Efficient Location Management Based on Moving Location Areas

Yigal Bejerano*
Bell Laboratories, Lucent Technologies,
600 Mountain Ave.
Murray Hill, NJ 07974
E-mail: *bej@research.bell-labs.com.*

Israel Cidon
Department of Electrical Engineering,
Technion - Israel Institute of Technology,
Haifa 32000, Israel
E-mail: *cidon@ee.technion.ac.il.*

Abstract— *Personal Communication Systems* **(PCS) maintain a** *Location Management* **mechanism for tracking the location of their mobile users. The increasing population of mobile users leads to congestion problems in these systems, and motivates the development of more efficient management schemes. This work presents a new mobility management scheme that integrates the location area approach with the location prediction idea. It is based on results from traffic flow theory and it is first that uses the concept of moving location areas. Traffic flow theory suggests that people tend to reside in specific places for long periods of time. Occasionally, they move to new locations and try to minimize the travel time using highways as much as possible. The scheme uses two complement sets of location areas that overlap each other. The first set contains small location areas and is designated for locating mobile users in a quasi-static state. The second set covers the highways and it is designated to track mobile users while they are traveling from place to place, where each highway is covered by a single location area. The dual set design enables tracking mobile users at a high degree of accuracy with low update cost while they are quasi-static state, and reduces the amount of update operations when they travel. For tracing mobile users on a highway, the scheme uses a system of moving location areas. A** *moving location area* **(MLA) is a small location area that defines the location of a group of mobile users, which are geographically concentrated and move in the same direction. The scheme guarantees low rate of update and search operations at each cell of the system and efficiently utilizes the radio spectrum and the network resources with low computational overhead. This advantages are also backed by simulation results.**

Keywords: Personal Communication Systems, Cellular Systems, Mobility Management, Location Management.

I. Introduction

Personal Communication Service (PCS) networks enable people to communicate independently of their location. For locating mobile users the system must maintain a *location management* mechanism, which maps subscriber numbers to the current location of the requested users. The current PCS systems, such as GSM [1] and IS-41 [2], use similar schemes for location management from the radio perspective. The coverage area of the system is divided into location areas (LA), each consists of a group of cells that forms a continuous geographic area. For each mobile user, the system keeps the LA where it resides. When an incoming call arrives, the system simultaneously pages the mobile user in all the cells of this LA. Each time a mobile user crosses a LA boundary, it updates the system with its new location. In current systems the LA coverage is determined in advance, based on static movement probabilities and remains unchanged.

In recent years the PCS networks are facing a rapid growth of mobile user population and coverage areas. The main solution for supporting the growing population is to reduce the cell sizes and to increase the bandwidth reuse [3].Therefore, the number of call delivery and location update operations increase dramatically and results in high load on the mobility management mechanism. This motivates the development of new mobility management schemes that use efficiently both the network resources and radio bandwidth.

The new schemes are generally divided into two categories. The first category includes incremental improvements to the current location area approach. It includes different schemes [4], [5], [6] for optimizing the location area design to reduce the cost of update and search operations. Other LA-based schemes consider profile information of each mobile user [7], [8], like the sequential paging method [9], [10], where the system searches for a mobile user by paging sequentially sub-areas of the LA where the mobile user resides. While some LA-based schemes consider the mobile user profile information, the LA approach is not flexible enough to adapt to different mobile users mobility patterns and communication requirements [11], [12].

The second category includes dynamic mobility management schemes [13], where mobile users perform update operations based on either the elapsed time, number of crossed cells or the traveled distance since the previous update. In all these methods the selected threshold for performing an update operation is adapted to the individual mobile user mobility patterns and communication traffic. Some schemes [11], [12], [14], also consider the speed and the trajectory of each mobile user at the time of its last update for predicting its location when needed. They present a lower mobility management cost than the location area and the non-predictive dynamic approach. However, the dynamic schemes have some inherent deficiencies. For each mobile user, they collect mobility information for predicting its location, requiring large databases and causes significant computational load on the system. This makes them complicated and difficult to implement [7], [12]. The prediction schemes generally assume that a mobile user tends to keep its trajectory and speed. In practice, a mobile user uses the existing road system for traveling from place to place. Its trajectory and speed are affected by the road topology and the traffic conditions. Moreover, the dynamic mobility management schemes attempt to reduce the total mobility management load. However, it is hard to determine the required bandwidth for supporting these operations at each cell.

In this work we present a new mobility management scheme

* While the author was at the Electrical Engineering Dept. at the Technion. Research supported by Eshkol Fellowship from the Israeli Ministry of Science.

that combines the concept of location areas with the location prediction idea. It guarantees a low rate of update and search operations at each cell of the system, and efficiently utilizes the radio spectrum and the network resources while maintaining low computational overhead. It is based on results from traffic flow theory, which deals with the design, operation, control and management of a national highway and street systems. The strong affect of the transportation system and the traffic flow on the behavior of mobile users has already mentioned in different papers [8], [16]. However, only few mobility management schemes [4], [5], [8]. are based on results from this research field.

Traffic flow theory suggests that people usually reside in specific places for long periods of time. Occasionally, they move to new locations and try to minimize the travel time using highways as much as possible. Based on this observation, the scheme uses two separate sets of location areas that overlap each other. The first set contains small location areas and is designated for locating mobile users in a quasi-static state. The second set covers the highways and it is designated to track mobile users while they are traveling from place to place. First, we assume that each highway is covered by a single LA. Thus mobile users need to perform update operations only upon entering and leaving a highway. The dual LA set design enables tracking mobile users at high degree of accuracy with low update cost while they are in a quasi-static state, and reduces the amount of update operations when they travel. In practice, most of the highways are long and support a dense population of mobile users. Hence, covering the entire highway with a single LA results in a high search cost as it requires paging all the cells along the highway for every incoming call. Our second solution considers an observation from traffic flow theory that drivers usually react similarly when facing the same road conditions and traffic flow. Consequentially, group of users that are close to each other at some time and are traveling in the same direction are expected to remain concentrated for a long period of time. For tracing mobile users on a highway, our proposed scheme uses a system of moving location areas. A *moving location area* (MLA) is a small LA that defines the location of a group of mobile users, which are geographically concentrated, and moves with them in the same direction. The movement speed and the size of an MLA are dynamically changed according to the traffic flow on the highway. This is done for adapting its position and speed to the group of mobile users it covers. The MLA system enables to trace the locations of user groups with a high degree of accuracy and with few update operations. This is in contrast to predicting the movement of individual users that requires much more overhead in the term of computation, storage and communication resources. These benefits are also backed by our simulations. In addition, the simulation results show that the proposed scheme produces lower update and search costs than the current location area approach [1], [2], and some dynamic mobility management schemes [13].

The work is organized as follows. Section II presents a short overview of traffic flow theory. The system model is described in Section III. Section IV presents the principles of the proposed scheme and Section V gives the optimal values of the scheme parameters. Section VI deals with implementation aspects of the scheme and simulation results are present at section VII.

II. A Short Overview of Traffic Flow Theory

Traffic flow theory is a branch of transportation theory. It deals with the design, operation, control and management of a national highway and street systems, for providing efficient and safe movement of people and goods [15]. This overview describes only the aspects of traffic flow theory that are relevant to this work. An extensive survey of this area can be found in [15], [16].

The highway and street system provides two different functions, through movement and land access. Every car journey starts and ends at a parking zone. The street system provides land access for enabling the car drivers to enter and to leave the system at different locations. Usually, the land access streets are not designed to support fast movement from place to place. Therefore, the street system also includes highways that are designed to support continuous flow of many vehicles for large distances at high speeds. In an urban area, the different streets are classified according to their through movement and land access functionality. Generally they are divided into three groups: *Arterial* streets primarily provide efficient through-traffic movement. *Collector* roads are the ”middle” classification and are intend to serve both as through movement and land access. Usually, they connect local streets with arterial roads. *Local streets* primarily provide land access functions and they are designed to enable vehicles to enter and leave the street system. In the following, we term arterial and collector roads as *highways*.

Let us turn to describe the mobility behavior of a car driver. His behavior is mainly guided by two rules, minimizing the travel time and safe driving. Once a driver enters the street system he seeks the shortest path to a highway and he uses highways for the longest possible portion of his trip to minimize the total travel time. When he approaches the destination, he selects collectors and local streets that bring him to his destination. Along the way, the driver's speed depends on the traffic stream that he experiences. The *traffic stream* is the outcome of the drivers' interaction with the road environment and with each other and it is described by three parameters, speed, density and flow. The traffic *speed*, s, is defined as the rate of motion in distance per unit of time. The mean speed is calculated by averaging the speeds of all the vehicles occupying a given section of the road at a specific time. This mean is termed the *space mean speed (SMS)* and it is denoted by $\bar{S}$. The *density*, D, is defined as the number of vehicles that occupy a given section of the road at a given time. There is a strong relation between car density and the car mean speed, because safe driving requires a safety distance between cars, which increases with car speed [16]. The traffic *flow*, Q, is defined as the number of vehicles that pass a point on the road in a given direction during a specific time interval and satisfies $Q = \bar{S} \cdot D$. The relations between the flow, speed and density are illustrated in Figure 1 from [15]. These relations depend on the *road environment* at each point, such as the number of lanes and the road slope.

In practice, different drivers react differently to road conditions. The speed of each car is a random process that depends on the traffic flow, the road environment and the driver characteristic. For a given location x and car density[1] D, the car

[1] The car density defines absolutely the traffic flow for a given speed-density

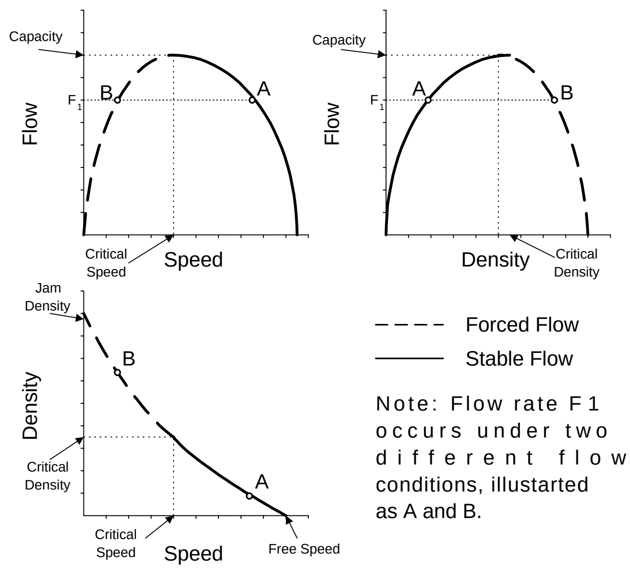

Fig. 1. The relations between flow, speed and density.

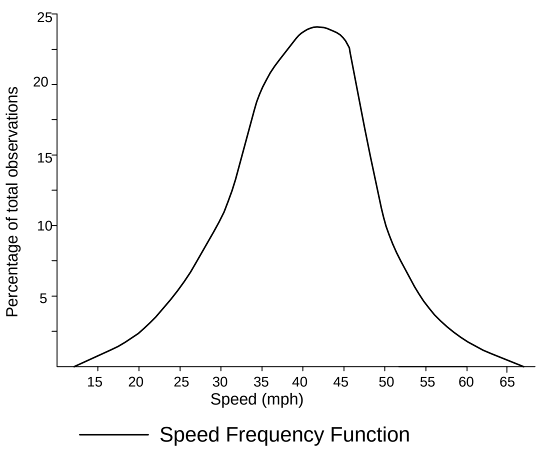

Fig. 2. A typical speed frequency function in a highway.

speeds are distributed according to some probability frequency function, $f_{S(x,D)}(s)$, that is called the *speed frequency function*. A typical speed frequency function is depicted in Figure 2, from [15]. It shows that most car speeds are in the central range of the function.

III. The Model

This work deals with a PCS network for a large urban area. The urban area is covered with radio cells that provide wireless communication to the mobile users in this area. The mobile users are free to move from place to place and they can determine their location, speed and trajectory, for instance by using the technique described in [17]. The PCS mobility management includes update and search procedures for tracking the location of mobile users. These procedures use dedicated control channels at each cell. In our work, λ denotes the average incoming call rate. C_s and C_u represents the cost of a search and update operations in a single cell, respectively, measured by the required radio resources. To simplify our analysis we assume that all the cells have the same diameter of one unit.

The different zones of the urban area are connected by a street system as described in Section II. In our model, every arterial road is considered as two uni-directional freeways that lead to opposite directions. A *freeway* is a multi-lane uninterrupted-flow road that does not contain traffic signals, STOP or YIELD signs or intersections that interrupt the flow. As a result, its traffic stream is a product of the vehicles interaction with each other and with the road environment. We assume that the system knows the traffic stream parameters, at each point along a highway and at any given time. Thus, it can calculate the speed frequency function, $f_{S(x,D)}(s)$, for every location x and density D.

relations.

IV. The Principles of the Proposed Scheme

This work presents a new mobility management scheme that is based on results from traffic flow theory and integrates the concepts of location areas with location prediction.

A. The Dual Sets of Location Areas

From traffic flow theory we learn that people usually reside in specific places for long periods of time. From time to time they move to new locations and attempt to minimize the travel time by using highways as much as possible. Based on this observation we classify mobile users to two types, *slow moving users*, like pedestrians or vehicular users traveling in local streets, and *fast moving users* that are vehicular users traveling on highways. The scheme uses two separate sets of location areas that overlap each other. The first set is designated for slow moving users and it is composed of small LAs. Each LA contains few cells that cover only a small urban zone and has access to a highway.

A location area design that is based only on small LAs brings fast moving users to perform an enormous amount of update operations. For these users we define an additional set of LAs that cover only the highways and are called *highway LAs*. Each highway is covered by a single LA that is used for locating fast moving users in this road. A fast mobile user performs update operations only when entering and leaving the highway. Consequently, a mobile user performs update operations during its travel; when it first enter a highway, moves from one highway to another and reaches its final destination. Figure 3 shows an example of the two LA sets and a trip of a mobile user from one end of the city to the other in which only three update operations were made. This design of two LA sets enables the system to keep the location of mobile users in high degree of accuracy with low update cost while they are in quasi static state, and reducing the amount of update operations when they travel from place to place.

B. The Concept of Moving Location Areas

Highway LAs are narrow since they cover only the highway routes. For relatively short streets these LAs contain a small number of cells and the cost of search operation in such LA is relatively low. However, arterial streets are usually long with a dense population of mobile users. Covering a long arterial street with a single LA results in a high search cost as it requires paging all cells along the highway for each incoming call. One

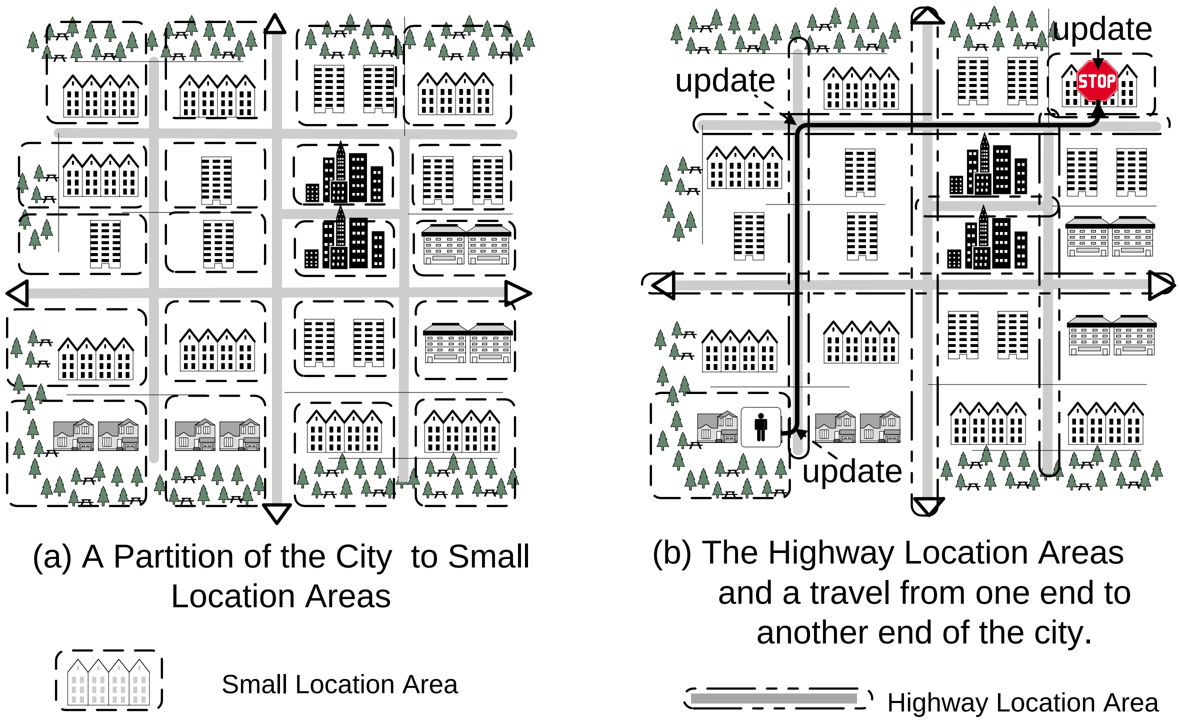

Fig. 3. An example of a city and its partition into location areas.

possible solution is covering the highway with several shorter LAs. This reduces the search cost but increases the amount of update operations. A different solution is to use a location prediction method [11], [14]. in which the location of a mobile user is predicted according to its location, trajectory and speed at its last update operation. Such a solution imposes a high computational load on the system and the mobile devices. In addition, it assumes that mobile users keep their speed and trajectory for estimating their locations and it ignores the changes of the road environment and the traffic flow on the mobile user locations

This work presents a new solution based on the observation that drivers usually react similarly to the road environment and the traffic flow. Consider a group of cars that are located at the same point at a given time and are traveling in the same direction. We expect this group to stay close to each other for a long period of time, since their drivers experience the same changes of the road conditions and react similarly to them. Consider a single traffic flow over a highway. The scheme traces the locations of the mobile users in this flow using a system of moving location areas. Each *moving location area* (MLA) operates like an ordinary LA. It defines the location of a group of mobile users that are geographically concentrated and moves with them in the same direction. At any given time, each MLA is associated with few cells and the current set of MLAs covers the entire highway. New MLA's are generated all the time at the highway starting point and move along its route until they reach the other end. During an MLA movement, its speed and size are dynamically changing according to the traffic mean speed and density at its current location. This is necessary for adapting the MLA position and speed to the group of mobile users it attempts to follow.

According to our model, a mobile user can track its current speed and trajectory. Thus, it determines that it is a fast moving user when its recent speed exceed a given threshold[2]. In our scheme, slow moving users use the static LA set. They perform update operations when they cross static LA boundaries and register themselves to the static LAs, in which they reside. Fast moving user use the highway MLA systems. They register themselves to MLAs upon entering or leaving highways, or when they cross MLA boundaries. Based on the similar reaction observation, we assume that fast moving users tend to remain attached to a single MLA for a long period of time. The separation between fast and slow moving users and the MLA approach enables the system to trace its mobile user efficiently with low communication and computational load.

C. Moving Location Areas with Soft Edges

This section develops an efficient application of the moving location areas concept. In practice, this concept can be implemented using disjoint MLAs, such that at every given time each cell along the highway is associated with a single MLA. The speed of each MLA is the flow mean speed at its current location. This simple implementation has two major drawbacks. A mobile user that travels near an MLA boundary may create a high number of update operations by crossing the MLA boundary back and forth many times as a result of minor changes in its speed. Moreover, at any given time each cell along the road is either an internal or edge cell of the MLA to which it is associated. During the time that a cell is internal it is not required to support update operations (except for the case of intersections

[2]This threshold may depend on the user location and the current traffic flow at this place.

where new cars enter and leave the highway). When a cell is at the MLA's edge it is required to support many update operations made by the mobile users that cross the MLA boundary. Each cell must hold enough control channels for supporting all the update operations when it becomes an edge cell, although these channels are not used when the cell is internal. This results in inefficient usage of the radio spectrum.

The proposed scheme solves these deficiencies and efficiently utilizes the control channels at all cells along the highway by using a method termed *soft edges*. In this method the front and back boundaries of each MLA are composed of two groups of cells at any given time in contrast to only two edge cells for the application described above. As a result, each cell faces low update rate. For simplifying the description, consider a highway with uniform road environment and flow. Each mobile user is associated with a number between 0 to 1, termed a *key*. The keys are uniformly distributed among all the mobile users. Every MLA is composed of three groups of cells, a *core* that contains L cells and represents the internal cells, and two groups of edge cells in which mobile users leave the MLA and register to a new one. The later groups are called *front edge* and *back edge*. They contain $\lceil \alpha \cdot L \rceil$ cells and $\lfloor (1 - \alpha) \cdot L \rfloor$ cells respectively, for a given constant $0 < \alpha < 1$. Thus the entire MLA contains $2 \cdot L$ cells. The values of L and α are addressed in section V. The set of cores covers all the highway cells without overlapping each other. However, adjacent MLAs are overlapping as a result of wrapping each core with front and back edges. At any given time each cell is included at core of one MLA and at a edge group of another MLA[3]. Consider the cells of an MLA front edge and enumerate them from 1 to $\alpha \cdot L$. Each cell, k, is associated with a set of numbers in the range $(\frac{k-1}{\lceil \alpha \cdot L \rceil}, \frac{k}{\lceil \alpha \cdot L \rceil}]$ and is the front edge cell to all the mobile users in this MLA that their keys are in this range. The sets are disjoint and include the entire range $(0, 1]$. Thus, every cell in the front edge group is an edge cell only for an expected fraction of $\frac{1}{\lceil \alpha \cdot L \rceil}$ of the mobile users and is an internal cell to the others. The back edge group is organized in a similar manner, and the cells are enumerated from 1 to $\lfloor (1 - \alpha) \cdot L \rfloor$. An example of the MLA's design is depicted in Figures 4-5, where $L = 7$ and $\alpha = \frac{4}{7}$. Figure 4 shows the design of a single MLA and the associated range of each edge cell. Figure 5 describes the design of several adjacent MLAs. Let u be the MLA's speed measured in cell size per time unit. It means that every $\frac{1}{u}$ time units every MLA moves one cell forward. A new cell is added to the MLA's front edge group, the last cell in the front edge becomes the first cell of the MLA's core and so on. In addition each front and back cell is associated with a new range according to its relative location in these groups.

We turn to describe the update and search procedures. Each MLA has a unique identification number, termed *MLA-id*. A mobile user that travels on the highway is registered at a single MLA, and keeps the corresponding MLA-id. At any given time, it considers a single front edge cell and a single back edge cell of the MLA, which are determined according to its key value. Every cell broadcasts to all the mobile users in its area

[3]Note that usually the cell is also included in a static LA and two MLA's that moving in opposite direction.

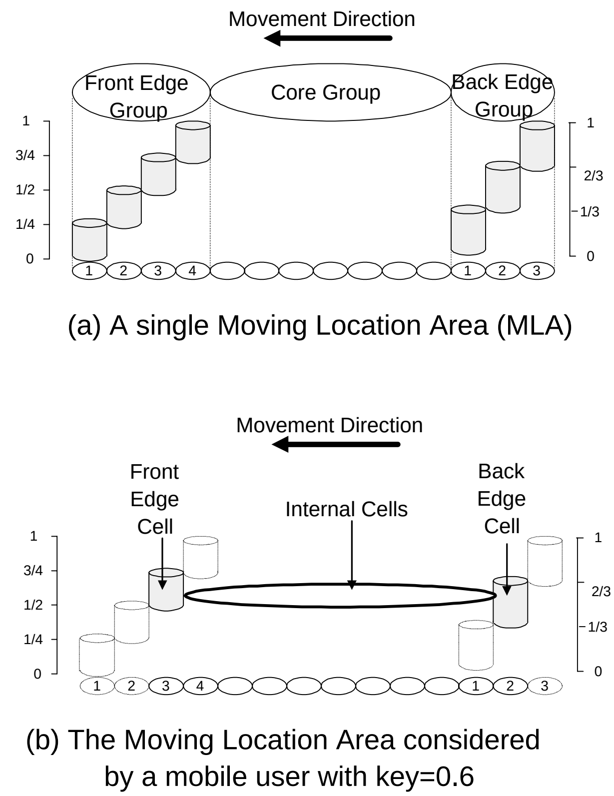

Fig. 4. An example of a single moving location area.

the MLA-id that it is included in its core, the MLA-id that it is included in one of its edge groups and its associated range of numbers, and also the MLA speeds and trajectories. According to this information and its key number, every mobile user determines whether it is attached to an edge cell and perform update operation. A mobile user, that continues to travel in the same direction, register itself to the MLA that it is included in its core. Upon leaving the highway, the mobile user registers itself to a static location area or to an MLA of an another highway, depends on its speed and trajectory. When an incoming call arrives, the system checks in which MLA the called user is registered, and pages this user at all the MLA cells between its two edge cells.

The soft edge method solves the two deficiencies presented above. When a mobile user performs an update operation it is located at the core of its new MLA and its distance from one of its edge cells is at least $min\{\lceil \alpha \cdot L \rceil - 1, \lfloor (1 - \alpha) \cdot L \rfloor - 1\}$. Thus, a mobile user cannot produce a high number of update operations by crossing MLA boundaries frequently. Moreover, it is clear from the MLA design that the load of the update operations is uniformly distributed amount all the highway cells at any given time. This enables supporting the update and search operations at each cell along the highway with limited number of control channels.

V. Optimal MLA Parameters

In this section we calculate the optimal values of the MLA parameters, u, L and α, (MLA speed, core length and ratio) for reducing the search and update costs at each cell along the high-

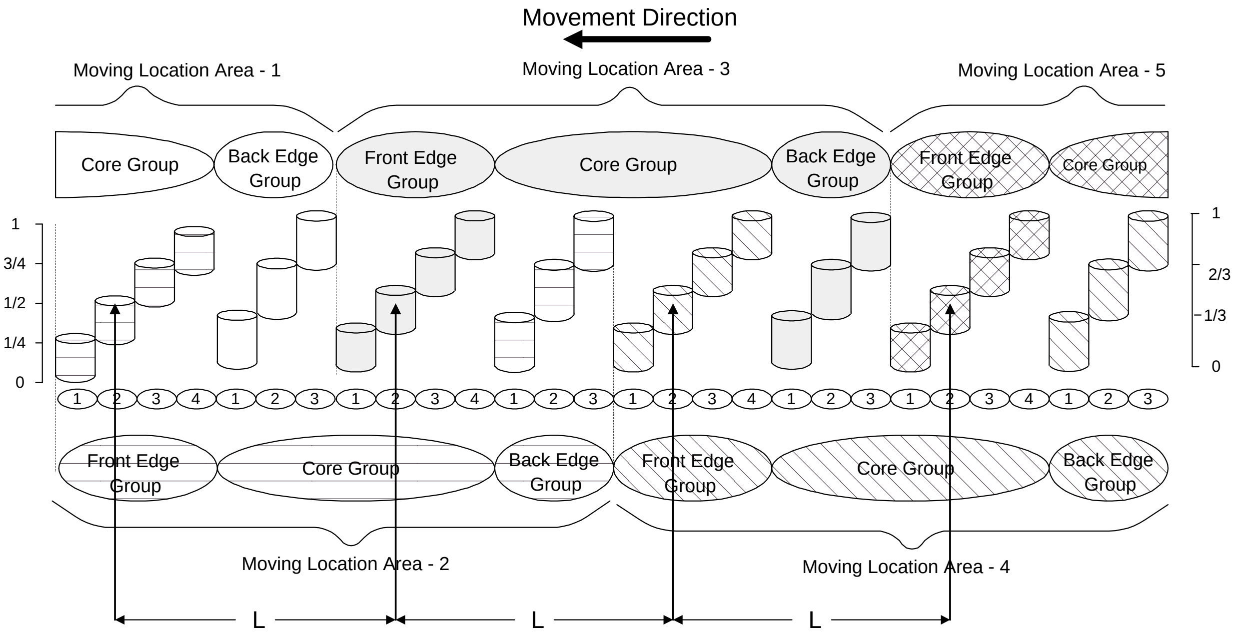

Fig. 5. An example of a set of moving location areas.

way. In our analysis we assume a multi-lane freeway without intersections and with a constant traffic flow, Q. In Section V-C we extend our results to highways with intersections. Due to space limitation, some proofs are omitted and can be found in the full version, [18].

A. A Freeway with a Constant Road Environment

First we consider the case of a constant road environment and invariant traffic flow along the freeway. Thus, at each point x and every time t the flow rate, Q, the density D and the car mean speed $\bar{S}$ are constants. We assume that the cars are uniformly distributed and the drivers are tending to keep their speeds. Consequently, the speed frequency function, $f_S(s)$, is the same at all cells.

Let $Cost(x,t)$ be the expected cost of the update and search operations at the cell in location x at a given time t during one time unit. $Cost(x,t)$ is time dependent since the cell changes from being a front edge cell to back edge cell along time. Notice that $Cost(x,t)$ is a periodic function with a period of $u \cdot L$ time units. Let $Peak\_Cost(x) = max_t Cost(x,t)$, be the maximal value of $Cost(x,t)$. Our goal is finding the MLA parameters that bring $Peak\_Cost(x)$ to minimum for every location x. The required number of control channels at each cell is proportional to this value. We first calculate the search cost at each cell. From symmetry considerations the expected number of mobile users at each MLA is $D \cdot L$. The incoming call rate of a mobile user is λ and the cost of a single search operation in a cell is C_s. Since each cell is associated with two different MLAs at any given time, the search cost is, $Search\_Cost = 2 \cdot \lambda \cdot D \cdot L \cdot C_s$.

Let turn to calculate the update cost. This cost depends whether the cell is a front or a back edge cell of an MLA. Front and back edge cells experience different update rates according to the speed frequency function, $f_S(s)$ and the MLA parameters. Let define,

$$I_f(u) = \int_u^{\infty} (s-u) \cdot f_S(s) ds$$

$$I_b(u) = \int_0^{u} (u-s) \cdot f_S(s) ds$$

The update rate at the cells of an MLA front edge group is $D \cdot I_f(u)$, and is $D \cdot I_b(u)$ at the cells of a back edge group. These operation are uniformly distributed along the cells of each edge group. The sizes of the front and back edge groups are $\alpha \cdot L$ and $(1-\alpha) \cdot L$, respectively. Hence, the update costs that a front and back edge cell experience are,

$$Front\_Update\_Cost = \frac{D \cdot I_f(u) \cdot C_u}{\alpha \cdot L}$$

$$Back\_Update\_Cost = \frac{D \cdot I_b(u) \cdot C_u}{(1-\alpha) \cdot L}$$

where C_s is the cost of a single update operation. Recall that the a cell is always at a core of a single MLA and at either the front or back of another MLA. Therefore, the peak cost of a cell is given by

$$Peak\_Cost = Search\_Cost + \quad (1)$$

$$+ MAX \{Front\_Update\_Cost, Back\_Update\_Cost\} =$$

$$= 2 \cdot \lambda \cdot D \cdot L \cdot C_s + \frac{D \cdot C_u}{L} \cdot MAX \left\{ \frac{I_f(u)}{\alpha}, \frac{I_b(u)}{1-\alpha} \right\}$$

From Equation 1 we see that u and α affect only the update cost. First we calculate the optimal values of these parameters. Let $G(u,\alpha) = MAX \left\{ \frac{I_f(u)}{\alpha}, \frac{I_b(u)}{(1-\alpha)} \right\}$. The optimal values of u and α are the ones that bring $G(u,\alpha)$ to minimum.

Lemma 1: $I_f(u) = \bar{S} - u + I_b(u)$.

Lemma 2: For any given value of α, $G(u,\alpha)$ is minimal only if $\frac{I_f(u)}{\alpha} = \frac{I_b(u)}{(1-\alpha)}$.

Let $\alpha_{opt}(u)$ denotes the optimal value of α for a given MLA speed, u.

Lemma 3: $\alpha_{opt}(u) = \frac{I_f(u)}{u - \bar{S} + 2 \cdot I_b(u)}$.

Corollary 1: If the MLA speed is $\bar{S}$, then $\alpha_{opt}(\bar{S}) = \frac{1}{2}$.

In the sequel we use the following notations. S_{mid} denotes the median of the speed frequency function and $\bar{S}_{mid} = E\{S \mid S \geq S_{mid}\}$.

Theorem 1: The optimal MLA speed is S_{mid} and

$$\alpha_{opt}(S_{mid}) = \frac{\bar{S}_{mid} - S_{mid}}{2 \cdot (\bar{S}_{mid} - \bar{S})}$$

Proof: Let $H(u) = G(u, \alpha_{opt}(u))$. According to Lemma 2, $H(u) = \frac{I_f(u)}{\alpha_{opt}(u)}$. From Lemma 3, $H(u) = u - \bar{S} + 2 \cdot I_f(u)$. Hence, $H(u) = u - \bar{S} + 2 \cdot \int_u^\infty (s-u) \cdot f_S(s) ds$

For finding the optimal value of u we calculate the derivative of $H(u)$ and equate it to zero.

$$\begin{aligned} \frac{dH(u)}{du} &= 1 + 2 \cdot \frac{d}{du} \left\{ \int_u^\infty (s-u) \cdot f_S(s) ds \right\} \\ &= 1 + 2 \cdot \left\{ -u \cdot f_S(u) - \int_u^\infty f_S(s) ds \right. \\ &\quad \left. + u \cdot f_S(u) \right\} \\ &= 1 - 2 \cdot \int_u^\infty f_S(s) ds = 0 \end{aligned}$$

As a result, the optimal value of u is achieved when $\int_u^\infty f_S(s) ds = \frac{1}{2}$, Thus $H(u)$ has an extreme point at $u = S_{mid}$. By calculating the second derivative of $H(u)$ we get that $\frac{d^2H}{du^2} = 2 \cdot f_S(u) \geq 0$, which proves that $u = S_{mid}$ is a minimum point. By setting S_{mid} into $\alpha_{opt}(u)$ we get that $\alpha_{opt}(S_{mid}) = \frac{\bar{S}_{mid} - S_{mid}}{2 \cdot (\bar{S}_{mid} - \bar{S})}$. □

Corollary 2: $G(S_{mid}, \alpha_{opt}(S_{mid})) = \bar{S}_{mid} - \bar{S}$.

Theorem 2: The optimal L is $L = \sqrt{\frac{C_s \cdot (\bar{S}_{mid} - \bar{S})}{2 \cdot C_u \cdot \lambda}}$ and the minimal $Peak\_Cost$ value is

$$Peak\_Cost = 2 \cdot D \cdot \sqrt{2 \cdot C_s \cdot C_u \cdot \lambda \cdot (\bar{S}_{mid} - \bar{S})}.$$

Proof: By substituting the optimal values of u and α from Theorem 1 into Equation 1 we get.

$Peak\_Cost = 2 \cdot \lambda \cdot D \cdot L \cdot C_s + \frac{D \cdot C_u}{L} \cdot (\bar{S}_{mid} - \bar{S})$

For finding the optimal value of L we calculate the derivative of $Peak\_Cost(L)$ and equate it to zero. $\frac{d\ Peak\_Cost}{dL} = 2 \cdot \lambda \cdot D \cdot C_s - \frac{D \cdot C_u}{L^2} \cdot (\bar{S}_{mid} - \bar{S}) = 0$

Hence, the optimal value of L and the minimal $Peak\_Cost$ are,

$L = \sqrt{\frac{C_s \cdot (\bar{S}_{mid} - \bar{S})}{2 \cdot C_u \cdot \lambda}}$

$Peak\_Cost = 2 \cdot D \cdot \sqrt{2 \cdot C_s \cdot C_u \cdot \lambda \cdot (\bar{S}_{mid} - \bar{S})}$ □

Our analysis shows that selecting MLA speed as the median speed, S_{mid}, guarantee the minimal peak cost. This may be counter intuitive as one may expect that the optimal MLA speed should be the average speed, $\bar{S}$. The following example demonstrates the correctness of our results.

Example 1: Consider a two-lane freeway that is used only by two types of vehicles. 80% of the vehicles are private cars with a constant speed of 100 Km/h, and the rest 20% are trucks with a constant speed of 50 Km/h. The mean speed is $\bar{S} = 0.8 \cdot 100 + 0.2 \cdot 50 = 90$ Km/h, and the median speed is $S_{mid} = 100$ Km/h. Let compare the peak cost of two MLA systems, both with the same L. In the first system the MLA speed $u = \bar{S} = 90$ and $\alpha_{opt}(\bar{S}) = 0.5$, while the MLA speed is $u = S_{mid} = 100$ and $\alpha_{opt}(S_{mid}) \to 0$ in the second system. We compare the two systems by calculating $G(u, \alpha_{opt}(u))$ of each one of them. In the first case $G(\bar{S}, \alpha_{opt}(\bar{S})) = MAX\{2 \cdot (100-90) \cdot 0.8, 2 \cdot (90-50) \cdot 0.2\} = 16$. But in the second system $G(S_{mid}, \alpha_{opt}(S_{mid})) = MAX\{\frac{1}{\alpha \to 0} \cdot (100-100) \cdot 0.8, 1 \cdot (100-50) \cdot 0.2\} = MAX\{0, 5\} = 10$. We can see that the second system produces less update operations per cell than the first one. In this system the speed of the private cars is the same as the MLA speed. Therefore, they don't perform update operations. The update rate of the trunks is 20% higher, however the size of the back edge group is twice than the corresponding group in the first system. Therefore the update rate per cell is 37.5% lower in the second system.

B. A Freeway with a Varied Road Environment

In the following, we consider the case of a freeway with a varied road environment. Consequently, the traffic speed and density are varied along the way. This requires adapting the MLA parameters to the traffic changes. The modifications of the MLA parameters should keep the system at low search and update costs. In addition, the transition stage from one road environment to another should be as smooth as possible with a low update cost.

We assume that the highway is composed of a set of intervals, each one with a constant road environment, and they are connected by transition sections. From Section V-A, we learn that the optimal MLA speed, u, is the car median speed S_{mid}, while the optimal core length, L, is proportional to $\sqrt{\bar{S}_{mid} - \bar{S}}$. This make it impossible to guarantee optimal MLA parameters along the highway without using different MLA system at each interval. The latter is unacceptable solution, since it force the mobile users to perform update operations each time the road environment changes and results with very high update cost.

Our solution is based on adapting the MLA speed to the traffic median speed at each location along the road. The scheme considers each MLA as a set of virtual points that may have various speeds simultaneously, where the speed of each point is the traffic median speed, $S_{mid}(x)$, at its current location, x. This is actually a linear transform that affects the MLA speed, u, and its core length, L. This method may leave the MLA length sub-optimal in some road intervals, but it guarantees a smooth transition between different road environments, as the following theorems prove. An example of the affects of traffic speed changes on an MLA shape is depicted in Figure 6. It shows that when the traffic is sparse and fast then the MLA is large and fast. When the traffic flow is dense and slow the MLA is small and slow.

For our analysis, we use the following assumptions. We consider a constant traffic flow, Q. Thus, the mean speed, $\bar{S}(x)$, and the density, $D(x)$, at every location x satisfy $D(x) \cdot \bar{S}(x) = Q$. Let $S(x)$ be the random variable describing the car speeds at location x with speed frequency function $f_{S(x)}(s)$. We assume that the different speed frequency functions along the way are linear transformations of a single frequency function, $f_{S(x_0)}$, of a given reference location x_0. Thus, $S(x) = \beta(x) \cdot S(x_0)$ and its frequency function satisfy $f_{S(x)}(s) = \frac{1}{\beta(x)} \cdot f_{S(x_0)}(\frac{s}{\beta(x)})$,

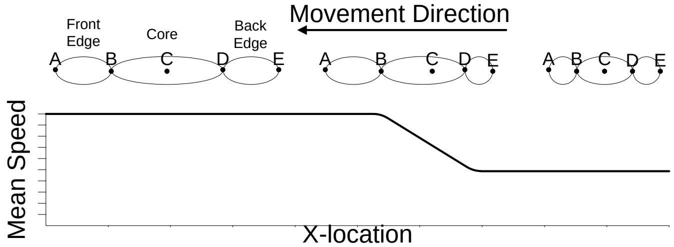

Fig. 6. An example of a movement of a moving location area over a highway with a varied road environment.

where $\beta(x) = \frac{\bar{S}(x)}{\bar{S}(x_0)}$. Recall that $\beta(x) = \frac{S_{mid}(x)}{S_{mid}(x_0)}$, since for every x, $S_{mid}(x) \equiv C \cdot \bar{S}(x)$, where C is a constant derived from the frequency function $f_{S(x_0)}$.

Consider a virtual object that moves along the highway and its speed at each location x is $\gamma \cdot \bar{S}(x)$, where γ is a positive constant. Let $r_{\gamma,f}(x)$ and $r_{\gamma,b}(x)$ be the rates of cars that this object passed and passed this object when its location is x, respectively.

Lemma 4: The rates $r_{\gamma,f}(x)$ and $r_{\gamma,b}(x)$ are constants independent of x.

Proof: Let calculate $r_{\gamma,f}(x)$ at any given location x.

$$r_{\gamma,f}(x) = \int_{\gamma \cdot \bar{S}(x)}^{\infty} D(x) \cdot (s - \gamma \cdot \bar{S}(x)) \cdot f_{S(x)}(s) ds$$

by using the assignment $w = \frac{s}{\beta(x)}$, we get that $r_{\gamma,f}(x) = \int_{\frac{\gamma \cdot \bar{S}(x)}{\beta(x)}}^{\infty} D(x) \cdot (\beta(x) \cdot w - \gamma \cdot \bar{S}(x)) \cdot f_{S(x)}(\beta(x) \cdot w) \cdot \beta(x) \cdot dw$ Since $D(x) \cdot \bar{S}(x) = D(x_0) \cdot \bar{S}(x_0) = Q$ and $\beta(x) = \frac{\bar{S}(x)}{\bar{S}(x_0)}$, we get that $D(x) = \frac{D(x_0)}{\beta(x)}$. In addition we the relation $S(x) = \beta(x) \cdot S(x_0)$. As a result,

$$\begin{aligned} & r_{\gamma,f}(x) = \\ &= \int_{\frac{\gamma \cdot \bar{S}(x)}{\beta(x)}}^{\infty} \frac{D(x_0)}{\beta(x)} \cdot (\beta(x) \cdot w - \gamma \cdot \beta(x) \cdot \bar{S}(x_0)) \cdot \\ & \quad \cdot f_{S(x_0)}(w) dw \\ &= \int_{\gamma \cdot \bar{S}(x_0)}^{\infty} D(x_0) \cdot (w - \gamma \cdot \bar{S}(x_0)) \cdot f_{S(x_0)}(w) dw \\ &= r_{\gamma,f}(x_0) \end{aligned}$$

Thus, foe erver x, $r_{\gamma,f}(x) \equiv r_{\gamma,f}(x_0)$. In a similar way we may prove that $r_{\gamma,b}(x) \equiv r_{\gamma,b}(x_0)$. □

Lemma 4 is in particular satisfied in the case of a moving point with speed $S_{mid}(x)$. It is the base for the following theorems. The full proofs of these theorems can be found in [18].

Theorem 3: The expected number of mobile user in an MLA is constant.

Theorem 4: The expected search cost of all the cells along the highway is the same, $Search\_Cost = 2 \cdot \lambda \cdot N \cdot C_s$, where N represents the expected number of mobile users in an MLA.

Theorem 5: The expected number of update operations in the front and back edge group of an MLA is constant independent of the MLA location.

Theorem 6: Let $Update\_Cost(x_0)$ be the expected update cost of the cell at location x_0 with length 1. Then the expected update cost of the cell between the coordinates x_{start} and x_{end} is $Update\_Cost(x_0) \cdot \int_{x_{start}}^{x_{end}} \frac{dx}{\beta(x)}$.

C. Handling Intersections by Using Temporary MLAs

So far, we have dealt with a single highway without intersections where all cars follow the same path without leaving or joining the traffic flow. In practice, highways contains intersections in which cars can travel to different directions. This raise a management problem how to efficiently support an MLA system with the present of intersections. Consider an intersection in which the highway splits to several directions. One solution is to duplicate every MLA as the number of egresses, such that each replica continues to travel through a distinct egress of the intersection, as described in Figure 7-(a). This enables the cars to continue their travels without performing update operations. When an incoming call arrive to a mobile user, the system pages the user at all the replicas of the MLA to which it is registered. Although this solution provides smooth transition between the ingress and the egress of the intersection, it increases the search cost and the management complexity, especially when these MLAs are duplicated again and again at each intersection. Another possibility is to select one egress of the intersection as the *main route*. The MLAs that reach the intersection continue to move through this egress and new MLA system are initiated at all other directions as depicted in Figure 7-(b). In this approach all the cars that do not travel over the main route are required to update the system about their movement direction. This causes the cell that cover an intersection to become a hot-spot with high update cost.

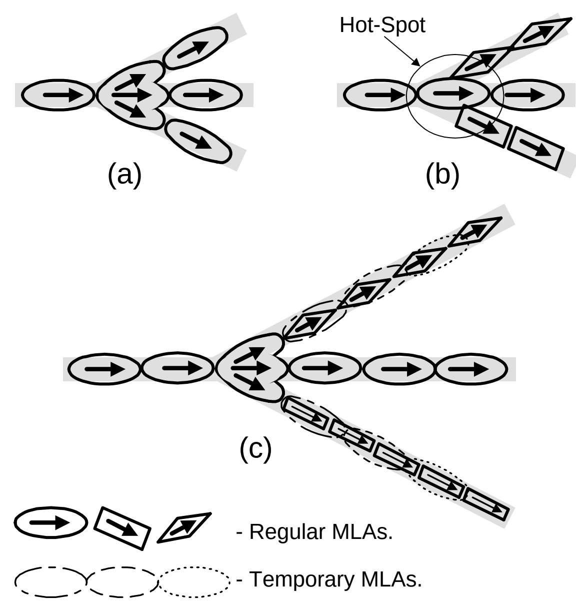

Fig. 7. Different solutions for dealing with intersections.

Our solution efficiently combines the two previous approaches without generating hot-spots and with modest incremental of the search cost and the management complexity. One of the egresses is selected as the *main route* and all the other are considered as *minor routs*. A new MLA systems is initiated at each minor route. When an MLA reaches the intersection, it is duplicated and one replica continues to move to each egress direction. However, an MLA replica that moves over a minor route is consider as a *temporary MLA*, it last for a short distance and than it is canceled. Each mobile user that is associated with a temporary MLA randomly selects a cell within this distance, in which it performs update operation and register itself to a regular MLA. A description of this scheme is given in Figure 7-(c). This solution provides smooth transition between the intersection ingress and the egresses without generating hot-spots.

VI. Implementation Aspects

This section deals with a practical implementation of the proposed scheme. This scheme can be consider as an extension of the mobility management mechanism of the current cellular system as GSM [1] and IS-41 [2]. In these systems a number of geographically adjacent base stations are grouped together and connected to a *base station controller* (BSC) that manages their radio resources. The later is attached to a *mobile switching center* (MSC) that connects the base stations to an infrastructure network and provides them a switching functionality for both communication and signaling. All the base stations that are connected to the same MSC define a location area (LA). Each MSC contains also a *visitor location register* (VLR) database that keeps records of all the mobile users that are currently located in its LA. Each mobile user is associate with a single *home location register* (HLR) server that stores its profile information and its location, by pointing on the VLR of the LA where the mobile user resides. When a mobile user crosses an LA boundary it updates its HRL and the relevant VRLs about its location. Upon a call arrival, the system pages the entire LA in which the mobile user resides.

This paradigm of cellular system is suitable to support our scheme. It can easily support the static LAs that are used for locating slow moving users. Here, the LAs should be smaller than the ones used in the current cellular systems for reducing the search cost. We also use the same paradigm to support the MLA systems. Each highway is covered with a dedicated LA termed a *highway LA*. Every highway LA is associated with a single VLR that keeps records of all the fast moving users that travel over this highway. A highway LA is divided into segments, each defines a continuous geographic path and it is included in a single static LA. The base stations at each segment are connected to a single base station controller, termed *highway BSC*. A highway BSC has a dual role, as a BSC of a static LA and a highway LA.

The highway BCSs are the elements that actually manage the MLA systems and it is done in a distributed manner. Consider a given highway and its highway LAs. The highway BCS, that covers the highway starting point, periodically initiates new MLAs and assign an unique MLA-id to each new MLA. Along the highway, every two adjacent highway BSCs are connected for synchronization purposes. This enables fluent movement of the MLAs through the segments. Each MLA may be included in one or few segments simultaneously. Its speed and size are determined by the highway BSCs of these segments in a distributed manner, as we described in Section V. In practice, a highway BSC is not required to know the exact traffic speed in its segment for determining the MLAs' speed. It may use the observation that the cells should experience the same update rate while they are front edge or back edge cells. Otherwise, the MLA speed should be increased or reduced depending when the update rate is higher.

The highway VLR keeps records of the current location of each MLA and the mobile users that are registered to this MLA. When a fast moving user enters the highway it registers itself to the highway LA and to a specific MLA. The mobile usr perform update operation when it crosses the MLA boundaries or when it leaves the highway. When an incoming call arrives, the highway VLR checks in which MLA the called mobile user resides. It send a "search" message to the relevant highway BSCs, that page the MLA for locating the user current cell.

The above description shows that the proposed scheme can be merged with the current cellular systems. It guarantees an efficient usage of the network resources and the management of the MLA systems can be implemented in a distributed manner with a low computational load on the different components of the system.

VII. Simulation Results

We compared by simulations the performance of the proposed schemes with two other schemes, the *location area* (LA) approach [2] and the *distance based* scheme [13]. In the distance based scheme a mobile user performs an update operation when its distance from its last update place exceeds a given threshold, h. When an incoming call arrives, the system pages all the cells inside a circle with radius h around the last update cell. In our simulation we distinguish between three variants of the proposed schemes. The first variant, termed the *dual LA sets*, uses two LA sets without using moving location areas. The second variant, uses in addition moving location areas (MLA) along the highways. The third variant, termed the *MLA + hot-spot removal*, merges the MLA approach with the temporary MLAs method that was presented in Section V-C. Recall that in our experiments, all the simulated schemes use simultaneous paging for locating a mobile user.

We simulated a large urban area covered by a square with $100 * 100$ cells, each cell with a diameter of 1 km. We assume that mobile users are most of the time in a quasi-static state and occasionally travel from place to place. During their journeys, they use highways most of the time. At all the highways, we use a normal speed frequency function with a mean speed of 80 kph and a standard deviation of 15. We assume that the average incoming call rate of a user is $\lambda = 2$ per hour. In the next examples we use the following parameters for the different schemes. At the location area approach, each LA has a square shape with 64 (8×8) cells. For the distance based scheme, we use a threshold $h = 5$. At the moving location area scheme, each small LA contains 10 cells, a highway LA contains 100 cells, and each MLA length is $L = 6$.

We made two types of experiments. First, we evaluate the average number of update operations made by a mobile user during a period of 100 hours, given that it travels p percent of the time. We also estimate the number of paged cells during that period. Selected typical results from our experiments are depicted in Figures 8 and 9. These figures show that the dual LA sets scheme produces low number of update operations but it

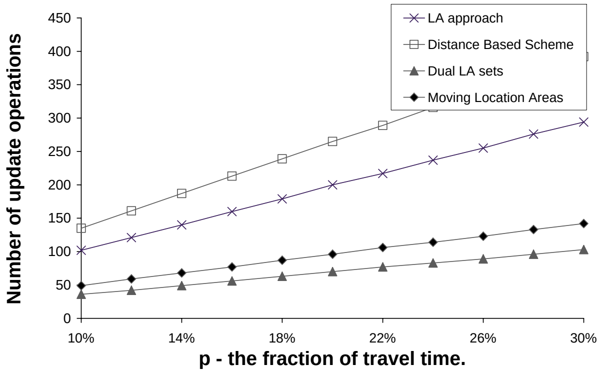

Fig. 8. The average number of update operations per mobile user.

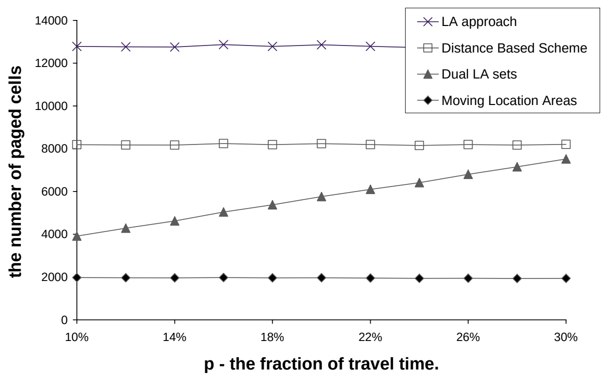

Fig. 9. The average number of paged cells per mobile user.

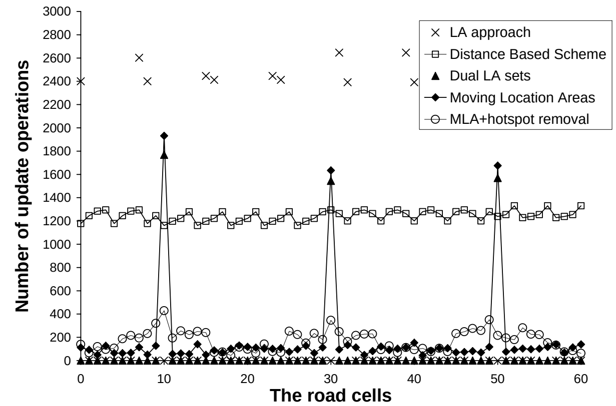

Fig. 10. The number of update operations along a highway segment covered by 60 cells.

relatively pages large number of cells for high values of p. While the moving location area scheme yields both a low number of update operations and paged a low number of cells.

In the second experiment, we simulated a circular highway with 6 intersections that is covered by 120 cells. We evaluated the number of update and search operations at each cell along the highway, after $10,000$ mobile users have traveled over it. This experiment considers only the load of searching mobile users while they have used the highway, and we selected scheme parameters that yield similar search load by all the evaluated schemes (beside the Dual LA sets method). This enables us to compare the update load generated by each scheme. The distribution of the update operations along a highway segment covered by 60 cells are depicted in Figure 10. The location area approach yields a high number of update operations at the border cells of the LA's (cell number 7, 8, 15, 16, 23, 24,...). The distance based approach generated a medium and uniform update load over all the highway cells. The MLA approach generates a low update cost along the highway cells, but also creates hot spot cells at highway intersections (cell number 10, 30, 50). However, the MLA + hotspot removal method yields low and uniform update load along all the highway cells without any hot spot cell. Thus, our experiments show that the moving location area scheme with the hot spot removal method outperform the other schemes for both the update and search operations perspective. Additional simulation results can be found in [18].

REFERENCES

[1] M. Mouly and M. B. Pautet. The GSM System For Mobile Communication. M. Mouly, 49 rue Louise Bruneau, Palaiseau, France 1992.

[2] I. F. Akyildiz and J. S. M. Ho. On Location Management For Personal Communications Networks. *IEEE communication magazine*, Vol 34 No 9. Sep. 1996.

[3] B. Jabbari, G. Colombo, A. Nakajima and J. Kulkarni. Network Issues for Wireless Communication. *IEEE communication magazine*, Vol 33 No 1. January 1995.

[4] J. G. Markoulidaskis, G. L. Lyberopoulos, D. F. Tsirkas and E. D. Sykas. Evaluation of Location Area Planning in Future Mobile Communication Systems. *ACM-Baltzer Journal of Wireless Networks*, Vol. 1, No. 1, March 1995, pp 17-29.

[5] C. U. Saraydar and C. Rose. Location Area Design Using Population and Traffic Data. *Proc. CISS 1998, March 1998*.

[6] S. Okasaka, S. Onoe, S. Yasuda and A. Maebara. A New Location Updating Method For Digital Cellular Systems. *Proc. IEEE VTC 41th, 1991*.

[7] S. K. Sen, A. Bhattacharya and S. K. Das. A Selective Location Update Strategy for PCS Users. *ACM-Baltzer Journal of Wireless Networks*, Vol. 5, No. 2, June 1999, pp 313-326.

[8] J. Scourias and T. Kunz. A Dynamic Individual Location Management Algorithm. *Proc. of IEEE PIMRC-97, 1997*.

[9] M. Shirota, Y. Yoshida and F. Kubota. Statistical Paging Area Selection Scheme (SPAS) for Cellular Mobile Communication Systems. *Proc. IEEE VTC 44th, 1994*.

[10] C. Rose and R. Yates. Minimizing the Average Cost of Paging Under Delay Constraints. *ACM-Baltzer Journal of Wireless Networks*, Vol. 1, No. 2, July 1995, pp 211-219.

[11] B. Liang and Z. J. Haas. Predictive Distance-Based Mobility Management for PCS Networks. *Proc. IEEE INFOCOM 1999*.

[12] G. Wan and E. Lin. Cost Reduction in Location Management Using Semi-Realtime Movement Information. *ACM-Baltzer Journal of Wireless Networks*, Vol. 5, No. 2, July 1999, pp 245-256.

[13] A. Bar-Noy, I. Kessler and M. Sidi Mobile Users: To Update Or Not To Update? *ACM-Baltzer Journal of Wireless Networks*, Vol. 1, No. 2, July 1995.

[14] Z. Lei and C. Rose. Probability Criterion Based Location Tracking Approach for Mobility Management of Personal Communication Systems. *Proc. IEEE GLOBCOM 1997*.

[15] W. R. McShane and R. P. Roess. Traffic Engineering. *Prentice Hall Polytechnic Series in Traffic Engineering*, 1990.

[16] J. G. Markoulidaskis, G. L. Lyberopoulos, D. F. Tsirkas and E. D. Sykas. Mobility Modeling in Third-Generation Mobile Telecommunication Systems. *IEEE personal communication*, August 1997.

[17] M. Hellebrandt, R. Mathar and M. Scheibenbogen. Estimating Position and Velocity of Mobiles in a Cellular Radio Networks. *IEEE Trans. on Vehicular Technology.* Vol. 46 No. 1 pp. 65-71. February 1997.

[18] Y. Bejerano and I. Cidon. Efficient Location Management for PCS NetworksBased on Moving Location Areas. Research Report. Center for Communication and Information Technologies, Technion Haifa, Israel, CC PUB No. 321, July 2000. URL: http://comnet.technion.ac.il/ bej