A Mobile User Location Update and Paging Mechanism Under Delay Constraints

Ian F. Akyildiz and Joseph S.M. Ho

Broadband and Wireless Networking Laboratory (B_W_N Lab)
School of Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, GA 30332, U.S.A.
Tel: (404)-894-5141; Fax: (404)-853-9410
E-Mail: ian@armani.gatech.edu; joseph@ee.gatech.edu

Abstract

A mobile user location management mechanism is introduced that incorporates a distance based location update scheme and a paging mechanism that satisfies predefined delay requirements. An analytical model is developed which captures the mobility and call arrival pattern of a terminal. Given the respective costs for location update and terminal paging, the average total location update and terminal paging cost is determined. An iterative algorithm is then used to determine the optimal location update threshold distance that results in the minimum cost. Analytical results are also obtained to demonstrate the relative cost incurred by the proposed mechanism under various delay requirements.

1 Introduction

Personal communication networks (PCNs) consist of a fixed wireline network and a large number of mobile terminals. These terminals include telephones, portable computers, and other devices that exchange information with remote terminals through the fixed network. The wireline network can be the telephone network in use today or the ATM network in the future. In this paper, we do not make a specific assumption on the wireline network. However, it must have sufficient capacity to carry the traffic generated by the terminals in an efficient manner.

In order to effectively utilize the very limited wireless bandwidth to support an increasing number of mobile subscribers, current PCNs are designed based on a cellular architecture. The PCN coverage area is divided into a large number of smaller areas called *cells*. Terminals within a cell communicate with the wireline network through a base station which is installed inside the cell. This base station serves as the network access point (NAP) for all the terminals within the cell. As soon as a terminal enters another cell, its NAP is switched to the base station of the newly entered cell. As terminals are free to travel from cell to cell, a mechanism is needed to effectively keep track of the location of each terminal. When an incoming call arrives, the wireline network must be able to determine the exact location of the destination terminal in a timely fashion without incurring excessive computation and communication costs.

Current cellular networks partition their coverage area into a number of *location areas* (LAs). Each LA consists of a group of cells and each terminal reports its location to the network whenever it enters an LA. This reporting process is called *location update*. When an incoming call arrives, the network locates the terminal by simultaneously polling all cells within the LA. This polling process is called *terminal paging*. Both the location update and the terminal paging processes require a certain amount of wireless bandwidth. In addition, significant power is consumed by the terminal to keep track of its location and to transmit update messages. As a result, *costs* are associated with both the location update and the terminal paging processes. It is clear that if each LA consists of only one cell, the network knows exactly the location of each terminal. In this case, the cost for terminal paging is minimal. However, the cost for location update will be very high as the terminal has to report its location whenever it enters a cell. A tradeoff, therefore, exists between the location update cost and the terminal paging cost. It is desirable to select a location update and terminal paging policy that can minimize the total cost.

A method for calculating the optimal LA size given the respective costs for location update and terminal paging is introduced in [8]. However, under the LA based scheme, terminals located close to an LA boundary may perform excessive location updates as they move back and forth between two LAs. Besides, the optimal LA size should be terminal dependent as mobility and calling patterns vary among users. It is not generally easy to use different LA sizes for different terminals.

Three location update schemes are examined in [3]: *time based*, *movement based* and *distance based.* Under these three schemes, location updates are performed based on the time elapsed, the number of movements performed, and the distance traveled, respectively, since the last location update. Results demonstrated that the distance based scheme produces the best result. However, the model considered in [3] is very simplified. For example, in [3] incoming calls are not taken into account. Moreover, paging delay is not constrained. A distance based location update scheme is introduced in [6] where an iterative algorithm that can generate the optimal threshold distance resulting in the minimum cost. However, the number of iterations required for the algorithm to converge varies widely depending on the mobility and call arrival probability parameters considered. Besides, as in [3], paging delay is not constrained. The time required to locate a mobile terminal is directly proportional to the distance traveled by the mobile terminal since its last location update. A dynamic location update mechanism is introduced in [1] where the location update time based on data obtained on-line is dynamically determined. It is demonstrated that the result obtained is close to the optimal result given in [6]. Computation required by this mechanism is minimal. It is, therefore, feasible to implement this scheme in mobile terminals that have limited computing power. Similar to other schemes described above, the drawback of this scheme is that paging delay is not explicitly considered.

In [7], paging subject to delay constraints is considered. Results demonstrate that when delay is unconstrained, the highest-probability-first scheme incurs the minimum cost. For the constrained delay case, the authors determine the optimal polling sequence that results in minimum cost. The authors, however, assume that the probability distribution of user location is provided. This probability distribution may be user dependent. A location update and terminal paging scheme that facilitates derivation of this probability distribution is needed in order to apply the paging scheme given in [7]. Besides, the trade off between location update and terminal paging is not considered in [7].

In this paper, we introduce a location management scheme that combines a distance based location update mechanism with a paging scheme which guarantees a pre-defined maximum delay requirement. We propose a Markovian model which captures the mobility and call arrival patterns of a particular terminal. Based on this model, we obtain the average location update and terminal paging costs. We use an iterative algorithm to determine the optimal location update threshold distance that will result in the minimum average total cost. We provide numerical results that demonstrate the trade off between high and low delay bounds under various parameter values.

This paper is organized as follows. In Section 2, we describe the mobility model and the location update and paging schemes. Sections 3 and 4 describe a Markov chain model for the one- and two-dimensional PCN coverage area, respectively. Solutions for the steady state probabilities of the Markov chain are also presented. Section 5 describes a method for deriving the total location update and paging cost. A method for obtaining the optimal location update threshold distance is given in Section 6. Section 7 presents the numerical results, the conclusion is given in Section 8.

2 System Description

2.1 Terminal Mobility

In this paper, we consider both one- and two-dimensional mobility models. The one-dimensional model is suitable for situations where the mobility of mobile terminals is restricted to two directions (forward and backward). Examples include roads, tunnels, train and train stations. The two-dimensional model is suitable for more general case where terminals can travel in any direction within a coverage area (such as a city). We assume that the PCN coverage area is divided into cells according to:

One-Dimensional Model: The one-dimensional space is divided into cells of the same length. Each cell has two neighbors. Figure 1(a) shows the cell partition in a one-dimensional coverage area. The numbers represent the distance of each cell from cell 0 and are discussed later in this section.

Two-Dimensional Model: The two dimensional space is divided into hexagonal cells of the same size. Each cell has six neighbors. Figure 1(b) shows the cell partition in the two-dimensional coverage area.

The size of each cell is determined based on the number of mobile subscribers, the number of channels available per cell and the channel allocation scheme used. In this paper we concentrate on finding the optimal location update distance assuming that the size of cells is given. Our scheme works both in the *macrocell* and the *mi-*

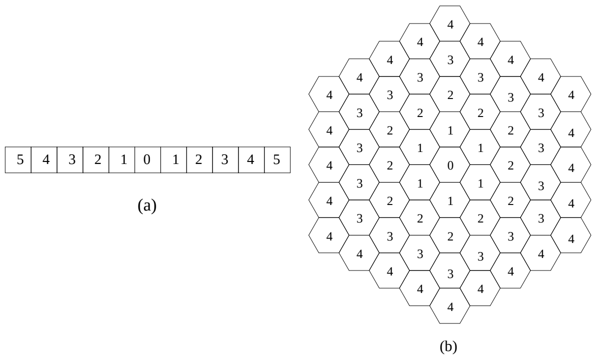

Figure 1: (a) One-Dimensional Model and (b) Two-Dimensional Model.

crocell environment. We note that if the size of cells is small, the probability of movement (to be discussed later in this section) will be high and vice versa. We consider a wide range of values for the probability of movement in the numerical examples given in Section 7.

Here we introduce the concept of a *ring*. As it is shown in Figure 1(b), each cell is surrounded by rings of cells. If we select cell 0 (as indicated) as the center, cells labeled '1' form the first ring around cell 0. Cells labeled '2' form the second ring around cell 0, and so on. Each ring is labeled according to its distance from the center such that ring r_1 refers to the first ring away from cell 0. In general, if we consider cell 0 as a ring itself, then ring r_i $(i = 0, 1, 2...)$ refers to the i^{th} ring away from the center. We assume that the distance of each cell from the center cell is measured in terms of number of rings such that cells in ring r_i are i rings away from the center. The distance (in terms of the number of rings) of each cell from cell 0 is given in Figure 1(b). For the one-dimensional model, even though cells do not form physical rings, we will use the same terminology as in the two-dimensional case. As shown in Figure 1(a), cells labeled i belong to ring r_i. The distances (in term of the number of rings) of each cell from cell 0 is indicated in Figure 1(a). Unless specified otherwise, all distances mentioned in this paper are measured in terms of rings. We will not explicitly indicate the unit of distance hereafter.

We define $g(d)$ to be the number of cells that are within a distance of d from any given cell (including this cell) in the coverage area:

$$g(d) = \begin{cases} 2d+1 & \text{for 1-dim. model, } d = 0, 1, 2... \\ 3d(d+1)+1 & \text{for 2-dim. model, } d = 0, 1, 2... \end{cases} \tag{1}$$

Mobile users travel from cell to cell within the PCN coverage area. In this paper, we assume the mobility of each terminal to follow a discrete-time random walk model as described below:

- At discrete times t, a user moves to one of the neighboring cells with probability q or stay at the current cell with probability $1 - q$.
- If the user decides to move to another cell, there is equal probability for any one of the neighboring cells to be selected as the destination. This probability is $\frac{1}{2}$ in the one-dimensional model and $\frac{1}{6}$ in the two-dimensional model.

As compared to the fluid flow model reported in [8], the random walk model is more appropriate as most of the mobile subscribers in a PCN are likely to be pedestrians. The fluid flow model is more suitable for vehicle traffic such that a continuous movement with infrequent speed and direction changes are expected. For pedestrian movements such that mobility is generally confined to a limited geographical area while frequent stop-and-go as well as direction changes are common, the random walk model is more appropriate. Similar random walk models are also reported in [1, 3, 6]. Incoming calls may arrive during each discrete time t. We assume that the incoming call arrivals for each mobile terminal are geometrically distributed and that the probability of a call arrival during each discrete time t is c. We also assume that the location identifier of a cell is broadcast by the base station periodically so that every mobile terminal knows exactly its own location at any given time. Each mobile terminal reports its location to the network according to the location update scheme to be described in Section 2.2. The network stores each mobile terminal's location in a database whenever such information is available.

2.2 Location Update and Terminal Paging Mechanism

We define the *center cell* (cell 0) of a mobile terminal to be the cell at which the terminal last reported its location to the network. A distance based location update mechanism is used such that a terminal will report its location when its distance from the center cell exceeds a threshold d. This location update scheme guarantees that the terminal is located in an area that is within a distance d from the center cell. This area is called the *residing area* of the terminal.

In order to determine if a terminal is located in a particular cell, the network performs the following steps:

1. Sends a polling signal to the target cell and waits until a timeout occurs.

2. If a reply is received before timeout, the destination terminal is located in the target cell.

3. If no reply is received, the destination terminal is not in the target cell.

We call the above process the *polling cycle*. For simplicity, we define a maximum paging delay of m to mean that the network must be able to locate the destination terminal within m polling cycles. When an incoming call arrives, the network first partitions the residing area of the terminal into a number of subareas (a method for partitioning the residing area into subareas will be described next) and polls each subarea one after another until the terminal is found.

We denote subarea j by A_j where the subscript indicates the order in which the subarea will be polled. Each subarea contains one or more rings and each ring cannot be included in more than one subarea. As discussed in Section 2.1, ring r_i $(i = 0, 1, 2...)$ consists of all cells that are a distance of i away from the center cell. For a threshold distance of d, the number of rings in the residing area is $(d + 1)$. If paging delay is not constrained, we assign exactly one ring to each subarea. The residing area is, therefore, partitioned into $(d + 1)$ subareas. However, if paging delay is constrained to a maximum of m polling cycles, the number of subareas cannot exceed m. Otherwise, we may not be able to locate the terminal in less than or equal to m polling cycles. Given a threshold distance of d and a maximum paging delay of m, the number of subareas, denoted by ℓ, is:

$$\ell = \min(d + 1, m) \tag{2}$$

We partition the residing area according to the following steps:

1. Determine the number of rings in each subarea by $\eta = \lfloor \frac{d+1}{\ell} \rfloor$.

2. Assign η rings to each subarea (except the last subarea) such that subarea A_j $(1 \le j \le \ell - 1)$ is assigned rings $r_{\eta(j-1)}$ to $r_{\eta j-1}$.

3. Assign the remaining rings to subarea A_ℓ.

The partitioning scheme described above is based on a shortest-distance-first (SDF) order such that rings closer to the center cell are polled first. It is shown in [7] that in order to minimize paging cost, the more probable locations should be polled first. Under the mobility model described in subsection 2.1, in most cases there is a higher probability of finding a terminal in a cell that is closer to the center cell than in a cell that is further away. Our partitioning scheme is, therefore, analogous to a more-probable-first scheme. It is shown in Section 7 that under our partitioning scheme, significant gain can be achieved when the maximum paging delay is increased only slightly from its minimum value of one polling cycle. We stress that our method of determining the optimal threshold distance is not limited to this SDF scheme. Since we can determine the probability distribution of terminal location in the residing area, our mechanism can determine the optimal threshold distance when any other partitioning schemes are used.

In the following sections, we determine the optimal threshold distance that results in the minimum total location update and terminal paging cost. The following two steps are taken in deriving this optimal threshold distance:

- Determine the probability distribution of terminal location within the residing area using a Markov chain model. This information is useful in finding the average total cost.

- Determine the average total cost as a function of threshold distance and maximum paging delay and locate the optimal threshold distance using an iterative algorithm.

3 One-Dimensional Mobility Model

3.1 Markov Chain Model

We setup a discrete-time Markov chain model to capture the mobility and call arrival patterns of a terminal. Figure 2 gives the Markov chain model when the location update threshold distance is d. The state of the Markov chain i $(i \geq 0)$ is defined as the distance between the current location of the mobile terminal and its center cell. This state is equivalent to the index of the ring in which the mobile terminal is located. As a result, the mobile terminal is in state i if it is currently residing in ring r_i. The transition probabilities $a_{i,i+1}$ and $b_{i,i-1}$ represent the probabilities at which the distance of the terminal from its center cell increases and decreases, respectively. As described before, c denotes the probability of a call arrival. Transitions from a state to one of its two neighboring states represent movements of the terminal away from a cell. A transition from any state to state 0 represents either the arrival of a call or the occurrence of a location update when the threshold distance d is exceeded. When a call arrives, the network determines the current location of the mobile terminal by paging and, as a result, the center cell is reset to the current cell location of the mobile terminal. Similarly, the center cell is reset when a location update occurs.

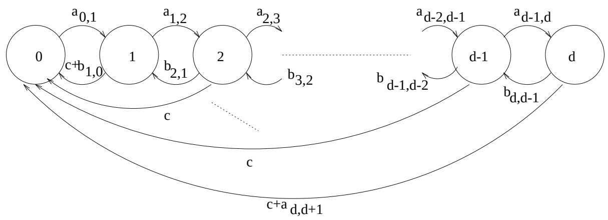

Figure 2: Markov chain model.

In both cases, the center cell becomes the current location of the mobile terminal and the new state of the terminal is therefore 0. The transition probabilities of the discrete-time Markov chain are given as:

$$a_{i,i+1} = \begin{cases} q & \text{if } i = 0 \\ \frac{q}{2} & \text{if } 1 \leq i \leq d \end{cases} \tag{3}$$

$$b_{i,i-1} = \frac{q}{2} \tag{4}$$

We assume $p_{i,d}$ $(0 \leq i \leq d)$ to be the steady state probability of state i when the maximum threshold distance is d. The balance equations for the Markov chain are given as:

$$p_{0,d}a_{0,1} = p_{1,d}b_{1,0} + p_{d,d}a_{d,d+1} + c\sum_{k=1}^{d} p_{k,d} \tag{5}$$

$$p_{d,d}(a_{d,d+1} + b_{d,d-1} + c) = p_{d-1,d}a_{d-1,d} \tag{6}$$

$$p_{i,d}(a_{i,i+1} + b_{i,i-1} + c) = p_{i-1,d}a_{i-1,i} + p_{i+1,d}b_{i+1,i} \quad \text{for } 0 < i < d \tag{7}$$

Given the balance equations (5-7) and the transition probability equations (3) and (4), the steady state probabilities of the Markov chain can be obtained. The closed form expressions for the steady state probabilities are given in the next subsection. The probability of state d is particularly important in determining the location update cost. As will be described in Section 5, the location update cost can be obtained given the value of $p_{d,d}$. In the following subsection, we will first solve for the equation of $p_{d,d}$. The steady state probabilities of other states will be derived in terms of $p_{d,d}$. These closed form expressions allow us to determine the location update and terminal paging costs in a more effective manner.

3.2 Steady State Probabilities

Based on equation (7), we can obtain the steady state probabilities for state $(i+1)$ $(1 \leq i \leq d-1)$ in terms of the steady state probabilities of states i and $(i-1)$ according to the following expression:

$$p_{i+1,d} = \frac{p_{i,d}(a_{i,i+1} + b_{i,i-1} + c) - p_{i-1,d}a_{i-1,i}}{b_{i+1,i}} \quad \text{for } 0 < i < d \tag{8}$$

Substituting the transition probabilities given by equations (3) and (4) in equation (8), the following equation for $p_{i+1,d}$ is obtained:

$$p_{i+1,d} = \alpha p_{i,d} - p_{i-1,d} \qquad \text{for } 1 < i < d \tag{9}$$

where α is given as:

$$\alpha = 2 + \frac{2c}{q} \tag{10}$$

Equation (9) can be applied recursively to obtain the steady state probabilities for state i (where $2 \leq i \leq d$) in terms of the steady state probabilities of states 1 and 2:

$$p_{i,d} = S_{i-2}p_{2,d} - S_{i-3}p_{1,d} \qquad \text{for } 2 < i < d \tag{11}$$

where S_i is defined recursively as:

$$S_i = \begin{cases} 1 & \text{if } i = 0 \\ \alpha & \text{if } i = 1 \\ \alpha S_{i-1} - S_{i-2} & \text{if } 2 \leq i \leq d \end{cases} \tag{12}$$

To solve for the closed form expression for S_i, we apply Z-transform to the last part of equation (12). Multiplying each side of equation (12) by z^i and summing both sides of the equation from 2 to ∞, we obtain the following:

$$\sum_{i=2}^{\infty} S_i z^i = \alpha \sum_{i=2}^{\infty} S_{i-1} z^i - \sum_{i=2}^{\infty} S_{i-2} z^i \tag{13}$$

We can now make use of the initial conditions of S_i as given in equation (12) to obtain the Z-transform of S_i, denoted by $S(z)$, as:

$$S(z) = \sum_{k=0}^{\infty} S_i z^i = \frac{1}{z^2 - \alpha z + 1} \tag{14}$$

Taking the inverse Z-transform of the above equation, the closed form equation for S_i can be obtained as:

$$S_i = \frac{e_1^{i+1} - e_2^{i+1}}{e_1 - e_2} \tag{15}$$

where e_1 and e_2 are the roots of the quadratic equation $z^2 - \alpha z + 1$, which is the denominator of $S(z)$, and is given as:

$$e_1 = \frac{1}{2}(\alpha + \sqrt{\alpha^2 - 4}) \tag{16}$$

$$e_2 = \frac{1}{2}(\alpha - \sqrt{\alpha^2 - 4}) \tag{17}$$

Making use of equations (11) and (15), we can obtain the steady state probabilities of states d and $(d-1)$ as follows:

$$p_{d,d} = \frac{e_1^{d-1} - e_2^{d-1}}{e_1 - e_2} p_{2,d} - \frac{e_1^{d-2} - e_2^{d-2}}{e_1 - e_2} p_{1,d} \quad \text{for } d > 1 \tag{18}$$

$$p_{d-1,d} = \frac{(e_1^{d-2} - e_2^{d-2})}{e_1 - e_2} p_{2,d} - \frac{(e_1^{d-3} - e_2^{d-3})}{e_1 - e_2} p_{1,d} \quad \text{for } d > 2 \tag{19}$$

These two equations will be used, together with the equations to be derived later, to solve for the steady state probabilities of the Markov chain.

Rearranging equations (5) and (7) with $i = 1$, results in the following expression for $p_{d,d}$ in terms of $p_{1,d}$ and $p_{2,d}$:

$$p_{d,d} = (\frac{1}{2}\alpha^2 - 1)p_{1,d} - \frac{1}{2}\alpha p_{2,d} - (\alpha - 2) \quad \text{for } d > 1 \tag{20}$$

Another expression for $p_{d,d}$ can be obtained by substituting the transition probabilities in equation (6) as follows:

$$p_{d,d} = \frac{1}{\alpha} p_{d-1,d} \qquad \text{for } d > 1 \tag{21}$$

Substitute $p_{d-1,d}$ as given by equation (19) in equation (21) results:

$$p_{d,d} = \frac{(e_1^{d-2} - e_2^{d-2})}{\alpha(e_1 - e_2)} p_{2,d} - \frac{(e_1^{d-3} - e_2^{d-3})}{\alpha(e_1 - e_2)} p_{1,d} \quad \text{for } d > 2 \tag{22}$$

Equations (18), (20) and (22) are three different expressions for $p_{d,d}$ as a function of $p_{1,d}$ and $p_{2,d}$. Since we have three equations and three unknowns, the solution for $p_{d,d}$ can be obtained:

$$p_{d,d} = K_1 \frac{R_1 R_3 - R_2^2}{K_2 R_1 + K_3 R_2 + K_4 R_3 + 2R_1 R_3 - 2R_2^2} \quad \text{for } d > 2 \tag{23}$$

where R_i and K_1 to K_4 are defined as:

$$R_i = e_1^{d-i} - e_2^{d-i} \tag{24}$$
$$K_1 = -2(\alpha - 2) \tag{25}$$
$$K_2 = (\alpha^3 - 2\alpha)(e_1 - e_2) \tag{26}$$
$$K_3 = (-2\alpha^2 + 2)(e_1 - e_2) \tag{27}$$
$$K_4 = \alpha(e_1 - e_2) \tag{28}$$

We can further obtain the expressions for $p_{0,d}$, $p_{1,d}$ and $p_{2,d}$ in terms of $p_{d,d}$ using equations (18), (20), (22) and (7). The results are given below:

$$p_{0,d} = \frac{R_1\alpha^2 - 2R_2\alpha + R_3}{2(R_2^2 - R_1 R_3)} R_{d-1} p_{d,d} \qquad \text{for } d > 2 \tag{29}$$

$$p_{1,d} = \frac{R_1\alpha - R_2}{R_2^2 - R_1 R_3} R_{d-1} p_{d,d} \qquad \text{for } d > 2 \tag{30}$$

$$p_{2,d} = \frac{R_2\alpha - R_3}{R_2^2 - R_1 R_3} R_{d-1} p_{d,d} \qquad \text{for } d > 2 \tag{31}$$

Substituting these $p_{1,d}$ and $p_{2,d}$ expressions in equation (11) gives the following equation for $p_{i,d}$:

$$p_{i,d} = \frac{S_{i-2}(R_2\alpha - R_3) - S_{i-3}(R_1\alpha - R_2)}{R_2^2 - R_1 R_3} R_{d-1} p_{d,d} \quad \text{for } 2 < i < d, \ \ d > 2 \tag{32}$$

Since the transition probability $a_{i,i+1}$ has a different value when $i = 0$ compared to when $i > 0$, the steady state probability equations (23)-(32) are valid only for threshold distance larger than 2. When the threshold distance is smaller than or equal to 2, we solve the steady state probabilities directly from the balance equations. The steady state probabilities for the threshold distances of 0, 1 and 2 are given below:

$d = 0$:

$$p_{0,0} = 1.0 \tag{33}$$

$d = 1$:

$$p_{0,1} = \frac{q + c}{2q + c} \tag{34}$$
$$p_{1,1} = \frac{q}{2q + c} \tag{35}$$

$d = 2$:

$$p_{0,2} = \frac{2c + q}{2c + 3q} \tag{36}$$
$$p_{1,2} = \frac{4q(c + q)}{9q^2 + 12qc + 4c^2} \tag{37}$$
$$p_{2,2} = \frac{2q^2}{9q^2 + 12qc + 4c^2} \tag{38}$$

When the threshold distance d is greater than or equal to 3, equation (32) gives the steady state probability for states 3 through $(d-1)$. Equations (29), (30), (31) and (23) give the steady state probabilities for states 0,1,2 and d, respectively. When the threshold distance d is smaller than 3, equations (33) to (38) are used to determine the steady state probabilities.

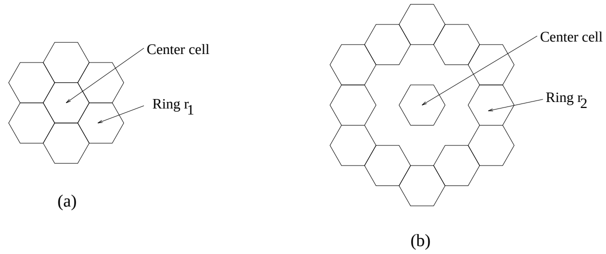

Figure 3: (a) Ring r_1 and (b) Ring r_2 in the Two-Dimensional Model.

4 Two-Dimensional Mobility Model

4.1 Markov Chain Model

The Markov chain model given in Figure 2 and the balance equations (5)-(7) are valid for the two-dimensional model. However, in the two-dimensional model, the transition probabilities $a_{i,i+1}$ and $b_{i,i-1}$ are state dependent. Figure 3(a) shows the cells in ring r_1. There are 6 hexagonal cells in ring r_1 and each cell has 6 edges. This results in a total of 36 edges. It can be seen from Figure 3(a) that 12 of these edges are between cells belonging to the same ring. There are 6 edges at the inside perimeter of the ring and there are 18 edges at the outside perimeter of the ring. Given that the terminal is at one of the cells in ring r_1, the probabilities that a movement will result in an increase or decrease of the distance from the center cell are $\frac{1}{2}$ and $\frac{1}{6}$, respectively. Similarly, Figure 3(b) shows the cells in ring r_2. If the terminal is located in ring r_2, the probabilities that a movement will result in an increase or decrease of the distance from the center cell are $\frac{5}{12}$ and $\frac{1}{4}$, respectively. In general, given that the mobile terminal is located in ring r_i, the probabilities that a movement will result in an increase or decrease in the distance from the center cell, denoted by $p^+(i)$ and $p^-(i)$, respectively, are given as:

$$p^+(i) = \frac{1}{3} + \frac{1}{6i} \tag{39}$$

$$p^-(i) = \frac{1}{3} - \frac{1}{6i} \tag{40}$$

The transition probabilities for the two-dimensional model are given as follows:

$$a_{i,i+1} = \begin{cases} q & \text{if } i = 0 \\ q(\frac{1}{3} + \frac{1}{6i}) & \text{if } 1 \le i \le d \end{cases} \tag{41}$$

$$b_{i,i-1} = q(\frac{1}{3} - \frac{1}{6i}) \tag{42}$$

Given the balance equations and the above transition probability equations, we can solve for the steady state probabilities recursively by expressing the probability of each state in terms of the probability of state d. The law of total probability can then be used to solve for the value of $p_{d,d}$. We refer to the solution obtained by this recursive method the *exact* solution. For the two-dimensional model, however, a simple closed form expression for the steady state probabilities may be difficult to obtain. In the next subsection, we will solve for the approximate closed form solutions for the steady state probabilities. This approximate solution is useful in determining the optimal threshold distance.

4.2 Approximate Solution for Steady State Probabilities

In order to obtain the approximate steady state probabilities for the two-dimensional model, we modify the transition probability equations (41) and (42) as follows:

$$a_{i,i+1} = \begin{cases} q & \text{if } i = 0 \\ \frac{q}{3} & \text{if } 1 \le i \le d \end{cases} \tag{43}$$

$$b_{i,i-1} = \frac{q}{3} \tag{44}$$

Here we omit the $\frac{q}{6i}$ terms in $a_{i,i+1}$ and $b_{i,i-1}$ expressions. The effect of this modification is more significant when i is small. When i is large, the value of $\frac{q}{6i}$ is negligible. For example, for $q = 0.1$ and $i = 5$, $\frac{q}{6i} = 3.33 \times 10^{-3}$. In the macrocell environment where the size of cells is large, the movement probability q is relatively small. The value of $\frac{q}{6i}$ is small even for small values of i. In the microcell environment where size of cells is small, the probability of moving to a neighboring cell, q, is relatively high. Because of the large probability of movement, cells are usually further away from the center cell. This means that the state, i, of a mobile terminal is relatively large and $\frac{q}{6i}$ is small because of a large value of i. In Section 7 we demonstrate that the optimal threshold distance obtained using these approximate transition probabilities differs from that obtained using the exact transition probabilities only by 1 ring or less most of the times.

Exactly the same steps (as in the one-dimensional model) can be used to solve for the approximate closed form expressions here. We, therefore, skip the steps and the approximate steady state equations are given as follows:
$d \ge 3$:

$$p_{d,d} = K_1 \frac{R_2^2 - R_1 R_3}{K_2 R_1 + K_3 R_2 + K_4 R_3 - 3R_1 R_3 + 3R_2^2} \tag{45}$$

$$p_{0,d} = \frac{R_1 \alpha^2 - 2R_2 \alpha + R_3}{3(R_2^2 - R_1 R_3)} R_{d-1} p_{d,d} \tag{46}$$

$$p_{1,d} = \frac{R_1\alpha - R_2}{R_2^2 - R_1 R_3} R_{d-1} p_{d,d} \quad (47)$$

$$p_{2,d} = \frac{R_2\alpha - R_3}{R_2^2 - R_1 R_3} R_{d-1} p_{d,d} \quad (48)$$

$$p_{i,d} = \frac{S_{i-2}(R_2\alpha - R_3) - S_{i-3}(R_1\alpha - R_2)}{R_2^2 - R_1 R_3} R_{d-1} p_{d,d} \quad \text{for } 2 < i < d \quad (49)$$

The definitions of R_i, e_1 and e_2 are the same as in the one-dimensional model. Their expressions are given by equations (24), (16) and (17), respectively. In this case, α and K_1 to K_4 are defined as:

$$\alpha = 2 + \frac{3c}{q} \quad (50)$$

$$K_1 = -3(\alpha - 2) \quad (51)$$

$$K_2 = (3\alpha - \alpha^2 - \alpha^3)(e_1 - e_2) \quad (52)$$

$$K_3 = (2\alpha^2 + 2\alpha - 3)(e_1 - e_2) \quad (53)$$

$$K_4 = -(\alpha + 1)(e_1 - e_2) \quad (54)$$

Boundary cases:
$d = 0$:

$$p_{0,0} = 1.0 \quad (55)$$

$d = 1$:

$$p_{0,1} = \frac{2q + 3c}{5q + 3c} \quad (56)$$

$$p_{1,1} = \frac{3q}{5q + 3c} \quad (57)$$

$d = 2$:

$$p_{0,2} = \frac{3c + q}{3c + 4q} \quad (58)$$

$$p_{1,2} = \frac{q(3c + 2q)}{4q^2 + 7qc + 3c^2} \quad (59)$$

$$p_{2,2} = \frac{q^2}{4q^2 + 7qc + 3c^2} \quad (60)$$

5 Location Update and Terminal Paging Costs

The steady state probabilities obtained in the previous section allow us to determine the cost associated with both location update and terminal paging for a given threshold distance and maximum paging delay parameters. Now we first look into the case where maximum paging delay is at its minimum of one polling cycle. This delay bound is of special interest because a delay of one polling cycle is also provided by the LA based scheme [8]. Results can be used to compare our scheme with the LA based scheme. We will also extend our result to obtain the terminal paging cost when the maximum delay is higher than one polling cycle.

Assume that the costs for performing a location update and for polling a cell are U and V, respectively. Given a threshold distance d, we denote the average location update cost by $C_u(d)$. The average location update cost can be determined by multiplying U by the probability at which the terminal exceeds the threshold distance given below:

$$C_u(d) = p_{d,d}\, a_{d,d+1} U \quad (61)$$

The average terminal paging cost, $C_v(d, m)$, is a function of both the threshold distance, d, and the maximum paging delay, m. When the maximum paging delay is one polling cycle, the network has to poll all cells in the destination terminal's residing area at the same time. The average terminal paging cost as a function of the threshold distance d for a maximum delay of one polling cycle is then:

$$C_v(d, 1) = c g(d) V \quad (62)$$

where c is the call arrival probability and $g(d)$ is the number of cells that are within a distance of d from the center cell. The expression of $g(d)$ is given in equation (1). When the maximum paging delay is higher than one polling cycle, the residing area is partitioned into a number of subareas such that each subarea consists of one or more rings (as described in Section 2.) A mobile terminal is at ring r_i when its distance from its center cell is i. The probability that the mobile terminal is in ring r_i is, therefore, equal to the steady state probability of state i, $p_{i,d}$. The probability that a terminal is located in subarea A_j given a threshold distance of d is:

$$\beta_{j,d} = \sum_{r_i \in A_j} p_{i,d} \quad (63)$$

We denote the number of cells in subarea A_j by $N(A_j)$. Given that the terminal is residing in subarea A_j, the number of cells polled before the terminal is successfully located is:

$$w_j = \sum_{k=1}^{j} N(A_k) \quad (64)$$

Then, the *average paging cost* can be calculated as follows:

$$C_v(d, m) = cV \sum_{i=1}^{\ell} \beta_{i,d} w_i \quad (65)$$

The *average total cost for location update and terminal paging* is:

$$C_T(d, m) = C_u(d) + C_v(d, m) \quad (66)$$

6 Optimal Threshold Distance

The results obtained in Section 5 let us determine the average total cost $C_T(d,m)$, equation (66), for location update and terminal paging given a threshold distance d and maximum paging delay m. Analytical results demonstrate that, depending on the method used to partition the residing area of the terminal, the total cost curve may have local minimum. We, therefore, cannot use gradient decent methods to locate the optimal distance. One way of determining the optimal threshold distance is to limit the location update distance to a predefined maximum D. The optimal location update distance can then be found by evaluating the total cost for each of the allowed threshold distances $0 \le d \le D$. The threshold distance that results in the minimum total costs is selected as the optimal distance d^*. Limiting the threshold distance to a maximum value may not have significant effect on the total cost. For typical call arrival and mobility values, the optimal distance rarely exceeds 50. Besides, the size of a local coverage area is usually limited to the boundary of a city, it is realistic to require the mobile terminal to perform a location update before it leaves the local coverage area. This method can always locate the optimal distance in $D+1$ iterations.

Another method for determining the optimal threshold distance is an iterative algorithm called simulated annealing [2, 5]. In simulated annealing, potential solutions are generated and compared in every iteration. This potential solution is accepted or rejected probabilistically consistent with the Boltzmann distribution law involving a *temperature* T. During the course of the annealing, T is reduced according to a *cooling schedule*. When a suitably low temperature is finally reached, the algorithm terminates on an optimal or near optimal solution. The algorithmic structure of simulated annealing is given as follows:

$T \leftarrow 1;$
$d \leftarrow Random\_Init();$
$k \leftarrow 1;$
do
  $d' \leftarrow generate(d);$
  $\Delta d \leftarrow total\_cost(d) - total\_cost(d');$
  $d \leftarrow replace((d,d'),\Delta d);$
  $T \leftarrow \frac{y}{y+k};$
  $k \leftarrow k+1;$
while($T > exit\_T$)

To start the annealing process, a random threshold distance is generated by the *Random_Init*() routine. The *generate*(d) routine returns a threshold distance value close to d. Based on equation (66), *total_cost*(d) returns the average total cost for location update and terminal paging given the threshold distance d. The values of y and *exit_T* are adjusted based on the required accuracy of the result. Assume that *rand*() returns a random number in [0,1), the *replace*($(d,d'),\Delta d$) routine is given as follows:

replace($(d,d'),\Delta d$)
  if $\Delta d \ge 0$ return d';
  if ($rand() < exp((\Delta d/T))$ return d';
  return d;

Due to space limitation, we will not explain in detail the theory behind simulated annealing. Interested readers may refer to [2, 5]. We note that simulated annealing is not the only method available for finding the global minimum of a discrete function. There are other optimization methods reported in literature [4]. These methods can find the optimal threshold distance at varying degrees of speed and efficiency.

7 Analytical Results

We first present the numerical results for both the one- and two-dimensional model using some typical parameters. For the two-dimensional model, the exact steady state probability solutions (discussed in Section 4.1) are used. Figures 4(a) and 4(b) show the average total cost as the probability of moving is varied from 0.001 to 0.5. The call arrival probability, location update cost and paging cost are fixed at 0.01, 100 and 1, respectively. We consider three maximum paging delay bounds (1,2 and 3 polling cycles, respectively) as well as the case of unconstrained delay. In all cases, the average total cost increases as the probability of moving increases. The cost is the highest when the maximum paging delay is 1 polling cycle. As the maximum paging delay increases, the average total cost drops. The reduction in cost is significant even for a maximum paging delay of only 2 polling cycles. Figures 5(a) and 5(b) present similar results for the one- and two-dimensional models, respectively. However, the probability of moving is fixed at 0.05 and the call arrival probability is varied between 0.001 and 0.1. The average total cost increases as the call arrival probability increases. Discontinuities appear in some curves due to the sudden changes in the optimal threshold distances. As in the case of Figures 4(a) and 4(b), the average total cost decreases as the maximum paging delay increases. The decrease is more significant when the maximum paging delay increases from 1 to 2 polling cycles than from 2 to 3 polling cycles. This demonstrates that a large maximum paging delay is not necessary to obtain significant reduction in cost. In most cases, the average total costs are very close to the minimum value (when there is no paging delay bound) when a maximum paging delay of 3 polling cycles is used.

Table 1 shows the optimal threshold distances and the associated average total cost values for the one-

	delay = 1		delay = 2		delay = 3		unbounded	
U	d^*	C_T	d^*	C_T	d^*	C_T	d^*	C_T
1	0	0.125	0	0.125	0	0.125	0	0.125
2	0	0.150	0	0.150	0	0.150	0	0.150
3	0	0.175	0	0.175	0	0.175	0	0.175
4	0	0.200	0	0.200	0	0.200	0	0.200
5	0	0.225	0	0.225	0	0.225	0	0.225
6	0	0.250	0	0.250	0	0.250	0	0.250
7	0	0.275	1	0.270	1	0.270	1	0.270
8	0	0.300	1	0.282	1	0.282	1	0.282
9	0	0.325	1	0.293	2	0.291	2	0.291
10	0	0.350	1	0.305	2	0.296	2	0.296
20	1	0.527	1	0.418	2	0.339	3	0.338
30	2	0.630	2	0.465	2	0.382	3	0.357
40	2	0.673	3	0.486	3	0.415	4	0.371
50	2	0.716	3	0.506	3	0.435	4	0.381
60	2	0.760	3	0.526	3	0.454	5	0.386
70	2	0.803	3	0.545	3	0.474	6	0.391
80	2	0.846	3	0.565	3	0.494	6	0.394
90	3	0.878	4	0.579	5	0.510	7	0.396
100	3	0.897	4	0.589	5	0.515	7	0.397
200	3	1.095	4	0.686	6	0.548	12	0.401
300	4	1.193	6	0.724	7	0.565	17	0.402
400	4	1.290	6	0.750	7	0.579	22	0.402
500	5	1.351	6	0.776	7	0.593	27	0.402
600	5	1.401	6	0.803	7	0.607	32	0.402
700	5	1.451	6	0.829	7	0.621	37	0.402
800	5	1.501	6	0.855	7	0.635	42	0.402
900	6	1.537	8	0.868	7	0.649	47	0.402
1000	6	1.563	8	0.876	7	0.663	52	0.402

Table 1: Optimal Threshold Distance and Average Total Cost for One-Dimensional Mobility Model

dimensional model as the location update cost varies. The call arrival and movement probabilities are set to 0.01 and 0.05, respectively. The location update cost U is varied from 1 to 1000 while the paging cost V remains at 10. It is demonstrated that both the optimal threshold distance d^* and the total cost C_T increase as the location update cost increases. This result is intuitive. When the location update cost is low, there is an advantage of performing location update more frequently in order to avoid paying the relatively high paging cost when an incoming call arrives. When the location update cost is high, the paging cost is relatively low. There is a cost advantage if location update is performed less frequently. Table 2 shows similar results for the two-dimensional model under the same parameter values. Define the near optimal threshold distance, $d^{*'}$, to be the optimal threshold distance obtained using the approximated steady state probabilities obtained in Section 4.2. Also define the near optimal average total cost, C'_T, to be the total cost obtained when the near optimal threshold distance is used. It can be seen from Table 2 that the differences between d^* and $d^{*'}$ are within 1 from each other almost all the time. In most cases, the two values are the same. Table 2 also demonstrates that the values of C_T and C'_T are very close to each other when the optimal threshold distance is higher than 1. The worst cases occur when the optimal threshold distance is 1 and the near optimal threshold value is 0. Under this situation, the value of C'_T can be double that of C_T. This happens because a threshold distance of 0 generally results in relatively high total cost if the optimal threshold distance is not 0. This situation can be alleviated by a simple modification to the mechanism used to locate the near optimal threshold distance. Assume C_T^0 and C_T^1 to be the exact average total cost when a threshold distance of 0 and 1 are used, respectively. When the near optimal threshold distance $d^{*'}$ is 0, we replace it by 1 if $C_T^1 < C_T^0$. Otherwise it stays unchanged. This method guarantees that a threshold distance of 0 will not be selected if a threshold distance of 1 can result in lower cost. Obtaining the C_T^0 and C_T^1 is straightforward and the additional computation involved is minimal. Since the computation of $d^{*'}$ is less involved as compared to d^* because of the availability of the closed form solution. The near optimal threshold distance is therefore useful in dynamic location update and paging schemes where computation allowed is very limited. The saving in computation cost may outweight the slight increase in total location update and terminal paging costs.

The numerical results obtained in this section demonstrate that:

- Significant reduction in average total cost $C_T(d, m)$ can be obtained by a small increase of the maximum paging delay m from its minimum value of one polling cycle.

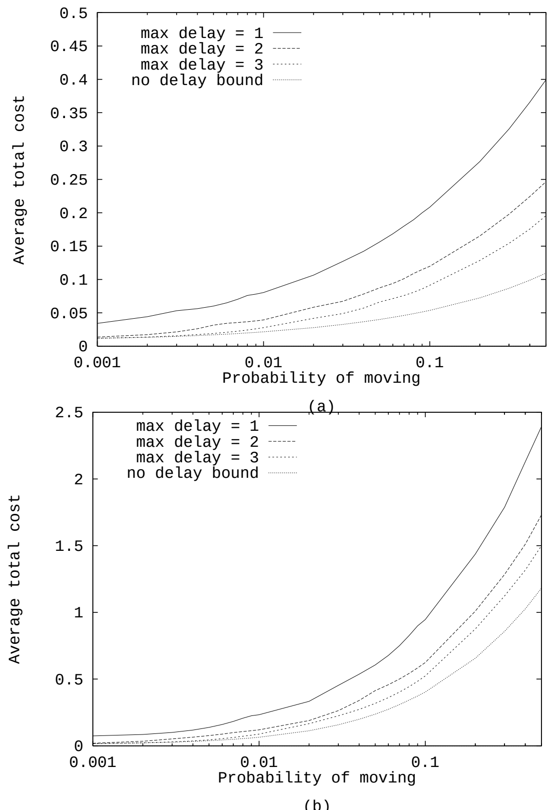

Figure 4: Average Total Cost versus Probability of Moving for (a) One-Dimensional Mobility Model and (b) Two-Dimensional Mobility Model.

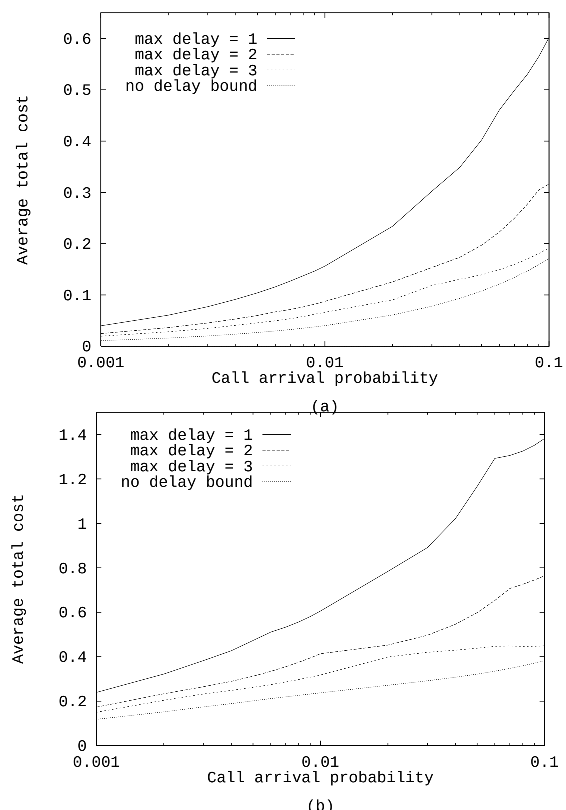

Figure 5: Average Total Cost versus Call Arrival Probability for (a) One-Dimensional Mobility Model and (b) Two-Dimensional Mobility Model.

- The optimal threshold distance d^* varies as the maximum delay m is changed. This means that the optimal threshold distance selected by a scheme that assumes unconstrained paging delay may not be appropriate when the paging delay is limited. Our scheme allows the determination of the optimal threshold distance at different values of the maximum paging delay m.

- The near optimal threshold distances obtained using the approximate transition probability equations (43) and (44) are shown to be accurate. This means that when computation cost is critical, the near optimal threshold distance can be used without causing significant increase in the total average cost $C_T(d, m)$.

8 Conclusions

In this paper we introduced a location management scheme that combines a distance based location update mechanism with a paging scheme subject to delay constraints. The mobility of each terminal is modeled by a Markov chain model and the probability distribution of terminal location is derived. Based on this, we obtain the average total location update and terminal paging cost under given threshold distance and maximum delay constraint. Given this average total cost function, we determine the optimal threshold distance by using an iterative algorithm. Results demonstrated that the optimal cost decreases as the maximum delay increases. However, a small increase of the maximum delay from 1 to 2 polling cycles can lower the optimal cost to half way between its values when the maximum delays are 1 and ∞ (no delay bound) polling cycles, respectively.

Most previous schemes assume that paging delay is either unconstrained (such as [1, 3]) or confined to one polling cycle (such as [8]). In the former case, it may take arbitrarily long to locate a mobile terminal. While in the latter case, the network cannot take advantage of situations when the PCN can tolerate delay of higher than one polling cycle. Our scheme is more realistic as the maximum paging delay can be selected based on the

	delay = 1				delay = 3				unbounded			
U	d^*	$d*'$	C_T	C_T'	d^*	$d*'$	C_T	C_T'	d^*	$d*'$	C_T	C_T'
1	0	0	0.150	0.150	0	0	0.150	0.150	0	0	0.150	0.150
2	0	0	0.200	0.200	0	0	0.200	0.200	0	0	0.200	0.200
3	0	0	0.250	0.250	0	0	0.250	0.250	0	0	0.250	0.250
4	0	0	0.300	0.300	0	0	0.300	0.300	0	0	0.300	0.300
5	0	0	0.350	0.350	0	0	0.350	0.350	0	0	0.350	0.350
6	0	0	0.400	0.400	0	0	0.400	0.400	0	0	0.400	0.400
7	0	0	0.450	0.450	0	0	0.450	0.450	0	0	0.450	0.450
8	0	0	0.500	0.500	0	0	0.500	0.500	0	0	0.500	0.500
9	0	0	0.550	0.550	1	0	0.542	0.550	1	0	0.542	0.550
10	0	0	0.600	0.600	1	0	0.555	0.600	1	0	0.555	0.600
20	1	0	0.968	1.100	1	0	0.689	1.100	1	0	0.689	1.100
30	1	0	1.102	1.600	1	0	0.823	1.600	1	0	0.823	1.600
40	1	0	1.236	2.100	1	0	0.957	2.100	1	0	0.957	2.100
50	1	0	1.370	2.600	2	2	1.074	1.074	2	2	1.074	1.074
60	1	0	1.504	3.100	2	2	1.126	1.126	2	2	1.126	1.126
70	1	0	1.638	3.600	2	2	1.178	1.178	2	2	1.178	1.178
80	1	1	1.771	1.771	2	2	1.231	1.231	2	2	1.231	1.231
90	1	1	1.905	1.905	2	2	1.283	1.283	2	2	1.283	1.283
100	1	1	2.039	2.039	2	2	1.335	1.335	2	2	1.335	1.335
200	2	1	2.945	3.379	2	2	1.858	1.858	3	3	1.683	1.683
300	2	2	3.468	3.468	3	2	2.372	2.381	4	3	1.912	1.918
400	2	2	3.991	3.991	3	3	2.608	2.608	4	4	2.025	2.025
500	2	2	4.514	4.514	3	3	2.843	2.843	4	4	2.138	2.138
600	2	2	5.036	5.036	5	3	2.955	3.079	5	5	2.204	2.204
700	3	2	5.349	5.559	5	5	3.011	3.011	5	5	2.260	2.260
800	3	2	5.585	6.082	5	5	3.066	3.066	5	5	2.315	2.315
900	3	2	5.820	6.604	5	5	3.122	3.122	6	6	2.346	2.346
1000	3	2	6.056	7.127	5	5	3.177	3.177	6	6	2.374	2.374

Table 2: Optimal Threshold Distance and Average Total Cost for Two-Dimensional Mobility Model

particular system requirement.

Results obtained in this paper can be applied in static location update schemes such that the network determines the location update threshold distance according to the average call arrival and movement probabilities of all the users. This result can also be used in dynamic schemes such that location update threshold distance is determined continuously on a per-user basis.

Future research includes the simplification of the threshold distance optimization process such that our mechanism can be implemented in mobile terminals with limited power supply. Also, an optimal method for partitioning the residing area of the terminal should be developed. In any case, our method for obtaining the optimal location update threshold distance is not limited to the partitioning scheme described in this paper. The total cost and hence the optimal threshold distance can be obtained by our method when other partitioning methods are used.

Acknowledgement

We would like to thank Zygmunt J. Haas, Jason Y.B. Lin and Kazem Sohraby for their constructive comments which lead to significant improvements of this paper.

References

[1] I.F. Akyildiz and J.S.M. Ho, "Dynamic Mobile User Location Update for Wireless PCS Networks," *ACM-Baltzer Journal of Wireless Networks*, April 1995.

[2] I.F. Akyildiz and R. Shonkwiler, "Simulated Annealing for Throughput Optimization in Communication Networks with Window Flow Control," *Proc. IEEE ICC 90', pp.1202-1209*, April 1990.

[3] A. Bar-Noy, I. Kessler and M. Sidi, "Mobile Users: To Update or not to Update?" *ACM-Baltzer Journal of Wireless Networks*, April 1995.

[4] R.D.Brent, *Algorithms for Minimization without Derivatives*, Prentice-Hall, 1973.

[5] S. Kirkpatrick, C.D. Gelatti and M.P. Vecchi, "Optimization by Simulating Annealing," *Science Journal*, 220, pp. 671-680, 1983.

[6] U. Madhow, M.L. Honig and K. Steiglitz, "Optimization of Wireless Resources for Personal Communications Mobility Tracking," *Proc. IEEE INFOCOM '94, pp. 577-584*, June 1994.

[7] C. Rose and R. Yates, "Paging Cost Minimization Under Delay Constraints," *ACM-Baltzer Journal of Wireless Networks*, April 1995.

[8] H. Xie, S. Tabbane and D. Goodman, "Dynamic Location Area Management and Performance Analysis," *Proc. IEEE VTC '93, pp. 536-539*, May 1993.