Movement-Based Mobility Management and Trade Off Analysis for Wireless Mobile Networks

Yuguang Fang, *Senior Member, IEEE*

Abstract—Mobility management plays a central role in providing ubiquitous communications services in future wireless mobile networks. In mobility management, there are two key operations, *location update* and *paging*, commonly used in tracking mobile users on the move. Location update is to inform the network about a mobile user's current location, while paging is used for the network to locate a mobile user. Both operations will incur signaling traffic in the resource limited wireless networks. The more frequent the location updates, the less paging in locating a mobile user; thus, there is a trade off in terms of signaling cost. Most trade off analysis in the literature is carried out under the assumption that some time variables are exponentially distributed. However, such assumptions will not be valid, particularly for the wireless Internet. In this paper, we present some general analytical results without these assumptions, which are essential for the general trade off analysis. Specifically, we analytically derive the average number of location updates during the interservice time for a movement-based location update scheme under fairly realistic assumptions, which are crucial for all trade off analysis. Our general analytical results make thorough numerical analysis for finding the optimal mobility management under various network operation scenarios possible.

Index Terms—Mobility management, location update, paging, wireless networks, mobile networks, cost analysis.

1 Introduction

In wireless mobile networks, in order to effectively deliver a service to a mobile user, the location of the called mobile user must be determined within a certain time limit (before the service is blocked). Due to the mobility of mobile users, the location of mobile users may constantly change, so the database for storing the location information (in the Home Location Register (HLR) or Visiting Location Register (VLR)) has to be constantly updated. *Location management* is used to serve this purpose to track mobile users that move from place to place in the coverage area of a wireless mobile network or in the coverage areas of multiple communications networks, which are working together to fulfill the grand vision of ubiquitous communications. Thus, location management (also called *mobility management*) is a key component for the effective operation of wireless networks ([5] and references therein).

There are two basic operations for tracking a mobile user: *location update* and *paging*. Location update is the process for the mobile terminals to report their locations to the network; thus all mobiles are actively sending location update messages to keep the network informed. When an incoming call to a mobile user (we will use mobile user and mobile terminal interchangeably) arrives, the wireless network simply routes the call to the last reported location of the mobile terminal. Intuitively, the location accuracy depends on the location update frequency; the more often the location updates are, the more accurate the location information is. However, frequent location updates increase the signaling traffic, which may lead to congestion in the signaling network. Besides, this may not be necessary if a mobile terminal moves slowly (such as a pedestrian) or if a mobile terminal does not have too many calls arriving. A passive approach, *paging* (also called *terminal paging*), may be much more appropriate. Paging is the process for the network to search for the called terminal by sending polling signals to cells close to the last reported location of the called terminal. Of course, the larger of the paging area, the more signaling traffic will result. If we define the region where a mobile stays as the *uncertainty region*, then location update is attempting to minimize the uncertainty region, while paging is to discover the called terminal in the uncertainty region. The larger the uncertainty region, the higher the paging signaling traffic. Therefore, there is a trade off between the location update traffic and the paging traffic. The cost trade off analysis has been intensively studied in the current literature ([1], [2], [3], [4], [6], [19], [20], [23]).

Surveying the literature, we have the following observation: Most cost analyses adopted some exponential assumptions for some time variables. For example, in [4], [19], the time between the previously served call arrival and the time of currently served call (t_c in their notation), also called *interservice time* in [12], is assumed to be exponentially distributed. In fact, both [4] and [19] identified the interarrival time of call arrivals to a mobile terminal with the interservice time, thus, when the call arrivals to a terminal are assumed to be Poisson, then the interservice time will be exponentially distributed. However, this

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

Manuscript received DD Mmmm, YYYY; revised DD Mmmm, YYYY; accepted DD Mmmm, YYYY.
For information on obtaining reprints of this article, please send e-mail to: tc@computer.org, and reference IEEECS Log Number 117653.

identification ignored the busy-line effect: Call arrivals to a mobile terminal will be blocked if the mobile terminal is busy serving another call (this will change if voice and data services are integrated; we will not address this situation in this paper), thus, the served calls by the mobile is a sampled process from a Poisson process, is most likely not a Poisson process. Moreover, both [4] and [19] ignored the service time for a served call, which is reasonable for PCS networks, because the (voice) calling time is negligible compared to the interarrival time between the call arrivals to the mobile terminal. However, in future wireless mobile networks, a call connection (particularly for data service such as wireless Internet) will tend to last longer than voice calls. During the connection period, the mobile does not need to do the location update! Thus, the "interservice time" is not the right time variable to be used for cost analysis. Rather, it is the *mobile terminal idling time*. We still use the term *interservice time* with the following modifying meaning: It is the time between the end of the previously served call and the arriving instant of the currently served call for a mobile terminal. Obviously, the interservice time may not be exponentially distributed, even if the call arrivals to a mobile terminal are Poisson!

In this paper, we will focus on tackling the problem with a more realistic assumption. In solving other problems in performance evaluation of wireless networks such as PCS networks ([10], [11], [12], [13]), we developed a new analytical technique by applying complex analysis and probability theory. This technique was also used to analyze two location management schemes, namely, Pointer-Forwarding Scheme and Two-Location Algorithm, in [9]. In this paper, we evaluate movement-based mobility management under more general assumptions. We first present a general framework for the trade off analysis in location management, then we develop techniques to handle the situation when the interservice time is generally distributed. Our main focus in this paper is to compute the location update cost under generally distributed interservice time. In the current literature, three location update schemes were proposed and studied ([3], [6]): *distance-based location update*, *movement-based location update*, and *time-based location update*. In the distance-based locate update scheme, location update will be performed when a mobile terminal moves d cells away from the cell in which the previous location update was performed, where d is a distance threshold. In the movement-based location update scheme, a mobile terminal will carry out a location update whenever the mobile terminal completes d movements between cells, where d is the movement threshold. In the time-based location update scheme, the mobile terminal will update its location every d time units, where d is the time threshold. It has been shown ([6]) that the distance-based location update scheme gives the best result in terms of signaling traffic, however, it may not be practical because a mobile terminal has to know its own position information in the network topology. The time-based location update scheme is the simplest to implement; however, unnecessary signaling traffic may result (imagine a stationary terminal for a long period may not need to do any update before it moves). The movement-based location update scheme seems to be the best choice in terms of signaling traffic and implementation complexity. We will concentrate on this scheme for our study in this paper.

This paper is organized as follows: In the next section, we present the general framework for the cost analysis for location management. In Section 3, we study the probability distribution of the number of cell boundary crossings under fairly realistic assumptions. We then give the analytical formula for the average number of location updates during the interservice time in Section 4. Trade off cost analysis is carried out in Section 5. We then conclude the paper in the Section 6.

2 A General Framework for Trade Off Analysis

Consider a wireless mobile network with cells of the same size. A mobile terminal visits a cell for a time interval which is generally distributed, then moves to the neighboring cell with equal probability (we are interested, in this paper, in the homogeneous wireless networks in which all cells are statistically identical). Consider the movement-based location update scheme. Let d denote the movement threshold, i.e., a mobile terminal will perform a location update whenever the mobile terminal makes d movements (equal to the number of serving cell switching) after the last location update. When an incoming call to a mobile terminal arrives, the network initiates the paging process to locate the called mobile terminal. Thus, both location update and terminal paging will incur signaling traffic; location updates consume the uplink bandwidth and mobile terminal power, while terminal paging mainly utilizes the downlink resource, hence the cost factors for both processes are different. When uplink signaling traffic is high or power consumption is a serious consideration, it may be better to use terminal paging instead. Different users or terminals may have different quality of service (QoS) requirements; location update and paging may be designed to treat them differently. All these factors should be considered and a more general cost function, which reflects these considerations, is desirable.

We observe that a mobile terminal engaging service does not need to perform location update since it is known to the network, thus we only need to consider the situation when a mobile terminal is idle, i.e., it does not have call connection with another terminal (stationary or mobile). A reasonable cost should consider two factors: location update signaling and paging signaling, which can be captured by the following general cost function:

$$\mathcal{C}(d) = \mathcal{C}_u(N_u(d), \lambda_u, q_s) + \mathcal{C}_p(N_p(d), \lambda_p, q_s), \qquad (1)$$

where $N_u(d)$ and $N_p(d)$ are the average number of location updates and the average number of paging messages under the movement-based location update and a paging scheme with movement threshold d, respectively, during a typical period of interservice time, λ_u is the signaling rate for location updates at a mobile switching center (MSC) and λ_p is the signaling rate for paging at an MSC, q_s indicates the QoS factor, $\mathcal{C}_u$ and $\mathcal{C}_p$ are two functions, reflecting the costs for location updates and paging. Depending on the choice

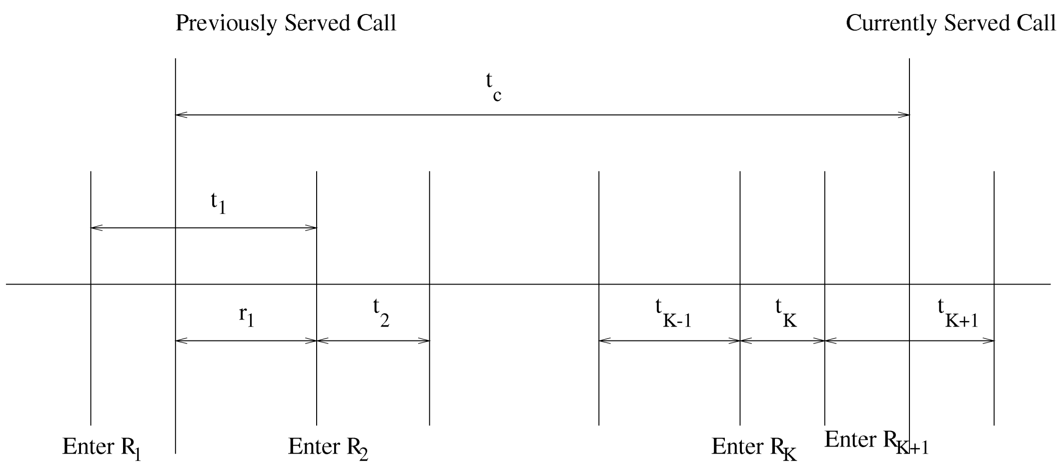


Fig. 1. The time diagram for K cell boundary crossings.

of the two functions, we can obtain different mobility management schemes, particularly, we can obtain the optimal movement threshold d to minimize the total cost $\mathcal{C}(d)$, which gives us the best trade off scheme.

In the current literature, the total cost is chosen to be the linear combination of the location update and the paging signaling traffic, i.e., $\mathcal{C}_c(N_u(d), \lambda_u, q_s) = UN_u(d)$, $\mathcal{C}_p(N_p(d), \lambda_p, q_s) = PN_p(d)$, where U is the cost factor for location update and P is the cost factor for terminal paging ([4], [19], [20]). When the signaling traffic for location updates in an area is too high, some mobile terminals may be advised to lower their location updates; this can be done by choosing a function $\mathcal{C}_u$ to contain a factor $1/(\lambda_{\max} - \lambda_u)$, where $\lambda_{\max}$ is the maximum allowable signaling rate for location update. For example, if we choose $\mathcal{C}_u = U/(\lambda_{\max} - \lambda_u)$, when λ_u is approaching $\lambda_{\max}$, the update cost will become huge, the optimization will prefer to use paging more, leading to fewer location updates. How to choose the appropriate cost functions is a challenging problem to be addressed in the future. Obviously, the total cost relies heavily on the computation of $N_u(d)$ and $N_p(d)$; we need to find the computational procedure for them. Since terminal paging schemes depend on detailed searching procedures (sequential, parallel, or selective), the number $N_p(d)$ is paging specific. Our focus in this paper is to present a general computational procedure for $N_u(d)$ under fairly general assumptions and carry out the cost analysis accordingly. We leave other related topics for future research. In what follows, we will carry out the computation of $N_u(d)$ in the following few sections.

3 Probability of the Number of Cell Boundary Crossings

In order to derive the average number of location updates under the movement-based location update scheme during a typical interservice time, we first need to know the probability distribution of the number of cell boundary crossings. In this section, we give the analytical results for this distribution.

Assume that the incoming calls to a mobile terminal, say, $\mathcal{T}$, form a Poisson process, the time the mobile terminal stays in a cell (called the *cell residence time*) has a general nonlattice distribution (a *lattice distribution* is the distribution of a random variable that only takes on integral multiples of some nonnegative number). We will derive the probability $\alpha(K)$ that a mobile terminal moves across K cells between two served calls arriving to the mobile terminal $\mathcal{T}$. The time between the end of a call served and the start of the following call served by the mobile terminal is called *the interservice time*. It is possible that a call arrives while the previous call served is still in progress ([20]). In this case, the mobile terminal $\mathcal{T}$ cannot accept the new call (the caller senses a busy tone in this case and may hang up). Thus, the interarrival (intercall) times for the calls terminated at the mobile terminal $\mathcal{T}$ are different from the interservice times. This phenomenon is called the *busy line* effect. Although the incoming calls form a Poisson process (i.e., the interarrival times are exponentially distributed), the interservice times may not be exponentially distributed. When the busy line effect is negligible and the interservice time is exponentially distributed, Lin ([20]) is able to derive an analytical formula for $\alpha(K)$. Most trade off analysis (e.g., ([4], [20])) for location update and paging used this formula, hence, the hidden assumption for such a trade off analysis is that the interservice time is exponentially distributed. To carry out more general trade off analysis, we need more general results for $\alpha(K)$. In this section, we assume that the interservice time is generally distributed and derive an analytic expression for $\alpha(K)$.

Let $t_1, t_2, \ldots$ denote the cell residence times of a mobile in cells $1, 2, \ldots$, respectively, and r_1 denote the residual cell residence time of t_1 (i.e., the time interval between the time instant the call registers to the network and the time instant the mobile terminal exits the cell). Let t_c denote the interservice time between two consecutive served calls to a mobile terminal $\mathcal{T}$. Fig. 1 shows the time diagram for K cell boundary crossings. In this figure, we omit the call service times for clarity, thus t_c shown in the figure is in fact the time interval between the start of the previously served call and the start of the currently served call. In reality, t_c should be the time between the end of the currently served call and the start of the next served calls. Suppose that the mobile terminal is in a cell R_1 when the previous call arrives and, accepted by $\mathcal{T}$, it then moves K cells during the interservice time and $\mathcal{T}$ resides in the jth cell for a period t_j $(1 \leq j \leq K+1)$. We consider a homogeneous wireless mobile network, i.e., all cells in the network are statistically identical. Assume $t_1, t_2, \ldots$ are independent and identically distributed (iid) with a general probability density function $f(t)$, let t_c be generally distributed with probability density function $f_c(t)$ and let $f_r(t)$ be the probability density function of r_1. Let $f^*(s)$, $f_c^*(s)$, and $f_r^*(s)$ denote the Laplace-Stieltjes (L-S) transforms (or simply Laplace transforms) of $f(t)$, $f_c(t)$, and $f_r(t)$, respectively. Let $E[t_c] = 1/\lambda_c$ and $E[t_i] = 1/\lambda_m$. From the residual life theorem ([18]), we have

$$\begin{aligned} f_r(t) &= \lambda_m \int_t^\infty f(\tau)d\tau = \lambda_m[1 - F(t)], \\ f_r^*(s) &= \frac{\lambda_m}{s}[1 - f^*(s)], \end{aligned} \tag{2}$$

where $F(t)$ is the distribution function of $f(t)$. It is obvious that the probability $\alpha(K)$ is given by

$$\alpha(0) = \Pr[t_c \leq r_1], \quad K = 0, \tag{3}$$

$$\begin{aligned} \alpha(K) = \Pr[r_1 + t_2 + \cdots + t_K < t_c \leq r_1 + t_2 + \cdots + t_{K+1}], \\ K \geq 1. \end{aligned} \tag{4}$$

We first calculate $\alpha(0)$. Since the Laplace transform of $\int_t^\infty f_r(\tau)d\tau$ is $(1 - f_r^*(s))/s$, from (3), the inverse Laplace transform and the independence of r_1 and t_c, we have

$$\begin{aligned} \alpha(0) &= \int_0^\infty \Pr(r_1 \geq t) f_c(t) dt \\ &= \int_0^\infty \int_t^\infty f_r(\tau) d\tau f_c(t) dt \\ &= \int_0^\infty \frac{1}{2\pi j} \int_{\sigma - j\infty}^{\sigma + j\infty} \frac{1 - f_r^*(s)}{s} e^{st} ds f_c(t) dt \\ &= \frac{1}{2\pi j} \int_{\sigma - j\infty}^{\sigma + j\infty} \frac{1 - f_r^*(s)}{s} \int_0^\infty f_c(t) e^{st} dt ds \\ &= \frac{1}{2\pi j} \int_{\sigma - j\infty}^{\sigma + j\infty} \frac{1 - f_r^*(s)}{s} f_c^*(-s) ds, \end{aligned} \tag{5}$$

where σ is a sufficiently small positive number which is appropriately chosen for inverse Laplace transform. We want to remark that the choice of such σ is possible for the validity of the inverse Laplace transformation. Recall that, for any probability density function, the Laplace transform is always analytic on the right half complex plane. Thus, if $f_c^*(s)$ has finite number of isolated poles (which is the case when it is a rational function), then $f_c^*(-s)$ will have finite number of isolated poles in the open right half complex plane, then we can choose σ to be less than the smallest of real parts of the poles of $f_c^*(-s)$. In this case, when we apply the Residue Theorem, we can use the semicircle in the right half complex plane as the integration contour.

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

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

Thus, the probability density function is given by

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

Also, the Laplace transform of $\Pr(\xi \leq t)$ (the distribution function) is $f_\xi^*(s)/s$. We have

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

Taking this into (4), we obtain

$$\begin{aligned} \alpha(K) &= \Pr(t_c \geq r_1 + t_2 + \cdots + t_K) \\ &\quad - \Pr(t_c \geq r_1 + t_2 + \cdots + t_{K+1}) \\ &= \frac{1}{2\pi j} \int_{\sigma - j\infty}^{\sigma + j\infty} \frac{f_r^*(s)[f^*(s)]^{K-1}[1 - f^*(s)]}{s} f_c^*(-s) ds. \end{aligned} \tag{6}$$

It is obvious that the integrand without term $f_c^*(-s)$ in (5) and (6) is analytic on the right half open complex plane. If $f_c^*(-s)$ has no branch point and has only a finite possible isolated poles in the right half plane (which is equivalent to saying that $f_c^*(s)$ has only a finite number of isolated poles in the left half plane), then the Residue Theorem can be applied to (5) and (6) using a semicircular contour in the right half plane. Indeed, if we use σ_c to denote the set of poles of $f_c^*(-s)$ in the right half complex plane, then, from (5) and (6) and the Residue Theorem ([18]), we obtain

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

$$\begin{aligned}\alpha(0) &= \frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty}\frac{1-f_r^*(s)}{s}f_c^*(-s)ds\\ &= -\sum_{p\in\sigma_c}\mathrm{Res}_{s=p}\frac{1-f_r^*(s))}{s}f_c^*(-s),\\ \alpha(K) &= \frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty}\frac{f_r^*(s)[1-f^*(s)][f^*(s)]^{K-1}}{s}f_c^*(-s)ds\\ &= -\sum_{p\in\sigma_c}\mathrm{Res}_{s=p}\frac{f_r^*(s)[1-f^*(s)][f^*(s)]^{K-1}}{s}f_c^*(-s),\\ &\quad K>0,\end{aligned} \tag{7}$$

where $\mathrm{Res}_{s=p}$ *denotes the residue at poles* $s=p$ *and* $f_r^*(s)=\lambda_m(1-f^*(s))/s$.

Proof. Choosing such a σ that all poles of $f_c^*(-s)$ in the right half plane are on the righthand side of a vertical line $s=\sigma$ and choosing the contour enclosed by the semicircle at center $s=\sigma+j0$ and with radius sufficiently large, then we can apply the Residue Theorem to complete the proof. □

4 Average Number of Location Updates

As we mentioned, the location update cost relies on the average number of location updates during the interservice time. In this section, we present an analytical result for this number under the movement-based location update scheme.

Let d be the threshold for the movement-based location update scheme. Then, the average number of location updates during an interservice time interval under the movement-based location update scheme can be expressed as ([4])

$$N_u(d)=\sum_{i=1}^{\infty} i\sum_{k=id}^{(i+1)d-1}\alpha(k). \tag{8}$$

In what follows in this section, we present the computation for $N_u(d)$. Let

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

then, from Theorem 1, we obtain

$$\begin{aligned}S(n) &= \sum_{k=1}^{n-1}\frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty}\frac{f_r^*(s)[1-f^*(s)][f^*(s)]^{k-1}}{s}f_c^*(-s)ds\\ &= \frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty}\frac{f_r^*(s)[1-f^*(s)]}{s}\left(\sum_{k=1}^{n-1}[f^*(s)]^{k-1}\right)f_c^*(-s)ds\\ &= \frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty}\frac{f_r^*(s)[1-(f^*(s))^{n-1}]}{s}f_c^*(-s)ds.\end{aligned} \tag{9}$$

Moreover, we have

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

Thus, from (8), (9), and (10), we obtain

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

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

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

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

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

In summary, we obtain:

Theorem 2. *The average number of location updates $N_u(d)$ is given by*

$$N_u(d) = \frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty} \frac{f_r^*(s)(f^*(s))^{d-1}}{s[1-(f^*(s))^d]} f_c^*(-s)ds. \tag{11}$$

If $f_c^*(s)$ has a finite number of poles, which is the case when it is a rational function, then we have

$$N_u(d) = -\sum_{p\in\sigma_c} \mathrm{Res}_{s=p} \frac{f_r^*(s)(f^*(s))^{d-1}}{s[1-(f^*(s))^d]} f_c^*(-s), \tag{12}$$

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

As a remark, we point out that, in the case when the service time is not negligible compared to the interservice time, we can easily apply our result to derive a similar result. Let $a(t)$ denote the probability density function of the time t_a between the beginning of a served call to the beginning of the following served call, let $s(t)$ denote the probability density function of the service time t_s for the served calls. Then, the interservice time is given by $t_c = t_a - t_s$ (i.e., $t_a = t_s + t_c$), thus t_c has the following L-S transform: $f_c^*(s) = a^*(s)/s^*(s)$, where $a^*(s)$ and $s^*(s)$ are the L-S transforms for $a(t)$ and $s(t)$, respectively. Thus, if we know the distribution $a(t)$ and $s(t)$, then we can find $f_c^*(s)$ and use it in Theorem 2 to obtain the corresponding analytical result.

5 Trade Off Cost Analysis: A Case Study

In this section, we present some results for the trade off analysis under the movement-based location update scheme and some paging schemes with the linear cost functional for illustration purpose.

5.1 Location Update Cost

If U denotes the unit cost for location update, then the total location update cost is given by

$$\mathcal{C}_u(d) = -U\sum_{p\in\sigma_c} \mathrm{Res}_{s=p} \frac{f_r^*(s)(f^*(s))^{d-1}}{s[1-(f^*(s))^d]} f_c^*(-s). \tag{13}$$

If the interservice time t_c is exponentially distributed with parameter λ_c, then $f_c^*(s) = \lambda_c/(s+\lambda_c)$, from (13), we can easily obtain

$$\mathcal{C}_u(d) = \frac{U f_r^*(\lambda_c)[f^*(\lambda_c)]^{d-1}}{1-[f^*(\lambda_c)]^d}. \tag{14}$$

This result was obtained in [19] (noticing that $f_r^*(s) = \lambda_m(1-f^*(s))/s$) via a different approach.

If the interservice time t_c is Erlang distributed with parameter (p, λ_c), i.e.,

$$f_c(t) = \frac{(p\lambda_c)^p t^{p-1}}{(p-1)!} e^{-p\lambda_c t},\ f_c^*(s) = \left(\frac{p\lambda_c}{s+p\lambda_c}\right)^p, \tag{15}$$

then the location update cost is given by

$$\mathcal{C}_u(d) = \frac{U(-1)^{p-1}(p\lambda_c)^p}{(p-1)!} g^{(p-1)}(p\lambda_c), \tag{16}$$

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

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

The hyper-Erlang distribution is a convex combination of multiple Erlang distributions, which has the following probability density function and Laplace transform:

$$\begin{aligned} f_{he}(t) &= \sum_{i=1}^M \alpha_i \frac{(m_i\eta_i)^{m_i} t^{m_i-1}}{(m_i-1)!} e^{-m_i\eta_i t} \quad (t \geq 0), \\ f_{he}^*(s) &= \sum_{i=1}^M \alpha_i \left(\frac{m_i\eta_i}{s+m_i\eta_i}\right)^{m_i}, \end{aligned} \tag{17}$$

where

$$\alpha_i \geq 0,\ \sum_{i=1}^M \alpha_i = 1.$$

This model contains the exponential distribution and Erlang distribution as special cases; it has been shown ([10], [17], [18]) that any distribution of a nonnegative random variable can be approximated by a hyper-Erlang distribution. We can easily find the mean and variance for this distribution as $\sum_{i=1}^M \alpha_i/\eta_i$ and $\sum_{i=1}^M \alpha_i/(m_i\eta_i^2)$, respectively. If the interservice time t_c is hyper-Erlang distributed as in (17), it is easy to derive the location update cost as

$$\mathcal{C}_u(d) = U\sum_{i=1}^M \alpha_i \frac{(-1)^{m_i-1}(m_i\eta_i)^{m_i}}{(m_i-1)!} g^{(m_i-1)}(m_i\eta_i), \tag{18}$$

where $g(s)$ is given earlier and

$$\lambda_c = \left[\sum_{i=1}^M \alpha_i/\eta_i\right]^{-1}.$$

5.2 Paging Cost

We consider the paging strategy used in [19] for our case study. Consider the hexagonal layout for the wireless network; all cells are statistically identical. According to the movement-based location update scheme, a mobile terminal moves at most d cells away from the previous position where it performs the last location update. Thus, a mobile terminal will surely be located in a cell which is less than d cells away from the previously reported position. If we page in the circular area with d cells as radius and with the previously reported position as the center, then we can definitely find the mobile terminal. Thus, this paging scheme is the most conservative among all paging. If we let P denote the unit

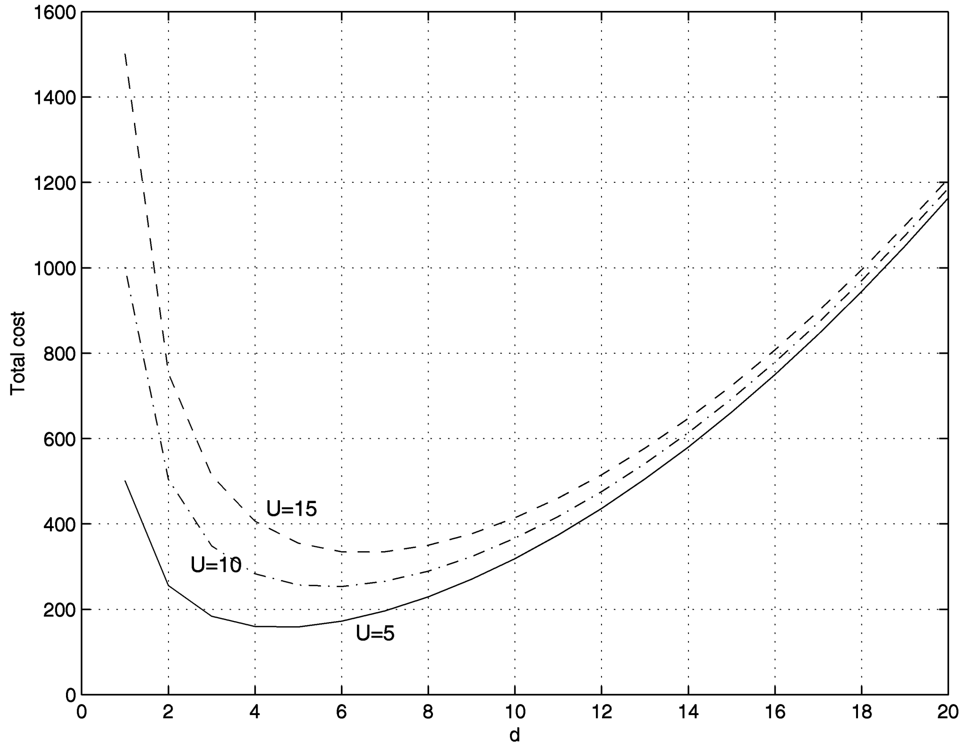


Fig. 2. Total cost versus the movement threshold: Varying unit cost U.

cost for each paging in a cell, the maximum paging cost for this paging scheme is given by ([19])

$$\mathcal{C}_p(d) = P(1 + 3d(d-1)). \tag{19}$$

As a remark, depending on the paging process, the delay can be calculated. For example, if parallel paging (i.e., paging all cells at once) is used, the delay is just one paging cycle, say, w_p; if we use the sequential paging (paging the inner ring first, then the second ring, and so on, which is a paging scheme according to the probability distribution), then the paging delay will be dw_p. We neglect such cost in our case study, only considering the trade off in terms of signaling traffic.

5.3 Total Cost

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

$$\begin{aligned}\mathcal{C}(d) &= \mathcal{C}_u(d) + \mathcal{C}_p(d)\\ &= -U \sum_{p\in\sigma_c} \mathrm{Res}_{s=p} \frac{f_r^*(s)(f^*(s))^{d-1}}{s[1-(f^*(s))^d]} f_c^*(-s)\\ &\quad + P(1+3d(d-1)).\end{aligned} \tag{20}$$

5.4 Numerical Analysis

In this subsection, we present some illustrative examples to show the behavior of overall cost function with respect to various parameters. We assume that the cell residence time is Gamma distributed with the following probability density function and the Laplace transform:

$$f(t) = \frac{(\gamma\lambda_m)^\gamma t^{\gamma-1}}{\Gamma(\gamma)} e^{-\gamma\lambda_m t},\ f^*(s) = \left(\frac{\gamma\lambda_m}{s+\gamma\lambda_m}\right)^\gamma. \tag{21}$$

The mean is $1/\lambda_m$ and the variance is $1/(\gamma\lambda_m^2)$. We assume that the interservice time is Erlang distributed with parameters (p, λ_c), its mean is $1/\lambda_c$, and its variance $1/(p\lambda_c^2)$. We first carry out similar studies as in [19] and observe the property of the total cost function with respect to the parameters such as the location update cost ratio P/U, the call-to-mobility ratio (CMR) $\theta = \lambda_c/\lambda_m$, and the variance of the cell residence time to compare our results with those given in [19]. We then present some more study for the case when the interservice time is not exponentially distributed.

Fig. 2 shows the plot for total cost versus the movement threshold d. We choose $P = 1$ and let $U = 5$, $U = 10$, and $U = 15$ (to reflect the ratio U/P). We observe that the cost function is a convex function with respect to the movement threshold, the same as for the exponential case (when $p = 1$). From the figure, we also observe that, as unit cost U for location update increases, the optimal movement threshold increases, which is very intuitive because the unit cost is like a penalty factor; if the penalty is high, the average number of location updates must be smaller, hence d should be greater.

Fig. 3 shows the cost function versus movement threshold for various call-to-mobility ratios. As we observe from the figure, the cost function is also convex, the global minimum (the optimal movement threshold) d decreases as the CMR increases. This is also intuitively understandable;

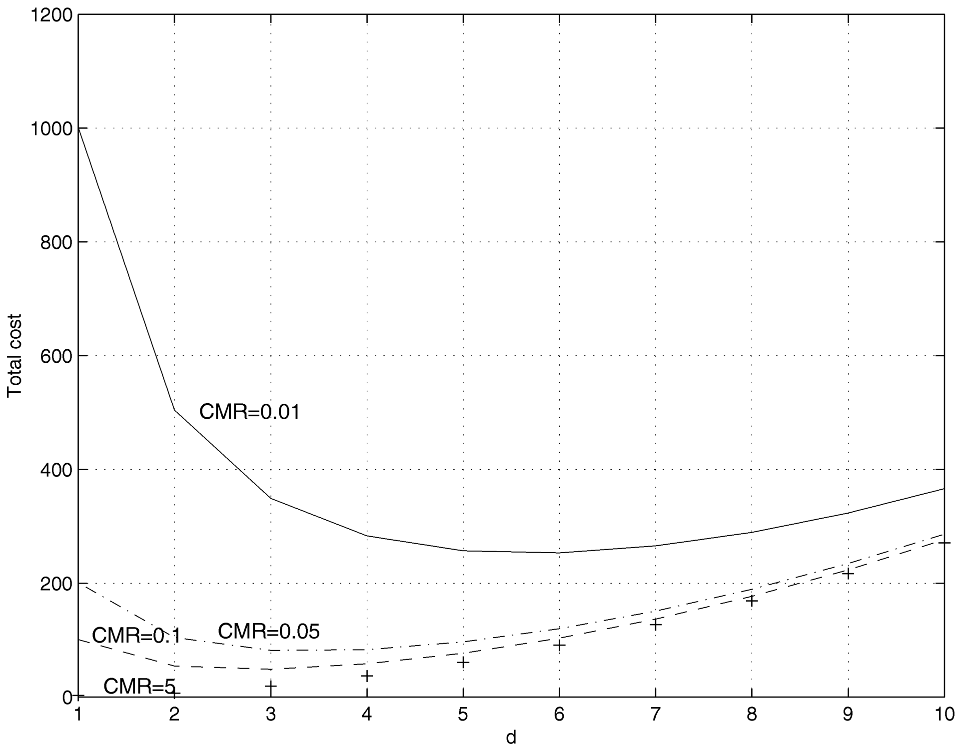


Fig. 3. Total cost versus the movement thresholds: Varying call-to-mobility ratio (CMR).

when the CMR is higher, more often the call arrivals are, resulting in more paging, hence more location updates may be necessary to balance the paging cost.

Figs. 4 and 5 show the cost function versus the movement threshold for varying variance of the cell residence time (or γ), where we choose $P = 1$ and $U = 20$. Fig. 4 shows a significant difference for different values of γ (hence, the variances of the cell residence time); the total cost is in fact sensitive to γ. This is also observed in Fig. 5 where the interservice time is Erlang distributed with $p = 5$. We do observe the insensitivity phenomena for certain parameter range: With the same set of parameters except

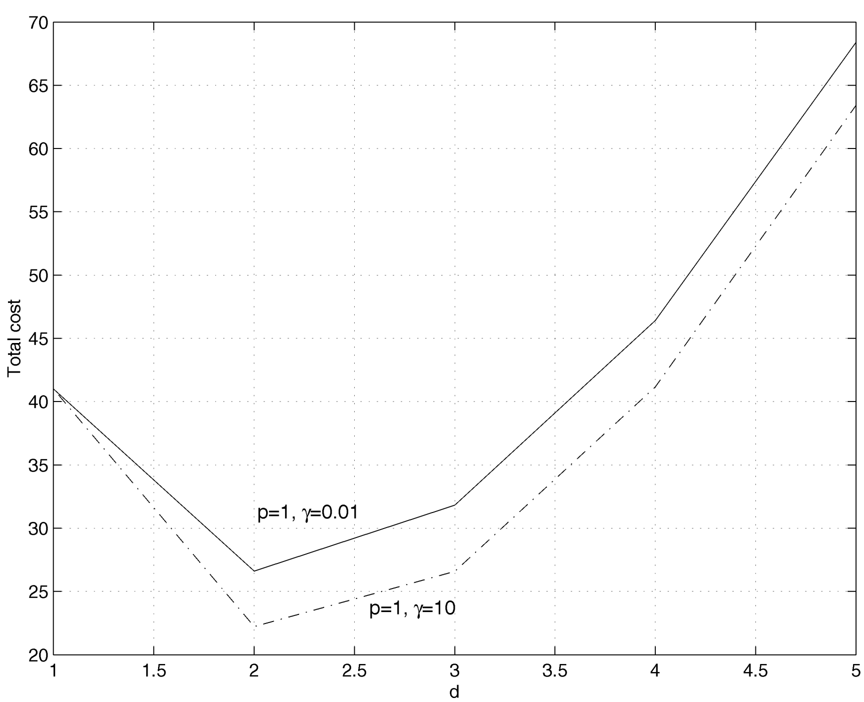


Fig. 4. Total cost versus the movement threshold for exponential interservice time: Varying variance of cell residence time.

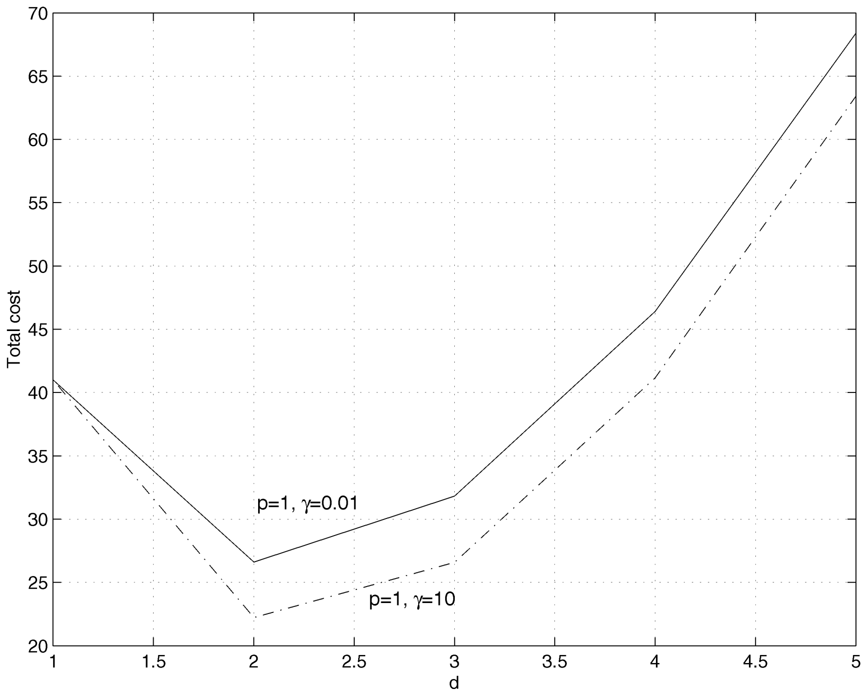


Fig. 5. Total cost versus the movement threshold for Erlang interservice time: Varying variance of cell residence time.

$U = 5$, the total cost is insensitive to γ, which is shown in Fig. 6. However, when $U = 20$, we will observe the significant differences for various γ (Fig. 7). Another question is whether the optimal threshold is sensitive to the variance of cell residence time. Fig. 8 shows the plot for varying the variance of the cell residence time when the interservice time is Erlang distributed. It is observed that the optimal thresholds for $\gamma = 0.01$ and $\gamma = 100$ are different. In many parameter settings we computed, the optimal costs for different variances of cell residence time are not significantly different. However, we are not able to draw the general conclusion, which makes the analytical formula we present more important.

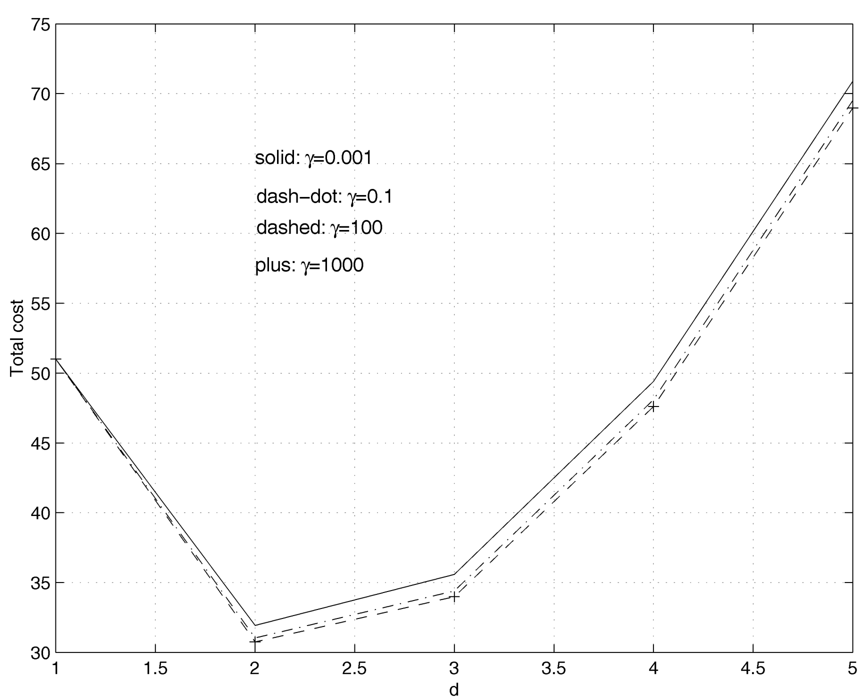


Fig. 6. Total cost versus the movement threshold for exponential interservice time: Varying variance of cell residence time.

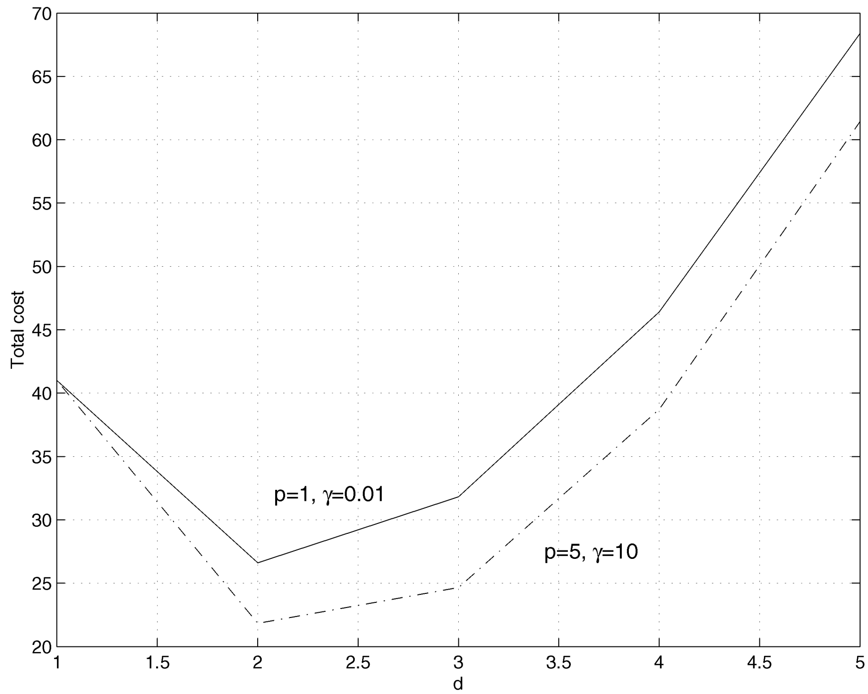


Fig. 7. Total cost versus the movement threshold for exponential interservice time: Varying variance of cell residence time.

In Fig. 9, we fix the average cell residence time and the average interservice time and change the variances of cell residence time and interservice time (i.e., γ and p), we again observe the significant differences between the total costs (even the optimal costs, i.e., the one corresponding to the optimal thresholds).

Next, we study the sensitivity issue of the total cost with respect to the variance of interservice time. Fig. 10 shows the plot for the total cost versus the

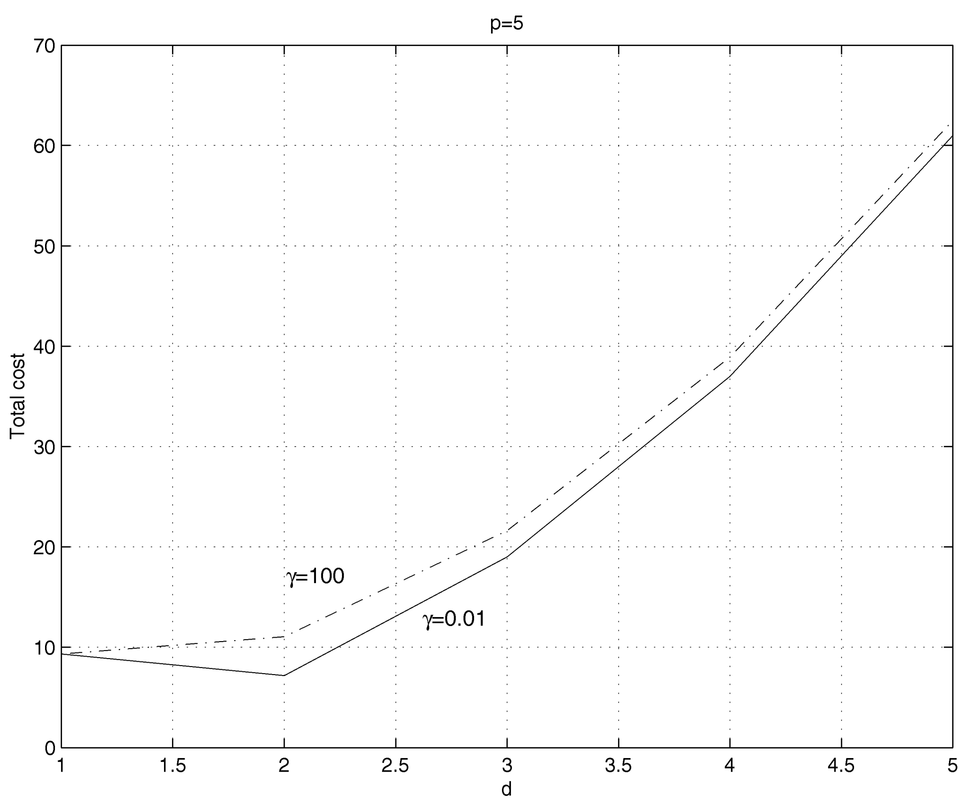


Fig. 8. Total cost versus the movement threshold: Varying variance of cell residence time.

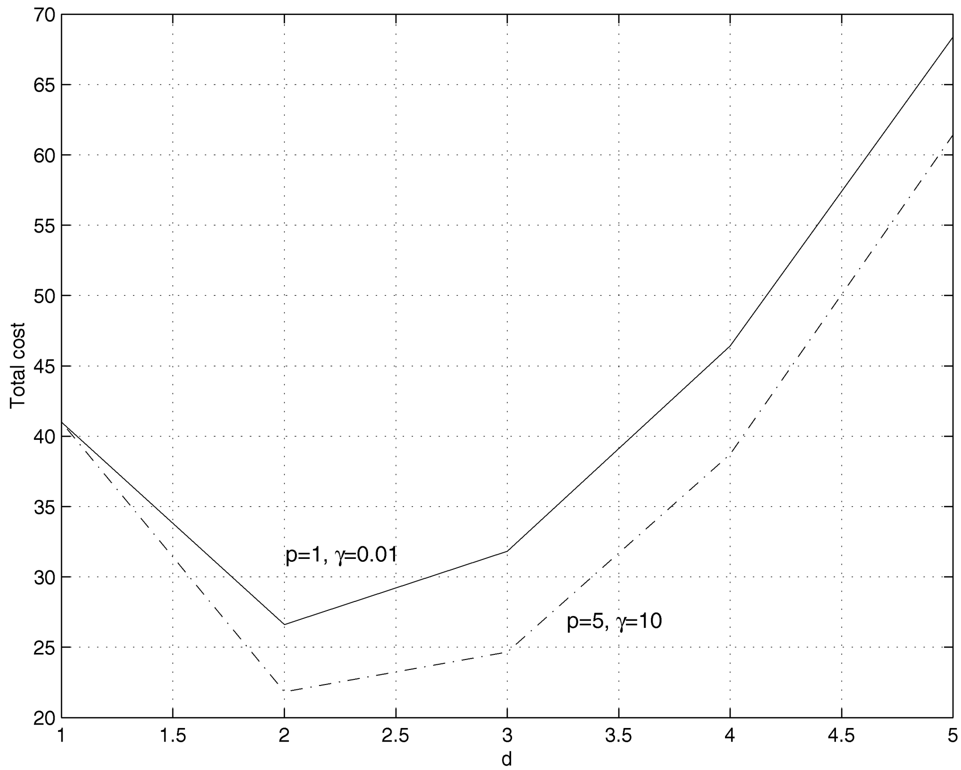


Fig. 9. Total cost versus the movement threshold for exponential interservice time: Varying variances of both cell residence time and interservice time.

movement thresholds with various values for p (corresponding to the variances of the interservice time), where we choose the $CMR = 0.5$. It is obvious that the total cost is insensitive to the variances when the interservice time is Erlang distributed. However, this is not always true. Fig. 11 shows the plots when

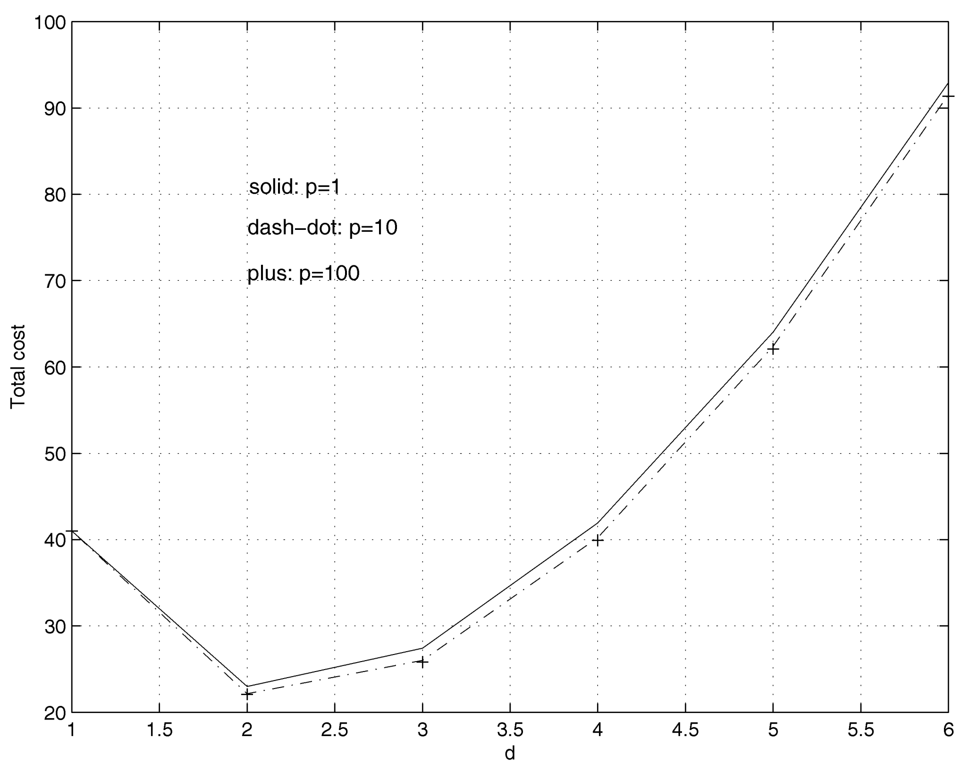


Fig. 10. Total cost vs. the movement threshold for Erlang distributed inter-service time: $p = 1, 10, 100$.

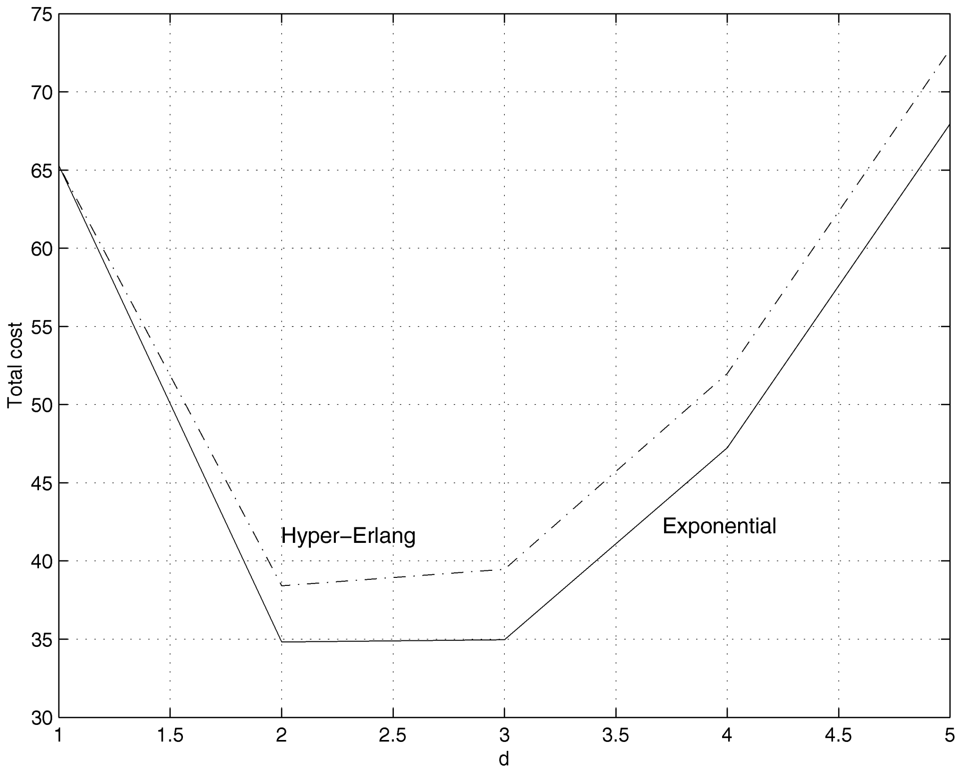


Fig. 11. Total cost versus the movement threshold when the interservice time is exponentially distributed and hyper-Erlang distributed, respectively.

the interservice time t_c is exponentially distributed and hyper-Erlang distributed with the same mean, respectively. In this example, we assume that hyper-Erlang distribution has the following parameters: $\alpha_1 = 0.15$, $\alpha_2 = 0.85$, $m_1 = 1$, $m_2 = 20$, $\eta_1 = 0.07$, $\eta_2 = 5000$, thus the hyper-Erlang distribution has mean $1/\lambda_c = \alpha_1/\eta_1 + \alpha_2/\eta_2 = 2.143$ and with variance $\alpha_1/(m_1\eta_1^2) + \alpha_2/(m_2\eta_2^2) = 30.6122$, while the exponential distribution will have mean $1/\lambda_c = 2.143$ and variance $1/\lambda_c^2 = 4.5926$. From the figure, we observe that total cost is indeed sensitive to the variance of the interservice time (the total cost for the hyper-Erlang case can be 12 percent higher.

One general observation is that our extensive numerical study shows that the total cost function is a convex function of the movement threshold and that the optimal total cost is not significantly sensitive to the variances of both cell residence time and interservice time. We conjecture that the total cost is a convex function of the movement threshold and that the optimal threshold can be appropriately approximated by the one when we use the exponential model for both cell residence time and interservice time. However, we have not found a rigorous proof yet. Fortunately, since our analytical results only involved the Laplace transforms of some nonnegative random variables, which can be approximated by rational functions, all involved computations are not computationally complex.

6 Conclusions

In this paper, we have performed the trade off analysis for the movement-based location update and paging for the wireless mobile networks. We have proposed a general framework for the study of such problems and analytically obtain the crucial quantity, the average number of location updates during an interservice time, used in all cost analysis, under fairly general assumptions. The analytical results and analytical approach for the cost analysis developed in this paper can be very useful in the design of mobility management for future wireless mobile networks. Our study shows that the total cost of the location update and paging is a convex function of the movement threshold, hence we can always find the unique optimal threshold. Our future research will focus on further study of the proposed general framework for the trade off analysis, particularly for the cases when call blocking due to paging delay is allowed.

Acknowledgments

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

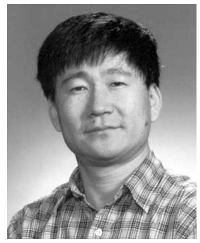
References

[1] A. Abutaleb and V.O.K. Li, "Paging Strategy Optimization in Personal Communication Systems," *Wireless Networks,* vol. 3, pp. 195-204, 1997.

[2] A. Abutaleb and V.O.K. Li, "Location Update Optimization in Personal Communication Systems," *Wireless Networks,* vol. 3, pp. 205-216 1997.

[3] I.F. Akyildiz and J.S.M. Ho, "Dynamic Mobile User Location Update for Wireless PCS Networks," *Wireless Networks,* vol. 1, pp. 187-196, 1995.

[4] I.F. Akyildiz, J.S.M. Ho, and Y.-B. Lin, "Movement-Based Location Update and Selective Paging for PCS Networks," *IEEE/ACM Trans. Networking,* vol. 4, no. 4, pp. 629-638, 1996.
[5] I.F. Akyildiz, J. McNair, J.S.M. Ho, H. Uzunalioglu, and W. Wang, "Mobility Management in Next-Generation Wireless Systems," *Proc. IEEE,* vol. 87, no. 8, pp. 1347-1384, Aug. 1999.
[6] A. Bar-Noy, I. Kessler, and M. Sidi, "Mobile Users: To Update or Not Update?" *Wireless Networks,* vol. 1, no. 2, pp. 175-186, July 1994.
[7] D.C. Cox, "Wireless Personal Communications: What Is It?" *IEEE Personal Comm. Magazine,* pp. 20-35, Apr. 1995.
[8] D.R. Cox, *Renewal Theory.* New York: John Wiley & Sons, 1962.
[9] Y. Fang, "General Modeling and Performance Analysis for Location Management in Wireless Mobile Networks," *IEEE Trans. Computers,* vol. 51, no. 10, pp. 1169-1181, Oct. 2002.
[10] Y. Fang and I. Chlamtac, "Teletraffic Analysis and Mobility Modeling for PCS Networks," *IEEE Trans. Comm.,* vol. 47, no. 7, pp. 1062-1072, July 1999.
[11] Y. Fang, I. Chlamtac, and H. Fei, "Analytical Results for Optimal Choice of Location Update Interval for Mobility Database Failure Restoration in PCS Networks," *IEEE Trans. Parallel and Distributed Systems,* vol. 11, no. 6, pp. 615-624, June 2000.
[12] Y. Fang, I. Chlamtac, and Y.B. Lin, "Portable Movement Modeling for PCS Networks" *IEEE Trans. Vehicular Technology,* vol. 49, no. 4, pp. 1356-1363, July 2000.
[13] Y. Fang, I. Chlamtac, and Y.B. Lin, "Call Performance for a PCS Network," *IEEE J. Selected Areas in Comm.,* vol. 15, no. 8, pp. 1568-1581, 1997.
[14] D. Hong and S.S. Rappaport, "Traffic Model and Performance Analysis for Cellular Mobile Radio Telephone Systems with Prioritized and Nonprioritized Handoff Procedures," *IEEE Trans. Vehicular Technology,* vol. 35, no. 3, pp. 77-92, 1986.
[15] B. Jabbari, "Teletraffic Aspects of Evolving and Next-Generation Wireless Communication Networks," *IEEE Comm. Magazine,* pp. 4-9, Dec. 1994.
[16] I. Katzela and M. Naghshineh, "Channel Assignment Schemes for Cellular Mobile Telecommunication Systems: A Comprehensive Survey," *IEEE Personal Comm.,* vol. 3, no. 3, pp. 10-31, June 1996.
[17] F.P. Kelly, *Reversibility and Stochastic Networks.* New York: John Wiley & Sons, 1979.
[18] L. Kleinrock, *Queueing Systems: Theory, Volume I.* New York: John Wiley & Sons, 1975.
[19] J. Li, H. Kameda, and K. Li, "Optimal Dynamic Mobility Management for PCS Networks," *IEEE/ACM Trans. Networking,* vol. 8, no. 3, pp. 319-327, 2000.
[20] Y.B. Lin, "Reducing Location Update Cost in a PCS Network," *IEEE/ACM Trans. Networking,* vol. 5, no. 1, pp. 25-33, 1997.
[21] Y.B. Lin, S. Mohan, and A. Noerpel, "Queueing Priority Channel Assignment Strategies for Handoff and Initial Access for a PCS Network," *IEEE Trans. Vehicular Technology,* vol. 43, no. 3, pp. 704-712, 1994.
[22] S. Mohan and R. Jain, "Two User Location Strategies for Personal Communications Services," *IEEE Personal Comm.,* pp. 42-50, first quarter 1994.
[23] C. Rose and R. Yates, "Minimizing the Average Cost of Paging under Delay Constraints," *Wireless Networks,* vol. 1, no. 2, pp. 211-219, 1995.
[24] S. Tabbane, "An Alternative Strategy for Location Tracking," *IEEE J. Selected Areas in Comm.,* vol. 13, no. 5, pp. 880-892, 1995.



Yuguang Fang (S'92-M'94-S'96-M'97-SM'99) received the BS and MS degrees in mathematics from Qufu Normal University, Qufu, Shandong, China, in 1984 and 1987, respectively, the PhD degree from the Department of Systems, Control, and Industrial Engineering at Case Western Reserve University, Cleveland, Ohio, in January 1994, and the PhD degree from the Department of Electrical and Computer Engineering, Boston University, in May 1997. From 1987 to 1988, he held research and teaching positions in both the Department of Mathematics and the Institute of Automation at Qufu Normal University. He held a postdoctoral position in the Department of Electrical and Computer Engineering at Boston University from June 1994 to August 1995. From June 1997 to July 1998, he was a visiting assistant professor in the Department of Electrical Engineering at the University of Texas at Dallas. From July 1998 to May 2000, he was an assistant professor in the Department of Electrical and Computer Engineering at the New Jersey Institute of Technology, Newark. Since May 2000, he has been an assistant professor in the Department of Electrical and Computer Engineering at the University of Florida, Gainesville. His research interests span many areas including wireless networks, mobile computing, mobile communications, automatic control, and neural networks. He has published more than 80 papers in refereed professional journals and conferences. He received the US National Science Foundation Faculty Early Career Development Award in 2001 and the US Office of Naval Research Young Investigator Award in 2002. He is listed in Marquis *Who's Who in Science and Engineering*, *Who's Who in America*, and *Who's Who in the World*. Dr. Fang has actively engaged in many professional activities. He is a senior member of the IEEE and a member of the ACM. He is an editor for the *IEEE Transactions on Communications*, an editor for the *IEEE Transactions on Wireless Communications*, an editor for *ACM Wireless Networks*, an area editor for *ACM Mobile Computing and Communications Review*, an associate editor for the *Wiley International Journal on Wireless Communications and Mobile Computing*, and feature editor for Scanning the Literature in *IEEE Personal Communications*. He was an editor for the *IEEE Journal on Selected Areas in Communications: Wireless Communications Series*. He has also been actively involved with many professional conferences such as ACM MobiCom'02, ACM MobiCom'01, IEEE INFOCOM'00, INFOCOM'98, IEEE WCNC'02, WCNC'00 (Technical Program Vice-Chair), WCNC'99, and the International Conference on Computer Communications and Networking (IC3N'98) (technical program vicechair).