IEEE Transactions on Vehicular Technology
v. 46, n. 2, pp. 269–278, 1997

Mobility Management for Personal Communications Systems

Timothy X Brown[1]
Seshadri Mohan[2]
[1]Department of Electrical and Computer Engineering
Interdisciplinary Telecommunication Program
University of Colorado
Boulder, CO, 80309-0530
`timxb@colorado.edu`
[2]Bell Communications Research
Morristown, NJ 07960
`smohan@bellcore.com`

Abstract

Using a combination of empirical data and theoretical models, this paper develops a model of user behavior for a personal communications system environment. This model is used to analyze a mobility management strategy that combines automatic updates by the users—either when they make significant moves or when they go extended periods without network interaction; multiple hysteresis in the form of dynamic registration areas and delayed updates; and a focused paging strategy that minimizes the number of pages given a constraint on the time allowed to page. Over a range of system and user parameters, the total paging and update traffic can be kept below 1 per 2000 user seconds. The impact on the user's handset is less than 10 brief updates per day. The total traffic is only a factor of three more than the minimum, immobile users case.

1 INTRODUCTION

Mobility management is the process of keeping track of and locating users so that calls arriving for them can be directed to their current location. For wireless systems it consists of two parts: users notifying the system of their current location (*updating*), triggered at certain events; and the system notifying the user that a call has arrived for them (*paging*).

Personal communications systems (PCS), such as those based on Universal Digital Portable Communications (UDPC) [Cox87], will provide wireless access to pedestrians over widespread areas using a dense grid of small low-power base stations mounted on utility poles and buildings. PCS differs from conventional cellular. Users will be more main stream, expecting service quality equivalent to conventional wireline service at costs approaching wireline. The number of users and their density will be much greater with much greater total call traffic than with today's cellular users. Cells will be smaller and denser to match the greater user and traffic density. Handsets will be smaller, and lighter with smaller batteries. The handset will be used indoors more than outdoors.

As a result: care must be taken to minimize the updating and paging traffic, micro-movements of users around a given home or work site in addition to the macro-movements by car and public transport become important, radio signal variability blurs the boundaries between the many small cells, delays should be small, and battery usage should be minimized.

This paper describes an empirically-based model of the system, handset, and users that captures these features, and a combination paging and location updating strategy optimized for PCS constraints. Applying the strategy to the model the paper shows that with minimal system impact, the paging and update traffic are a factor of 3.5 larger than the stationary-user lower bound. Simple parameter variation analysis, gives interesting insights into the structure of the traffic. This analysis is also applicable to dense cellular environments where cell sizes approach that in PCS. The analysis is impossible with simpler models described in the next section.

2 BACKGROUND

A widely used terminal/user mobility model is the fluid flow model [Tho88]. This model has been used to analyze the signaling system performance of the GSM system for PCS [Mei92], and study the impact of voice-based PCS on network databases [Lo92], [Moh94]. The model assumes that PCS users carrying terminals are moving at an average velocity of v and their direction of movement is uniformly distributed over $[0, 2\pi]$. Assuming that the PCS users are uniformly populated with a density of ρ and the registration area boundary is of length L, it has been shown that the rate of registration area crossing, C, is given by

$$C = \frac{\rho v L}{\pi}. \tag{1}$$

The simulation studies in [Ses92] show close agreement to the above model for boundary crossing rate for a Manhattan-style layout of streets.

The model in [Hon86] assumes that the speed and direction of a mobile terminal are independent random variables and that when they cross cell boundaries their speed and direction are independently and randomly regenerated, the direction being uniformly distributed in the range $[-\frac{\pi}{2}, \frac{\pi}{2}]$ with respect to the tangent at the point of boundary crossing. This model is then used to study some handoff procedures.

The model in [Gue87] is similar to that in [Hon86], with the difference the speed and direction are regenerated independently and randomly not at the time of boundary crossing but after a random time interval. The model is then used to calculate the channel occupancy time distribution.

The above three models do not take into account the fact that user mobility characteristics derive from movement on two different scales using distinct modes: the users are likely at the micro-scale to move about at low speeds within their cells and at the macro-scale move at vehicular speeds within and between the cells. Consequently, this paper studies a multi-scale mobility model, a combination of a random walk model suitable for the micro scale and a big move model suitable for the macro scale.

3 MODEL ASSUMPTIONS

We consider a system with a dense grid of base stations, such as UDPC [Cox87], targeted primarily at pedestrian users. Thus, we don’t consider effects of users communicating while in transit at vehicular speeds, although the model could be extended. In addition we are concentrating on paging and user updates, and not handoffs or the traffic generated by the calls themselves. As a result, we focus on calls and movement as events ignoring their duration and structure.

3.1 PCS SYSTEM MODEL

Users and the system are connected by radio links between base station *ports* and hand held radio *handsets*. The system is a large planar area divided into disjoint paging areas (PA). The PA are roughly circular with area $A = \pi R^2$. The ports can page users within the same PA with certainty. The ports regularly broadcast an (x, y) coordinate of their location in the system as part of their control signals. Frequency reuse techniques limit spectrum usage, (i.e. we only have to consider the bandwidth for paging traffic generated in one reuse area regardless of system size).

3.2 HANDSET MODEL

The handset is able to automatically call the system and identify itself and its current PA. It is able to use the (x, y) coordinates broadcast by the base stations to calculate the distance from the PA covering its current position to the PA covering the position where it last interacted with the system. It also has a clock which measures the time since the user last interacted with the system.

3.3 USER MODEL

The user calls out and receives incoming calls as Poisson processes with rates ρ_{CO} and ρ_{CI} respectively and cumulative density

$$\text{Prob}\{\text{event } X \text{ occurs in time interval } t\} = 1 - e^{-\rho_X t} \tag{2}$$

where X is either CO or CI. Occasionally the user cannot respond to a page request due to travelling at high speed, the handset turned off, etc. resulting in a given page being ignored and having no response with probability p_{NR}.

The user mobility consists of micro and macro movements (see Figure 1). In general, at the micro level, the user is moving around in the vicinity of a fixed location such as a home, work site, or shopping mall. This is well modeled by a random walk, s.t. given a user at the origin at $t = 0$:

$$\text{Prob}\{\text{user distance to origin} < x \text{ at time } t\} = 1 - e^{\frac{-x^2}{2\alpha t}} \tag{3}$$

where α is the random walk speed parameter s.t. if the user makes S size steps every T units of time $\alpha = S^2/T$ [Dog90].

Motivated by empirical data described in Appendix A, the user occasionally makes very large, macro moves, e.g. via automobile. The moves are radially symmetric and exponentially distributed:

$$\text{Prob}\{\text{distance of a big move} < d\} = F_{BM}(d) = 1 - e^{-\gamma d} \tag{4}$$

where $1/\gamma$ is the average distance moved. The big moves occur as a Poisson process with arrival rate ρ_{BM} and obey (2). Given vehicular speeds relative to the PA size, we assume that the big moves are instantaneous.

4 MOBILITY MANAGEMENT STRATEGY

Mobility management consists two components, one is the *updating* strategy whereby users notify the system of their position, and the other is the *paging* strategy for the system to find the users. Our strategy consists of four elements: automatic periodic updates, dynamic registration areas, delayed updates, and focused paging.

4.1 UPDATING

The user's handset is at all times registered to a particular PA. Whenever the user makes or receives a call it is assumed that enough information is automatically passed so that the user is effectively updated. To maintain an accurate location in the system's database, the handset updates if more than time T_{U} has passed without any interaction with the system. A similar mechanism exists in the UK's TACS [Bed93]. We assume that the user is uniformly distributed throughout the PA at update time, although in some instances for simplicity the user is considered at the center of the PA.

There exists a boundary distance, D, such that if the user ever moves more than D away from the center of the PA where the user most recently registered, then it automatically updates. The registered PA and all PA within D are the user's registration area (RA). In the method described here RAs are dynamically assigned to users and can be disjoint or overlap with another user's RA. This provides a hysteresis such that a user that leaves their RA and updates, creating a new RA, will not immediately re-update if it crosses back into the old RA. With fixed RA, a user wandering near the border could produce many updates as they crisscross (either physically or due to spurious radio effects) the border. Also with fixed RA, update traffic is disproportionately distributed to PA on the border [Bau94]. The concept of dynamic registration area is, however, not new. In [Xie93], the authors propose and analyze a scheme in which the RA for a user is dynamically determined based on the rate of incoming calls to that user and the user's mobility rate. However, unlike the multi-scale mobility model proposed in our paper, the authors use the simple fluid flow model of [Tho88].

The user does not update as soon as they leave their RA, rather they check their position periodically with period T_{min}, and updates only then. Furthermore it updates if it detects itself in a new PA that is the same for two consecutive periods. This has two effects, it avoids generating a rapid stream of updates when the user is travelling at high speed. In particular, we assume that a big move generates at most one move no matter what the distance (see [Bau94] for a more detailed analysis). The second effect is that for fixed RA or when the RA is the same size as the PA (i.e. no hysteresis), it helps to minimize many updates from users random walking near the border.

4.2 PAGING

When a call request arrives for a user, instead of immediately paging all cells within the RA, the system initiates a paging sequence. A sequence of disjoint regions are defined and the PA in each region in turn page the user. this continues until the user responds, or all PA within the RA are paged. Symmetry and logic dictate that the regions are successively larger concentric rings around the position where the user last updated. Given the competing constraints of minimizing the time to find the user, and minimizing the number of PA paged, [Dog90] derives that for a user random walking for time t, at speed constant α, in PA of radius R, to minimize the average number of PA paged, P, given a constraint on the average number of different regions paged, N, yields areas of successive rings that are constant area and:

$$\bar{P}(t) = \bar{N}\ln\left(\frac{\bar{N}}{\bar{N}-1}\right)\frac{2\alpha t}{R^2} \tag{5}$$

So, P/N PA are paged at a time in successively larger concentric rings until all of the PA within D of the last place of update have been tried.

For typical parameters, (5) yields less than 1 PA paged per page. This then can be treated as a guide, and the following paging strategy is used. First, the PA where the user is registered is paged, then all adjacent PA's, then the next adjacent PA so that the successive rings match the limits of the PA sizes. This could be refined so that not all adjacent PA are paged at once, but for simplicity and as an upper bound, we will ignore this. If $\{r_0 = 0, r_1, r_2, \ldots, r_k = D\}$ are the successive ring sizes that are paged and p_i is the probability that the user is more than r_i away from the center at the time of the page, then the average number of cells paged, P, is:

$$\bar{P} = \sum_{i=1}^{k} p_{i-1}\frac{(r_i^2 - r_{i-1}^2)}{R^2} \tag{6}$$

5 TRAFFIC GENERATED

Combining the above models we calculate the following traffic components. Refer to Figure 2. Let b_{U} and b_{P} be the units of updates and PAs paged respectively and note that a sum of Poisson processes is another Poisson process with rate equal to the sum of the individual processes.

5.1 UPDATE TRAFFIC

Update traffic consists of three components, assuming that the automatic updates at calls in and out are free. Users that make a big move will, if they move further than D away from their current registered PA, trigger an update. Using (4):

$$\mathrm{Traffic}_{\mathrm{BM}} = (\text{Rate of BM}) \cdot \mathrm{Prob}\{\mathrm{BM} > D\} \cdot b_{\mathrm{U}} = \rho_{\mathrm{BM}} \cdot (1 - F_{\mathrm{BM}}(D)) \cdot b_{\mathrm{U}} \tag{7}$$

where we ignore the effect of the much smaller random walk component.

If the time since the last interaction with the system reaches T_U, then the handset initiates a time out update. For this we define the effective number of non-time out update events, $\lambda' = \rho_{CI}(1 - p_{NR}) + \rho_{CO} + \rho_{BM}(1 - F_{BM}(D))$ which is the sum of the rate of calls that the user responds to or initiates plus the rate of big moves that cross the RA border. This should also include the rate of random walk updates (below), but as this will be small compared to the other values it can be ignored. To count the number of time out updates, at each non-time-out update we look at the time, t, since the last non-time-out update, and count the number of integral periods of length T_U:

$$\begin{aligned} \text{Traffic}_{\text{TO}} &= (\text{Rate of other updates}) \cdot \sum_{i=0}^{\infty} i \text{Prob}\{ iT_{\text{U}} \le t < (i+1)T_{\text{U}} \} \cdot b_{\text{U}} \\ &= \lambda' \cdot \sum_{i=0}^{\infty} i\left(e^{-\lambda' i T_{\text{U}}} - e^{-\lambda'(i+1)T_{\text{U}}}\right) \cdot b_{\text{U}} \\ &= \frac{\lambda'}{e^{\lambda' T_{\text{U}}} - 1} \cdot b_{\text{U}} \end{aligned} \tag{8}$$

Users who make a random walk across a PA boundary trigger an update. If $p'_{RW}(z, R, D, t)$ is the probability that after time t a user initially uniformly distributed in a PA of radius R has made random walk more than z away from the PA center given that they will update if further than D away, and $p_{RW}(z, R, D, \lambda', T_U)$ is $p'_{RW}(z, R, D, t)$ averaged over inter-update times up to T_U, using Poisson update rate λ', then:

$$\begin{aligned} \text{Traffic}_{\text{RW}} &= (\text{Non-timeout update rate} \cdot p_{\text{RW}}(D, R, D, \lambda', T_{\text{U}}) + \text{Timout update rate} \cdot p'_{\text{RW}}(D, R, D, T_{\text{U}})) \cdot b_{\text{U}} \\ &= \left(\lambda' \cdot p_{\text{RW}}(D, R, D, \lambda', T_{\text{U}}) + \frac{\lambda'}{e^{\lambda' T_{\text{U}}} - 1} \cdot p'_{\text{RW}}(D, R, D, T_{\text{U}})\right) \cdot b_{\text{U}} \end{aligned} \tag{9}$$

This factor, p_{RW} and p'_{RW}, are solutions to a heat equation involving a Fourier series of Bessel functions and is derived in Appendix B. It turns out that it is typically small which has been assumed in this derivation. See Appendix B for a discussion if it is not small.

5.2 PAGING TRAFFIC

The paging traffic also consists of three components. If a call arrives, and the handset does not respond, then all of the PA within radius D are paged:

$$\begin{aligned} \text{Traffic}_{\text{P|NR}} &= (\text{Rate of Calls}) \cdot \text{Prob}\{\text{No Response}\} \cdot (\#\text{ PA within } D) \cdot b_{\text{P}} \\ &= \rho_{\text{CI}} \cdot p_{\text{NR}} \cdot \frac{D^2}{R^2} \cdot b_{\text{P}} \end{aligned} \tag{10}$$

If a big move is less than D, then the number of PA paged is a function of the size of the big move. We use (4) and average over all big moves less than D. In this case we need to know the probability of an undetected big move. We can determine the joint probability that the last move was a big move and not a time-out event (or any other detectable event) from the conditional of the event being a big move given it wasn't a time-out times the probability of it not being a time-out. We assume that the big move is much larger than any random walk component, and the probability of two consecutive big move events less than D is negligible:

$$\begin{aligned}
\text{Traffic}_{\text{P|BM}} &= (\text{Rate of Calls}) \cdot \text{Prob}\{\text{Response}\} \cdot \text{Prob}\{\text{Last Event not Timeout}\} \\
&\quad \cdot \text{Prob}\{\text{Last Event was BM} \mid \text{Last Event was not Timeout}\} \\
&\quad \cdot \text{Prob}\{\text{BM} < D\} \cdot (\text{Average \# Cells Paged }) \cdot b_{\text{P}} \\
&= \rho_{\text{CI}} \cdot (1 - p_{\text{NR}}) \cdot (1 - e^{-\lambda T_{\text{U}}}) \cdot \frac{\rho_{\text{BM}}}{\lambda} \cdot F_{\text{BM}}(D) \cdot \frac{\int_0^D r^2 \gamma e^{-\gamma r} dr}{R^2 F_{\text{BM}}(D)} \cdot b_{\text{P}} \qquad (11) \\
&= \rho_{\text{CI}} \cdot (1 - p_{\text{NR}}) \cdot (1 - e^{-\lambda T_{\text{U}}}) \cdot \frac{\rho_{\text{BM}}}{\lambda} \cdot \frac{2}{R^2}\left(\frac{1}{\gamma^2} - \left(\frac{D^2}{2} + \frac{D}{\gamma} + \frac{1}{\gamma^2}\right)e^{-\gamma D}\right) \cdot b_{\text{P}} \\
&\cong \rho_{\text{CI}} \cdot (1 - p_{\text{NR}}) \cdot (1 - e^{-\lambda T_{\text{U}}}) \cdot \frac{\rho_{\text{BM}}}{\lambda} \cdot \frac{\gamma D^3}{3R^2} b_{\text{P}} \qquad \gamma D << 1
\end{aligned}$$

where $\lambda = \rho_{\text{CI}}(1 - p_{\text{NR}}) + \rho_{\text{CO}} + \rho_{\text{BM}}$ is the rate of detectable events.

In the case of a user who has only random walked since the last update, we get:

$$\begin{aligned}
\text{Traffic}_{\text{P|RW}} &= (\text{Rate of Calls}) \cdot \text{Prob}\{\text{Response}\} \cdot \text{Prob}\{\text{Last Event not BM} < D\} \\
&\quad \cdot (\text{Average \# Cells Paged }) \cdot b_{\text{P}} \qquad (12) \\
&= \rho_{\text{CI}} \cdot (1 - p_{\text{NR}}) \cdot \left[1 - (1 - e^{-\lambda' T_{\text{U}}})\frac{\rho_{\text{BM}}}{\lambda} F_{\text{BM}}(D)\right] \cdot \bar{P}(R, D, \lambda', T_{\text{U}}) \cdot b_{\text{P}}
\end{aligned}$$

where $P(R, D, \lambda', T_{\text{U}})$ is defined in (6), using $r_i = (2i - 1)R$ (see next section), $r_k = D$, and from Appendix B, $p_i = p_{\text{RW}}(r_i, R, D, \lambda', T_{\text{U}})$.

6 OPTIMIZING PARAMETERS

To the system paging designer, we assume that the only parameters under their control are the paging boundary D, and the update period, T_{U}. Table 2 summarizes the traffic components and the direction change needed on these variables to reduce the traffic (single and double arrow indicate weak and strong desired change respectively). These indicate that the strongest components are for small D and perhaps also T_{U}, but we first consider system effects not modeled directly.

The size of the PAs are a given, so as shown in Figure 3, D must be $(2i - 1)R$ for some positive integer i. Further, due to multiple paths and shadowing in the complex radio environment, a user isn't always located in the PA that he registers[Ber88], and small movements can change the perceived location. To be robust to these variations, D must be at least $3R$ or $5R$ depending on a more precise analysis of these radio effects.

The utility of the portable increases greatly with the time it can be used without charging. The number of updates should be minimized implying T_{U} should be large. As will be seen in the following, this will not be a great constraint since the time out update time will be long, on the order of hours. Maximum utility dictates that T_{U} be as large as possible. An expression can be garnered for the T_{U} that minimizes the traffic as a function of the other parameters, but it is elaborate and not very insightful without making distorting simplifications, so we will instead find the T_{U} that minimizes the traffic numerically.

7 EXAMPLE RESULTS

7.1 ASSUMPTIONS

We choose the following values for model parameters. From Appendix A, we have the big move parameters, $\gamma = 6800\mathrm{m}$ and $\rho_{\mathrm{BM}} = 4.5/\mathrm{Day}$. For the random walk speed parameter we choose $\alpha = 2\mathrm{m}^2/\mathrm{sec}$. This implies—assuming no big moves—that in one hour, a random walker has a 50% probability of being within 100m of where they started. For the PA size, we consider a dense system like UDPC [Cox87] where base station spacing is 600m, so $R = 300\mathrm{m}$. For calls in and calls out, we assume the rate of calls is $\rho_{\mathrm{CO}} = 5\mathrm{Calls/Day}$ and, by symmetry, $\rho_{\mathrm{CI}} = \rho_{\mathrm{CO}}$. The proba-

Table 1: Summary of Traffic Components and Desired Direction of Change in D and T_{U}.

Traffic Type		Expression	Desired Change	
			D	T_{U}
Page	No Response	$\rho_{\mathrm{CI}} \cdot p_{\mathrm{NR}} \cdot \frac{D^2}{R^2} \cdot b_{\mathrm{P}}$	⇓	-
	Big Move	$\rho_{\mathrm{CI}} \cdot (1 - p_{\mathrm{NR}}) \cdot (1 - e^{-\lambda T_{\mathrm{U}}}) \cdot \frac{\rho_{\mathrm{BM}}}{\lambda} \cdot \frac{\gamma D^3}{3R^2} \cdot b_{\mathrm{P}}$	⇓	↓
	Random Walk	$\rho_{\mathrm{CI}}(1 - p_{\mathrm{NR}})\left[1 - (1 - e^{-\lambda T_{\mathrm{U}}})\frac{\rho_{\mathrm{BM}}}{\lambda}F_{\mathrm{BM}}(D)\right] \cdot \bar{P}(R, D, \lambda', T_{\mathrm{U}}) \cdot b_{\mathrm{P}}$	↑	⇓
Update	Big Move	$\rho_{\mathrm{BM}} \cdot (1 - F_{\mathrm{BM}}(D)) \cdot b_{\mathrm{U}}$	↑	-
	Time Out	$\frac{\lambda'}{e^{\lambda' T_{\mathrm{U}}} - 1} \cdot b_{\mathrm{U}}$	-	⇑
	Random Walk	$\left(p'_{\mathrm{RW}}\ (D, R, D, T_{\mathrm{U}}) + \frac{p_{\mathrm{RW}}(D, R, D, \lambda', T_{\mathrm{U}})}{e^{\lambda' T_{\mathrm{U}}} - 1}\right) \cdot \lambda' \cdot b_{\mathrm{U}}$	↑	↓

bility of not responding to a page we attribute mainly to being in transit. From [Kea71] on average people drive one hour per day implying p_{NR} = 5%. Studies of mobile systems suggest as high as 50% of calls to mobiles are never terminated [Meh94]. We assume that the system will mark handsets that miss their update time as out of service, eliminating most of the useless pages. The minimum inter-update time is 5 minutes. This is a balance between effectively keeping track of users and not limiting the number of updates.

Finally we consider two measures for the paging and update costs. One is data base accesses, and the other is spectrum usage. For the former we assume $b_P = b_U = 1$ access unit. For the latter, we assume each update or page uses 100bits to transmit the portable ID plus coding and protocol overhead. Each of these need to be translated differently into system impact. For the accesses, we consider a database serving 10^6 customers, while for the bandwidth, we consider a system where the number of users per frequency reuse area is 10^4. This corresponds to about 1500 users per Km^2 for R = 300m and 25 frequency channel groups. We remark that as long as $b_P = b_U$, the traffic results scale linearly with the cost. Table 2 summarizes the parameter values.

7.2 RESULTS

Figures 4a–4d show four plots where we varied one set of parameters while holding the others constant. The results are in terms of database accesses per second per 10^6 users, and by multiplying by the factor (#User/Reuse)(#bits/page)/(#User/DB) can be converted to bits per second. Using Table 2 this factor happens to be 1.

Figure 4a shows varying the calls per day (both in and out) from 0.5 to 30. Figure 4b shows varying the random walk constant, α, and number of big moves, from 0.1 to 10 times their Table 2 values. Figure 4c shows varying D from R to 30R. Figure 4d shows varying R from 30m to 3,000m while holding D/R =3. The main result is that the system load is small, less than 1,000 accesses per second (per 1,000,000 users) over a wide range of values and more typically is less than 500 accesses per second.

Table 2: Model Parameters and Their Values.

$\rho_{CO} = \rho_{CI}$ = 5/Day	α = 2m^2/sec	#User/DB = 10^6
ρ_{BM} = 4.5/Day	γ = 1/6800m	$b_P = b_U$ = 1access
p_{NR} = 0.05	R = 300m	#User/Reuse = 10^4
T_{min} = 300sec	$D = 3R$	$b_P = b_U$ = 100bits

As call rates increase in Figure 4a, most of the traffic results from the pages to find the users, while update traffic remains nearly constant. As users move more in Figure 4b, the numerous updates triggered by moves grows quickly, while paging traffic increases slightly to find users who make a big move within the same RA. As the registration area size increases relative to the paging area in Figure 4c traffic is dominated by the pages to the whole RA when users don't respond and the greater proportion of spurious big moves that fall within the same RA. Time out updates increase to better detect these spurious big moves. As the registration area shrinks updates due to the random walk component dominate. Also from Figure 4c, we see that given the choice of $D \in \{R, 3R, 5R, \ldots\}$, $3R$ minimizes the traffic although 5R is not much larger. As the PA size and RA size increase together in Figure 4d, the PA size grows until most movement is within the same RA and few updates occur, while with smaller PA and RA, mobility is more important requiring more updates and pages to locate users.

Table 3 indicates, for the parameters in Table 2 and four variations based on Figure 4, the total traffic, the update time and the breakdown of the traffic into the 6 components of Table 2. The variations were chosen so that the total traffic doubled. We observe that each scenario has different characteristics. In the nominal case, most of the traffic is from normal paging of random walking users, and updates triggered by big moves. The former may be reduced by paging PA within each successive concentric ring in steps (e.g. one PA at a time) although constraints on acceptable delays may limit this. When call traffic increases, the traffic is mostly from normal paging, while when mobility increases, the traffic shifts towards the big move updates. Interestingly, when we increase the RA size relative to the PA size, the largest traffic component is pages to the many PA in the RA when a user is not responding. When the PA size decreases updates due to random walking users become significant. We observe that the random walk component is a

Table 3: Total Traffic (in Database Accesses per 10^6 User Sec.), Update Time, and Traffic Breakdown for Nominal (Table 2) Parameters and Four Variations.

Difference from Nominal	Total Traf.	T_U (Hr)	Paging			Update		
			NR	BM	RW	BM	TO	RW
Nominal	280	7.4	9%	2%	71%	16%	1%	0%
(a) $\rho_{CO} = \rho_{CI} =$ 14/Day	560	3.4	13%	1%	77%	8%	1%	0%
(b) $\alpha = 11\text{m}^2/\text{sec}$ $\rho_{BM} = 25/\text{Day}$	560	6.6	5%	2%	45%	45%	0%	3%
(c) $D = 8.6\text{R}$	560	3.5	38%	19%	31%	6%	5%	0%
(d) $R = 82\text{m}$ $D/R = 3$	560	3.1	5%	0%	67%	9%	3%	16%

significant component of the update traffic only with extremely small PA, while the paging with no response is always significant. Efforts such as keeping *D* relative to *R* small, and a mechanism for tagging as unavailable users that miss their time-out update would control this latter wasted traffic. This analysis shows that the strategy is robust to a wide range of values, from busy users that call or receive more than a call per hour; to active mobile users making more than two trips an hour; to large registration areas; to paging areas that range down to pico-cell sizes.

From Table 3 we see that the update period is long relative to the time necessary to make an update indicating duty cycles less than 1/1000. In practice, the task of setting the update time could fall either on the user, or the system. Since the cost in terms of battery energy is not significant relative to the energy spent in other activities whereas the cost in terms of aggregate system impact can be high, the responsibility falls on the system. The period could be either fixed (although different values for different service levels), or set dynamically based on observed user behavior with the system notifying the user when their next update should be.

The number of updates per day for each handset can be calculated by multiplying the graph of update rates by (86,400 sec/day)/(10^6 Users) ~ 1/10. From Figure 4, the number of updates except for the most active users, is less than 10 updates per day.

7.3 EFFECTIVENESS OF STRATEGY ELEMENTS

To understand the significance of these results and the relative importance of the different strategy elements we consider the following experiments summarized in Table 4. As a lower bound on the traffic, we note that for users who never move and never update, they still must be paged in their PA for every incoming call and they still may not respond. Plugging $\alpha = \rho_{BM} = 0$ into the model we see that this requires 80 updates per million user seconds. This implies that with this paper's strategy the full mobility case increases the traffic by a factor of 3.5.

Table 4: Minimal Traffic and Traffic Effects of Different Elements of Paging Strategy.

Difference from Nominal	Total Traf.	T_U (Hr)	Paging			Update		
			NR	BM	RW	BM	TO	RW
Nominal	280	7.4	9%	2%	71%	16%	1%	0%
Minimal (No Moves)	80	∞	31%	0%	69%	0%	0%	0%
No Periodic Updates	280	∞	9%	2%	72%	16%	0%	0%
Fixed RA	470	2.8	6%	1%	38%	10%	2%	43%
Immediate Updates	670	7.4	4%	1%	30%	65%	0%	0%
Page All of RA	700	∞	4%	22%	68%	6%	0%	0%

Without automatic periodic updates (i.e. set T_U to infinite), the paging is less precise for random walkers while the time-out update traffic is eliminated. The effect on total traffic though is negligible questioning the importance of the mechanism. But, the unmodeled benefit of keeping track of which users are active makes it a significant component. Future work will quantify this.

With fixed instead of dynamic RA, a user who enters a new registration area could be anywhere as opposed to in the central PA as with the dynamic case. For simplicity, we assume that big moves dominate the update traffic and users are located uniformly within their RA. To model this: for random walks we set $R = D$ in (18); for big moves, we assume that the number of undetected big moves halves; and for the paging model we assume no significant difference. The traffic in this case nearly doubles, with random walk updates nearly half of the update traffic. This is due to the many users crisscrossing the fixed border despite the minimum inter-update time, T_{min}. If fixed RA must be used then closer attention must be paid to T_{min}, choosing a larger value.

With immediate updates, a big move would trigger an update not once at the end of the move, but once for each RA radius, D, of distance travelled. Note that this is actual distance travelled and not just the displacement from trip start and end. In this case, given an average trip distance of $1/\gamma'$, and since at update the RA of radius D centers on the user, the big move triggers $(\gamma' D)^{-1}$ updates, where from Appendix A, $\gamma' = \frac{\pi}{4}\gamma$. This more than doubles the traffic.

If instead of the focused paging, all of the PA in the RA are always paged, then the traffic for each of the three paging components is the probability of that component times the number of cells in the RA. We see that this increases traffic by a factor of almost three.

8 MODEL REFINEMENTS

Several assumptions were made in this model that could be improved upon. For the strategy defined and parameters used, the minor improvements in these refinements are offset by their added complexity, but in different regimes these may be important. Throughout a circular cell geometry was assumed, whereas in practice it may be hexagonal, square, or varied. In general, the big moves could account for the combined effect of big moves together with the random walking, and other big moves smaller than D. The probability of no response was set assuming a mechanism for tagging as unavailable users that miss their time-out update. This is clearly a function of T_U, and should be modeled directly. The random walk component was either assumed negligible or (in Appendix B) generating updates as a

Poisson process, both imprecise assumptions. Finally, the model assumes that all of the users are identical to a single average user instead of averaging the traffic over a distribution of users. As Figure 4 indicates, users can have dramatically different impact based on their characteristics, and a few marginal users may dominate the total load.

9 CONCLUSION

Using a combination of empirical data and theoretical models, this paper develops a model of user behavior for a PCS environment. With this model we analyze a mobility management strategy that combines automatic updates by the users—either when they make significant moves or when they go extended periods without network interaction—multiple hysteresis in the form of dynamic registration areas and delayed updates, and a focused paging strategy.

The model shows that over a range of system and user parameters, the total paging and update traffic can be kept below 1 per 2000 user seconds. The impact on the user's handset is—except for the most active users—less than 10 brief updates per day. We also showed that the total traffic is within a factor of four of the stationary user minimum. The three most significant elements of the strategy are the dynamic registration areas, delayed updates, and the focused paging while the periodic updates may be important for avoiding wasted pages on users who are not in contact with the system.

It should be pointed out that the strategy discussed in this paper would not be possible to analyze using typical fluid flow models. As examples, the fluid flow model assumes that users move continuously so that a strategy that has only one update per trip would be meaningless. Dynamic registration areas would actually increase traffic in these models as users when updating would be at the center of a RA heading out instead of at the edge heading in. For this reason comparison with other models is difficult.

The model could easily be extended to cover other mobility management strategies. It provides a powerful tool for forecasting a system's performance and resource requirements. The results here, although they were targeted at PCS could apply to a variety of dense mobile radio systems.

Acknowledgment

We thank E. C. Posner and P. T. Porter for their useful comments on earlier versions of this work. This work was done in part at the California Institute of Technology.

APPENDIX A: DERIVATION OF BIG MOVE MODEL

This section is based on an extensive study of 30,000 trips by 1000 vehicles in a 1971 study sponsored by the Environmental Protection Agency and others[Kea72]. Although the study is older, it was thorough, recording details of every vehicle trip, its start time, duration, and distance and covered six different metropolitan areas. In addition, our main purpose here is to develop general models of user's mobility in which the parameters may vary, but the general patterns are constant.

We base the model on three figures from the report. For consistency we concentrate on the data from Los Angeles while the other cities look very similar. The first is the distribution of trip distances. Plotting with logarithmic y-scale in Figure 5, we see that the distribution is nearly linear, with a slightly heavier tail. We are most interested with the behavior at short distances (i.e. within the RA) so we fit an exponential to the first 18 miles to find a mean distance of 5.4 miles = 8600m. This needed to be modified slightly. If we assume that travel is restricted to a regular grid of streets, then the actual displacement from the start of the trip is not the distance driven, rather it is something less. To compensate, we assume that a driver wants to move to a place that is *r* away with every direction equally likely. The expected distance travelled is then:

$$E[\text{distance}] = \frac{1}{2\pi}\int_0^{2\pi}(|r\sin\theta| + |r\cos\theta|)d\theta = \frac{4}{\pi}r \tag{13}$$

Thus, converting from average distance traveled to average displacement is 8600m $\pi/4$ = 6800m.

We now turn to the trip arrival process. Figure 6a shows the distribution of the number of arrivals in a 24 hour period. This is reasonably well fit by a Poisson distribution with mean 4.5 trips. Since most trips come in pairs (there and back), odd numbers of trips are reduced in probability. Figure 6b plots the inter-trip times. Note that it is the time between when one trip ends and the next begins and not what we need—the time between the start of trips—but we will use it unmodified. Despite distortions at 9–10 and 13–15 hours corresponding to typical intra-work and inter-work periods, it is reasonably well fit with an exponential distribution with mean 24 hours per 4.5 trips.

This data indicates that we can treat both trip start times, and trip lengths as Poisson processes. Since a Poisson process is a memoryless (Markov) process, one interesting interpretation is that a user is just as likely to start a trip in the next minute no matter how long since their last trip, and similarly, each car driving down a street is just as likely to stop in the next block no matter how long it has been driven. The analysis in this section has not been rigorous, but sufficient to justify the use of our simple model.

APPENDIX B: DERIVATION OF RANDOM WALK MODEL

This is based on a heat equation that describes the diffusion of the probability of a random walking user initially uniformly distributed in its registered PA, out to a heat sink at the edge of the larger circle of radius D [Kre72]. The basic function is, $U(r,t)$, the probability density function at radius r from the center of the registered PA at time t since registration of users which have not random walked to the edge of their RA and reregistered. We use the boundary conditions:

$$\begin{aligned} U(D, t) &= 0 \qquad \forall t \\ U(r, 0) &= \frac{1}{\pi R^2} \qquad \forall |r| < R \end{aligned} \tag{14}$$

The first implies that any user wandering to the edge of their RA will re-register and be ‘absorbed’. The second assumes that initially a user is anywhere within his PA with uniform probability. By making equivalence with temperature distribution dynamics with a heat sink boundary in cylindrical coordinates [Kre72], we see that U must obey the differential equation:

$$\frac{\partial U}{\partial t} = \alpha\left(\frac{\partial^2 U}{\partial r^2} + \frac{1}{r}\frac{\partial U}{\partial r}\right) \tag{15}$$

where α is the random walk speed constant. The solution to this equation is:

$$U(r, t) = \sum_{i=1}^{\infty} \frac{2J_1(a_{i0}R/D)}{\pi D R a_{i0}(J_1(a_{i0}))^2} e^{-\alpha\left(\frac{a_{i0}}{D}\right)^2 t} J_0(a_{i0}r/D) \tag{16}$$

where J_n is the nth Bessel function, and a_{i0} is the ith zero of J_0. The probability that someone’s position is more than z away from the center of the PA at time t is

$$\begin{aligned} p'_{\mathrm{RW}}(z, R, D, t) &= 1 - \int_0^z 2\pi r U(r, t) dr \\ &= 1 - \sum_{i=1}^{\infty} \frac{4zJ_1(a_{i0}z/D)J_1(a_{i0}R/D)}{R(a_{i0}J_1(a_{i0}))^2} e^{-\alpha\left(\frac{a_{i0}}{D}\right)^2 t} \end{aligned} \tag{17}$$

Averaging over all possible times since the last update:

$$\begin{aligned} p_{\mathrm{RW}}(z, R, D, \lambda', T_{\mathrm{U}}) &= \int_0^{T_{\mathrm{U}}} p'_{\mathrm{RW}}(z, R, D, t) \frac{\lambda' e^{-\lambda' t}}{1 - e^{-\lambda' T_{\mathrm{U}}}} dt \\ &= 1 - \sum_{i=1}^{\infty} \frac{4zJ_1(a_{i0}z/D)J_1(a_{i0}R/D)\lambda'\left(1 - e^{-(\lambda' + \alpha(a_{i0}/D)^2)T_{\mathrm{U}}}\right)}{R(a_{i0}J_1(a_{i0}))^2(\lambda' + \alpha(a_{i0}/D)^2)(1 - e^{-\lambda' T_{\mathrm{U}}})} \end{aligned} \tag{18}$$

This is the final expression for the probability that a user has random walked further than z away averaging over inter-update times up to T_{U}.

The derivations in Section 5.1 assume that (17) and (18) are both small, i.e. a few percent of λ'. For intermediate values, a simple approximation is to calculate the rate of random walk updates, ρ_{RW}, (i.e. (9) without the factor b_{U}) and assume that they occur as a Poisson process and add this to λ and λ'. Since ρ_{RW} depends on λ', this can be iterated until ρ_{RW} converges. In this paper this procedure is used even if ρ_{RW} is small.

For large values of ρ_{RW} a different approach must be taken. Since (9) only checks if someone random walk updated at other update events, a large ρ_{RW} implies that a significant fraction of users have already updated and may have updated many times but are only counted once. The most significant case where this occurs is when the RA and PA are the same, i.e. $z = R = D$. The instantaneous update rate is the derivative of (17). Evaluating this at $t = 0$ for $z = R = D$, the rate is infinite (this makes sense since (3) is the limit of infinitely small random steps made infinitely often, thus at least half the users at the border will step across the border instantaneously). As noted in Section 4.1, we assume that users do not update instantaneously when they cross the border, but at set periodic intervals of length T_{min}. In the large ρ_{RW} case we can assume that the greatest proportion of the crossings occur in the first T_{min} period. This reduces the random walk traffic to

$$\mathrm{Traffic}_{\mathrm{RW}} = \rho_{\mathrm{RW}} b_{\mathrm{U}} = \frac{1}{T_{\mathrm{min}}} p'_{\mathrm{RW}}(D, R, D, T_{\mathrm{min}})\, b_{\mathrm{U}}. \tag{19}$$

In this paper (19) and (9) are calculated and the largest chosen.

References

[Bau94] Baumann, F.V., Niemegeers, I.G., “An Evaluation of Location Management Procedures,” *Proc. 3rd Int. Conf. on Universal Personal Communications,* IEEE, Piscataway, NJ. 1994. pp. 359–364.

[Bed93] Beddoes, E.W., Easteal, J.R., “TACS: The U.K. Approach,” in Balston, D.M., Macario, R.C.V., ed., *Cellular Radio Systems,* Artech House, Boston. 1993. p. 123.

[Ber88] Bernhardt, R.C., “User Access in Portable Radio Systems in a Co-Channel Interference Environment,” *Proc. IEEE Inter. Commun. Conf.,* 1988. pp. 768–773.

[Cox87] Cox, D.C., “Universal Digital Portable Radio Communications,” *Proc. IEEE,* Vol. 75, 1987. pp. 436–477.

[Dog90] Doganata, Y.N., Brown, T.X, Posner, E.C., “Call Setup Strategy Trade-offs for Universal Digital Portable Communications,” *Computer Networks and ISDN Systems 20,* North-Holland. 1990. pp. 455–464.

[Hon86] Hong, D., Rappaport, S.S., “Traffic model and performance analysis for cellular mobile radio telephone systems with prioritized and non-prioritized handoff procedures,” *IEEE Tran. on Vehicular Technology,* Vol. VT-35, No. 3, August 1986, pp. 77–91.

[Gue87] Guérin, R., “Channel occupancy time distribution in a cellular radio system,” *IEEE Tran. on Vehicular Technology,* Vol. VT-35, Aug. 1987, pp. 89–99.

[Kea71] Kearin, D.H., Lamoureux, R.L., Goodwin, B.C., *A Survey of Driving Patterns in Six Urban Areas of the U.S.: Summary Report*, National Technical Information Service. Report No. CRC-APRAC-CAPE-10-68-7, Jan. 29, 1971

[Kre72] Kreyszig, E., *Advanced Engineering Mathematics, 3rd Ed.*, Wiley, NY. 1972. Chapter 3 and p. 430.

[Lo92] Lo, C.N., Wolff, R.S., Bernhardt, R.C., “An estimate of network database transaction volume to support universal personal communications services,” *Proc. 1st. Int. Conf. on Universal Personal Commun.*, Dallas, Texas, Sep. 18 – Oct. 1, 1992.

[Meh94] Mehrotra, A., *Cellular Radio Analog & Digital Systems,* Artech House, Boston. 1994. p. 94.

[Mei92] Meiller-Hellstern, K., Alonso, E., “The Use of SS7 and GSM to Support High Density Personal Communications,” *Proc. ICC ‘92*, Chicago, Illinois, Paper 356.2, June 15–18, 1992.

[Moh94] Mohan, S., Jain, R., “Two User Location Strategies for Personal Communications Services,” *IEEE Personal Commun.,* First Quarter, 1994 Vol. 1, No. 1, pp. 42–50.

[Ses92] Seskar, I., Maric, S., Holtzman, J. Wasserman, J., “Rate of location area updates in cellular systems,” in *Proc. IEEE VTC’92*, Denver, May 1992.

[Tho88] Thomas, R., Gilbert, H., Mazziotto, G., “Influence of the movement of the mobile station on the performance of the radio cellular network,” *Proc. 3rd Nordic Seminar*, Paper 9.4, Copenhagen, Sep. 1988.

[Xie93] Xie, H., Tabbane, S., Goodman, D.J., “Dynamic Location Area Management and Performance Analysis,” in *Proc. IEEE VTC’93*, 1993.

Figure Captions:

Figure 1: User Mobility Model

Figure 2: Updating and Paging Traffic Components.

Figure 3: Discrete Nature of Update Trigger Distance, *D*.

Figure 4: Pages and Updates per Second per 1,000,000 Users. Four sets of parameters are varied: call rates (in and out) (a), users local and long distance mobility (b), update trigger distance (c), and paging area size (d).

Figure 5: Distribution of Trip Distance in Los Angeles, Weekday Trips (from [Kea71, Figure 5a]).

Figure 6: Distribution of Number of Trips per Day (a), and Elapsed Time Between Trips (b) in Los Angeles, Weekday Trips (from [Kea71, Table 1 and Figure 11a]).

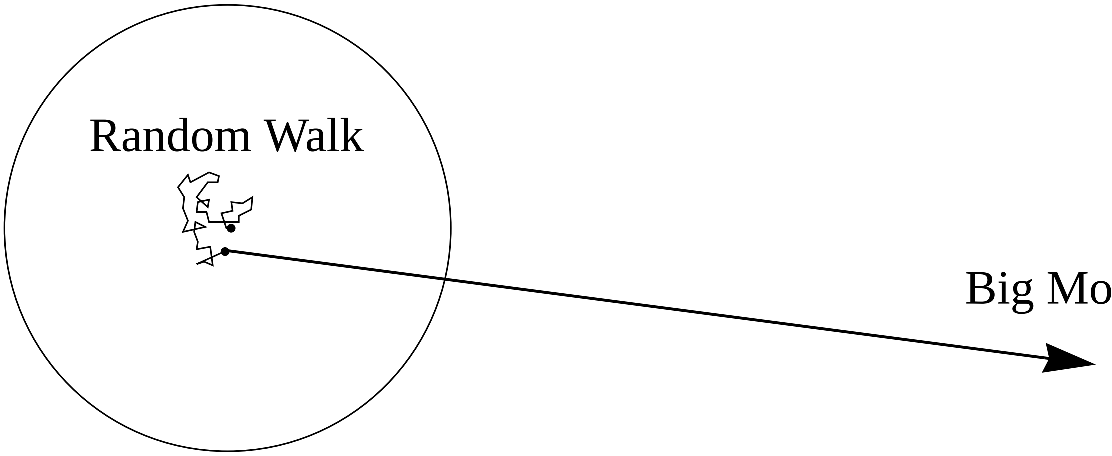


Figure 1

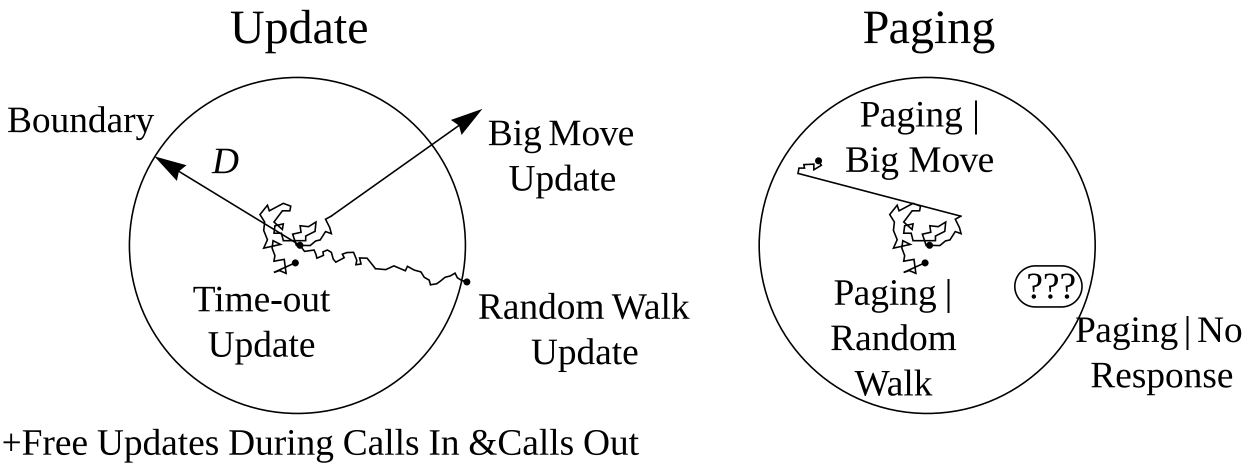


Figure 2

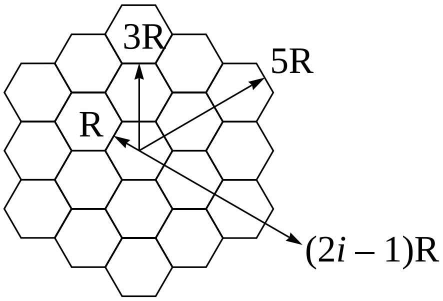


Figure 3

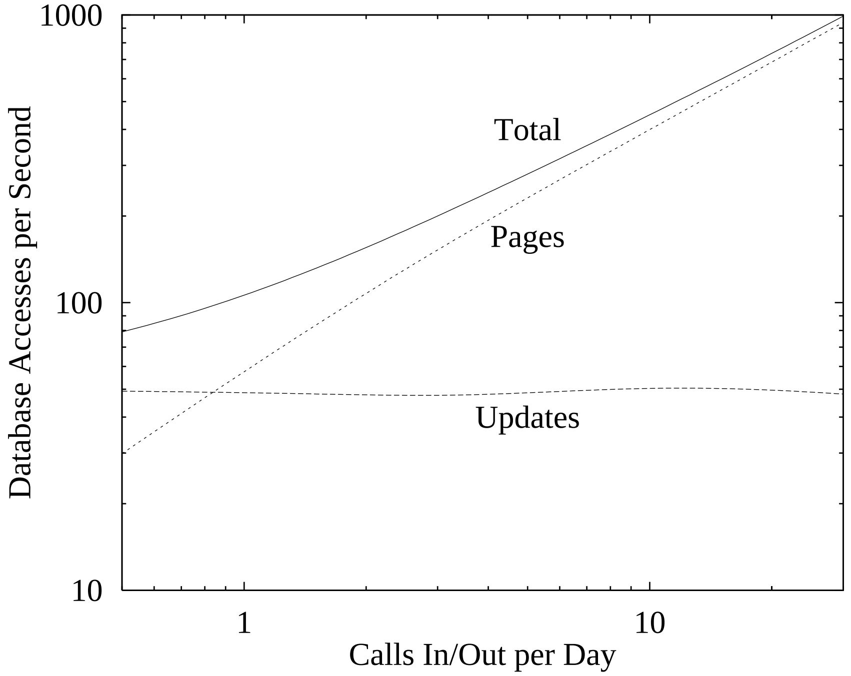


Figure 4a

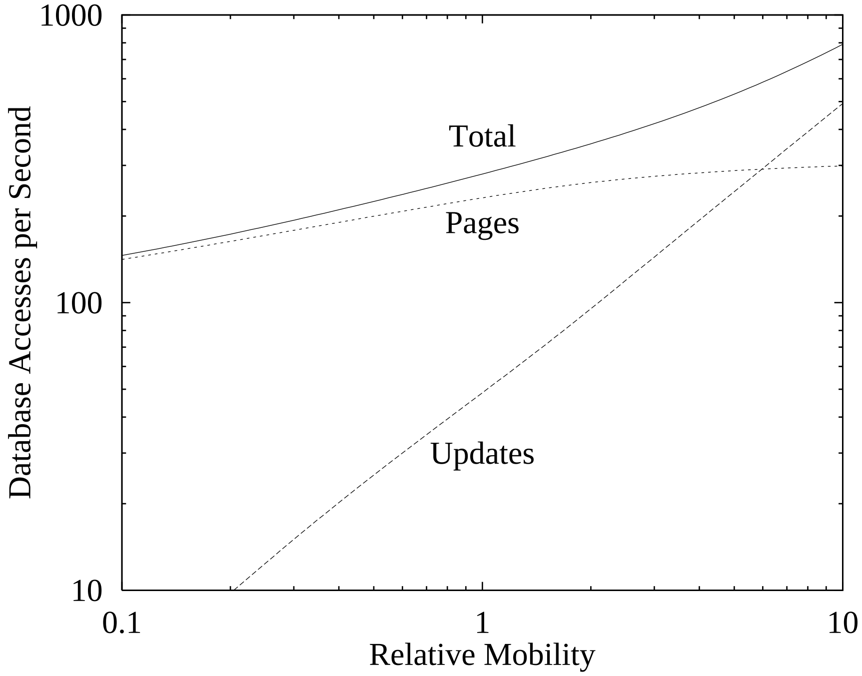


Figure 4b

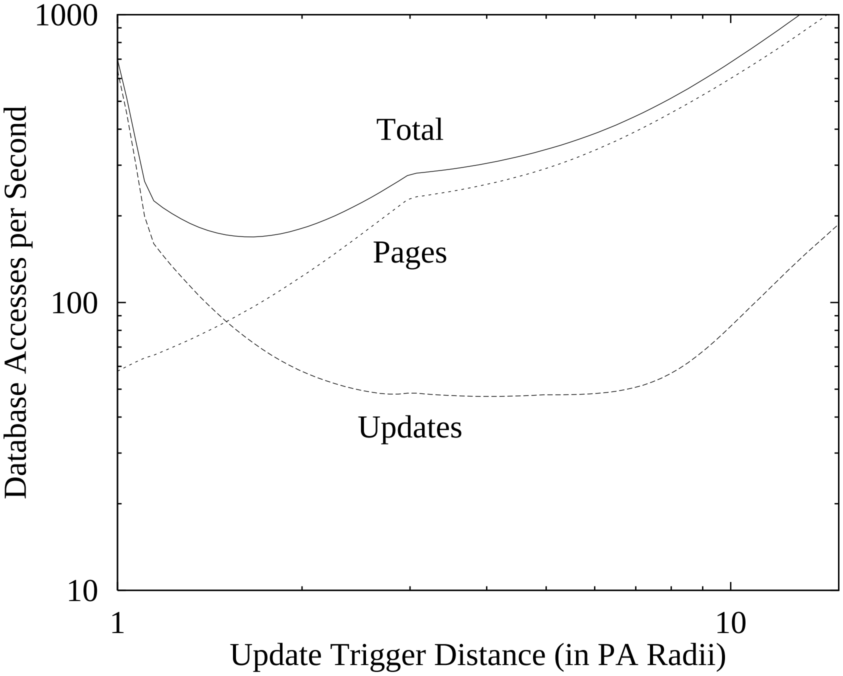


Figure 4c

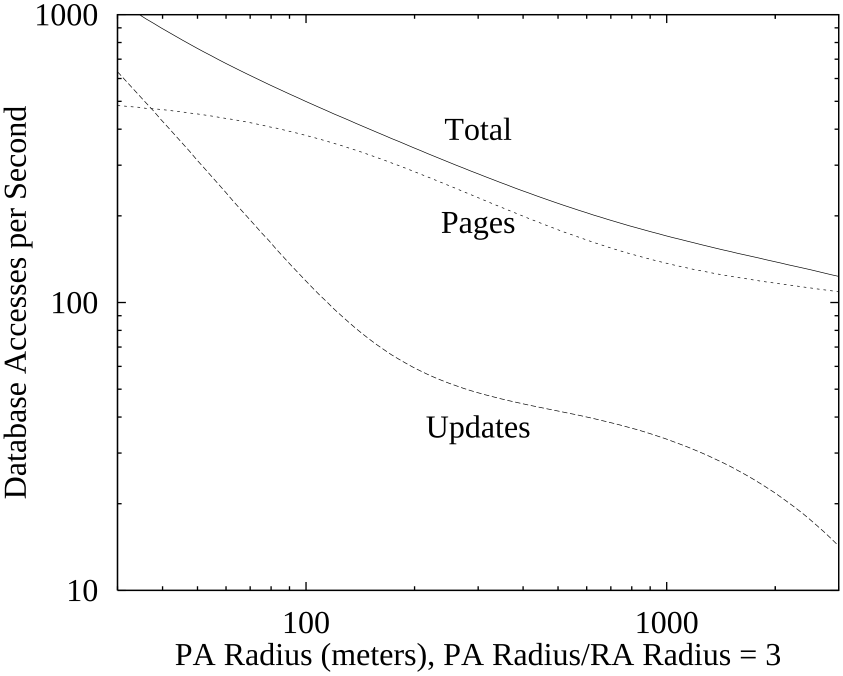


Figure 4d

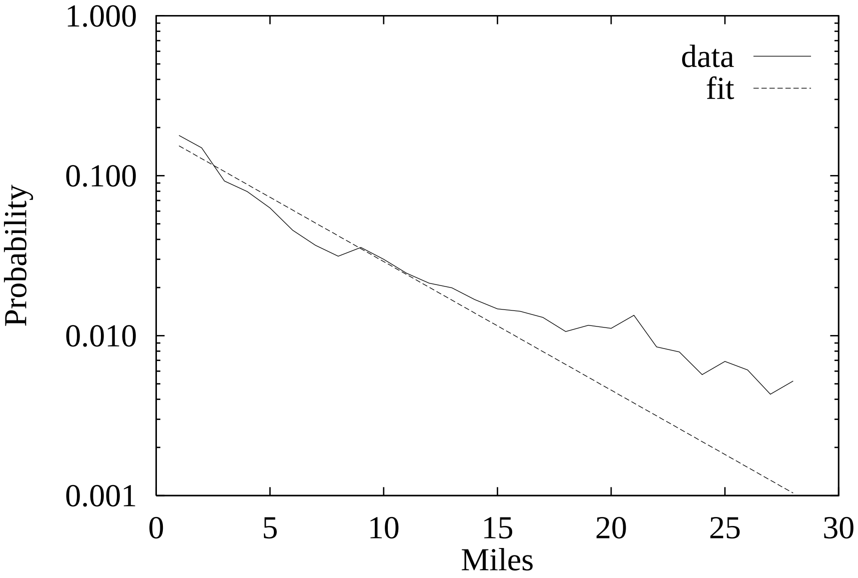
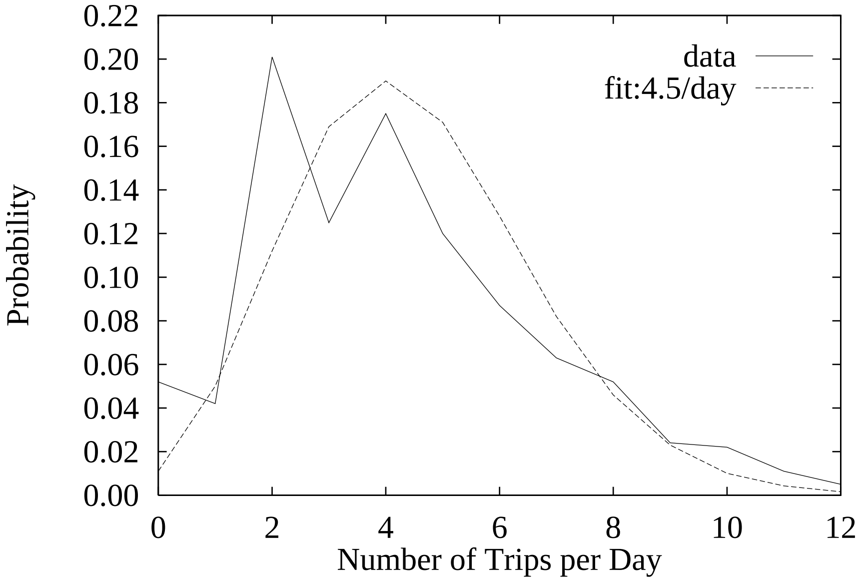

1.000
0.100
0.010
0.001
Probability
data
fit
0
5
10
15
20
25
30
Miles

Figure 5



0.22
0.20
0.18
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
Probability
data
fit:4.5/day
0
2
4
6
8
10
12
Number of Trips per Day

Figure 6a

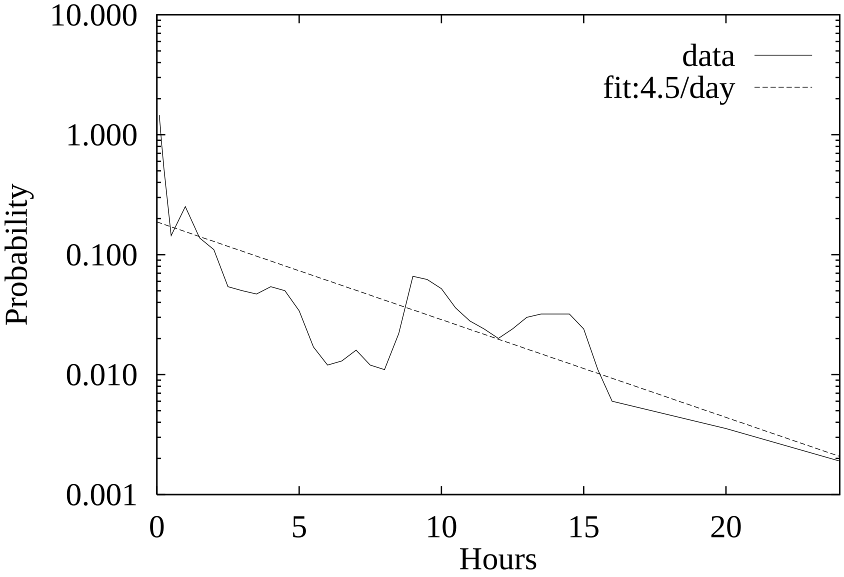

10.000
1.000
0.100
0.010
0.001
Probability
0
5
10
15
20
Hours
data
fit:4.5/day

Figure 6b