Location Management in Cellular Networks: Classification of the Most Important Paradigms, Realistic Simulation Framework, and Relative Performance Analysis

Kyandoghere Kyamakya and Klaus Jobmann

***Abstract*—This paper actualizes the classification of location management methods published up to now and presents results of a related extensive performance comparison of the most important paradigms for location management in cellular networks. First, a universal structure of a performance analysis framework for location management methods is defined and analyzed. Then, both a user mobility model and a specific simulation environment claiming to be as realistic as possible are suggested and implemented. Finally, the simulation framework obtained is used for a systematic comparative performance analysis of a representative sample of the most important location management schemes. The overall conclusion is that a location-area based location management method designed around a profile or history-based direction information offers the absolute best performance, in terms of location management cost, compared to all other alternative approaches. The key difference from previous works in literature is clearly underlined concerning both mobility modeling and novel location management schemes.**

***Index Terms*—Location management paradigms, mobility modeling, realistic simulation framework, relative performance comparison.**

I. Introduction

THERE is a significant growth of mobility inside the information society today. This growth of mobility can be seen at three levels: spatial level (national and international roaming), penetration rate (most optimistic surveys have nearly always underestimated the fantastic growth of cellular systems), and the constantly growing traffic generated by each wireless user. From these considerations, it is easy to predict that the generalized mobility features will have a serious impact on future wireless communications networks. One aspect of this impact is related to the question of how to realize a cost-effective location management in terms of necessary signaling effort while coping with highly mobile users in cellular networks.

A cellular network has the following architecture. The geographical coverage area is generally logically partitioned into location areas, which consist of a group of cells. Each cell is served by a base station. Mobile users and their terminals are connected to the network via the base stations. Several base stations are connected to a base station controller, and a number of base station controllers are then connected to a mobile switching center [1], [2].

Mobility can be categorized in two areas: a) radio mobility, which mainly consists of the handover process, and b) network mobility, which mainly consists of location management (location updating and paging). While handover processes are essentially coping with radio aspects (and the main difficulties in improving handover procedures performance come from unpredictable and highly fluctuating radio channel behavior), location management is rather mainly influenced by both user mobility and call patterns [2].

The focus in this paper is on location management. Location management has the function of tracking the location of users (better of their terminals) inside a cellular network. The granularity of this tracking has two levels: a) in the time between calls, the tracking is performed by the location update procedure at the location area (LA) granularity and b) at call arrival, the tracking is performed inside the location area by the paging procedure at the cell granularity. Note that a paging area is generally defined as a subset of the location area. Besides, during a call, in case the serving cell is changed due, amongst other things, to user mobility or fluctuating signal levels, the cell selection is handled by the handover process.

Considering paging (PA) and location update (LU) effort, there are two conflicting trends. To reduce the paging effort to a minimum, the user tracking in the time between calls should be performed at cell granularity (that is, the location area should contain just one cell). So doing, the system would know at each time in which cell the user is residing and the resulting paging effort would be the lowest possible minimum. However, this user tracking at cell level would result in too many LUs if the user has a significant level of mobility compared to the average cell size. On the other hand, when the LA is large, the LU rate does diminish, since it will take longer until a user can cross a LA border (one extreme case is to define the whole network as a single LA; then there will be no LU traffic at all). In this case, however, the paging effort will increase, since at call arrival, a larger number of cells will have to be paged. That is why it is generally stated, in the literature [1]–[7], that if the location cost is high (and thus user location knowledge is accurate), the

Manuscript received March 16, 2002; revised November 24, 2003; August 4, 2004. This work was supported in part by the German Research Foundation under Grant Jo 281/4-1. The review of this paper was coordinated by Prof. Y.-B. Lin.

The authors are with the Department of Electrical and Computer Engineering Department, University of Hannover, D-30167 Hannover, Germany (e-mail: kyandogh@ant.uni-hannover.de).

Digital Object Identifier 10.1109/TVT.2004.842459

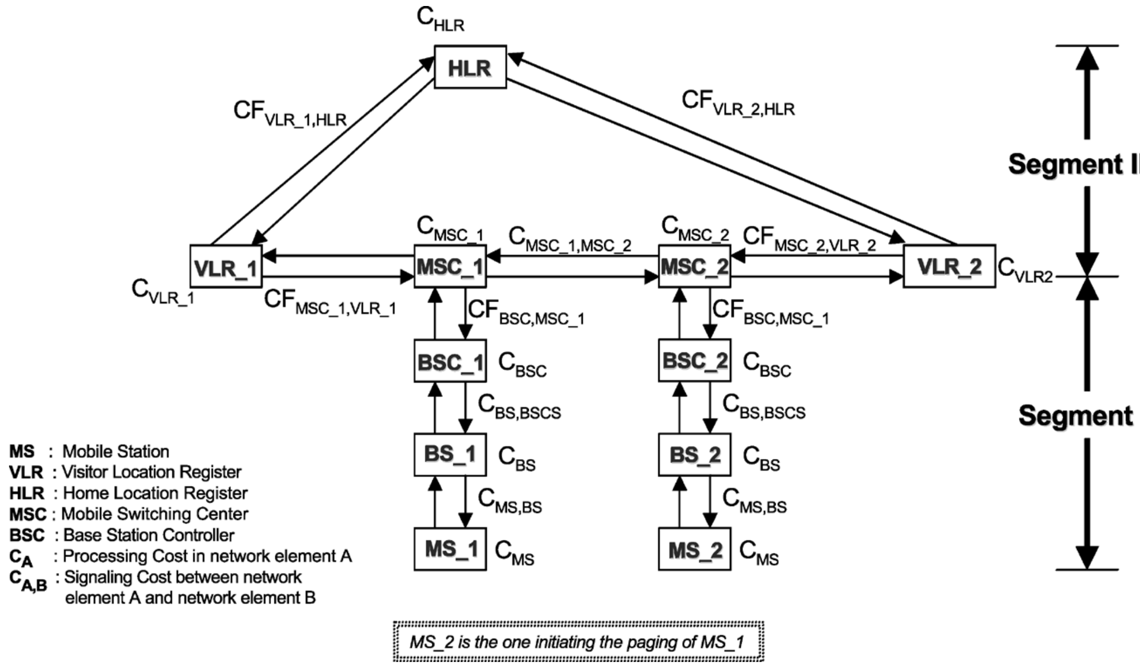


Fig. 1. Cost components (or structure) of the overall location management effort (including location updates and paging).

paging cost will be low (paging messages will only be transmitted over a small area). If the location cost is low (and thus user location knowledge is fuzzy), the paging cost will be high (paging messages will have to be transmitted over a wide area). Therefore, the central problem or challenge in location management is to design algorithms that minimize the overall effort for both LU and PA.

A series of good survey papers on location management in wireless networks [1]–[3] do provide a comprehensive overview of various handover and location management techniques published to date. One overall notice is that all optimization strategies for location management do turn around two key issues that all have an impact on the overall location management effort: the design of LU and PA algorithms and the design of appropriate database architectures and related management and caching strategies. Fig. 1 shows the structure of the overall location management effort, which consists of signaling costs on the one hand and processing costs in different network elements on the other hand. Depending on the specific scheme (LU, PA, database management, caching scheme), the overall cost will be the sum of all related costs related to the network elements and signaling links involved in the process.

While database architectures have an impact solely on the cost elements of Segment II (in Fig. 1), PA and LU schemes do have an impact on both Segments I and II (of Fig. 1).

This paper's focus is on PA and LU schemes and less on database architectures. Nevertheless, just to give a short indication, as far as database architectures are concerned, different schemes aiming at a reduction of related signaling traffic have been published to date. The following related approaches have retained attention so far: hierarchical partitioning [4], load balancing [1], data replication [5], and caching strategies [8], [9].

In essence, the main focus of this paper is on the following issues.

1) Define a universal structure for a performance analysis framework for both location update and paging schemes.
2) Suggest a simulation/performance analysis platform for both location update and paging schemes. This platform should contain the following modules: a realistic (or rather, for analysis purposes, sufficient) and/or parameterized model for mobility in time and space; a realistic (or rather, for analysis purposes, sufficient) model of the geographic context; and a realistic (or rather, for analysis purposes, sufficient) model of user call behaviors. This platform should be a simulation framework which has the claim of being nearly realistic (or rather for analysis purposes sufficient) and satisfies a very important concern formulated in [2]: although various location update and paging schemes have been proposed in the literature, their relative performance was not yet clear. This is mainly due to the fact that different models and assumptions were being used to evaluate different schemes. Due to the difficulty of developing test beds or field trials simply for performance comparisons, it is obviously desirable to have a generic analytical or simulation model that can analyze various update and paging schemes in a realistic environment.

   A similar concern has been expressed in [6], where the difficulty of comparing different mobility management schemes is underlined. It is in fact difficult to compare various mobility management schemes, since as yet there exist no absolute bounds on optimum performance that can be applied to any procedure regardless of assumptions about mobility, network structure, etc.
3) Using the simulation framework so defined, determine or generate a series of scenarios that will be the basis for sensitivity studies, which will help finding out under which circumstances each of the location update and paging approaches performs best.

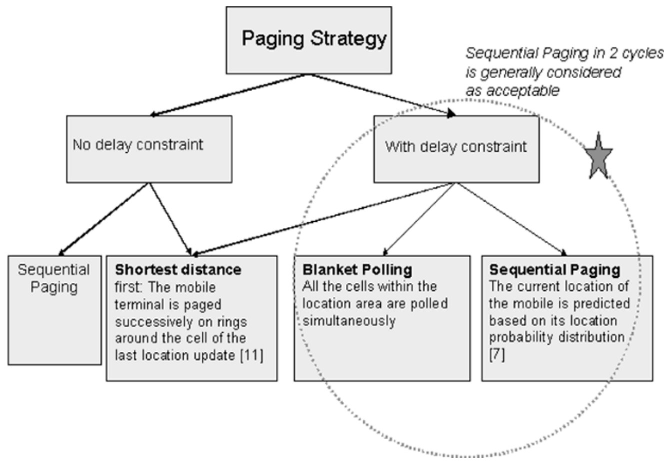


Fig. 2. A classification of different paging schemes [2], [7], [10], [11] ([11] can also be used with delay constraint. In this paper we do, however, focus on blanket polling and sequential paging).

4) Present and discuss simulation performance evaluation results for a representative sample of location management schemes (some being already known and some being our own design).

In the recent literature, a series of static and dynamic location update algorithms and various selective paging algorithms have been proposed. Section II contains a summary of the most important ones. An overview of relevant teletraffic issues and different user mobility models that have been used for performance analysis of location update and paging is presented in Section III. Section IV presents our concept of a nearly realistic (or rather, for analysis purposes, sufficient) simulation framework, while Section V focuses on the description of the sample location management methods implemented in this paper. Section VI presents and comments on simulation results, and Section VII summarizes the main results.

II. Classification of Location Management Methods in Cellular Networks

Since location management consists of two main procedures, a location update scheme and a paging scheme, a classification of location management approaches should consider these two components.

A. Paging Schemes for Mobile Terminals

The paging process aims to find the cell in which the mobile terminal (MT) is actually located. Since MT location is known by the network at the LA granularity, the paging process consists of now polling the cells inside the LA until the target MT replies to the page message. The LA can be divided in cell subsets that are called paging areas, which will be polled sequentially [2]. However, since paging must be performed within a fixed time constraint (called maximum paging delay), each polling cycle should be limited by a timeout period. In the recent literature, it has been often stated that one to three polling cycles are acceptable for practical implementations. Furthermore, the paging cost is proportional not only to the number of polling cycles but also to the number of cells polled in each cycle. The paging area can be either set statically or be dynamically determined by a prediction based on available profile information (related to short- and/or long-term movement history) provided by the location update function. A classification of various paging strategies is presented in Fig. 2. For the rest of this paper we will concentrate on the paging approach with delay constraint, as it has been started to be more realistic [2].

B. Location Update Schemes

Location update is performed either from time to time or whenever the mobile terminal crosses the LA borders. The LU procedure (also called registration) begins with an update message sent by the mobile terminal over the uplink control channel, which is followed by some signaling procedures to update the databases [2] (see also Fig. 1).

Location update algorithms are either static or dynamic. In a static algorithm, the location areas are static and are the same for all users. Thus, LU is triggered based on the sole network topology (that is, the LA's geographical structure and size). Note that the LA's structure and extent are the same for all users, independently of eventual differences in their call and mobility patterns.

In a dynamic algorithm, however, LU is personalized in that it is rather based on the user's call and mobility patterns. This is due to the fact that the LA structure is user specific and changes whenever the user changes his mobility and call patterns. Fig. 3 classifies various LU schemes published in the most recent literature. Note that both the classification of mobility management methods by Tabane (see [1, p. 75]) and that of Biesterfeld (see [24, p. 72]), which are almost solely "location-area-based," are not as complete as that presented in Fig. 3. Reference [2] presents a more complete classification, whereby, however, the direction-based approach is missing. Thus, Fig. 3 is an exhaustive presentation of all possible classification criteria for location update schemes known to date.

III. Proposal of a Mobility Model for Cellular Networks

Mobility models play an important role while examining different issues in wireless networks, including resource allocation, handoff, and location management. In general, the mobility models depend on speed, direction, and movement history of the mobile users.

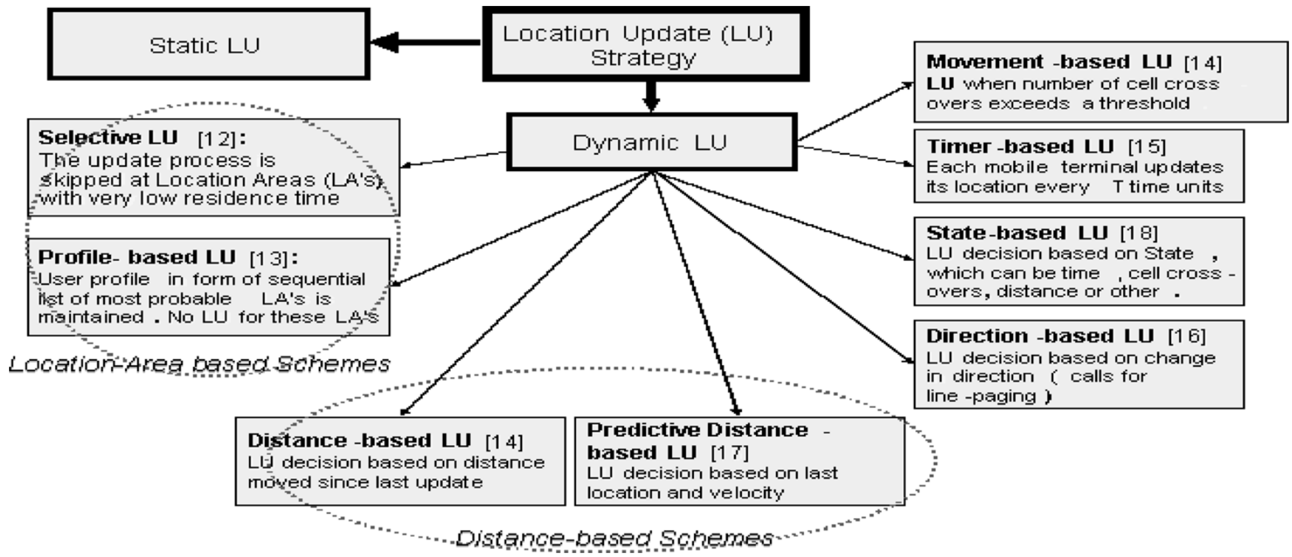


Fig. 3. LU method classification [12]–[18]. (Compared to the classification in [2], this classification adds a new approach, the direction-based one.)

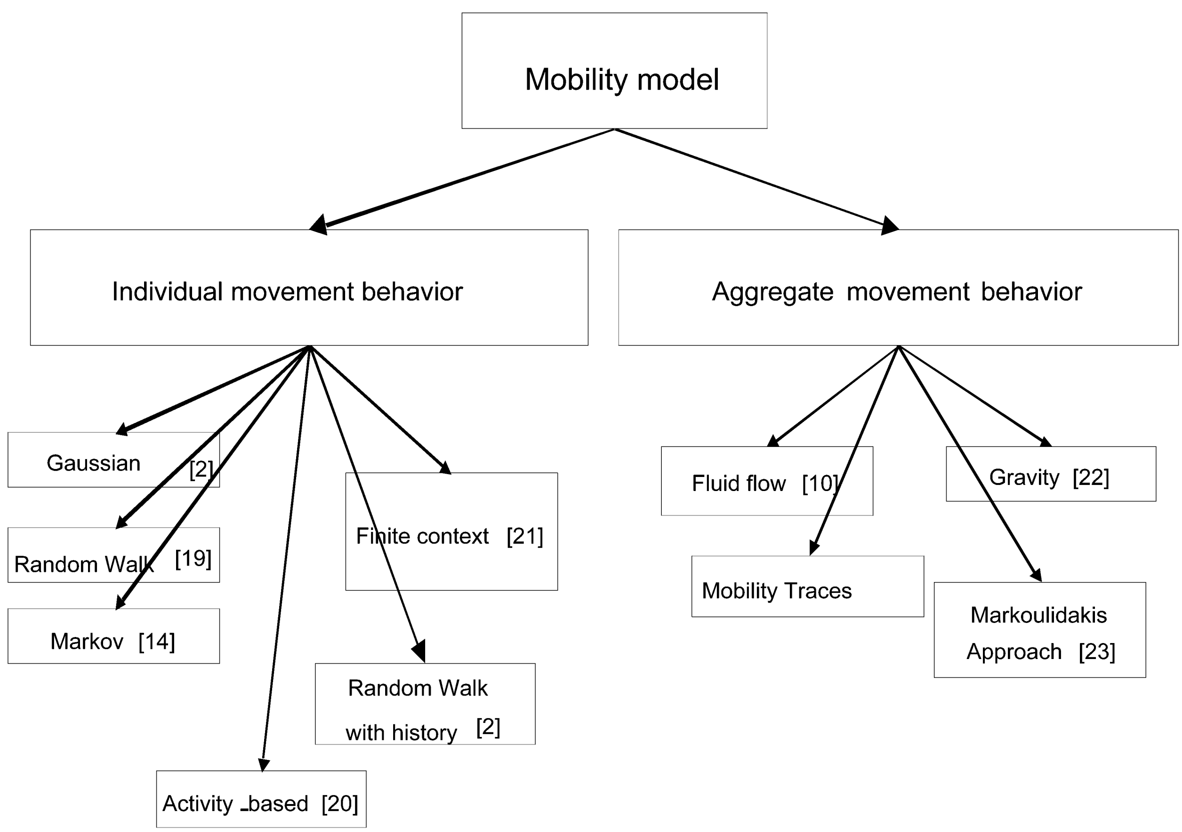


Fig. 4. A classification of different mobility models [2], [10], [14], [19]–[23]. (The method "Random Walk with history" is different from "Pure Random Walk" in that it has memory, since the last one is rather memoryless.)

A series of previous research works on the teletraffic in mobile communication networks have tried to seize the mobility characteristics as stochastic distributions of speed, cell dwell time, and cell crossover rates of mobile terminals [10], [25], [26]. A related classification of different mobility models is presented in Fig. 4. On the one hand we do have the "individual movement behavior"-based models, and on the other hand the "aggregate movement behavior"-based models.

In this section, we propose a mobility model that:

a) adopts the activity-based approach for the movement control logic of the individual behavior;
b) takes the fact into consideration that aggregate traffic characteristics in the space do influence the evolution of the individuals trips: congestion, velocity profiles, routing, etc.;
c) generates reference mobility profiles which can then be used to develop a Markov model with history.

This is a summary of the new mobility model we have proposed in [27]. The goal of the modeling of the space dimension is to create a virtual space/area (region, city area, etc.) with a complete representation of the space's individuals, their activities, and the transportation infrastructure. Trips are planned to satisfy the individuals' activity patterns. The model then simulates the movement of individuals across the transportation network, including their use of vehicles such as cars or buses, on a second-by-second (or even coarser if necessary) basis. This virtual world of travelers mimics the traveling and driving behavior of real people in the area/region. The interactions of individual vehicles produce realistic traffic dynamics from which mobile network designers, analysts, and operators can estimate the mobility-related performance parameters of their networks.

After a thorough review of the models described in the literature (see [27]) we could sort out elements that can sustain the design of a mobility model suitable for the analysis of location management schemes in cellular networks. The model integrates those aspects from already published approaches that, from our point of view, better fit the general requirements of location management. So doing, we reduce to a minimum the potential drawbacks of this integrated approach. In fact, as extensively discussed in the literature, a mobility model should generally fit to the application context. That is why models for both mobility and the geographic context have been optimized for different application scenarios; some examples follow.

1) Ad hoc networks (a microscopic model is needed, with an accuracy of some 10–100 meters, since issues like radio propagation, energy control, and radio-resource management are of particular importance here) [28].

2) Cellular networks, but with focus on the analysis of radio-propagation-related aspects (e.g., slow fading analysis, handover-related studies, analysis of offered traffic loads in micro cells), where the user location is considered at the accuracy of a few wavelengths. Microscopic models are also needed for this case [23].

3) Cellular networks, but with focus on location management. The required accuracy for user location is at the cell level [20]–[23].[1]

To fit to the reality, our model has been adapted to the specific context of location management in that it does integrate both aggregate and individual/microscopic considerations. Indeed, the model developed involves the following dimensions.[2]

- The space dimension, where user movements take place. The modeling of the geographic space of interest is up to the accuracy of street segments. The issue here is to seize the topological data/constraints of the geographic area in which user movement is taking place. Appropriate models are fortunately already available for many parts of the world. In Europe, for example, just for illustration, road map data of more than nine millions links are available.[3] This is also the case for other industrial nations (United States, Canada, Japan, etc.). These data are already intensively used in GPS-assisted navigation systems for example.

  A simulation tool commercially available that can be used for this modeling is VISUM.[4] Other good tools worth mentioning are VISSIM[5] and TRANSIMS,[6] for example. There are many other commercially available tools that are very good for the task.
- The modeling/estimation of aggregate traffic state profiles on the streets, while the user is moving from a place A to a place B in the city. Even if the user individual movements are the ones that are of interest, they are highly influenced by the situation created on roads by the aggregate flows. For example, the occurrence of congestion on a road segment can result in a rerouting or in the delaying of a trip. Here too, adequate commercially available tools and data should be used for the task (see Fig. 5). VISUM, or the already cited tool TRANSIMS, does this task as well. For VISUM, the input data for this purpose (i.e., the estimation of the daily profiles of the total number of trips originating from an "area zone i" to an "area zone j") must be generated by another tool called VISEM.[7]
- Location, timing, and sequencing of individual user movements. As the activity-based sequencing approach is generally accepted as the more realistic one [29]–[32], we take it as one of the pillars of our model.

  Our user mobility modeling (activity-based approach) is based on the following elements.

  a) Number of activities of interest for a user, which generally take place at different spatial locations.

  b) Time zones for each activity: during a day, it can be observed that there are time periods during which

[1]The Sumatra Homepage: http://www-db.standford.edu/sumatra.

[2]Modeling of the space dimension is traditionally an area of expertise of traffic planners and geographers. Note that several commercially available tools have been developed for this purpose. Because of the complexity of the modeling process of the geographic dimension, mobile network planners do not have the necessary expertise to develop other specific tools to perform the same task. It is therefore recommended to use already well-established and specialized tools for this process. The data produced can then be imported to other mobile-network-specific tools. This is what has been done during this work.

[3]See http://www.english.ptv.de/cgi-bin/produkte/demand.pl.

[4]VISUM is an information and planning system for network analysis and forecast. It combines all relevant aspects of private and public transport planning in one comprehensive transportation model. The application area reaches from strategic network planning to operative service planning and detailed network optimization. A vast data supply (network and geographic database covering all of Europe and consisting of over nine million different routes) and a huge number of interfaces enable the creation of powerful platforms for information systems. See http://www.english.ptv.de/cgi-bin/produkte/visum.pl.

[5]VISSIM can model more details than other microscopic traffic simulators. Doing this, all motorized road users as well as crossing pedestrians are considered. VISSIM is suitable for both inner- and outer-urban traffic, so that the application possibilities are enormous. See http://www.english.ptv.de/cgi-bin/produkte/vissim.pl.

[6]The Transportation Analysis and Simulation System (TRANSIMS) is an integrated system of travel forecasting models designed to give transportation planners accurate, complete information on traffic impacts, congestion, and pollution. Los Alamos National Laboratory is leading this effort to develop new transportation and air quality forecasting procedures required by the Clean Air Act, the Intermodal Surface Transportation Efficiency Act, and other regulations. It is part of the Travel Model Improvement Program sponsored by the U.S. Department of Transportation, the Environmental Protection Agency, and the Department of Energy. See http://transims.tsasa.lanl.gov/.

[7]VISEM is a Transport Information System—Generation Model. VISEM supports the transport planner in travel demand calculation based on a demand model. VISEM is applied for generation of travel demand matrices based on a demand model that basically considers activities and connects these with a mobility program. This model distinguishes between several behavior-homogeneous population groups and generates a specific set of activity chains for each of them. VISEM runs under Windows(95, 98 up to XP). See http://www.english.ptv.de/cgi-bin/produkte/visem.pl.

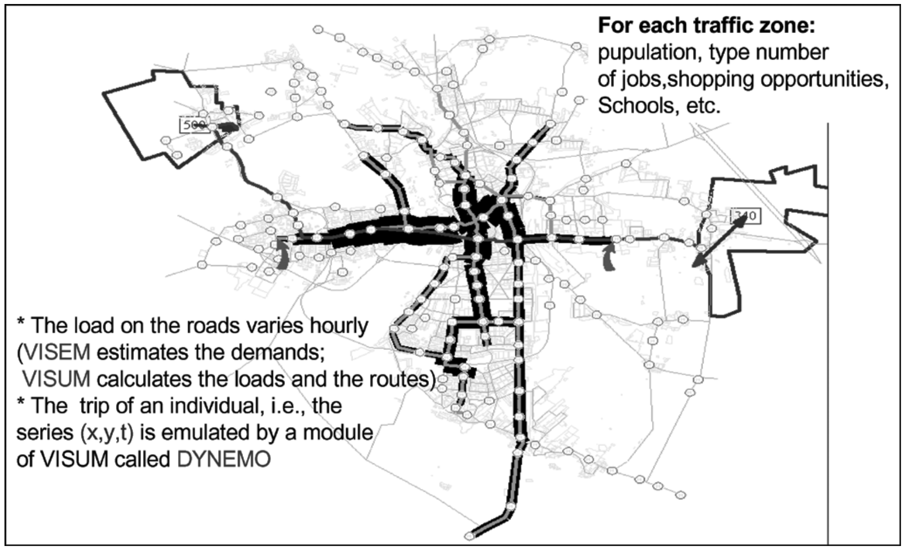


Fig. 5. Modeling of the space dimension—town of Hannover and surroundings, city area modeling level. Impact of aggregate flows on individual movements.

certain types of movement take place (e.g., movement toward work places) and time periods where certain population groups reside at certain attraction points (e.g., working hours, shopping hours, etc.).

c) Activity duration profile: for each activity, the duration profiles can either be obtained from panel survey data (in this case, they are recorded in what we call "classical activity duration matrix") or set by the analyst (in this case they are recorded in what we call "general activity duration matrix"). The classical activity duration matrix is thus just a particular constellation or configuration of the general activity duration matrix.

d) Activity sequencing profile: an activity is selected based on the previous activity and the current time period. The probability of transition from one activity to another is determined by (or extracted from) the activity transition matrix. In case the transition matrix has been calculated from panel survey data, we call it "classical activity transition matrix." Otherwise, in case it is freely set by the analyst, we call it "general activity transition matrix." The classical activity transition matrix is thus just a particular constellation of the general activity transition matrix.

e) Geographic location of activities: for each activity, we must determine its location. The location can be estimated using the geographic distribution of the different movement attraction points (MAPs)—for example, shopping centers, work places, residences, hospitals, schools, etc. [30]. Furthermore, the location of an activity can be set to be either deterministic or stochastic. An example of a stochastic location for an activity is the following: for shopping, I can use many different shops located in different parts of the town depending on what I want to actually buy or where I am actually or just depending on mood or preferences.

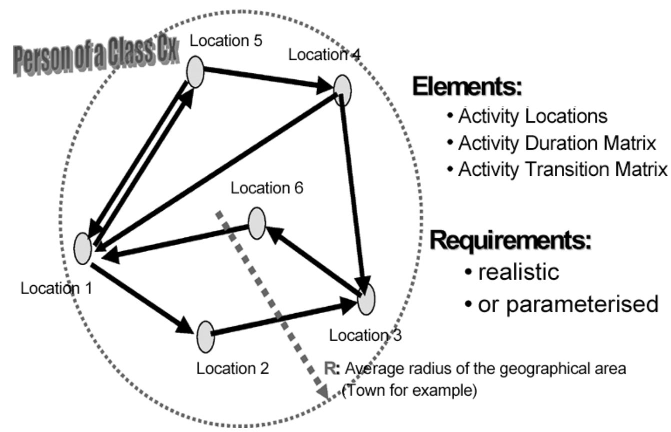


Fig. 6. Illustration of the activity-based mobility model used in this paper.

- The average diameter/radius of the geographic area where the long-term user movement is taking place. This area is for example, in our case, the town where the user is living.
- The emulation of realistic user movements. Based on the preceding elements, it becomes possible to emulate realistic user movements in space and time.
- The call model. This model should be able to emulate all possible incoming call arrival patterns and not only the generally assumed Poisson distribution. Both the distribution and the intensity of the call patterns should be easily variable.

The mobility model obtained is summarized in Fig. 6. The radius of the geographic area limits the movement space. Activity locations are set either randomly or according to structure data of the area of concern (i.e., the geographical distribution of the so-called MAPs). The model not only has the necessary degree of randomness but also reflects the routine in daily activity observed by social and transportation scientists.

For a systematic user classification according to their mobility behavior (number of activities per day, activity timing), most of the studies have relied on survey data, and the classification criteria in the latest studies (see [31], for example) has been

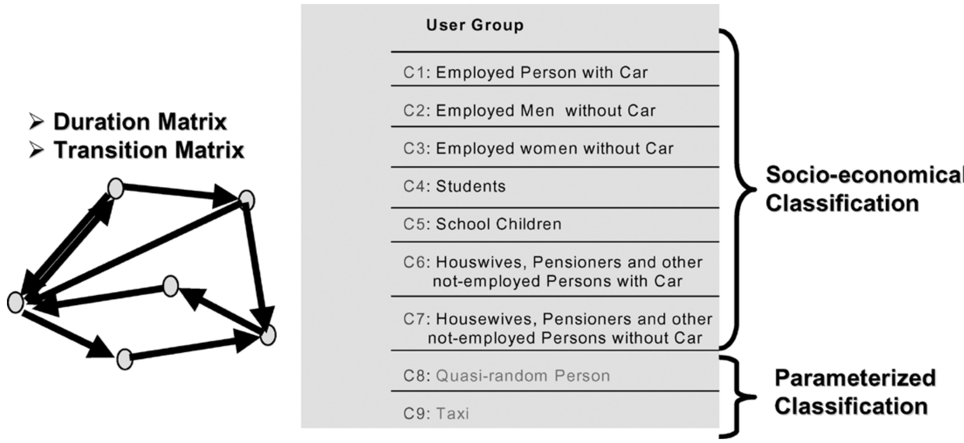


Fig. 7. Illustration of the user classification used in this paper.

(Note: for Groups C1–C7, Duration Matrix and Transition Matrix are extracted from survey data; for Groups C8 and C9, both matrixes are randomly selected, parameterized.)

TABLE I

PROFILE TYPES FOR RECORDING MOVEMENT HISTORY. THIS TABLE DESCRIBES THE DIFFERENT FORMATS THAT HAVE BEEN NECESSARY FOR THE ADEQUATE RECORDING OF THE MOVEMENT HISTORY INFORMATION (THAT HAS TO BE RECORDED BY THE MOBILE DEVICE AND TRANSMITTED TO THE NETWORK DURING THE LOCATION UPDATE PROCESS) FOR THE DIFFERENT LOCATION MANAGEMENT METHODS

Profile type or identification	Profile content description (i.e., data that are recorded)
0	No profile (any data recorded).
1	For every visited cell, the updated average dwell time is recorded.
2	In essence, a Markov model, i.e., for every visited cell, both the updated average cell dwell time and the updated transition frequency to neighboring cells are recorded.
3	In essence, it is a Markov model with history, i.e., similar to "*Profile Type 2*", however, with the restriction that the recent history is considered, i.e., the cell previously visited is associated correlatively (see Fig.8). In other words, a '*profile-based direction*' is considered. Note that "*Profile Type 2*" does consider long term movement history, whereas "*Profile Type 3*" does consider both long term and recent (that is, short term) history.
4	Beside call history and '*Profile type 1*' related data, also movement history information in form of *trip segments* [30] are recorded.

the so-called time budget in either activity locations or on trips between activity locations. The classification process is, basically, a user population segmentation. It is, in essence, a segmentation on the basis of attributes describing the decision-maker's preferences. It is used to create groups that are assumed to be homogenous in taste, and therefore their members are assumed to react in a similar manner to travel-related policy changes.

This socioeconomic user classification corresponds to the classes C1–C7 of Fig. 7. One key question to be answered in this paper is whether this classification is of any relevance from the point of view of the mobility-dependent performance metrics of a cellular network. The answer is given in Section VI.

IV. Representative Sample of Location Management Methods Implemented and Considered in This Paper

This section presents a sample of location management methods implemented in this paper. While some of them have been already published in the literature, the remaining have been developed by us, either as an extension of known and published ones or as fully new concepts. One concern has been to select a representative sample of the most important schemes for location management in cellular networks. The recording of movement history is an integral part of many of the location management methods. The different forms of the profiles describing movement history are presented in Table I.

The modification of published methods or the design of new ones has been motivated by a thorough analysis of performance simulation results of the original ones (known and published). The identification of weaknesses of the known methods, as depicted by the analysis, has opened the door to either an extension of the original schemes or the design of completely new ones. As such, the whole process has been rather an iterative one, which has led to the selection and implementation of the methods summarized in Table II.

TABLE II
KEY FEATURES OF THE LOCATION MANAGEMENT SCHEMES IMPLEMENTED IN THE FRAME OF THE PERFORMANCE ANALYSIS CONDUCTED IN THIS PAPER (LU: LOCATION UPDATE)

Method ID	LU scheme	Paging / Polling scheme	Profile Type used (see Table I)	Author(s)
GSM -Classic	Static LU	Blanket	0	GSM Standard
GSM+Profiles	Static LU	Sequential, 2 steps	1	Our extension of "GSM-Classic"
Direction -based	Direction -based	Blanket	0	Hwang et al. [16]
Direction -based_Sector	Direction -based	Blanket	0	Our first extension of "Direction -based"
Direction -based+History	Direction -based	Sequential, 2 steps	1	Our second extension of "Direction -based"
Direction -based_Sector + History	Direction -based	Sequential, 2 steps	1	Our third extension of "Direction -based"
Movement -based	Movement -based	Blanket	0	Bar-Noy et al. [14]
Movement -based+ History	Movement -based	Sequential, 2 steps	1	Our extension of "Movement -based"
Distance -based	Distance -based	Blanket	0	Bar-Noy et al. [14]
Distance -based + History	Distance -based	Sequential, 2 steps	1	Our extension of "Distance -based"
BIEST	LA-based	Sequential, 2 steps	4	Biesterfeld [24]
BIEST_KYA	LA-based	Sequential, 2 steps	3	Our modification of "BIEST"
SCOURIAS	LA-based	Sequential , 2 steps	2	Scourias et al. [20]
SCOUKYA	LA-based	Sequential, 2 steps	2	Our modification of „SCOURIAS "
SCOUKYA_V2	LA-based	Sequential, 2 steps	3	Our modification of „SCOUKYA "
KYAMA	LA-based + Timer-based	Sequential, 3 steps	3 (extended)	Our own method

Before we start presenting the different methods, it is helpful to recall the key components needed for efficient location management.

a) Location update: this is performed according to one of the schemes mentioned in Fig. 3.
b) Paging: either blanket or sequential polling.
c) Movement profiles/history: methods using sequential paging need information about movement history, which is recorded in a way that must be consistent with the location update scheme. For this reason, we have defined different formats/types for the movement profiles, as they can be needed by the different location update schemes. Especially in the case of location-area-based schemes, the location area design process strongly imposes a particular format for the profile or movement history information.

Table I presents the different formats that have been necessary for an adequate formulation of the profile/movement history information, as required by the different location management schemes studied. Table II summarizes the key features of the selected methods, that is, the respective location update and paging schemes, and the profile information type respectively required or used.

Furthermore, let us point out a particular aspect. Some of the location management methods require information about movement history to perform properly. However, it can happen that this information is not yet available, especially for areas/cells that have not yet been visited by the user. For these cases, a backup or fallback location method is needed, that is, an alternative location management scheme that can perform without movement history information. Some methods take the "GSM-Classic" scheme as fallback, whereby others rather take an improvement of this last one, which is called here "GSM+Profile" (see a description of this scheme in the next section).

A. The GSM-Classic and GSM+Profiles Methods

The GSM+Profiles method is proposed to improve the performance of the GSM-Classic (defined in the GSM standard) method by involving mobility history. So doing, it becomes possible to use a better paging scheme than the blanket polling, i.e., a sequential paging (in two cycles). During the daily user movements, the average cell dwell time Tc that the user spends in a cell c is maintained. As the user enters a cell, a timer is started. The final value of the timer is used to update the average value Tc. Only cells already visited have a data entry.

The location update is performed as in the Classical GSM method, that is, whenever the user crosses LA borders. The LAs are similar to the GSM-classic ones, that is, LAs are static and the same for all users.

For paging (sequential paging in two steps), however, the question is that of determining which specific cells inside the LA to page first, given a corresponding list of Tc values. The average[8] value of Tc among all cells in the current LA is calculated, and cells where Tc is greater than or equal to the average form the paging area used in the first round (or cycle) of paging. If this paging attempt is not successful, all remaining cells of the location area will be paged in the second round of paging.

B. Scourias Method

This method has been selected because of the comparative simulations conducted by its authors [20], under their given simulation context, which show that it performs very well while compared to the classical GSM-classic method. Further, the method attempts to utilize mobility history, or user profile, to dynamically create location areas for individual subscribers and to determine the paging areas for a two-step polling scheme.

The user profile contains the number of transitions the subscriber has made from cell to cell, and the average duration of visits to each cell. Specifically, a counter Ca,b is kept for each cell a of the number of times the subscriber has moved from cell a to cell b. The average time Tb that a subscriber spends in a cell b is also recorded or updated. The information stored for the whole cellular network can be represented by a directed graph (i.e., a Markov model). The nodes represent visited cells and the links represent transitions between cells. The weight of a link (a,b) is the value of Ca,b and the weight of node b is the value of Tb.

LAs are user-specific and LU is triggered, of course, whenever a user enters a new cell not belonging to the previous location area. The important question is how location areas are designed. The location area is determined in the following manner (starting from the cell the user is actually located in after a location update; the movement history information needed is of "profile type 2").

Step 1) All neighboring cells corresponding to links whose weights are greater than or equal to the average weight W are added to the location area in an order according to the link weight. Note that paths/routes that the subscriber often takes correspond to higher values of the link weights.

Step 2) Once the first ring of neighboring cells has been added to the LA, the above step is repeated using the newly selected cells by decreasing link weight order.

Step 3) These steps are repeated until no other cells are left for inclusion. Note that there is no limitation of the number of cells per LA.

For circumstances (visited areas/cells) where movement history information is not yet available, GSM-Classic is taken as fallback scheme.

Concerning paging, it is performed similarly to the way it is done for GSM-Classic + Profiles, i.e., sequential paging in two steps.

C. The SCOUKYA Method

This is our improvement of the SCOURIAS method in two aspects. 1) The fallback method: instead of using the GSM-classic as fallback, in cases where movement history information is not yet available, we rather use GSM+Profiles as fallback (this has the potential of reducing the paging effort, which has also been confirmed by the simulations). 2) The location area is designed differently, in that the Steps 1) and 3) of the original location area design process (i.e., of the method SCOURIAS) are modified. The new LA design is explained in the following.

The location area is determined in the following manner (starting from the cell in which the user is actually located, while using "profile type 2" movement history information).

Step 1) *Modified:* All neighboring cells corresponding to links whose weights are greater than or equal to one are added to the location area in an order according to the link weight. This modification is sustained by simulation results that have been partially presented in [33]. In fact, the authors of the SCOURIAS method [20] did realize that their method performed quite well for users with a high degree of regularity in their mobility profile and quite bad for users having a certain degree of randomness in their mobility behavior. However, they failed to give the reasons for this. We could find an explanation of the phenomenon; we explain it as follows. Prediction (as the LA design process is in essence a prediction of future user movement) is basically based on mobility history. A deterministic user is allegiant to his movement history, and prediction methods using this history will be successful. However, for a random or fuzzy user, it is rather per se misleading to base prediction on history. Since a random movement behavior tends, in essence, to contradict history, if one blindly bases a prediction on history—and Scourias *et al.* do this in their prediction approach [20]—one will surely fail.

Therefore, the challenge is to design a location prediction method that performs well for both random and deterministic users.

The modification of Step 1), as done here, has the purpose of reducing the strong dependence on movement history (simulation results have confirmed the rightness of this approach).

[8]It is recommended to rather take the quadratic average.

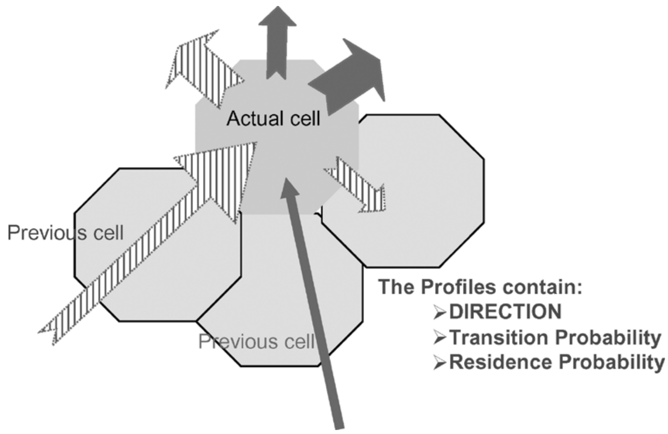


Fig. 8. Illustration of the profile-based direction concept.

Step 2) *Unchanged:* Once the first ring of neighboring cells has been added to the LA, the above steps are repeated using the newly selected cells by decreasing link weight order.

Step 3) *Modified:* These steps are repeated until the location area size has reached a predefined size (for example four, eight, 16 or 25 cells) or no other cells are left for inclusion. The modification here consists in the limitation of the location area size.

For circumstances (i.e., visited areas/cells) where movement history information is not yet available, GSM+Profiles is taken as fallback scheme.

D. The SCOUKYA_v2 Method

This is our modification of the SCOUKYA method in a way such that instead of using a "type 2 profile," a "type 3 profile" is used for movement history information. Apart from this, the formulation of the LA design process and the fallback scheme are the same as for SCOUKYA.

In this approach, a Markov model with history (see Fig. 8) is used and expresses what we call in this work, a "profile-based direction." In fact, the movement prediction will now depend on the cell that has just been previously visited.

E. The BIEST Method

As explained in [24], this method integrates the use of "type 4 profile" for movement history information. The LA is built in an iterative process in the following manner (see Fig. 9).

Step 1) One starts from the actual cell after an LU.

Step 2) Add a cell from the profile to the LA.

Step 3) Repeat Step 2) until the total resulting paging cost exceeds LU cost. Note that call history is used to estimate the expected paging cost [24]. Besides, there is no limitation of the LA size.

The paging is also performed sequentially in two steps, similarly to GSM+Profiles.

For circumstances (visited areas/cells) where movement history information is not yet available, GSM-Classic is taken as fallback scheme in the original formulation of this method. In our implementation, however, we take GSM+Profiles as fallback scheme.

F. The BIEST_KYA Method

This is our modification of the BIEST method in that we do the following.

- Rather use the "profile type 3" for movement history information: this change has been motivated by the fact that simulations have shown that BIEST performs well only when the user behavior contains sufficient regularity in both movement and call patterns. For a random/fuzzy user (in both movement and call pattern), however, the method has a rather bad performance, whereby the dependence on call history appeared to be most destructive.
- The LA size is no more determined by the involvement of call history. Instead we fix a constant LA size (16 or 25 cells/LA, for example) independently of both movement and call history.

Thus, Step 3) of the LA design process in BIEST is modified in the following way.

Step 3) *Modified:* Repeat Step 2) until maximum LA size has been reached.

Apart from this, everything is similar to BIEST, including the fallback scheme (GSM+Profile).

G. The KYAMA Method

This method is an integration of both LA-based and timer-based location management paradigms. The LAs considered are static ones, similar to GSM+Profiles. It introduces a two-level hierarchy of the LA architecture. Further, it essentially builds on the "profile-based direction" idea, similarly to SCOUKYA_v2.

The method consists of the following elements:

1) The first level of the LA architecture consists of static LAs similar to GSM+Profiles. The static LAs are the same for all users.
2) The concept of macro LAs is introduced. The macro LA is user specific (contrary to the static LAs). For a mobile station, at a particular time, a macro LA consists of two (or, in principle, even more) LAs. Thus, the macro LA is made of the "actual LA" (let us call it LA1) and the "next LA" (let us call it LA2). The "next LA" is the LA that, according to a prediction based on movement history (of profile type 3), is assumed to be visited next by the mobile station (see Fig. 10).
3) While the user is moving, movement history in the form "profile type 3" is recorded. "Profile type 3" information is recorded not only for the cell level but also for the LA level.

   For an LA, the so-called "Input Cell" and "Output Cell" are defined (see Fig. 10). For the Markov model with history at LA level, the previously visited element not only is the previous LA but also the corresponding last visited cell (of that previous LA) is integrated in the records. The "Input Cell" for the actually visited LA is at the same time "Output Cell" for the previously visited LA.

   The respective minimum and maximum LA dwell times are also recorded. At the end we have, in fact, an extended "profile type 3" format for the movement history information records.

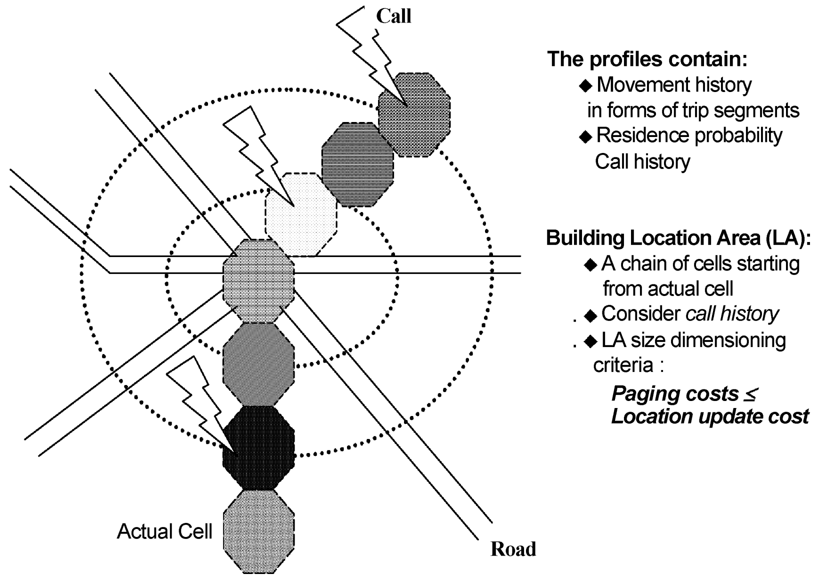


Fig. 9. Illustration of the BIEST method.

4) An LU is generated whenever the border of the macro LA is crossed.
5) After LU, a new determination of the macro LA is performed, i.e., in addition to the actual LA, select the "next LA" based on the movement history information recorded in "extended profile type 3."
6) For paging, we introduce the concept of "paging area" that is a subset of the macro LA and that will be sequentially paged at call arrival. The time Tp that has passed since the last LU is used to cut down (subdivide) the macro LA.
    a) If Tp is higher than the maximum dwell time in LA1, then

$$\text{paging area} = \text{LA2}.$$

    b) If Tp is smaller than the minimum dwell time in LA1, then

$$\text{paging area} = \text{LA1}.$$

    c) Otherwise

$$\text{paging area} = \text{LA1} + \text{LA2}.$$

    d) After the "paging area" is determined, paging is performed sequentially in two steps (over the "paging area," similarly to GSM+Profiles).
7) In cases where sufficient movement history is not yet available, GSM+Profiles is taken as the fallback scheme.

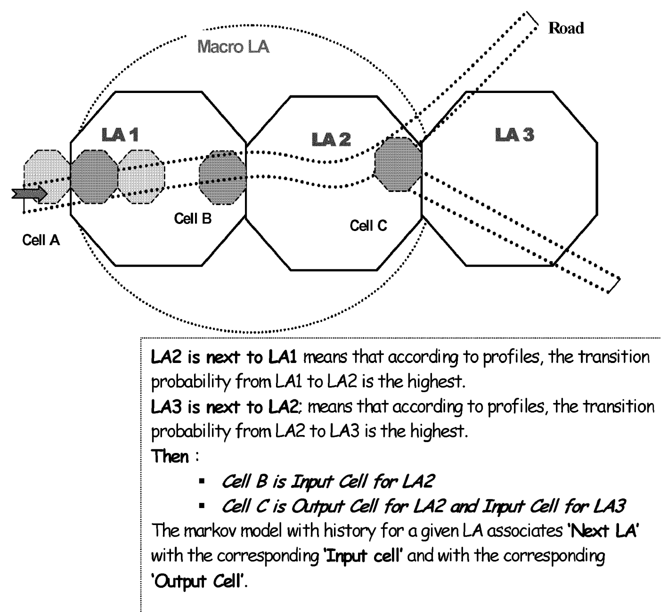


Fig. 10. Basic principle of the KYAMA method.

H. The Movement-Based Method (Without or With History)

In this approach [14], each mobile terminal counts the number of boundary crossings between cells incurred by its movements. Location update is performed whenever this number exceeds a predefined movement threshold ($D = 3$ for example).

For paging, in the original method, blanket polling is performed over the area determined by the movement threshold.

In the version with "history" (our extension of the original method), however, a type-1 profile is recorded during the user movement. Thus, a sequential paging in two steps, similar to the one in GSM+Profiles, is possible.

I. The Distance-Based Method (Without or With History)

In this scheme [14], each mobile terminal tracks the distance it has moved (in number of cells) since the last LU and transmits a location update whenever the distance exceeds a certain threshold ($D = 3$ cells, for example).

For paging, in the original method, blanket polling is performed over the area determined by the movement threshold.

In the version with "history" (our extension of the original method), a type-1 profile is recorded during user movement. The purpose of this extension is to reduce the paging effort. Thus, a sequential paging in two steps, similar to the one in GSM+Profiles, is possible.

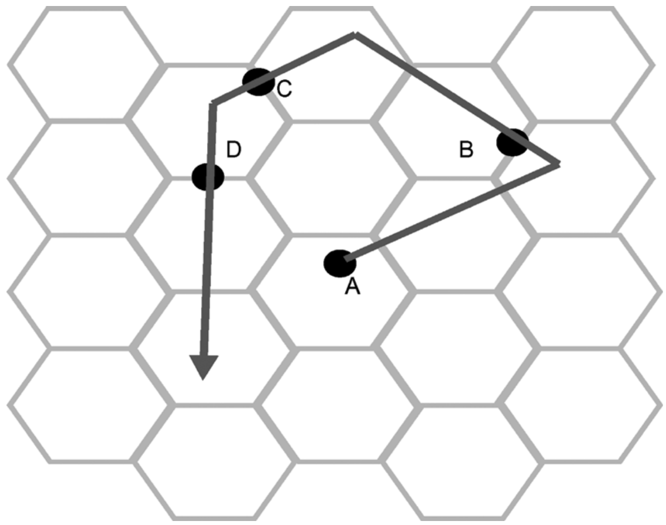


Fig. 11. Direction-based location update: example of a moving mobile terminal and positions of movement change and location update.

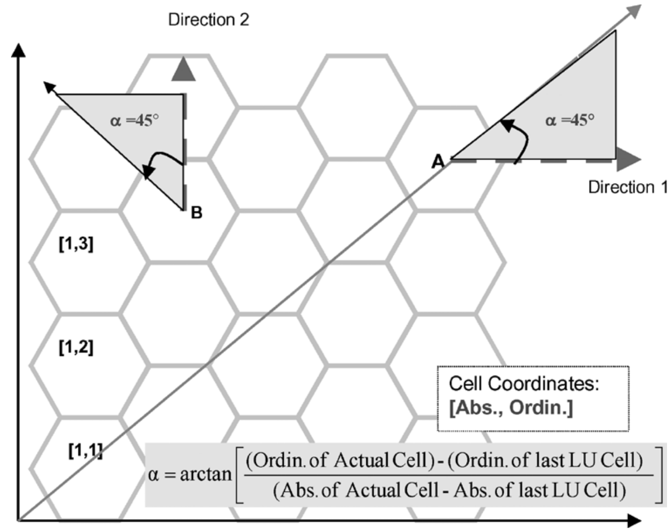


Fig. 12. Direction-based sector location update scheme: illustration of how the sector angle is calculated.

J. The Direction-Based Scheme (Without or With History)

In this scheme (see Fig. 11) [16], a mobile unit performs a location update whenever its moving direction changes. In Fig. 11, for example, the mobile terminal (MT) is moving and its moving direction changes at points B, C, and D. After passing each direction-changing point, MT performs location update and informs the network of its new moving direction. So doing, the network can always keep track of the MT's moving direction. When a call is destined to MT, the system only pages the cells (or LAs) on the line of the MT's moving direction; this is called "line-paging."

For paging, in the original method, blanket polling is performed (i.e., blanket line-paging) over the area determined by the movement direction.

In the version with "history" (our extension of the original method), a type-1 profile is recorded during the user movement. The purpose of this extension is to reduce the paging effort. Thus, a "line"-sequential-paging in two steps is possible.

K. The Direction-Based Sector Method (With or Without History)

This scheme is a modification of the direction-based approach in that, instead of a single direction, a sector of a given angle is rather used (45°; for example, see Fig. 12). The purpose of this extension is to reduce the sensitivity of the LU approach to frequent direction changes for users having a fuzzy movement pattern. In our implementation, we do consider a sector angle of 45°.

V. Simulation Context and Parameter Settings

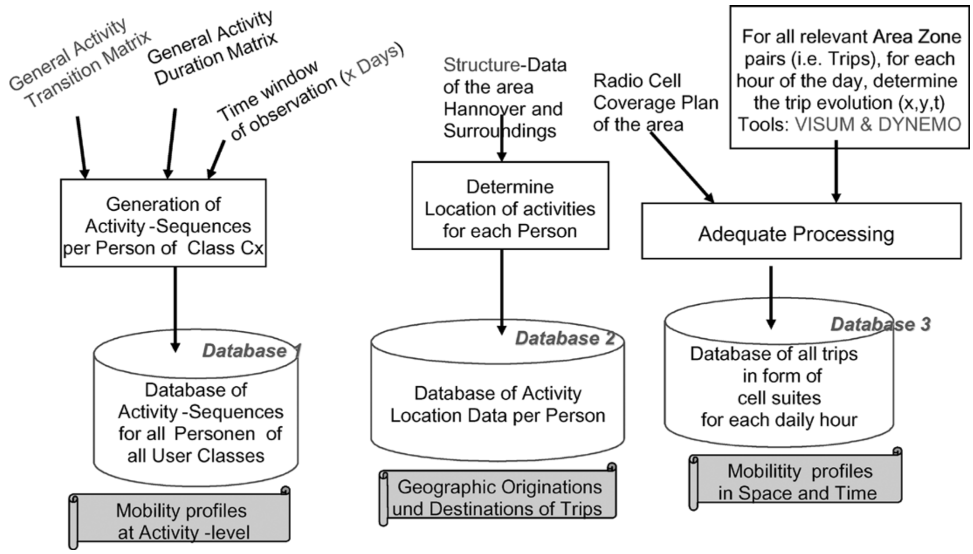
We have implemented the model and schemes described in Sections III and IV. This implementation has been used, amongst other things, in the frame of a series of performance analysis and simulation studies: location management, radio resource management/channel allocation schemes, and position-based networking concepts and services.

In order to a) validate and fine-tune the mobility model and b) perform extensive performance analysis, a simulation environment has been developed and implemented. In the following, we provide the parameter settings that have been used. The overall scheme of the process is summarized in Fig. 13(a). In fact, the mobility data to be used in the frame of the performance analysis of the different location management schemes are expected to be as realistic as possible concerning both the time and the geographic dimension. Three databases are needed, in which data related to a representative population of all user types are stored: 1000 samples for each type. The first database contains the user activity sequences, whereby the second one does cover user activity locations. Its data have been generated by C++ coded programs implementing the logic extracted from published German survey data, except for the random user type for which the sequence logic is random. The third database relates to the possible trips between the different activities locations. It has been generated in two steps: a) in a first phase, the possible trips are generated for all elements of the samples set, using specialized commercial tools (VISUM and VISUM; see [29]) and b) in a second phase, using a C++ coded program, the position sequences related to each of the trips are mapped to a radio cell coverage plan. The following sections do provide additional insight into the details.

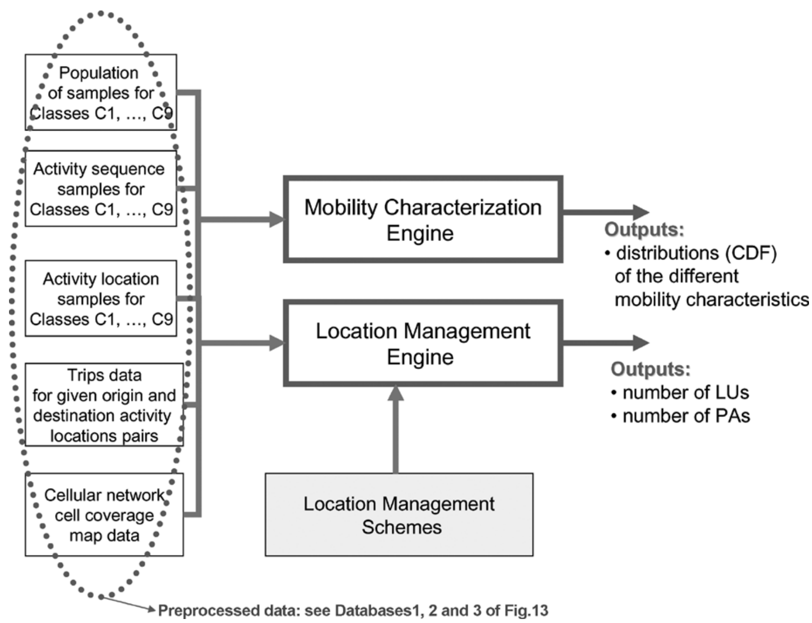
Before going into details, a big picture of the simulation process is appropriate at this position [see Fig. 13(b)]. The basic data concerning classified user population, activity sequences, activity locations, trip data, and cellular network cell coverage maps are preprocessed and prerecorded [using partly commercial tools and partly C++ coded programs; see Fig. 13(a)]. In run time either the mobility characterization engine or the location management engine will emulate individual trips on the basis of the prerecorded data for each sample individual of the population and compute corresponding output values [see Fig. 13(b)].

A. Geographical Context and Cell Structure and Sizes

The data related to the modeling of the space dimension have been obtained from the administration of the town Hannover,



(a)



(b)

Fig. 13. (a) Generation of reference databases in the frame of the mobility modeling. (b) Functional structure of the simulation process (The basic data are preprocessed; the engines emulate the trips in runtime and compute the outputs).

Germany[9] and from the Institute for Traffic Planning of the University of Hannover.[10] These data have been imported and adapted for the tools VISEM and VISUM [29]. Fig. 5 shows a snapshot of the space dimension modeling.

Consider velocity distribution on roads in an urban context. As we simulate the area Hannover, which is a urban context, we use the velocity profiles recommended in the related manuals of DYNEMO as input for the tool VISUM. The maximum speed in the urban context, in Germany, lies in the range of 60–70 km/h. In order to seize the influence of the geographical area size or scale (see parameter R in Fig. 6) on both mobility characteristics and key performance metrics from the point of view of the cellular networks, we have considered two scale scenarios for the geographical area: a small-scale (R around 7 km—this corresponds to the center of the town Hannover) and a large-scale area (R around 17 km diameter—this corresponds to the whole town of Hannover and part of the region around it, which is called, in German, Großraum-Hannover).[11]

Concerning the (radio) cell structure, regularly arranged square cells are hypothetically laid over the geographical area where the user movements take place. The (radio) cell size (diameter) considered in the simulations has been varied from 100 to 7000 m.

[9]Kommunal Verband Großraum Hannover, Niedersachsen, Germany.

[10]Institut für Verkehrswirtschaft, Straßenwesen und Städtebau, University of Hannover, Germany.

[11]The people living in Großraum-Hannover are very linked to Hannover concerning their mobility pattern (as some of their activities are located in the center of Hannover).

TABLE III
APPROXIMATED PROBABILITY FOR USERS OF CLASSES C6–C7 (ACCORDING TO FIG. 7) TO BE ATTRACTED TO SPATIAL ZONES, MAPs (SOURCE: GERMAN SURVEY DATA KONTIV'89 AND [31])

Time Zone			Attraction probability (%)				
#	**Hours**	**Description**	**R**	**W/S**	**B**	**S**	**L**
1	06-08	Morning rush hours	90	00	00	00	10
2	08-12	Morning working hours	20	30	40	10	00
3	12-14	Day free hours	40	10	40	10	00
4	14-18	Afternoon working hours	40	10	10	10	00
5	18-20	Night rush hours	60	00	00	10	20
6	20-24	Night free hours	80	00	00	00	20
7	24-06	Sleep hours	100	00	00	00	00

B. Timing of User Movements

In order to simulate the individual movements across the transportation network, including the use of vehicles such as cars or buses, on a second-by-second basis, the following data have been used.

1) Activity Locations
   a) For users of classes C1–C7 (according to Fig. 7), use either a distribution of user activity locations over the town according to measured data of the MAPs (in our case we have used data obtained from the administration of the town Hannover-Germany) or distribute them randomly on the city surface.
   b) For users of classes C8 and C9 (of Fig. 7), a random distribution over the city surface has been used.
2) Activity Sequencing and Timing
   a) For users of classes C1–C7, classical transition and duration matrixes calculated from German survey data of the series KONTIV'89 (part of which has been summarized in [31]) have been used (the corresponding data, however rounded, are condensed in Tables III and IV); alternatively, a "random" generalized transition matrix (i.e., the next activity is randomly selected among the remaining activities) has been considered.
   b) For users of classes C8 and C9, both the generalized transition and the generalized duration matrixes are random.

For the random activity duration matrix, the possible values presented in Fig. 14 are considered. For all user classes, but especially for classes C8 and C9, we consider a series of activity duration scenarios (ADSs) in order to conduct sensitivity studies—see Fig. 14.

Two groups can be distinguished.

- Group A: For the scenarios ADS1,..., ADS6, the activity duration is selected randomly in an interval between 0 min and a maximum that is set for each ADS.
- Group B: For the scenarios ADS7 up to ADS12, the activity duration is set constant, i.e., a constant value for each group.

From the point of view of the cellular network performance metrics on a long-term perspective, the consideration of both group scenarios A and B is helpful while determining whether it is the "effective distribution" of the activity durations or rather the "average" activity duration that is of interest. Average activity time budgets for the different user classes (of the socioeconomic classification) are presented in Table V.

User groups C8 and C9 are very interesting in that they provide the analyst with the flexibility of freely setting values for both the general activity duration matrix and the general activity transition matrix. Hence, it is possible to emulate the whole range of mobility patterns. The attraction probability of an activity location is related to the activity transition matrix and does express the probability that, for a given time interval, this location will be the next to be visited by the user (see Tables III and IV).

C. Call Arrival Profiles

Concerning the user call behavior, we still use the Poisson distribution of the call arrival process. First, we fix the number of calls per day and we distribute them over the day, while copying the daily call traffic profiles as observed in published data from telephone network carriers. During the day, the call intensity is

TABLE IV
APPROXIMATED PROBABILITY FOR USERS OF CLASSES C1–C5 TO BE ATTRACTED TO SPATIAL ZONES, MAPs (SOURCE: GERMAN SURVEY DATA KONTIV'89 AND [31])

Time zone		Attraction probability (%)					
#	Hours	Description	R	W/S	B	S	L
1	06-08	Morning rush hours	100	00	00	00	00
2	08-12	Morning working hours	10	80	10	00	00
3	12-14	Day free hours	40	10	30	20	00
4	14-18	Afternoon working hours	10	80	05	05	00
5	18-20	Night rush hours	60	00	00	20	20
6	20-24	Night free hours	80	00	00	00	20
7	24-06	Sleep hours	100	00	00	00	00

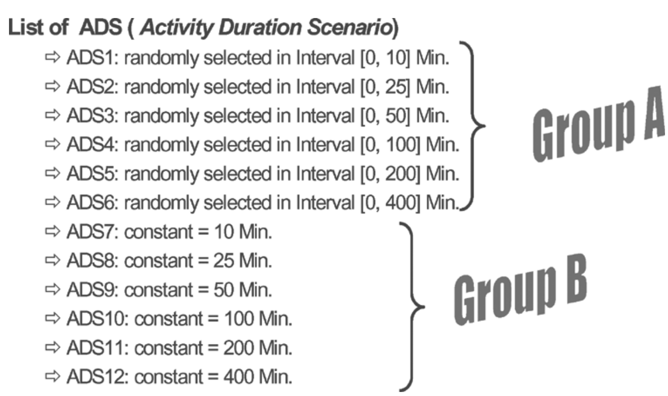


Fig. 14. Design of activity duration scenarios for user groups C8 and C9 (see classification of Fig. 7).

TABLE V
AVERAGE TIME BUDGET (IN HOURS FOR EACH ACTIVITY VISITED OF USERS OF CLASSES C1–C7. FOR THE LEGEND, REFER TO TABLES III AND IV [31]

User Class	Average time spent in activity [hours]				
	R	W/S	B	S	L
C1	14:31	6:09	0:57	0:20	1:21
C2	14:50	6:38	0:39	0:14	1:12
C3	16:46	5:04	0:36	0:30	1:0
C4	16:21	2:43	1:54	0:24	1:29
C5	17:01	3:46	1:06	0:13	1:07
C6	20:05	0:39	1:03	1:03	1:02
C7	21:14	0:20	0:41	0:54	0:47

generally much higher than in the night, for example. Further, in order to seize the influence of the way calls are distributed over the day, we also consider a uniform distribution. Therefore, we use two call distribution profiles: a nearly realistic one and a uniform one (see Fig. 15).

Therefore, the call profile is characterized by the following elements.

- The call distribution profile over the day: we define two distributions (a) and (b)—see Fig. 15.
- The average call holding time.
- The average call intensity expressed in "number of calls per day."

For the different experiments that have been necessary for the diverse performance analysis (see Fig. 15), six call holding time (CHT) scenarios and eight call intensity scenarios (CIS) have been considered.

VI. PERFORMANCE ANALYSIS

In a first step, we focus on the mobility characterization, which results in the simplification of the mobility model designed. Then a systematic performance analysis follows. It is based on two metrics and a benchmark mobility scenario, which are all defined.

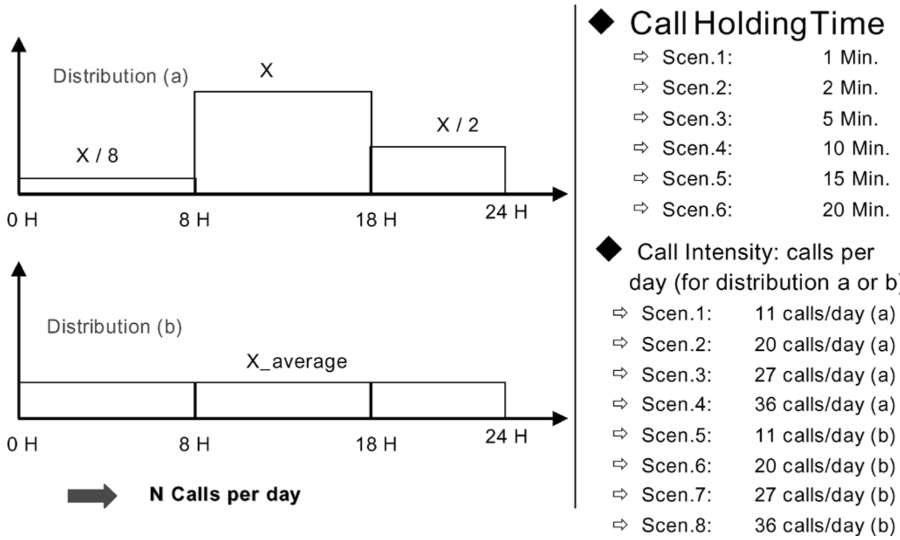


Fig. 15. Call profile structure, different scenarios—averages considered for call holding time and for call intensity.

A. Mobility Characterization and Simplification of the Mobility Model

A series of systematic and extensive experiments has been conducted using the simulation environment developed. One first task has been to perform sensitivity studies with the goal of examining which parameters of the mobility model do have a significant (or relevant) influence on the so-called mobility characteristics (of a cellular network). From the point of view of a cellular network, mobility characteristics are related to either the rate or the distribution of following metrics: user cell dwell time [the cell dwell time is the period time a user (mobile station) enters a certain cell to the time it moves into another cell); cell crossover (or handover) rate (per time unit per hour, for example) or per call]; and user velocity probability density function.

A simplification of the mobility model is possible if the following questions are answered.

- Do the activity duration profiles have any influence on the mobility characteristics and which one (i.e., does the form of the general duration matrix have any influence on the mobility characteristics)?
- Does the activity transition profiles (this refers to the general activity transition matrix) have any influence on the mobility characteristics?

Does the deterministic or stochastic nature of the activity locations have any significant influence on the mobility characteristics?

- Does the regularity in mobility profiles (some determinism) have any influence on the mobility characteristics?
- Does the scale of the geographic area considered (small or large scale) have any significant influence on the overall form (or pattern or signature) of the mobility characteristics?
- What is the impact of the form of the call distribution profile (over the day)?
- Is the socioeconomical part of the classification of Fig. 7 of any relevance from the point of view of the mobility characteristics, and subsequently of all performance-related metrics of cellular networks? If this is not the case, which new classification criteria can be suggested?

A huge number of results have been generated to answer these questions. However, because of the large number of scenarios to be considered in the sensitivity studies, we cannot present all the results (space constraint). We can just provide a sample of them while giving the answers found for the series of questions formulated above.

Consider cell dwell time. When using logarithmic axes, for a given cell size and for a given geographic area (i.e., given radius R, as indicated in Fig. 6) the general trend/form of the cell dwell time cumulative probability distribution function (cdf) is almost the same for all users and is therefore independent of the specific form of both activity duration and transition matrixes. This is illustrated in Fig. 16. Even a change of R (that is, the radius of the geographic area of concern) from 7 to 17 km does not result in a significant change in the overall form of the curves/diagrams (compare Figs. 16 and 17). In other words, all users within a given geographic area where movement takes place do have nearly the same cell dwell time cdf, independently of their specific mobility pattern (within the variation frame determined by our simulation model).

Consider the impact of both the form of both activity duration and transition matrixes on the mobility characteristics (especially, exemplary, on cell crossover rate). The results show (see sample results in Figs. 18 and 19 for illustration) that only the activity duration matrix has a relevant and significant impact and that, hereby, not the specific distribution is of relevance, but rather the average activity duration value, from a long-term perspective. To make this clear, let us take an example based on Figs. 14, 18, and 19. First, notice that (see Fig. 14) “ADS3 and ADS8” have (or correspond to) the same “average activity duration.” We do have the same situation for ADS4 and ADS9; ADS5 and ADS10; and ADS6 and ADS11. And Figs. 18 and 19 clearly show that the cell crossover cdf is the same for ADS scenarios that possess the same “average activity duration.” The same has been observed for the other mobility characteristics.

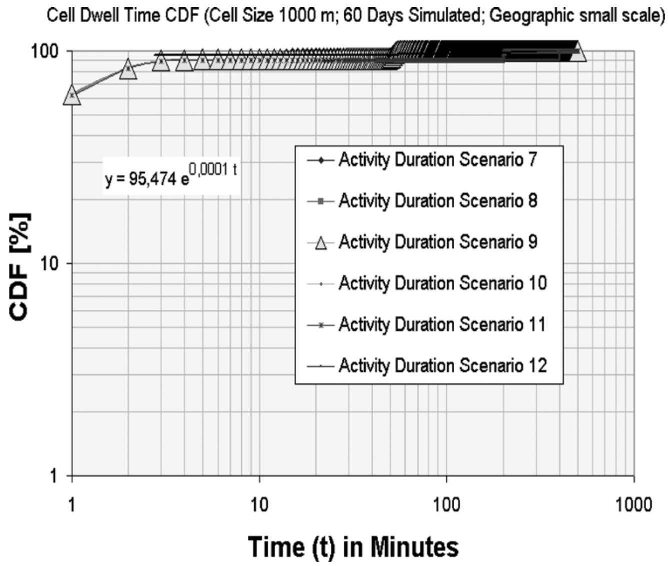


Fig. 16. Cell dwell time cdf for activity duration scenarios 7 to 12. Geographic small-scale scenario and logarithmic axes.

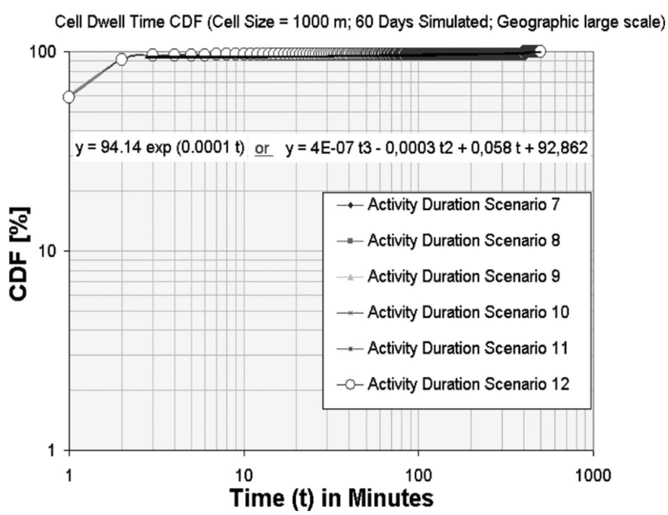


Fig. 17. For the geographic large-scale scenario, cell dwell time cdf for activity duration scenarios 7 to 12. Logarithmic axes.

Further, we could make the observation that neither the deterministic nor the stochastic nature of activity locations has a significant impact on the mobility characteristics.

Consider call profiles. Concerning the distribution of calls over the day, from a long-term perspective, only the average call per day has been found to be relevant and not the real distribution profile of these calls over the day (results not presented due to space shortage).

Consider the socioeconomic classification of users according to Fig. 7, i.e., user groups C1, . . . , C7. A thorough analysis of the diverse results obtained for the various scenarios thinkable (many results are not presented due to space shortage) leads to the conclusion that this classification is of no significant relevance from the point of view of the location management schemes. The only relevant user classification criteria appears to be rather, as has been discussed below, the (long-term) average activity duration, independently of the form of the activity transition matrix. In fact, the essential difference between the different socioeconomic groups from the mobility pattern point of view resides mainly in the transition matrixes (see Tables III-V).

A thorough review of the study leads to a simplified user classification—this corresponds to a simplification of both user classification and mobility model. The simplified model states that the relevant user classification criteria are the following.

1) Mobility model—it is mainly activity based, however, with the following elements.
   a) The "average activity duration" is the most relevant criteria; the activity duration can be taken random, however, with a predetermined average value, from a long-term perspective.
   b) Activity locations can be randomly distributed over the surface of the geographic area of interest (of radius R—see Fig. 6).
   c) The activity transition matrix can be taken random. This constitutes a "benchmark" from the point of view of the location management methods. The more randomness the user movement possesses, the more difficult it should be to predict user location.
2) The radius of the geographic area of interest R.
3) The average call intensity (CIS), which is the average number of calls per day that can be randomly distributed over the day.
4) The average CHT.
5) The average cell size (cell radius or diameter).

This model simplifies the elements presented in Fig. 6 in that even a parameterization is possible. We can just consider a single user type (corresponding to class C9 of Fig. 7) that has, however, the elements described in this simplified model. From the activity transition point of view, it corresponds to the random walk. The activity duration is random, however, with a predetermined average value, from a long-term perspective (called ADS, see Fig. 14). Further, the movement between two activity locations is determined by the simulation or better the mimic of a realistic user movement (using appropriate tools—VISUM/DYNEMO, for example) inside the town (in this case, the town of Hannover) between two locations A and B.

Finally, the strong dependence of the mobility characteristics and of other performance results on the average activity duration (ADS) highlights the superiority of our mobility model (see Fig. 6), while compared to previous traditional statistical models that do not differentiate users from the point of view of their related average activity duration.

B. *Performance Analysis of the Diverse Location Management Methods*

The task of a location management method is per se to predict user location. To predict the location of a deterministic user is rather a trivial task. Therefore, we need a "benchmark scenario" that offers a really random or fuzzy user (in the frame of the "simplified" mobility model developed) and that will be used for a systematic and extensive performance analysis of the diverse location management methods. This "benchmark scenario" is based on the simplified model obtained in the last subsection.

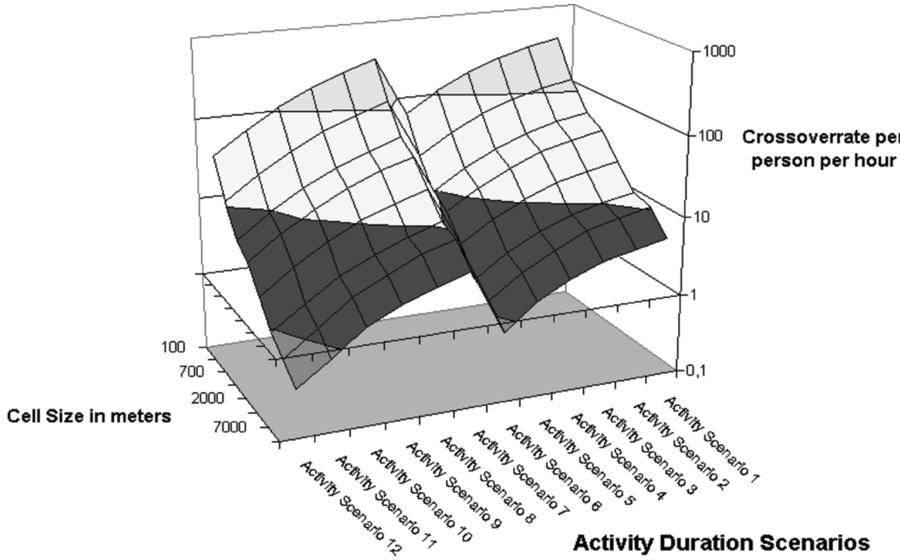


Fig. 18. Cell crossover rate per hour (average over all terminals, communicating or not). Geographical large-scale scenario. CIS2, i.e., 20 calls/day.

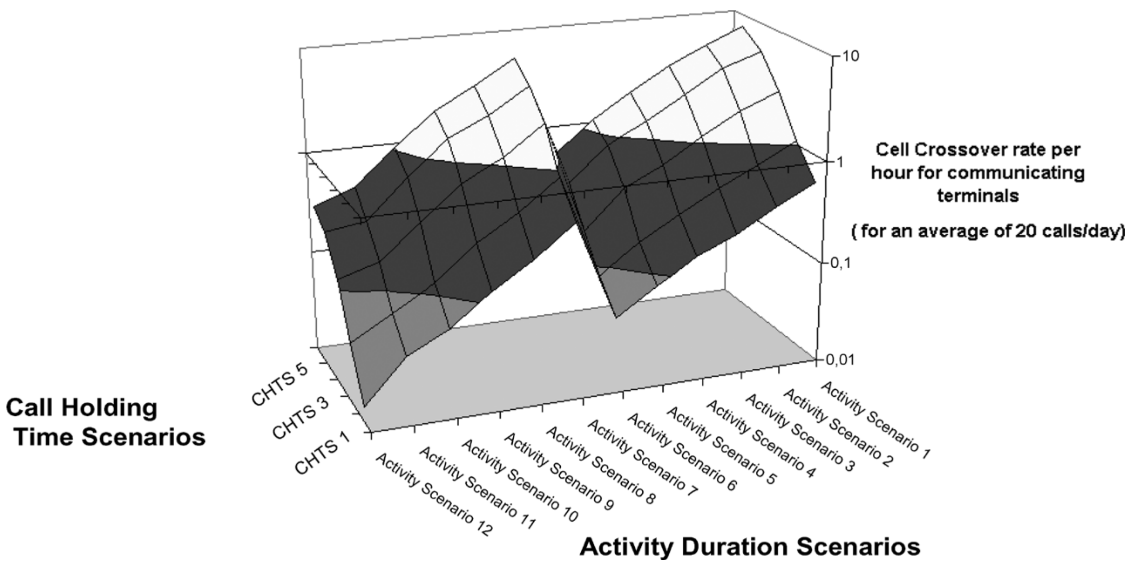


Fig. 19. Cell crossover rate per call for communicating terminals, independent of the call holding time. (Geographic small scale scenario. CIS2, i.e., 20 calls/day).

Furthermore, it has been seen that both the average activity duration and the average call profile are the key differentiating factors. And in order to cover the full range of values that can be taken by these two parameters, we define a parameter that we call "call-to-mobility ratio" (CTM) by the following formula:

$$\mathrm{CTM} = \text{Average Number of Calls per Dayover Average Activity Duration [Min]} * 100. \quad (1)$$

For the various performance analyses conducted, CTM has been varied from 0 to about 720 (this covers all the different mobility and call scenarios presented in Figs. 14 and 15).

Note that the lower CTM is (i.e., high ADS), the more deterministic the user is (respectively, the less fuzzy his location is). For location management, the challenge is to design methods that perform well (that is, low management cost) for both highly random and less random (nearly deterministic) users, i.e., the average performance over the possible range of CTM should be the best possible.

Consider location management performance. The performance of a location management method is determined by its cost. The average number of location updates (nLU) and that of cells paged (nPA) are the two key components of the location management cost (see also Fig. 1)

$$\mathrm{Cost} = \mathrm{nPA} + \mathrm{c} * \mathrm{nLU}. \quad (2)$$

Thus, the location management cost is a linear combination used to quantify the net effect of location updates and paging messages. A location update is more expensive than a paging, due to the need to establish a signaling channel, but the exact cost can be quantified in different ways (see Fig. 1) [24]. In this paper, the ratio c [see (2)] of the cost of a location update to the cost of a paging has been taken in the range between five and ten. These values roughly take into consideration the number and size of messages exchanged for each procedure (according to an

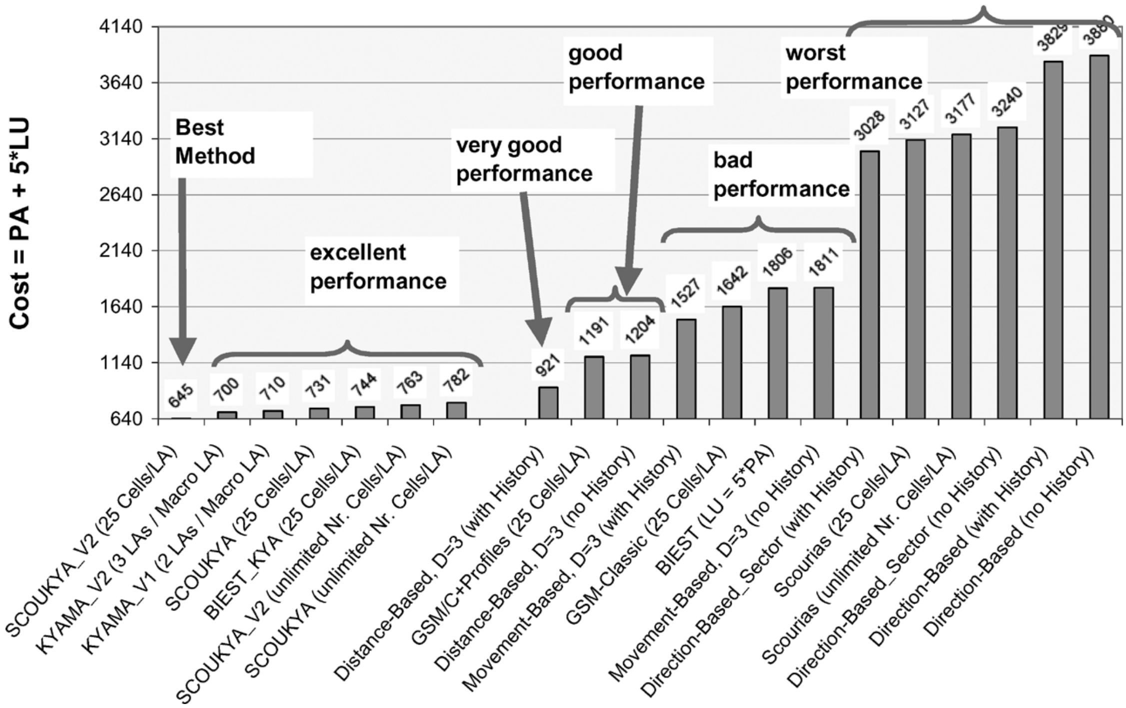


Fig. 20. Performance comparison of the different location management methods (average over CTM range 0 to 720)—average cost per day. (Simulation context: benchmark; geographical data of the town Hannover; average cell diameter 1000 m; user movements simulated over 60 days; D (in cells) is the threshold considered for the approaches distance-based and movement-based; location management cost = PA + 5^*LU.)

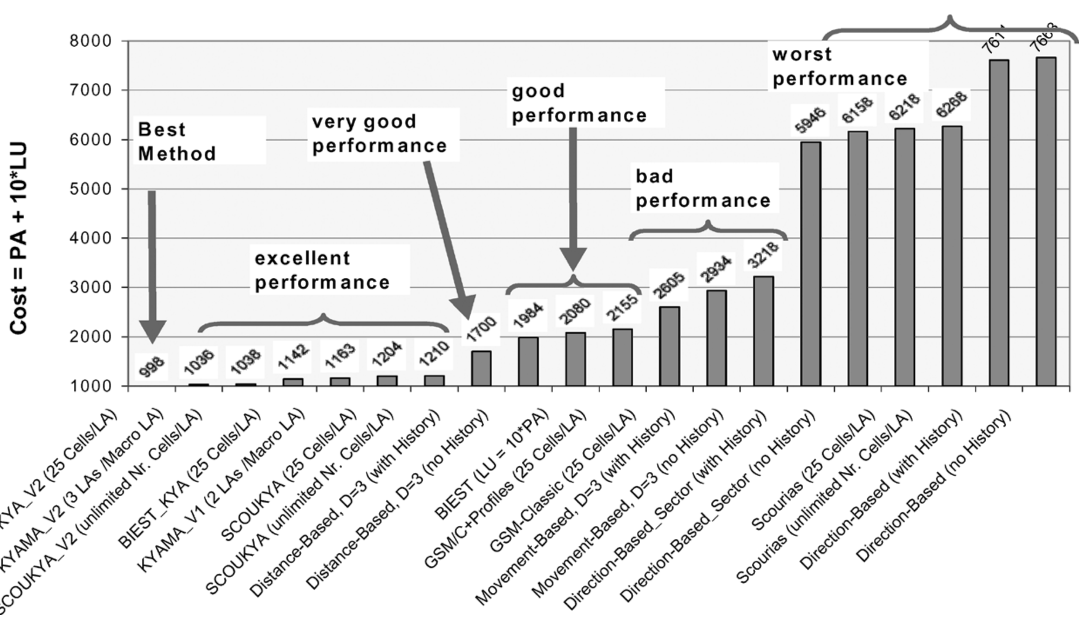


Fig. 21. Performance comparison of the different location management methods (average over CTM range 0 to 720)—average cost per day. (Simulation context: benchmark; geographical data of the town Hannover; average cell diameter 1000 m; user movements simulated over 60 days; D (in cells) is the threshold considered for the approaches distance-based and movement-based; location management cost = PA + 10^*LU.)

estimation done in [24]). And since the location update process involves the exchange of more data/messages, it is generally correspondingly more expensive.

Figs. 20–22 present a representative sample of the results obtained, which are basically self-explained. Note that we have simulated about three months of user movement. The results showed that a longer simulation does not bring any change. To perform a fair comparison of the different methods (for Figs. 20 and 21), the number of cells per location area (or equivalent; this refers to the nonlocation area based schemes where the selected thresholds D corresponds approximately to an area of similar size) has been taken approximately 25 for all, except for the cases, where this number is not limited (according to the concerned location management approaches). Except for the location management scheme BIEST, where the location management process does depend on the relation between LU and PA costs, the only difference between Figs. 20 and 21 for the other schemes lies in the way the total cost is computed.

TABLE VI
RANKING OF THE DIFFERENT LOCATION METHODS BASED ON THEIR RESPECTIVE PERFORMANCE, THE FORM OF PROFILE INFORMATION USED TO RECORD USER MOVEMENT HISTORY, AND THE LOCATION UPDATE SCHEME

RANKING	FORM OF PROFILE INFORMATION (ACCORDING TO TABLE I)	LOCATION UPDATE SCHEME(S)	LIST OF METHODS, SORTED AFTER RESPECTIVE PERFORMANCE
I (Best)	Type 3	LA-based	SCOUKYA_V2
II (Excellent)	Type 3 (extended)	LA-based + Timer-based	KYAMA
	Type 2	LA-based	SCOUKYA
	Type 3	LA-based	BIEST_KYA
III (Very Good)	Type 1	Distance-based	Distance-based (with History)
IV (Good)	Type 0	Distance-based	Distance-based (no History)
	Type 4	LA-based	BIEST
	Type 1	Static LU	GSM/C+Profiles
V (Bad or Acceptable)	Type 0	Static LU	GSM-Classic
	Type 1 or 0	Movement-based	Movement-based
VI (Worst or Unacceptable)	Type 1 or 0	Direction-based	Direction-Based_Sector
	Type 2	LA-based	Scourias
	Type 1 or 0	Direction-based	Direction-based

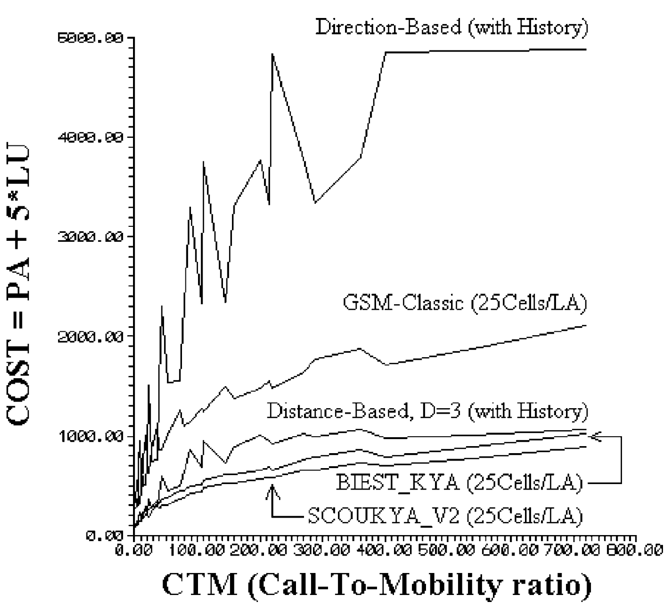


Fig. 22. Performance comparison of a sample of the different location management methods (for each of the performance groups indicated in Figs. 20 and 21 a representant has been selected)—cost per day versus CTM. (Simulation context: benchmark; geographical data of the town Hannover; average cell diameter 1000 m; user movements simulated over 60 days; D (in cells) is the threshold considered for the approaches distance-based and movement-based.)

A ranking of the methods according to their performance is presented in Table VI. This ranking also takes results into account that are not presented here due to space shortage.

An appreciation of the different performance results should take the GSM-Classic method as basic reference. A method that performs worse than this method is definitively a bad one. Furthermore, good methods should significantly perform better than the reference one. The minimal performance improvement should be in an order of magnitude of around 20% compared to that of the reference method.

As indicated in Table II, a location management method's essence is strongly dependent on the profile type necessary to record the necessary movement history information. A ranking of the different location methods that takes the form of the profile information used into consideration is presented in Table VI.

The major conclusion to be drawn from the results obtained is that location management methods designed around profile information of type 2 or 3 have the highest potential to offer the best performance. In other words, location management involving the idea of "profile-based direction" (this is expressed in the movement history information of types 2 and 3 of Table I) has the potential to realize the best performance compared to all other approaches.

Another important observation is that, from all approaches that do not involve movement history information, the distance-based approach performs at best, and can perform up to the ranking "Good." This is important information due to the fact that the recording of movement history information is a process that require smemory in both mobile terminals and the network.

The movement-based approach seems to perform at best just as well as the GSM-Classic approach. Its performance is also to be qualified bad. As far as the direction-based location management approach is concerned, it overall performs worse than the GSM-Classic. In principle, a direction-based approach has the best performance in cases where the movement follows a constant geographic direction (scenario like on a highway).

However, if the movement direction often changes, and this is the case in the "benchmark mobility scenario" considered in this paper, then the performance drastically decreases. Overall, the movement-based approach is marginally acceptable, whereby the direction-based one is unacceptable.

The observations further highlight two extreme situations: in a case of very deterministic movement by following a line of constant direction, the direction-based approach is the best; in a case of "benchmark mobility" pattern, the "profile-based direction" location management approach performs absolutely the best. Besides, one should notice that, from a long-term perspective, the profile-based direction location approach, which is more general, also includes the geographic direction-based one in the segments of time and space where it appears.

Therefore, and this is the key teaching from this paper, a location-area based location management method designed around a profile or history-based direction information offers the absolutely best performance in terms of location management cost, compared to all other alternative approaches (see Table VI).

VII. Conclusions and Outlook

The scope of this paper has been very broad. First, the key issues in location management have been formulated. Then, a classification of the different location management paradigms has been performed by means of the respectively followed schemes for location updates and for paging. The location update scheme appears to be the most significant differentiating factor. One first distinguishes two cases. In a first case, the location area is static and the same for all users. In a second case, however, the location area (or better, the condition for a location update) is set user-specific, depending on his own mobility pattern. The second case is more general and flexible and offers space for more design options. It also requires, in many cases, the availability of movement history information. Thus, one could distinguish here between a series of different user-specific location management schemes: location-area based, timer-based, movement-based, distance-based, and direction-based.

Concerning movement history information, its formatting is of great importance. In fact, the grade of detail and the specific form of available movement history information allow the design of diverse, respectively, adapted prediction methods for user location. We have defined five different formats for movement history information: from type 0 to type 4. Type 0 refers to "no movement history information" recorded or needed, while type 4 is the most complex one in that, beside pure movement history, it also contains call history information. From the point of view of pure movement history, type 3 is the best one, as it formulates or expresses the notion of a "profile or history-based direction" that can be extracted from recorded movement history data.

Concerning users, the traditional socioeconomic classification has been considered, besides a more general and stochastically parameterizeable one.

Furthermore, a mobility modeling framework aiming to be as realistic as possible has been developed. The final mobility model obtained is activity-based and comprises a list of key elements that have been defined. The settings of the different elements are, in principle, to be extracted from survey data or set randomly by the analyst.

The modeling of the geographic dimension has been done considering the major streets of the town and the simulation or emulation of user movement between two given points of the town, while taking all possible effects of aggregate traffic flows into account. This has been performed with the help of some specialized commercial tools.

To simplify the mobility model and to answer the question of whether the socioeconomic user classification is of any relevance here, we have conducted a first series of simulations, whereby the different cellular network mobility characteristics have been calculated. The observations made upon the results have led to a simplification of the mobility model, whereby the average activity duration appears to be the most significant key differentiating factor between users from the point of view of their mobility pattern. Therefore, the socioeconomic classification has been found to be of no relevance.

The simplified model has the following components:

a) average radius of the geographic area where user movement take place;
b) a given number (between four and eight) activity locations, which are randomly distributed over the geographic area surface;
c) the activity duration is a random variable with a given average (which is the key differentiating parameter between different user classes);
d) a random transition between the activity locations;
e) average number of calls per day;
f) the average cell size.

For a systematic comparative performance analysis of a representative sample of all location management paradigms, whereby some of the methods are already published and some have been extended or designed by us, we have defined two metrics: a cost function and a call-to-mobility ratio. In addition, we have designed a "benchmark simulation framework" under which a large number of simulations have been conducted in the framework of a systematic performance analysis. Very interesting and significant results have been obtained, on which basis a ranking of the different methods could be performed.

One of the most significant key teachings from this paper is that a location-area based location management method designed around a profile or history-based direction information offers the absolutely best performance in terms of location management cost (over the full range of the call-to-mobility ratio), compared to all other alternative approaches.

References

[1] S. Tabbane, "Location management methods for third-generation mobile systems," *IEEE Commun. Mag.*, vol. 35, pp. 72–84, 1997.
[2] V. W. S. Wong and V. C. M. Leung, "Location management for next-generation personal communications networks," *IEEE Commun. Mag.*, vol. 14, pp. 18–24, 2000.
[3] I. F. Akyildiz, "Mobility management in next-generation wireless systems," *Proc. IEEE*, 1999.
[4] Y. Bejerano and I. Cidon, "An efficient mobility strategy for personal communication systems," in *Proc. ACM/IEEE MobiCom'98*, Dallas, TX, 1998.
[5] J. Jannink, "Efficient and flexible location management techniques for wireless communication systems," *ACM/Baltzer J. Wireless Networks*, vol. 3, pp. 361–374, 1997.
[6] C. Rose and R. Yates, "Location uncertainty in mobile networks," *IEEE Commun. Mag.*, pp. 94–101, 1997.
[7] ——, "Minimizing the average cost of paging under delay constraints," *ACM/Baltzer J. Wireless Networks*, vol. 1, 1995.

[8] R. Jain, Y. B. Lin, and S. Mohan, "A caching strategy to reduce network impacts of PCS," *IEEE J. Sel. Areas Commun.*, vol. 12, no. 8, pp. 1434–1444, 1994.

[9] K. Ratnam, I. Matta, and S. Rangarajan, "Analysis of caching-based location management in personal networks," in *IEEE 7th Int. Conf. Network Protocols*, 1999.

[10] D. Hong and S. S. Rappaport, "Traffic model and performance analysis for cellular mobile radio telephone systems with prioritized and nonprioritized handoff procedures," *IEEE Trans. Veh. Technol.*, vol. VT-35, pp. 77–92, 1986.

[11] I. F. Akyildiz and J. S. M. Ho, "Movement-based location update and selective paging for PCS networks," *IEEE/ACM Trans. Netw.*, vol. 4, 1996.

[12] S. K. Sen, A. Bhattacharya, and S. K. Das, "A selective location update strategy for PCS users," *ACM/Baltzer J. Wireless Networks*, vol. 5, pp. 313–326, 1999.

[13] G. P. Pollini, "A profile-based location strategy and its performance," *IEEE J. Sel. Areas Commun.*, vol. 15, pp. 1415–1424, 1997.

[14] A. Bar-Noy, I. Kessler, and M. Sidi, "Mobile users: to update or not to update?," *ACM/Baltzer J. Wireless Networks*, vol. 1, pp. 175–195, 1995.

[15] C. Rose, "Minimizing the average cost of paging and registration: a timer-based method," *ACM/Baltzer J. Wireless Networks*, vol. 2, pp. 629–638, 1996.

[16] H.-W. Hwang, M.-F. Chang, and C.-C. Tseng, "A direction-based location update scheme with a line-paging strategy for PCS networks," *IEEE Commun. Lett.*, vol. 4, pp. 149–151, 2000.

[17] B. Liang and Z. Haas, "Predictive distance-based mobility management for PCS networks," in *Proc. IEEE INFOCOM'99*, New York, 1999.

[18] C. Rose, "State-based paging/registration: a greedy technique," *IEEE Trans. Veh. Technol.*, vol. 48, pp. 166–173, 1999.

[19] U. Madhow, M. Honig, and K. Steiglitz, "Optimization of wireless resources for personal communications mobility tracking," *IEEE/ACM Trans. Netw.*, vol. 3, pp. 698–707, 1995.

[20] J. Scourias and T. Krunz, "An activity-based mobility model and location management simulation," in *Proc. PIMRC'97*, Helsinki, Finland, 1997.

[21] A. Bhattacharya and S. K. Das, "LeZi-update: an information-theoric approach to track mobile users in PCS networks," in *Proc. ACM/IEEE MobiCom'99*, Seattle, WA, 1999.

[22] D. Lam, D. C. Cox, and J. Widom, "Teletraffic modeling for personal communications services," *IEEE Commun. Mag.*, vol. 35, no. 2, pp. 79–87, Feb., 1997.

[23] J. G. Markoulidakis, G. L. Lybeopoulos, D. F. Tsirkas, and E. D. Sykas, "Mobility modeling in third-generation mobile telecommunication systems," *IEEE Personal Commun. Mag.*, pp. 41–56, 1997.

[24] J. Biesterfeld, *Adaptive Verfahren zur Aufenthaltsverwaltung in Telekommunikations netzen, in Series Hannoversche Beiträge zur Nachrichtentechnik, Band 2.2 (book in German)*: SHAKER Verlag, 2001.

[25] P. Orlik and S. S. Rappaport, "Traffic performance and mobility modeling of cellular communications with mixed platforms and highly variable mobility," *Proc. IEEE*, vol. 86, no. 7, pp. 1464–1479, Jul., 1998.

[26] F. Khan and D. Zeghlache, "Effect of cell residence time distribution on the performance of cellular mobile networks," in *Proc. IEEE VTC*, 1997.

[27] K. Kyamakya and K. Jobmann, "Contribution to user mobility modeling for cellular networks," in *Proc. 6th Int. Conf. CDMA Networks (CIC'2001)*, Seoul, Korea, 2001.

[28] T. Camp, J. Boleng, and V. Davies, "A survey of mobility for ad hoc network research," *Wireless Commun. Mobile Comput.*, pp. 483–502, 2002.

[29] W. Y. Oi and P. W. Shuldiner, *An Analysis of Urban Travel Demands*. Evanston, IL: Northwestern Univ. Press, 1962.

[30] S. Schönfelder, "Some notes on space, location and travel behavior," in *Proc. 1st. Swiss Transport Research Conf.*, Monte Verita, Ascona, 2001.

[31] R. Schmiedel, Determination of person groups with similar mobility pattern for the traffic planning (Title in German: Bestimmung Verhaltensähnlicher Personenkreise für die Verkehrsplanung), in Department of Civil Engineering, Univ. Karlsruhe, Karlsruhe, 1984.

[32] J. Scourias and T. Kunz, "A dynamic individualized location management algorithm," in *Proc. PIMRC'97*, Helsinki, Finland, 1997.

[33] K. Kyamakya and K. Jobmann, "Classification of mobility management methods, realistic simulation framework and relative performance analysis," in *Proc. MWCN'2001—3rd IEEE Int. Conf. Mobile and Wireless Communication Networks*, 2001.

Kyandoghere Kyamakya received the doctorate degree in electrical engineering from FernUniversität-Hagen University, Germany, in 1999.

Since 2000, he has been with the Faculty of Electrical and Computer Engineering, University of Hannover, Germany, as a Postdoctoral Researcher and Lecturer (Junior Professor). His areas of research and teaching include, amongst others, network reliability and survivability, sensor networks, location and mobility management, positioning, and location-based services.

Klaus Jobmann received the diploma and the doctorate degree in communications engineering from the Technical University of Hannover, Germany, in 1974 and 1979, respectively.

He is a Professor in the Faculty of Electrical and Computer Engineering, University of Hannover, Germany. From 1981 to 1992, he was with Royal Philips Electronics in various positions, including Head of marketing network and system planning, Head of international sales, and Head of "Competence Centre PTS (Public Telecom Systems) Spain." In 1994 he joined the Department of Electrical Engineering at the University of Applied Sciences, Magdeburg, Germany, as a Professor before accepting his current position in 1995. He is a Consultant to various industry, government, and public institutions. His research and teaching interests include management/QoS of public telecommunication networks, ad hoc networks, resource control in mobile networks, and teleeducation.

Prof. Jobmann is a member of the Association for Electrical, Electronic and Information Technologies, the Information Technology Society, and Engineers for Communication (Ifkom).