Location Management in Wireless Cellular Networks

Travis Keshav -- traviskeshav@hotmail.com

Abstract

Cellular networks are spreading rapidly, leading to overloaded systems, unacceptable delays, and increasing computational costs due to inefficient Location Management (LM). This paper evaluates both past and present cellular practices and shortcomings in LM, presents recent theoretical proposals to overcome current flaws, indicates probable trends towards future cellular LM, and covers the E911 service and its use of LM.

Table Of Contents

1.0 Introduction

Cells in a network are grouped into Location Areas (LAs). Users can move within these LAs, updating their location with the network based upon some predefined standard. When a user receives a call, the network must page cells within the LA (also referred to as polling) to find that user as quickly as possible.

This creates the dynamics behind much of Location Management (LM), and many of the reports and theories discussed within this paper. The network can require more frequent Location Updates (LUs), in order to reduce polling costs, but only by incurring increased time and energy expenditures from all the updates. Conversely, the network could only require rare LUs, storing less information about users to reduce computational overhead, but at a higher polling cost. Additionally, LAs themselves can be optimized in order to create regions that require less handoff and quicker locating of users. The goal of LM is to find a proper balance between all of these important considerations.

This paper discusses LM schemes, from past and current static LM methods, to current progress and advances in dynamic LM. Much emphasis is placed on the many theoretical implementations of dynamic LM. Also discussed are future LM developments and the Enhanced 911 (E911) service, a real-world example demonstrating applications of LM and its importance.

2.0 Static Location Management

Presently, most LM schemes are static, where LUs occur on either periodic intervals or upon every cell change. However, static LAs incur great costs with the ping-pong effect. When users repetitively move between two or more LAs, updates are continuously performed unnecessarily. In these static LAs, cells are constant in size, uniform, and identical for each user. The current static LM standards are IS-41 and GSM MAP, which use a hierarchical database structure and are described in 2.3.

2.1 Static Location Update

Three simple static Location Update schemes exist in static LM, being always-update, never-update, and static interval-based. The third of these is the most commonly used in practical static LM systems.

One scheme involves the user updating its location upon every inter-cell movement, and is named always-update. This will incur significant energy and computational costs to both the network and the user, especially to the most mobile users. This may be particularly wasteful, as if a user makes frequent, quick movements within an LA, beginning and ending at the same location, many LUs will occur that might be unnecessary, especially if few or no calls are incoming. However, the network will always be able to quickly locate a user upon an incoming call, and extensive paging will not be necessary.

The converse method would be to never require the user to inform the network of inter-cell movements, only updating on LA changes, and is named never-update. In this scheme, resources are saved as constant updates are not required, but paging costs rise substantially. This occurs as every cell within the user's LA may need to be checked during paging due to the lack of information, which causes excessive overhead for users with a high incoming call frequency.

These two schemes are generally unused in real-world systems, but help to provide an illustration to network administrators as to the costs of LM, the problems that occur when thoughtless LU methods are used, and a baseline that every newly developed LU scheme must show improvements over.

The final static LM technique discussed requires each user within the network to update at static, uniform intervals. This attempts to provide a balance between the extremes of the previous schemes, as the network will neither be overwhelmed with LUs nor wholly unaware of users' locations. However, users with rapid rates of movement may move into new LAs between updates, which causes locating that user to be very difficult. Conversely, an inactive user will not move at all, but will still regularly be sending unneeded LUs. [Cowling04] While LA optimization could mitigate these problems, as discussed in the following section, such improvements are impossible under static LM schemes where LAs are uniform and constant.

2.2 Static Location Areas

Location Areas in static LM are themselves static as well. They are effectively the easiest solution to physically dividing a network, providing the same LA to every user, without any customization.

These function as static LU schemes do: suboptimally, but sufficiently for most networks. However, their perhaps most egregious flaw is their vulnerability to the ping-pong effect. Given that these static LAs are set and cannot change, users may repetitively move between two or more adjacent LAs, which for many LU schemes will cause a large number of LUs with a small or zero absolute cell distance moved. Figure 1 demonstrates

such an example, where the user may simply be moving around a city block, but may be incurring an LU on every inter-cell movement due to each movement crossing an LA boundary.
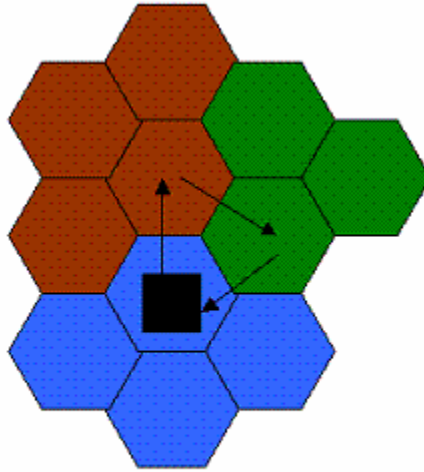

Figure 1: Low absolute movement, high number of LA changes

The optimal static LA size algorithm, which uses a Fluid-Flow mobility model, states that in a network with uniform cell size, cell shape, and user movement speed, the ideal number of cells per LA is

$$N_{opt} = \sqrt{\frac{vC_{lu}}{\pi R C_{pg}}}$$ [Giner04]

where R is the cell radius, v is the speed of a user, C_{lu} is the LU cost, and C_{pg} is the paging cost per call. This equation states that high user speed and LU costs indicate that having a large number of cells per LA is preferable, while a large cell radius and high paging costs imply that a small number of cells per LA is optimal. Obviously, users are not homogeneous, but with sufficient data collection and analysis of aggregate user movement patterns, Fluid-Flow is a relatively successful method to optimize static LAs.

As will later be discussed, dynamic LA schemes can both decrease occurrence of ping-pong effects and provide more flexible and personalized LAs, albeit at their own costs. However, as would logically follow, current static LM standards and schemes do not use these modifications, instead being implemented as seen in the following section.

2.3 Static LM Standards/Implementation

In current cellular telephone usage, there are two common standards: the Electronic and Telephone Industry Associations (EIA/TIA) Interim Standard IS-41, and the Global System for Mobile Communications (GSM) Mobile Application Part (MAP). Both of these are quite similar, having two main tasks of Location Update and Call Delivery.

Currently, a two level hierarchical database scheme is used. The Home Location Register (HLR) contains the records of all users' services, in addition to location information for an entire network, while Visitor Location Registers (VLRs) download data from the HLR concerning current users within the VLR's specific service areas. Each LA has one VLR servicing it, and each VLR is designed to only monitor one LA. Additionally, each VLR is connected to multiple Mobile Switching Centers (MSCs), which operate in the transport network in order to aid in handoffs and to locate users more easily. For LUs, IS-41 and GSM both use a form of the always-update method. All inter-cell movements cause an update to the VLR, while the HLR does not need any modification, as both the MSC and VLR that the user resides in remains constant. Inter-MSC movements within the same LA cause the VLR to be updated with the new cell address, and also cause an update to the HLR to modify the stored value of the user's MSC. Finally, Inter-VLR movements cause the new VLR to create a record for the user, as well as causing an update to the HLR where both MSC and VLR fields are updated. After this occurs, the old VLR's record for the user is removed. Figure 2 displays a symbolic high-level view of the HLR/VLR architecture, as well as demonstrating the methods of communication on a call. [Giner04]
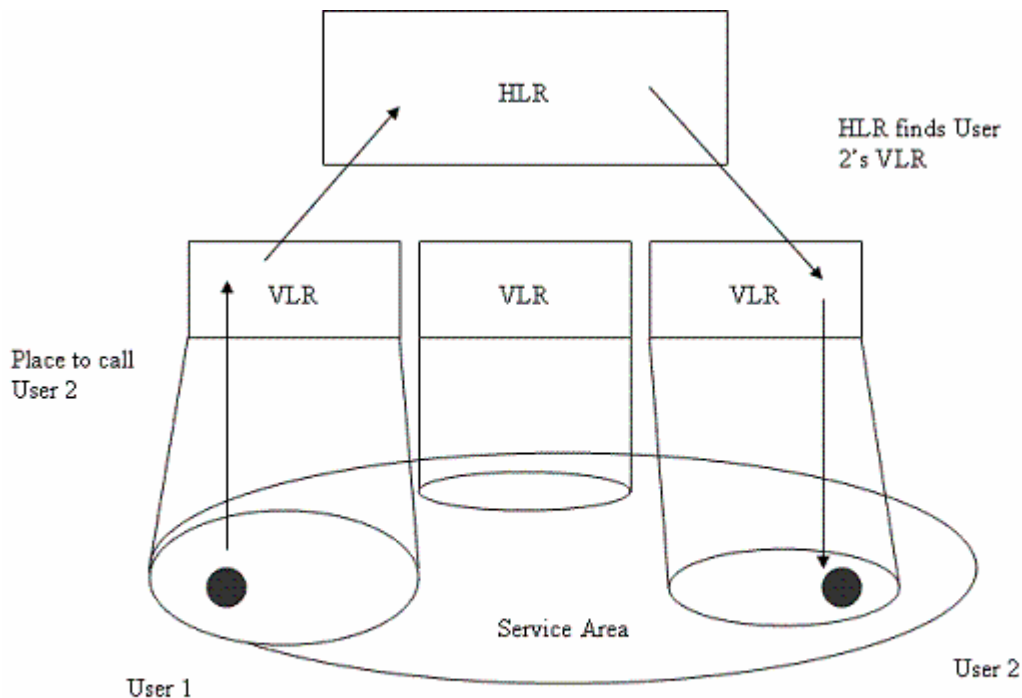

Figure 2: HLR/VLR architecture, Example of Call Process

In Call Delivery, the system database is queried to determine the LA or registration area

of the user. This is split into two steps; a fixed network interrogation is performed to find a region of cells containing the target, in which the HLR is queried to obtain the VLR of the called user. Next, paging is used to poll cells within this region until the user is found. Currently, a system called Signaling System 7 (SS7) is used to transmit these locating messages. In this system, Signal Transfer Points (STPs) route messages through the network (as the HLR), while Service Control Points (SCPs) maintain the databases and information (as VLRs), and Service Switching Points (SSPs) help route calls (as MSCs). Figure 3 provides an illustration that demonstrates how SS7 further aids the HLR/VLR architecture. [Giner04]
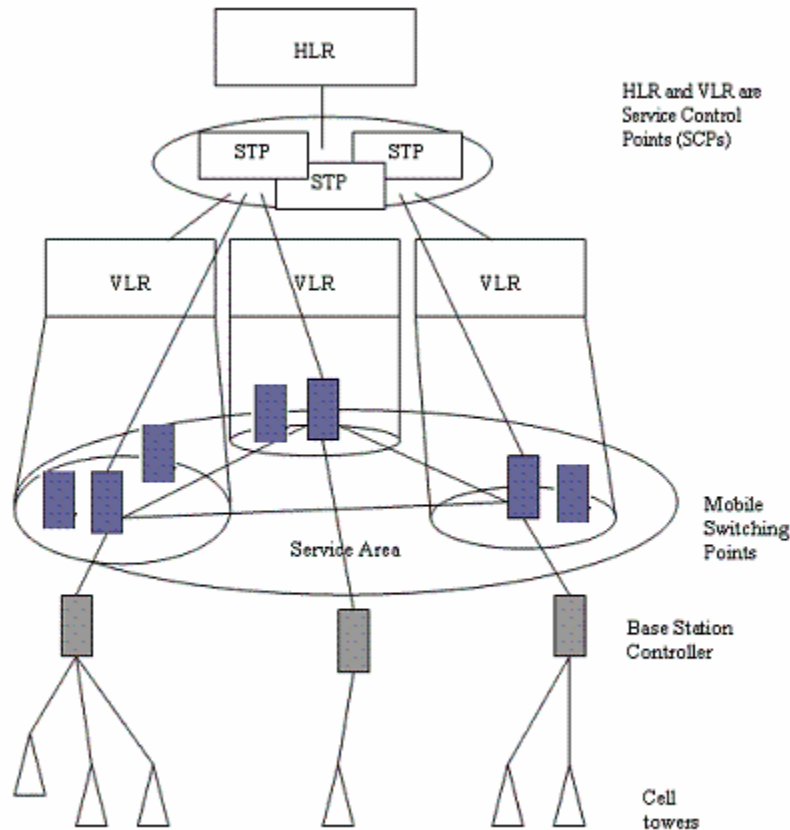
Figure 3: HLR/VLR architecture with SS7

This architecture, while designed and used in static LM, has many theoretical extensions, which will be briefly discussed in 5.6. However, although these architectures and the distinction between static and dynamic schemes comprise much of LM analysis, other parameters in LM must be considered in order to derive and comprehend the optimal overall model.

3.0 Location Management Parameters/Analysis

While evaluating schemes to design LAs or determining the optimal updating standard of LUs for users may seem of higher importance, paging and user mobility can also be

briefly examined to assist in improving LM and refining models. Although LU costs are generally higher than paging costs, these paging costs are not insignificant. Additionally, poor paging procedures and ineffective mobility modeling and prediction may lead to either significantly delayed calls or decreased QoS, neither of which are acceptable to a user.

3.1 Paging

In the attempt to locate recipients of calls as quickly as possible, multiple methods of paging have been created. The most basic method used is Simultaneous Paging, where every cell in the user's LA is paged at the same time in order to find the user. Unless there are a relatively low number of cells within the LA, this will cause excessive amounts of paging. Although this method will find the user quicker than the following scheme of Sequential Paging, the costs make Simultaneous Paging rather inefficient.

An alternative scheme is Sequential Paging, where each cell within an LA is paged in succession, with one common theory suggesting the polling of small cell areas in order of decreasing user dwelling possibility. Unfortunately, this was found to have poor performance in some situations, as if the user was in an infrequently occupied location, not only might every cell be paged, but a large delay could occur in call establishment. Additionally, this method requires accurate data gathering concerning common user locations, which necessitates more frequent LUs and thereby increased costs. Consequently, most real-world Sequential Paging methods simply poll the cells nearest to the cell of the most recent LU, and then continue outward if the user is not immediately found. However, such a method will still be inefficient if the user's velocity is high or an LM scheme is used which specifies infrequent LUs.

As an attempt to improve on previous models, another design called Intelligent Paging was introduced, which calculates specific paging areas to sequentially poll based upon a probability matrix. This method is essentially an optimized version of Sequential Paging. However, this scheme has too much computational overhead incurred through updating and maintaining the matrix, and although perhaps optimal in theory, is effectively impossible for commercial cellular network implementation. Therefore, most current schemes use one of the first two methods presented. [Cowling04] While the best paging methods can significantly speed the locating of a user, significant further improvements are possible by combining them with user mobility predictions.

3.2. User Mobility

For aid in effectively predicting the user's next location, user movement patterns are analyzed and mobility models are designed. Many such mobility models exist and can be used by networks in LM.

The simplest of these models is random-walk, where user movements are assumed to be entirely random. While this is clearly going to lead to inaccurate predictions, it does require no knowledge of the individual user, and can be effective as a simulation tool.

Frequently, random-walk is used to demonstrate the improvements a given scheme makes in comparison to this random method.

A very general scheme, ignoring individual users but considering the network as a whole, is called fluid-flow. This method aggregates the movement patterns of users, and consequently can help optimize the network's utilization and design at a macroscopic level. However, fluid-flow provides no insight on a smaller scale, nor will it give any predictions as to specific user movements for any specific user.

Markovian mobility models also exist, where user movements are predicted through past movements. At large computational cost, every inter-cell movement probability is defined for each user. An extension of the Markovian model, created at perhaps even greater cost, is the activity-based model. In this model, parameters such as time of day, current location, and predicted destination are also stored and evaluated to create movement probabilities. However, for all the resource expenditures required in implementing these methods, in a test of a simple activity-based scheme, unstable results were returned. An even more complex activity-based scheme might provide better results, but would not be implementable on a large scale due to its immense costs. [Cowling04]

In fact, research on all current models shows that none truly does a satisfactory job of predicting user movements, demonstrating the need for further research in this area. Consequently, [Cowling04] describes a possible enhancement as a scheme called selective-prediction, where predictions are only made in regions where movements are easily foreseeable, and a random prediction method is used elsewhere. To further this scheme, Cowling advocates a network where the base station (BS) learns the mobility characteristics of the region, in addition to the cell movement probabilities. This learning network provides the basis of a Markov model, with the full knowledge of movement probabilities and theoretically incurring low overhead.

Additionally, the paper [Halepovic05] provides an in-depth analysis of sample user movement and call traffic. Empirical data gathered in these experiments reveals that the 10% most mobile users account for approximately two-thirds of the total number of calls within the network. Consequently, such users must be given appropriate consideration involving resource allocation. Over half of users appeared to be stationery, and most (but not all) such users generated much less cellular activity. Therefore, only a weak correlation between user mobility and data traffic exists, as users with few cell changes do have a large variance in the number of calls they made. Further tests within the paper reveal that a majority of users have a home cell. This is a location, whether it be an actual home, office, or other place, from which a majority of their calls originate. This characteristic can be exploited by paging techniques and the customization possible in dynamic LM, as less updates may be necessary if a user is within their home area.

4.0 Dynamic Location Management

Dynamic Location Management is an advanced form of LM where the parameters of LM can be modified to best fit individual users and conditions. Theories have been and continually are being proposed regarding dynamic LUs and LAs. Additionally, many are reexamining paging and mobility parameters based upon these developments. Many of these proposals in dynamic LM attempt to reduce computational overhead, paging costs, and the required number of LUs. However, many of these proposals are excessively theoretical and complex, and are difficult to implement on a large scale.

4.1 Dynamic Location Update

Many dynamic LU schemes exist, in order to improve upon excessively simple and wasteful static LU schemes. Additionally, these schemes are intended to be customizable, such that each user will have their own optimal LU standard, greatly reducing the overall number of LU updates.

One of these dynamic LU formats is threshold-based, where updates occur each time a parameter goes beyond a set threshold value. One possible threshold is time, where users update at constant time intervals. This saves user computation, but increases overhead significantly if the user does not move. This time-based scheme is very similar to the common static LU scheme, with the important difference of the time value being modifiable. Another threshold-based scheme requires a user update each time they traverse a certain number of cells. This was found to work better than the time-based scheme, unless the users were constantly moving. In such a case, this method becomes quite similar to the static always-update scheme, where many unnecessary updates might occur. Consequently, a preferable scheme was found, called distance-based. This called for an update only if the user moved a certain radial length of distance. However, this scheme is not perfect, as it requires the cellular device to keep track of such distances, which added much computational complexity. [Cowling04] Figures 4, 5 and 6 provide illustrations of these threshold-based LU methods.
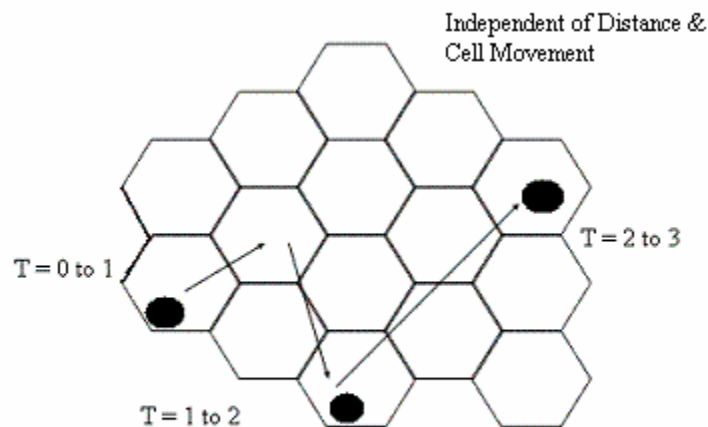

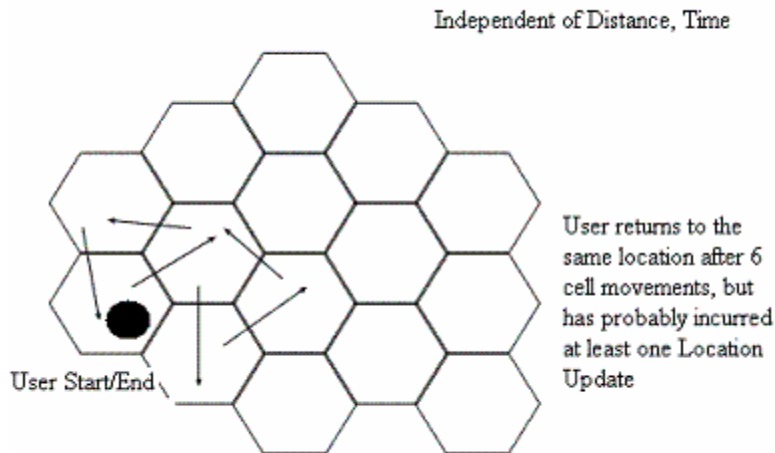Figure 4: Time-based Threshold Location Update

Independent of Distance, Time

User returns to the
same location after 6
cell movements, but
has probably incurred
at least one Location
Update

User Start/End

Figure 5: Cell-based Threshold Location Update

Independent of Cell
Movement & Time

D=2, LU
occurs
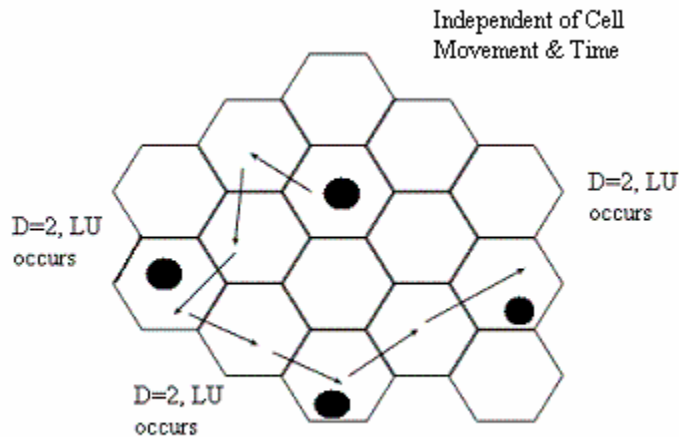
D=2, LU
occurs

D=2, LU
occurs

Figure 6: Distance-based Threshold Location Update

Another dynamic LU scheme is profile-based. This functions by the network compiling a list of the most frequently accessed cells by the user, and only requiring an LU if the user moves outside of these common cells. As would be expected, this scheme is only effective if these predictions can be made accurately and without excessive overhead, but is otherwise inefficient. Additionally, this list must be relatively small, or else paging will become costly.

A more advanced scheme, built upon the efforts of previous methods, is called adaptive. Adaptive LU schemes are very flexible and even may differ from each other, as such schemes are designed to take multiple parameters, such as velocity and mobility patterns, to determine the most efficient LAs. In such an example, having knowledge of a user's past movements combined with the user's current speed and direction allows strong predictive power when determining a possible future location for paging. Therefore, LUs

may not need to be as frequent, thereby reducing the overall LM costs. However, although these adaptive LM schemes are highly successful in terms of reducing LU costs, they are generally too difficult to implement for large networks, requiring excessive computational overhead. [Cowling04] Consequently, dynamic LA schemes must be examined as a possibly preferable solution.

4.2 Dynamic Location Area

While static LA schemes are restrictive and inefficient, dynamic LA designs offer much more flexibility, allowing much more customization. To improve on the past schemes, [Cowling04] proposes several changes to the static LA methodology. Instead of viewing the network as an aggregation of identical cells, it is now viewed as a directed graph, where nodes represent cells, with physical adjacency shown through graph edges. General movement patterns and probabilities, updated at predetermined intervals, are recorded between these cells based upon handoff information. This allows low overhead while still providing predictive power. Additionally, a smoothing factor of k is implemented to allow individual networks to weight new data as desired, where a low k-value causes the network to highly weight new data, and a high k-value causes the network to give precedence to previous data. These adapting patterns can be stored, in order to allow further prediction based upon other data such as time.

A similar parameter examined is dwell time, which is defined as the length of a user's stay within a cell. This is used to dynamically determine the appropriate dimensions of LAs. A smoothing parameter is also used for dwell times to weight past collected data against new data. The preferable method of collecting dwell times is by having the cellular device report its dwell time to the network upon a handoff.

Within a dynamic scheme, LAs, instead of being constant and circular, can be diverse and may take different shapes, in order to be optimal for individual user parameters and network characteristics. As well, cells are organized within these LAs based upon frequency of use, with the most frequently visited cells being placed in an ordered array. This array can be used in conjunction with known physical topology to design optimal user LAs, constructed such that the users will change LAs and make handoffs as infrequently as possible. To further optimize service for the individual user, call interval times and relative mobility rates are calculated to make low-overhead predictions concerning when LA and cell changes will occur.

4.3 User Characterization

The primary goal of dynamic LM is to provide LU and LA algorithms to each individual user that minimizes the costs for them. To do this, call rate and movement factor estimations must be made. To determine the call rate, the network first determines the call interval, by taking the average interval between calls (in seconds), and adding new call intervals in a weighted manner. The call rate, defined as λ, is then calculated by dividing 3600 by the call interval, in order to obtain the number of calls per hour.

Given that cells will be of different sizes and covering different types of areas in dynamic LA, it is incorrect to base user movement properties on raw user cell dwell times. Instead, the movement factor γ is used, where this factor specifies a user's movement speed relative to other users. This factor is further classified based upon specific regions, as the following equation demonstrates:

$$\gamma = \frac{\text{Time Taken for Average User to Move Through Region}}{\text{User Time Taken to Move Through Region}}$$

[Cowling04]

The above equation simply defines a parameter that characterizes user speed, with a large γ value indicating a high-velocity user, and a small γ indicating a low-velocity user. Note that this calculation is per-region rather than per-LA or per-cell in order to avoid the ping-pong effect skewing a user's perceived speed. However, the network must also determine a method of calculating these times. This is solved through using dwell times and data received from LUs on LA changes. Further cost analysis, which factors in the movement factor γ, will be described in the following section.

4.4 Cost Analysis

As seen before, the total cost of LM is equal to the LU cost added to the paging cost. This equation is always true, regardless of which system is used. However, different systems will have different values for these, and the goal is to find an implementable system that minimizes the total.

The paging cost can be fairly simply defined, as the previously calculated call rate λ, multiplied by the number of cells in the paging area, multiplied by a constant representing the cost per paging message. Consequently, reducing any of these parameters will reduce the overall paging cost. These variables can be seen below. However, the LU cost calculation is somewhat more complicated. Although we can simply say that the LU cost is equal to the cost per LU divided by the estimated dwelling time within the current LA, T_{LA}, it is somewhat difficult to calculate this dwelling time. Through analysis and justified by simulation, [Cowling04] defines the average user's cost as the sum of the mean dwelling times within the cell multiplied by the probability of residing within the cell for all cells, as seen below. Therefore, as also seen below, to determine an individual user's T_{LA}, the average T_{LA} is divided by the movement factor γ.

$$C_{TOTAL} = \frac{\varsigma_{LU}}{T_{LA}} + \lambda \cdot \varsigma_P \cdot N_C$$

$$\widehat{T_{LA}} = D_{S_1} \cdot p_{S_1} + D_{S_2} \cdot p_{S_2} + D_{S_N} \cdot p_{S_N}$$

$$T_{LA} = \frac{\widehat{T_{LA}}}{\gamma}$$

[Cowling04]

Additionally, the C_{lu} and C_p costs can be examined in order to determine whether the LAs and system created should favor more or less LUs, in order to determine an appropriate balance between LU and paging costs. In real-world applications, these parameters are based upon the available resources of the specific network provider being examined.

This concludes a fairly comprehensive introduction to conventional static and dynamic LM. However, many new theories have recently been presented. Although these are primarily theoretical, they merit examination in the following sections, as they may be the basis for further developments. Additionally, a comparative analysis between static and dynamic LM will be discussed in 5.1, in order to observe costs and view the relative advantages and disadvantages of realistic methods.

5.0 Dynamic Location Management Analysis and Developments

In the world of dynamic LM, new theories are constantly being proposed to reduce LM costs, or otherwise improve the network quality. Additionally, an experiment was conducted to compare generally static industry standards and a simple dynamic LM scheme, in order to clearly show how dynamic LM is preferable to static LM. The following sections provide this comparative analysis as well as an overview of several of these wide-ranging developments.

5.1 Static/Dynamic Location Management Comparison

The paper [Toosizadeh05] summarizes four commonly used LM methods and compares them through C++ simulation. The first method examined was GSM Classic, an older scheme with fixed location areas, LUs occurring on each LA transition, and blanket polling of all cells on an incoming call. Next was GSM+Profiles, a method containing the same LU scheme as GSM Classic, but where sequential paging is performed on of cell areas of decreasing dwell times. The third is the Kyama Method, where both static and dynamic LAs are used in order to provide easier paging. These dynamic LAs in Kyama are a combination of the standard static LA and an extra LA created dynamically for predicting the user's future movements. LUs occur each time the user moves out of their combined LA. Paging is benefited by using dwell times to determine high probability polling regions within these dynamic LAs. Finally, the dynamic distance-based method was also tested, where LUs occur only on exceeding user movement thresholds, and paging follows the GSM+Profiles paging scheme.

For testing, the authors of this paper ran experiments assuming both a random-waypoint

model, which assumes the parameters speed, pause time, and direction to all be random, and an activity-based model, where users move as if accomplishing normal tasks, such as moving from home to work and back at predicted intervals. The simulated space was 100 cells in a 10x10 grid, serving 100 users. Traffic was statistically generated through Poisson calculation of random traffic.

Results showed that while the GSM and GSM Profile methods incurred the largest LU costs, due to their static LA and LU methods, they had the lowest paging costs. Conversely, the Kyama method, while requiring a relatively low number of LUs, had extremely high paging costs, as their combined static/dynamic LA method proved to be somewhat unsuccessful. Overall, due to having a very low total LU cost, and also having fairly low paging costs, the best method was the distance-based algorithm. Additionally, the activity-based model proved far superior to the random-waypoint model in all performance categories concerning user movement prediction, as would be expected. [Toosizadeh05]

This experiment demonstrates that dynamic LM schemes will generally be superior to static LM schemes, but also that care must be taken when creating dynamic LAs. If proper analysis is not made, these dynamic LAs may become overcomplicated, and thereby can incur excessively high paging costs. While paging costs are generally significantly less than LU costs, it is impossible to state that the Kyama method's low LU costs will entirely offset its paging costs. Regardless, the success of the distance-based method in this experiment demonstrates that proper dynamic LM schemes will generally be preferable to static LM schemes. Furthermore, the distance-based dynamic LM scheme used is relatively simple; the theories presented in the following sections include additional improvements that might lead to even higher performance in tests.

5.2 Further Cowling Theories

[Cowling04] briefly presents several possible developments and hypotheses, which, while believed to be worth researching, have not been implemented in large real-world systems. The first is a timer-based LU in conjunction with other LU schemes. This helps to rectify unnecessarily large LAs set for users moving frequently and unpredictably within a relatively small area of space. Assume a scheme is used where LUs occur only on LA changes. Consequently, a user frequently moving between boundaries of LAs that represent physically large areas would be assigned a large LA. However, if the user did not make further movements that would go outside of this LA, the user would unnecessarily retain the large area, within which paging might be costly. Adding a timer-based mechanism would cause an additional regular update to occur, causing the LA to appropriately shrink, as the network would determine the user's movement habits and optimal LA size on these updates.

One simple thought was to reduce the number and complexity of parameters considered in user movement. Although this would reduce predictive power, it is possible that the corresponding reduction in computational costs would be significant enough to offset these losses. Additionally, considering average user speed, instead of individual user

speed, would give lower computational costs, while still providing some predictive power.

Finally, [Cowling04] suggests further examination of the ratio of LU cost to paging cost. This is believed to be approximately 10:1; however, this is only a rough conjecture. If a precise ratio could be determined, it would allow dynamic LM schemes to determine an optimal LU method, which would update in a manner that minimized overall costs. However, this ratio is often based not only on the size and dimensions of LAs, but also on network-specific packet formats and database resources, which companies do not provide to the public. Consequently, this analysis may not be possible, and so the optimization of the parameters of user movement and mobility, as previously discussed and further used in the following topic, may be preferable.

5.3 Hand-off Velocity Prediction Location Management

[Lam05] proposes a probabilistic approach to paging called Hand-off Velocity Prediction (HVP). This algorithm attempts to efficiently analyze and accurately predict user mobility for paging purposes, based upon user mobility parameters.

In HVP, the paging area for a user is calculated based upon handoff statistics, the location of the last LU, the time since the last LU, and the velocity of the user. HVP then creates handoff graphs, indicating the most likely location of the user. Complex calculations beyond the scope of this paper are used to determine the precise probabilities of user movements, but can be examined in [Lam05]. If these predictions can be made accurately, LUs would not need to be so frequent, as less user data could still have enough predictive power using HVP to avoid excessive paging costs. However, a sequential paging scheme used with HVP may delay some calls, in cases where users move in an unpredictable manner. Therefore, the authors propose a group-paging method to meet any QoS requirements, where multiple cells are polled at once to rapidly find the user.

A network designed for HVP use incorporates both distance-based and time-based LUs. Users are classified into groups, depending on their approximate velocities. These velocities are found by calculating the change of signal strength of the user's cellular device over time, hence the need for time-based LUs. Through statistics regarding handoffs and velocity, the network can calculate acceleration, which allows higher-precision predictions of movement. Probabilistic equations are used to further improve these predictions, by considering the parameters cell size, acceleration, and the number of cells previously traversed. The author's experiments demonstrate a significant decrease in LU and paging costs, although the authors note themselves that this system is very computationally complex.

This last point is unfortunately the same for many Dynamic LM algorithms. While HVP is theoretically successful in simulations and small experiments, giving results superior to those of current systems, it is impractical to be implemented for today's large commercial networks. However, if user movements cannot directly be used to predict handoffs, as in

HVP, these movements can be used for improved LA construction as seen in the next section.

5.4 Set-Covering Based Location Area Planning

[Lo04] advocates designing LAs in a set covering method, specifically creating Set-Covering-Based Location area Planning (SCBLP). The goal of this method is to group the cells into appropriate LAs, facilitating LM by decreasing the LU and paging costs. The network is viewed as a set of location databases, and these databases are used to create a virtual tree structure. This structure provides a model for examining database changes on user movements, where a user effectively leaves one tree branch to move to another on an LA change. Consequently, the costs incurred from the frequently made call locations and destinations can be reduced, by using this set method to determine appropriate organization of nodes within the network. SCBLP makes set-covering specifically linked to LM, and is used to divide the network into multiple LAs, by determining which sets, given movement and handoff data, will require the least number of LUs. This is calculated by recording the number of boundary crossings between cells, and analyzing possible network configurations to determine which will cause the least number of required updates.

Once the best LA partitions are determined, they are further examined in respect to their cohesion, coupling and cost-benefit. Cohesion is defined as the sum of border crossings contained with a particular LA, and is seen in Figure 8. A higher cohesion is desired, as these intra-LA border crossings occurring more often is preferable to border movements causing LA changes. Coupling is defined as the inverse of the number of border crossings defined to other LAs, and is seen in Figure 9. In this case, we desire to minimize the LA changes, and consequently maximize the inverse. Cost benefit is simply defined as the cohesion benefit multiplied by the coupling costs.
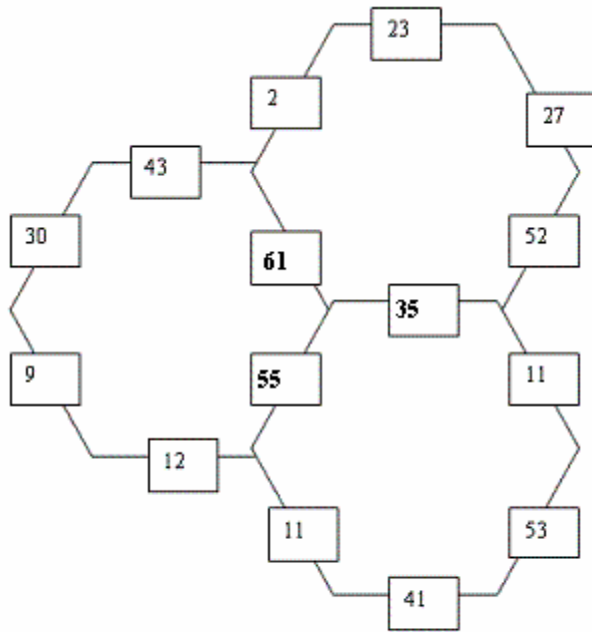
Figure 7: Bolded entries indicate cohesion.
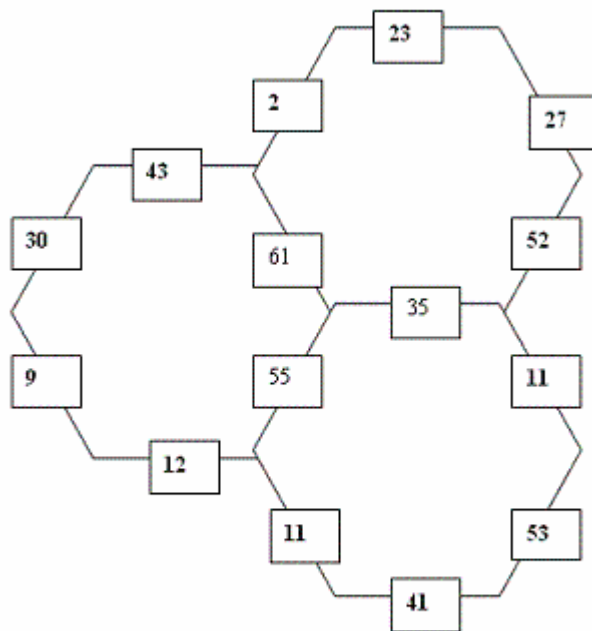Total Cohesion: $61 + 55 + 35 = 151$

Figure 8: Bolded entries indicate coupling values.
Sum of values = 314, therefore coupling is 1 / 314

This algorithm is somewhat complicated, but simulations within the paper demonstrate that mobility management and LM costs are reduced in comparison to common greedy or

random set creation algorithms used to create LAs. [Lo04] However, as possibly effective as this and the previous algorithms may be, they all fail to wholly address the potential risks of the ping-pong effect. The following subject addresses this.

5.5 Triple-Layer Location Management

Although most efforts involving dynamic LM focus on improving LU schemes rather than creating efficient LAs, [Fan02] proposes a scheme where three location layers are used to ensure optimal user coverage. The main issue addressed and solved by this paper is the ping-pong effect.

Traditional (Figure 9) and generalized (Figure 1) ping-pong effects cause a significant number of unnecessary LUs, combining to equal over 20% of the total LU cost. Two insufficient schemes to counter this are the Two Location Area (TLA) scheme and Virtual Layer (VLA) Scheme. The TLA scheme remembers the most recently visited areas, such that the network will know not to update if it is simply moving back and forth between the same two LAs. However, this scheme still has problems with the generalized ping-pong effect. VLA Schemes add a second LA layer to prevent unnecessary LUs from occurring when movements take place within a small region of cells. However, not only is this not fully sufficient, it simultaneously adds complexity that may end up causing unexpected ping-pong effects. Alternatively, extensive cell overlapping can be implemented, but only at a large paging cost.
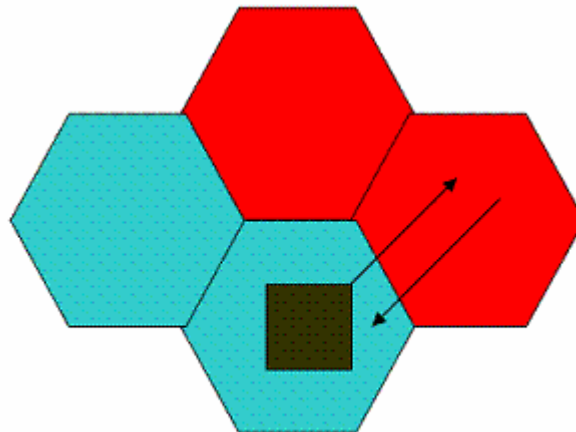

Figure 9: Traditional Ping-Pong Effect

The triple-layer scheme builds upon these, and is designed such that whenever a user moves out of one layer, they are able to selectively register with either of the two other layers, optimally choosing the layer with the closest center to the user. These three layers are also constructed such that layer boundaries will not overlap each other over multiple adjacent cells, and that all three layers will not overlap on the same cell boundary. This occurs in order to ensure that users will be able to connect to an appropriate new layer on handoff. Through several simulations comparing results such as LU costs, overall costs and reduction of ping-pong costs, triple-layer network architecture is shown by the authors to be more efficient than current methods and those previously proposed.

However, computational overhead in this scheme might be high for large networks, and therefore this scheme will not be seen often in real-world applications. Similarly, many of the following theories may also be somewhat unsuitable for real-world implementation, but still provide valuable insight as to the types of modifications that can lead to improvements in LM.

5.6 Additional Dynamic LM Theories

Many other innovative Dynamic LM concepts exist, which although perhaps excessively complex, theoretical, or systems-based to be fully explored, deserve brief mention.

[Lee04] proposes a scheme where LU and lookup costs are balanced, minimizing the overall performance penalty and overhead. Location information is stored within location databases called agents, which themselves are placed within BSs. Each agent contains the address of several users or other agents, thereby allowing for easier forwarding of data within the network. Instead of searching through large tables of entries or large numbers of cells, these agent trees can quickly be traversed, with the addresses stored providing quick reference for locating the recipient of a call. Additionally, through the grouping of users at similar physical locations to similar logical address, simpler paths can be created, reducing the necessary number of agents. However, this algorithm proves effective only for networks with a low call-to-mobility ratio.

[Giner04] proposes several improvements to the commonly used HLR/VLR architecture, mainly involving methods of increasing the speed of finding a called user's VLR. One proposed theory involves using pointer chains within the HLR to store successive user VLR locations, such that updates can be made to the end of the chain, rather than requiring HLR memory to be directly accessed and updated on each movement. A similar scheme involves using local anchoring, where each user has their most frequently occupied LA designated the anchor. If the user moves, the anchor VLR is updated instead of the HLR, again saving HLR resources. Then, upon call arrival, the anchor is checked first, and then the pointer list is traversed to locate the user if necessary. Both of these schemes were shown to only be somewhat effective, working best when calls are infrequent and user mobility is high. Giner suggested using per-user location caching to fix this weakness, storing probable user locations within the MSC, but this scheme fails when users frequently move to new locations.

[Hassan03] suggests that instead using the conventional HLR/VLR architecture, all operations should be routed through several BSs connected by wireless inter-BS links, in a system called cell hopping. This system is not as decentralized as ad-hoc routing, as users do not make connections directly with each other. Users simply roam freely, registering dynamically with the nearest BS, with user requests flowing through these BSs. Cell-Based On-Demand Location Management is used, where Membership Request Query (MRQ) is used with querying and caching to determine the location of a desired user. This process may be slower than paging done in HLR/VLR architecture, but this lightweight scheme requires no central storage system. Although [Hassan03] makes an interesting proposition, it is highly unlikely that current cellular companies would desire

to make this radical change and decentralize to this degree.

Additionally, [Xiao03] proposes a fractional movement-based LU scheme for cellular networks. This paper cites a deficiency of standard cell threshold-based LU methods, in that such schemes only assume VLR updates will occur after the specified number of cells are traversed, ignoring the possibility that an LU will otherwise occur due to the user crossing an LA boundary. Consequently, Xiao introduces a fractional threshold value r in addition the standard value. This is used such that after the standard threshold is reached, an LU is performed on this step with probability $1/r$, or on the next cell movement with probability $1/(1 - r)$. [Giner05] uses a somewhat similar fractional distance-based method. In this scheme, the distance counter is reexamined on every cell movement, and zeroed if a cached cell is reached. Giner uses a fractional threshold value q similarly to Xiao's methodology, where after the standard threshold is reached, an LU is either performed immediately with probability $1/q$, or on the next movement with probability $1/(1 - q)$. Giner also provides further research and analysis as to optimal threshold and q values.

While these improvements and theories provide a sampling of current research, the continued evolution of technology and cellular networks themselves will also cause significant changes to and improvements in LM, as discussed in the following sections.

6.0 Future LM Developments and E911

The world of cellular networks and LM is certainly not static. Even the most recent theories and developments may become outdated within months. With the increased use of 3G Cellular Networks, and the E911 service providing a national standard in which LM is absolutely essential, it is critical to observe what is to come.

6.1 Location Tracking Evolution

Currently, the market standard for location tracking uses the cell ID to track cellular devices through many of the methods as described previously, primarily including the HLR/VLR architecture. However, it is not as precise as many would desire, as this system is not always able to locate users to within a few meters.

Future possibilities for location tracking in cellular systems include Global Positioning Systems (GPS), which use line-of-sight satellite position calculation, giving five-meter range precision, but encounter problems when physical obstructions block the connection. Another possibility is Assisted Global Positioning Systems (AGPS), which is similar to normal GPS, but uses advanced searching techniques to determine user position. Unfortunately, AGPS is quite expensive, and still requires line-of-sight. Finally, there is the potential to use Broadband Satellite Networks, which use the low-earth-orbit satellites to create a global network. These give relatively high precision without line-of-sight, but are very complex to manage. [Rao03]

An example of a current commercial use for GPS is cellular phone games. These games are a larger part of the industry than might be expected, accounting for approximately $4.8 billion of revenue to providers [Rashid06]. In order to approximate the true location of the call, these networks generally use Time of Arrival (TOA) measurements and examine the difference of arrival times for multiple incoming signals, while considering signal transmission location. These calculations can be used with GPS to further resolve user location. However, these methods have not been implemented on a large scale, mainly due to GPS's costs and limitations. Alternative methods include approximating cellular device location based upon the device's relative location to other objects within the cellular infrastructure, but this is generally imprecise and unreliable. Games may seem to be a somewhat trivial example of GPS's usage and of advanced location tracking, but actually includes some methods used in the E911 service, as will be seen later.

Regardless, these are future plans, currently only used for military, private, and small-network settings, and may not fully enter the cellular phone marketplace in the near future. Most of these are excessively expensive or are too difficult to maintain for a large number of cellular subscribers at this point in time. Still, for users willing to pay additional fees for Location-Based Services, such as games as described above, these technologies can be of great use and entertainment value. As technology improves and the transition to 3G continues, perhaps such services will become common.

6.2 Location Management in 3G

3G Cellular Networks, while providing significant improvements over 2G Networks, share many similarities in LM with its predecessor. However, there is one significant difference in the HLR/VLR architecture worth noting.

This new 3G addition is the GLR (Gateway Location Register). GLRs are used to handle VLR updates in place of the HLR. The network in 3G is partitioned into Gateway Location Areas (G-LAs) that contain normal LAs. Crossing a G-LA causes an HLR update, while crossing an LA causes a GLR location update, and making a predetermined number of movements between cells causes a VLR update. It is essentially a three-level hierarchical scheme that, while significantly adding to the complexity of the network, equally improves efficiency. Also, analysis shows that dynamic and static 3G networks using this scheme outperform dynamic and static 2G networks, especially if the remote-local-cost ratio is high, and also that both static and dynamic 3G have their purposes; static 3G is preferable if mobility is high or polling/paging cost is low, with dynamic 3G superior in other cases. [Xiao04] However, in the case of a service such as Enhanced 911, the cost of LM is less relevant; the safety of the user transcends monetary considerations.

6.3 E911

Enhanced 911 is an emergency calling system that uses advanced LM techniques to find the physical location of a user in distress. Public Safety Answering Points (PSAPs), specific physical locations designed to center E911 activities around, are staffed by 911

operators who use AGPS or TDOA to aid in locating users in need of assistance. Additionally, Wireless Location Signatures (WLS) are used with path loss and shadow fading data, and are combined with Geographical Predicted Signature Databases (PSD) and statistical analysis to make the most accurate predictions of user locations possible.

In E911, calls follow this process: the BS forwards the call to the MSC, where a voice path is set up to the access point. The MSC requests the BS locate the caller, which forwards the location request to the Serving Mobile Location Center (SMLC). The SMLC uses network measurement reports to approximate the location of the user, and then forwards the approximation to the Gateway Mobile Location Center (GMLC), which then finally transfers a refined estimate to the PSAP. It is a very complex process, unusable for normal communication, but essential in these extreme cases.[Feuerstein04]

However, as recent a development as E911 is, may already be in the process of being phased out. The National Emergency Number Association (NENA) report [NENA05] states that E911 must be upgraded to a Next Generation 911 (NG911) scheme in the near future. New devices must implement systems to handle NG911 calls, and LM must continue to be improved. The new NG911 must be expanded to VoIP architecture, and better integration of national databases and stations must occur. However, E911 still stands as an example of the importance of LM as well as the rapidity of change within cellular networks.

7.0 Conclusion

As the comparisons within the papers included in this survey indicate, static LM schemes are becoming increasingly out of date. While they are still used where cost or resource availability is an issue, upgrading to dynamic schemes is preferable. However, dynamic LM schemes have not yet become a panacea, as many schemes are still only theories, being insightful developments, but not useful under real-world conditions. Consequently, dynamic LM must continue to be researched and improved.

Given the increasing number of cellular users in an on-demand world, and transitions occurring from E911 to NG911 and 2G to 3G and beyond, Location Management must and will continue to improve both Location Update and paging costs, while allocating appropriate Location Areas in a practical, implementable fashion.

8.0 References

Please note - some references may require user authentication.

[Cowling04] James Cowling, "Dynamic Location Management in Heterogeneous Cellular Networks," MIT Thesis. http://people.csail.mit.edu/cowling/thesis/jcowling-dynamic-Nov04.pdf
Extensive overview of Static and Dynamic LM as well as paging and several other parameters.

[Giner04] Vicenete Casares Giner, "State of the art in Location Management procedures," Eurongi Archive. http://eurongi.enst.fr/archive/127/JRA151.pdf
Provides an informative view of the HLR/VLR architecture, as well as providing an overview of LM techniques.

[Kyantakya04] K. Kyantakya and K. Jobmann, "Wireless Networks - Location Management in Cellular Networks: Classification of the Most Important Paradigms, Realistic Simulation Framework, and Relative Performance Analysis."IEEE transactions on vehicular technology 2005: v.54, no.2 687-708.
http://ieeexplore.ieee.org/search/wrapper.jsp?arnumber=1412086
Similar to Cowling's Paper, provides an extensive overview of Static and Dynamic LM schemes.

[Toosizadeh05] Navid Toosizadeh and Hadi Bannazadeh, "Location Management in Mobile Networks: Comparison and Analysis," University of Toronto Report.
http://www.eecg.toronto.edu/~navid/ProjectReport.pdf
Compares 4 Static/Dynamic LM schemes.

[Halepovic05] Emir Halepovic and Carey Williamson, "Characterizing and Modeling User Mobility in a Cellular Data Network," University of Calgary.
http://portal.acm.org/ft_gateway.cfm?id=1089969&type=pdf&coll=GUIDE&dl=ACM&CFID=67915996&CFTOKEN=36591408
Provides an analysis of user mobility, including examination of data traces.

[Lo04] Shi-Wu Lo, Tei-Wei Kuo, Kam-Yiu Lam, Guo-Hui Li, "Efficient location area planning for cellular networks with hierarchical location databases," Computer Networks. Amsterdam: Aug 21, 2004.Vol.45, Iss. 6; pg. 715.
http://dx.doi.org/10.1016/j.comnet.2004.02.012
Discusses the SCBLP theory.

[Xiao04] Y. Xiao, Y. Pan, J. Li, "Design and analysis of location management for 3G cellular networks," IEEE Transactions on Parallel and Distributed Systems, v.15 i.4 April 2004 p.339-349. http://ieeexplore.ieee.org/search/wrapper.jsp?arnumber=1271183
Provides insight as to 3G-specific LM modifications.

[Giner05] Vicenete Casares Giner and Pablo Garcia-Escalle, "On the Fractional Movement-Distance Based Scheme for PCS Location Management with Selective Paging," Lecture Notes in Computer Science, Volume 3427, Jan 2005, Pages 202 - 218.
http://www.springerlink.com/index/M87GJA20CLFMVX9Q.pdf
Discusses a Fractional-Movement scheme using Distance as the main parameter.

[Xiao03] Yang Xiao "Optimal fractional movement-based scheme for PCS location management ," Communications Letters, IEEE, v.7 i.2 Feb 2003 p.67-69. http://ieeexplore.ieee.org/Xplore/login.jsp?url=/iel5/4234/26480/01178889.pdf?tp=&arnumber=1178889&isnumber=26480
Discusses a Fractional-Movement scheme using Cell Movement as the main parameter.

[Lee04] Kevin Lee, Hong Wing Lee, Sanjay Jha, and Nirupama Bulusu, "Adaptive, Distributed Location Management in Mobile, Wireless Networks." http://www.cse.unsw.edu.au/~sjha/papers/icc04lee.pdf
Proposes an agent-based LM system.

[Lam05] Kam-Yiu Lam, BiYu Liang, ChuanLin Zhang, "On Using Handoff Statistics and Velocity for Location Management in Cellular Wireless Networks," The Computer Journal Jan 2005: 48, 1.
http://vnweb.hwwilsonweb.com/hww/jumpstart.jhtml?recid=0bc05f7a67b1790e3761dd0af148832703413c5fdbb17b22a1747616f6aa0b3ead887cd28ffac928
Proposes HVP, discussing use of Velocity and other parameters for user movement prediction.

[Rao03] Bharat Rao, Lous Minakakis, "Evolution of Mobile Location-Based Services." http://portal.acm.org/citation.cfm?doid=953460.953490
Discusses Evolution of Location-Tracking mechanisms, such as GPS/AGPS.

[Rashid06] Omer Rashid, Ian Mullins, Paul Coulton, Reuben Edwards, "Extending Cyberspace: Location Based Games Using Cellular Phones." http://portal.acm.org/ft_gateway.cfm?id=1111302&type=pdf&coll=GUIDE&dl=ACM&CFID=67915996&CFTOKEN=36591408
Discusses LM and systems used in current cell-phone games.

[Feuerstein04] Marty Feuerstein, "The Complex World of Wireless E911," Polaris Wireless. http://search.epnet.com/login.aspx?direct=true&db=buh&an=14114819
Provides an overview of E911 service and call process.

[NENA05] Unsigned, "NENA NG E9-1-1 Program 2005 Report," National Emergency Number Association Report. http://www.nena.org/media/files/ng_final_copy_lo-rez.pdf
Presents NENA's opinions on E911 in 2005 and future recommendations.

[Hassan03] Jahan Hassan and Sanjay Jha, "Cell Hopping: A Lightweight Architecture for Wireless Communications," IEEE Wirelesss Communications October 2003. http://www.comsoc.org/livepubs/pci/private/2003/oct/hassan.html
Presents an LM scheme which avoids centralization and focuses on BS use.

[Fan02] Guangbin Fan, Ivan Stojmenovic, and Jingyuan Zhang, "A Triple Layer Location Management Strategy for Wireless Cellular Networks." http://www.site.uottawa.ca/~ivan/TripleIC3N.pdf

Presents an LA scheme which uses three layers to avoid the ping-pong effect.

[Yu05] F. Yu, V.W.S. Wong, and V.C.M. Leung, "Performance Enhancement of Combining QoS Provisioning and Location Management in Wireless Cellular Networks," IEEE transactions on wireless communications 2005: v.4 no.3 943-953.
http://ieeexplore.ieee.org/search/wrapper.jsp?arnumber=1427684
Presents a scheme where QoS and LM are considered simultaneously.

[Varshney03] Upkar Varshney, "Location Management for Mobile Commerce Applications in Wireless Internet Environment."
http://www.sis.pitt.edu/~dtipper/3955/acm_paper.pdf
Presents an LM scheme for Commercial (high-priority) applications.

---

9. List of Acronyms

AGPS: Assisted GPS
BS: Base Station
E911: Enhanced 911
EIA/TIA: Electronic and Telephone Industry Associations
HLR: Home Location Register
HVP: Hand-off Velocity Prediction
IS: Interim Standard
G-LA: Gateway Location Area
GLR: Gateway Location Register
GMLC: Gateway Mobile Location Center
GPS: Global Positioning Systems
GSM: Global System for Mobile Communications
LA: Location Area
LM: Location Management
LU: Location Update
MAP: Mobile Application Part
MRQ: Membership Request Query
MSC: Mobile Switching Center
NENA: National Emergency Number Association
NG911: Next-Generation 911
PSAP: Public Safety Answering Point
PSD: Predicted Signature Databases
QoS: Quality of Service
SCBLP: Set-Covering Based Location Area Planning
SCP: Service Control Point
SMLC: Serving Mobile Location Center
SSP: Service Switching Point
SS7: Signaling System 7
STP: Signal Transfer Point

TOA: Time of Arrival
TDOA: Time Difference on Arrival
TLA: Two Layer Area
VLA: Virtual Layer Area
VLR: Visitor Location Register
VoIP: Voice over IP
WLS: Wireless Location Signatures

Paper available at http://www.cs.wustl.edu/~jain/cse574-06/cellular_location.htm