Zero configuration adaptive paging (zCap)

Per Kreuger and Daniel Gillblad
Swedish Institute of Computer Science (SICS)
Box 1263, SE-164 29 Kista, Sweden
Email: {Per.Kreuger,Daniel.Gillblad}@sics.se

Åke Arvidsson
Ericsson AB
SE-164 83 Kista, Sweden
Email: Ake.Arvidsson@ericsson.com

***Abstract*—Today, cellular networks rely on fixed collections of cells (*tracking areas*) for handset localisation. This management parameter is manually configured and maintained and is not regularly adapted to changes in use patterns. We present a decentralised approach to localisation, based on a self-adaptive probabilistic mobility model. Estimates of model parameters are built from observations of mobility patterns collected online using a distributed algorithm. Based on these estimates, dynamic local neighbourhoods of cells are formed and maintained by the mobility management entities of the network. These neighbourhoods replace the static tracking areas used in current implementations by using the tracking area list facility of LTE. The model is also used to derive a multi phase paging scheme, where the division of cells into consecutive phases is optimal with respect to a set balance between response times and paging cost. The approach requires no manual tracking area configuration, and performs localisation efficiently in terms of number of location updates, page messages per localisation request and response times.**

I. Introduction

A. Mobility management in cellular networks

In LTE, the wireless part of cellular systems is confined to connecting users and their closest *cell* or small group of adjacent cells (which we here, for brevity, will refer to simply as a cell). To this end user terminals, *user equipments* (UEs), keep track of their closest cell at all times whereas network nodes, *mobility management entities* (MMEs), keep track of sets, *tracking areas* (TAs), of possible cell associations.

The cells of the network are statically partitioned into TAs and regularly broadcast their respective TA identifiers (TAIs). UEs compare the TAI of their closest cell to their current TAI lists. If a UE cannot find the broadcasted TAI in its TAI list, it will perform a TA update (TAU) to inform the MME about its new location and to obtain a new TAI list. To find the exact cell at which a UE resides, the network broadcasts page messages in all cells of all TAs in the TAI list of that UE and detects the cell from which a response is received.

Noting that large TAs and long TAI lists reduce the need for TAU messages while small TAs and short TAI lists reduce the need for paging messages, it is clear that optimal TAs and TAI lists are a matter of striking a balance between TAU messages (which imply large TAs and long TAI lists) and paging messages (which imply small TAs and short TAI lists). As noted by many authors, [1]–[3], these choices can be critical to avoid problems like massive signalling at TA borders.

While the above description applies to LTE systems, similar concepts were used in 2G (GSM/GPRS *etc.*) and 3G (WCDMA/HSPA *etc.*) systems, where the network nodes and location sets are known as MSCs (mobile switching centres) and LAs (location areas) for circuit switching and SGSNs (serving GPRS service nodes) and RAs (routing areas) for packet switching respectively. A major difference, however, is that the TAI *lists* in LTE UEs correspond to *single* LAIs/RAIs in 2G/3G UEs. In our proposal, the TAI lists (or similar) facility is essential to represent the dynamic local neighbourhoods we use instead of static tracking areas for paging.

B. Optimised mobility management

Several alternative update schemes and paging strategies have been proposed. The updating schemes in [4] and [5] propose that networks trace two and three static cell sets (LA/RA) respectively while a completely different approach is taken in [6]–[10] where static cell sets are replaced by or combined with a maximum allowed distances Δ from the last reported location ℓ.

Paging strategies based on ranking cells in a cell set (current LA/RA or cells less than Δ away from ℓ) according to assumed likelihood of response is proposed in [11]–[16] (likelihood estimated from UE specific statistics) and [17], [18] (estimated from current UE direction and/or velocity).

The goal of the present work is to optimise updating and paging in 4G systems with respect to system efficiency and management complexity. Comparing LTE to the earlier proposals, it is noted that the two or three cell sets discussed in [4] and [5] have become (up to) 16 cell sets through TAI lists while the distance based approach [6]–[10] has not been adopted at all. The use of UE specific information for paging as in [11]–[18], is not only complex to collect and maintain, but also comes with significant privacy issues. Proposals requiring UE modifications are generally hard to push through in practice.

The approach taken in this work is to construct dynamic *local neighbourhoods* (LNs) based on collective mobility patterns as observed through successful paging attempts logged in the MME or similar. The LNs, which are unique for each cell, are used to form and maintain the TAI lists distributed to UEs. Update rates in this approach depend only on the the probability mass represented by the cells denoted by each TAI list in our model. For any given list, page sequences are derived on cell level with the object of balancing the expected no.

978-1-4673-1881-5/12/$31.00 ©2012 IEEE

of page messages required against response times in a multi stage paging scheme. The resulting method, which is patent pending, is implementable within present standards and offers a self tuning and distributed method to obtain efficient TAI lists and optimised paging of the cells denoted by these lists.

C. Outline

The details of the proposed mechanism are presented in section II. In II-A we discuss how to maintain distributed traces of cells recently associated with specific UEs, in II-B we describe how to propagate UE observations along these traces and how to update counters to support local estimates of UE mobility, and in II-C we show how these counters can be used to maintain, for each cell and time frame, a number of Bayesian estimates of the probability of UEs having moved to other cells. In II-D we define how to discount older data at a rate which captures both stable and dynamic properties of local mobility patterns, in II-E we use the mobility estimates to form LNs which are transferred as TAI lists to UEs and in II-F we compute optimal page sequences for any given LN, time frame, and current estimate. Finally, section III provides a summary of some empirical results obtained from an implementation of the method in a relatively straightforward scenario after which we conclude in section IV.

II. Main mechanisms of the proposed method

A. Distributed trace and LN records

In order to maintain estimates used to compute and update the LNs we maintain a short distributed trace per UE of the cells with which the UE has been associated through location updates, connections and hand-overs. At each such event, the time and previous cell for the UE is recorded at the MME of the new cell. The current LN of that cell is associated with and transferred to the UE as a TAI list.

This "chain" of previous association records constitutes a distributed, short term location trace of the UE which, after a time related to the periodic location update (PLU) timer, is purged from the MME. From a privacy (and management) perspective, it is noted that neither long term individual associations nor any profiling of individual UE behaviour are required.

Due to the limit (16) on the size of the TAI list in LTE, we propose using small groups of *e.g.* 7-19 adjacent cells as static TAs, and forming the LNs as the union of the cells in TAs referred to by TAI lists. This will yield LNs consisting of up to a few hundred cells. If larger LNs are desired, some more involved mechanism to construct the static TAs, may be used. In future realisations this type complication can be avoided by using longer TAI lists and identifying TAs with single cells.

B. Observation propagation

UE mobility data is collected and distributed through the network using a distributed algorithm as follows:

At each successful page of a UE U at cell C_j, each cell C_i in the distributed trace for U starting at C_j is recursively informed of the observation. The MME of each cell C_i where

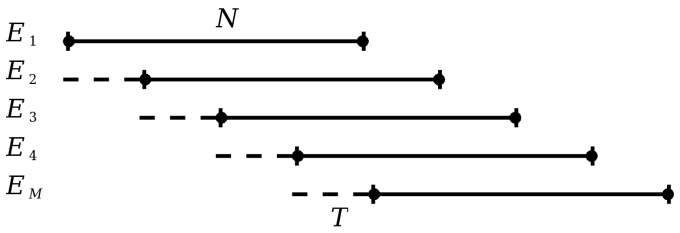

Figure 1. Overlapping estimators

U has previously been observed at time t_i, on receiving the observation (t_j, C_j) updates, if $t_j - t_i \leq \langle\text{PLU time}\rangle$, counters n_i^k and n_{ij}^k for $(t_j - t_i) \in k$, where k belongs to a partitioning of $(0, \langle\text{PLU time}\rangle]$ into distinct time frames. I.e. for each cell C_i we maintain, for each time frame k, one counter n_i^k and one n_{ij}^k for each other (known) cell C_j. With a PLU timer of two hours we may *e.g.* use, $k \in \{(0,2], (2,5], (5,15], (15,45], (45,120]\}$ minutes.

PLU times, the number of and durations of time frames can be local to each cell, and, if desirable, to types of UEs. If PLU times differ between cells, some care need to be taken to account for this when purging association records. The only overhead associated with this scheme is MME memory and fairly minimal and easily automated management.

C. Bayesian estimation

The observation counters n_i^k and n_{ij}^k are parameters in the following estimation of local user mobility.

For each cell C_j known to a cell C_i and each time frame k, we estimate the probability p_{ij}^k, of an UE associated with C_i at t_i to be located at C_j at a later time t_j s.t. $t_j - t_i \in k$ by

$$p_{ij}^k = \frac{q_{ij}^k \alpha + n_{ij}^k}{\alpha + n_i^k}$$

where n_{ij}^k is the number of observations of UEs at C_j associated with and recorded by C_i for k, n_i^k is the total number observations of UEs recorded by C_i for k, q_{ij}^k is a prior expectation of a UE to move to C_j from C_i for k. E.g.

$$q_{ij}^k = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

for zero knowledge of the network topology and where α is a weight reflecting our confidence in the prior.

Note that we compute separate estimates for each time frame k and use the same prior for all of them. For faster convergence, a prior expectation encoding known distance relations between cells, and separate priors for each time frame may also be used.

D. Parallel evolving estimates

In order to account for long-term development of mobility patterns we use multiple overlapping estimators E_η for UE destination distributions, and use the corresponding, previous model as prior, as illustrated in figure 1.

The mechanism is implemented as a circular estimation scheme using $M = N/T$ models, each based on N observations with T degree of overlap and where the ηth estimate is

$$p_{ij}^{k\eta} = \frac{p_{ij}^{k(\eta-M)}\alpha + n_{ij}^{k\eta}}{\alpha + n_i^{k\eta}}$$

while the first M priors are

$$(\forall \eta < M)\, p_{ij}^{k(1-M)} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

Note that each estimator represents the state of the observation counters n_{ij}^k and n_{ij}^k (of section II-C) for any given observation period η. This means that the first M estimates will be calculated with the zero knowledge prior q_{ij}^k as in section II-C while all following estimates will use the corresponding, earlier estimate based on a complete set of N observations as prior. We increment η after a fixed number T of observations which gives us a continuously evolving current estimate that can be tuned to discount older information at any rate we choose. A similar mechanism was presented for another network management domain in [19].

E. Local Neighbourhoods

On each association (location update, (dis)connection, handover etc.), a new LN for each time frame k is computed using the current estimate as in II-D, using either a cut-off residual probability or by selecting the D most likely UE destinations up to a maximum LN size. The LN is transferred to the UE as a TAI list and a record of the association of the UE with this LN is kept until the next location update, connection or handover. The restriction, in LTE, of having to express the LNs as (short) TAI lists referring to static TAs is suboptimal, but still a substantial improvement on current practice. Note that, in contrast to current and previously proposed paging mechanisms, the LNs may be local to each cell and the LNs of nearby cells can be expected to overlap. This reduces the risk of signalling overload at TA borders.

F. Optimal page page phase partitioning

Upon receiving a request for UE U at time t which was last associated with a cell C_i at time t_i, the MME retrieves the LN L_U^k for U and $t - t_i \in k$, pages each cell $C_j \in L_U^k$ in a sequence of phases. Paging all cells in a single phase would minimise response time (since success in an early phase eliminates need for further paging) but maximise number of paging messages sent (since all cells are affected). We thus prefer to, for any given number H of phases, page in the most likely ones first in order to reduce the expected number of paging messages.

The probability of success in a phase h, $h : h = \{1, \ldots, H\}$, is a function of the cumulative probability mass of all cells paged in h and the probability of reaching phase h is the cumulative probability mass of all cells $h, \ldots, H$ not yet paged. Associating costs for paging messages and response times, optimal partitioning of cells into phases can be expressed as a set cover problem which is solved by a constraint program using a global cardinality constraint.

We express the cost of a partition of the cells in a LN L_U^k in terms of boolean variables b_{jh}^k representing the fact the that the cell C_j is paged in phase h. Since each cell $C_j \in L_U^k$ is paged exactly once,

$$\sum_{0 \leq h \leq H} b_{jh}^k = 1 \tag{1}$$

for each j and k. We express the cost of delaying the paging in a cell C_j to phase h as

$$\sum_{j:C_j \in L_U^k} w_h b_{jh}^k p_{ij}^k \tag{2}$$

where p_{ij}^k is the estimated probability that U is in C_j at time t and w_h is a weight specific to phase h. w_h should be 0 for $h = 1$ and grow with h to reflect our estimate of the delay cost. The delay cost for a complete partition is the sum over all phases. We also express a discounted paging cost for a phase h as the number of page messages c_h performed in that phase minus a discount based on the probability of having successfully located U in an earlier phase:

$$c_h - \sum_{1<g<h} \sum_{j:C_j \in L_U^k} b_{jg}^k p_{ij}^k \tag{3}$$

and summed over all phases for a complete partition. The cost function for the set partition problem is a weighted sum of these two components (equations 2 and 3) under the constraint equation 1 for all j and booleans b_{jk}^k. Note that this optimisation problem takes into account only the expected number of page messages required to locate a user U and the response time. The particular balance between these opposing objectives is determined by the choice of weights and does not take into account cost for location updates. The number of location updates depends only on the choice of cut-of probability for the inclusion of a cell in LN, which is an input parameter to the optimisation algorithm.

For a small number of phases and medium sized LNs the solver provides proven optimal solutions. We note that H should in any case be kept small, on the order of 2–4 say, since each phase takes a significant amount of time, which contributes to response time. For larger LNs approximations of the optimal page sequence can be used. In the paging stage we are not restricted to paging cells in groups corresponding to static TAs but can base the sequence on individual estimates for each cell. This reduces, to some extent, the disadvantage of using short TAI list noted above.

Re-computation of page sequences is still fairly costly computationally, and for this reason, we filter triggering events by requiring a sufficiently large (Kullback-Leibler) divergence of the current estimate to the set of probabilities supporting each LN. Thus accuracy and timeliness of the estimate can be traded for computational complexity.

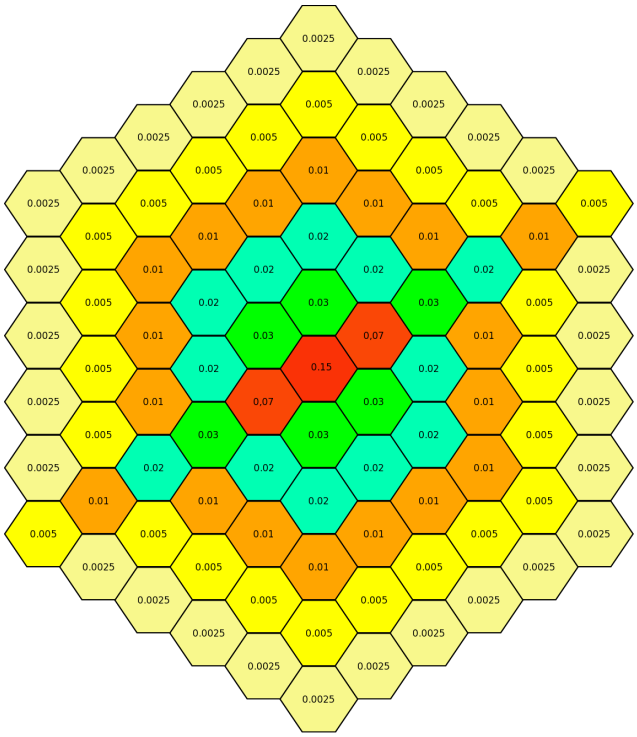

Figure 2. A network with 91 nodes with next observation probabilities relative to the centre node.

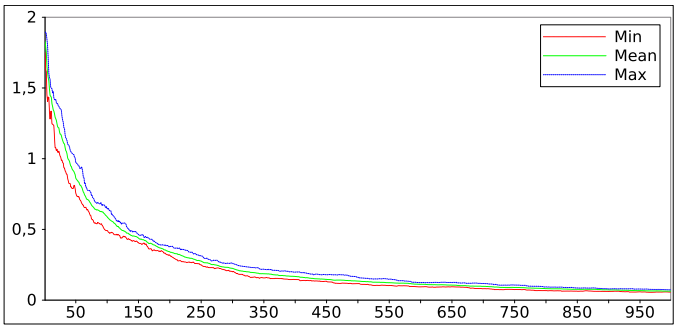

Figure 3. Estimated information gain from current estimate to stationary distribution used for simulation for a 91 node sample set and measured over 5 consecutive runs of 999 observations.

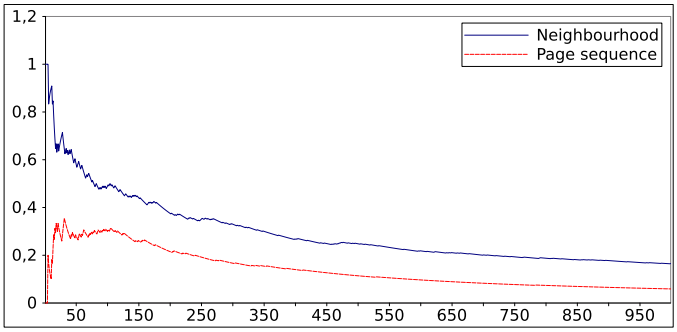

Figure 4. Probability of LN and page sequence updates as a total of 999 observations of a 91 node sample set accumulate during simulation. This is from a single run with 3 page phases and a maximum LN size of 32.

III. Experimental results

To verify the correctness and scalability of the proposed method we have undertaken a series of experiments using a fairly complete implementation of the method and a simple simulator. The scenarios are based on UEs registering at a given cell or group of cells A and then being observed exactly once at any cell B out of a given number of other nodes according to a given stationary probability distribution. An example of the type of distributions used is illustrated in 2 where A is in the centre and B can be anywhere.

The distribution is designed to capture the dependence on distance from the first node A with two directions having a slightly higher probability than others as expected, *e.g.*, around a major road. For the sake of simplicity we assume here that all of this happens with in one single time frame k and that the cut-off for inclusion in the TAI list is 0 but that it instead has a maximum size of 64.

Figure 3 illustrates the convergence of the estimated distribution built up inside the node where the UE is registered, to the stationary distribution used in the experiment. As we can see the proposed method converges nicely and for this sample set size (91 cells), errors become insignificant after around 1000 observations (page responses).

As observations accumulate, nodes gradually update their LNs to include the most probable, other cells that its UEs are likely to be observed at. This means that, as information about the distribution is built up, the LN will first grow to a fixed maximum size and then fluctuate slightly as the estimate of the stationary distribution converges. Figure 4 illustrates how the probability of encountering a new LN decreases from 1.0 to about 0.2 over a run of 999 observations.

As mentioned above, page sequence computations may be computationally costly hence new page sequences for existing LNs are computed only when the Kullback-Leibler divergence exceeds a certain threshold. Figure 4 also plots the probability of such updates for a given cut-off divergence.

The kind of gains that can be expected by implementing the proposed method can be illustrated in several ways. In figure 5 we show the number of cells included in each page phase (top) and the the sum of (estimated) probabilities of the nodes included in each phase (bottom). To read the graphs, note that, *e.g.*, at 750 observations phase one (blue) includes nine nodes (top) and the probability of finding the user in these cells is 43% (bottom), phase two (purple) contains 24 nodes (bottom) and the probability of finding the user in this cell is 23% *etc*. Similarly, it can be seen that the 64 cells included in the LN of the centre cell (top) in total cover about 84% of the users (bottom) hence a residual probability of 16% for users who would reside at cells outside the LN of the centre cell or group of cells (hence they have made one or more location updates to obtain new TAI lists).

Another illustration of possible gain is to plot the expected number of page messages in order to reach a randomly selected UE from the sample set. In figure 6 we see that, for this example, the expected number of paging messages for a random incoming call stabilises at around 26.5 (out of the

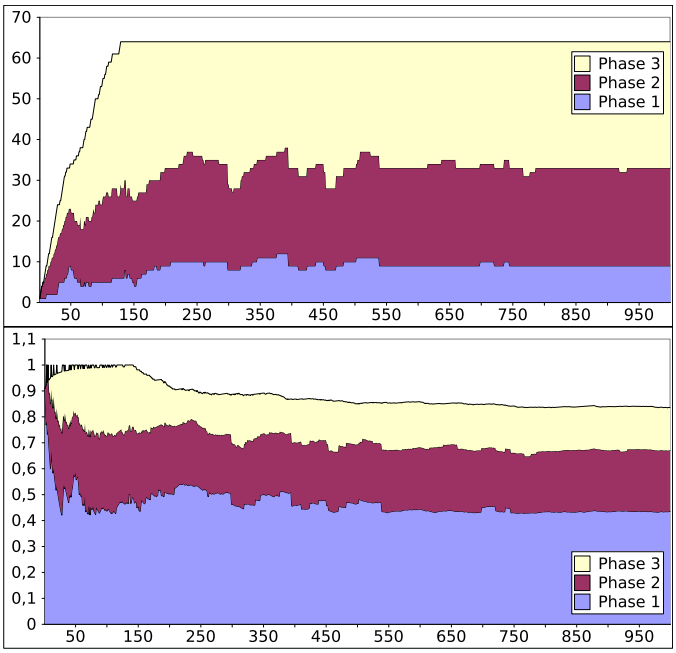

Figure 5. The number of cells and estimated cumulative probability for these nodes in each of 3 phases over 999 samples of a stationary distribution over 91 nodes and a maximum LN size of 64.

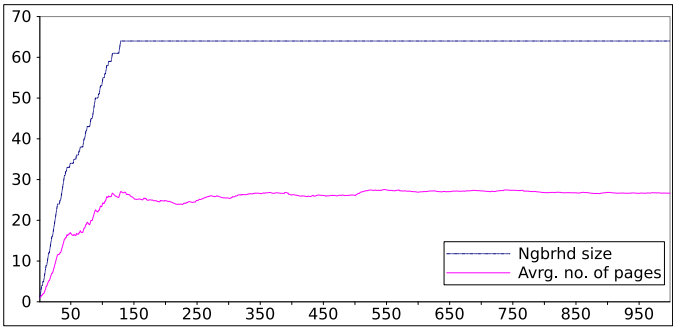

Figure 6. Expected number of page messages for a random connection for 3 phases over 999 samples of a stationary distribution over 127 cells and a maximum LN size of 64.

64 maintained in the LN). This should be contrasted to the currently common situation where paging takes place in all 64 cells included in the LN (TAI list). This means that our method reduces the number of paging messages by 58%. For variants (not shown in the figure) with two phase (shorter response time) and four phases (longer response time) we save 48% and 63% respectively.

IV. Conclusions

We have presented a novel method to automatically manage TAI lists and paging strategies in 4G networks. It relies on a distributed mechanism to compute and maintain local estimates of statistical distributions of UE movements and page sequences for these lists are computed by solving set partition problems using constraint programming. We have also evaluated the proposed mechanism in a simple scenario. We describe the mechanism in the context of an implementation within current a LTE system, but note that lifting the length restriction on TAI list in LTE would allow a somewhat simpler mechanism, and improve the expected gains even further. The mechanism is Patent Pending by Ericsson AB.

For future work we plan to measure the actual gains in much larger scenarios and, if possible, on real data sets.

References

[1] E. Cayirci and I. F. Akylildiz, "Optimal location area design to minimize registration signalling traffic in wireless systems," in *IEEE Trans. Mob. Comput.*, vol. 2, 2003, pp. 76–85.

[2] J. Taheri and A. Zomaya, "A genetic algorithm for finding optimal location area configurations for mobility management," in *IEEE Conf. Local Comput. Netw.* IEEE, 2005, pp. 577–566.

[3] S. M. Razavi, D. Yuan, F. Gunnarsson, and J. Moe, "Optimizing the tradeoff between signaling and reconfiguration: A novel bi-criteria solution approach for revising tracking area design," in *IEEE 69th Veh. Technol. Conf.* IEEE, Apr Spring 2009.

[4] Y.-B. Lin, "Reducing location update cost in a pcs network," *IEEE/ACM Trans. Netw.*, vol. 5, pp. 25–33, Feb 1997. [Online]. Available: http://dx.doi.org/10.1109/90.554719

[5] P. Escalle, V. Giner, and J. Oltra, "Reducing location update and paging costs in a PCS network," *IEEE Trans. Wireless Commun.*, vol. 1, no. 1, pp. 200–209, Jan 2002.

[6] I. F. Akyildiz, J. S. M. Ho, and Y. B. Lin, "Movement based location update and selective paging for PCS networks," *IEEE/ACM Trans. Networking*, 1996.

[7] M. Verkama, "A simple implementation of distance-based location updates," in *IEEE 6th Int. Conf. on Universal Personal Communications Record*, vol. Vol. 1. San Diego, CA , USA: IEEE, Oct 1997, pp. 163–167.

[8] B. Liang and Z. J. Haas, "Predictive distance-based mobility management for PCS networks," in *IEEE INFOCOM'99*, New York, NY, 1999, pp. 1377–1384.

[9] Y. Xiao and K. Wu, "Location update for PCS networks with a fractional movement threshold," in *23rd Int. Conf. on Distributed Computing Systems*. Washington DC: IEEE, May 2003, pp. 825–829.

[10] C. K. Ng and H. W. Chan, "Enhanced distance-based location management of mobile communication systems using a cell coordinates approach," *IEEE Trans. Mobile Comp.*, vol. 4, pp. 41–55, Jan 2005. [Online]. Available: http://dx.doi.org/10.1109/TMC.2005.12

[11] G. Lyberopoulos, J. Markoulidakis, D. Polymeros, D. Tsirkas, and E. Sykas, "Intelligent paging strategies for third generation mobile telecommunication systems," *IEEE Trans. Veh. Technol.*, vol. 44, no. 3, pp. 543–554, 1995.

[12] J. Scourias and T. Kunz, "A dynamic individualized location management algorithm," in *IEEE PIMRC-97*, 1997, pp. 1004–1008.

[13] G. Pollini and I. Chih-Lin, "A profile-based location strategy and its performance," *IEEE J. Sel. Areas Commun.*, vol. 15, no. 8, pp. 1415–1424, Oct 1997.

[14] H.-K. Wu, M.-H. Jin, J.-T. Horng, and C.-Y. Ke, "Personal paging area design based on mobile's moving behaviors," in *IEEE INFOCOM'01*, vol. Vol. 1, Anchorage, AK , USA, Apr 2001, pp. 21–30.

[15] J. Zhang and L. Gruenwald, "Spatial and temporal aware, trajectory mobility profile based location management for mobile computing," in *13th Int. WS on Database and Expert Systems Applications*, ser. DEXA'02. Washington, DC, USA: IEEE Comp. Soc., 2002, pp. 716–720. [Online]. Available: http://dl.acm.org/citation.cfm?id=646130.679831

[16] H. Zang and J. Bolot, "Mining call and mobility data to improve paging efficiency in cellular networks," in *MOBICOM'07*, 2007, pp. 123–134.

[17] G. Wan and E. Lin, "A dynamic paging scheme for wireless communication systems," in *3rd annual ACM/IEEE international conference on Mobile computing and networking*, ser. MOBICOM'97. New York, NY, USA: ACM, 1997, pp. 195–203. [Online]. Available: http://doi.acm.org/10.1145/262116.262147

[18] H.-W. Hwang, M.-F. Chang, and C.-C. Tseng, "A direction-based location update scheme with a line-paging strategy for PCS networks," *IEEE Commun. Lett*, vol. 4, no. 5, May 2000.

[19] R. Steinert and D. Gillblad, "Long-Term Adaptation and Distributed Detection of Local Network Changes," in *IEEE Global Telecomm. Conf. GLOBECOM*. IEEE, 2010, pp. 1–5.