# Last Interaction Based Paging in Mobile Cellular Networks

I.Z. Koukoutsidis and M.E. Theologou

National Technical University of Athens,
School of Electrical and Computer Engineering,
9 Heroon Polytechneiou st., Zographou 157 73, Athens, Greece
`koukou@telecom.ntua.gr`

**Abstract.** In a mobile network, the paging operation is used to indicate the subscriber's position in order to establish a connection with another user, calling from a fixed or mobile environment. The currently employed scheme, blanket paging, incurs a significant cost in radio bandwidth utilization and is cost-effective only in small networks. In this paper, we introduce a new strategy that combines an optimal partitioning of the location area with a model to predict user movements, based on the additional information of the cell where the last interaction between the network and the terminal took place. We study the performance of the strategy under delay bounds and different mobility and call arrival characteristics. It is found that the new scheme further minimizes signaling costs and enhances previous sequential paging algorithms.

## 1 Introduction

Tracking the location of a mobile terminal (MT) is one of the fundamental operations in a wireless system. Current mobile networks implement this operation by dividing the system coverage area into geographical location areas (LAs). Each LA consists of a number of cells whose base stations are connected to the same service switching center (Mobile Switching Center, MSC). The MSC is itself connected to the telephone network (PSTN). The mobile network is always aware of the LA the user currently resides in. Upon an LA crossing, it is the responsibility of an MT to initiate a location update (LU) or registration procedure to inform the network of the change and update the appropriate databases [1]. When an incoming call or messaging service arrives for a specific subscriber, the network must determine the exact cell where he currently resides in order to deliver the service. For this, a search is conducted in all the cells of the registered location area, during which the network sends polling messages and awaits for a response from the right cell. The process is called *paging* (PG) and is shown graphically in Fig. 1.

This scheme has been eloquently named *blanket paging* because it covers every possible cell where the terminal might be located. The location update and paging procedures both define a *location management* (LM) scheme. The LM scheme described above is simple and easy to implement, however it is not cost-effective, especially under heavy traffic conditions and high user densities. A significant amount of bandwidth and processing power (both at the air-interface and the fixed

network) is consumed for the location updates, and unnecessary signaling load is generated for paging all cells in a location area.
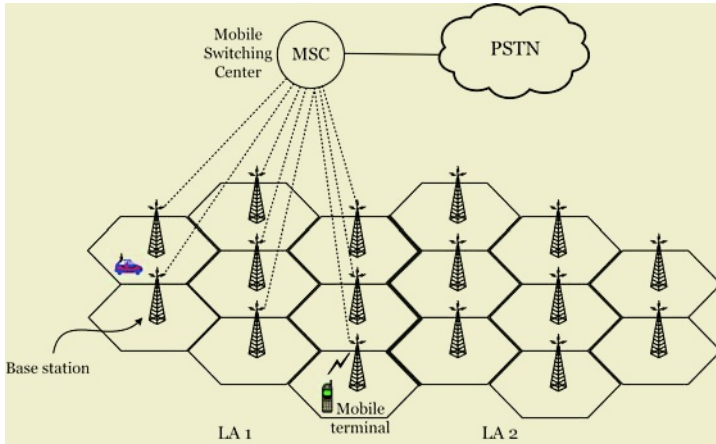


**Fig. 1.** Paging procedure in a mobile cellular network (the figure shown here is simplified, as other components involved with the paging procedure are omitted, such as the SS7 network and HLR/VLR databases. Additionally, the BSC/RNC is omitted, since its primary function is to maintain the radio resources of its base stations and is not directly involved in paging)

In general, the objective to minimize the total signaling cost of the location update and paging procedures constitutes a complex optimization problem: "Given the system area size and shape, the mobility behavior of a specific terminal and the incoming call characteristics, determine an optimal LA design and corresponding LU and PG mechanisms such that the total signaling cost incurred by location management procedures is minimized, and the delay for finding a mobile user is kept under a reasonable constraint". In [2], it was shown that the problem in general is *NP-complete*, as it is reduced to a *min-cut clustering* problem. In this way, an exact solution is amenable only for small networks. A heuristic algorithm applied to the problem can be found in [3], where a sub-optimal LA design is acquired by initially merging cells to groups by means of a greedy method and then exchanging cells between the groups until no further reduction in the signaling cost can be attained.

Assuming that a favorable partition of the coverage area into location areas is acquired, we can further reduce total costs by employing a different paging strategy. In a foundation work in [4], Rose and Yates formally showed the intuitive idea that polling groups of cells sequentially in order of decreasing location probabilities minimizes paging cost. The penalty paid for reducing the number of paging messages is the increase in the average delay for finding a mobile user, which is proportional to the number of groups the location area is divided into. The strategy described above is also referred to as *selective paging* in [5]. A similar algorithm presented in [6] is called *multi-step paging*, where each group of cells is called a paging area (PA). Optimal sequential paging is extensively studied in [7], where dynamic programming algorithms for partitioning network areas are derived.

Our work is motivated by the ideas in [4,7]. The aim is also to derive optimal partitions to improve the paging mechanism, but this time using conditional location

probabilities from the time of the last interaction with the network to the next incoming call arrival. A *last interaction* (LI) is defined as any transaction that can provide incidental information regarding the cell position of a mobile user, such as incoming or outgoing call setup, location update or registration at switch-on, execution of messaging services, etc. It is shown that by employing the extra information of the last interaction cell, short-term positions of the user can be derived with the use of a location prediction model. Thus we can largely diminish the uncertainty about the user's position and a user is paged with greater accuracy. It is demonstrated that for the same number of partitions, a large signaling saving can occur compared to the case where normal unconditional probabilities are used. For conciseness, our scheme will be referred to as LIBP (Last Interaction Based Paging) in the paper.

The remainder of the text is organized as follows. In Section 2, we describe the paging strategy in three parts: first we introduce some necessary notation; the essentials of the strategy are described next; the last part reports algorithms for deriving optimal partitions. A location prediction model is derived mathematically and the system is studied in Section 3. In Sections 4 and 5 we apply the analysis to a simple paging example and evaluate the scheme's performance. We end in Section 6 with a brief review of the main points of our strategy and the major conclusion.

## 2 Paging Strategy

### 2.1 Preliminary Definitions and Notation

Let us denote by $S = \{1, 2, \ldots, n\}$ the set of $n$ cells in a location area, such that a mobile is guaranteed to be in one of those cells at the time of a call arrival. We suppose that a model for conditional user location can be acquired either mathematically or by gathering statistics. The model describes a time-varying conditional distribution $p(i \mid u, t)$, which gives the probability of a user being located in cell $i$ at time $t$, provided he was in cell $u$ at $t = 0$. The probability takes distinct values for every $i \in S$, such that $\sum_{i \in S} p(i \mid u, t) = 1$. By ordering these conditional probabilities in decreasing values, we can apply an algorithm to determine the optimum number of paging areas $N$ and their associated sizes, $n_i$ $(i = 1 \ldots N)$, to minimize the number of messages sent subject to a delay constraint. For $N > 1$, the size of $PA_i$ is $0 < n_i < n$, and it holds that

$$\forall \, i \neq j \;\; \rightarrow \; PA_i \cap PA_j = \varnothing$$

$$\text{and } \sum_{i=1}^{N} n_i = n$$

,i.e. the partitions are mutually exclusive and collectively exhaustive. The conditional location probability of a user in paging area $PA_i$ is then defined as the sum of individual location probabilities, as

$$q_i(u, t) = \sum_{j \in PA_i} p(j \mid u, t). \tag{1}$$

## 2.2   Algorithm Description

After a last interaction event where the position of an MT in the location area is updated in the network database, a timer is set on the network side. When an incoming call request arrives for a specific user, the conditional location probabilities $p(i \mid u)$ are generated by the system for the time instance of the arrival, $t_a$. Based on these probabilities, cells are grouped into paging areas according to an optimal partitioning scheme.

Assuming that $PA_1, PA_2, \ldots, PA_N$ is an optimal *N*-partition of the set $S$, then the paging algorithm works by first sending polling messages to all cells in $PA_1$. If there is a response from one of the polled cells, the first step is successful and the algorithm stops. Otherwise, a second step is conducted that queries cells in the second partition $PA_2$. In case of another failure, the third partition is queried, then the third, and so on. Each step in the algorithm corresponds to a polling cycle, from the time the paging message is sent until a response is received. If the mobile is not in one of the cells in the paging area, then the polling cycle corresponds to the duration of a timeout interval. Assuming the cost for polling one cell is 1 (e.g. measured in bits/message), the average paging cost based on this algorithm is

$$C = \sum_{i=1}^{N} \sum_{j=1}^{i} n_j \cdot q_i(u, t_a),$$  (2)

i.e. it is the average number of paged cells. The average delay, in number of polling cycles is

$$d = \sum_{i=1}^{N} i \cdot q_i(u, t_a).$$  (3)

The main effort in implementing this scheme is required at the network to calculate the conditional location probabilities at the time of an incoming call arrival –based on the last interaction cell– and derive the optimal partitions. This is not inhibitive considering that modern networks have increased computational capabilities. In addition, a small extra cost occurs from setting the timer anew each time a transaction between the user and the network occurs, and storing and retrieving the identity of the last interaction cell when necessary. This information is actually the base station id, and is usually transmitted but not exploited in message transactions between the MSC and cell base stations.

## 2.3   Optimal Partitions

The objective of a partitioning algorithm is to choose the optimum $N$ and $n_i$ to minimize the number of paging messages subject to a delay constraint. In [7], the authors have shown that an optimal partition (OP) of a set $S$ must be *non-increasingly ordered*. This means that for $i, j \in S$ with $i \in PA_k$, $j \in PA_l$ such that $k \leq l$, it must hold that $p(i \mid u, t_a) \geq p(j \mid u, t_a)$. From now on, we consider all cells in the location area to be numbered in order of decreasing conditional location

probabilities. Hence, an optimal *N*-partition will consist of contiguously numbered cells, that is

$$PA_1 = \{1, 2, \ldots, n_1\}, \ PA_2 = \{n_1 + 1, \ldots, n_1 + n_2\}, \ldots, PA_N = \{n_{N-1} + 1, \ldots, n_{N-1} + n_N\}.$$

In this work we only consider OPs subject to a *worst case delay* constraint. For worst case delay, the total number of partitions equals the delay upper bound $D$, i.e. $N = D$. A dynamic programming algorithm can be constructed according to [7] which provides the partition into $N$ zones that minimizes the paging cost. Let $g[n', N']$ be the minimum average paging cost achievable by partitioning the first $n'$ cells into $N'$ paging areas. Then we have the following recursion:

$$g[n', N'] = \min_{j=N'-1}^{n'} \left( g[j, N'-1] + n' \sum_{i=j+1}^{n'} p(i \mid u, t_a) \right). \tag{4}$$

The recursive formula exhibits the fact that an optimal partition of size $N'$ that consists of $n'$ numbered cells can be found by the minimum of an expression that allocates $j \leq n'$ cells to an optimal partition of size $N'-1$ and the overhead $n' - j$ cells to the last partition $N'$. The initial conditions are

$$g[j, 1] = j \sum_{i=1}^{j} p(i \mid u, t_a). \tag{5}$$

With the help of these conditions we can apply the recursive equation to gradually build the two-dimensional matrix $g_{n \times N}$. Then by looking up the entries in the table, partition sizes can be found for increasing number of partitions. The time complexity of this algorithm is $O(Nn^2)$, which is the time to calculate all entries in the table.

Alternatively to the dynamic programming approach, we can employ a heuristic algorithm introduced in [8] to find the optimal allocation of cells to paging areas. If $p_j^s$ and $p_j^l$ denote the smallest and largest probability cell, respectively, in a partition $j$ for conditional probabilities calculated at time instant $t_a$, then the algorithm starts with an initial partition and proceeds to move partition boundaries left or right, depending on the following conditions:

- *Backward Boundary Condition*: if the total conditional location probability of a partition $j$, $q_j(u, t_a) > p_j^s(u, t_a) \cdot (n_{j+1} + 1)$, then the cell with probability $p_j^s$ will be moved to the next PA, i.e. the partition boundary is moved left.
- *Forward Boundary Condition*: if the total conditional location probability of a partition $j$, $q_j(u, t_a) < p_{j+1}^l(u, t_a) \cdot (n_{j+1} - 1)$, then the cell with probability $p_{j+1}^l$ will be moved to the previous PA, i.e. the partition boundary is moved right.

The algorithm checks all partitions in the area until forward and backward boundary conditions have been satisfied. This is the strategy we employ to derive optimal partitions in the paper. Practically this algorithm is very effective; for small scale problems and depending on the form of the location probabilities, one can cleverly choose the initial partition so that relatively few boundary movements are necessary to achieve the optimal configuration.

Partitioning under average delay constraints is not considered here, as this is very difficult for $N > 2$. However, a similar polynomial time dynamic programming algorithm can be used to derive optimal paging areas in the average delay case [7].

For the case of $N = 2$ it is relatively easy, and a corresponding algorithm can be found in [9].

## 3   Mathematical Model

In order to model the proposed scheme mathematically, we assume that the incoming call interarrival time is an exponentially distributed random variable $t_a$ with rate $\lambda_a$. Then the probability density function (pdf) is given by

$$f_{t_a}(t) = \lambda_a \cdot e^{-\lambda_a t} . \tag{6}$$

The residence time of a mobile user in a cell $i$ is also assumed to be exponentially distributed with mean $1/\lambda_{m_i}$ and density $f_{m_i} = \lambda_{m_i} \cdot e^{-\lambda_{m_i} t}$. We can safely postulate that in most cases, a user movement direction depends only on the cell he is currently located. That, together with the cell dwell times being exponentially distributed, permits us to model the system as a continuous time *Markov chain*, as shown in Fig. 2.
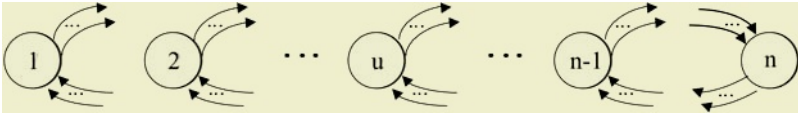


**Fig. 2.** Markov chain diagram for modeling MT movement in a location area

Each state of the diagram corresponds to a cell in the LA. The transition rates depend on the movement rates $\lambda_{m_i}$ and the transition probabilities $p_{ij}$, $\forall\, i, j \in$ S. The latter provide the routing probability that an MT currently in cell $i$ will move towards an adjacent cell $j$. Accordingly, rates $q_{ij}$ are calculated as

$$q_{ij} = \lambda_{m_i} p_{ij},$$
$$q_{ii} = -\sum_{j \neq i} q_{ij} , \tag{7}$$

for $i, j = 1, 2, \ldots, n$. The $p_{ij}$ probabilities are influenced by the topological structure of the LA and the mobility pattern of the terminal. Here we follow a *symmetric random walk* model, so that a mobile moves with equal probabilities to each of its neighboring cells. Moreover, for convenience the mean residence times in every cell will be taken equal, $\lambda_{m_i} = \lambda_m$ $\forall\, i$, although different values can be applied provided the required data is in hand. It is worth noting that the estimation of the mean residence time is rather difficult since its value depends on the individual behavior and mobility pattern of a mobile user, as well as the size and shape of each cell. For implementing the LIBP scheme, estimates may be acquired either by a user-supplied profile, or by gathering statistics regarding the crossings of the MT from certain points in the network. These would be logged periodically either via the existing structure or an autonomous sub-network, and communicated to the MSC/VLR. Secondarily, we note

that a model usually classifies objects into classes. Since it would be too arduous to estimate mobility parameters for each and every subscriber of a network provider, we may say that the mean residence time and call rate are representative of a given class of subscribers.

Using expressions in (7), we create the transition rate matrix (generator matrix), $\mathbf{Q} = [q_{ij}]$, $i, j \in S$. Then by solving the system $\pi \cdot \mathbf{Q} = 0$ with $\sum_{i \in S} \pi_i = 1$, we obtain the steady-state probabilities for the continuous time case. A probability $\pi_i$ denotes the proportion of time the mobile spends in state $i$, *provided that* it resides in the specified LA. Then equivalently, using the PASTA theorem [10], $\pi_i$ is the probability that an incoming call will find a user in cell $i$. Therefore, steady-state probabilities can be used to derive optimal partitions for a general unconditioned case, where no knowledge of the last interaction cell is employed.

If the additional information of the last interaction cell is available, we can calculate the likelihood of reaching state (cell) $i$ after a period $t$, based on the transition matrix $\mathbf{Q}$. If $X_t$ is the state of the Markov chain at time $t$, by applying the Chapman-Kolmogorov equations it is possible to derive the following property [11]:

$$p(i \mid u, t) = \Pr\{X_t = i \mid X_0 = u\} = e^{t\mathbf{Q}}(u, i) \qquad (8)$$

The analysis from now on considers only incoming calls and LA crossings as last interaction events. However, the model can be generalized to include more transaction types if we embed the frequency of interactions to the call arrival rate. We can imagine that raising the call arrival rate has nearly the same effect as the addition of more interactions in the model.

The timing diagram describing a possible sequence of events as the mobile moves in the network area is shown in Fig. 3.
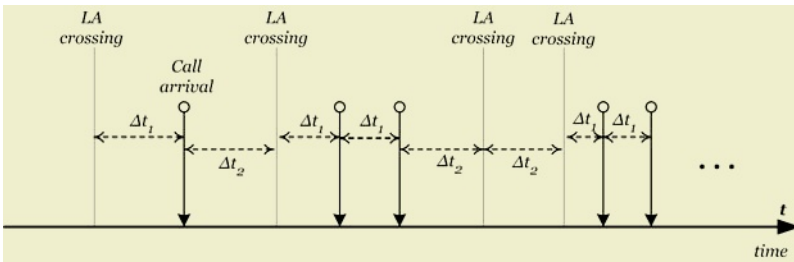


**Fig. 3.** Timing diagram of modeled user-network interactions during MT movement

Each one of the events in the figure is a random renewal point and corresponds to an update of the user position in the network database. The periods $\Delta t_1, \Delta t_2$ define the roaming intervals from the most recent interaction to the next event —a call arrival or normal registration event respectively. Any incoming call request for a specific subscriber arrives at the end of a *paging interval*. This is defined here as the period between the last interaction and the next call arrival. Since we always assume that a call will arrive *while* the MT is in the LA, this is equal to $\Delta t_1$ *if and only if* $\Delta t_1 < \Delta t_2$. Due to the memoryless property of the exponential, the distribution of $\Delta t_1$ is the same

as the interarrival distribution, even if the last interaction was an LA crossing. Consequently, the distribution of a paging interval is

$$f_p(t) = f_{\Delta t_1}(t \mid 0 \le t < \Delta t_2) = \frac{\lambda_a \cdot e^{-\lambda_a t}}{\int_{t_2=0}^{\infty} \int_{t_1=0}^{t_2} \lambda_a \cdot e^{-\lambda_a t_1} \cdot f_{\Delta t_2}(t_2) dt_1 dt_2} . \tag{9}$$

From eqs. (8), (9) using the continuous version of the total probability theorem [12], the probability of the user residing in cell $i$ when the next call arrives is

$$p(i \mid u) = \int_0^{\infty} p(i \mid u, t) f_p(t) dt = \int_0^{\infty} e^{tQ}(u, i) \frac{\lambda_a \cdot e^{-\lambda_a t}}{\int_{t_2=0}^{\infty} \int_{t_1=0}^{t_2} \lambda_a \cdot e^{-\lambda_a t_1} \cdot f_{\Delta t_2}(t_2) dt_1 dt_2} dt . \tag{10}$$

The distribution of $\Delta t_2$, $f_{\Delta t_2}$ is the distribution of the residence time in a location area, when the user initiates his movement from the cell of the most recent interaction. Suppose the MT resides in the $n$-cell LA for a period $T$, during which it visits $m$ cells, and the time spent at each cell is $t_i$, $i \in [1, m]$. Then the total sojourn time $T = t_1 + t_2 + \ldots + t_m$ has density function

$$f_{LA}^{(m)}(T) = \int_{t_1=0}^{T} \int_{t_2=0}^{T-t_1} \cdots \int_{t_m=0}^{T-t_1-\ldots-t_{m-1}} f_m(t_1) f_m(t_2) \ldots f_m(t_m) dt_m \ldots dt_2 dt_1 , \tag{11}$$

i.e. is the convolution of $f_m(t_1), \ldots, f_m(t_m)$.

In order to calculate the number of cells visited until exiting the LA, we can use an embedded version of the Markov chain in Fig. 2, with stepwise transition probabilities $p_{ij}$ and an absorbing state $E$ to mark transitions to cells in the external boundary of the area. If $\mathbf{P} = (p_{ij})$ is the $(n+1) \times (n+1)$ transition matrix, then its power is defined recursively as

$$\mathbf{P}^{[k]} = \begin{cases} \mathbf{P}, & \text{if } k = 1 \\ \mathbf{P} \times \mathbf{P}^{[k-1]}, & \text{if } k > 1 \end{cases}, \tag{12}$$

where an element $p_{i,j}^{[k]}$ of $\mathbf{P}^{[k]}$ is the probability that an MT initially in cell $i$ moves into cell $j$ after $k$ steps. Likewise, $p_{u,E}^m$ is the probability that a mobile exits the LA after $m$ steps. It follows that the probability that an MT exits the LA in *exactly* $m$ steps after the terminal was last seen in cell $u$ is given by

$$p_{m,u,E} = \begin{cases} p_{u,E} & , \text{if } m = 1 \\ p_{u,E}^{(m)} - p_{u,E}^{(m-1)}, & \text{if } m > 1 \end{cases} \tag{13}$$

With the help of (11), the theorem of total probability gives us the density function of the residence time in the LA, conditioned on the last interaction:

$$f_{\Delta t_2}(t) = \sum_{m=0}^{\infty} p_{m,u,E} \cdot f_{LA}^{(m)}(t) . \tag{14}$$

To avoid the difficult algebraic computation of $f_{\Delta t_2}(t)$, we may approximate the distribution of a paging interval with an unconditional pdf, so that (10) becomes

$$p(i \mid u) = \int_0^{\infty} e^{tQ}(u, i) \lambda_a e^{-\lambda_a t} dt . \tag{15}$$

This approximation is valid when the last interaction cell is an inner cell in the LA and/or the *call-to-mobility ratio* (CMR) $\lambda_a/\lambda_m$ takes relatively high values. In this case, the probability of an incoming call arriving before the user exits the LA approximates unity. On the other hand, when the last interaction cell is near the LA border or the CMR is low, the approximation is not very accurate. Then $\Pr\{0 \le t < \Delta t_2\} < 1$ and higher values near the LI cell are expected. As it is shown in the next section, because of this fact paging costs calculated using the simplified formula (15) provide a lower bound on the signaling reduction achieved with the proposed scheme.

It is worth noting that since we restrict the model to a location area, the total number of states in the Markov chain is not too large and the algorithm maintains a relatively low complexity, both in space and time. Fast numerical methods may be used to facilitate the computation of the matrix exponential in (8) for application to a system with limited processing capabilities. A reference is cited in [13] for further details. We also remark that a similar analysis like the one presented above can be made when the cell residence times have a general distribution; for interested readers, details are included in [14].

Besides the framework presented here, any other model that derives a predictive probability of user location can be used to examine the algorithm. For instance, a natural extension of the presented work could be an isotropic diffusion model [15], where the last interaction considers the exact position in the *xy*-plane and drift velocities are $v_x = v_y = 0$. Further, a Gauss-Markov model can be used as in [16]. Such a model describes a mobile's movement pattern more accurately, as it better captures the essence of the correlated behavior in time.


# 4   Application

We apply the analysis above to a simple paging example. The location area consists of 31 cells in a hexagonal configuration. Cells are numbered from 0 to 30, as shown in the simplified representation in Fig. 4. Our aim is to derive optimal partitions of the LA based on the call arrival and movement characteristics. The symmetric random walk model discussed in the previous section is followed and equation (15) is used to derive the conditional location probabilities. Results are derived for several values of the call-to-mobility ratio. Tables 1-3 depict the allocations of cells to optimal partitions, for N=2, 3 and 4, in the general OP case and when LI cell=15 in the LIBP scheme, for different CMR values. A higher number of partitions is not considered; as it was demonstrated in [5] for an LA of similar size, the advantages gained from paging cost reduction for higher values of $N$ become less significant compared to the increase in the average delay. The delay in locating an MT and thus establishing a call connection is a quality of service parameter for a subscriber, therefore we want to sustain it beneath acceptable levels. Moreover, it will be demonstrated that for common CMR values the behavior of the LIBP scheme does not change by adding more partitions.
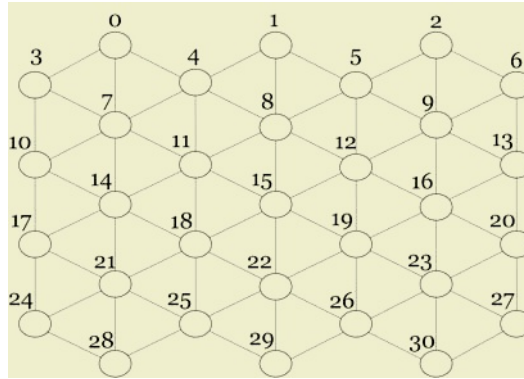
**Fig. 4.** Simplified view of the 31-cell hexagonal location area in our example

**Table 1.** Optimal partitions of the location area in the general OP case

| Partitions | Cell Selection |
|---|---|
| N=2 | {(21,18,14,22,7,11,15,8,12,19,9,16,25,4,23); (5,17,10,26,29,13,20,28,24,3,0,1,2,6,30,27)} |
| N=3 | {(21,18,14,22,7,11,15,8,12); (19,9,16,25,4,23,5,17,10,26); (29,13,20,28,24,3,0,1,2,6,30,27)} |
| N=4 | {(21,18,14,22,7,11,15); (8,12,19,9,16,25,4); (23,5,17,10,26,29,13,20); (28,24,3,0,1,2,6,30,27)} |

**Table 2.** Optimal partitions of the location area in the LIBP scheme (LI cell=15, CMR=$10^{-1}$)

| Partitions | Cell Selection |
|---|---|
| N=2 | {(15,22,8,18,11,12,19,14,25,16,21,7); (4,5,9,23,26,29,1,17,10,13,20,28,0,2,24,3,6,30,27)} |
| N=3 | {(15,22,8,18,11,12,19); (14,25,16,21,7,4,5,9,23,26,29); (1,17,10,13,20,28,0,2,24,3,6,30,27)} |
| N=4 | {(15,22,8,18,11,12); (19,14,25,16,21,7,4); (5,9,23,26,29,1,17,10); (13,20,28,0,2,24,3,6,30,27)} |

**Table 3.** Optimal partitions of the location area in the LIBP scheme (LI cell=15, CMR=1)

| Partitions | Cell Selection |
|---|---|
| N=2 | {(15,22,8,19,18,12,11); (25,5,4,16,14,26,29,21,9,7,23,1,20,13,17,10,28,2,0,30,24,6,3,27)} |
| N=3 | {(15); (22,8,19,18,12,11); (25,5,4,16,14,26,29,21,9,7,23,1,20,13,17,10,28,2,0,30,24,6,3,27)} |
| N=4 | {(15); (22,8,19,18,12,11); (25,5,4,16,14,26,29,21,9,7); (23,1,20,13,17,10,28,2,0,30,24,6,3,27)} |

By observing the tables, we remark that for the unconditioned case paging areas have almost equal sizes. This happens because the steady-state probabilities do not differ very much from one cell to another. As it has been proven in [7], "if each cell has equal probability of user location then the *N*-partition of an *n*-cell LA which minimizes the average cost of paging is balanced such that the difference in the

number of cells between any two partitions is no more than one", specifically $\forall\, i \in \{1, 2, \ldots, N\}$, $\left\lfloor \frac{n}{N} \right\rfloor \leq n_i \leq \left\lceil \frac{n}{N} \right\rceil$.

When the information of the last interaction cell is used, the derived conditional probabilities have higher values around this cell and partition sizes become smaller for the first partitions. This can be interpreted as an increased "concentration" of location probabilities at and around the LI cell. The degree of concentration depends on the duration of the paging interval. In our model, if the CMR is small, the paging interval increases and location probabilities are less concentrated around the LI cell. Theoretically, if the paging interval is infinite, the conditional probabilities become equal to the probabilities in the steady state. On the other hand, when the CMR takes higher values, the $p(i\,|\,u)$ probabilities are much more concentrated in a small portion of the LA. For a high number of partitions and a high CMR, the first partition may consist of only one cell (usually, the LI cell) as is shown in Table 3. The concentration of user location probabilities with increasing values of CMR compared to the steady-state probabilities is shown in the stack diagram of Fig. 5.
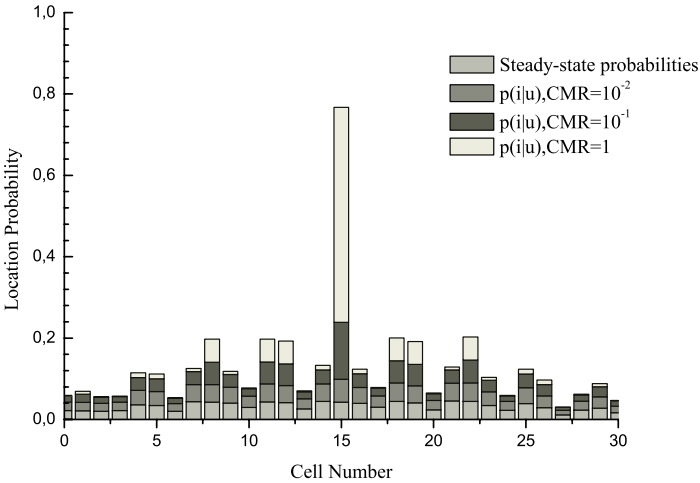


**Fig. 5.** Concentration of user location probabilities with increasing values of CMR in the LIBP scheme, when LI cell=15, compared to the steady-state probabilities

The more concentrated the location probabilities are, the higher is the signaling load reduction by the application of an optimal partitioning scheme. This is an essential conclusion first drawn explicitly in [7], and constitutes the rational basis of our work. It also explains why the simplified formula (15) in our LA model provides a lower bound on the signaling savings achieved with the proposed method. If the condition that a call arrives while an MT is in the LA were taken into consideration, the individual probabilities would be higher around a smaller number of cells. Therefore the savings using the approximate formula are always less than the exact savings.

# 5  Performance Evaluation

To evaluate the performance of the proposed method, we calculate the cost in number of paging messages to reach a user and the resulting average paging delay, for the LA of Fig. 4. Graphic results are shown in Figs. 6,7. We have considered two cases of last interaction cells: the first when LI cell=15, the center cell in the area, and second for LI cell=1, a cell near the border of the LA.

The results in terms of paging cost verify the following:

- In all cases, as the call-to-mobility ratio increases, the relative mobility of the terminal decreases or, equivalently, the paging interval diminishes. Thus the location uncertainty is reduced and the signaling savings increase. In Fig. 6(a), for $N$=2 partitions the percentage saving is 9.68% for CMR=0.1, 38.16% for CMR=0.5, 51.56% for CMR=1 and 63.72% for CMR=3. For higher values of the call-to-mobility ratio results are not shown here, however it is evident that more gains are to be expected, until eventually the curve reaches the saturation point of $C$=1.

- As the number of partitions increases, the saving in messaging cost is higher. Initially the most significant reduction occurs when we move from the conventional blanket polling strategy to a 2-partition scheme. Then the slope of the curve is smaller for higher number of paging areas; hence our gains are also smaller, until there is no improvement by increasing the number of partitions. Beyond $N$=3 the reduction is very slight for the LIBP scheme with high CMR values.

- The LIBP scheme is always better than the general optimal partitioning strategy with unconditional location probabilities. Even for small CMR values where the location database is not updated very often or the mobile moves rapidly, the saving in signaling cost from the general OP case can be worthwhile. For CMR=0.1 in Fig. 6(a), the percentage saving is 9.68% for $N$=2, 15.07% for $N$=3 and 15.22% for $N$=4.

- The LIBP scheme produces better results quicker than the general OP scheme, so that more gains can be achieved with a smaller price to pay for increasing delays. In the most commonly studied case where CMR is around 1, the optimum value of $N$ that can achieve significant traffic reduction with acceptable delay is $N < 4$, smaller or at least equal than the one for the general unconditioned OP case.

- When the LI cell is near the border of the LA, the paging cost is lower for the same number of partitions and the same CMR. By comparing Figs. 6(a)-(b) we note that, for example, when $N$=2 and CMR=1, the average decrease is 51.56% when LI cell=15 and 61.79% when LI cell=1. This happens because a near border cell has always fewer adjacent cells or fewer 'escape routes' to move, therefore chances are that it will be closer to its origination point at the end of a paging interval. So, in correspondence with our central idea, the conditional probabilities are more concentrated around the LI cell and savings come increased.
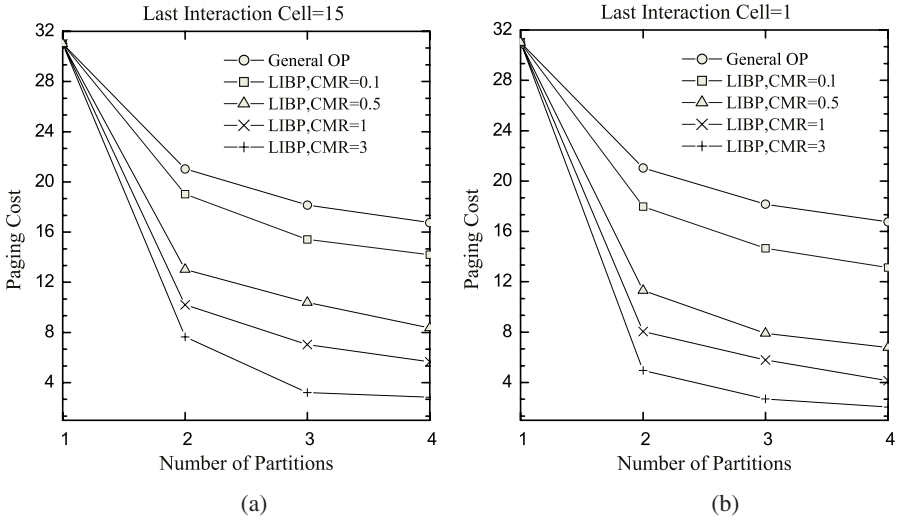
**Fig. 6.** Paging cost vs. number of partitions for the general OP scheme and the LIBP algorithm with different values of the call-to-mobility ratio, when (a) LI cell=15 and (b) LI cell=1
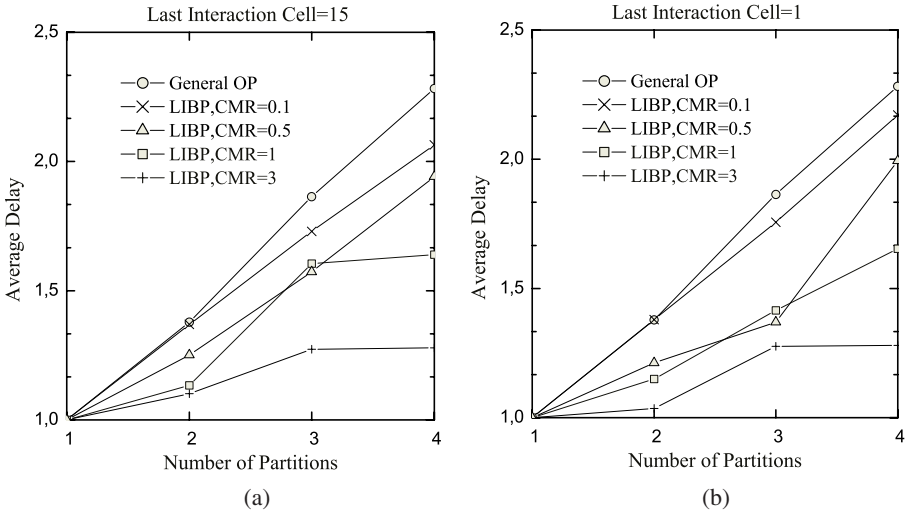


**Fig. 7.** Average delay for the general OP scheme and the LIBP algorithm with different values of the call-to-mobility ratio, when (a) LI cell=15 and (b) LI cell=1

Regarding the metric of average delay to locate a mobile user, we note that the LIBP scheme is again superior to a general OP scheme. For instance, in Fig. 7(a), when $N$=3 the delay for the LIBP scheme is 1.60 polling cycles when CMR=1, while for the general OP case it amounts to 1.86 cycles. This reduction can be quite

significant when translated to real time. As an example, if the duration of a polling cycle is 500 ms, the extra delay when 3 partitions are used is 300 ms for the LIBP scheme and 430 ms for the unconditioned strategy. Thus the overhead delay deteriorates performance much less when the LIBP scheme is employed. Moreover, as for the average cost, the LIBP scheme performs better for higher CMR values. This happens because the location database is updated more often, and thus the probability of finding the user in one of the first steps increases.

We also notice that delay increases approximately linearly for a scheme using the unconditional location probabilities. This is in accordance with previous results in [5,7]. However, in the LIBP scheme this is not true for high CMR; a proportionally greater or lower increase may be observed by adding one more partition. This depends on the exact allocation of cells –and hence location probabilities– to paging areas, which is not symmetrical for increasing number of partitions. In addition, we note that choosing a different starting cell does not affect the results clearly as in the case of paging cost. Delays may increase when the last interaction cell is near the border of the LA.

Finally, when the CMR is very high, a mobile terminal is almost certainly in the last interaction cell, with a small probability to be found elsewhere. In our test cases, when CMR>10, the LI cell probability was >0.9. For such values, partitioning the LA into more than two paging areas is almost useless and does not improve the performance of the algorithm or affect delays. Then an optimal partition of the LA would be first to poll the last interaction cell, and secondly the remaining cells in the area.

# 6   Summary and Conclusion

This paper has presented a novel paging strategy that helps in reducing the signaling load generated by the search operations in a wireless network. The new strategy enhances current partitioning algorithms by exploiting the location information acquired through recent interactions with the system. It employs a mathematical or statistical model to predict conditional location probabilities based on the cell where the last interaction between the network and the terminal took place. The derived location probabilities are more concentrated in a small portion of the location area, which results in greater savings by the application of an optimal partitioning algorithm. Numerical results in a simple LA paradigm have shown that significant savings with smaller delays can be accomplished compared to a conventional partitioning scheme, especially for high usage or low mobility subscribers. Our scheme is particularly flexible in that all or some transaction types may be used to update the system or terminal database about the last interaction position, depending on implementation, protocol and processing cost issues. However, the more frequent transactions are, the higher are the gains of the LIBP scheme.

A symmetric random walk model was used to describe MT movements in a location area and the obtained results correspond to lower bounds on the specific system performance. More realistic mathematical or statistical models can be employed in compliance with the general framework described in the paper, which will help in further attesting the effectiveness of the proposed scheme.

# References

1. Akyildiz, I.F., Ho, J.S.M.: On Location Management for Personal Communications Networks. IEEE Communications Magazine **9** (1996) 138–145
2. Ali, S.Z.: Location Management in Cellular Mobile Radio Networks. In Proc. 13th IEEE PIMRC, Lisbon, Portugal, Sept. 2002, 745–749
3. Plehn, J.: The design of Location Areas in a GSM-network. In Proc. 45th IEEE Veh. Tech. Conf., Chicago, USA, July 1995, 871–875
4. Rose, C., Yates, R.: Minimizing the Average Cost of Paging Under Delay Constraints. Wireless Networks **1** (1995) 211–219
5. Abutaleb, A., Li, V.O.K.: Paging Strategy Optimization in Personal Communication Systems. Wireless Networks **3** (1997) 195–204
6. Markoulidakis, J.G., Sykas, E.D.: Performance Bounds of a Multiple-Step Paging Strategy in Future Universal Mobile Telecommunication Systems. Int. J. Wireless Inf. Networks **3** (1995) 133–147
7. Krishnamachari, B., Gau, R.-H., Wicker, S.B., Haas, Z.J.: Optimal Sequential Paging in Cellular Networks. To appear in Wireless Networks
8. Wang, W., Akyildiz, I.F., Stüber, G.L.: An Optimal Paging Scheme for Minimizing Signaling Costs Under Delay Bounds. IEEE Comm. Letters **2** (2001) 43–45
9. Koukoutsidis, I.Z., Demestichas, P.P., Papaioannou, P.I., Theologou, M.E.: Optimal Decision Strategies for Paging in Wireless Networks. Submitted to Wireless Personal Communications (2003)
10. Wolff, R.W.: Stochastic Modeling and the Theory of Queues. Prentice-Hall, Englewood Cliffs, NJ (1989)
11. Feldman, R.M., Valdez-Flores, C.: Applied Probability and Stochastic Processes. PWS Publishing Co., Boston (1996)
12. Papoulis, A., Pillai, U.S.: Probability, Random Variables and Stochastic Processes. 4th Int. edn. McGraw-Hill, New York (2002)
13. Stickel, E.U.: An Algorithm for Fast High Precision Computation of Matrix Exponential and Logarithm. Analysis **10** (1990) 85–95
14. Koukoutsidis, I.Z., Theologou, M.E.: A Combination of Optimal Partitioning and Location Prediction to Assist Paging in Mobile Cellular Networks. Submitted to Int. J. Wireless Information Networks (2003)
15. Lei, Z., Rose, C.: Probability Criterion Based Location Tracking Approach for Mobility Management of Personal Communications Systems. In Proc. IEEE Globecom, Phoenix, USA, Nov. 1997, 977–981
16. Liang, B., Haas, Z.J.: Predictive Distance-Based Mobility Management for PCS Networks. In Proc. 18th IEEE Infocom, New York, USA, March 1999, 1377–1384