Mining Call and Mobility Data to Improve Paging Efficiency in Cellular Networks

Hui Zang
Sprint, California, USA
hui.zang@sprint.com

Jean C. Bolot
Sprint, California, USA
bolot@sprint.com

ABSTRACT

Locating mobile users and devices efficiently is a critical operation in cellular networks. This is done using a combination of location update (by the mobile) and paging (by the network). The paging scheme determines how and where to search for a mobile user given the latest location update information from that user. In this paper, we consider how to increase the efficiency of the paging scheme.

Much previous work has relied on simulation or modeling to design and evaluate the performance of proposed paging schemes. We take a different, data-driven approach in how we design and evaluate our solution. Specifically, we mine more than 300 million call records from a large cellular operator to characterize user mobility and create mobility profiles. We then develop a family of profile-based paging techniques, considering both static schemes and dynamic schemes which adapt as user profiles continuously get updated. We find that our paging techniques can dramatically reduce signaling load (up to 80%) with minimal increase in paging delay (usually less than 10%).

Categories and Subject Descriptors

C.2.3 [**Computer-Communication Networks**]: Network Operations

General Terms

Measurement, Performance, Design, Security

Keywords

Paging, Mobility, Data Mining, CDMA

1. INTRODUCTION

Location management is a key component in the operation of cellular networks. Two basic operations are involved in managing mobile users and tracking their location, namely location update and paging. The location update scheme determines when it is necessary for mobile users and devices to report their location. The paging scheme determines how and where to search for a mobile user given the latest location update information from that user. In the extreme case when the mobile updates its location every time it enters a new cell, there would be no need for paging since the network would know exact which cell to direct calls or connections to in order to reach that mobile device or user. In practice, however, updating location at every cell boundary is too costly. More generally, there is a fundamental tradeoff between location update and paging. As the frequency of location update increases, the uncertainty around the exact location of the mobile decreases and therefore the cost of paging the mobile decreases. Conversely, as the frequency of location update decreases, the uncertainty around the location of the mobile increases, and so does the paging cost.

It is critical for cellular operators to implement paging schemes which locate mobile devices quickly (so as to keep delay low) and at low cost (because paging consumes valuable spectrum and signalling resources and because the paging channel is a low bandwidth channel). Typically, the approach taken is to minimize paging cost, and in particular to reduce network signalling, in exchange for an increase in paging delay (see for example [1] and Section 2). We follow this approach in the paper. Specifically, we attempt to increase the efficiency of the signaling process in general, and thereby decrease the utilization of the signaling channels. We do this developing very effective location management and paging techniques. As a result, we obtain a significant increase in the average available capacity of the signaling channels.

Note that efficient paging schemes that decrease the utilization of the signaling channels also increase the robustness of the network to paging overloads or attacks. This is an important feature of paging schemes as recent research has shown that control channels such as the paging channel can easily become overwhelmed by targeted denial of service (DoS) attacks including SMS attacks [2] or UDP scanning attacks [3]. Therefore, it has become critical to develop techniques to prevent signaling attacks or to minimize their impact. Several approaches have been considered, ranging from filtering and rate limiting [2] to improved resource allocation and scheduling [4]. Another approach could be taken, based on the results in this paper, namely to increase the average available capacity of signaling channels by minimizing paging traffic, and therefore to increase the robustness of the channels to overloads or DoS attacks. Of course, this approach does not prevent or completely eliminate, but only

MobiCom'07, September 9–14, 2007, Montréal, Québec, Canada.

decreases or delays, the impact of attacks. Still, it increases the intensity required for an attack to disrupt the network and as such can be valuable component in the panoply of security solutions against signaling DoS attacks.

Much previous work on paging schemes and location management has relied on simulation or modeling to design and evaluate the performance of their proposed schemes (however, see for example [5]). In this paper, we take a different, data-driven approach in how we develop and evaluate our solution: we mine extensive cellular call record data (so called PCMD, or Per Call Measurement Data) from a large cellular operator to characterize the behavior of mobile users. Specifically, *we analyze more than 300 million call records collected in three US metropolitan areas to characterize the calling activity and mobility patterns of 2 million users in more than 400 cells. We then use those characteristics to design location management schemes that minimize signaling traffic given the observed user profiles. We consider both static schemes as well as dynamic schemes which adapt as user profiles get updated on a continuous basis.*

Thus, we make two main contributions in this paper. First, we present what we believe is one of the first reports analyzing large scale call record data (PCMD) as well as calling and mobility patterns of mobile users in a nationwide cellular network. Second, we use that data to develop a dynamic profile-based paging/location management technique which significantly increases the effectiveness of the location management process, with average paging success rates across voice/data/SMS calls above 85% and paging success rates for specific applications such as data calls above 95%. Deploying such paging schemes will correspondingly reduce signaling load by 85% or 95%, at a cost of a slight increase in paging delay.

The rest of the paper is organized as follows. In Section 2, we discuss related work and approaches. In Section 3, we present background information on PCMD and location management techniques, and we describe our methodology, i.e. the methods and techniques used to analyze the data and to design our location management solution. In Section 4, we analyze the PCMD traces in detail and derive the key mobility and locality characteristics of users. In Section 5, we develop our dynamic profile-based paging/location management technique, analyze its performance and compare it to other techniques in the literature. Section 6 concludes the paper.

2. RELATED WORK

Location management refers to the set of techniques used to locate users efficiently. The key characteristic of a location management scheme involves the trade-off between the search or paging cost (how much work is involved in finding the mobile user) and the update cost (how much work is involved in the system keeping track of moving users). The paging cost is a function of the number of cells broadcasting a search message, also referred to as the size of the paging area. The update cost is the cost for mobiles to update their location registry in the network through the access channel. It is typically measured by the average number of registrations per user per unit of time.

A large number of location management schemes have been described in the literature [6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16]. The schemes can be divided into several categories, such as location-based schemes, in which mobiles update whenever they enter/leave a location area [17]; time-based schemes, in which updates are triggered periodically by a pre-defined timer [6, 8]; distance-based schemes, in which mobiles update when they move beyond a distance threshold from the location where they most recently registered [15, 10]; movement-based schemes, in which mobile register after they cross a certain number of cell boundaries [13, 1]; velocity-based schemes, in which mobiles are paged in an area calculated based on their velocity [18, 19, 20]; state-based schemes, in which mobiles estimate the paging cost of the system in each state and select registration points to minimize the total of registration and paging cost [21]; and profile-based schemes which keep track of user mobility patterns in profiles and configure paging areas (in either a static or dynamic fashion) based on profile information [22, 12, 23, 24]. Other related approaches use techniques such as caching to reduce network and database resource requirements [25]. The paging schemes described in this paper fall in the general category of profile-based schemes.

Location management schemes are typically examined and evaluated using simulation and models. For example, distance-based paging has been shown to perform better than time-based and movement-based in the case of a memoryless mobility model [6]. Much research investigates optimizations such as concurrent paging [26], or develops methods to select the appropriate thresholds, such as the timer value, the size of the location and paging area, so as to minimize reasonable cost functions [27, 12, 28, 16]. User mobility is modeled using probability vectors describing the probability of user i of being in cell j with various independence assumptions.

We take a different approach in this paper. *Our goal is not to optimize a known scheme using different mobility or paging hypothesis, or to develop a performance model of a specific location management scheme. Rather, we start from operational, large scale mobility data, and we determine which scheme leads to the largest reduction of signaling traffic.* Our results show that schemes based on dynamic mobility profiles perform best.

Our results also show the importance of building schemes based on empirical and measured user behavior. We mine PCMD databases from a nationwide cellular carrier to characterize the behavior of mobile users, in particular their mobility profiles. Mobility profiling or mobility modeling has been the subject of much research over the past several years. However, much of the work has focused on two areas, namely small to medium scale wireless networks, typically campus or small metro networks, and wireless networks operating in unlicensed bands such as WiFi networks (e.g. [29, 30]). This is not surprising since those networks are easier to instrument and they often include users (such as students or corporate employees) willing to have their movements monitored and analyzed (at least in an aggregate way). The interest in ad-hoc or opportunistic networks has also led researchers to investigate mobility patterns in those networks [31, 32, 33].

In contrast, little has been published on the analysis of mobility patterns in large scale cellular networks. The most relevant related work, published recently in [5, 34], considers traces from a CDMA2000 1x network, similar to the network examined in this paper. Reference [34] focuses on the call arrival processes and the daily/weekly periodicity of user behavior. Reference [5] analyzes the same kind of data, namely PCMD, also analyzed in this paper. There are several differences, however. In particular, the work in [5] only considers data sessions, whereas we consider data, voice, and SMS sessions. Also, it considers ”static” or aggregate properties, such as the distribution of visited cells or the correlation between mobility and activity level, whereas we consider spatial and temporal correlations to characterize trajectories (e.g. correlation between cell visits). Finally, its intent is different - our goal is to analyze user mobility not to derive a model of mobility, but to derive profiles of user behavior and use those profiles to design and develop a location management system which minimizes the utilization of signaling channels.

3. BACKGROUND AND METHODOLOGY

In this section, we provide relevant background information on location management and on network monitoring with Per Call Measurement Data (PCMD).

3.1 The Paging Channel

The paging channel is used to carry signaling messages to a mobile when the mobile has no dedicated channel assigned, i.e., no traffic channel assigned. The operation of the paging channel is similar in GSM [35] and in CDMA networks [36]. In this paper, we consider a CDMA2000 network. The problem and solutions apply to other cellular networks including 3G and 4G networks.

The paging channel carries messages from a base station to mobiles and the access channel carries messages from mobiles to a base station. Both channels are shared by all mobiles in the same cell. Whenever a mobile is not assigned a traffic channel, it monitors the paging channel for both system parameters and paging requests. For power saving purposes, a mobile device does not monitor the paging channel continuously. Instead, it stays in so-called sleep mode most of the time and wakes up periodically to check the paging channel for messages. Therefore, the mobile and the network must agree on a schedule so that a message to a mobile is only transmitted on the paging channel when the mobile is due to wake up.

This is done as follows: The paging channel is divided into 80ms slots. The time between two consecutive slots monitored by the same mobile is referred to as a cycle and it is set in most commercial networks to 32 or 64 slots, or 2.56 or 5.12 seconds, respectively. A mobile is assigned to a particular slot in a cycle based on one of its identifiers called the International Mobile Station Identifier (IMSI) [36]. The base station knows each mobile's IMSI and hence can infer when a mobile is due to wake up and can hold messages for that mobile until then.

A network usually does not have accurate information about which cell the mobile is located when a call termination request arrives.[1] To locate the mobile, the network broadcasts a paging message to all the cells in a group of cells called the paging area. When a mobile receives a paging message, it replies with its location and then stays awake until either a call goes through and is completed, or no new message is received in a while (in practice, before a timer expires). While the mobile is awake, signaling messages to that mobile are not broadcast but only sent to the cell the mobile replied from.

[1]“Call termination”, as opposed to “call origination”, refers to the type calls of which the receiving party is a mobile.

3.2 Paging Schemes for Location Management

The performance of a paging scheme can be characterized by different measures, including:

- Paging delay: the delay calculated from the time a call termination request arrives at the base station (BSC) till the time a reply is received by the BSC from the mobile containing its cell location. Paging delay directly affects the caller's experience and it should be kept as low as possible. If the paging delay is too long, the caller party may hang up.
- Success rate: the likelihood of paging the precise cell where the mobile recipient of the call is located. If not all cells are paged, a paging message may not be reach the mobile. Clearly, a high success rate is desired.
- Cost of paging: the total number of cells paged for each established call.
- Cost of location management: the amount of resources required to perform location management. Relevant resources include communication bandwidth, data storage, CPU cycles, etc.

Let us examine paging delay in more detail. Paging delay is determined by the number of rounds of paging performed before a mobile is successfully located. Recall that paging messages need to be sent within designated slots in a paging cycle, with each cycle 5.12 seconds long in most modern cellular systems. Assume there is zero delay between the time a paging request is received by the mobile and the time its reply arrives at the BSC. Then, if a mobile is successfully located with the first page, the expected paging delay is half of a paging cycle, namely 2.56 seconds. Otherwise, each additional page adds 5.12 seconds to the paging delay.

For broadcast paging, the delay is fixed at half of a paging cycle (on average) if the mobile is in the paging area and has signal quality good enough to decode the page. Otherwise, subsequent broadcast page(s) may be sent until the mobile is located or the maximum number of paging rounds has been reached.

For profile-based paging, the first round of paging is sent to a set of cells derived from the mobility profile of the target. If no reply to that round is received by the BSC, a regular broadcast paging is carried out. Therefore, the paging delay with profile-based paging is larger than with regular broadcast paging. Furthermore, the lower the success rate of profile-based paging, the longer the expected paging delay.

In this paper, we consider two families of profile-base paging strategies:

- Fixed-profile based: A profile containing a list of locations or cells is pre-computed using on mobility data. The profile remains unchanged for a relatively-long period of time, at which point it “expires” and a new profile is created.
- Dynamic-profile based: A list of cells is dynamically adjusted to track a mobile user's activities, with cells that have not been visited in a long time possibly expiring from the list. Profiles based on that list are then created and updated dynamically as well.

In both cases, the profile is created based on the location list but it may not be the same as the location list. For example, heuristics can be applied to select most-likely cells from the location list and include only those cells in the profile.

3.3 Per Call Measurement Data (PCMD)

In this paper, we examine Per Call Measurement Data (PCMD) traces from an operational nationwide CDMA2000 network which supports voice, data, and SMS services. Several traces were collected simultaneously at Base Station Controllers (BSC) during February 2006. The collection function was provided by the equipment deployed in the network. Each trace contains records of calls established through the BSC. Each record includes numerous fields about the call event itself, the mobile device, the base station, and the BSC. The fields of interest in this paper are:

- Call starting time: measured with a granularity of 100 ms.
- Call duration: measured with a granularity of 1 ms.
- Mobile identification number (MIN): globally-unique identifier of the mobile involved in the call.
- Initial cell: cell number where the mobile is when the call is established.
- Final cell: cell number where the mobile is when the call is completed.
- Service type: identifies a voice call, a data call, or a SMS.
- Call direction: identifies incoming (mobile terminated) or outgoing (mobile originated) call.
- Number of pages: for incoming calls only, records how many times the mobile was paged for a call. If a mobile is already in a call at the time that call is made, or the mobile just finished another call (depending on the network configuration, within the last 20 seconds), the mobile will not be paged and the call setup request is sent to the last cell where the mobile was seen.

There are other fields in the call record database, such as the Mobile Serial Number (MSN) which identifies the device, the caller information for a mobile-terminated call, the sequence of dialed digits in case of a mobile-originated call, whether the call is completed or dropped, causes for a dropped call, etc., all of which might be of interest to other studies but are not considered in this paper.

We examined the mobility patterns of cellular users on three month-long traces from three BSCs: Philadelphia, Brisbane in the San Francisco Bay area, and Manhattan in New York City. All those traces were collected between Feb. 2 and Feb. 28, 2006. Table 1 shows high-level statistics of the traces, identified by the corresponding BSC names. Note that the Manhattan trace has twice as many users as the other two traces. For reasons of space, we describe in this paper the results obtained from the Brisbane and Manhattan traces only[2].

[2]The results from the Philadelphia trace are consistent with those from the two traces above.

Table 1: Statistics of three PCMD traces

BSC	Nb. of records	Nb. cells	Nb. users
Manhattan	120×10^6	139	1061×10^3
Philadelphia	140×10^6	150	543×10^3
Brisbane	50×10^6	144	404×10^3

4. ANALYZING MOBILITY AND CALLING PATTERNS WITH PCMD

We mentioned earlier in Section 3 that PCDM traces include a wide range of variables that characterize the calling and mobility patterns of users. In this section, we analyze a subset of those variables that are relevant to paging performance. We describe results, such as those in in Sections 4.1 and 4.2, which are not necessarily new, but which are interesting nevertheless because they confirm and tie in well with those obtained by others, for example in [5, 34].

4.1 Statistics of user mobility and activity

We consider first the number of cells visited by users in a month. Figure 1 shows the cumulative distribution function (CDF) of the number of cells visited by a user in a month. We observe a Pareto-type behavior in which most users visit a small number of cells. For example, 96% of the users in Manhattan or in Brisbane visit fewer than 40 cells.

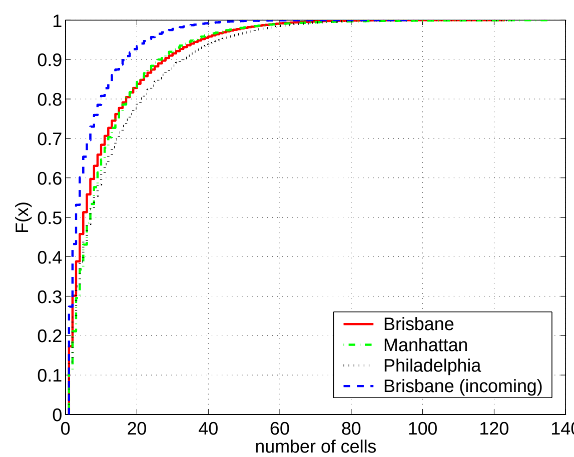


Figure 1: CDF of the number of cells visited in a month.

The impact of this result on paging is clear. Assume a paging area of 120 cells (consistent with Table 1) and a profile-based paging scheme where the mobility profile for a user includes all the cells visited by that user in the previous month. Then the result above shows the for 96% of the users, the cost of profile-based paging will be only 1/3 the cost of broadcast paging. However, it remains to be determined how many calls are made to those 96% of users and how successful the profile-based paging really is.

We observe a Pareto-type behavior similar to that above when considering calling patterns. For example, we find that 60% of the users call fewer than 26 times, which amounts to less than one call per day.

Next, we examine joint calling and mobility patterns, for example whether highly mobile users make more or fewer calls than static users. Figure 2 shows the CDF of the num-

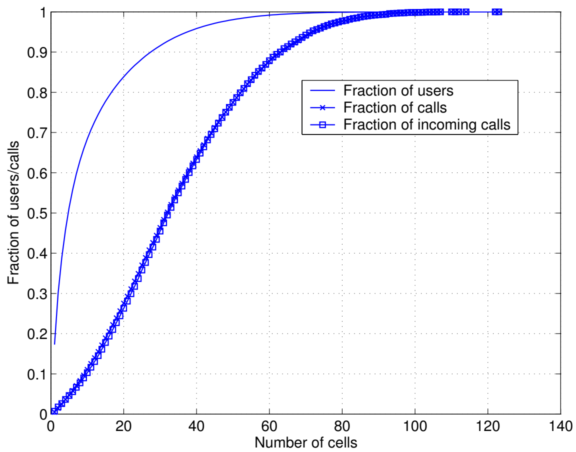


Figure 2: CDF of calls made/received by users ranked by their mobility.

ber of calls made or received by users ranked by their mobility, specifically by the number of different cells they visit in a month. We observe that although 96% of the users visit fewer than 40 cells, those users make or receive only 75% of the total number of calls.

Therefore, highly mobile users tend to make and receive more calls than static users. This fact makes paging challenging since a higher paging cost (to locate highly mobile users) may have to be paid frequently (since those highly mobile users need to be paged often as they make a large fraction of the calls). Furthermore, an expected high paging success rate for static users (since they do not move much) might not translate into paging resource savings since those users less frequently make or receive calls.

In summary, we make the following observations:

- Most users visit a small number of cells: 80% of the users visit fewer than 20 cells, and 96% of the users visit fewer than 40 cells.
- More than half of the users make less than a call per day.
- The 4% most mobile users make a disproportionately large number of calls, namely 25% of the total number of calls.

4.2 Dominant type of activity: voice, data, SMS

Table 2 shows a breakdown of the number of calls in Brisbane for each service type, namely voice, data, and SMS. Note that the direction (incoming or outgoing) of a small fraction (about 3%) of calls cannot be determined because these calls were handed over from another Mobile Switching Center (MSC).

Table 2: Call break-down for the Brisbane trace.

Type	Total calls	Incoming calls	Paged incoming calls
Voice	25.5×10^6	9.8×10^6	9.1×10^6
Data	16.7×10^6	4.0×10^6	3.8×10^6
SMS	5.6×10^6	3.4×10^6	2.9×10^6
Other	2.5×10^6	2.3×10^6	0.3×10^6

We observe that more than 1/3 of the voice calls and more than 1/2 of the SMS calls are are incoming calls. However, the majority of the data calls are outgoing, which is expected because most existing data applications in cellular networks are initiated by the mobile. In the future, as more applications are available, some of which may be initiated by another peer or an application server, we expect the fraction of incoming data calls to increase.

The breakdown data in Table 2 can be leveraged in the paging process because roughly two of every three calls can be used to update the location of a user before that user is paged for the third call. In the case of data calss, most of which are outgoing, a user's location is available and updated more frequently, leading us to anticipate a higher success rate with profile-based paging. We also note that more than half of the incoming calls are voice calls. Therefore, a good paging scheme must work well with voice calls and voice users.

Figure 3 shows the CDF of the number of calls of different types in the Brisbane trace. We find that the data calls have a much longer tail than the voice and SMS calls.

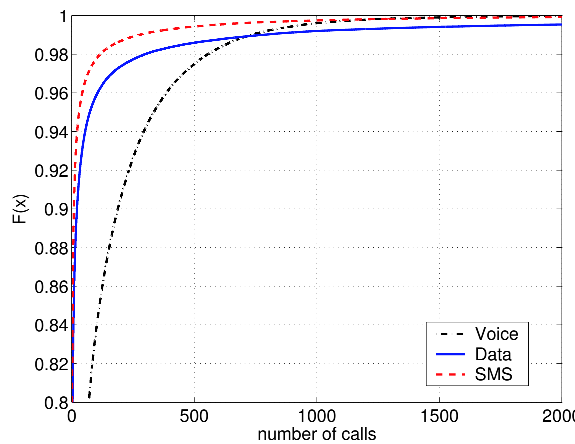


Figure 3: CDF of number of calls of each type in the Brisbane trace.

At this point, it is interesting to classify users based on their dominant service type. If more than half of the calls of a user u are of a certain service type X, say voice, we say that user u is a X-dominant user (say, a voice-dominant user) and call X the dominant service type of user u. If a user does not have a dominant service type, we classify her a "balanced user". We found in all 3 traces that the vast majority of users are voice-dominant users, followed by a smaller fraction of data-dominant users. SMS-dominant users are next, except, maybe surprisingly, in Manhattan where balanced users are more numerous than SMS-dominant users.

We examine dominance in more detail in Figure 4. We rank users by increasing number of calls and divide them evenly into 1000 groups; for each group, we compute the fraction of voice/data/SMS-dominant users and plot the fractions against the highest number of calls made or received by users in this group. We observe that in groups with fewer than 300 calls, 80% users are voice-dominant users. However, in groups with more than 300 calls, the fraction of voice-dominant users quickly drops and the fraction of the data-dominant users increases dramatically. The fraction of

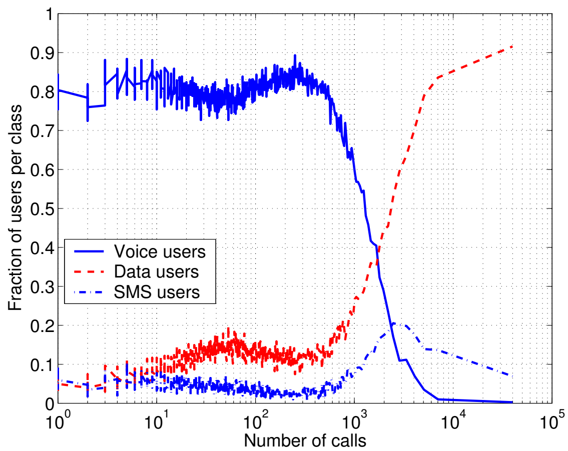


Figure 4: Dominant users in the Brisbane trace.

SMS-dominant users increases until about 3000 calls then drops.

Users with more than 2000 calls are mostly data-dominant users. These high-volume or highly active users frequently report their locations (each time they engage in a call) which means that we can estimate their location accurately. Therefore, we expect the success rate of profile-based paging for these users to be higher than that of the other, low activity, users.

We summarize the number of calls made or received by each category of users in Table 3. We see that voice-dominant users have the most calls in each type, including data calls. Therefore, it is important to optimize the paging performance for this group of users although optimizing for other groups, e.g., data-dominant users, might be easier.

Table 3: Number of calls of different types (voice, data, SMS) for different categories of users (voice, data and SMS-dominant users) in the Brisbane trace.

	Voice dominant	Data dominant	SMS dominant	Balanced
Voice	321113	27656	12301	16653
Data	51435	41640	3891	8853
SMS	76940	13085	19492	11653

In summary, we make the following observations regarding services and users:

- The majority of incoming calls are voice calls.
- We expect a high paging success rate for data calls and data-dominant users because 1) more data calls are outgoing than incoming and 2) high-volume users tend to be data-dominant.

4.3 Correlating the locations of successive calls

Next, we examine whether or not knowing the cells visited by a user in the past can help us determine the current location of that user. We do this by computing the entropy of the locations (cells) visited by each each user and the conditional entropy of a user's location given its previous N locations.

Let us consider first the case $N = 1$ and $N = 2$. Let X be the random variable representing a user's locations. If user u has had M calls over a total of K locations, the time series of user u's locations can be represented by a vector $L = (l_0, l_1, \ldots, l_{M-1})$ where l_i $(0 \leq i \leq M-1)$ denotes the location where call i occurred. There are K distinct locations in $\{l_i\}$. Assume each of these K locations appeared x_j times in L, $0 \leq j \leq K-1$. Therefore, $\sum_j x_j = M$. Then the probability of the user being in location l_i such that location l_i appeared x_j in L can be computed as x_j/M. Therefore the entropy of X is:

$$H(X) = \sum_{j=0}^{K-1} (x_j/M) log \frac{1}{x_j/M} \qquad (1)$$

Assume that $N = 1$. Let Y be the random variable for a user's next location given the immediately previous location X'. Y and X' have the same distribution as X when M is large enough. Recall that $L = (l_0, l_1, \ldots, l_{M-1})$ is the mobility vector, and let $Z = \{(l_i, l_{i+1}) : 0 \leq i \leq M-2\}$. Then the joint entropy of Y and X' is:

$$H(X', Y) = \sum_{(x',y) \in Z} P(x', y) log \frac{1}{P(x', y)}$$

where $P(x', y)$ is the number of times (x', y) appears in Z divided by the total number of elements in Z, which is $M-1$.

The conditional entropy of Y given X' is:

$$H(Y|X') = H(X', Y) - H(X') = H(X', Y) - H(X)$$

When $N = 2$, let X'' denote the random variable representing the distribution of the previous *two* locations. Thus, X'' is the joint variable (X', Y). Then the conditional entropy of Y given X'' is:

$$H(Y|X'') = H(X'', Y) - H(X'') = H(X'', Y) - H(X', Y)$$

The joint entropy $H(X'', Y)$ can be computed similarly and the details are omitted.

Figure 5 shows the CDF of the entropy and the conditional entropies, for $N = 1$ and $N = 2$, for all users in the Brisbane trace.

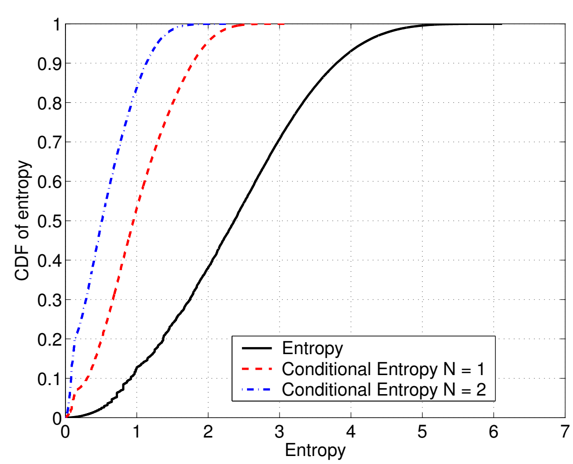


Figure 5: CDF of entropy and conditional entropies of user locations in the Brisbane trace.

Not surprisingly, we observe that the conditional entropy of location variable Y is much smaller than the original entropy of Y and that the conditional entropy with $N = 2$ is smaller than with $N = 1$. This means that the uncertainty about a user's location decreases when we know that user's previous location, and decreases further when its previous two locations are known.

Next, we show in Figure 6 the entropy and conditional entropy for users with different dominant service types. We only show the results for $N = 1$. Interestingly, the entropy distribution of data-dominant users is close to that of voice-dominant users. However, the conditional entropy of data-dominant users is much smaller than that of voice-dominant users and it is actually close to that of SMS users. SMS users have the smallest entropy and conditional entropy. Therefore, although data-dominant users can be as mobile as voice-dominant users, the locations of successive calls by those users are more highly correlated for voice-dominant users.

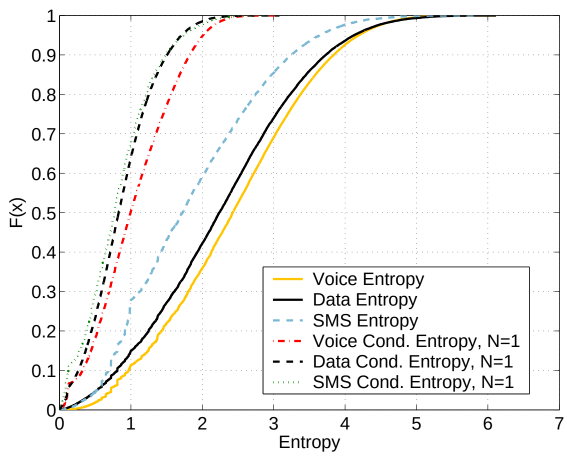


Figure 6: CDF of entropy and conditional entropy of user locations by dominant service type.

4.4 Correlating mobility patterns over time

In order to construct mobility profiles, we need to understand the correlation between the locations of successive calls (examined above in Section 4.3), but we also need to understand how mobility patterns evolve over time. We need to answer two specific questions, namely 1) how is historical data related to current data and 2) how much history data do we need to track to capture the full range of user mobility patterns over time. We examine these questions next.

First, we note that we can look at history data in terms of volume or age. We decide to look at age because previous studies have shown that call volume at switches demonstrates periodic patterns with periods of a day and a week [34]. Furthermore, WiFi users have been shown to demonstrate periodic mobility patterns over periods of a day and a week as well [30]. Therefore, we can safely assume that cellular users also exhibit this periodic mobility behavior. The study described below validates this assumption and at the same time answers the two questions above.

We quantify the correlation between two traces as follows. We take the trace from day x, compare it with the trace from day $x - n$, and vary n from 1 to a large number. We use *mutual information* to quantify the correlation. Specifically, the trace from day x contains a number of calling records, each of which can be represented by a (user, cell) pair. For a particular user u, we consider the mobility vector $T_x(u) = (a_u^0, a_u^1, \ldots, a_u^{c-1})$ where c is the total number of cells in the location area and a_u^i is the number of calls occurred at cell i $(0 \le i \le c-1)$ associated with user u on day x. We also have the vector for user u from day $x - n$, $T_{x-n}(u)$. We can then compute the entropy of the joint distribution of $T_x(u)$ and $T_{x-n}(u)$ over cells 0 through $c-1$. The mutual information of $T_x(u)$ and $T_{x-n}(u)$, $I(T_x(u), T_{x-n}(u))$, can be obtained via the joint entropy $H(T_x(u), T_{x-n}(u))$ and the entropies of $T_x(u)$ and $T_{x-n}(u)$ as follows:

$$I(T_x(u), T_{x-n}(u)) = H(T_x(u)) + H(T_{x-n}(u)) - H(T_x(u), T_{x-n}(u)) \quad (2)$$

where $H(\cdot)$ is computed using Eqn. (1). We define the *Normalized Mutual Information (NMI)* by

$$NMI(T_x(u), T_{x-n}(u)) = \frac{I(T_x(u), T_{x-n}(u))}{H(T_x(u)}.$$

For simplicity, we write $NMI(T_x(u), T_{x-n}(u))$ as $NMI_x^n(u)$.

Figure 7 shows the NMI for $x =$ Feb. 28 and $x =$ Feb. 26, for $n = 1$ through 27, in both the Brisbane trace and the Manhattan trace, averaged over all users on day x. We pick those two days because Feb. 28 is a weekday (Tuesday) while Feb. 26 is a weekend (Sunday).

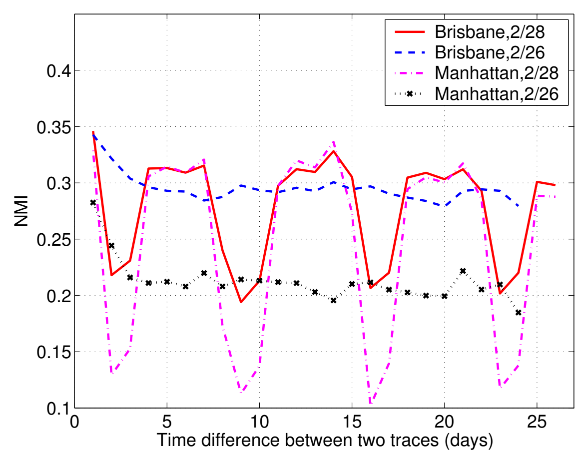


Figure 7: Average NMI vs. n between mobility vectors n days apart.

We make the following observations:

- A weekday trace has higher correlation with other weekday traces than with weekend traces. The peak correlation is obtained for values of n multiples of 7, *i.e.*, $n = 7, 14, 21, \ldots$
- A weekend trace also has slightly higher correlation with weekend traces than with weekday traces. However, this difference is less significant than that observed with weekday traces.
- As the time differences between two traces (n) increases, the amount of correlation, not surprisingly, decreases. However, this decrease is fast at the beginning (around one day) then slows down.

To better understand how many days should be considered in constructing mobility profiles, we examine the change in the NMI when we make more historical data available. Instead of using solely the data from day $x-n$, we aggregate the data from day $x-1$ through day $x-n$. The resulting mobility vector is thus $\sum_{i=1}^{n} T_x(u-i)$. Then, we consider the mutual information between $T_x(u)$ and $\sum_{i=1}^{n} T_x(u-i)$ normalized by $H(T_x(u))$. We show the results in Fig. 8.

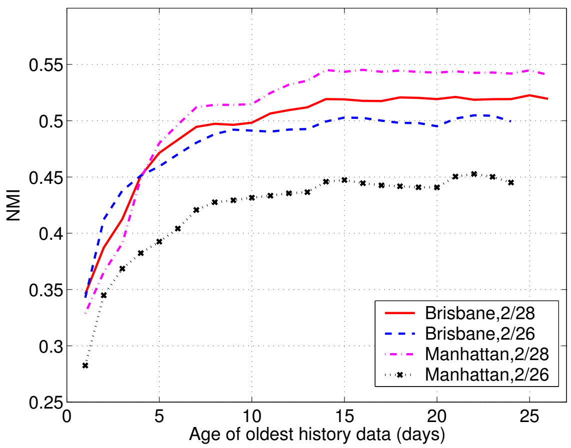


Figure 8: Average NMI vs. n between point and cumulative mobility vectors.

It is quite clear that the NMI increases until about $n = 14$, when it hits a plateau and stabilizes. This means that adding more data beyond 14 days does not help (but does not hurt either) in buiding mobility profiles. This is an important observation that we will use later to design paging profiles.

5. EVALUATING PROFILE-BASED PAGING SCHEMES

The analysis in Section 4 suggests that we should construct a the mobility profile of a user based on two characteristics of that user, namely 1) the locations of that user during recent calls and 2) the pattern of calls made by that user in the previous $n = 14$ days. The first characteristic leads us to consider dynamically updated profiles. The second characteristic is applicable to both static and dynamic profiles. Next, we study the performance of profile-based paging schemes with both fixed and dynamic profiles.

5.1 Profile-Based Paging with Fixed Profile

Recall that profile-based paging includes two steps. The first step involves paging only the cells in a user's profile. The second step is optional and is performed only if and after the first step fails, and it involves broadcasting the page to the entire paging area. The overall paging delay and paging cost are computed over the combination of these two steps. The paging delay is the delay between a paging request arriving at the BSC and the time a paging message is sent and received by the mobile successfully (either in the first or the second step). The paging cost is the total number of cells paged in both the first step and the optional second step.

The delay from the time a paging request arrives at the BSC till the time a paging message is sent is on average equal to half of a paging cycle. If the first paging fails, a subsequent page can only be sent after an entire paging cycle. Therefore, the average paging delay can be computed in terms of the number of paging cycles as:

$$0.5 + (1 - R_{succ}), \qquad (3)$$

where R_{succ} denotes the success rate of the first step.

We study the performance of fixed-profile-based paging using 14-day profiles on both the Brisbane trace and the Manhattan trace. The results are shown in Fig. 9(a) through Fig. 9(c). We built per-user profiles using the data from Feb. 2 through Feb. 15 and simulated the paging activity using the data from the 13-day period from Feb. 16 through Feb. 28. In each figure, the x-axis represents the number of days from the time the profile is created, i.e., the number of days elapsed between the simulated day and Feb. 15. The y-axis shows paging success rate in (a), paging delay in (b) and paging cost in (c).

Considering first the paging success rate in Fig. 9(a), we observe two clear dips, with a sharper dip for the Manhattan trace. By examining the dates on the x-axis, we find that the two dips correspond to two weekends. This makes sense and ties in with our earlier observation of lower NMI between weekday traces and weekend traces (Fig. 7) and lower NMI also between aggregated history data and a weekend trace (Fig. 8).

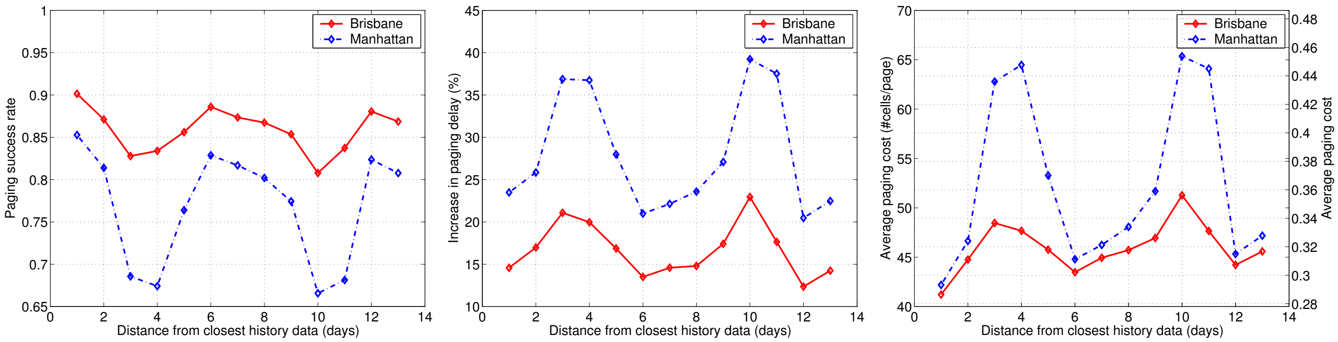
We next examine the paging delay, and we use as reference the delay of broadcast paging. Assume that all mobiles being paged are inside the paging area under study. Then, a broadcast paging scheme always has a paging delay of 0.5 paging cycle and a paging cost equivalent to the total number of cells in the paging area. Figure 9(b) shows the supplemental paging delay incurred by profile-based paging (with a fixed profile). Figure 9(c) shows both the number of cells paged (left axis) and the fraction of cells paged over the entire location area (right axis).

We summarize the results as follows:

- The paging performance with fixed profiles remains relatively stable over a period of two weeks.
- The success rate for weekday paging is about 85% in the Brisbane trace and 80% in the Manhattan trace. The success rate for weekend paging is about 80% to 85% in the Brisbane trace and only 65% to 70% in the Manhattan trace.
- In the Manhattan area, the fixed profile-based paging can incur a paging delay increase from 20% up to 40%. In the Brisbane area, the increase is only 10% to 20%.
- The fraction of cells paged in the Manhattan area is 45% through 65% while in the Brisbane area 40% to 50%. Therefore, on average, this fixed profile-based paging can save about 50% of the paging cost of broadcast paging.

5.2 Profile-Based Paging with Dynamic Profile

One way to improve the performance of fixed profile-based paging might be to build several fixed profiles. For example, we saw earlier that mobility patterns differ significantly between weekdays and weekends. One possibility then would



(a) Success rate (b) Paging delay (c) Paging cost

Figure 9: Profile-based paging with fixed profile - Brisbane and Manhattan traces

be to build two fixed profiles, one for weekdays and one for the weekend. We do not pursue this approach in the paper, but intend to do so in the future. In this section, we consider paging schemes with dynamic profiles. By "dynamic" we mean that for each paged or un-paged call associated with a user, the location of the call will be used to update the profile on the fly. We fix a start day and vary the number of days in history that are used to create the dynamic profiles. We pick two start days in particular, Feb. 28 which is a weekday, and Feb. 26 which is a weekend.

Figure 10 shows the average paging success rate for those days for both traces. Again, we find that the success rate in the Brisbane trace is higher than that in the Manhattan trace. The paging success rate increases with the number of days in history, faster when the number of days in history is small and slower afterwards.

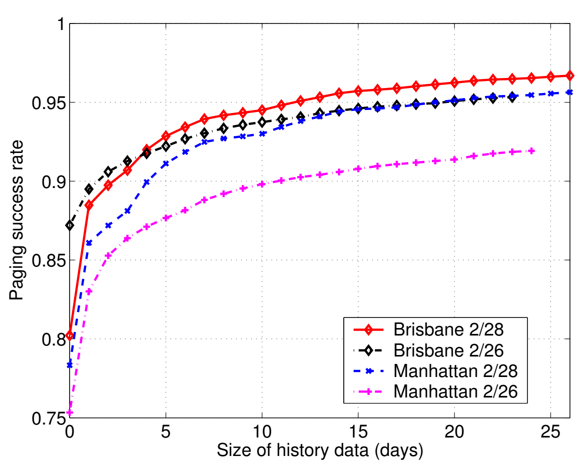


Figure 10: Dynamic profile: Average success rate.

We next compare the paging success rate with dynamic and with fixed profiles. The paging success rate for weekday Feb. 28 in Brisbane area with 14-day *dynamic* profiles is about 96% while it is about 90% with fixed 14-day profiles created *one day* earlier (Fig. 9(a)). For the weekend of Feb. 26 in the Manhattan area, the success rate with dynamic profile is about 91% at 14 days while it is only 68% with a fixed profile created *three days* before. Clearly, dynamic profiles can significantly improve the paging success rate especially when the data has less correlation with history data, i.e., on weekends.

We pick Feb. 28 for the Brisbane trace and Feb. 26 for the Manhattan trace to further investigate the paging performance for each service type (Fig. 11) and the trade-off between paging cost and paging delay (Fig. 12).

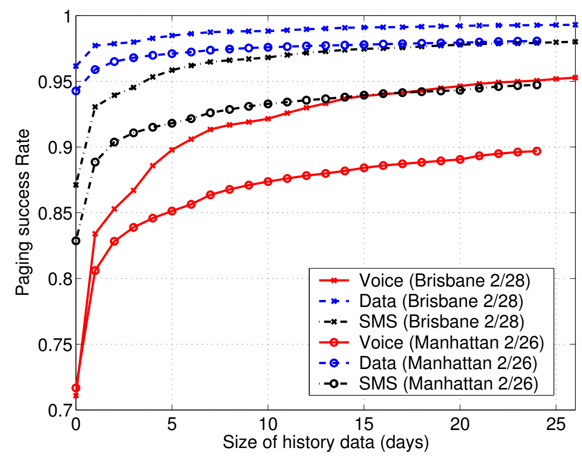


Figure 11: Dynamic profile: Success rate per service type.

The paging success rate for data service is much higher than the success rate for voice or SMS services, both for the weekday (Feb. 28) and for the weekend (Feb. 26). Even without any history data ($x = 0$ in Fig. 10), the paging success rate for data is about 94% for Feb. 26 in Manhattan. This agrees with our expectation that data calls are easier to locate (refer to Section 4). From Fig. 12, we see that with 14 days of history, the excess paging delay is less than 10% for Feb. 28 in Brisbane and less than 20% for Feb. 26 in Manhattan, with an average of 35 cells involved for each page. Both paging delay and paging cost are significantly lower with dynamic profiles than with fixed profile-based paging.

5.3 Smart Paging

As we have seen above, dynamic profile-based paging can achieve superior paging performance in terms of paging delay and paging cost compared to fixed profile-based paging. However, the paging cost associated with dynamic paging is

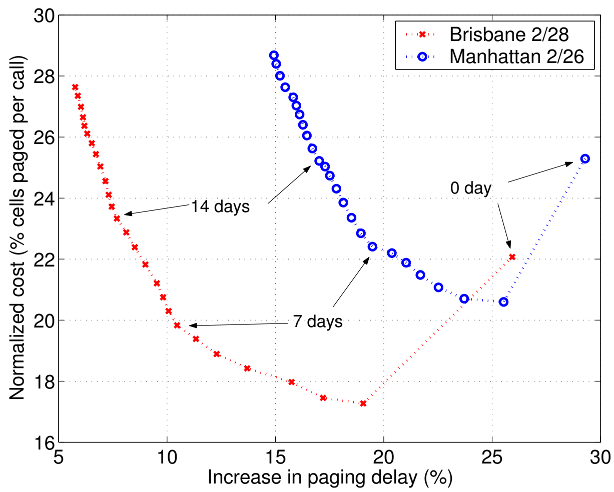


Figure 12: Dynamic profile: Paging cost vs. delay.

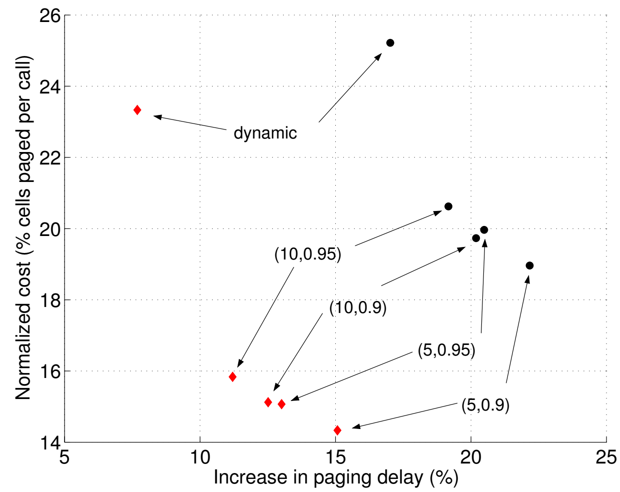


Figure 13: Performance of smart paging scheme - diamonds for Brisbane Feb 28 trace, dots for Manhattan 2/26 trace.

still quite high, namely about 25% of the total number of cells in the paging area with a 14-day profile.

There are various ways to optimize the paging process to further reduce paging cost and delay. One way would be to identify users with very low success rate under profile-based paging and always broadcast paging message to those users without paging the cells in the profile. Because the success rate of these users is low, the (broadcast) paging cost would not increase much (compared to the cost of profile-based paging), yet paging delay would likely drop significantly.

Another way to optimize the paging process is to try to "guess" within the profile the most likely locations of the user and page these cells only in the first phase. We describe and examine one such scheme next. We refer to that scheme as the *smart paging* scheme.

Recall that data pages have a high success rate compared to voice or SMS pages. Furthermore, we found that, on average, and for both the Brisbane and Manhattan traces, over 90% of the calls are associated with only 40% of all the cells the user visited over a month. We use these two results to design the smart paging scheme as follows:

- For data page requests, page the *last N most recently* cells visited by the user.
- For other page requests, page both the N most recently visited cells *and* the top X fraction of most visited cells by the user.

Both N and X are configurable parameters. We vary N between 5 and 10 and X between 90% and 95% and examine the performance of the smart paging scheme. We used both the Brisbane trace to page the users on Feb. 28 and the Manhattan trace to page the users on Feb. 26. Figure 13 shows the delay and cost of our scheme and compares it to the delay and cost of the dynamic profile scheme of Figure 12 (using a 14-day history).

The results from the dynamic paging scheme are labeled "dynamic"; the results from the smart paging scheme are labeled with the appropriate (N, X) pairs. We observe that with $N = 10$ and $X = 0.95$, the smart paging scheme provides both very low paging cost at the price of only a small increase in paging delay. Specifically, its paging cost is 30% lower in Brisbane on Feb. 28 (compared to dynamic paging) and 17% lower in Manhattan on Feb. 26. In practice, operators can configure N and X to achieve desired cost and delay trade-offs.

6. CONCLUSION

Locating mobile users efficiently and quickly is a critical operation in cellular networks. In this paper, we have developed dynamic profile-based paging schemes which significantly increase the effectiveness of the location management process, with average paging success rates across voice/data/SMS calls above 90% in Brisbane and 85% in Manhattan (see Figure 10 above) and paging success rates for specific applications such as data calls above 95% in both locations (Figure 10). Deploying such paging schemes would correspondingly reduce signaling load by 85% or 90%, at a cost of a slight increase in paging delay. Furthermore, we observed earlier that an interesting benefit of such schemes is to increase the robustness of cellular networks to signaling DoS attacks by increasing the available capacity of signaling channels, and thus by increasing the intensity required for an attack to disrupt the network.

The approach we take in the paper is to design and evaluate the performance of paging schemes using call record data about mobile users of a large cellular operator. Specifically, we analyze more than 300 million call records collected in three US metropolitan areas to characterize the calling activity and mobility patterns of 2 million users in more than 400 cells. We use this data to build mobility profiles of those users and to develop and validate the performance of our static and dynamic profile-based paging schemes.

We are extending the work presented in this paper in several directions. One is to increase the flexibility of our paging schemes by incorporating additional state variables relevant to paging performance, for example network load. When the load is low, increasing the paging cost to reduce the delay might be acceptable, however when the load is high increasing the delay is more justifiable. Another is to quantify the specific benefits in terms of security (specifically increased robustness to paging attacks) of increased paging efficiency. Another one is to collect, store and analyze PCMD data for a complete, nationwide cellular network.

Acknowledgements

We wish to thank Ashwin Sridharan for help with PCMD collection and analysis. We also wish to thank the anonymous reviewers as well as Ravi Jain for their feedback and suggestions which improved the quality of the paper.

7. REFERENCES

[1] I. F. Akyildiz, J. S. M. Ho, and Y.-B. Lin, "Movement-based location update and selective paging for pcs networks," *IEEE/ACM Transactions on Networking*, vol. 4, no. 4, pp. 629–638, 1996.

[2] W. Enck, P. Traynor, P. McDaniel, and T. LaPorta, "Exploiting Open Functionality in SMS-Capable Cellular Networks," in *Proc. 12th ACM Conf. on Computer and Communications Security*, Alexandria, VA, Nov. 2005.

[3] J. Serror, H. Zang, and J. Bolot, "Impact of Paging Channel Overloads or Attacks on Cellular Networks," in *Proc. ACM Workshop Wireless Security WiSe'06*, Los Angeles, CA, Sept. 2006.

[4] P. Traynor, W. Enck, P. McDaniel, and T. L. Porta, "Mitigating attacks on open functionality in sms-capable cellular networks," in *Proc. ACM MobiCom '06*, 2006, pp. 182–193.

[5] E. Halepovic and C. Williamson, "Characterizing and modeling user mobility in a cellular data network," in *Proc. ACM Intl. Workshop on Modeling Analysis and Simulation of Wireless and Mobile Systems PE-WASUN'05*, 2005, pp. 71–78.

[6] A. Bar-Noy, I. Kessler, and M. Sidi, "Mobile users: to update or not to update?" *Wireless Networks*, vol. 1, no. 2, pp. 175–185, 1995.

[7] I. F. Akyildiz and J. S. M. Ho, "Dynamic mobile user location update for wireless PCS networks," *Wireless Networks*, vol. 1, no. 2, pp. 187–196, 1995.

[8] C. Rose, "Minimizing the average cost of paging and registration: a timer-based method," *Wireless Networks*, vol. 2, no. 2, pp. 109–116, 1996.

[9] J. S. M. Ho and I. F. Akyildiz, "Mobile user location update and paging under delay constraints," *Wireless Networks*, vol. 1, no. 4, pp. 413–425, 1995.

[10] M. Verkama, "A simple implementation of distance-based location updates," in *IEEE 6th Intl Conf. Universal Personal Communications Record*, San Diego, CA, USA, 1997, pp. 163–167.

[11] S. K. Sen, A. Bhattacharya, and S. K. Das, "A selective location update strategy for PCS users," *Wireless Networks*, vol. 5, no. 5, pp. 313–326, 1999.

[12] H. Wu, M.-H. Jin, and J.-T. Horn, "Personal paging area design based on mobile's moving behaviors," in *Proc. IEEE Infocom '01*, Anchorage, AK, Apr. 2001, pp. 21–30.

[13] Y. Xiao and K. Wu, "Location update for PCS networks with a fractional movement threshold," in *Proc. Intl. Conf. Dist. Computing Systems '03*, Washington, DC, 2003, p. 825.

[14] P. Mutaf and C. Castelluccia, "Hash-based paging and location update using bloom filters," *Mobile Networks and Applications*, vol. 9, pp. 627–631, 2004.

[15] C. K. Ng and H. W. Chan, "Enhanced distance-based location management of mobile communication systems using a cell coordinates approach," *IEEE Trans. Mobile Computing*, vol. 4, no. 1, pp. 41–55, 2005.

[16] A. Bar-Noy, Y. Feng, and M. Golin, "Paging mobile users efficiently and optimally," *Proc. IEEE Infocom 2007*, Anchorage, AK, Apr. 2007.

[17] "Cellular Radiotelecommunications Intersystem Operations," EIA/TIA IS.41.

[18] G. Wan and E. Lin, "A dynamic paging scheme for wireless communication systems," in *Proc. MobiCom '97*, 1997, pp. 195–203.

[19] B. Liang and Z. J. Haas, "Predictive distance-based mobility management for PCS networks," in *Proc. IEEE Infocom'99*, New York, NY, pp. 1377–84.

[20] H.-W. Hwang, M.-F. Chang, and C.-C. Tseng, "A direction-based location update scheme with a line-paging strategy for PCS networks," in *IEEE Communications Letters*, 2000, pp. 149–151.

[21] C. Rose, "State-based paging/registration: a greedy technique," *IEEE Transactions on Vehicular Technology*, no. 1, pp. 166–173, 1999.

[22] G. P. Pollini and C.-L. I, "A profile-based location strategy and its performance," *IEEE Journal on Selected Areas in Communications*, vol. 15, no. 8, pp. 1415–1424, 1997.

[23] P. G. Escalle, V. C. Giner, and J. M. Olta, "Reducing location update and paging costs in a pcs network," *IEEE Transactions Wireless Communications*, vol. 1, no. 1, Jan. 2002.

[24] J. Zhang and L. Gruenwald, "Spatial and temporal aware, trajectory mobility profile based location management for mobile computing," *LNCS - Proc. Intl. Database and Expert Systems Applications (DEXA) Workshop.*, pp. 716–720, Sept. 2002.

[25] R. Jain, Y.-B. Lin, C. Lo, and S. Mohan, "A caching strategy to reduce network impacts of pcs," *IEEE Journal on Selected Areas in Communications*, vol. 12, no. 8, Oct. 1994.

[26] R.-H. Gau and Z. J. Haas, "Concurrent search of mobile users in cellular networks," *IEEE/ACM Transactions in Networking*, vol. 12, no. 1, pp. 117–130, 2004.

[27] F. Baccelli and S. Zuyev, "Poisson-voronoi spanning trees with applications to the optimization of communication networks," *Operations Research*, vol. 47, no. 4, 1999.

[28] G. Y. Lee and Y. Lee, "Numerical analysis of optimum timer value for time-based location registration," *IEEE Communications Letters*, vol. 6, no. 10, pp. 431–433, 2002.

[29] M. Kim, D. Kotz, and S. Kim, "Extracting a mobility model from real user traces," *Proc. IEEE Infocom'06*, April 2006.

[30] M. Kim and D. Kotz, "Periodic properties of user mobility and access-point popularity," *Journal of Personal and Ubiquitous Computing*, vol. 11, no. 6, August 2007.

[31] F. Bai, N. Sadagopan, B. Krishnamachari, and A. Helmy, "Modeling path duration distributions in

Manets and their impact on routing performance," in *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 7, Sept. 2004.

[32] A. Chaintreau and et al., "Impact of human mobility on the design of opportunistic forwarding algorithms," in *Proc. IEEE Infocom'06*, Barcelona, Apr. 2006.

[33] A. Jardosh, E. Belding-Royer, K. Almeroth, and S. Suri, "Towards realistic mobility models for mobile ad hoc networks," in *Proc. ACM Mobicom'03*, San Diego, CA, Sept. 2003.

[34] C. Williamson, E. Halepovic, H. Sun, and Y. Wu, "Characterization of CDMA2000 cellular data network traffic," in *Proc IEEE Conf. Local Computer Networks*, Washington, DC, USA, 2005, pp. 712–719.

[35] "Mobile radio interface layer 3 specifications," 3GPP GSM 04.08, Version 7.8.0, Oct. 2000.

[36] "Upper layer (layer 3) signaling standard for CDMA2000 spread spectrum systems, release D," 3GPP2 C.S0005-D, Version 1.0, Feb. 2004.