i|u_0 - \mu_{u_i|u_0}\right)^2\right] \\ &= \sum_{i=0}^{m-1} \mathrm{E}\left[(u_i|u_0 - \mu_{u_i|u_0})^2\right] + 2\sum_{i=0}^{m-1}\sum_{j=i+1}^{m-1} \\ &\qquad \mathrm{E}\left[(u_i|u_0 - \mu_{u_i|u_0})(u_j|u_0 - \mu_{u_j|u_0})\right] \\ &= \sum_{i=0}^{m-1} \sigma^2_{u_m|u_0} + 2\sum_{i=0}^{m-1}\sum_{j=i+1}^{m-1} \sigma^2(1-\alpha^{2i})\alpha^{j-i} \\ &= \frac{-(1-\alpha^m)^2 + m(1-\alpha^2) - 2\alpha(1-\alpha^m)}{(1-\alpha)^2}\sigma^2 . \end{aligned} \tag{72}$$

Finally, the covariance factor between u'_m and y_0 is

$$\gamma = \frac{C_{(u'_m,y_0)|u_0}}{\sigma^2_{u'_m|u_0}\sigma^2_{y_0|u_0}} , \tag{73}$$

where

$$\begin{aligned} C_{(u'_m,y_0)|u_0} &= \mathrm{E}\left[\left(u'_m|u_0 - \mu_{u'_m|u_0}\right)\left(\sum_{i=0}^{m-1} u_i|u_0 - \mu_{u_i|u_0}\right)\right] \\ &= \sum_{i=0}^{m-1} \mathrm{E}\left[(u'_m|u_0 - \mu_{u'_m|u_0})(u_i|u_0 - \mu_{u_i|u_0})\right] \\ &= \sum_{i=0}^{m-1} \sigma^2(1-\alpha^{2i})\alpha^{m-i} \\ &= \frac{(1-\alpha^m)(\alpha-\alpha^m)}{1-\alpha}\sigma^2 . \end{aligned} \tag{74}$$

REFERENCES

[1] EIA/TIA, "Cellular radio-telecommunications intersystem operation: Automatic roaming," Technical Report IS-41.3-B, EIA/TIA, 1991

[2] I. F. Akyildiz and J. S. M. Ho, "Dynamic Mobile User Location Update for Wireless PCS Networks," *ACM/Baltzer Wireless Networks*, vol. 1, no. 2, pp. 187-196, 1995

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

[4] I. F. Akyildiz et al, "Mobility Management in Next-Generation Wireless Systems," *Proceedings of the IEEE*, vol. 87, no. 8, pp. 1347-1384, August 1999.

[5] A. Bar-Noy, I. Kessler, and M. Sidi, "Mobile Users: To Update or Not to Update?" *ACM/Baltzer Wireless Net.,* vol. 1, no. 2, pp. 175-185, 1995

[6] A. Bhattacharya and S. K. Das, "LeZi-Update: An Information-Theoretic Approach to Track Mobile Users in PCS Networks," *Proceedings of ACM/IEEE Mobicom*, August 1999.

[7] J. Broch, D. A. Maltz, D. B. Johnson, Y.-C. Hu, and J. Jetcheva, "A performance comparison of multi-hop wireless ad hoc network routing protocols," *Proceedings of the ACM/IEEE MOBICOM*, 1998.

[8] T. X. Brown and S. Mohan "Mobility Management for Personal Communication Systems," *IEEE Transactions on Vehicular Technology*, vol. 46, no. 2, pp. 269-278, May 1997

[9] S. R. Das, C. E. Perkins, and E. M. Royer, "Performance comparison of two on-demand routing protocols for ad hoc networks," *Proceedings of the IEEE INFOCOM*, 2000.

[10] A. Ganz, Z. J. Haas, C. M. Krishna, and D. Tang, "On Optimal Design of Multi-Tier Wireless Cellular Systems," *IEEE Communications Magazine,* pp. 88-93, February 1997

[11] A. Gelb, *Applied Optimal Estimation*, The M.I.T Press, 1974

[12] Z. J. Haas and Y.-B. Lin, "On Optimizing the Location Update Costs in Presence of Database Failures," *ACM/Baltzer Wireless Networks,* vol. 4, no. 5, pp. 419-426, 1998

[13] M. Hellebrandt, R. Mathar, and M. Scheibenbogen, "Estimating Position and Velocity of Mobiles in a Cellular Radio Network," *IEEE Transactions on Vehicular Technology*, vol. 46, no. 1, pp. 65-71, February 1997

[14] J. S. M. Ho and I. F. Akyildiz, "Mobile User Location Update and Paging Under Delay Constraints," *ACM/Baltzer Wireless Networks*, vol. 1, no. 4, pp. 413-425, 1995

[15] L.-R. Hu and S. S. Rappaport, "Adaptive Location Management Scheme for Global Personal Communications," *IEEE Proceedings on Communications*, vol. 144, no. 1, pp. 54-60, 1997

[16] Y.-C. Hu and D. B. Johnson, "Caching Strategies in On-Demand Routing Protocols for Wireless Ad Hoc Network," *Proceedings of the ACM/IEEE Mobicom*, 2000.

[17] K.-R. Koch, *Parameter Estimation and Hypothesis Testing in Linear Models*, Springer-Verlag, 1999.

[18] Z. Lei and C. Rose, "Wireless Subscriber Mobility Management Using Adaptive Individual Location Areas for PCS Systems," *Proceedings of IEE ICC*, vol. 3, pp. 1390-1394, June 7-11, 1998

[19] 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, June 2000.

[20] B. Liang and Z. J. Haas, "Predictive Distance-Based Mobility Management for PCS Networks," *Proceedings of the IEEE Infocom*, New York, NY, March 1999.

[21] U. Madhow, M. L. Honig, and K. Steiglitz, "Optimization of Wireless Resources for Personal Communications Mobility Tracking," *IEEE/ACM Transactions on Networking*, vol. 3, no. 6, pp. 698-707, December 1995

[22] S. Mohan and R. Jain, "Two User Location Strategies for Personal Communications Services," *IEEE Personal Comm.,* First Quarter, 1994

[23] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, Third Edition, McGraw-Hill, 1991

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

[25] C. Rose, "Minimizing the Average Cost of Paging and Registration: A Timer-Based Method," *ACM/Baltzer Wireless Networks*, vol. 2, no. 2, pp. 109-116, 1996.

[26] C. Rose, "State-Based Paging/Registration: A Greedy Technique," *IEEE Transactions. on Vehicular Technology*, vol. 48, no. 1, pp. 166-173, January 1999.

[27] D. Tang and M. Baker, "Analysis of a Metropolitan-Area Wireless Network," *Proceedings of the Fifth ACM/IEEE MobiCom*, 1999.

[28] D. Tang and M. Baker, "Analysis of a Local-Area Wireless Network," *Proceedings of the Sixth ACM/IEEE MobiCom*, 2000.

[29] E. Wong and B. Hajek, *Stochastic Processes in Engineering Systems*, Springer-Verlag, 1985

[30] V. W. S. Wong and V. C. M. Leung, "Location Management for Next-Generation Personal Communications Networks," *IEEE Network*, pp. 18-24, September/October 2000.

[31] V. W. S. Wong and V. C. M. Leung, "An Adaptive Distance-Based Location Update Algorithm for Next-Generation PCS Networks," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 10, pp. 1942-1952, October 2001.

[32] H. Xie, S. Tabbane, and D. J. Goodman, "Dynamic Location Area Management and Performance Analysis," *Proceedings of 43rd IEEE VTC*, Secaucus, NJ, pp. 536-539, May 1993

[33] H. Xie and D. J. Goodman, "Mobility Models and Biased Sampling Problem," *IEEE ICUPC*, Ottawa, Canada, vol. 2, pp. 803-807, October 1993