[12] C. Rose. Minimizing the average cost of paging and registration: A timer-based method. WINLAB-TR 83, WINLAB, Rutgers University, December 1994. (submitted for publication).

[13] C. Rose. State-based paging/registration: A greedy technique. WINLAB-TR 92, WINLAB, Rutgers University, December 1994. (submitted for publication).

[14] D. P. Bertsekas. *Constrained Optimization and Lagrange Multiplier Methods.* Academic Press, San Diego, CA, 1982.

[15] F.B. Hildebrand. *Advanced Calculus for Applications.* Prentice Hall, Englewood Cliffs, NJ, 1976.

[16] D. P. Bertsekas. *Dynamic Programming.* Prentice Hall, Englewood Cliffs, NJ, 1987.

[17] G.F. Simmons. *Differential Equations with Applications and Historical Notes.* McGraw-Hill Book Company, New York, NY, 1972.

[18] A. Papoulis. *Probability, Random Variables, and Stochastic Processes.* McGraw-Hill, New York, 1965.

References

[1] K.S. Meier-Hellstern, G. P. Pollini, and D. J. Goodman. Network protocols for the cellular packet switch. *IEEE Transactions on Communications*, 42(2–4/2):1235, 1994.

[2] K.S. Meier-Hellstern, E. Alonso, and D. O'Neil. The use of SS7 and GSM to support high density personal communications. In *Proceedings of the International Conference on Communications ICC'92*, 1992.

[3] R. Thomas, H. Gilbert, and G. Mazziotto. Influence of the movement of the mobile station on the performance of a radio cellular network. In *Proc. 3rd Nordic Seminar*, September 1988. Paper 9.4.

[4] H. Xie, S. Tabbane, and D. J. Goodman. Dynamic location area management and performance analysis. In *Proc. of IEEE Vehicular Technology Conference*, 1993.

[5] S. Tabbane. Comparison Between the Alternative Location Strategy and the Classical Location Strategy. *WINLAB Technical Report no. 37*, 1992.

[6] A. Bar-Noy and I. Kessler. Tracking mobile users in wireless networks. In *INFOCOMM'93*, pages 1232–1239, March 1993. San Francisco, CA.

[7] S. Tabbane. Evaluation of an Alternative Location Strategy for Future High Density Wireless Communications Systems. *WINLAB Technical Report no. 51*, January 1993.

[8] G. Pollini and S. Tabbane. The Intelligent Network Signaling and Switching Cost of an Alternative Location Strategy Using Memory. In *IEEE VTC'93*, May 1993. Seacaucus, NJ.

[9] J.Z. Wang. A Fully Distributed Location Registration Strategy for Universal Personal Communication Systems. *IEEE J. SAC*, 11(6):850–860, August 1993.

[10] S. Okasaka, S. Onoe, S. Yasuda, and A. Maebara. A New Location Updating Method for Digital Cellular Systems. In *IEEE VTC'91*, pages 345–350, May 1991. Denver, CO.

[11] B. Awerbuch and D. Peleg. Concurrent Online Tracking of Mobile Users. In *Proc. ACM SIGCOMM Symposium on Communication, Architecture and Protocols*, October 1991.

Let $\mathbf{\Delta} = \mathbf{x} - \mathbf{y}$ so that $\mathbf{z} = \lambda\mathbf{\Delta} + \mathbf{y}$. We then have

$$\frac{\partial^2 \mathcal{G}(\mathbf{z})}{\partial \lambda^2} = \sum_{i,j=1}^{N} \frac{\partial^2 \mathcal{G}(\mathbf{z})}{\partial x_i \partial x_j} \Delta_i \Delta_j \tag{19}$$

Using equation (18) we obtain,

$$\begin{aligned}
\frac{\partial^2 \mathcal{G}}{\partial \lambda^2} &= \sum_{i=1}^{N} \frac{\partial^2 \mathcal{G}(\mathbf{z})}{\partial x_i^2} \Delta_i^2 + 2\sum_{i=1}^{N-1} \frac{\partial^2 \mathcal{G}(\mathbf{z})}{\partial x_i \partial x_{i+1}} \Delta_i \Delta_{i+1} \\
&= \sum_{i=1}^{N} (z_i - z_{i+1} - \alpha) g'(z_i) \Delta_i^2 + 2\sum_{i=1}^{N} g(z_i)\Delta_i^2 - 2\sum_{i=1}^{N-1} g(z_i)\Delta_i\Delta_{i+1} \\
&= \sum_{i=1}^{N} (z_i - z_{i+1} - \alpha) g'(z_i) \Delta_i^2 + g(z_1)\Delta_1^2 + g(z_N)\Delta_N^2 \\
&\quad + \sum_{i=1}^{N-1} g(z_i)(\Delta_i - \Delta_{i+1})^2
\end{aligned} \tag{20}$$

For $\alpha \geq 0$ we have $z_n - z_{n+1} - \alpha \leq 0$ which implies that $\frac{\partial^2 \mathcal{G}}{\partial \lambda^2} \geq 0$ and $\mathcal{G}$ is convex when $\alpha \geq 0$. The same holds true for α chosen to satisfy equation (11) owing to the positivity of $g()$. □

Proof: Theorem 6 For optimality of y_n we examine

$$\frac{\partial \mathcal{G}'}{\partial y_n} = (y_n^* - y_{n+1}^* - \frac{\alpha}{\kappa}) g'(y_n^*) + \int_{y_{n-1}^*}^{y_n^*} g'(\omega) d\omega \tag{21}$$

Substitution of $\frac{x_n^*}{\kappa}$ for y_n^* and $\kappa g(\kappa z)$ for $g'(z)$ yields,

$$\frac{\partial \mathcal{G}'}{\partial y_n} = (x_n^* - x_{n+1}^* - \alpha) g(x_n^*) + \int_{\frac{x_{n-1}^*}{\kappa}}^{\frac{x_n^*}{\kappa}} \kappa g(\kappa\omega) d\omega$$

Letting $\omega = z/\kappa$ yields

$$\frac{\partial \mathcal{G}'}{\partial y_n} = (x_n^* - x_{n+1}^* - \alpha) g(x_n^*) + \int_{x_{n-1}^*}^{x_n^*} g(z) dz \tag{22}$$

which is identically zero owing to the assumed optimality of $\mathbf{x}^*$. Thus, $y_n^* = \frac{x_n^*}{\kappa}$ optimizes

$$\mathcal{G}'(\mathbf{y}^*) = \mathcal{L}(\mathbf{y}^*) + \frac{\alpha}{\kappa} \mathcal{D}'(\mathbf{y}^*)$$

Now, by analogy to the reduction of the integral term in equation (21) to that in equation (22) and the definitions of $\mathcal{L}$ and $\mathcal{D}$ from equations (7) and (8), we can see that $\mathcal{L}'(\mathbf{y}^*) = \mathcal{L}(\mathbf{x}^*)/\kappa$. Likewise, $\mathcal{D}'(\mathbf{y}^*) = \mathcal{D}(\mathbf{x}^*)$. Thus, $\mathcal{G}'(\mathbf{y}^*) = \mathcal{G}(\mathbf{x}^*)/\kappa$ □

at most M non-zero elements and such that $\overline{F}_X(i) > \overline{F}_U(i) = 1 - i/M$ for some $i \in \{1, \ldots, M\}$. Let i_1 be the first such i. Since $\overline{F}_X(i_1 - 1) \leq \overline{F}_U(i_1 - 1)$ we have $p_{i_1} < 1/M$. Since p_i is decreasing, $p_i < 1/M$ for all $i \geq i_1$. Thus, $\overline{F}_X(i_1) = \sum_{j=i_1+1}^{M} p_j < (M - i_1)/M$, which is a contradiction. □

Proof: Theorem 3 We have $E[L] = \sum_{n=1}^{\infty} s_n q_n$. If we search $A_r = \{r_\ell | \ell = 1, \ldots, k_r\}$ sequentially, the paging cost becomes

$$\begin{aligned} E[L'] &= \sum_{n \neq r} s_n q_n + \sum_{\ell=1}^{k_r} (s_{r-1} + \ell) p_{r_\ell} \\ &\leq \sum_{n \neq r} s_n q_n + (s_{r-1} + k_r) \sum_{\ell=1}^{k_r} p_{r_\ell} \\ &= E[L] \end{aligned}$$

Thus, sequential search always reduces $E[L]$, and by Theorem 1, the optimal sequential search is in order of decreasing probability. □

Proof: Theorem 4 Suppose a set $\{A_n\}$ maximizes $E[D]$ and $k_r > 1$. We have $E[D] = \sum_{n=1}^{\infty} n q_n$. If the set A_r is searched sequentially then $E[D]$ becomes

$$\begin{aligned} E[D'] &= \sum_{n=1}^{r-1} n q_n + \sum_{\ell=0}^{k_r - 1} (r + \ell) p_{r_{\ell+1}} + \sum_{n=r+1}^{\infty} (n + k_r - 1) q_n \\ &= \sum_{n=1}^{\infty} n q_n + \sum_{\ell=0}^{k_r - 1} \ell p_{r_{\ell+1}} + \sum_{n=r+1}^{\infty} (k_r - 1) q_n \\ &\geq \sum_{n=1}^{\infty} n q_n \\ &= E[D] \end{aligned}$$

Thus, sequential search maximizes D. □

Proof: Theorem 5 The second partials of $\mathcal{G}$ are

$$\frac{\partial^2 \mathcal{G}}{\partial x_i \partial x_j} = \begin{cases} (x_i - x_{i+1} - \alpha) \frac{dg(x_i)}{dx_i} + 2g(x_i) & j = i \\ -g(x_i) & j = i + 1 \\ -g(x_{i-1}) & j = i - 1 \\ 0 & otherwise \end{cases} \tag{18}$$

Given $\mathbf{x}$ and $\mathbf{y}$, let $\mathbf{z}(\lambda) = \lambda \mathbf{x} + (1 - \lambda) \mathbf{y}$. We will show that $\mathcal{G}(\mathbf{z}(\lambda))$ is convex in λ over $0 \leq \lambda \leq 1$ for all admissible $\mathbf{x}$, $\mathbf{y}$.[2]

[2]It is easily shown that if $\mathbf{x}$ and $\mathbf{y}$ are admissible (i.e., $x_n \leq x_{n+1}$ and $y_n \leq y_{n+1}$) then $\mathbf{z}$ is admissible as well.

A Proofs

Proof: Theorem 1 Suppose the set $(A_1, A_2, \ldots)$ is optimal but there exists $i \in A_l$ and $j \in A_m$ with $p_i < p_j$ but $l > m$. Let $(A'_1, A'_2, \ldots)$ denote a new paging sequence derived from $(A_1, A_2, \ldots)$ in which i and j are swapped so that $i \in A'_m$ and $j \in A'_l$. For the modified paging sequence, we define the paging cost and paging delay by L' and D'. We note that

$$\begin{aligned} E[D] - E[D'] &= lp_i + mp_j - (lp_j + mp_i) \\ &= (l-m)(p_i - p_j) \\ &> 0 \end{aligned}$$

This is a contradiction of the assumed optimality of $\{A_n\}$. Likewise for $E[L]$ we have

$$\begin{aligned} E[L] - E[L'] &= s_l p_i + s_m p_j - (s_l p_j + s_m p_i) \\ &= (s_l - s_m)(p_i - p_j) \\ &> 0 \end{aligned}$$

which also contradicts the assumed optimality. □

Proof: Theorem 2 First, we verify that $D(X) \stackrel{\text{st}}{>} D(Y)$ since

$$P\{D(X) > n\} = P\{X > s_n\} > P\{Y > s_n\} = P\{D(Y) > n\}$$

Given $l \geq 1$, there exists n such that $s_n \leq l < s_{n+1}$, so that

$$P\{L(X) > l\} = P\{X > s_n\} > P\{Y > s_n\} = P\{L(Y) > l\}$$

Thus,

$$L(X) \stackrel{\text{st}}{>} L(Y)$$

□

Proof: Corollary 1 Let $X = \{x_i\} \stackrel{\text{st}}{>} Y = \{y_i\}$ be two different distributions over an index set $A = \{a_i\}$ with $a_i \leq a_{i+1}$. We have

$$E[A(X)] = \sum_{n=1}^{\infty}(a_n - a_{n-1})\overline{F}_X(n-1) > \sum_{n=1}^{\infty}(a_n - a_{n-1})\overline{F}_Y(n-1) = E[A(Y)]$$

Thus by Theorem 2 we must have $E[D(X)] > E[D(Y)]$ and $E[L(X)] > E[L(Y)]$ since $D(X) \stackrel{\text{st}}{>} D(Y)$ and $L(X) \stackrel{\text{st}}{>} L(Y)$. □

Proof: Lemma 1 Suppose X has distribution $P\{X = i\} = p_i$ with $p_i \geq p_{i+1}$ and that the relative $\mathcal{L}^*$ remains virtually constant under fixed delay constraints. For example, with $\mathcal{D} = 2$ we can achieve $\mathcal{L}^* = 13$ for $U = 20$. For a distribution with $U = 200$ we can expect through application of Theorem 6 that $\mathcal{L}^* = 130$ with the same $\mathcal{D} = 2$. Since the absolute minimum $\mathcal{L}$ for $U = 200$ is 100.5, the relative $\mathcal{L}^*$ of the absolute minima are roughly equal.

In conclusion, for cases where the system need not find the user immediately, the optimal paging strategies presented here afford a means to significantly reduce the average amount of signaling necessary to locate a user, while maintaining modest average polling delay D. This work is applicable to any and all types of user motion for which a probability distribution on location can be measured or derived. In addition, since not all locations are polled simultaneously, a parallel search for multiple users can be mounted thereby increasing the potential paging rate and/or reducing the overall system paging delay. These ideas are the subject of current investigations.

Acknowledgements

The authors would like to thank David Goodman of the Wireless Information Network Laboratory (WINLAB) at Rutgers University for bringing the problem of location area management to the attention of the authors. We would also like to thank Uzi Timor, Zhuyu Lei and Sudarshan Rao of Rutgers University, WINLAB for interesting discussions.

signaling load. Of course, such a scheme requires some knowledge of where any particular user is likely to be. Here, we assume that such a probability distribution on user location can be derived from past measurements of user motion[1] and possibly stored with user profiles.

Given these assumed probability distributions on the user location, we have derived optimal polling strategies which minimize the average number of locations polled, L. The procedure which minimizes L, polls locations sequentially in decreasing order of probability and the L achieved is thus the mean of the ordered distribution, L^*. If we assume that each polling event requires unit time, then the mean polling delay D is equal to L^*. We also found that the uniform distribution achieved the worst performance (maximum L^* and D) of any distribution with the same number of non-zero elements. In addition, a uniform distribution with fewer elements can be used to overbound performance as well. Thus, the uniform distribution is a useful surrogate for understanding the behavior of arbitrary distributions.

For large numbers of location areas, $D = L^*$ may be unacceptably large. We therefore also considered the problem of minimizing L under constraints on D. Problems such as constraining the maximum delay or weighted average delay can be solved exactly using dynamic programming (DP). Problems involving constraints on the mean delay, however, are not amenable to DP solution. However, a continuous formulation which may be applied to all the constrained D cases was derived and variational techniques applied. The solution to the continuous problem overbounds the best performance of any discrete solution since the discrete solution is a subset of possible continuous solutions.

It was found that the discrete solutions obtained by rounding the continuous solutions lie close to this bound. Thus, the analytically tractable continuous formulation seems to provide a good approximate solution to the discrete case. The uniform distribution is especially tractable and, as previously mentioned, serves as an underbound on performance over all finite distributions of the same length. The uniform distribution was therefore used as a worst case to illustrate the gains possible using optimal paging strategies.

In the continuous case, it is generally seen that $\mathcal{L}^*$, the average number of locations polled declines rapidly with $\mathcal{D}$ the average number of polling events. This result implies that near-optimum $\mathcal{L}^*$ can be obtained even under relatively severe constraints on $\mathcal{D}$. Specifically, for the uniform distribution of 20 elements, the unconstrained minimum $\mathcal{L}^*$ is 10.5. However, even when a mean polling delay of $\mathcal{D}^* = 2$ is required, $\mathcal{L}^* = 13$ can still be achieved.

It is also noteworthy that the scaling properties of the continuous solutions implies

[1]The measurements might come specifically from the user in question or might be compiled from an aggregate of users with similar motion characteristics. Estimation of location probability distributions is the subject of current work.

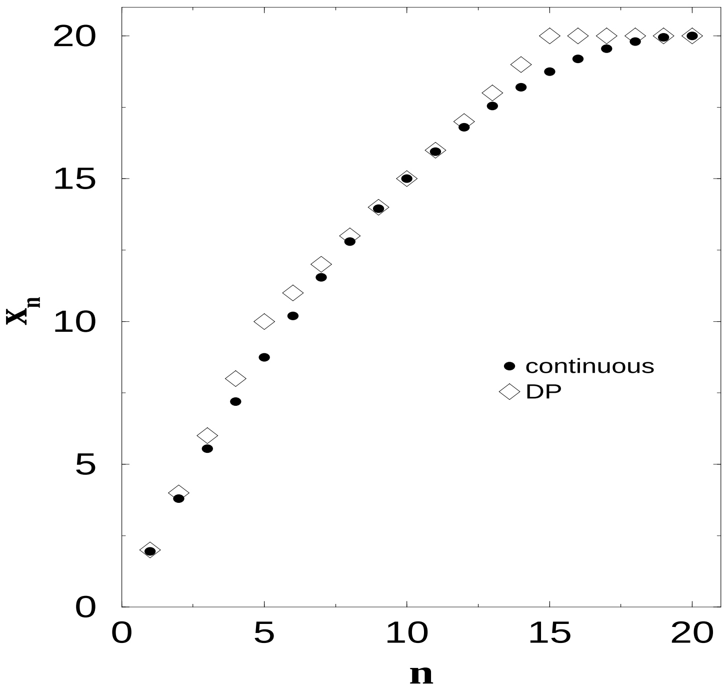

Figure 4: Location area groupings x_n obtained using a uniform location distribution and a weighted delay criterion $\alpha = 0.1$. Continuous and discrete dynamic programming (DP) solutions shown. $N = 20$ locations.

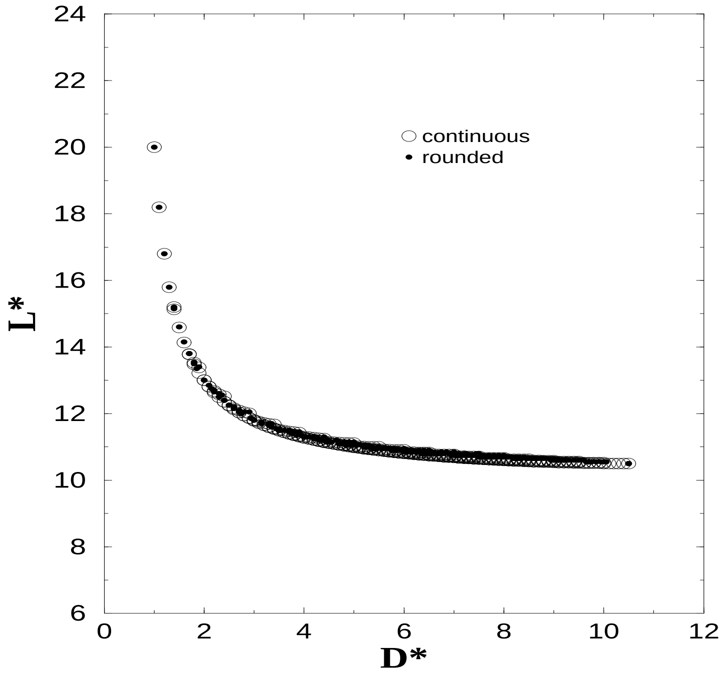

Figure 5: Montage of minimum paging cost $\mathcal{L}^*$ versus fixed mean polling $\mathcal{D}^*$ for a uniform distribution with $U = 20$ and $N = 1, 2, \ldots 20$. Both continuous and rounded solutions shown.

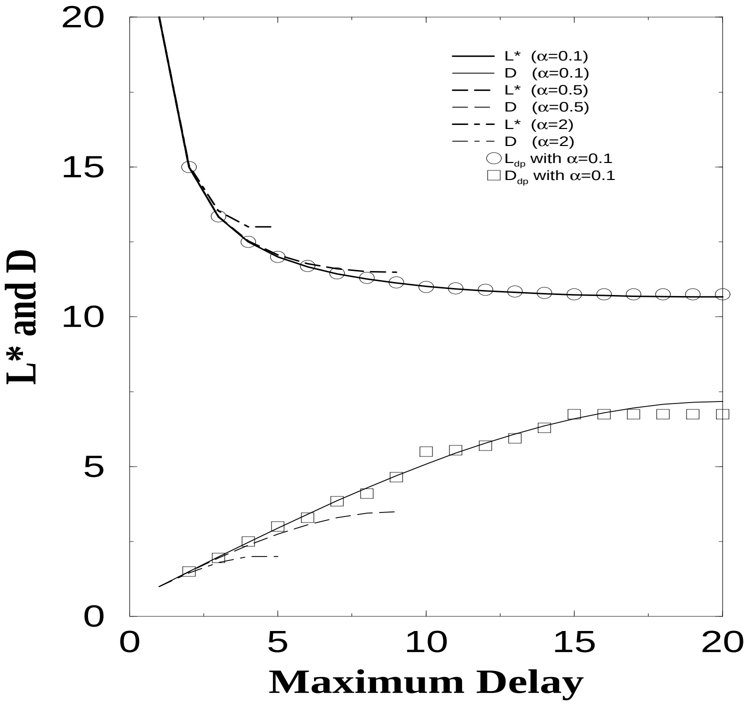

Figure 3: Family of minimum paging delay L^* versus mean polling delay $\mathcal{D}$ curves, parametrized in delay weight, α, for a uniform location distribution using the continuous formulation. Discrete solution via dynamic programming (DP) also shown. $N = 20$ locations.

maximum D constraint (equation (14)), the x_n^* for $\alpha = 0.1$ with $N = 20$ are shown in FIGURE 4. Notice that the size of the groups decreases with increasing n.

5.2.3 Mean D Constraints

We plot $\mathcal{L}^*$ versus $\mathcal{D}^*$ for fixed $U = 20$ and $N = 1, 2, ...20$ in FIGURE 5. The range of N is necessary since not all $\mathcal{D}^*$ are achievable for a given value of N. Also shown in the figure are the performance of discrete solutions ($\{s_n\}$) to the problem obtained by rounding the continuous solutions ($\{x_n\}$). The two solutions are virtually identical.

6 Discussion and Conclusion

In current systems, a connection request results in the polling of all cells in the so–called "location area" where the user is registered. These location areas are fixed and independent of individual users and usually rather large; sometimes the size of an entire city. Since each location polled requires the use of signaling resources, it would be useful to minimize the average amount of polling required to locate users.

If the user normally moves only among a fraction of cells in the location area, then some savings can be had by first searching in likely locations. This suggests that the personalized *user location areas* derived here might be useful in reducing the overall

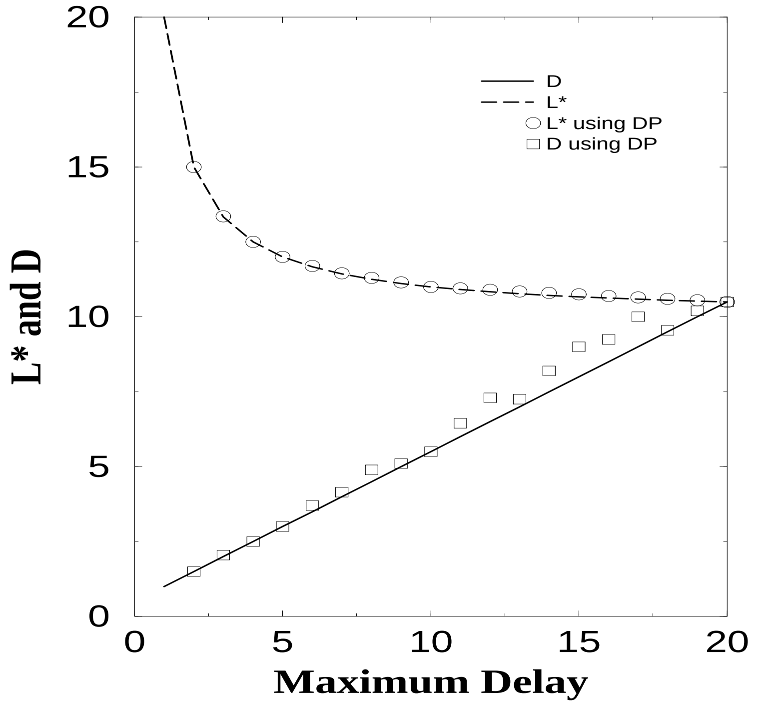

Figure 2: Minimum paging cost L^* and mean polling delay $\mathcal{D}$ versus maximum delay N, for a uniform location distribution using the continuous formulation. Discrete solution via dynamic programming (DP) also shown. $N = 20$ locations.

$\mathcal{L}$ may be then calculated as

$$\mathcal{L} = \sum_{n=1}^{N} n \left[\frac{U}{N} + \frac{\alpha}{2}(N-n)\right] \left[\frac{1}{N} + \frac{\alpha}{2U}(N+1) - n\frac{\alpha}{U}\right] \tag{16}$$

For mean constraints we find α in terms of D^*, U and N as

$$\alpha = (\frac{N+1}{2} - \mathcal{D}^*)\frac{12U}{N^3 - N} \tag{17}$$

5.2.1 Maximum D Constraints

For the case of maximum $\mathcal{D} \leq N$ we have $\alpha = 0$. Thus, $x_n = nU/N$ with $\mathcal{D} = (N+1)/2$ and $\mathcal{L}^* = U\frac{N+1}{2N}$. In FIGURE 2 we plot $\mathcal{L}^*$ and $\mathcal{D}$ as N ranges from 1 to U for $U = 20$. Also shown for comparison are the comparable discrete solutions obtained using dynamic programming. Notice the relatively close agreement.

5.2.2 Weighted D Constraints

For the case of weighted $\mathcal{D}$ we may plot a family of curves parametrized in α, the weighting factor. This was done in FIGURE 3 using $U = 20$ and $1 \leq N \leq 20$. We also show the close agreement of typical discrete solutions obtained via dynamic programming for the $\alpha = 0.1$ case. For comparison to the uniform location area groupings obtained using a

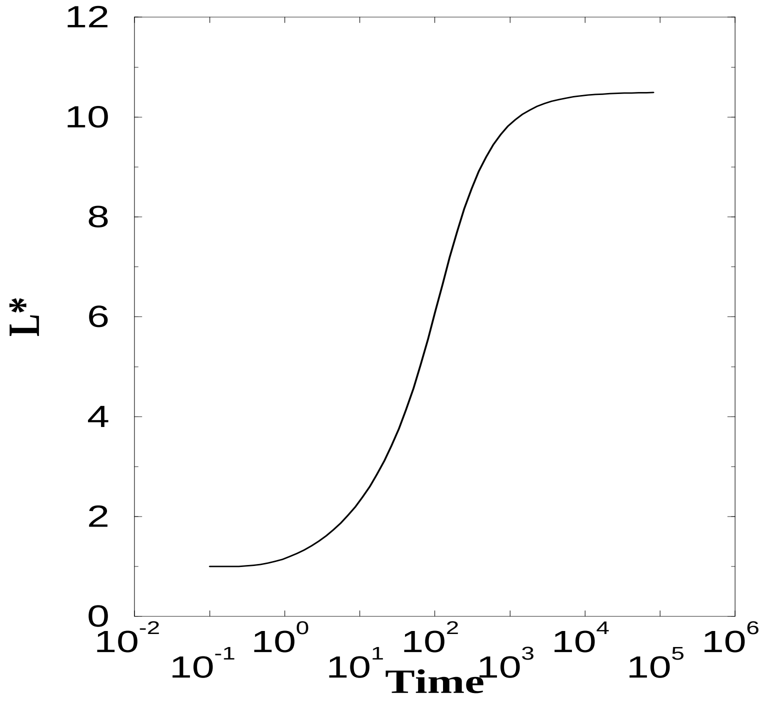

Figure 1: Minimum paging cost L^* for truncated time-varying Gaussian location distribution. The polling delay D is unconstrained. $N = 20$ locations.

- Through Corollary 1, a uniform distribution with sufficiently few elements may be used to underbound the minimum $E[L]$ and $E[D]$ of any finite location random variable distribution.

Thus, we can begin to understand the behavior of L^* and D^* for arbitrary distributions in terms of the uniform distribution.

We derive continuous solutions (which underbound the discrete solutions) for maximum, weighted and mean $\mathcal{D}$ constraints. Note that the maximum $\mathcal{D}$ and weighted $\mathcal{D}$ solutions are simply the constrained mean $\mathcal{D}$ solution with fixed Lagrange multiplier, α. These solutions will later be compared to their discrete counterparts obtained via dynamic programming.

For a continuous uniform distribution defined on $[0, U]$, equation (11) yields,

$$x_{n+1} - 2x_n + x_{n-1} = -\alpha$$

For $n = 1, 2 \ldots N$ with $x_0 = 0$ and $x_N = U$. we have,

$$x_n = \left(\frac{U}{N} + \frac{\alpha}{2}N\right) n - \frac{\alpha}{2}n^2 \tag{14}$$

Thus,

$$\mathcal{D} = \sum_{n=1}^{N} n \left[\frac{1}{N} + \frac{\alpha}{2U}(N+1) - n\frac{\alpha}{U}\right] \tag{15}$$

arises naturally for Gaussian user location distributions with time-dependent variances. We will show that if x_n^* is an optimal solution for $g(x)$ then $y_n^* = x_n^*/\kappa$ is an optimal solution for $g'(x)$. We will also show the relationship between the $\mathcal{G}$, $\mathcal{L}$ and $\mathcal{D}$ achieved by $\mathbf{x}^*$ and $\mathbf{y}^*$.

Theorem 6 *If x_n^* minimizes $\mathcal{G} = \mathcal{L} + \alpha\mathcal{D}$ for some probability density function $g(x)$, then $y_n^* = \frac{x_n^*}{\kappa}$ minimizes $\mathcal{G}' = \mathcal{L}' + \frac{\alpha}{\kappa}\mathcal{D}'$ for a scaled probability function $g'(x) = \kappa g(\kappa x)$. Furthermore, if $\mathcal{G}(\mathbf{x}^*) = \mathcal{G}^*$ then $\mathcal{G}'(\mathbf{y}^*) = \mathcal{G}^*/\kappa$.*

5 Application of Results

5.1 Unconstrained D

We showed in section 4.1 that for a non-increasing distribution, the minimum achievable mean number of locations searched is the mean of the distribution. For distributions which are not non-increasing, the minimum L is the mean of the reordered distribution.

The Gaussian distribution, $\mathcal{N}(0, \sigma^2 t)$ is a typical time-varying location probability distribution for systems under isotropic random motion [18]. We used a discretized and truncated version of the distribution defined as,

$$y_n = \frac{1}{\mathrm{erf}(\frac{N}{\sqrt{t\sigma^2}})} \frac{2}{\sqrt{2\pi t\sigma^2}} \int_{n-1}^{n} e^{-\frac{x^2}{2t\sigma^2}} dx \tag{13}$$

with $1 \leq n \leq N$. Notice that as t varies from 0 to infinity, y_n varies between a deterministic and uniform distribution on the N possibilities.

Under classical polling strategies, $L^* = N$ and $D = 1$. In FIGURE 1 L^* is shown as a function of time for the distribution y_n with $N = 20$ and $\sigma = 1$. Notice that at all times, $L^* \leq (N+1)/2$. Thus, optimal polling substantially reduces the average number of locations searched, even for a uniform distribution, by almost half.

However, the unconstrained polling delay D is identical to L^* and increases monotonically as the distribution approaches uniformity. In the next section we show that moderate constraints on D still result in L^* reasonably close to those obtainable using unconstrained D.

5.2 The Uniform Distribution and Constrained D

We pursue analytic results for the uniform distribution since,

- They are simple to derive in closed form.
- As shown in Corollary 2, the uniform distribution supplies an overbound on the minimum $E[L]$ and $E[D]$ of any finite location random variable distribution.

carry over to the continuous case if we define the appropriate complementary density function

$$\bar{\mathcal{F}}_X(x) = P\{X > x\}$$

We can then consider minimizing

$$\mathcal{G} = \mathcal{L} + \alpha(\mathcal{D} - D^*) = \mathcal{L} + \alpha\mathcal{D} + constant \tag{9}$$

For minimization with a maximum $\mathcal{D}$ constraint we have $\alpha = 0$. For the weighted mean problem, α is some constant greater than zero, and for constrained mean problems, α is the Lagrange multiplier. Differentiation of equation (9) with respect to the x_n yields,

$$\frac{\partial \mathcal{G}}{\partial x_n} = (x_n - x_{n+1} - \alpha)g(x_n) + \int_{x_{n-1}}^{x_n} g(\omega)d\omega \tag{10}$$

Setting equation (10) to zero yields,

$$(x_{n+1} - x_n + \alpha)g(x_n) = \int_{x_{n-1}}^{x_n} g(\omega)d\omega \tag{11}$$

Since $x_0 = 0$ and $x_N = X$, this second order difference equation has a unique solution [17]. Note that equation (11) may be rewritten as a recursion in x_n,

$$x_{n+1} = x_n - \alpha + \frac{1}{g(x_n)} \int_{x_{n-1}}^{x_n} g(\omega)d\omega \tag{12}$$

which given α, allows the $\{x_n\}$ to be found iteratively via a choice of x_1. All that remains is to determine whether $\mathcal{G}$ is convex.

Theorem 5 *$\mathcal{G}$ is convex in $\mathbf{x}$ both for $\alpha \geq 0$ and when α is the Lagrange multiplier chosen to satisfy equation (11).*

Thus, through appropriate choice of α, the continuous formulation can be used to perform three different minimizations:

- Minimize $\mathcal{G}$ subject to $\mathcal{D}_{max} \leq N$
- Minimize $\mathcal{G}$ with $\mathcal{D}$ weighted by $\alpha > 0$
- Minimize $\mathcal{G}$ subject to $\mathcal{D} = D^*$.

4.3 Scaling of Continuous Solutions

Suppose we have obtained a set of optimal x_n for a particular probability density function $g(x)$ and wish to find the optimal y_n for a scaled density $g'(x) = \kappa g(\kappa x)$. This situation sufficiently large M. Otherwise there must be some location area set with non-zero q_n but infinite cardinality k_n and $E[L]$ will be infinite. $E[L]$ may be rewritten as

$$E[L] = \sum_{n=1}^{N} s_n q_n \tag{5}$$

and we seek s_n which minimize it. It is also possible to add a function which penalizes large D by minimizing

$$G = \sum_{n=1}^{N} (s_n + \alpha n) q_n \tag{6}$$

where $\alpha \geq 0$ is defined as the delay weighting factor.

Notice that equations (5) and (6) are both linear superpositions of incremental costs of the form $\chi_n q_n$. Thus, using boundary conditions $s_N = M$ where M is the number of nonzero p_i and $s_0 = 0$, the problem may be solved numerically using standard finite-horizon dynamic programming [16].

Now consider the problem of minimizing L subject to a constraint on D. Specifically,

$$\begin{aligned} \text{minimize} \quad & E[L] \\ \text{subject to} \quad & s_n > 0 \\ & E[D] = D^* \end{aligned}$$

This problem is *not* amenable to solution via dynamic programming owing to the constraint on $E[D]$; i.e., the total cost is not additive. Specifically, although the cost is still composed of increments depending only on the s_n and q_n, if the delay constraint is not met, then we impose an effectively infinite cost for infeasibility. However, we can reformulate all the constrained problems using continuous distributions. The resulting solutions provide an underbound to the achievable L^*, and in addition, offer a means of obtaining an approximate solution to the discrete problem.

Consider then a non-increasing probability density function $g(x)$ defined for $0 \leq x \leq X$ and comparable to the non-increasing discrete distribution. We define $\mathcal{L}$ as

$$\mathcal{L} = \sum_{n=1}^{N} x_n \int_{x_{n-1}}^{x_n} g(\omega) d\omega \tag{7}$$

where the $x_n \geq x_{n-1}$ are analogous to the s_n for the discrete case and $x_N = X$. Likewise, we define

$$\mathcal{D} = \sum_{n=1}^{N} n \int_{x_{n-1}}^{x_n} g(\omega) d\omega \tag{8}$$

As an aside for completeness, notice that we can make the analogy to the discrete case as precise as necessary by setting $x_n = s_n \delta$ for some $\delta > 0$. Therefore, the discrete theorems which relate L and D for various location distributions via stochastic ordering

This result has the following simple corollary.

Corollary 1 *If $X \overset{\text{st}}{>} Y$, then $E[L(X)] > E[L(Y)]$ and $E[D(X)] > E[D(Y)]$.*

This permits us to find the finite distribution with the poorest performance (maximum $E[L]$ and $E[D]$) for any given paging strategy.

Corollary 2 *Given any location area set $\{A_n\}$, the uniform distribution, $P\{U = i\} = 1/M$ for $i = 1, \dots, M$, maximizes both $L(X)$ and $D(X)$ over all location random variables X having at most M non-zero elements. Thus, the uniform distribution affords the worst L and D performance of any finite distribution with M elements.*

The proof follows directly from Theorem 2 and the following lemma.

Lemma 1 *Over all location random variables X such that $P\{X > M\} = 0$, the uniform random variable U has maximum stochastic order; i.e., $U \overset{\text{st}}{\geq} X$.*

4 Minimizing Paging Costs

4.1 Unconstrained Delay

Here we prove that $E[L]$ is minimized by sequential search over the user locations in decreasing order of probability; i.e., $k_n = 1$ for all n. We likewise show that D is maximized by this choice of the $\{k_n\}$.

Theorem 3 *For an LRV X, Searching locations sequentially in decreasing order of probability minimizes the expected number of locations searched over all possible choices of location area set $\{A_n\}$. Thus,*

$$L^* = \min_{\{s_n\}} E[L] = \sum_{n=1}^{\infty} n p_n = E[X]$$

Theorem 4 *For an LRV X, the ordered sequential paging algorithm of Theorem 3 maximizes D over all choices of location area set $\{A_n\}$ which satisfy $A_n = \{s_{n-1} + 1, s_{n-1} + 2, \dots s_n\}$; i.e., less probable not searched before more probable. Thus,*

$$D^* = \max_{\{s_n\}} E[D] = \sum_{n=1}^{\infty} n p_n = E[X]$$

4.2 Maximum, Weighted and Mean Delay Constraints

Here we seek to minimize $E[L]$ while fixing the total number of location area sets, N. Notice that in this case the distribution p_i must be finite; i.e., $p_i = 0$, $i > M$ for some

$P\{D=n\}=q_n$ and that

$$E[D]=\sum_{n=1}^{\infty} n q_n \tag{4}$$

Our basic problem will be the minimization of $E[L]$ subject to constraints on D or $E[D]$ over the set of polling strategies.

3.2 General Results

We first establish the following results. Proofs are deferred until Appendix A.

Theorem 1 *To minimize $E[L]$ or $E[D]$, more probable locations must not be searched after less probable locations. Formally, if i and j are locations with $p_i > p_j$, then the location area sequence $(A_1, A_2, \ldots)$ that minimizes either $E[D]$ or $E[L]$ must satisfy $i \in A_l$ and $j \in A_m$ for some $l \leq m$.*

Since the ordering of locations within a location area A_n does not affect L or D, Theorem 1 implies that we need only consider orderings of the location distribution $\{p_i\}$ which are decreasing. We define random variable X to be a *location random variable* (LRV) if X takes on values from the positive integers and $P\{X=i\} \geq P\{X=i+1\}$ for all $i \geq 1$. The remainder of this work will consider only the paging problem for which the user location X is specified by an LRV.

In addition, Theorem 1 implies that given an LRV X, an optimal paging strategy that minimizes either $E[D]$ or $E[L]$ is of the form $A_1=\{1,\ldots,s_1\}$ and $A_n=\{s_{n-1}+1,\ldots,s_n\}$. That is, for the appropriate choice of $k_1, k_2, \ldots$, we should first page the k_1 most probable locations, followed by the next k_2 most probable remaining locations and so on. A few theorems relating L and D achievable by different distributions follow. We will make use of the following definitions:

> **Definition:** Let the complementary distribution function of a random variable X be $\overline{F}_X(i)=P\{X>i\}$. A random variable X is said to be stochastically greater than random variable Y, written $X \overset{\text{st}}{>} Y$, if $\overline{F}_X(i) > \overline{F}_Y(i)$ for all i. Likewise if $\overline{F}_X(i) \geq \overline{F}_Y(i)$ for all i then $X \overset{\text{st}}{\geq} Y$.

> **Definition:** Let $L(X)$ and $D(X)$ be random variables associated with the paging cost and delay respectively for a given paging strategy on location random variable X.

Theorem 2 *Given a paging strategy $(A_1, A_2, \ldots)$ and two location random variables X, Y respectively, if $X \overset{\text{st}}{>} Y$ then $D(X) \overset{\text{st}}{>} D(Y)$ and $L(X) \overset{\text{st}}{>} L(Y)$; i.e., increasing stochastic order of the location distribution increases the stochastic order of both L and D.*

sequential polling in decreasing order of probability minimizes this mean. Then, delay constraints are introduced; i.e., maximum and mean constraints are imposed on the number of polling events. The maximum constraint and weighted mean cases can be solved via dynamic programming. However, minimization subject to a mean constraint is not amenable to dynamic programming solution. Therefore, a continuous formulation is developed so that variational [14], [15] principles can be applied. The continuous formulation can also be used to approximately solve the maximum and weighted delay problems. We then apply the results to a simple location probability distribution which arises for a number of motion models and to the worst-case uniform distribution.

3 Preliminaries

3.1 Definitions and Problem Formulation

We enumerate the paging locations by $1, 2, \ldots$ such that the user is at location i with probability p_i. We can associate a user location with a random variable X such that $P\{X = i\} = p_i$.

Location areas are disjoint sets of locations whose members are to be polled simultaneously Let A_n be the set of location indices covered by location area n. The subscript n denotes the order in which the location areas will be searched so that a polling strategy $\mathcal{A}$ is an ordered sequence $(A_1, A_2, \ldots)$ of location areas to be paged. We will use k_n to denote the cardinality of A_n.

The probability of a user residing in location area A_n is then

$$q_n = \sum_{i \in A_n} p_i \tag{1}$$

If the user is found in location area A_n, then the number of locations searched to find the user is

$$s_n = \sum_{j=1}^{n} k_j \tag{2}$$

Therefore, we can now define the cost of paging, L, as the number of locations searched to find the user. We observe that $P\{L = s_n\} = q_n$ and that

$$E[L] = \sum_{n=1}^{\infty} s_n q_n \tag{3}$$

Since all locations within a location area are polled simultaneously, the paging delay D equals number of location *areas* searched before the user is found. We note that

1 Introduction

Paging and registration are necessary features of wireless communication networks because user locations vary as a function of time. Since paging and registration impose burdens on both the switching system and radio resources[1,2,3], some effort has been devoted to minimization of their use [4]-[11]. However, minimization of paging and registration consists of several distinct fundamental problems. Former work has blended these problems and therefore obfuscated them slightly.

Specifically, optimal paging and registration, whether explicitly stated or not, is predicated on location estimation based on some notion of user location probability. It therefore makes sense to explicitly separate the paging, registration, and probability distribution estimation problems. Three basic questions result:

1. Given a probability distribution, what is the least average amount of effort necessary (number of locations searched) to find a user? What is the effect of delay constraints?

2. Given a time-varying distribution known both by the user and the system, what are the optimal paging procedures based on information available at the mobile? I.e., location-based, timer-based or "state" based registration/paging.

3. How can these time-varying location probabilities be efficiently estimated based on measurement and/or models of user motion?

In this paper we consider the problem of item 1 and derive optimal and near-optimal paging strategies for minimizing the average number of locations (base-stations) which must be polled. We assume that some probability distribution on user location can be provided either through measurement or analysis of motion models. Even for a uniform distribution where the user is equally likely to be anywhere in the coverage area, reductions of at least 50% can be had in the average number of locations polled. For other distributions the improvement is larger. Of course, the improvement comes at the cost of increased delay; i.e., not all locations are polled at once. However, even under relatively strong delay constraints, substantial improvements can still be had.

This work provides a foundation for studying the joint optimization of paging and registration [12,13] as well as motivation for future work addressing question 3.

2 Overview

After formally specifying the problem and establishing some general results, we consider the minimization of mean locations polled without a delay constraint. It is shown that

Minimizing the Average Cost of Paging Under Delay Constraints

C. Rose & R. Yates

May 24, 1995

Department of Electrical and Computer Engineering
Rutgers University
PO Box 909
Piscataway NJ 08855
email: crose@ece.rutgers.edu, ryates@ece.rutgers.edu

ABSTRACT

Efficient paging procedures help minimize the amount of bandwidth expended in locating a mobile unit. Given a probability distribution on user location, it is shown that the optimal paging strategy which minimizes the expected number of locations polled $E[L]$ is to query each location sequentially in order of decreasing probability. However, since sequential search over many locations may impose unacceptable polling delay, D, optimal paging subject to delay constraints is considered. It is shown that substantial reductions in $E[L]$ can be had even after moderate constraints are imposed on acceptable D (i.e., $D \leq 3$).

Since all methods of mobility management eventually reduce to considering a time-varying probability distribution on user location, this work should be applicable to a wide range of problems in the area. most notably those with additive cost structures.