

Individual Profile Graphs for Location Management in PCS Networks^{*}

Chansophea Chuon^{a,*} Sumanta Guha^a

^a*Computer Science and Information Management Program, School of Engineering and Technology, Asian Institute of Technology, P.O. Box 4, Klong Luang, Pathumthani, 12120, Thailand.*

Abstract

A user's profile, for the purpose of location management in a personal communication service (PCS) network, is formalized as a subgraph of the network graph. This subgraph, the so-called individual profile graph (IPG), is determined after a period of observation with the intent of predicting and codifying the user's diurnal routine. The IPG is easily-motivated, robust, straightforwardly computed from observed data, and, under fairly intuitive assumptions, provably predictive of the user's diurnal routine. An IPG-based paging and update strategy is analyzed. It is shown to significantly improve a straight location area (LA) based strategy that ignores user profiles.

Key words: Individual profile graph, location tracking, mobile communication, mobility management, PCS networks, user profiling

1 Introduction

Subscription to personal communication services (PCS) is growing at an exponential rate owing to the increasing portfolio of services offered over wireless channels. A challenge facing PCS providers is to fulfil demand within a limited frequency spectrum. As cell sizes become smaller and the number of mobile

^{*} Preliminary version appeared in Proc. 2005 IEEE International Conference on Wireless Networks, Communications and Mobile Computing (MobiWac 2005).

^{*} Corresponding author.

Email addresses: chansophea.chuon@ait.ac.th (Chansophea Chuon),
guha@ait.ac.th (Sumanta Guha).

URL: <http://www.cs.ait.ac.th/~guha> (Sumanta Guha).

users increases, the signaling cost incurred in both location update and paging increases. Efficient location management per user to minimize update and paging costs, accordingly, becomes vital and has been studied extensively [1–4,6–10,12–16].

The geographical coverage area of a PCS network is, typically, partitioned into a number of location areas (LAs), each composed of a group of cells. A mobile user sends a location update when it crosses the boundary of an LA. Upon call arrival, the user is simultaneously paged in all cells of its currently visited LA. Unfortunately, there are two significant inefficiencies associated with the LA-based update scheme. Firstly, an LA comprising a large number of cells consumes significant radio bandwidth while paging. Secondly, location update by users at the time of crossing LA boundaries generates significant data traffic as well, particularly, for those users located near the boundaries.

Recently, location management schemes have been proposed to reduce signaling traffic based upon profiling the mobility pattern of individual users. We discuss here those that motivated the current work.

Xie et al [16] propose a scheme in which LAs are no longer static but, rather, determined dynamically per user. The authors apply the fluid flow model to represent user mobility. Further, each user’s call arrival rate, as well as parameters determined by its mobility, enter into a signaling cost function for that particular user, and an optimal LA size is determined for the user in terms of this cost function.

Tabbane [14] proposes an alternative strategy (AS) for mobile radio communication that focuses on reducing signaling traffic. In AS, the system maintains a profile for each user that consists of a list of LAs ordered according to the system’s expectation that the user will be found in each. Upon call arrival the LAs in the list are paged sequentially in order of descending expectation. AS actually maintains two sets of profiles – the long-term profile and the dynamic profile. The long-term profile is obtained after a lengthy period of observation, while the dynamic profile changes according to the user’s recent call history. The system chooses between either profile to locate the user. Typically, the dynamic profile is used in case of frequent incoming calls.

Pollini et al [7] propose a profile-based strategy (PBS) that evolved from AS. The authors classify users into three categories, in particular, deterministic, quasi-deterministic and random, depending on the predictability of their diurnal routine.

The definitions of user profiles proposed in the literature to date are mostly heuristically motivated. As the authors do not give a rigorous definition of the user profile, obviously they cannot prove that it captures aspects of a user’s routine. This is the issue that we address here. Our goal is a long-term profile

(as in Tabbane [14]), based on a lengthy period of observation, presumed static after formulation, which can, thereafter, be applied to track the user. However, our approach is different from existing ones.

We define, for each user, a graph that we call the individual profile graph (IPG) for that user, whose intent is to predict and codify the user’s diurnal routine. The IPG itself is determined as a subgraph of the network graph – the latter consisting of all network cells as vertices and edges representing physical adjacencies of cells.

Our definition of the IPG is simple, robust, easily-motivated, and allows for its straightforward computation from observed data. Further, we prove, under fairly intuitive assumptions, that the IPG indeed captures the diurnal routine of a user. We also show how to incorporate the IPG into a paging and update strategy.

We believe that the IPG is the most natural formulation of a user’s profile in the context of location management: on the one hand it arises as a subgraph of the existing network graph, and, on the other, its topology is maximally influenced by the user. This is not the case with LA-based profiling where, for example, the shape constraint on an LA is, typically, system-dependent.

In Section 2 we motivate and define the IPG and prove our claim that it captures a user’s diurnal routine. In Section 3 we describe IPG-based paging and update and analyze its performance, showing it to significantly improve a straight LA-based strategy that ignores user profiles. We conclude in Section 4.

2 Individual Profile Graph

We assume that the network is modeled as an undirected graph $G = (V, E)$ whose vertices V represent cells and edges E represent the adjacency of cells, with no restriction on G other than that its vertices all be of degree $O(1)$. This frees us from the common but somewhat unrealistic constraint that cells in a network are homogeneous in size and shape. Henceforth, we shall use the terms “cell” and “vertex” interchangeably.

A user U that newly enters the system is monitored for some N consecutive days and a daily log is kept for each of these days that records the duration that U spends in each cell. We ignore the technicalities and cost of monitoring U as this is a one-time constant cost incurred per user. Instead, we study how the resulting data can be exploited to develop an efficient update/locate strategy for U .

2.1 Definition

Our approach is to use the log data to define a so-called *individual profile graph* (IPG) of U , denoted G^U , which is subgraph of G that “captures” the diurnal routine of U . The procedure to determine G^U is as follows:

For each vertex $v \in V$, let N_v be the number of days of the N -day monitoring period that U visits v at least once. Normalize, dividing by N , to obtain $n_v = N_v/N$, so that $0 \leq n_v \leq 1$. The value n_v , that we call the *diurnal weight* of v in U 's routine, is the probability that user U visits cell v on a random day.

Given a subset $H \subset V$, let the *induced subgraph* of H , denoted $in(H)$, be the subgraph of G whose vertex set is H , and whose edge set consists of edges in E both of whose end vertices lie in H .

For $0 \leq \mu \leq 1$, define the subgraph G_μ of G as follows:

$$G_\mu = in(\{v \in V : n_v \geq \mu\}) \quad (1)$$

In other words, G_μ consists of cells with diurnal weight at least μ and edges connecting such cells. Let the sequence

$$\mu_1 > \mu_2 > \dots > \mu_r \quad (2)$$

be the sequence, in descending order, of distinct values of n_v , for $v \in V$.

We get a corresponding sequence

$$G_{\mu_1} \subset G_{\mu_2} \subset \dots \subset G_{\mu_r} \quad (3)$$

of subgraphs of G , where the successive containments are strict.

These subgraphs may not all be connected – an evident requirement of an IPG. Accordingly, let

$$G_{\mu_{i_1}} \subset G_{\mu_{i_2}} \subset \dots \subset G_{\mu_{i_s}} \quad (4)$$

be the subsequence of Sequence (3) of connected graphs.

We wish to choose U 's IPG to be the one of $\{G_{\mu_{i_k}} : 1 \leq k \leq s\}$ that represents a “boundary” beyond which diurnal weights drop significantly – intuitively, that $G_{\mu_{i_k}}$ outside of which U rarely travels on a daily basis. Moreover, we wish

to make this choice in a robust manner. Accordingly, consider the successive differences

$$\mu_{i_1} - \mu_{i_2}, \mu_{i_2} - \mu_{i_3}, \dots, \mu_{i_{s-1}} - \mu_{i_s} \quad (5)$$

and let $\mu_{i_K} - \mu_{i_{K+1}}$ be the *maximum* of these (if there is more than one maximum term in the sequence, choose $\mu_{i_K} - \mu_{i_{K+1}}$ to be the leftmost one).

Define the user's IPG to be

$$G^U = G_{\mu_{i_K}} \quad (6)$$

Figure 1 shows a network graph G that is an 8×8 grid graph and the IPG G^U (shown bold) of a hypothetical user U whose diurnal weights are indicated in each cell. The derivation of G^U is indicated in Figure 2. In particular, the ten distinct values of the diurnal weights in G are:

$$0.96 > 0.92 > 0.82 > 0.76 > 0.70 > 0.66 > 0.64 > 0.18 > 0.06 > 0.00$$

Figure 2 shows separately the ten subgraphs

$$G_{0.96} \subset G_{0.92} \subset G_{0.82} \subset G_{0.76} \subset G_{0.70} \subset G_{0.66} \subset G_{0.64} \subset G_{0.18} \subset G_{0.06} \subset G_{0.00}$$

of which the following subsequence is of connected subgraphs

$$G_{0.96} \subset G_{0.92} \subset G_{0.66} \subset G_{0.64} \subset G_{0.18} \subset G_{0.06} \subset G_{0.00}$$

which proves that the IPG G^U is $G_{0.64}$.

Observations:

- (1) Our current definition of the IPG is coarse in that it uses only per-day data and not intra-diurnal data, that is assumed available as well, in particular, the user's mobility pattern during a day. We do, however, utilize intra-diurnal data in our paging strategy discussed in the next section.
- (2) Our current definition of the IPG is independent of any grouping of cells into LAs, if, in fact, the network be configured in such a manner. A straightforward extension is to simply replace cells with LAs in the construction of the IPG. However, more sophisticated approaches are possible – we discuss LAs in the context of IPGs in the concluding section.

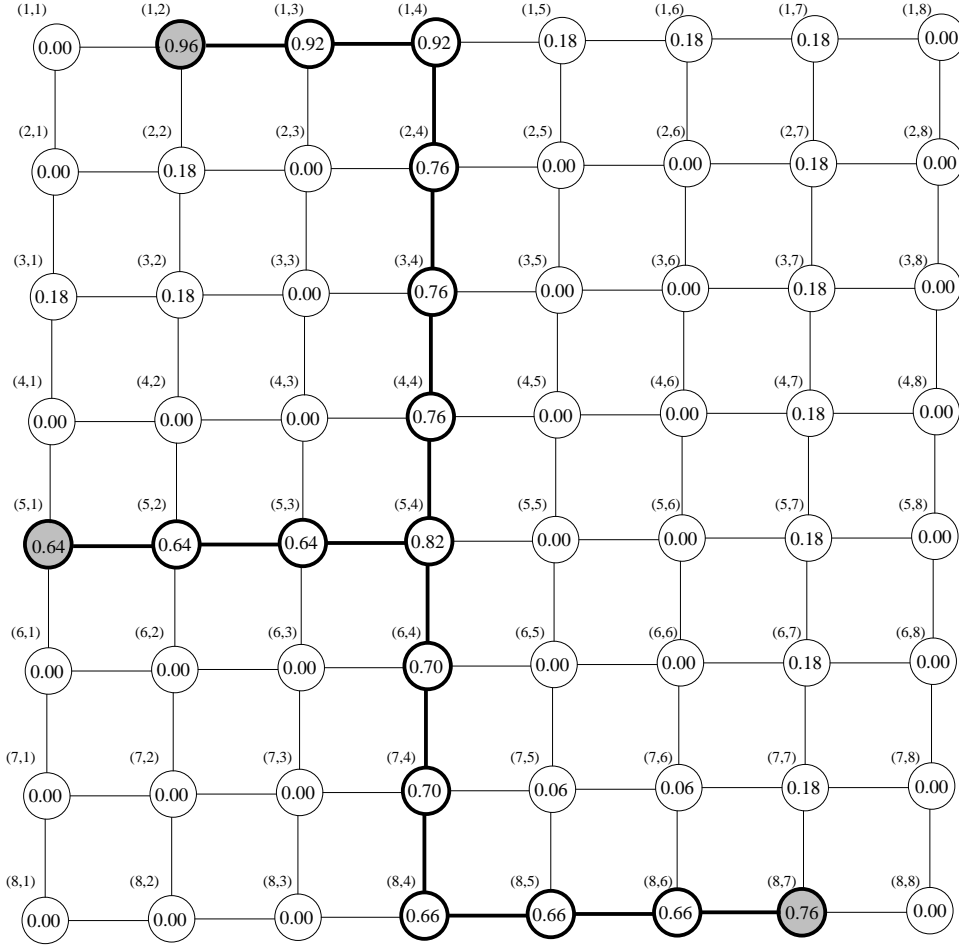


Fig. 1. Individual profile graph (bold) as a subgraph of the network graph – diurnal weights are shown in each cell. Possible anchor cells are shaded.

- (3) IPGs are particularly suitable for users that would be classified as either deterministic or quasi-deterministic in the taxonomy introduced by Pollini et al [7]: these are users whose daily mobility pattern tends to be repetitive. Random users, on the other hand, almost by definition, do not yield an IPG of any value.

2.2 Validity

We prove next that, under certain fairly intuitive assumptions, the IPG G^U indeed captures U 's diurnal routine.

We are told, say, by an oracle, that U 's diurnal routine is “actually” captured by the subgraph G_*^U of G , called the *true IPG* of U , which is specified in the following manner:

A set A , of size $O(1)$, of vertices of G is given as the set of *anchor cells*, or,

simply, *anchors*, of U 's diurnal routine. Intuitively, an anchor cell is one in which U tends to linger after entering (*anchor stickiness condition*), and that U visits frequently (*anchor popularity condition*). For example, cells containing U 's home and office are likely to be anchors, but a cell on his route from home to office is likely not, as it's popular but not sticky. On the other hand, the cell containing his dentist's chamber is probably sticky, but not particularly popular.

In addition, a set P , of size $O(1)$, of directed paths in G – each path starting from an anchor and ending at an anchor – is given as the set of *connector paths*, or, simply, *connectors*, of U 's diurnal routine. Intuitively, a connector is a path that U tends to take when traveling from one anchor to another (*connector preference condition*). For every anchor cell v , we assume that there is at least one connector starting from v and at least one ending at v .

Finally, the graph G_*^U is defined, simply, as the subgraph of G that consists of all anchor cells of U 's diurnal routine, together with all cells and edges that lie on some connector. Note that, though each connector path is directed, G_*^U itself is an undirected subgraph of G . We make the reasonable assumption that, as an undirected subgraph of G , G_*^U is connected.

In Figure 1, three possible anchor cells (1, 2), (5, 1) and (8, 7) are shown shaded, and corresponding connector paths may be chosen in an obvious manner.

We wish to prove that $G^U = G_*^U$, but, of course, this is not possible without certain assumptions about how the oracle deemed G_*^U as the true IPG in the first place. We formalize next a set of assumptions about G_*^U .

Say, that a day is divided into T equal time slots, called, simply, *slots*, e.g., 24 one-hour slots or 48 half-hour slots, and, further, we make the simplifying assumption that U spends each slot entirely in one cell (this is a fair assumption if slots are short periods of time). Call a maximal interval of slots spent by U in a cell v to be a *block* spent by U in v ; in other words a block B consists of some number of consecutive slots t_1, t_2, \dots, t_m such that U is in v for all slots t_i , for $1 \leq i \leq m$, but is not in v in either the slot preceding t_1 or the slot succeeding t_m . The *size* of a block B is the number of slots comprising it. We'll assume that each interval of time spent by U in any cell v follows an exponential distribution, which is a common assumption [5]. As the exponential distribution is continuous, in subsequent calculations we'll discretize it over the integral block size.

The stickiness of an anchor cell v is straightforwardly estimated by the sizes of the blocks U spends in v . One has to be more careful in gauging popularity. Say, that the outdegree of an anchor cell v in G_*^U is d_v , in particular, that there are d_v connectors leaving it (by an earlier assumption, $d_v \geq 1$). As these connectors are supposedly paths that U takes on a daily basis, one would

expect, on a typical day, that the number of blocks spent by U in v is at least as large as d_v , in order to allow him to exit along each connector at least once.

The preference of U for a connector p starting at an anchor cell v is measured by the frequency with which he takes p on completion of a block spent at v . We'll make the simplifying assumption that connectors leaving v are each about equally preferable; further, we'll assume that when U takes a particular connector p out of v he traverses it fully to the anchor at the other end of p .

A last assumption that we'll make is that cells not belonging to the true IPG G_*^U are not popular in that U visits them infrequently (*non- G_*^U cell unpopularity condition*).

Accordingly, our assumptions of G_*^U is the following list:

- (a) (Anchor Stickiness) The mean size s_v of blocks spent by U in an anchor cell v is large; in particular, $s_v \geq S$, where S is a constant.
- (b) (Anchor Popularity) The mean number of (distinct) blocks b_v per day spent by U in an anchor cell v is large; in particular, the time spent by U in anchor cell v follows an exponential distribution with mean b_v , where $b_v \geq d_v$ (recall that d_v is the outdegree of v) – again we'll discretize the distribution over the integral block size.
- (c) (Connector Preference) At the end of a block spent by U in anchor cell v , U exits v to travel along each of the d_v connectors, say, p_1, p_2, \dots, p_{d_v} , that leaves v , with probability at least $1/d_v - \delta$, where δ is a small constant. If U exits along connector p_i , for some i , where $1 \leq i \leq d_v$, then U traverses all the vertices on path p_i to reach the anchor v' at the other end.
- (d) (Non- G_*^U Cell Unpopularity) The mean number of (distinct) blocks b'_v per day spent by U in a cell v not belonging to G_*^U is small; in particular, the time spent by U in a cell v , not belonging to G_*^U , follows an exponential distribution with mean b'_v , where $b'_v \leq \delta'$, and where δ' is a small constant – we'll discretize the distribution over the integral block size.

Proposition 1 *With the assumptions (a)-(c) above, the diurnal weight of any cell in the true profile graph G_*^U is at least $0.36787 - \epsilon$, where $\epsilon \rightarrow 0$ as $\delta \rightarrow 0$.*

Proof. The exponential distribution with mean d is

$$f(x) = \frac{1}{d} e^{-x/d}$$

Since block sizes are integral we discretize to obtain the probability $P_d(n)$ that the block size is n by integrating the exponential distribution over the interval $[n, n + 1]$. Accordingly,

$$\begin{aligned}
P_d(n) &= \int_n^{n+1} \frac{1}{d} e^{-x/d} \\
&= e^{-n/d} (1 - e^{-1/d})
\end{aligned} \tag{7}$$

We shall prove the proposition for anchor and non-anchor cells of G_*^U separately.

First, consider an anchor cell v . Since the outdegree of v is d_v , by assumption (b), U spends n blocks in v with probability $P_{b_v}(n)$. Therefore, the probability that U visits v at least once in a day is

$$\begin{aligned}
n_v &= \sum_{n=1}^{\infty} P_{b_v}(n) \\
&= 1 - P_{b_v}(0) \\
&= e^{-1/b_v} \quad \text{using Equation (7)} \\
&\geq e^{-1/d_v} \\
&\geq e^{-1} \\
&> 0.36787
\end{aligned}$$

for every anchor v , as $b_v \geq d_v \geq 1$.

Next, consider a cell w of G_*^U that is not an anchor. Accordingly, w must lie on a connector p that starts from, say, anchor cell v . Note that w could lie on more than one such connector. Therefore, the probability n_w that U visits w at least once in a day is at least

$$1 - \sum_{n=0}^{\infty} \text{Prob}(U \text{ spends } n \text{ blocks in } v) \times \text{Prob}(U \text{ does not take } p \text{ each of the } n \text{ times he leaves } v)$$

Therefore,

$$\begin{aligned}
n_w &= 1 - (P_{b_v}(0) + P_{b_v}(1)(1 - 1/d_v - \delta) + P_{b_v}(2)(1 - 1/d_v - \delta)^2 + \dots) \\
&= 1 - ((1 - e^{-1/b_v}) + e^{-1/b_v}(1 - e^{-1/b_v})(1 - 1/d_v - \delta) \\
&\quad + e^{-2/b_v}(1 - e^{-1/b_v})(1 - 1/d_v - \delta)^2 + \dots) \\
&= 1 - (1 - e^{-1/b_v}) \times \\
&\quad (1 + e^{-1/b_v}(1 - 1/d_v - \delta) + e^{-2/b_v}(1 - 1/d_v - \delta)^2 + \dots) \\
&= 1 - \frac{1 - e^{-1/b_v}}{1 - e^{-1/b_v}(1 - 1/d_v - \delta)} \\
&\geq e^{-1} - \epsilon
\end{aligned}$$

$$> 0.36787 - \epsilon$$

where $\epsilon \rightarrow 0$ as $\delta \rightarrow 0$, as $d_v = O(1)$, by our assumption that the size of the set P of connectors is $O(1)$.

Corollary 2 *With the assumptions (a)-(d) above, if δ and δ' are sufficiently small, then*

$$G_*^U = G^U$$

Proof. Using the theorem, assume that δ is small enough that $0.36787 - \epsilon > 0.36$, implying that the least of the diurnal weights of any cell v in the true profile graph G_*^U is the term, say, $\mu_{K'}$ in Sequence (2), where $\mu_{K'} > 0.36$.

The probability that U visits a cell v not in G_*^U at least once in a day is

$$\begin{aligned} n'_v &= \sum_{n=1}^{\infty} P_{b'_v}(n) \\ &= 1 - P_{b'_v}(0) \\ &= e^{-1/b'_v} \\ &\leq e^{-1/\delta'} \\ &\leq 0.1 \end{aligned}$$

as $b'_v \leq \delta'$, assuming that δ' is sufficiently small. This implies that the largest of the diurnal weights of any cell v not in G_*^U is the term $\mu_{K'+1}$ of Sequence (2), where $\mu_{K'+1} < 0.1$.

It follows that the graph $G_{\mu_{K'}}$ of Sequence (3) is, in fact, equal to G_*^U . Therefore, $G_{\mu_{K'}}$ is connected and appears in Sequence (4) as, say, the term $G_{\mu_{i_K}}$.

From the inferences drawn above about the least and largest values of the diurnal weights of cells in G_*^U and not in G_*^U , respectively, we note the following of Sequence (5):

$$\begin{aligned} \mu_{i_K} - \mu_{i_{K+1}} &> 0.26 \\ \mu_{i_k} - \mu_{i_{k+1}} &< 0.26, \quad \text{if } k < K \\ \mu_{i_k} - \mu_{i_{k+1}} &< 0.1, \quad \text{if } k > K \end{aligned}$$

It follows that $\mu_{i_K} - \mu_{i_{K+1}}$ is the unique maximum of terms in Sequence (5). Then, from the definition of G^U in Equation (6) it is seen that, indeed, $G^U = G_*^U$.

3 Paging and Location Update Strategy

We discuss next how to apply the IPG to paging and location update. Our current strategy is somewhat simplistic but the analysis points to possible refinements.

The user-end of our current strategy is straightforward. User U is not required to update his location as long as he is in a cell of G^U ; however, U is required to update at *each* cell he visits that does not belong to G^U . This strategy, of course, assumes that at the end of the tracking period, when the PCS network system has determined the user's IPG G^U , G^U is cached in the user's device with help of appropriate user-end software so that the user has knowledge of its own IPG.

We explain next the system-end of our strategy. Assume that each day is divided into T equal slots (refer to the definitions of slot and block in the preceding section). The probability α_v that U is in cell v in a random time slot is, clearly, given by

$$\alpha_v = \frac{b_v s_v}{T}$$

as b_v is the mean number of blocks spent by U in v and s_v is the mean block size – we assume that the additional data of b_v and s_v is stored and pointed at from the vertices of G^U .

We follow Pollini et al [7] in paging the cells of G^U in order of decreasing probability. Assume, accordingly, that

$$\alpha_{v_1} \geq \alpha_{v_2} \geq \dots \geq \alpha_{v_m}$$

where v_1, v_2, \dots, v_m are the vertices of G^U .

Let C_p and C_f denote the cost of successful and unsuccessful paging in a cell, respectively. The expected cost of paging per call when U is inside the IPG is, accordingly, given by

$$\alpha_{v_1} C_p + \alpha_{v_2} (C_f + C_p) + \alpha_{v_3} (2C_f + C_p) + \dots + \alpha_{v_m} ((m-1)C_f + C_p) \quad (8)$$

On the other hand, U is outside the IPG with probability $1 - \sum_{i=1}^m \alpha_{v_i}$. In this case, by requirement of our strategy, the system has already been updated as to the cell where U is currently located. Accordingly, the expected cost of paging per call is, simply, C_p .

We conclude that the total expected paging cost per call is given by

$$\begin{aligned}
C_p &= \alpha_{v_1} C_p + \alpha_{v_2} (C_f + C_p) + \alpha_{v_3} (2C_f + C_p) + \dots \\
&\quad + \alpha_{v_m} ((m-1)C_f + C_p) + \left(1 - \sum_{i=1}^m \alpha_{v_i}\right) C_p \\
&= C_p + (\alpha_{v_2} + 2\alpha_{v_3} + \dots + (m-1)\alpha_{v_m}) C_f
\end{aligned} \tag{9}$$

Next, consider location update. If a single location update cost is C_u , then the total expected location update cost using our strategy is, simply,

$$C_u = \left(1 - \sum_{i=1}^m \alpha_{v_i}\right) C_u \tag{10}$$

because the user updates only at a cell outside G^U .

3.1 Comparison with Traditional LA-Based Methods

We compare the IPG-based strategy with traditional LA-based location management. For this purpose we assume that a node in of the IPG is a location area, rather than a cell. This is realistic because in an LA-based scenario locations are the geographical unit and, moreover, the IPG-method is independent of whether nodes are cells or location areas, and the analysis above goes through as well.

In order to determine the update cost in the LA-based scheme observe that the mean number of times that a user U leaves the LA v is identical to the mean number of blocks b_v that he spends in v (as on completion of each block in v , by definition U leaves v). Therefore, as long as U is in the IPG the expected LA-based update cost is

$$\left(\sum_{i=1}^m b_{v_i}\right) C_u$$

because the scheme requires an update whenever U crosses an LA boundary.

Multiplying the above expression by the probability that U is in the IPG we get

$$\left(\sum_{i=1}^m \alpha_{v_i}\right) \left(\sum_{i=1}^m b_{v_i}\right) C_u$$

as the expected update cost incurred while the user moves within the IPG.

Outside the IPG the LA-based method costs exactly as much as the IPG-based one, given by Equation (10):

$$\left(1 - \sum_{i=1}^m \alpha_{v_i}\right) C_u$$

Therefore, the total expected update cost in an LA-based strategy is the sum of the previous two expressions:

$$\begin{aligned} C'_u &= \left(\sum_{i=1}^m \alpha_{v_i}\right) \left(\sum_{i=1}^m b_{v_i}\right) C_u + \left(1 - \sum_{i=1}^m \alpha_{v_i}\right) C_u \\ &= \left(1 + \sum_{i=1}^m \alpha_{v_i} \left(\sum_{i=1}^m b_{v_i} - 1\right)\right) C_u \end{aligned} \quad (11)$$

Denoting $\sum_{i=1}^m \alpha_{v_i}$, the probability that U is in the IPG, by P , and denoting $\sum_{i=1}^m b_{v_i}$, the frequency with which U moves in and out of LA's, by X , a simple calculation using Equation (10) shows that:

$$C'_u = C_u + PXC_u \quad (12)$$

Equation (12) bears out the intuition that the more time U spends within the IPG (in fact, if the IPG is being deployed, by definition it captures a substantial part of the U 's diurnal routine, so that P is likely to be high), and the more frequently he travels within the IPG, the more efficient is the IPG-based update scheme is likely to be compared with the naive LA-based scheme.

The expected paging cost in the LA-based scheme is simply

$$C'_p = C_p \quad (13)$$

because it is the same in each LA regardless of if that LA belongs to the IPG or not. Using Equation (9), we see that

$$C'_p = C_p - (\alpha_{v_2} + 2\alpha_{v_3} + \dots + (m-1)\alpha_{v_m})C_f \quad (14)$$

The paging cost in the naive LA-based strategy is lower than that of the IPG-based one – as expected because of the investment that the LA-based strategy makes in updating. However, as we'll see next, the difference in paging costs may be minimized by optimizing the IPG.

From Equation (14) we find that

$$C_p \rightarrow C'_p \text{ as } \alpha_{v_1} \rightarrow 1 \quad (15)$$

as the co-efficient of C_f in Equation (14) tends to 0 as α_{v_1} increases to 1.

So, if α_{v_1} is nearly 1 then the paging costs of both LA-based and IPG-based strategies are comparable. Unfortunately, there is no *a priori* reason to expect that α_{v_1} to be large. Intuitively, this means that there is no reason to expect one of the anchors (or, in fact, any LA with a high diurnal weight) to be more popular than another *over the entire diurnal routine*. However, if the time of day is taken into account then users with predictable routines are likely to be found in certain unique LA's with a high probability (implying that α_{v_1} is indeed high). The refinement suggested, therefore, is the following:

The user's daily log (which, as indicated earlier, includes information about his location in each time slot) be applied to determine the most likely LA as a function of the current time slot. Say, this LA is v_t for time slot t . If a call arrives in time slot t , the first LA paged is v_t . Subsequently, LA's are paged outwards along connectors passing through v_t – these connectors themselves may be ranked in order of decreasing likelihood.

Note that v_t need not be an anchor (e.g., if the time of the day happens to be one when U is typically en route from one anchor to another). The optimal paging strategy evidently depends upon the information that is recorded and associated with users' IPGs. We have yet to investigate more sophisticated paging strategies based on such refined IPGs.

4 Conclusions and Future Work

The most significant contribution of the current work is a rigorous formulation of a PCS user's individual profile that is easily computed, robust, and provably predictive of the user's diurnal routine from the point of view of location management.

We believe that the IPG-based location management can eventually prove superior to LA-based management. Because of the way the topology of the IPG is extracted from that of the network in a manner driven by the user's routine, we argue that IPGs are "natural" as profiles, more so than even dynamically-determined LAs (e.g., Xie et al [16]). For example, Figure 3 shows a typical location area in a network of hexagonally-shaped cells, indicating clearly a system-dependent shape constraint that is absent from IPGs.

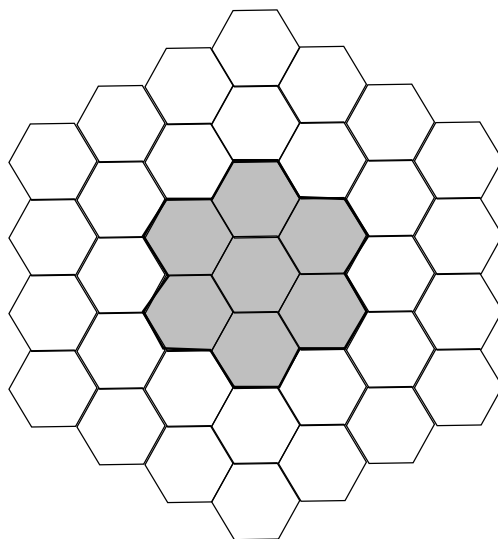


Fig. 3. LA (shaded) in a network of hexagonal cells.

Work remains to be done to further solidify the theory and address implementation issues. We list those that seem to be the most important:

- (1) The current definition of the IPG is “theoretically sound”, but, admittedly, may not be the best in certain practical scenarios. For example, if the maximum of Sequence (5) is not a “clear winner”, then the IPG may exclude cells that possibly should be included in the diurnal routine for the purpose of efficient paging, or, conversely, include unimportant cells.

Accordingly, an interesting question is to formulate a robust definition of the IPG, in the case that the Sequence (5) is noisy.

- (2) Refinement and enhancement of the IPG, suggested in the previous section, based upon time-slot based location data, rather than, simply, diurnal averages, is, to our minds, critical to successfully applying IPGs to efficiently page and locate users. Merely paging cells along a slope of decreasing likelihood is not adequate, and far more powerful performance is to be had by recording and associating intra-diurnal data with the IPG.

In fact, Proposition 1 and, in particular, Corollary 2, relating the IPG with the true IPG, suggest refinements using intra-diurnal data, as well.

From an implementation point of view, a user whose profile alters substantially (say from the week to the weekend) can be assigned more than one profile to take such behavior into account. The cost to do this is simply one-time in the initial observation phase.

- (3) A natural companion to the definition of the IPG is that of a metric to measure its performance. Other than the obvious utility of such a metric in analyzing the system’s performance, it could conceivably be used as a source of feedback to modify the IPG over time. An incidental utility could, as well, be to enable the system to rigorously classify users according to predictability (following, say, the taxonomy of Pollini et al [7]). Users would vary from deterministic to random according to the

accuracy with which their routines are predicted by their respective IPGs.

Acknowledgments. We would like to thank the anonymous referees for suggestions that significantly helped improve the paper.

References

- [1] I.F. Akyildiz, J. Ho, and Y.B. Lin, Movement-Base Location Update and Selective Paging for PCS Networks, *IEEE/ACM Transactions on Networking*, Vol. 4, No. 4, 1996, 629-638.
- [2] I.F. Akyildiz and W. Wang, The Predictive User Mobility Profile Framework for Wireless Multimedia Networks *IEEE/ACM Transactions on Networking*, Vol. 12, No. 6, 2004, 1021-1035.
- [3] A. Bar-Noy, I. Kessler, and M. Sidi, Mobile Users: To Update or Not to Update?, *Wireless Networks*, Vol. 1, No. 2, 1995, 175-195.
- [4] A. Bhattacharya and S.K. Das, LeZi-Update: An Information-Theoretic Approach to Track Mobile Users in PCS Networks, *Proceedings ACM/IEEE MOBICOM*, 1999, 1-12.
- [5] F. Khan and D. Zeghlache, Effect of cell residence time distribution on the performance of cellular mobile networks, *Proceedings IEEE Vehicular Technology Conference*, Vol. 2, 1997, 949 – 953.
- [6] B. Liang and Z.J. Haas, Predictive distance-based mobility management for multidimensional PCS networks, *IEEE/ACM Transactions on Networking (TON)*, Vol. 11, No. 5, 2003, 718–732.
- [7] G.P. Pollini and I. Chi-Lin, A Profile-Based Location Strategy and Its Performance, *IEEE Journal on Selected Areas in Communications*, Vol. 15, No. 8, 1997, 1415-1424.
- [8] A. Quintero, A User Pattern Learning Strategy for Managing Users' Mobility in UMTS Networks, *IEEE Transactions on Mobile Computing*, Vol. 4, No. 6, Nov/Dec, 2005, 552–566.
- [9] A. Quintero and O. Garcia, A Profile-Based Strategy for Managing User Mobility in Third-Generation Mobile Systems, *IEEE Communication Magazine*, Vol. 42, No. 9, Sep 2004, 134-139.
- [10] C. Rose, R. Yates, Minimizing the Average Cost of Paging Under Delay Constraints, *Wireless Networks*, Vol. 1, No. 2, 1995, 211-219.
- [11] S.M. Ross, *Introduction to Probability Models*, Amsterdam, Academic Press, 2003.
- [12] S.K. Sen, A. Bhattacharya, and S.K. Das, A Selective Location Update Strategy for PCS Users, *Wireless Networks*, Vol. 5, No. 5, 1999, 313-326.

- [13] D. Senzaki, H.G. Chakraborty and M. Matsuhara, Distance Based Location Management in Cellular PCS Network -a Critical Study, *Proceeding 18th International Conference on Advanced Information Networking and Applications (AINA '04)*, 2004, 95.
- [14] S. Tabbane, An Alternative Strategy for Location Tracking, *IEEE Journal on Selected Areas in Communications*, Vol. 13, No. 5, 1995, 880-892.
- [15] V.W.-S Wong and V.C.M Leung, Location Management for Next-Generation Personal Communication Networks, *IEEE Networks*, Vol. 14, No. 5, 2000, 18-24.
- [16] H. Xie, S. Tabbane, and D.J. Goodman, Dynamic Location Area Management and Performance Analysis, *Proceedings 43rd IEEE Vehicular Technology Conference*, 1993, 536-539.