IEEE INFOCOM'99, New York, NY, March 21-25, 1999

Predictive Distance-Based Mobility Management for PCS Networks

Ben Liang and Zygmunt J. Haas
School of Electrical Engineering, Cornell University, Ithaca, NY 14853
e-mail: liang@ee.cornell.edu, haas@ee.cornell.edu

Abstract— **This paper presents a mobile tracking scheme that exploits the predictability of user mobility patterns in wireless PCS networks. Instead of the constant velocity fluid-flow or the random-walk mobility model, a more realistic Gauss-Markov model is introduced, where a mobile's velocity is correlated in time to a various degree. Based on the Gauss-Markov model, a mobile's future location is predicted by the network based on the information gathered from the mobile's last report of location and velocity. When a call is made, the network pages the destination mobile at and around the predicted location of the mobile and in the order of descending probability until the mobile is found. A mobile shares the same prediction information with the network and reports its new location whenever it reaches some threshold distance away from the predicted location. We describe an analytical framework to evaluate the cost of mobility management for the proposed predictive distance-based scheme. We then compare this cost against that of the regular, non-predictive distance-based scheme, which is obtained through simulations. Performance advantage of the proposed scheme is demonstrated under various mobility and call patterns, update cost, page cost, and frequencies of mobile location inspections.**

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 ([12]), 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. Although dynamic LA management is possible ([16]), LA-based schemes, in general, are not flexible enough to adapt to different and differing user traffic and mobility patterns.

Dynamic mobility management schemes ([4] – [5]) discard the notion of LA borders. A mobile in these schemes updates its location 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. When the LA-s are not defined, upon call arrival, the network pages the destination mobile using a selective paging scheme ([14]), 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 the same as a largest-probability-first algorithm, which 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 memoryless random-walk mobility pattern, the distance-based scheme has been proven to result in less mobility management cost (location update cost plus paging cost) than schemes based on time or number of cell boundary crossings ([4]).

However, in practical systems, a mobile user usually travels with a destinations in mind, therefore mobile's location and velocity in the future are likely to be correlated with its current location and velocity. 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 ([16] and [17]). The fluid-flow model is suitable for vehicle traffic in highways, but not pedestrian movements with frequent stop-and-go interruptions. A discrete Markovian model is reported in ([4]). However, in this model, the velocity of the mobiles is overly simplified and characterized by three states only. In this paper, we introduce a Gauss-Markov ([15] and [13]) mobility model, which captures the essence of the correlation of a mobile's velocity in time. The Gauss-Markov model represents a wide range of user mobility patterns, including, as the two extreme cases, the random-walk and the constant velocity fluid-flow models.

In systems with correlated velocity mobility patterns, unlike those with random-walk mobility patterns, the largest-probability location of a mobile is generally not the cell where the mobile last reported. Thus a mobility management scheme that takes advantage of the predictability of the mobiles' location can perform better.

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, in turn, given by the Gauss-Markov model based on its location and velocity at the time of the last location update. 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 in time. The mobile checks its position periodically (*location inspection*) and performs location update whenever it reaches some threshold distance (*update distance*) away from the predicted location. To locate a mobile, the network pages the mobile starting from the predicted location and outwards, in a shortest-distance-first order, and until the mobile is found.

The cost of mobility management is defined as the sum of a mobile's location update cost and the cost incurred in paging the mobile. We derive an analytical framework to evaluate the mobility management cost of the predictive distance-based scheme, which is a function of the traffic and mobility patterns, the update distance, the relative cost of location update versus paging, and the location inspection frequency. The cost of the non-predictive distance-based scheme is evaluated through simulations, and is compared with the proposed predictive scheme. We find the performance gains of the predictive scheme based on the optimal updating distance that results in the minimum mobility management cost.

In Section II, we describe the Gauss-Markov mobility model and the prediction algorithm, showing that the proposed scheme follows the rule of largest-probability-first. Section III presents the analytical framework for evaluating the predictive scheme. The numerical results based on the analysis are presented in Section IV, where we compare the mobility management costs of both schemes and show the performance gains achieved by prediction. Finally, the concluding remarks are discussed in Section V.

II. System Description

A. The Gauss-Markov Mobility Model

We consider here a one-dimensional cellular system to demonstrate the cost-effectiveness of the predictive scheme. The multi-dimensional extension to this mobility model and the associated analysis framework can be developed similarly by substituting a vector of random processes for the single random process used in this paper.

A mobile's velocity is assumed to be correlated in time and modeled by a Gauss-Markov process. In continuous-time, a stationary Gauss-Markov process is described by the autocorrelation function ([15])

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

where σ^2 is the variance, and $\beta \geq 0$ determines the degree of memory in the mobility pattern. Equation (1) is also sometimes called the Ornstein-Uhlenbeck solution of the Brownian motion with zero restoring force ([13].

We define a discrete version of the mobile velocity with

$$v_n = v(n\Delta t) , \qquad (2)$$

and

$$\alpha = e^{-\beta \Delta t} , \qquad (3)$$

where Δt is the clock-tick period (normalized to 1 throughout this paper). Then the discrete representation of (1) is ([7])

$$v_n = \alpha v_{n-1} + (1-\alpha)\mu + \sqrt{1-\alpha^2}\, x_{n-1} , \qquad (4)$$

where $0 \leq \alpha \leq 1$, μ is the asymptotic mean of v_n when n approaches infinity, and x_n is an independent, uncorrelated, and stationary Gaussian process, with mean $\mu_x = 0$ and standard deviation $\sigma_x = \sigma$, where σ is the asymptotic standard deviation of v_n when n approaches infinity.

Define the initial $n = 0$ as the time when a mobile last updates its location and velocity. We can recursively expand (4) to express v_n explicitly in terms of the initial velocity v_0,

$$v_n = \alpha^n v_0 + (1-\alpha^n)\mu + \sqrt{1-\alpha^2} \sum_{i=0}^{n-1} \alpha^{n-i-1} x_i . \qquad (5)$$

Define s_n as the displacement of a mobile, at time n, from its last updated location. By definition, $s_0 = 0$. Furthermore,

$$s_n = \sum_{i=0}^{i=n-1} v_i . \qquad (6)$$

B. Mobility Tracking

In most practical systems, a mobile cannot continuously monitor its location or velocity[1]. Assuming that each mobile performs location inspection periodically, the optimal value of this location inspection frequency is a variable, which depends on the cost of the location and velocity checking process, which, in turn, depend upon many other factors such as the tracking and paging methods involved, the communication channel usage, and the computational power of the mobile's CPU.

Suppose the mobile examines its location every m clock ticks. We define

$$y_k = \sum_{i=km}^{km+m-1} v_i . \qquad (7)$$

Then

$$\begin{aligned} y_k &= \sum_{i=km}^{km+m-1} \left(\alpha^i v_0 + (1-\alpha^i)\mu + \sqrt{1-\alpha^2} \sum_{j=0}^{i-1} \alpha^{i-j-1} x_j\right) \\ &= \alpha^{km} \frac{1-\alpha^m}{1-\alpha} v_0 + \left(m - \frac{1-\alpha^m}{1-\alpha}\right)\mu \qquad (8) \\ &\quad + \sqrt{1-\alpha^2} \sum_{i=km}^{km+m-1} \sum_{j=0}^{i-1} \alpha^{i-j-1} x_j . \end{aligned}$$

Since x_j above is Gaussian with zero mean, for any constant v_0, y_k is a Gaussian process with mean

$$\mu_{y_k} = \alpha^{km} \frac{1-\alpha^m}{1-\alpha} v_0 + \left(m - \frac{1-\alpha^m}{1-\alpha}\right)\mu . \qquad (9)$$

Thus, a mobile's location displacement from its last updated location at its k^{th} location inspection since the last location update is

$$s_{km-1} = \sum_{i=0}^{k-1} y_i , \qquad (10)$$

[1]We will not address in this paper the exact mechanism by which a mobile monitors its location and velocity. One possibility is for a mobile to use base-station beaconing signals to determine its position, and to average position displacement over time to find its velocity. Other methods and related references on mobile location and velocity determination can be found in ([9]).

which is also a Gaussian random variable with mean

$$\mu_{s_k} = \sum_{i=0}^{k-1} \mu_{y_i} \ . \tag{11}$$

In the regular, non-predictive distance-based mobility management scheme, a mobile transmits a location update to the PCS network at the k^{th} location inspection if $|s_{km-1}|$ is greater than a distance threshold, N. 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.

In the proposed predictive distance-based scheme, a mobile reports both its location and velocity as part of the location update process. Thus, both the PCS network and the mobile make the same prediction of the probability distribution of s_{km-1} for any k. A mobile transmits a location update to the PCS network at the k^{th} location inspection if $|s_{km-1} - \mu_{s_k}|$ is greater than a distance threshold , N. When a call is made to a mobile, the system pages for the mobile in cells at and around μ_{s_k}, in a shortest-distance first order, and until the mobile is found. Here, we have assumed that the call arrival intervals are much greater than the location inspection interval, such that we can assume that calls arrive at the end of location inspection intervals.

Let s'_{km-1} be a random variable representing a mobile's location displacement, at the k^{th} location inspection, from its last reported location, given that the mobile has not performed location update up to the k^{th} location inspection. Since, without consideration of location updates, s_{km-1} is Gaussian with mean μ_{s_k}, and the location update process affects the probability density function (PDF) of a mobile's location symmetrically around μ_{s_k}, the optimal prediction of s'_{km-1}, in terms of the largest probability, is still μ_{s_k}. Furthermore, since the probability density of s'_{km-1} ramps down symmetrically around μ_{s_k}, the predictive distance-based paging scheme is a largest-probability-first scheme, which satisfies the requirement of minimum cost selective paging. Therefore, we can expect the predictive scheme to incur lower mobility management cost than the non-predictive distance-based scheme[2].

In the next section, we introduce an analytical framework to evaluate the mobility management cost of the predictive distance-based scheme.

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

A. *PDF of Time Interval between Two Consecutive Autonomous Location Updates*

Since within a phone call duration, the position of a mobile is closely monitored, a call arrival has the same effect as a mobile location update. Here we distinguish a location update based on distance as *autonomous update*. We first consider the time interval between two consecutive autonomous location updates without the interruption of phone calls.

Shifting the center of the PDF of y_k and s_{km-1} to the origin, we define

$$\begin{aligned} w_k &= y_k - \mu_{y_k} && (12)\\ r_k &= s_{km-1} - \mu_{s_k} \ . && (13) \end{aligned}$$

Then, in the predictive distance-based scheme, the mobile update condition at the k^{th} location inspection becomes

$$|r_k| > N \ , \tag{14}$$

and the following recursive equation holds between location updates:

$$r_k = r_{k-1} + w_{k-1} \ . \tag{15}$$

Next, we derive a recursive formula to find w_k. Defining $C_{i,j}$ as the auto-covariance of w_k,

$$C_{i,j} = \mathrm{E}[w_i w_j] \ , \tag{16}$$

the PDF of w_k is completely determined by $C_{k,k}$.

Furthermore, we have the following recursive relation

$$w_k = \frac{C_{k-1,k}}{C_{k-1,k-1}} w_{k-1} + \sqrt{1 - \frac{C_{k-1,k}^2}{C_{k-1,k-1}C_{k,k}}}\, z_{k-1} \ , \tag{17}$$

where z_k is a Gaussian process, independent of w_k and with variance equal to $C_{k,k}$. Equation (17) can be justified by verifying that

$$\mathrm{E}[w_k] = 0 \ , \tag{18}$$

$$\begin{aligned} \mathrm{E}[w_k w_{k-1}] &= \frac{C_{k-1,k}}{C_{k-1,k-1}} C_{k-1,k-1} && (19)\\ &= C_{k-1,k} \ , \end{aligned}$$

and

$$\begin{aligned} \mathrm{E}[w_k^2] &= \frac{C_{k-1,k}^2}{C_{k-1,k-1}^2} C_{k-1,k-1} + \left(1 - \frac{C_{k-1,k}^2}{C_{k-1,k-1}C_{k,k}}\right) C_{k,k} \\ &= C_{k,k} \ . && (20) \end{aligned}$$

The auto-covariance of w_k in (17) can be computed as follows. Since

$$\begin{aligned} C_{k,k} &= \mathrm{E}[y_k^2] - \mu_{y_k}^2 && (21)\\ &= \mathrm{E}\left[\left(\sqrt{1-\alpha^2} \sum_{i=km}^{km+m-1} \sum_{j=0}^{i-1} \alpha^{i-j-1} x_j\right)^2\right] \ , \end{aligned}$$

exchanging the order of the above double summation, and then dividing it into two independent parts, we have

$$C_{k,k} = (1-\alpha^2)\mathrm{E}[(A+B)^2] \ , \tag{22}$$

[2]Paging delay constraints and reliability considerations are not addressed here. For systems with paging delay constraints, the proposed scheme can be easily extended using the methods similar to those reported in ([10], [3]), and ([14]). For reliability considerations, methods similar to those described in ([8] could be used.

where

$$A = \sum_{j=0}^{km-1} x_j \alpha^{-j-1} \sum_{i=km}^{km+m-1} \alpha^i \, ;$$

$$B = \sum_{j=km}^{km+m-2} x_j \alpha^{-j-1} \sum_{i=j+1}^{km+m-1} \alpha^i \, . \quad (23)$$

Since both A and B are summations of independent zero-mean Gaussian variables,

$$\mathrm{E}[A^2] = \sum_{j=0}^{km-1} \sigma^2 \alpha^{-2j-2} \left(\sum_{j=km}^{km+m-1} \alpha^i \right)^2 , \quad (24)$$

and

$$\mathrm{E}[B^2] = \sum_{j=km}^{km+m-2} \sigma^2 \alpha^{-2j-2} \left(\sum_{j=j+1}^{km+m-1} \alpha^i \right)^2 . \quad (25)$$

Therefore,

$$C_{k,k} = (1-\alpha^2)(\mathrm{E}[A^2] + \mathrm{E}[B^2]) \quad (26)$$
$$= \sigma^2 \frac{-\alpha^{2km}(1-\alpha^m)^2 + m(1-\alpha^2) - 2\alpha(1-\alpha^m)}{(1-\alpha)^2} \, .$$

Equation (27) suggests that $C_{k,k}$ is approximately a bell-shaped function of α, starting from $m\sigma^2$ when $\alpha = 0$, ending at 0 when $\alpha = 1$, and reaching the maximum at some $\alpha \in (0,1)$. This fact is useful when, in the numerical analysis section, we study the effect of memory in a user's mobility pattern on the cost of the predictive scheme.

To compute $C_{k-1,k}$, we have

$$\begin{aligned} C_{k-1,k} &= \mathrm{E}[y_{k-1} y_k] - \mu_{y_{k-1}} \mu_{y_k} \quad (27) \\ &= \mathrm{E}\left[\left(\sqrt{1-\alpha^2} \sum_{i=km-m}^{km-1} \sum_{j=0}^{i-1} \alpha^{i-j-1} x_j \right) \right. \\ &\quad \left. \left(\sqrt{1-\alpha^2} \sum_{i=km}^{km+m-1} \sum_{j=0}^{i-1} \alpha^{i-j-1} x_j \right) \right] . \end{aligned}$$

Exchanging the order of the above double summations, then dividing each into two independent parts, we have

$$C_{k-1,k} = (1-\alpha^2) \times$$
$$\mathrm{E}\left[\left(\sum_{j=0}^{km-m-1} \sum_{i=km-m}^{km-1} \alpha^{i-j-1} x_j + \sum_{j=km-m}^{km-2} \sum_{i=j+1}^{km-1} \alpha^{i-j-1} x_j \right) \right.$$
$$\left. \left(\sum_{j=0}^{km-1} \sum_{i=km}^{km+m-1} \alpha^{i-j-1} x_j + \sum_{j=km}^{km+m-2} \sum_{i=j+1}^{km+m-1} \alpha^{i-j-1} x_j \right) \right] .$$

Cancelling out uncorrelated products and combining the independent Gaussian random variables, we have

$$C_{k-1,k} = (1-\alpha^2)\sigma^2 \times$$
$$\left(\sum_{j=0}^{km-m-1} \left(\sum_{i=km-m}^{km-1} \alpha^{i-j-1} \cdot \sum_{i=km}^{km+m-1} \alpha^{i-j-1} \right) \right.$$
$$\left. + \sum_{j=km-m}^{km-2} \left(\sum_{i=j+1}^{km-1} \alpha^{i-j-1} \cdot \sum_{i=km}^{km+m-1} \alpha^{i-j-1} \right) \right)$$
$$= \sigma^2 \frac{(1-\alpha^m)^2(\alpha - \alpha^{2km-m})}{(1-\alpha)^2} \, . \quad (28)$$

Defining $\alpha_k = \frac{C_{k-1,k}}{C_{k-1,k-1}}$ and $u_k = \sqrt{1 - \frac{C_{k-1,k}^2}{C_{k-1,k-1} C_{k,k}}} z_{k-1}$, from (17), we have

$$w_k = \alpha_k w_{k-1} + u_k \, . \quad (29)$$

Let f_{w_k} and f_{u_k} be the PDF of w_k and u_k respectively. Since u_k is independent of w_{k-1},

$$f_{w_k}(w) = \frac{1}{\alpha_k} f_{w_{k-1}}(w/\alpha_k) * f_{u_k}(w) \, , \quad (30)$$

where $*$ denotes one-dimensional convolution.

We further define the following functions:

- $p_{r_k w_{k-1}}(r,w)$: Probability that $r_k = r$ and $w_{k-1} = w$, and that there is no update up to time $k-1$
- $q_{r_k w_{k-1}}(r,w)$: Probability that $r_k = r$ and $w_{k-1} = w$, and that there is no update up to time k
- $q_{r_k w_k}(r,w)$: Probability that $r_k = r$ and $w_k = w$, and that there is no update up to time k
- $h_N(r) = \begin{cases} 1, & -N \le r \le N \\ 0, & \text{otherwise} \end{cases}$, where N is the distance to update
- $\overline{F}(k)$: Probability that there is no update up to time k
- $f(k)$: PDF of time between two consecutive updates

The initial distribution of mobile displacement w_0 can be determined from (9) by setting $v_0 = 0$. The location distributions at time $k > 0$ can then be found with the following iterative steps:

Step 0. $p_{r_1 w_0}(r,w) = f_{w_0}(w)\delta(r-w)$; set $k = 1$;
Step 1. $q_{r_k w_{k-1}}(r,w) = p_{r_k w_{k-1}}(r,w) h_N(r)$;
Step 2. $\overline{F}(k) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} q_{r_k w_{k-1}}(r,w) dw dr$
Step 3. $q_{r_k w_k}(r,w) = \frac{1}{\alpha_k} q_{r_k w_{k-1}}(r, \frac{w}{\alpha_k}) * f_{u_k}(w)$
Step 4. $p_{r_{k+1} w_k}(r,w) = q_{r_k w_k}(r-w, w)$
Step 5. Set $k = k+1$, and repeat from Step 1.

The above iteration ideally goes to $k = \infty$. However, it can be terminated when $\overline{F}(k)$ is sufficiently small.

Then, the PDF of the update time interval is given by

$$f(k) = \overline{F}(k-1) - \overline{F}(k) \, . \quad (31)$$

Since $f(k)$ does not depend on v_0, the update time intervals are independent and identically distributed.

B. Cost of Mobility Management

We consider location updates between two successive call arrivals. As shown in Figure 1, the i.i.d. location update time intervals comprise a renewal process with the probability density function $f(t)$.

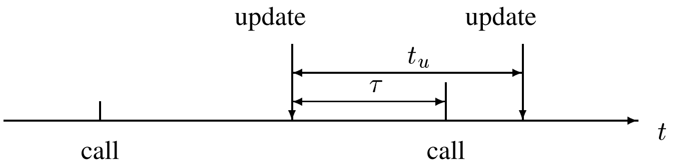

Fig. 1. Renewal Process between Two Successive Call Arrivals

Let's consider one location update. Assume that the next autonomous location update is at time t_u and that the next call arrives at time τ, where $\tau < t_u$. The probability density function of τ given that τ is less than t_u is

$$p_{\tau|\tau<t_u}(t) = \frac{\lambda e^{-\lambda t} u(t_u - t)}{1 - e^{-\lambda t_u}} , \tag{32}$$

where λ is the call arrival rate, and $u(t)$ is the step function.

Thus, when a call arrives, the PDF of time elapsed since the mobile's last location update is

$$\begin{aligned} p_\tau(t) &= \int_0^\infty p_{\tau|\tau<t_u}(t) f(t_u) dt_u \\ &= \lambda e^{-\lambda t} \int_t^\infty \frac{f(t_u)}{1 - e^{-\lambda t_u}} dt_u . \end{aligned} \tag{33}$$

Given that a call arrives at $t = \tau$ after the last location update, the PDF of the mobile's distance to the predicted location is

$$g_{r_\tau}(r) = \frac{\int_{-\infty}^{\infty} q_{r_\tau w_{\tau-1}}(r, w) dw}{\int_{-N}^{N} \int_{-\infty}^{\infty} q_{r_\tau w_{\tau-1}}(r, w) dw dr} . \tag{34}$$

Using a paging scheme that searches for the mobile in cells based on the shortest-distance-first rule, the paging cost associated with locating the mobile at distance r from the predicted location is $(2\lfloor r/S \rfloor + 1)C_p$, where S is the cell size, and C_p is the cost of paging one cell. Thus, the average paging cost per unit time is

$$C_{page} = \lambda C_p \int_{\tau=0}^{\infty} p_\tau(\tau) \int_{-N}^{N} g_{r_\tau}(r)(2\lfloor r/S \rfloor + 1) dr d\tau . \tag{35}$$

Next, we will find the cost associated with location updates. Let $u(t)$ denote the number of location updates within the time interval of length t between two successive call arrivals. Then,

$$\Pr[u(t) = i] = F^{(i)}(t) - F^{(i+1)}(t) , \tag{36}$$

where

$$F^{(i)}(\tau) \triangleq \int_{t=0}^{\tau} f^{(i)}(t) dt , \tag{37}$$

and $f^{(i)}(t)$ is defined by the following recursive relation

$$f^{(1)}(t) = f(t) \tag{38}$$

$$f^{(i)}(t) = f^{(i-1)}(t) * f(t) . \tag{39}$$

Let $M(t)$ be the expected value of $u(t)$. Then,

$$\begin{aligned} M(t) &= \sum_{i=1}^{\infty} i(F^{(i)}(t) - F^{(i+1)}(t)) \\ &= \sum_{i=1}^{\infty} iF^{(i)}(t) - \sum_{i=2}^{\infty} (i-1)F^{(i)}(t) \\ &= \sum_{i=1}^{\infty} F^{(i)}(t) . \end{aligned} \tag{40}$$

From the equation (40), the average updating cost per unit time can be obtained by

$$C_{update} = \lambda C_u \int_0^\infty \lambda e^{-\lambda t} M(t) dt , \tag{41}$$

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

Finally, the total cost of mobility management per unit time is

$$C_{total} = C_{update} + C_{page} . \tag{42}$$

IV. Numerical Results and Comparisons

In the following numerical analysis, we normalize all costs to the unit of the paging cost C_p. We also normalize distance to the unit of the cell size. We are interested in understanding how the remaining variables, namely, the memory factor exponent β, the average velocity μ, the standard deviation σ, the cost per location update C_u, the call arrival rate λ, and the location inspection period m, affect the performance gain of the predictive scheme, as compared with the non-predictive distance-based scheme, in terms of the optimal C_{total}.

For the non-predictive distance-based scheme, we use computer simulations to determine its cost with the above six parameters taking various combinations of values. In these simulations, we assume an infinite one-dimensional space that is divided into cells of size 1, where a mobile travels according to the 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 ($s_0 = 0$), and has initial velocity with Gaussian distribution defined by μ and σ. The time of the next call arrival is randomly generated following the exponential distribution with 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 the call arrives, the paging cost is computed based on the mobile's distance from the origin¿ Also, the updating cost is computed based on the total number of location updates performed since time initiation. The above experiment is repeated 10^5 times, and the average is taken for each set of parameters.

For each combination of the above six parameters, the minimum cost for each scheme is obtained by searching over different update distance thresholds.

We define the performance gain of prediction as the ratio between the minimum C_{total} for the non-predictive scheme to the minimum C_{total} for the predictive scheme. 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, (β, μ, σ) and (λ, C_u, m). For each group of variables, we study the effect of these variable on mobility management cost in detail, while the variables in the other group are fixed. Due to space limitations, we show here only the plots of the performance gain.

In Figures 2-3, we study the effect of mobility pattern, namely, (β, μ, σ), on the performance gain. The other parameters are set to $(\lambda, C_u, m) = (0.01, 10^{0.5}, 10)$.

Figure 2 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 $\{0, 10^{-1}, 10^{-0.5}, 10^{0}, 10^{0.5}, 10^{1}\}$. These plots demonstrate that the performance gain is a concave function of β. On the one hand, when β is small, the user mobility has high memory level, which favors the predictive scheme. On the other hand, when β is large, α is small, and from (27), $C_{k,k}$ reaches a local minimum, $m\sigma^2$, at $\alpha = 0$. Therefore, in this case, the disadvantage of the non-predictive scheme is mainly determined by a mobile's average velocity, when μ is not too small (larger than 0.1 in this case). Since μ does not affect the cost of the predictive scheme, the predictive performance gain is larger for larger β, which leads to smaller $C_{k,k}$.

When $\mu = 0$, the performance gain decreases from about 2 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. 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 is easily predictable. In this case, the only cost incurred using the predictive scheme, is the cost of paging once in the cell of the predicted location, since the mobile never needs to update its location.

Figure 3 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 $\{0, 10^{-1}, 10^{-0.5}, 10^{0}, 10^{0.5}, 10^{1}\}$. For the various curves, σ takes a value from $\{0.1, 0.5, 1, 5, 10\}$. These plots demonstrate that the performance gain is an increasing function of μ, for μ not too much larger than σ. When μ is much larger than σ, as discussed in the previous figure, the mobility pattern is close to the fluid-flow model, and the savings by prediction are maximal and independent of the mobile's velocity. The asymptotic standard deviation, σ, represents the uncertainty in a mobile's velocity. Here we see that the performance gain is maximum when σ is small, but there is little performance gain when the magnitude of σ is close to or larger than μ. For example, when $\sigma = 0.1$ and $\mu = 10$, a performance gain greater than 10 is achieved, but when $\sigma > 5$, the performance gain is

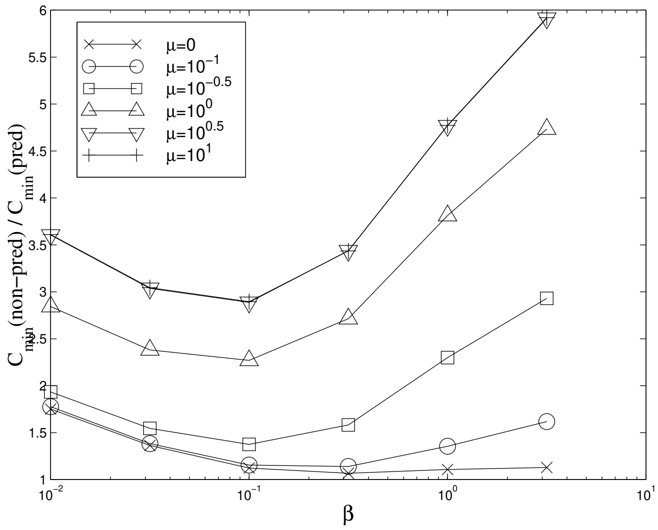

Fig. 2. Performance gain vs. β and μ for $\lambda = 0.01$, $C_u = 10^{0.5}$, $m = 10$, and $\sigma = 0.5$

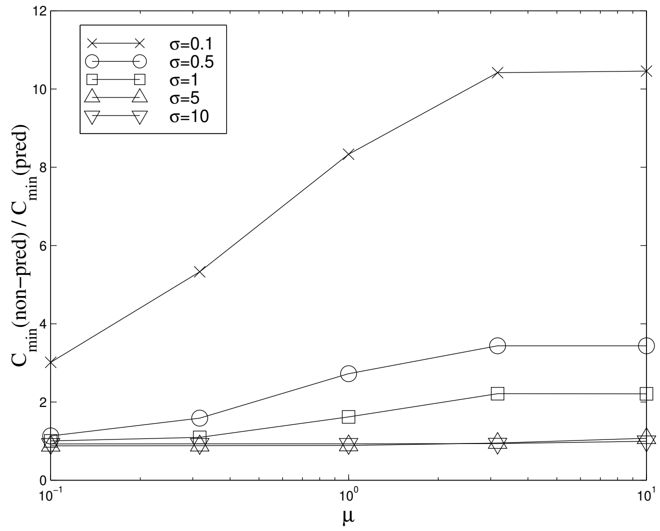

Fig. 3. Performance gain vs. μ and σ for $\lambda = 0.01$, $C_u = 10^{0.5}$, $m = 10$, and $\beta = 10^{-0.5}$ ($\alpha = 0.73$)

unity for all μ in the interval [0.1, 10].

Figure 4 shows the plots of the performance gain versus σ, for various values of β, and with fixed $\mu = 1$. Here σ takes the values $\{0.1, 0.5, 1, 5, 10\}$. For the various curves, β takes a value from $\{10^{-2}, 10^{-1.5}, 10^{-1}, 10^{-0.5}, 10^{0}, 10^{0.5}\}$. These plots demonstrate that the performance gain is a faster-than-exponential decreasing function of σ; this observation is in agreement with the results from the previous graph. Here we see again that the cost gain is a concave function of β, and is large when β is either large or small.

In Figures 5-7, we study how the parameters (λ, C_u, m) af-

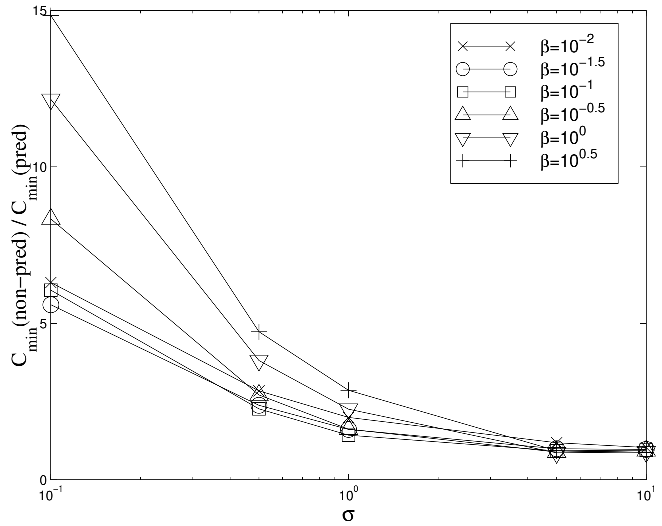

Fig. 4. Performance gain vs. σ and β for $\lambda = 0.01$, $C_u = 10^{0.5}$, $m = 10$, and $\mu = 1$

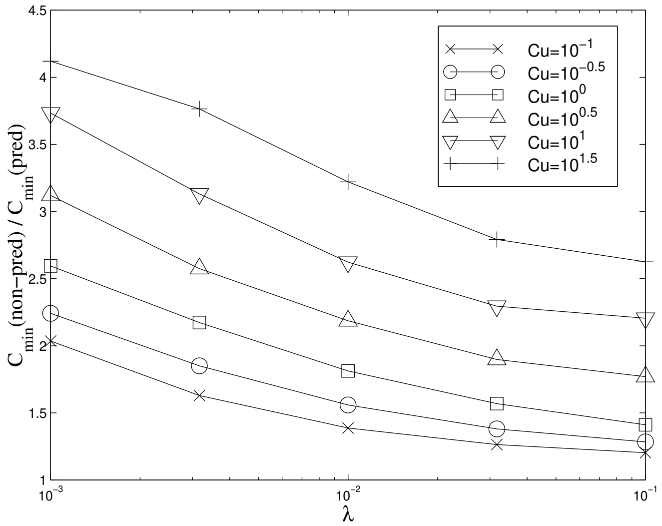

Fig. 5. Performance gain vs. λ and C_u for $\mu = 0.5$, $\sigma = 0.5$, $\beta = 10^{-0.5}$, and $m = 5$

fect the performance gain. We set the mobility parameters $(\beta, \mu, \sigma) = (10^{-0.5}, 0.5, 0.5)$.

Figure 5 presents the plots of the performance gain versus λ, for various values of C_u, and with fixed $m = 5$. Here λ takes the values $\{10^{-3}, 10^{-2.5}, 10^{-2}, 10^{-1.5}, 10^{-1}\}$. For the various curves, C_u takes a value from $\{10^{-1}, 10^{-0.5}, 10^{0}, 10^{0.5}, 10^{1}, 10^{1.5}\}$. These plots demonstrate that the predictive performance gain is an approximately exponential decreasing function of the call arrival rate. However, the rate of decrement is not very steep. For example, with $C_u = 10$, the performance gain is 3.7, 2.7, and 2.2, when $\lambda = 0.001$, $\lambda = 0.01$, $\lambda = 0.1$, respectively. These plots also suggest that the predictive scheme gives larger cost savings when the cost per location update is larger.

Figure 6 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 $\{10^{-1}, 10^{-0.5}, 10^{0}, 10^{0.5}, 10^{1}, 10^{1.5}\}$. For the various curves, m takes a value from $\{2, 5, 10, 20, 40\}$. These plots demonstrate that, except in the extreme case when m is very large, the predictive performance gain is an approximately linearly increasing function of the location update cost; this observation is in agreement with the results from the last figure. There are sharp turns in the curves with $m = 20$ and $m = 40$. This is because, for these two values of m, when C_u is relatively small, the obtained optimal update distance is 1 cell size for both the predictive scheme and the non-predictive scheme. In this case, since, in our analysis, the distance is quantized in units of a cell, the exact update distance is not well defined here. These curves also suggest that the performance gain is larger for smaller location inspection periods. This is studied in more detail in the next figure.

Figure 7 presents the plots of the performance gain versus m, for the various values of λ, and with fixed $C_u = 1$. Here m takes the values $\{2, 5, 10, 20, 40\}$. For the various curves, λ takes a value from $\{10^{-3.5} 10^{-3}, 10^{-2.5}, 10^{-2}\}$. These plots demonstrate that the performance gain is an approximately linearly decreasing function of the location inspection period. Therefore, the prediction scheme favors a system where the mobiles frequently monitor and update their locations. However, since using prediction incurs more computation and communication cost for each location inspection, there exists a trade-off between the extra cost due to frequent location inspection and the amount of performance gain, which should be considered when designing the optimal location inspection frequency for the proposed predictive system.

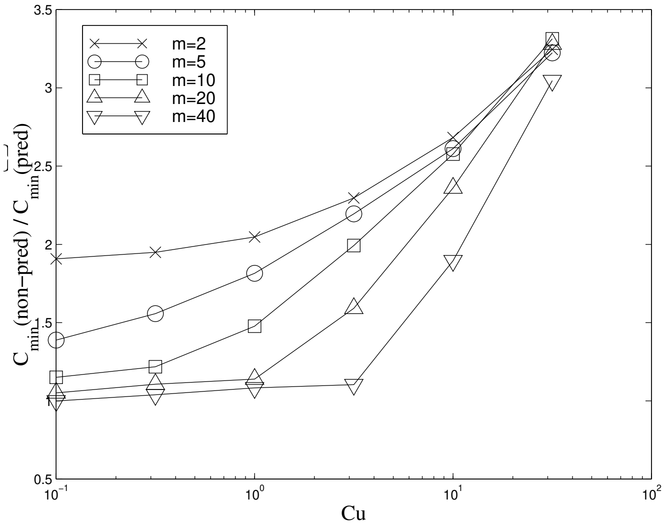

Fig. 6. Performance gain vs. C_u and m for $\mu = 0.5$, $\sigma = 0.5$, $\beta = 10^{-0.5}$, and $\lambda = 0.01$

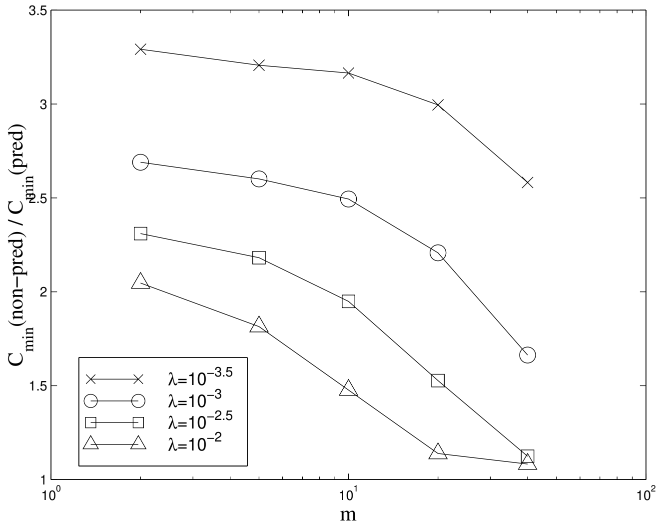

Fig. 7. Performance gain vs. m and λ for $\mu = 0.5$, $\sigma = 0.5$, $\beta = 10^{-0.5}$, and $C_u = 1$

V. Conclusions

Mobile users in PCS networks move with wide variety of mobility patterns, especially in networks with multilayer macro-cellular and micro-cellular infrastructures ([6]). The 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 random-walk or constant-velocity fluid-flow as the user mobility model. However, neither of these mobility models can practically represent mobiles' movements in a PCS network. In this paper, we introduce a mobility model based on the Gauss-Markov random process, which more realistically captures the various degrees of correlation of a mobile user velocity in time.

With the Gauss-Markov mobility model, we present 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. An analytical framework is introduced to evaluate the performance of the predictive scheme, which allows us to study the effects of various parameters on the mobility management cost. This cost is then compared with the cost of the regular, non-predictive distance-based scheme obtained from simulations. In the span of parameter values under consideration, the performance improvement by the predictive scheme ranges from unity to a factor of more than 10.

Numerical results suggest that the predictive scheme gives better gains in systems with larger μ and smaller σ; i.e., when user mobility pattern is closer to the constant-velocity fluid-flow model. It gives the largest gain when $\mu >> \sigma$. As an example, with the parameter values as shown in Figure 2, when $\mu = 1$ and $\sigma = 0.1$, a cost reduction of 8.5 times is achieved.

When μ is small, and σ and β are large, the user mobility pattern is close to the random-walk model. In this case, the predictive scheme does not perform any better than the non-predictive one. For example, when $\mu = 0$, $\sigma = 0.5$, and $\beta = 1$, the performance gain is close to unity.

When μ is not very small relative to sigma (for example, with the range of parameter values considered in our paper, when $\mu > 0.1$ and $\sigma = 0.5$), a large β, which indicates low memory level in the user mobility pattern, not necessarily lead to low performance gain. This is due to the non-monotonicity of $C_{k,k}$ as a function of β. As a point of reference, for $\lambda = 0.01$, $C_u = 10^{0.5}$, $m = 10$, $\mu = 1$, and $\sigma = 0.5$, the performance gain is 2.8, 2.4, and 3.8, for $\beta = 0.01$, $\beta = 0.1$, and $\beta = 1$, respectively.

The parameters other than those associated with the mobility pattern also affect the performance gain. Numerical results show that the performance gain is, approximately, an exponentially decreasing function of the call arrival rate, a linearly increasing function of the cost per location update, and a linearly decreasing function of the mobile location inspection period. There exists a trade-off between the extra cost due to the predictive location inspections and the amount of performance gain, which should be considered when designing the optimal location inspection frequency for the proposed predictive system.

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] 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

[5] 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

[6] 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

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

[8] 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

[9] 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

[10] 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

[11] 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

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

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

[14] 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

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

[16] 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

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