Location Management Proposals:

Survey and Evaluation

by

Atif Akram Siddiqi, B.Sc. Eng.

University of Engineering & Technology, Lahore, Pakistan

A thesis submitted to

the Faculty of Graduate Studies and Research

in the partial fulfillment of

the requirement for degree of

M.Sc. in Information and Systems Science

Department of Systems and Computer Engineering

Carleton University

1125 Colonel By Drive

Ottawa, Ontario, K1S 5B6

Canada.

April 1, 1999

The undersigned recommend to the Faculty of Graduate Studies and Research

acceptance of the thesis

Location Management Proposals:

Survey and Evaluation

submitted by Atif Akram Siddiqi, B.Sc. Eng.

in partial fulfillment of the requirements for

the degree of M.Sc. in Information and Systems Science

Professor Thomas Kunz, Thesis Supervisor

Chair, Department of Systems and Computer Engineering

Carleton University

April 1, 1999

Abstract

The challenge of supporting rapidly growing numbers of mobile users, while constrained by limited radio spectrum is being faced by cellular network operators worldwide. The standard technique for cellular networks is to decrease radio cell size, thereby maintaining a supportable subscriber density. However, smaller size result in increased signaling for location management procedures, which reduces the bandwidth available for user traffic.

Several location management schemes have been proposed to improve the performance of such networks, but a fair assessment and comparison of their performance is difficult without a realistic mobility model, since the performance of location management schemes depends considerably on subscriber mobility patterns. Some of the methods used in the literature are reviewed and commented upon. The strategies are then classified in several categories on the basis of their functionality. Finally, some proposals are selected that concentrate on the location updating and paging overhead.

To analyze the performance of the selected proposals, two mobility models were used in this thesis. The first was a realistic, measurement-based, place and time-dependent individual mobility model to simulate the mobility behavior of actual people. The second was a random mobility model based on simplified assumption, such as a uniform distribution of the direction of travel for each mobile subscriber, as being used in a number of proposed schemes.

From the results it is clear that the mobility model has a significant impact and the results described in various schemes using a random mobility model may not reflect the relative performance when deploying schemes in actual systems.

Acknowledgements

I would like to express my sincere thanks to my thesis supervisor, Dr. Thomas Kunz, for his invaluable guidance and helpful suggestions during the course of this thesis. This thesis would not have been possible but for his enthusiasm for research and his remarkable reviewing speed.

I would like to thank all the students of the Broadband Network Lab for putting up the nice working atmosphere and the staff of the Department of Systems and Computer Engineering for their help in the development of this thesis.

I would like to thank my parents and brothers for their prayers and support, encouragement and guidance throughout my academic pursuits.

Last, but not the least, I would like to thank my wife who provided me with a continuous source of inspiration and for her patience and strong support without which the research would not have been possible.

Table of Contents

List of Tables

List of Figures

List of Acronyms

PCS	Personal Communication Services
MT	Mobile Terminal
HLR	Home Location Register
VLR	Visitor Location Register
BS	Base Station
BTS	Base Transceiver Station
RA/LA	Registration Area in IS-41 / Location Area in GSM
PA	Paging Area
MSC	Mobile Switching Center
LU	Location Update
PSTN	Public Switching Telephone Network
PFSHLR	Pointer Forwarding with Single HLR
PFDHLR	Pointer Forwarding with Distributed HLR
OSPFDHLR	One-Step Pointer Forwarding with Distributed HLR
GPC	Global Personal Communication
MAP	Mobile Application Part
TLA	Two-Location Algorithm
LMTs	Location Management Techniques
LCMR	Local Call to Mobility Ratio
HOPPER	Hierarchical Online Parametric ProfilE Replication
HIPER	Hierarchical ProfilE Replication
DR	Directory Registers
STP	Signal Transfer Point
RDB	Regional DataBase
HDB	Home DataBase
TGMTS	Third Generation Mobile Telecommunication Systems
PHS	Personal Handyphone System
PACS	Personal Access Communication System

Chapter 1

Introduction

1.1 General Concepts in Location Management

In the past few years, market demand has grown rapidly for Personal Communication Services (PCSs) such as digital cellular mobile radio systems, the Personal Access Communication System (PACs), and the Japanese Personal Handyphone System (PHS). This can be attributed to several factors, including decreasing prices, improved radio coverage, and compact lightweight terminals. These second-generation wireless communication systems are designed for particular applications and environments. Customer demand for these systems has increased dramatically every year. In Japan, for example, the subscriber growth rates from September 1995 to September 1996 have risen by 129% for digital cellular mobile radio systems (PDC) and 2889% for PHS [23]. In general, as the number of subscribers increases given a fixed radio spectrum allocation, the size of radio coverage cells must decrease, in order to accommodate the higher subscriber densities. This creates many challenges, especially concerning location management. PCS, which are higher-capacity and lower-power cellular networks with smaller cells, face similar location management issues.

Compared to second-generation systems and apart from the increased traffic demands, the employment of location management and handoff procedures in a microcellular environment, in conjunction with the huge number of Location Updates (LUs), will generate a considerable “mobility-related signaling” load. The increase of mobility-related signaling apart from the radio link will have a major impact on the number of database transactions, thus causing the database to be a possible bottleneck at the fixed network side. Consequently, given the scarcity of radio resources, methods for signaling load reduction are emerging for Third Generation Mobile Telecommunication Systems (TGMTS). It is obvious that optimization techniques and efficient network planning algorithms are critical issues concerning overall TGMTS performance.

Location management is concerned with the procedures required to enable the network to maintain location information for each subscriber, or more specifically, for each active mobile terminal with a registered subscriber, and to efficiently handle the establishment of incoming calls.

The location management procedures are required in cellular networks to manage subscriber mobility. In general, mobility management can be divided into three different aspects: terminal, service and personal mobility, with the latter being the goal of modern cellular systems. Terminal mobility is a network capability that enables mobile terminals to access telecommunication services from any location, while in motion, and the capability of the network to locate and identify a mobile terminal as it moves. In contrast, personal mobility allows customers to access the network, independent of the location and type of terminal that a customer happens to use at a given time. This is accomplished by allocating an identity number not only to the terminal, but also to the subscriber, and keeping an association between the subscriber and the terminal. In practice, this involves the use of a personalized smart card which can be inserted into any compatible terminal, and which allows calls to be made and received at that particular terminal, with proper billing and call routing. Service mobility enables customers to access the same type of network services, regardless of their current locations and terminals in use, as if the customers were at home or office.

The two fundamental procedures that comprise the basis of location management are location updating and paging. For the network to route incoming calls to a mobile user, each mobile user is required to report its location to the network at specific points in time. This reporting is known as Location Update (LU). Terminal paging is the process of contacting a specified user at its current address. Paging messages are broadcast in one or more paging areas, contained within the current location area, and inform the target user of the incoming call. These concepts will be defined in more detail below.

The smallest unit of location information that cellular networks are concerned about is a radio cell. A cell is the geographical area over which the average signal strength from the radio transmitter that defines the cell is over some threshold, as compared to neighboring cells and frequencies. Cells are ideally represented as hexagons, although in reality, interference and path losses due to irregular terrain and neighboring cells, as well as power control algorithms, smear cell boundaries into fuzzy, irregular shapes.

With the conventional location management strategies, the network coverage area is divided into smaller cell clusters called Location Areas (LAs) and Paging Areas (PAs). LA is the granularity at which the network keeps track of the location of the Mobile Terminal (MT), i.e., LA is that group of cells in which the MT must be located (provided that the system is well-functioning and the MT is powered-on). PA is the set of cells over which a paging message is sent to inform the subscriber of an incoming call. In most operational systems, LA and PA are identical. For this reason, any grouping of cells for location management purposes is usually called a LA.

The relationship between location updates, paging, location areas, and paging areas can best be illustrated by the two extreme approaches to location management. The first approach is known as the Always-Update, in which each mobile user transmits an update message whenever it moves into a new cell, and the Never-Update, in which the users never send update messages regarding their location. Clearly, under the former strategy the overhead due to transmissions of update messages is very high, especially in networks with small cells and a large number of highly mobile users, but the overhead of finding users is zero since the current location of each user is always known. With the latter strategy there is no overhead for updating but whenever there is a need to find a particular user, a network-wide search is required, and this overhead is very high.

1.2 Problem Definition

Personal communication service provides seamless and uninterrupted communication to mobile subscribers. PCS users carrying MTs can communicate with a remote terminal (mobile or static) regardless of its current location and mobility pattern. However, unlike static networks, such as the Internet, where routing information is embedded in the address of each node on the network, the current location of a MT cannot be obtained from its identification number. A location management scheme, therefore, is necessary to effectively keep track of the MTs and to locate a called MT when a call is initiated.

Consider the problem of routing a call to a mobile unit M. A request for a connection to M is generated from some point P in the network. Since mobile units change locations, a physical address must be found. The means by which the network obtains information about unit location is through some combination of paging (issuing polling signals in likely locations) and registration (the unit identifying its location). A fair amount of previous work develops paging and registration procedures that minimize some cost function (usually signaling over the wireless and/or fixed network) for simple mobility models or specific systems.

Another body of work addresses the problem of how to optimally store and update location information throughout the network. Specifically, network resources at point P must have some knowledge of how to route a call to unit M. A number of possibilities exist. For example, should there be a central database, which contains the exact location of every mobile terminal, should every location have its own copy of the database, or should locations have partial information? If so, how should it be ordered and updated?

Thus one body of literature optimizes location information gathering (paging and registration); whereas the other assumes terminal locations are known and optimize dissemination of location information over the network. Both, however, make assumptions about the terminal mobility either through the use of simple mobility models or by more general intuitive notions such as locality of movement.

1.3 Motivation

As described in Chapter 2, location management as currently performed is suitable for a relatively small number of subscribers, in a cellular environment that uses relatively large cells. The definition of location areas, which can obviously greatly influence the location updating and paging traffic, is performed statically, using aggregate statistics and traffic patterns. While the overall number of subscribers is relatively low, and the location areas are moderately large, the signaling traffic volume remains acceptable.

One of the important issues in cellular networks is the design and analysis of strategies for tracking the mobile users. In these networks, whenever there is a need to establish communication with any particular user, the network has first to find out which base station can communicate with that user. This is due to the fact that the users are mobile and could be anywhere within the area covered by the network.

In recent years, many sophisticated location management schemes [1-27] have been proposed to reduce profile lookup and update response times and signaling traffic. These methods utilize techniques such as data replication, caching, pointer forwarding and hierarchical database schemes. Review of several papers is given in chapter 3. It is important to note that the actual performance of these proposals depends strongly upon user behavior. For example, the merits of caching and data replication schemes are functions of user mobility and calling patterns. Furthermore, different schemes optimize their performance to different aspects of user behavior. As a result, accurate user behavior models are required when evaluating the performance of location management schemes.

The motivation behind this thesis is the realization that in order to quantitatively and realistically assess any location management scheme, an accurate and plausible mobility model must be used. A mobility model is an analytical or statistical representation of the daily movements of mobile subscribers. In cellular mobile communication networks the number of users allocated to a particular base station is a random variable due to the mobile nature of the users. This inherent feature affects the behavior of the system in terms of the offered traffic. So, the influence of the mobility and traffic of users must be studied.

1.4 Basic Proposal and Thesis Outline

As users become accustomed to wireless communications, they expect to be reachable anywhere, anytime, and in any communication medium. This places a burden on the fixed network to keep track of the user's locations to provide seamless communications as users move from one place to another. A number of location management schemes have been proposed in order to improve upon the performance of existing location management algorithms, especially in second-and-third-generation cellular systems. Some are relatively easy to implement, some might provide high signaling savings on the radio link, and some seem to be interesting from a theoretical/conceptual point of view. The first contribution of this thesis is the detailed survey of some of the recent location management proposals. The second contribution is the use two mobility models to compare the performance of different location management algorithms and showing the impact of the mobility model on the relative performance of these schemes.

Chapter 2 discusses the procedures involved in location management, both in general terms and as implemented in the GSM cellular network standard and a brief review of IS-41 standard. In chapter 3, previous work on location management is reviewed by arranging different proposals in different categories. The advantages and disadvantages of various approaches are discussed. Furthermore several algorithms will be selected amongst these proposals for further evaluation using two mobility models: a realistic mobility model and a random mobility model.

The mobility models that formed the basis of the simulation study are described in Chapter 4. A description of selected algorithms and their implementation under both mobility models is also described in this chapter. In Chapter 5, the results from the simulation study are presented and analyzed. Conclusions from the simulation and possible future work are discussed in Chapter 6.

Chapter 2

Location Management Procedures

2.1 GSM Location Management Fundamentals

GSM is the second generation Pan European Digital cellular standard that has rapidly gained market share worldwide since its debut in 1991. It has been chosen to provide PCS in Europe, partly because widespread roaming is a key requirement in PCS. Unlike analog first-generation systems such as AMPS (Advanced Mobile Phone Services) used in the United States, GSM transmits digital signals over the radio interface as well as internally between different network elements. It provides ISDN compatibility in terms of user services offered and general protocol structure, although not in terms of data rates, since the allocated radio spectrum is limited.

The location management concepts and procedures will be explained in greater detail by looking at their implementation in the GSM digital cellular standard. For many location management proposals, a scheme similar to the one used in GSM is used as the standard of comparison.

2.1.1 The Architecture of GSM Network

The functional architecture of a GSM network can be divided into two levels: the stationary and mobile level. The stationary level comprises of the base station subsystem and the network switching system while the mobile level consists of mobile terminals, as

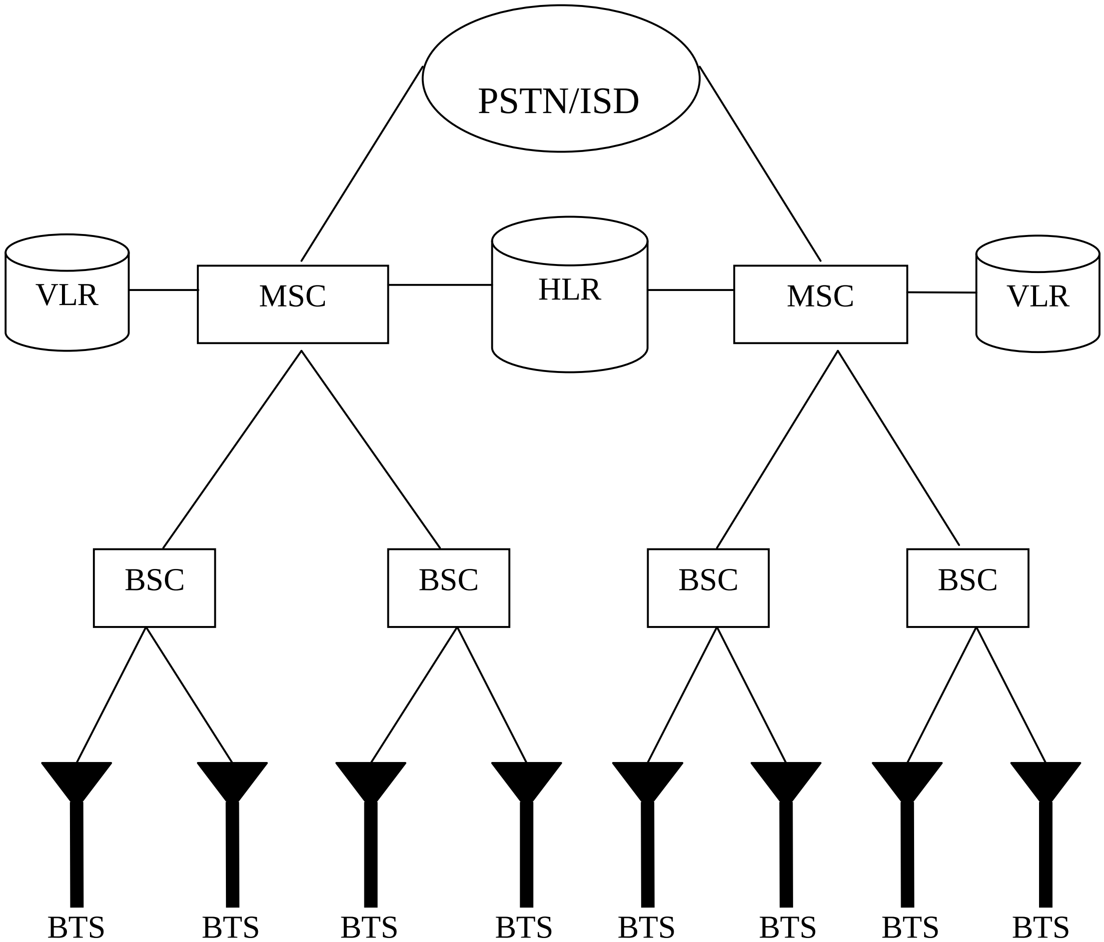


Figure 2.1: Architecture of a GSM Network.

shown in *Figure 2.1*. Each subsystem is comprised of functional entities that communicate through various interfaces using specified protocols.

Mobile Level:

Mobile level consists of Mobile Terminals (MT) that communicate with base stations via wireless links. Mobile terminals are either vehicle-mounted or hand-held terminals used by subscribers to send and receive information. The GSM mobile terminal consists of two entities: the mobile equipment itself and a smart card called the Subscriber Identity

Module (SIM). The SIM card is an integral part of the GSM design, which allows the implementation of personal mobility. A subscriber can plug the SIM card into any GSM terminal and is able to place and receive calls, with proper routing and billing.

Stationary Level:

Base Station Subsystem:

The Base Station Subsystem is composed of two parts, the Base Transceiver Station (BTS) and the Base Station Controller (BSC) and handles almost all radio resource aspects of GSM. The Base Transceiver Station houses the radio transceivers that define a cell and it also handles the radio-link protocols with the MT. A BTS is assigned a set of frequencies in such a way that interference from other cells in the network reusing the same frequencies is kept to a minimum, while maintaining enough radio channels in a cell to handle the required subscriber density. The Base Station Controller manages the radio resources for one or more BTS. It handles setup and release of radio channels, frequency hopping, and inter BTS handoff.

Mobile Switching Center:

At the stationary level, the Mobile Switching Center (MSC) is a central office, which provides the call processing and channel setup functionality. The MSC interconnects Mobile Terminals (MTs) with other MTs through Base Stations (BSs) or interoperates with other MSCs and public fixed networks. The functionality of the MSC is provided in conjunction with the following databases defined in the GSM architecture, which together with the MSC form the Network Switching Subsystem:

- Home Location Register (HLR)
- Visitor Location Register (VLR)
- Authentication Center (AuC)

The MSC and the databases communicate using the Mobile Application Part (MAP) protocol, which uses the services of Signaling System No. 7 (SS7) that is used as a transport mechanism for call control and database transactions.

<u>Home Location Register:</u>

The Home Location register (HLR) in GSM contains the permanent subscriber parameters such as the type of services subscribed, the quality of service (QoS) requirements, the billing information, and the current locations of the MTs. For each subscriber there is a pointer to a Visitor Location Register to assist routing incoming calls. There is logically one HLR per GSM network, although it may be implemented as a distributed database.

<u>Visitor Location Register:</u>

The Visitor Location register (VLR) is a dynamic database, whose entries change as subscriber move within one or more location areas (group of cells). VLR contains detailed information on location and service data regarding those MTs entering its coverage area for the routing of incoming and outgoing calls. The VLR also obtains subscriber parameters from the HLR and updates the HLR regarding the status of supplementary services, if needed. Although the VLR can be implemented as an independent entity, all manufacturers of switching equipment to-date implement the VLR together with the MSC. The combination of the VLR and MSC speeds up access times to information that the MSC requires during call processing.

Authentication Center:

The Authentication Center (AuC) is a strongly protected database that handles the authentication and encryption keys for every single subscriber in the HLR and VLR. The Authentication Center contains a register called the Equipment Identity Register (EIR) which identifies stolen or fraudulently altered phones that transmit identity data that does not match with information contained in either the HLR or VLR.

2.1.2 Location Management Procedures in GSM

For mobility management, GSM takes the intermediate approach between full-system paging and per-cell location updating, using location areas and paging areas (which are equivalent in GSM) made up of one or more cells [25]. The actual assignment of cells to location areas is operator-defined, and is constrained only by the fact that a location area cannot contain cells belonging to more than one MSC.

Location Update and Registration:

Location updating is the procedure for keeping the network informed of where the mobile is roaming. Location updating is always initiated by the mobile terminal on either detecting that it is in a new location area by periodically monitoring the location information broadcast by the network on the broadcast channel, and comparing it to the information previously stored in its memory, or by receiving an indication from the network that it is not known in the VLR upon trying to establish a mobility management connection. No location updates are required while roaming within the same location area.

Location updates and location registrations are very similar, and differ only in additional data transferred on the network side for location registrations. If the mobile terminal registers at a VLR for the first time, no subscriber information exist there yet, and it must be requested form the HLR. For location updates between location areas under the same VLR, the subscriber information already exists.

The steps involved and messages exchanged for location updating are illustrated in *Figure 2.2*, and explained below:

1. The mobile terminal must first be synchronized to the base station, to be able to receive and decode Broadcast Control Channel (BCCH). The BCCH broadcasts information necessary for MT to be able to properly access the network, including the identity of the network, location area, and cell, information for handoff and cell selection, information describing the current control structure, parameters for properly transmitting on the random access channel, and other cell options.

2. The MT transmits a CHANNEL_REQUEST message on the Random Access Channel (RACH). This message contains only 8 information bits in total, and includes the reason for the mobile access (3 to 6 bits) and a random discriminator used to prevent duplicate channel assignments. In this case, the reason for access is location updating.

3. If there is no collision of the access burst, and the MT is not barred from accessing the LA or cell, the Base Transceiver Station (BTS) measures the transmission delay, and

sends a CHAANE_REQUIRED message to the Base Station Controller (BSC).

4. The BSC selects a signaling channel (possibly full rate) and sends its description in a CHANNEL_ACTIVATION to the BTS. If the BTS is able to allocate that channel, it responds with a positive CHANNEL_ACTIVATION acknowledgement.

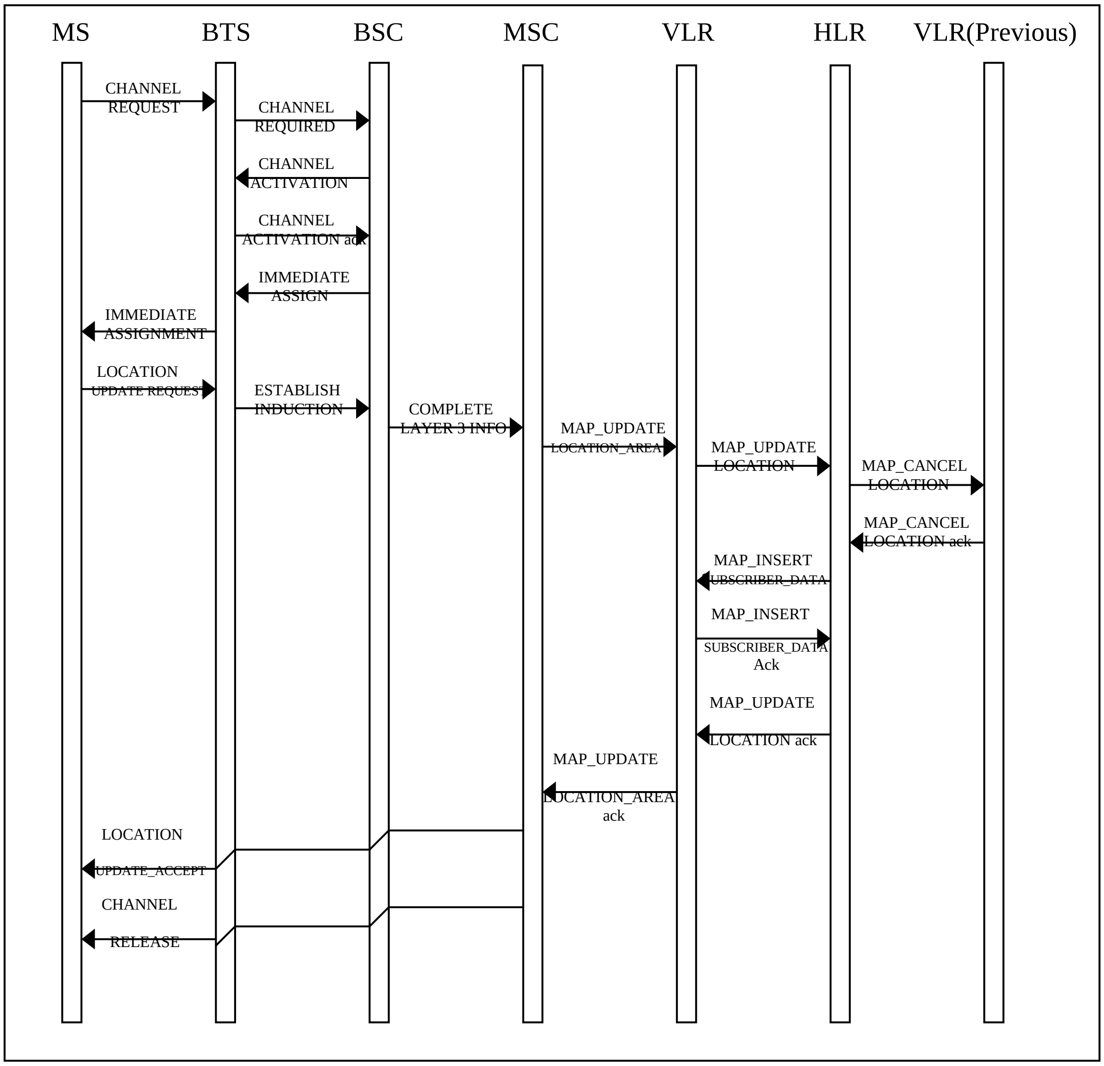


Figure 2.2: Message Exchange for GSM Location Updates.

5. If the BTS channel activation was successful, the BSC will forward an

IMMEDIATE_ASSIGN command to the BTS, which forwards it to the MT as an IMMEDIATE_ASSIGNMENT message, describing the assigned radio channel, on the Access Grant Channel (AGCH).

6. The MT accesses the assigned channel, using the given timing advance and power control parameters, and transmits an initial message indicating its identity (IMSI or TMSI), the mobile equipment classmark (indicating its capabilities), and the type of location update in a LOCATION_UPDATE_REQUEST message. (This layer 3 message is actually piggybacked on a layer 2 frame, whose acknowledgement establishes the layer 2 connection and resolves any access contention issues that may exist.)

7. The BTS sends an ESTABLISH INDICATION message to the BSC confirming the setup of the channel, and carrying the layer 3 location update request information.

8. The BSC sets up an SCCP (Signaling System No.7) connection with the Mobile services Switching Center (MSC). At this point, the MT can communicate with the MSC. Communication between the MSC and the MT is forwarded transparently by the base station subsystem. The BSC forwards a COMPLETE_LAYER 3_INFO message to the MSC, which contains the location update request information form the MT.

9. The MSC uses the Mobile Application Protocol (MAP) to communicate with the appropriate location registers to complete the location update. The MSC sends a MAP_UPDATE_LOCATION_AREA message (service request, technically) to its VLR, which includes the information received from the MT, such as the IMSI or

TMSI, the location update type, and the new Location Area Identifier (LAI).

10. At this point, a few optional procedures may be invoked. These include authentication of the mobile, activation of the ciphering mode, and verification of the International Mobile Equipment Identity (IMEI) status to ensure that the MT itself is not listed as being stolen or incompatible. These procedures will not be described here, but their presence, and additional signaling load, should be noted.

11. If the MT is allowed to register with the VLR, the subscriber's HLR is notified of the change in VLR area through the MAP_UPDATE_LOCATION service request. Some variations are possible here, such as if the MT included its TMSI in the location update request the previous VLR would be queried to obtain the subscriber's IMSI (since TMSI is only valid in conjunction with a LAI). In that case, subsequent procedures, as described here, would be slightly modified, since a TMSI reallocation may be invoked. The service identification to the HLR includes the IMSI and VLR address.

12. Upon receiving the MAP_UPDATE_LOCATION service identification, the HLR sends a MAP_CANCEL_LOCATION request to the subscriber's previous VLR. This deletes any subscriber information from the previous VLR, which then sends a MAP_CANCEL_LOCATION acknowledgement to the HLR.

13. The HLR then sends a MAP_INSERT_SUBSCRIBER_DATA request to the new VLR. Clearly, if the subscriber registers for the first time in the new VLR, no prior subscription information will exist. This request sends information such as the subscriber's MSISDN, the type of subscribed services, details on subscribed

supplementary services (such as call forwarding address), and any restrictions on the subscriber (such as roaming, or operator determined barring). This information is necessary for service provision by the new VLR and MSC. Upon receipt of this information, the request will be acknowledgement by the VLR.

14. Once the subscription information is slightly acknowledgement by the new VLR, the HLR will have completed the location updating, and will acknowledge the MAP_UPDATE_LOCATION request previously sent by the new VLR, which in turn will acknowledge the previous MAP_UPDATE_LOCATION_AREA request that the VLR had received from the MSC.

15. Upon completion of the network side location updating functions, indicated by the MAP_UPDATE_LOCATION_AREA acknowledgement, the MSC will send a LOCATION_UPDATE_REQUEST message to the MT. This message is transparently forwarded by the BSC and BTS, using the Direct Transfer Application Part (DTAP) over the existing dedicated signaling connection to the MT.

16. The location update is now complete. Unless the MT had previously indicated that it wants to maintain the existing signaling link, or the network requires it for any additional signaling, the link is released when the MT receives a CHANNEL_RELEASE request from the network (or the MT timer for receiving this message expires).

Interrogation and Paging:

By definition, a mobile subscriber is not directly associated with a fixed telephone link. When an incoming call from the fixed network (PSTN) is directed at the mobile network, the currently serving MSC for the target subscriber must be found, and a connection must be established to that switch for the duration of the call. This is the general call routing procedure, and interrogation is the mobile network function that determine what MSC the incoming call should be routed to.

As described previously, the mobile network maintains the current location of the subscriber in the HLR. Typically, because of legal, administrative, or technical reasons, only certain switches are designated as Gateway MSCs (GMSCs) which are able to directly query the HLR to obtain information for mobile terminating calls. GMSCs, usually in the destination subscriber's home mobile network, receive incoming calls from the PSTN, query the appropriate HLR for the intended subscriber's current location, and forward the call to the proper destination MSC. Although not technically impossible, PSTN switches are generally not able to directly query GSM network databases, such as the HLR.

As for the description of location updates, the mobile terminating call setup procedure described below and shown in *Figure 2.3,* is for the most case and assumes a successful outcome.

1. The originating PSTN switch will be able to derive address of the GMSC from the dialed digits (the intended mobile subscriber's MSISDN), and will send an ISUP Initial Address Message (IAM) to the GMSC indicating an incoming call.

2. The GMSC is able to derive the destination subscriber's HLR from the first few digits

of the MSISDN. The GMSC will send a MAP_SEND_ROUTING_INFORMATION request, with the destination MSISDN, to the HLR. The GMSC ultimately requires a roaming number (MSRN) which it can use to route the call to the destination MSC.

3. The HLR maps the MSISDN to the IMSI, and looks up the current location of the destination subscriber.. This address is normally in the form of the address of the subscriber's current VLR. The HLR sends a MAP PROVIDE ROAMING_NUMBER request to the current VLR, giving the IMSI of the subscriber and the current MSC (which is generally ignored, since there is in practice a one-to-one relationship between VLR and MSC).

4. The VLR allocates a roaming number (MSRN) from its assigned pool, and maps the IMSI to the MSRN. A response with the allocated roaming number, to MAP_PROVIDE_ROAMING NUMBER request is sent to the HLR.

5. The HLR forwards the MSRN to the GMSC in response to the MAP_SEND_ROUTING INFORMATION request. The GMSC is the able to continue with the call routing, and sends an IAM to the destination MSC, with the new MSRN.

6. Upon receipt of the IAM, the MSC sends a MAP_SEND_INFO_FOR_INCOMING CALL request to the VLR, which indicates the MSRN and any other information that may have been included in the original IAM message, such as bearer service type. A response to this request is not forwarded to the MSC until the subscriber has successfully responded to the paging.

7. Receipt of the MAP_SEND_INFO_FOR_INCOMING_CALL request triggers the

VLR to issue a MAP_PAGE request back to the MSC. This request to page the mobile station includes the IMSI of the destination subscriber and the LAI that is stored in the VLR.

8. The MSC sends a PAGING message to the appropriate BSCs, which includes the IMSI, TMSI (if one has been allocated), a list of the cells to be pages, and, if applicable and known, the type of channel required for he incoming call (e.g., speech or data).

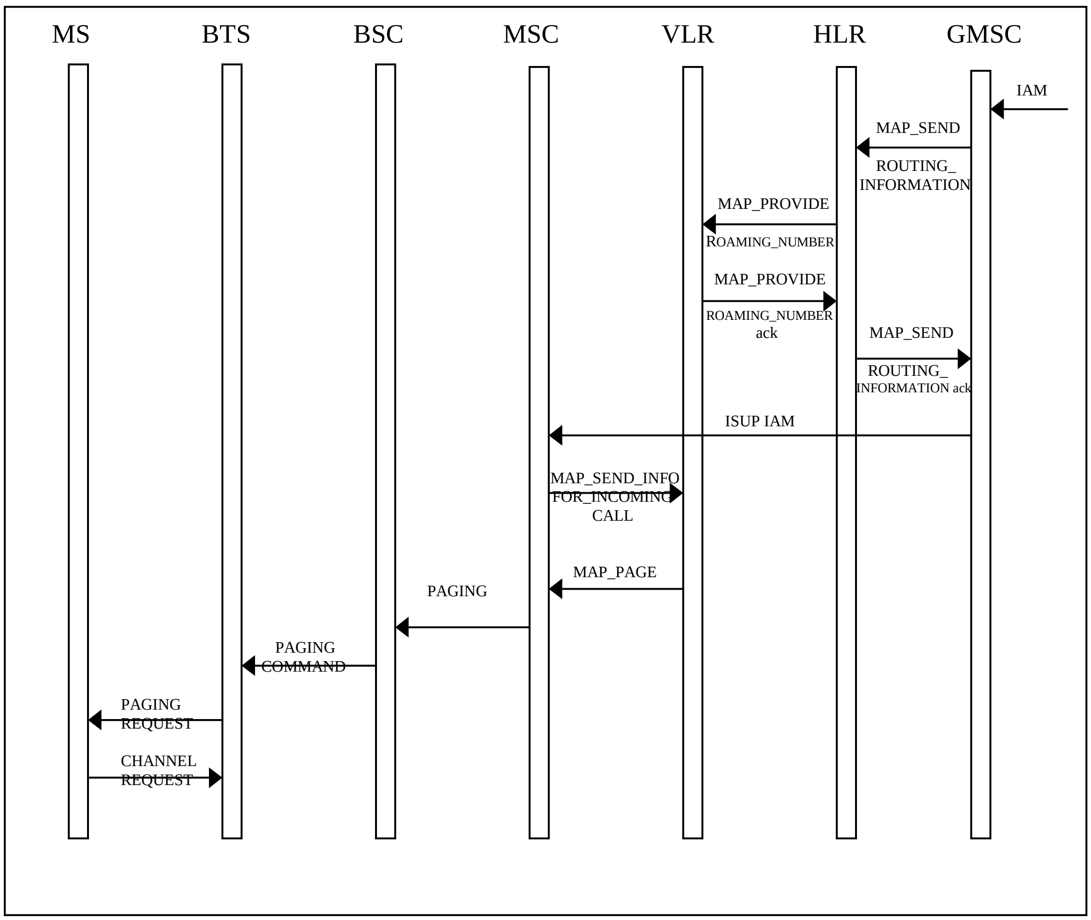


Figure 2.3: Message Exchange for GSM Interrogation and Paging.

9. The BSC sends PAGING_COMMAND messages to each BTS specified in the

previous PAGING message. The PAGING_COMMAND message includes the identity of the mobile subscriber (IMSI or TMSI) to be used, the paging group of the subscriber, and the type of channel required, if applicable.

10. Paging groups are sections of the complete paging channel, in which all paging messages for particular groups of subscribers are normally sent. Since mobile terminals, in general, need to listen only to their own paging group, and a paging group is much shorter (in terms of the number of transmitted timeslots) than the complete paging channel, less time is spent listening to the paging channel, thus conserving battery power. This concept is known as discontinuous reception.

11. There are three types of PAGING_REQUEST messages that may be transmitted over the radio interface. These differ in the type of identifier (IMSI or TMSI) used and the number of paging requests that may be sent in one message. All messages are 23 octets (184 bits) long, and since a TMSI is shorter than an IMSI, more TMSI paging requests can be sent per message (up to 4 with a PAGING_REQUEST_TYPE 3 message). Due to the unpredictability of the radio environment, paging messages are normally repeated an operator–configurable number of times, to decrease the chance of lost pages (for example, if the subscriber was temporarily in a shadowed area.)

12. When the MT successfully receives a paging message, it initiates a channel assignment procedure, identical to the first few steps of the location update procedure, with the exception that the establishment cause is ‘response to paging’ instead of ‘location update’. A context is maintained on the network side to complete the circuit for the call once the MT has responded to the page and a radio-signaling link has been

established.

2.2 IS-41 Location Management Fundamentals

Until the early 1990s U.S. cellular customers that roamed between different cellular systems had to register manually each time they entered a new market covered by a different service provider, during long distance travel. This required the user to call an operator to request registration. In the early 1990s, U.S. cellular carriers implemented the network protocol standard IS-41 to allow different cellular systems to automatically accommodate subscribers who roam into their coverage region. This is called interoperator roaming. IS-41 allows MSCs of different service providers to pass information about their subscribers to other MSCs on demand.

IS-41 relies on a feature of AMPS (Advance Mobile Phone Services) called autonomous registration. Autonomous registration is a process by which a mobile notifies a serving MSC of its presence and location. The mobile accomplishes this by periodically keying up and transmitting its identity information, which allows the MSC to constantly update its customer list. The registration command is sent in the overhead message of each control channel at five or ten minutes intervals, and includes a timer value that each mobile uses to determine the precise time at which it should respond to the serving BS with a registration transmission so that the MSC can validate and update the customer list. IS-41 allows the MSCs of the neighboring systems to automatically handle the registration and location validation of roamers so that users no longer need to manually register as they travel. The visited system creates a VLR record for each new roamer and notifies the home system via IS-41 so it can update its own HLR.

The IS-41 and GSM location management strategies are very similar, except for the following differences. In GSM, when the new VLR receives the registration affirmation from the HLR, it assigns a new TMSI to the MT for the new location area. The HLR also provides the new VLR with all relevant subscriber profile information required for call handling as part of the affirmation massage. Thus, in contrast with IS-41, authentication and subscriber profile information are obtained from the HLR and the old VLR, not just the HLR.

The procedure for delivering calls to mobile users in IS-41 is very similar to that in GSM. The sequence of messages between the caller’s and called party’s MSC/VLRs and the HLR is identical to that shown in call flow diagrams for GSM, although the names and contents, and lengths of messages may be different, see [26] for more detailed descriptions of the IS-41 mobility management strategy.

2.3 Conclusions

The detailed operation of the location management procedures in a practical mobile network, specifically a GSM network, serves to show the considerations that must be taken into account when developing new location management algorithms. For example, many proposals do not consider that at present, and likely for several years, mobile networks and fixed networks are disjoint. They do not have integrated switching and signaling functions, due to not only technical constraints, but also legal and administrative ones.

At this point, it should be noted that location updating, interrogation, and paging are relatively complex in terms of signaling on both the radio interface and within the fixed network, and that there are several factors that should be considered for an efficient location management scheme. Some of the practical issues will be discussed in the literature review in Chapter 3.

Chapter 3

Previous Work on Location Management Proposals with Classification

Emerging mobile and PCS enable communication with a person at any time and at any place. PCS introduce three capabilities: terminal mobility provided by wireless access, personal mobility based on personal numbers, and service mobility through management of user service profiles. Mobility management is an important issue in PCS. One of the main tasks of mobility management is location tracking. Location tracking functions provide PCS with the methods to keep the location of mobile users and terminals. In PCS, subscriber mobility is controlled through location registration and update. The overall location management functionality is comprised of three inter-related aspects of location updating, interrogation, and paging. Different researchers have concentrated on different system resources (such as database storage capacity), quality of service parameters (such as call setup delay) etc. The proposals are evaluated in terms of the overall goals of location management:

- Minimize signaling, both over the radio interface and within the network, to reduce bandwidth and increase capacity.
- Minimize call setup delays in order to maintain acceptable quality of service.
- Minimize the amount of data stored for each subscriber, to reduce equipment cost.

- Minimize the computational complexity of algorithm, to reduce hardware costs, conserve battery power, and improve response times etc.

In general these goals are conflicting and trade-offs must be made. Typically, proposals concentrate on improving one particular aspect by compromising on another.

3.1 Classification of Location Management Strategies

Different proposals exist with the goal of improving some aspect of location management. Some concentrate on improving signaling and querying efficiency of the network-side while other concentrate on reducing radio interface signaling in terms of paging. Thus location management proposals fall in two broad categories: those that concentrate on the interrogation aspect of location management and those that focus on location updating and paging algorithms. *Figure 3.1* shows our classification of location management schemes.

3.1.1 Location Updating and Paging Proposals

Location updating and paging algorithms focus on reducing the overall number of location updating and paging messages sent over the air interface to minimize the total location management cost. This can be achieved if the location of the mobile user is known in advance, so that instead of sending paging messages to all the cells in the LA, only the cell hosting the subscriber is paged.

Current PCS networks partition their coverage areas into a number of LAs. Each LA consists of a group of cells, and each MT performs a location update when it enters a LA.

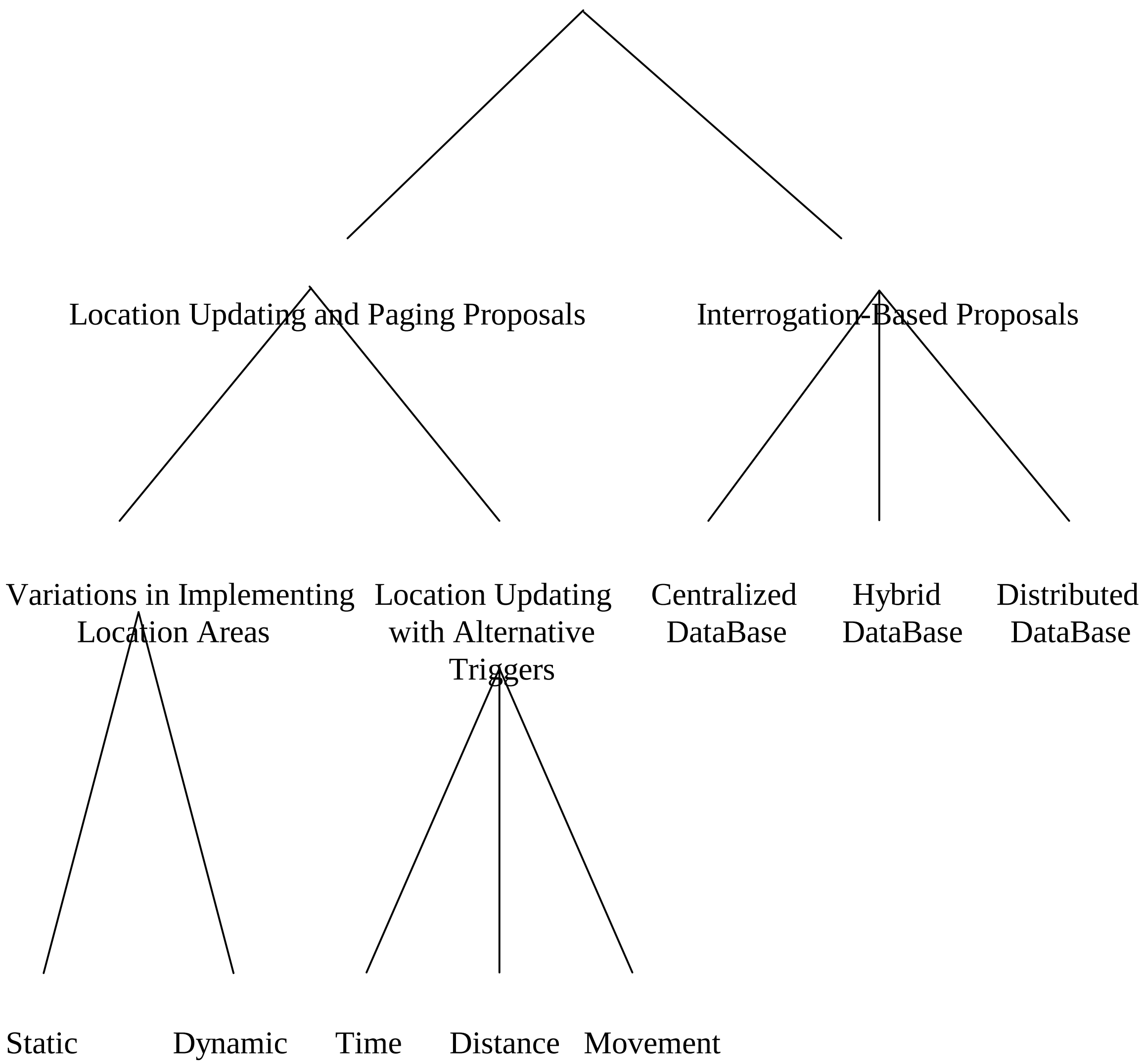


Figure 3.1: Classification of Location Management Proposals.

When an incoming call arrives, the network locates the MT by simultaneously paging all cells within the LA. There are a number of inefficiencies associated with this location update and paging scheme:

- MTs that are located around LA boundaries and make frequent movements back and forth between two LAs may perform excessive location updates.

- Requiring the network to poll all the cell within the LA each time a call arrives may result in an excessive volume of wireless broadcast traffic.
- The mobility and call arrival patterns of each MT vary, and it is generally difficult to select a LA size that is optimal for all users.

The location updating and paging strategies can be divided into two categories; one that makes use of LAs while the other one does not.

3.1.1.1 Variations in Implementing Location Areas

In the conventional location-updating methods like GSM, cells are grouped to form disjoint LAs; these are disjoint in the sense that one cell is included in only one LA. Location areas allow the system to track the mobiles during their roaming in the network(s): subscriber location is known if the system knows the LA in which the subscriber is located. When the system has to establish a communication with the mobile (to route an incoming call, typically), the paging only occurs in the current LA. Thus resource consumption is limited to this LA; paging messages are only transmitted in the cells of this particular LA. To provide additional capacity, while constrained by the fixed available radio spectrum, operators have been using progressively smaller cells, through cell splitting and sectorization. To accommodate this trend, either the number of cells and subscribers per location area must increase (to maintain the same geographical size for the location areas), or the geographical size of location areas must decrease (to maintain the same number of cells per location area). Assuming the same subscriber mobility patterns and location management algorithms, the number of paging messages will increase in the former case, and the number of location updates will increase in the latter case.

The above situation arises because location areas are geographically fixed, and apply to all subscribers, regardless of the individual subscriber mobility patterns. This can be illustrated by considering an example with a fast subscriber (for example, driving a car) and a slow subscriber (for example, a pedestrian). The fast subscriber may cross several location areas during a trip, while the pedestrian may not even cross one. The fast subscriber can be handled efficiently by creating large location areas to avoid numerous location updates. However, if these large location areas would also apply to the pedestrians subscribers, significant signaling would be wasted paging over the entire location area, when paging over a much smaller location area (or smaller paging areas within the location area) would suffice.

Static LAs:

The LA-based location update and paging scheme is a static scheme because it cannot be adjusted based on the parameters of a MT from time to time. One of the problems with this scheme is that location and paging areas are statically defined. The determination of which cell belong to which location areas (and implicitly, paging areas, since the two are usually identical) is through a spatially and temporally static procedure using aggregate data.

S. Tabbane:

In [1], an intelligent method for locating users, called an Alternative Strategy (AS) is introduced. This proposal is based on the observation that the mobility behavior of a majority of the users can be predicted. The strategy can reduce signaling messages involved in mobility management procedures, leading to savings in system resources. This approach reinforces the information capabilities of the system, and provides tools for prediction by integrating (in the system) more information about the external world.

The user profile may be built up by monitoring the user over some period of time. This long-term user profile seldom changes, although the actual updating procedures are not specified. The location accuracy may be improved by reordering the paging areas in the long-term profile according to short-term or medium-term algorithms, or by paging concentrically around the last known location area. The former approach attempts to reorder the paging areas so that areas associated with a higher probability of containing the subscriber are paged first, reducing signaling and delay. Call arrival and origination rates from particular paging areas may be used for a medium-term reordering of the user profile. Alternatively, estimated distance traveled since the last known location may be used at the time of an incoming call to determine which areas to page first. The paper mentioned that with this approach, and using different mobility models, parameters and evaluation methods, very high signaling savings have been reported.

<u>Gregory P.Pollini and Chih-Lin I:</u>

The strategy in [2] proposed a Profile-Based Strategy (PBS) to reduce the signaling traffic on the radio link by increasing the intelligence within the fixed network. The system maintains a sequential list of the most likely places where each user is located (in terms of a given set of LAs). The list is ranked from the most to the least likely place where a user is found. When a call arrives for a mobile, it is paged sequentially in each location within the list. When the user moves between location areas in this list, no location update is required. The list may be provided by the user or may be based on each user's past mobility history.

The Profile-Based strategy relies on the observation that most users follow regimented daily schedules that can be exploited to estimate their current location. If the system knows that at a given time the user is most likely to be in a certain location area, second most likely to be in a second location area, etc., the system can maintain a list of likely locations ordered from most to least likely. When a call arrives for a user, it is paged sequentially in the order indicated by the list.

<u>Ozan K. Tonguz et al:</u>

In [3], an analytical framework is presented for analyzing the performance of different schemes for location management. This paper considers two paging methods together with location updating techniques. The paging techniques considered are simultaneous paging and sequential paging. In case of simultaneous paging, all cells inside a LA will be paged simultaneously when there is a call to any one cell in the LA. In sequential paging, paging is done in k-cell PAs sequentially until the called MT is found. The paging traffic depends on the average number of paged cells till the MT is located. This formulation leads to a closed form expressions for the cost of location management in terms of the call-arrival rate, boundary crossing rate, the size of the LA, the maximum acceptable average paging delay, the total number of cells controlled by the Mobile Switching Center (MSC), etc.

It is suggested that the best strategy is sequential paging combined with some intelligent features. As suggested, useful information related to the MT, such as the most recent interaction area and its speed, can be kept track of, and this information can be used in carrying out the paging process. This extra information will increase the probability of finding the MT in the first area paged. Of course, this depends on the user speed and the time lapse between the last update of user location and call received. Paging is still done sequentially, and the difference resides in the way in which the paging base stations are selected (i.e., the added 'intelligence'). This intelligent paging does lead to a cost reduction compared to random sequential paging at an expense of extra complexity.

Yi-Bing Lin:

The proposal in [4] studies a special case of AS [1] called the Two-Location Algorithm (TLA). In TLA, two locations to track the location of the MT are considered. On the other hand, AS allows more than two user locations in the profile. When a phone call arrives, two LA identifiers are used to find the actual location of the portable. The order of the identifiers selected to locate the MT affects the performance of the algorithm. The MT has a small built-in memory to store the identifiers for the two most recently visited LA's. The record of the MT in the Home Location Register (HLR) also has an extra field to store the corresponding two locations. The first memory location stores the most recently visited LA.

When the MT joins the network, the location is stored in its memory, and a registration operation is required to modify the HLR record. When the MT moves to a new location, it checks if the new location is in the memory. If the new location is not found, the identifier for the LA that the MT just left is kept, and the other address is replaced by the address for the new LA. A registration operation is required to make the same modification in the HLR record. If the address for the new location is already in the memory, no registration operation is performed. The order of the addresses selected to locate the portable affects the performance of the algorithm. If the MT is located in the first try (referred to as location hit), then the paging cost for the TLA is the same as the IS-41 algorithm. Otherwise, an extra penalty is incurred by TLA to locate the MT for the second try (referred to as location miss; the second try is always successful). After the second try, the HLR identifies the LA where the MT is visiting.

Dynamic LAs:

Recent research efforts focus primarily on dynamic location update mechanisms, which perform location update based on the mobility of the MTs and the frequency of incoming calls. Several proposals suggest the use of user profiles to individualize the location areas. A personal, or individualized, location area, anchored around the cell in which the location update occurs, would consist of the most likely cells that the subscriber could traverse. The information to create the personal location area is derived from the user profile, which contains a record of previous transitions from cell to cell, together with the length of time spent in each cell. Some location management proposals are based on grouping subscribers with similar behavior. Other proposals consider the use of multiple LA levels to track locations of mobile terminals. The nominal LA sizes at each level are different. Each mobile terminal is registered at a LA of suitable level, and the registered level is dynamically changed according to its past as well as present mobilities.

Anna Hac and Xian Zhou:

The proposed algorithm in [5] allows a MT to transmit update messages only at specific cells (the number of which is small, compared to the total number of cells in the network), while restricting every search for a mobile user to a small subset of cells.

In a simple cellular network, there is one base station per cell and a large number of base stations in the system. In the proposed scheme, a subset of base stations called reporting centers is selected among all base stations. The cells associated with these base stations are referred to as reporting cells, while other cells are called non-reporting cells. The vicinity of reporting cell *i* is the collection of all non-reporting cells that are reachable from cell *i* without crossing another reporting cell. Also, cell *i* is in the vicinity of itself.

The frequency with which the mobile users enter a reporting cell corresponds to the updating cost for the cell. The higher the frequency, the higher the updating cost. If the cell with highest frequency is selected as a non-reporting cell, the users do not need to update their location whenever they enter this cell. This reduces the total updating cost by the largest amount for the entire network. This step is continued, and the cell with the highest frequency among all remaining reporting cells is reselected. This procedure continues until all cells are tried.

The proposed strategy provides a partition scheme to optimize the size of the location area domain. Simulations are performed to compare the performance of three schemes: always update, always search and the proposed scheme. The results show that the proposed scheme performs better than the other two.

L.R. Hu and S.S.Rappaport:

The paper in [6] proposes an efficient multilayer mobility management scheme that is especially suitable for hierarchical cellular structures. The scheme accounts for past as well as present terminal mobilities and dynamically changes the hierarchical level of the location area to which the mobile terminal registers. The nominal LA sizes at each level are different. Each mobile terminal is registered at a LA of suitable level; highly mobile terminals such as an aircraft will tend to register at the highest level where location areas are formed by spotbeams. Low mobility terminals such as pedestrians will tend to register at the lowest level where location areas are formed by microcells or clusters of microcells. The location-updating rate that is generated by a mobile terminal tends to remain fixed over a wide range of terminal mobilities.

The system uses multiple registration levels to track the locations of mobile terminals. Based on the dwell time, time that an active MT stays in the LA of a given registration level, of a mobile terminal in the previous LA, a mobile terminal is switched to a suitable level of location management. Mobile terminals tracked by the proposed scheme are seen to generate signaling traffic at a very stable rate over a broad class of terminal mobility. This feature may prevent the system from excessive signaling traffic due to high mobility users and can be used to set the required long-term total signaling traffic.

<u>Thomas Kunz and John Scourias:</u>

The proposal in [7] discusses a stochastic mobility model based on daily activity patterns, providing a realistic balance between completely deterministic and completely random mobility models. Activities, together with their location and duration, are selected using subscriber-specific data, as well as randomly selected data from tables derived from traffic planning survey data. The path to the next activity's destination, represented as a sequence of traversed radio cells, generates cell crossings used for location update measurements. The current location of simulated subscribers is continuously tracked and used for paging measurements.

The model was used as a tool for developing a location management algorithm utilizing user profiles for the dynamic creation of location and paging areas. Based on the number of times a subscriber had moved from the current cell to the neighboring cells, a personal location area is dynamically defined during a location update. A history of the average time spent in each cell is used dynamically to determine a subset of cells within the personal location area most likely to contain the subscriber, and which is paged first.

The proposed algorithm is compared with the algorithm currently in systems like GSM. It is shown that the proposed algorithm significantly reduce the total signaling cost, for nearly all different subscriber types defined, levels of incoming calls, work location variability, and dynamic location area size as compared to currently used techniques.

3.1.1.2 Location Updating with Alternative Triggers

All previously discussed variations in LA implementations rely on the same location-updating mechanism: a location updating procedure is performed if the MT exits its current LA; paging is always performed inside the LA. Other location-update schemes have been proposed that do not require the division of the network coverage area in LAs. The only location information available to the network is the last point of contact with the MT, either because of a location update, a connection, or another MT-Network interaction. These non-LA-oriented location-updating methods are inherently individualized, which would permit parameter optimization for each user individually. Several dynamic location update and paging schemes are described below:

<u>Time-Based:</u>

The time-based location update method [8] is the simplest because it just requires the mobile to periodically transmit its identity to the network without regarding to its mobility and incoming call arrival rate. Each MT has a timer running. On expiration of the timer, a location update is performed and the timer is set to zero, and then begins to run again. Whenever a user is looked for by the system, it is first searched at the cell it is last reported to, say i. If it is not found there, then it is searched in cells $i \pm j$ for every j – 1,2…,N/2, where N is number of cells, starting with $j = 1$ and continuing upward until it is found.

<u>Movement-Based:</u>

According to the movement-based strategy [8], each mobile user counts the number of cell boundary crossings and updates its location when the count exceeds a predefined parameter M (this number is referred to as movement threshold). When a mobile is paged, it can be found within a neighborhood of atmost M cells away from the previous reporting cell. In this case, each mobile needs to keep track of the number of cell boundary crossings since its last location update.

Distance-based:

In a third proposal, a MT performs a location update when its distance from the cell where it performed the last location update exceeds a predefined value D (this distance value is referred to as the distance threshold). A distance-based location update scheme is considered in [9]. The authors introduce an iterative algorithm that can generate the optimal threshold distance, which results in minimum cost. When an incoming call arrives, cells are paged in a shortest-distance-first-order such that cells closest to the cell where the last location update occurred are polled first. The delay in locating a MT is therefore proportional to the distance traveled since the last location update. Results demonstrated that depending on the mobility and call arrival parameters, the optimal movement threshold varies widely.

Comparison:

One study [10] compared three different location updating strategies with respect to the update costs, paging costs, and total costs of tracking mobile users by constructing an object-oriented environment in Java. In their comparison of the different update strategies, the authors claim that the distance-based method gives the lowest combined location updating and paging cost.. The time-based strategy causes updates periodically independent of how a user moves, so the update cost of this strategy is steadier than that of the other strategies. However the paging cost increases rapidly in proportion to MR, the Mobility Rate, which is defined as the average number of cells changed by the user per unit time.

In case of movement and distance-based strategies, for low MR, the update costs of these strategies are smaller than, and the paging costs are larger than that of the time-based strategy. While, for a high MR, these strategies result in larger update cost and smaller paging cost than the time-based strategy.

3.1.2 Interrogation-Based Proposals

Several proposals have attempted to improve the signaling and querying efficiency of the network-side database querying required by location management, with a secondary goal of reducing radio interface signaling in terms of paging. The driving requirement is to remove the bottleneck that may arise as a result of a centralized HLR, which must be queried for every incoming call and every location update associated with the subscriber.

LA partitioning, and thus mobility management cost, partly relies on the system architecture (e.g., database locations). Designing an appropriate database organization can reduce this signaling traffic.

Research in the area of database architectures generally falls into two categories. First, extensions to the GSM location management strategy are developed which aim to improve the GSM scheme while keeping the basic database architecture unchanged. This type of solution has the advantage of easy adaptation to the current PCS networks without major modification. Another category of research results in completely new database architectures which require a new set of schemes for location registration and call delivery.

Three kinds of network database architectures are closely related to mobility management schemes. The first is a centralized database architecture. All user information is stored in a central database. This architecture is simple and has the advantage of easy operation and control, if it is applied to a system that provides comparatively small-scale service in a limited area.

The second architecture is a hierarchical (distributed) database architecture which uses several independent databases according to the geographical proximity or service providers. The two level HLR/VLR database architecture as described in GSM [25] and IS-41 [26] standards are replaced by a large number of location databases. The location databases are organized as a tree with the root at the top and the leaves at the bottom. Each leaf-level database serves a location registration area and stores user information. Each database in the higher-levels stores pointers to the next lower-level database that stores user information or has a pointer to a lower-level database.

The third case is the hybrid database architecture that combines the centralized and distributed database architectures. In this case, a central database (HLR-like) is used to store all user information. Other smaller databases (VLR-like) are distributed all over the network. These Visitor Location Register (VLR) databases store portions of HLR user records. This can lower the congestion probability in the HLR to some degree but has overheads in multiple database operation because the VLR has to obtain MT information from the old VLR and delete it in the old VLR. A single GSM network is an example of such architecture. In some cases there can be more than one HLRs distributed in the network.

3.1.2.1 Distributed Database Architecture

Jan Jannink, Derek Lam, Narayanan Sivakumar, Jennifer Widom and Donald C.Cox:

The basic idea behind [11] is the concept of life-long numbering without making use of HLR/VLR. In IS-41 and GSM, each subscriber is assigned a phone number that contains enough information for the network to derive the location of the HLR. These phone numbers are, therefore, geographical and are tied to specific HLRs.

In a basic hierarchical model, each leaf-level database services a zone and stores profile of users located in that zone. Each database in the higher levels of the hierarchy stores "pointers" (user ID + database ID) to the next lower level database that stores the user's profile or has a pointer to a lower level database. There is a conceptual "root" database that stores a pointer for every user. In practice, the root may be distributed among several databases so that no one database needs to store all user pointers or service all root level queries and updates. This simple hierarchical LMT provides life-long numbering since there is no concept of a "home site" for a user as in HLR/VLR.

When user A calls user B, the location lookup for B's profile first propagates up the hierarchy from A's zone to the first database that contains a pointer to B's profile. The query then propagates down the hierarchy following pointers to B's profile until it is found in the leaf-level database that services the zone in which B is currently located. When user A moves from the current zone to another zone, the current zone ships a copy of A's profile to the next zone, and deletes its own copy. The databases along the path to the least common ancestor of both zones are then updated so that the pointers to A's profile reflect the new location of A. The idea behind profile replication is to reduce the latency of profile lookup at the expense of increased update and storage cost. Essentially, a user's profile is replicated at a set of databases such that the benefit of local lookups at these databases outweighs the update cost due to the user's mobility. For this the algorithm computes the Local Call to Mobility Ratio (LCMR) of the subscriber, defined as the number of calls from a particular switch to a particular subscriber, relative to the number of location updates performed by the subscriber. If the LCMR is greater than some predefined minimum threshold value, the user profile is replicated.

In [12], the authors developed another LMT, Hierarchical Online Parametric ProfilE Replication (HOPPER), that uses a hierarchical organization of databases and on-line profile replication to achieve life-long numbering with scalability and efficiency. In contrast to off-line schemes, the on-line replication technique uses on-line LCMR estimates and dynamically migrates replicas to appropriate databases. On-line schemes have the advantage that they can adapt to changes in user calling and mobility patterns by dynamically migrating replicas around the network. Due to practical storage constraints, it may not be possible to place replicas at all locations where there are expected benefits. By dynamically adjusting the locations of replicas, on-line schemes can increase the "usefulness" of a replica and achieve the same lookup performance of off-line schemes with smaller storage requirements.

Shlomi Dolev et al:

In [13], the authors suggested a data structure for location management in mobile networks. The data structure is based on the tree location database structure. The algorithm replaces the root and some higher levels of the tree with another structure that balances the average load of search request. The new structure of the upper levels improves the performance of the tree structure by spreading the load of queries. For this modification the authors use a set-ary butterfly network which is a generalization of the k-ary butterfly network. The root of the modified tree has to maintain the location information for only some portion of the mobile terminals. This inturn keeps the number of update messages small; only the embedded tree for the mobile terminal for which the location update message has been sent is involved in the update. The proposal also suggested modifying the lower level of the tree to reflect better neighbor relation between geographic cells to support simple location data management. This is done by adding connections to the lowest level of the tree such that any two MSCs that correspond to neighboring geographic regions will have either a common parent or a direct link connecting them. This may result in efficient local broadcasts and may support handoffs. The modification of the lower level supports a handoff procedure called ‘follow-me handoff’. The update of the proposed location database ensures correct location data following any number of transient faults that corrupt the location database information, and thus is self-stabilizing.

<u>P. Krishna et al:</u>

The goal of a location management scheme should be to provide efficient searches and updates. In [14] the authors present strategies for updates, search-updates and a search protocol. A search-update occurs after a successful search, when location information corresponding to the searched terminal is updated at some terminals. The authors present several location management strategies based on a hierarchical tree structure database.

The structure consists of MTs, BSs, and MSCs. The base stations are located at the leaf level of the tree. Each BS maintains information of the terminals residing in its cell. There is a unique 'home' address for every mobile terminal. The home address is the identifier or name of the mobile terminal. The physical address of a mobile terminal might change, but its home address remains the same irrespective of the terminal's location. Each MSC maintains an address-mapping table that maps the home address to the physical address of the MT residing in the subtree beneath it.

The paper presents the results of simulations carried out to evaluate the performance of various static location management strategies for various call-mobility patterns. The cost metric is the number of messages required for each operation (search, update and search-update). The performance parameter of interest is the total cost, defined as the sum of average update cost, average search cost, and the average search-update cost. Simulations show that performing search-updates significantly reduces the total costs.

3.1.2.2 Hybrid Database Architecture

<u>Shivakumar and Widom:</u>

The proposal by Shivakumar and Widom [15] discusses replication of the user profile, containing the subscriber's current location, in several location registers. Assuming an integrated network architecture, outgoing calls from a particular zone (area controlled by a VLR) would first check the local VLR, in an attempt to avoid querying a distant centralized database, i.e., the HLR, for the current location of a specific mobile subscriber. The decision of where to place the replicated profiles, given the costs for updating the replicated records, is based on a minimum-cost, maximum-flow algorithm described in the paper. The HLR maintains the current location of its mobile subscribers, and manages the updating of all their replicated records.

The algorithm requires the construction of a flow network, which is a graph containing nodes for each subscriber and each zone in the network. The edges between the subscriber nodes and the zone nodes have attributes that require estimates for the number of location updates performed by a particular subscriber. These estimates, together with estimates on the replication costs and database capacities, are used by the minimum-cost maximum-flow algorithm to determine for which subscribers, and in which nodes, replication of the user profile would be cost-efficient. The possibility of varying mobility and call arrival patterns is considered, and four algorithms are presented for modifying the flow network and minimum-cost, maximum-flow algorithm to avoid having to recompute the entire algorithm, which is exponentially bound.

Chih-Lin, Gregory P.Pollini and Richard D. Gitlin:

The Reverse Virtual Call (RVC) setup is a scheme for delivering mobile terminal call [16]. It allows, under the constraint that the LA is not smaller than the VLR area, a reduction in the number of signaling messages exchanged between the called and calling databases and switches. The primary motivation is that too much effort is being exerted by the fixed network to track users who seldom receive calls. An added factor is that a very high percentage of call attempts fail and consume much of the network resources unnecessarily. The goal of this paper is to deliver calls without keeping track of user locations and to minimize the cost of failed call attempts. RVC combines paging with physically reversed call setup to form a logically forward call setup.

As in GSM, all existing cellular standards allocate network resources (voice trunks) prior to paging a mobile terminal. If the terminal does not respond, the network must tear down these resources. Users who simply do not leave their phones turned on cause large portions of failed call attempts. In RVC setup, voice trunks are not reserved until the mobile is found and willing to accept the call. Moreover, the signaling load and the delay incurred by the currently standardized HLR/VLR method of tracking mobile users are expected to be high. RVC can function within the existing cellular paging network or within an integrated overlaid paging network. In RVC, instead of setting up voice trunks to the terminating switch where the callee last registered and instructing the terminating switch to page the mobile, the signaling system is used to instruct the base stations of the global paging network to page the network. The paging signal contains the ID of the caller. If the callee decides to accept the call, the mobile terminal automatically initiates a call setup to the caller. The call is routed from the terminating switch to the originating switch and is bridged to the caller. The call is completed in such a way that the RVC feature is transparent to both the caller and the callee.

<u>Joseph S. M. Ho and Ian F. Akyildiz:</u>

A Dynamic Hierarchical Database Architecture for Location Management in PCS Networks is proposed in [17]. The proposed scheme allows the dynamic adjustments of user location information distribution based on the mobility and calling patterns of the MT. A unique distribution strategy is determined for each MT, and location pointers are set up at selected remote locations that indicate the current location of the MTs.

Both the IS-41 and GSM standards employ a two-level database architecture consisting of the HLR and the VLRs. The proposed algorithm introduces a location management scheme based on a three-level database architecture. Under this approach, an additional level of databases called the Directory Registers (DRs) is introduced. Each DR determines the location information distribution strategies for its associated MTs. Location pointers are then set up at selected remote DRs indicating the current location of the MT. The functions of the HLR, the VLRs and the MSCs remain primarily unchanged. The distribution strategy for each MT is unique and is determined based on its mobility and call arrival parameters. The proposal introduces an algorithm for determining the per-user location information distribution strategy, which reduces the signaling, and database access costs for location management. The cost reduction is most significant when the Local Call to Mobility Ratio (LCMR) is low and the cost of accessing the HLR is high.

It is demonstrated that the proposed location management scheme results in significant saving in both the signaling and database access costs for location registration and call delivery. Based on the cost parameters considered in the analysis, the signaling and database access cost can be reduced by as much as 70% and 50% respectively, compared to the IS-41 standard.

R.Jain and Y-Bing Lin:

For a mobile which frequently moves across LAs but seldom has an incoming call, traditional IS-41 registration operation is wasteful in network cost since every registration operation must update the HLR to record the most recent location of a mobile. Infact, access to the HLR is much more expensive than access to VLR. Based on these observations, the pointer forwarding with single HLR (PFSHLR) [18] scheme is proposed to reduce the waste.

The basic idea is that instead of reporting a location change to the HLR, every time the MT moves to an area belonging to a different VLR, the reporting can be eliminated by simply setting up a forwarding pointer from the old VLR to the new VLR. A mobile arriving in a new LA sends a registration message to the VLR of the new LA. Instead of sending a message to inform the HLR, the new VLR sends a message to inform the old VLR about the mobile's departure. On receiving this message, the old VLR deletes the obsolete record for the mobile and creates a forwarding pointer pointing to the new LA. When a call for the MT is initiated, the network locates the MT by first determining the VLR at the beginning of the pointer chain and then follows the pointers to the current serving VLR of the MT. To minimize the delay in locating a MT, the length of the pointer chain is limited to a predefined maximum value. When the length of the pointer chain reaches the maximum value, additional forwarding is not allowed, and the location change must be reported to the HLR, when the next movement occurs. It is demonstrated that, depending on the mobility and call arrival parameters, this scheme may not always result in a reduction in cost from the original IS-41 scheme.

Yi-Bing Lin and Wen-Nung Tsai:

Another scheme combining pointer forwarding with distributed HLRs (PFDHLR), was proposed in [19]. The basic idea is that in IS-41 the HLR must determine the VLR currently serving the callee and query it by sending a route-request signal. Therefore, the HLR may become a bottleneck due to heavy signaling traffic generated by location tracking. *Lin* proposes a pointer forwarding with a distributed HLR (PFDHLR) scheme to eliminate this overhead in a distributed HLR environment. In this scheme, HLRs are distributed in remote Public Switching Telephone Networks (PSTNs). Like pointer forwarding with a single HLR strategy, every HLR in PFDHLR records a LA where the mobile may exist. When an incoming call is originated from some remote PSTN, the HLR will be directly queried. Next, the HLR at the PSTN queries the VLR recorded in the entry associated with the callee. If the mobile is found in this LA, a call will be delivered, otherwise, the system will trace the mobile through the forwarding pointer chain starting from that LA. This scheme effectively and efficiently implements the distributed HLR notation. However, it has some potential side effects. The length of the pointer chain may increase. Because of the distributed nature of the HLRs, the total incoming calls will be served by different HLRs. Therefore, the call-to-mobility ratio for each distributed HLR is I/N times that in PFSHLR, where N is the number of remote PSTNs. The expected length of each pointer chain increases considerably when the call-to-mobility ration is small.

Kuen-Liang Sue and Chien-Chao Tseng:

A scheme considering a one-step pointer forwarding strategy for location tracking in a distributed HLR environment is considered in [20]. In this scheme, every distributed HLR has one record for an authorized mobile. This records the individual LA/VLR where the mobile was found the last time by an individual HLR. Different HLRs may record different LA/VLRs as the head of their own locating paths, so they may have different locating paths; therefore one cellular system with N distributed HLRs may have *n* different locating pointer chains, $n < N$, for the same mobile. Whenever a mobile registers in a new LA, the VLR for the new LA will maintain an entry for this mobile. This entry is the common rear of the *n* locating pointer chains. Similar to a double-linked list, every such entry must record the previous nodes in the n locating pointer chains. In this manner, the VLR can send a deregistration message through all locating paths. This message is forwarded to the proper VLR via the previous link in each entry. Next, all locating paths are migrated to guarantee that each pointer chain's length among the VLRs does not exceed one.

Simulations were carried out to compare the three pointer forwarding strategies namely, PFSHLR, PFDHLR and OSPFDHLR. Simulation results demonstrate that, compared with PFDHLR, OSPFDHLR can reduce the call delivery time/ cost significantly, particularly when the number of distributed HLRs is growing.

Kin K. Leung and Yonatan Levy:

The proposal in [21] explores the use of replicated databases for management of customer data (e.g., mobility data, call routing logic) in global, intelligent, and wireless networks.

Two schemes namely full and partial data replication schemes, which are compatible with industry protocol standards (IS-41, GSM etc.), are analyzed and a comparison is made with the traditional, centralized database scheme. In centralized database design, all call setup and other signaling functions access the centralized database in the signaling network.

The authors considered such mobility services across multiple, international networks. Depending on the anticipated traffic load and other engineering and economic considerations, the world is divided into two or more regions, each of which covers a group of countries. It is assumed that the global customers are based in the United States and possibly are traveling in a foreign country. A centralized design, where the customer location information is stored in a centralized database is physically located in the home country (USA). An identical pair of Regional Databases (RDBs) is installed in each region and connected to the U.S. signaling network, e.g., via additional Signal Transfer Points (STPs).

The replicated database can be accessed for call setup and other signaling purposes as follows. Calls initiated in the U.S.A. will query the Home Database (HDB). On the other hand, for calls originated from a visited country, queries are first launched to the RDB. If a call is destined for a customer visiting the region, the RDB has the customer record, and is capable of handling the associated queries. Otherwise, when the RDB cannot find the needed records, the queries are forwarded over the signaling network to the HDB for further processing.

The replicated database design is further divided into two, full and partial, data replication schemes. In the full replication scheme, each customer visiting a foreign region has two complete replicated copies of customer data: one in the HDB and the other in the RDB. In contrast, in the partial replication scheme, only the RDB keeps a complete set of customer data, whereas the HDB maintains (replicates) only a subset of customer data that are infrequently updated. For example, the HDB does not keep the location data for highly mobile, wireless customers visiting a foreign region. The idea of the partial data replication scheme is to reduce the overhead of updating certain customer data that are frequently changed.

The performance results indicate that the choice of the database architecture for global services depends on the specific traffic and mobility parameters such as the proportion of customers visiting a foreign region, the fraction of calls generated overseas, the query-to-update ratio, and the query-to-registration ratio. Numerical results reveal that the full replication scheme outperforms the centralized one over a wide range of parameters mentioned before. Furthermore, if some customer data, such as location data for highly mobile customers in wireless networks changes frequently, and if each call launches multiple queries into the databases, the partial replication scheme offers further performance improvements.

3.2 Proposals Evaluation

It is difficult to compare different location management schemes since as yet there exists no absolute bounds on optimum performance which can be applied to any procedure regardless of assumptions about mobility, network structure, etc. In particular, there is a real need to obtain realistic performance evaluation models based on actual users' movements taking into account network structure. Different researchers have concentrated on different system resources (such as database storage capacity), Quality of Service parameters (such as call setup delay) etc. There have been several different approaches to mobility modeling for location management, and some of the methods used in the literature are reviewed.

Location update rates, call to mobility ratios, and other metrics used in the proposals depend on some description or model of subscriber movements between location areas. In many cases, some type of random mobility model is used. In certain papers, where the mobility model is explicitly required in mathematical derivations, a one-dimensional model is used to simplify the calculations.

Static LAs:

Static schemes have the advantage that the LA sizes are pre-defined and there is no need to implement complex algorithms to calculate dynamic LA size for each subscriber according to its characteristics. The disadvantages are that they cannot be adjusted according to the parameter of the individual user. For example, both slow-mobility pedestrians and fast-moving vehicles are assigned the same size location areas. Similarly, high-mobility time periods (e.g., morning rush hour) have the same location area assignments as low-mobility time periods, even though significant signaling savings could be realized by using larger location areas during high-mobility time periods.

S. Tabbane:

An analytical evaluation of Tabbane's proposal [1] is given which assumes typical values for certain parameters, such as cell size, average subscriber velocity, and average number of call arrivals and call origination. The size and number of signaling messages generated over each interface per unit time is the chosen comparison metric. A simulation is also used for comparison. The mobility model is a simple one, with a subscriber moving with an average velocity and random direction, having a certain probability of remaining in a certain paging area. The call arrival and origination rates are uniformly distributed. The Mobility Predictability Level (MPL) is a key parameter used in the comparisons to give an estimate of the randomness of the mobility patterns, and is given by MPL = $\Sigma^{k}_{i=1} p_i$, where k is the number of paging areas for the current time period, and p_i is the probability of locating a subscriber in area i. Results were plotted comparing the number of messages across various network interfaces, and the ratio of total cost of the proposed algorithm versus the standard GSM algorithm, as a function of the MPL.

The main advantage of the AS scheme is to save location updates by storing a user's mobility information in profiles. So, the more predictable the user's mobility, the lower the mobility management cost. Drawbacks are a possibility of increased paging delay (when the system must page a user in different locations in a sequential manner), increased memory requirements for storing each user's characteristics and LA parameters, and profile updating costs for users, which do not have long-term periodic mobility patterns. In addition, the methods for initializing the user profile, as well as long-term updates, are not clearly specified.

<u>Gregory P.Pollini and Chih-Lin I:</u>

The analysis in [2] is the extension of Tabbane proposal [1] that is based on the fixed network performance. Using a general cost formulation, the paper extends that methodology to study the radio link bandwidth usage, fixed networks signaling load, and call setup delay.

The proposal group people into classes depending upon the predictability of their daily routine. Although plausible, such grouping is fairly arbitrary since subscriber mobility characteristics may change. In addition the profile-based strategy of this paper is not applicable for the random users and is only applicable to either deterministic or quasi-deterministic users.

The system maintains a list of likely locations ordered from most to least likely. When a call arrives for a user, it is paged sequentially in the order indicated by the list. If a user moves between LAs in the list, a location update is not required. When a user leaves the list, it will be required to manually register with the system. The paper does not discuss the algorithm for updating the list. The benefits achieved by the profile-based strategy are partially offset by the costs associated with updating the list. The cost of list maintenance is the cost associated with processing the user's billing records to form the list; transferring the list to the location where it is stored, i.e., the switch that will conduct the search and notifying the mobile of any additions and deletions to the list. In addition, the authors assumed that for a particular user, all the location areas in the list are under the control of the same switch, but in a real system, the search list may span multiple switches. The authors also assumed that all the cells are identical, and all location areas have the same number of cells. In practical systems, these assumptions are not true; each system will have to be evaluated independently.

Considering the above assumptions, the authors showed that the location update traffic is reduced on both the radio and fixed networks links, the paging traffic is increased on the radio link, and the paging traffic remains the same in the fixed network when all of the location areas are in the same Mobile Switching Center (MSC) coverage area and the list is maintained local to the MSC. These savings are realized at the expense of an increase in call setup delay for mobile terminated calls.

Ozan K. Tonguz et al:

In [3], two different paging techniques namely, simultaneous and sequential paging are introduced and are compared in terms of relative paging delay and paging traffic. The authors assumed that the LA is composed of fixed number of square-shaped cells, although in reality, these assumptions are not true; each system will have to be evaluated independently. It is also assumed that the call-arrival rate to each cell is homogeneous which is again not true since different cells have different call arrival rates depending on the users residing in.

One of the main advantages of the analytical framework presented in this paper is the closed-form expression it provides for the cost of location management in terms of the call-arrival rate, boundary crossing rate, the size of the LA, the maximum acceptable average paging delay, the total number of cells controlled by the MSC, etc., thereby allowing valuable insight into the impact of each of these system parameters on the overall cost of location management.

Y-Bing Lin:

Paper [4] studies a special case of AS called the Two-Location Algorithm (TLA). It analyzes the Signaling System no. 7 (SS7) traffic cost of registration and call delivery for a basic scheme based on IS-41 and the TLA. An analytical approach assuming Poisson call arrivals and general roaming times in the LA was used to determine a probability $\alpha(K)$, the probability that a mobile user moves across K LAs between two phone calls. The evaluation of the algorithm uses different LA residence time distribution for different terminals and plots normalized cost functions, comparing the proposed scheme with IS-41 scheme.

The main advantage of this method is the reduction of Location Updates (LUs) when a mobile goes back and forth between two LAs. The algorithm shows that TLA can outperform IS-41 if the user mobility is higher than the call frequency or the user tends to move back to the previously visited location areas. But the performance of the TLA is significantly lower than IS-41 when the variance of the LA residence times is small.

Dynamic LAs:

Dynamic schemes allow on-line adjustments based on the characteristics of each individual MT. For example, when the distance-based location update scheme is used, a different distance threshold can be assigned to each MT based on its mobility and call arrival pattern. However, some of these schemes require information, such as the distance between cells, that is not generally available to the MTs. Besides, the operation of dynamic schemes may require significant computing power.

Anna Hac and Xian Zhou:

In the evaluation of [5], the metric used is total cost in terms of the subscriber's call to mobility ratio (incoming calls to location updates). Total cost is the sum of the total searching and updating costs. The simulation results compare the total cost for the three algorithms: always update, always search and the proposed one, for a number of different call to mobility ratios.

The proposed algorithm does not define the method to gather information regarding the frequency of a MT entering the cell. Another problem is the selection of reporting cells. The problem of selecting reporting centers is based on the volume of control traffic about the mobility of users and the frequency with which they are located. If these rates change dramatically, the reporting centers have to be reselected. It is suggested that in case of reselection of reporting centers, a dynamic allocation of the reporting centers is done, but no mention is made of how to do this dynamic allocation.

The main advantage of this scheme is the reduction of total cost in finding the user and the reduction in the consumption of the wireless channels. The algorithm to solve the reporting centers problem is well defined and has the potential to simultaneously reduce the radio link bandwidth usage and fixed network signaling.

L. R. Hu and S.S.Rappaport:

The analysis in [6] depends on the dwell time, defined as the time the mobile terminal stays within the LA of a given registration level. The system uses multiple registration levels to track the locations of MT. Based on the dwell time of an MT in the previous LA, an MT is switched to a suitable level for location management. Mobile terminals tracked by the proposed scheme are seen to generate signaling traffic at a very stable rate over a broad class of terminal mobility. This feature prevents the system from excessive signaling traffic due to high mobility users and can be used to set the required long-term total signaling traffic.

In the proposed scheme, there are no fixed LA boundaries and, for each LA boundary crossings, the mobile terminal is always placed at the center of a new LA. Consequently, three major benefits are obtained:

1. The location-updating rate is greatly reduced because a suitable LA is provided for each mobile, and the mobile terminal is always placed at the center of a new LA.

2. The location updating signaling traffic is evenly distributed because mobile terminals roaming in the system will cross the LA boundary randomly at different cells.

3. The zigzag effect (frequent updates) is eliminated. Since the mobile terminal is always placed at the center of a new LA on each location updating, it is unlikely to cross a LA boundary shortly after an updating even if its path is very irregular.

The proposed strategy considers homogeneous systems in terms of LA radius. In general the cells may have irregular shapes and different sizes. Extra computational complexity is required to calculate the dwell time.

Thomas Kunz and John Scourias:

The proposal in [7] attempts to utilize the mobility history of the subscriber to dynamically create location areas for individual subscribers, and to dynamically determine the most probable paging area. The mobility model used to evaluate the proposed algorithm is based on activity theory borrowed from related work in traffic engineering and social science, and using raw data from regional planning travel surveys.

In providing the location management cost savings, the dynamic algorithm requires some additional logic and memory on the mobile station and network (base station and/or switch). In terms of computation, the proposed algorithm requires the mobile station to record the average time spent in each cell, and the number of transitions from one cell to another. In addition, the comparison process when entering a new cell is more involved, requiring the comparison of a cell identifier against a list of cell identifiers (the personal location area), as opposed to the comparison of a Location Area Identifier (LAI) against a broadcast LAI. The number of comparisons however is on the order of ten or twenty.

Another additional cost of the dynamic algorithm is the transmission of the user profile to the network. Only the information related to paging needs to be transferred. This transfer would generally be done during a location update, making use of the existing signaling channel.

Location Updating With Alternative Triggers:

One of the advantage of non-LA-oriented methods is that the location updates can be spread over virtually all the cells of the network, resulting in a more evenly distributed signaling load as compared to LA-oriented proposals, which would tend to cause a higher (location updating) signaling load mainly on the LA border cells.

Time-Based:

This strategy is simple and each mobile can update its location according to its local clock. The advantage of this strategy is that the frequency of executing location update process can be controllable regardless of movement of each subscriber, and update cost does not jump compared with other strategies. Its drawback is its resource consumption, which is user-independent and can be unnecessary if the user does not move for several hours. A MT that returns to its starting location (the location where it had its last interaction with the network) or a stationary MT will update its location while the network already knows this location.

Movement-Based:

One disadvantage of this strategy is that when a MT moves into and out of a particular cell repeatedly (e.g., a mobile located at the boundary of two cells), it has to initiate an update for every M boundary crossings although no location update is necessary. Extra logic and memory is required to be aware when the MT cross boundaries between cells.

Distance-Based:

The implementation of the distance-based strategy is more difficult than the other two since the mobile need information about the topology of the cellular network to derive the distance information.

Interrogation-Based Proposals:

Centralized Database Architecture:

The centralized database structure that records all movements of MT in a central database has a simple algorithm to locate the MT. However, its implementation is impractical because the centralized database cannot support such a large number of MTs in a global system. With this scheme all users have to access only one central database in all situations even if both the caller and the callee are in the same location area, and thus it is impractical to serve all users in real time using a single database, and it is impossible to connect a call in case that problems occur in an HLR. Moreover, the cost (i.e., number of operations in the database) of an update for a relocation is fixed, and does not depend on the distance between the previous and current locations. For example, a movement within a city and a movement to a new continent are treated the same, yet many movements are likely to take place within a city, while movements to another continent are rare. Due to the dominating disadvantages of centralized database architecture, most of today's proposals are based on either the distributed or the hybrid database architecture.

Distributed Database Architecture:

In Distributed Database Architecture, location information of the MTs is distributed among a large number of location databases, organized as tree structure. The hierarchical architecture helps to reduce the total network signaling traffic and travelling distance when most calls received by the users and most registration area crossings are geographically localized. However, a large number of location databases have to be accessed during location registration and call delivery, which increases the complexity of the management procedure. For example, in a distributed algorithm, all related layer nodes have to execute complex logic depending upon the received message type whenever an update request or connection request is received. So the main drawbacks of this architecture are the cost of database system acquisition, implementation, and management. Moreover, the performance of this architecture strongly depends on the design of the tree structure and mobility characteristics of the users.

Jan Jannink, Derek Lam, Narayanan Sivakumar, Jennifer Widom and Donald C.Cox:

The analysis in [11] depends on the Local Call to Mobility Ratio (LCMR) defined as the number of calls from a particular switch to a particular subscriber, relative to the number of moves of the subscriber in a given time period. This ratio is computed to decide whether to replicate the user profile at a set of databases or not such that the benefit of local lookups at these databases outweighs the update cost due to the user's mobility. From the simulations it is showed that more than 90% of the calls can be serviced by local database lookups. The penalty paid for the latency improvement is an increase in bandwidth requirement, and near doubling of memory requirements and update rates.

One of the desirable feature of [11] is that instead of assuming a nonrealistic mobility model that incompletely characterizes human calling and mobility patterns, the authors used a realistic call and movement model of the subscribers to simulate accurately the performance of the LMTs in a close approximation of a real geographical topology.

The algorithm for selecting databases where profile replicas are most advantageous to be placed is well defined. With the advances in fiber optic technology and the dropping of storage prices, the flexibility of life-long numbering with fast lookups will become very important.

P. Krishna et al:

In the evaluation of [14], the metric used is total cost in terms of the subscriber's call to mobility ratio. Total cost is the sum of average update cost, average search cost, and the average search-update cost. The cost of a location update and search is characterized by the number of messages sent, size of messages and the distance the messages need to travel. Simulations were carried out to evaluate the performance of various location management strategies.

The Logical Network Architecture (LNA) considered in the proposal is a hierarchical tree structure consisting of BSs and MSCs which maintains the information of the MT residing underneath them. The paper does not discuss how to gather information regarding the mobility behavior of the users. It is assumed that, initially, the location information of the MTs is stored in the corresponding MSCs i.e., each MSC should have the correct location information for all the MTs residing in the cells in its subtree. It is also assumed that the message delays in case of searching the MT are negligible since message delays are small compared to the time between calls and moves.

Considering the above assumptions, the authors showed that performing search-updates significantly reduced the search and total costs without significantly increasing the cost of updates. For the logical network architecture assumed, it is shown that the combination of Lazy Updates (LUs) and Path Compression Update (PCU), namely LU-PC performs better than the other strategies for most call-mobility patterns.

Hybrid Database Architecture:

The hybrid database structure consists of HLRs, which store the permanent MT information within any network and VLRs in charge of roaming MTs in one or more location areas. This can lower the congestion probability in the HLR somewhat but has overheads in multiple database operation because the VLR has to obtain MTs information from the old VLR and delete it in the old VLR. Also, the access delay of VLR is not negligible because the number of MTs registered in a VLR varies over time and all contents in the VLR have to be searched in order to query a MT's information.

***Shivakumar and Widom*:**

According to the scheme proposed in [15], user profiles are replicated at selected locations for faster location lookup. The replication decision is made by a centralized system that must collect the mobility and calling parameters of the whole user population from time to time. This may not be feasible in current PCS networks because of the large number of PCS network providers involved. Besides, generating and distributing the replication decision for a large user population is a computationally intensive and time-consuming process that may consume a significant amount of network bandwidth.

The algorithm is analytically compared to [24] and another similar approach, based on the network query size, and the location lookup delay. An actual mobility model is not used to obtain any quantitative results, but rather variables are used to represent the number of location updates and the number of incoming calls from a particular zone, for generic subscribers.

Chih-Lin, Gregory P.Pollini and Richard D. Gitlin:

In the evaluation of [16], the metrics used is the radio link costs, fixed network signaling load, and call setup delay. Conventional approaches like GSM reserve voice trunks prior to knowing whether a call is successful or not. In Reverse Virtual Call setup, voice trunks are not reserved until the MT is found and willing to accept the call. Alternatively, the called party may choose to have the call routed to a voice mailbox or simply return the call via a wired phone, if one is nearby. Thus in addition to saving the unnecessary trunk allocation, there is a saving in signaling messages. The call setup delay is shown to reduced by about 50 percent when using RVC scheme.

The proposed RVC scheme is compared with the GSM scheme and it is shown that the RVC call setup delay is less than the GSM call setup delay. One of the features of the RVC strategy is that it can function within the existing cellular paging network like GSM. The page is sent as usual and may be acknowledged. However the originating switch does not setup the call, but instead, waits for a call to arrive from the called party, and then bridges the two connection. Only a few new fixed network SS7 signaling messages are needed to implement this protocol, which are mentioned in the algorithm.

Joseph S. M. Ho and Ian F. Akyildiz:

In [17], the authors introduced a location management scheme based on a three-level database architecture. The analysis depends on the Local Call to Mobility Ratio (LCMR). The signaling and database access overhead for call delivery can be reduced if a remote pointer exists at the DR of the calling MT. However, additional processing is required after each movement to update the remote pointers. Similarly, the signaling and database access overhead for location registration can be reduced if location change is reported to the HLR only after inter-DA movements. However, this will increase the database access overhead for call delivery, as an additional DR access is necessary to locate a called MT. The paper assumes that the cost for database access and the cost for transmitting signaling messages are available. An enhancement to the proposed location management scheme is introduced which lowers the inter-DA movement probability and thus results in additional reduction in the signaling and database access costs.

One of the features of the proposed algorithm mentioned by the authors is that it can be employed with the existing cellular system like IS-41. The functions of the network elements like HLR, VLR etc remain mostly unchanged while the DRs handle additional processing. The paper gives a comparison of the IS-41 scheme with the proposed one. Based on the cost parameters considered in the analysis, it is shown that the signaling and database access cost is reduced by as much as 70% and 50% respectively, compared to the IS-41 standard. The steps for implementing the algorithm are well defined in the paper.

R. Jain and Yi-Bing Lin:

In [18], a pointer-forwarding scheme with single HLR is proposed and is compared with IS-41 scheme. When an incoming call arrives in a single HLR environment, the forwarding pointers are traced to find the actual location of the MT. Each time a user moves, the user’s HLR must initiate updates to all the replicas of the user’s profile, and the network will have to carry the packets generated by the updates. Hence, the larger the number of replicas, the more work performed by the HLR and the network. Because of the heavy signaling traffic generated by PCS location tracking, the HLR may become a bottleneck. An additional consideration is the higher storage requirement of databases in zones in order to store the replicated profiles.

The PFSHLR scheme is compared to the IS-41 scheme and it is proved that when the call frequency is much lower than the MT move frequency (LCMR), pointer forwarding may outperform IS-41. When the call frequency is higher than the MT moves frequency, pointer forwarding behaves identically to IS-41.

Yi-Bing Lin and Wen-Nung Tsai:

In [19] a pointer forwarding strategy has been proposed to avoid the expensive HLR access each time a mobile moves to a new LA. Since no registration to the HLR is performed when a user moves to the new locations, the cost of updating multiple HLRs is eliminated in PFDHLR. Thus, the advantage of using distributed HLR for pointer forwarding is obvious from the aspect of database delay. However, it is difficult to implement distributed HLRs in IS-41. For MT registration, it may be required to update some or all HLRs. Thus, extra traffic is generated for multiple HLR updates.

The PFDHLR has some potential side effects. The length of the forwarding pointer chain may increase in a distributed HLR environment. Because of the distributed HLRs in several regions, the total incoming calls will be shared by all HLRs. Therefore, the average call arrival rate in each distributed HLR is 1/N times that in PFSHLR, where N is the number of HLRs. The reason why N divides the call arrival rate is due to the assumption that the incoming calls are uniformly shared by distributed HLRs. This decreases LCMR and the expected number of forwarding pointers increases significantly.

Kuen-Liang Sue and Chien-Chao Tseng:

The analysis in [20] is the extension of the proposal in [19]. Similar to PFDHLR, OSPFDHLR also incorporates the concept of distributed HLRs with pointer forwarding. This strategy attempts to migrate the locating chains in a distributed HLRs system when a mobile issues a registration operation so that the forwarding pointer chain does not exceed one.

The proposed strategy provides an upper bound on the location tracking time owing to the fact that the length of any locating path does not exceed one. Furthermore, obsolete entries in local databases VLRs can be reclaimed in this strategy.

One disadvantage of the proposed algorithm is the increase in the data structure of each entry in VLR. This data structure depends on the number of distributed HLRs in the cellular network. The more HLRs, the bigger will be the data structure that will increase the complexity of the system.

Another potential problem is that adjusting the forwarding pointers needs some extra signaling capacity between VLRs. For a mobile which often moves across the LA boundary, but does not frequently receive calls, considerable signaling capacity will be wasted.

Kin K. Leung and Yonatan Levy:

One of the advantages of the proposal in [21] is that call setup time and response times for signaling functions can be reduced. Due to the local availability of the necessary customer data, transmission cost for signaling messages for calls originating from the visited country or region, and possibly signaling resources can be saved. However these savings and the reduced load on the HDB have to be considered together with the cost of additional transmission and database management in the foreign region. Another factor that can degrade the performance of the replicated design is the query-to-update ratio. If the ratio is high (i.e., the customer information is seldom changed), the overhead in updating the replicated records will be minimal. Otherwise (e.g., for a mobile customer moving from location to location), the update overhead can be significant.

3.3 Conclusions

Locating mobile objects in a worldwide system requires a scalable location service. An object can be a telephone or a notebook computer, but also a data object (Web page may move as its owner changes computers), a file or an electronic document (a shared electronic document may travel between its users). Another example is a mobile agent that moves through the network in search of specific resources for its owner.

As mentioned earlier, mobile radio communications raise two major problems. First, we need to improve the signaling and querying efficiency of the network-side database. Second, we need to reduce radio interface signaling in terms of location updating and paging. This thesis will focus on the second issue and compare the performance of several proposals on the basis of total signaling cost, i.e., location updating and paging costs.

Several recent location management strategies have been commented upon in this chapter. They all seem to have their own advantages and disadvantages. Some are relatively easy to implement, some might provide high signaling savings on the radio link, and some seem to be interesting from a theoretical/conceptual point of view. At best, the proposals above were evaluated based on some version of a random mobility model, typically one-dimensional. Clearly, such simplifications for as critical a component of location management as subscriber mobility is not satisfactory. A random mobility model fails to take into account certain aspects of travel such as trip purpose, and so-called attraction zones, where people spend a large proportion of their time. Random mobility models may have provided a rough model for certain early users of cellular telephones, such as travelling salespersons, whose movements were fairly random. However, as wireless technology increasingly becomes a customer item, such mobility patterns will no longer reflect reality. There is a need to analyze the performance of the proposed strategies by using a more realistic user mobility behavior model instead of the random one.

After a careful review of the above-mentioned schemes, several proposals have been selected for comparison against the general goals of location management. The reason for selecting these proposals is that these strategies have the same goal of reducing the signaling traffic due to location management procedures: location updating and paging. The selection consists of those proposals that can minimize signaling, both over radio interface and within the network, to reduce bandwidth usage and increase capacity with least computational complexity. The performance measures that are considered in this theses are the number of location update messages, the number of cells paged, and the total location management cost. One of the desirable feature in the evaluation of these algorithms is that instead of assuming a nonrealistic mobility model that incompletely characterizes human mobility patterns, realistic movement model of the subscribers is used to simulate accurately the performance of the location management techniques in a close approximation of a real geographical topology.

The selected location updating and paging algorithms for performance comparison are as follows:

- *TLA Algorithm [4]*.
- *Reporting Centers Algorithm [5]*.
- *Dynamic Location Management Algorithm [7]*.
- *Distance-Based Algorithm [9]*.
- *Fixed Location Area Strategy (as implemented in Current Cellular Networks) [25]*.

The algorithm in [4] is the extension of Tabbane's proposal in [1] with a difference that two locations are considered to track the location of the mobile user as compared to AS which allows more than two user locations in the profile for each arbitrary-defined time period. The locations for the two most recently visited LAs are stored in the MT. Same two LA addresses are also stored in the HLR in an order depending upon the user's movement. Location update is only performed if the LA address of the new cell is not found in the memory. This may result in less location updating cost as compared to the GSM, especially when the user moves back and forth between two LAs. When a phone call arrives, the two addresses in the HLR are used to find the actual location of the MT. If the MT is located in the first try, then the paging cost for TLA is the same as the GSM algorithm. Otherwise extra penalty incurs in TLA to locate the MT for the second try.

The proposal in [5] addresses the major problem of reducing the use of radio resources for location updating and paging activities, thus minimizing the traffic on the radio link which is considered as the most scarce resource. The algorithm in the paper uses a partition scheme to optimize the size of the location area domain that minimizes the total cost for the entire network rather than for each individual user. After a careful selection of reporting centers (cells with the lowest frequency of mobile users), the MT is allowed to transmit update messages only at these reporting centers, while restricting every search for the mobile user to a small subset of cells. Whenever there is a need to establish communication with a particular mobile user, the user is searched for in all cells that are in the vicinity of the reporting center to which the user last reported. This may reduce the signaling cost in terms of location updating and paging.

Ideally, the mobility tracking scheme used for each user should depend on the user’s call to mobility pattern, so the standard approach, in which all cells in the location area are paged when a call arrives, may be wasteful of wireless resources (especially if the user is not very mobile and gets frequent calls of short duration). In order to conserve these resources, the network must have the capability to selectively page to a group of cells in which the probability of finding the user is maximum. On the other hand, keeping closer track of the user’s location involves additional expenditures since the MT uses power for listening to the beacons broadcast by the radio channel (to detect changes in his location) and for alerting the network of a location change. These reflections lead us to include the Distance-Based Strategy [9] for further evaluation. By using this strategy the system knows that the user is within distance D from the cell where he has last reported. This may result in less paging signaling in case of a call arrival. Also the user only has to send a location update signal when he crosses the threshold distance D from the cell where he last updated his location. This will result in the signaling reduction for location updating.

The location management strategies in [7] and [25] are selected for comparison purposes. The algorithm in [7] is based on a dynamic selection of cells to form a LA each time a user moves to a new cell and is claimed to be better than the GSM algorithm. The GSM approach is the digital cellular standard currently under use in Europe, and is based on a static set of LAs. Due to the diverse nature, both of these algorithms will contribute in the comparison of the above mentioned proposals.

Chapter 4

Mobility Models and Experimental Setup

Location management has received a fair amount of attention recently, as higher capacity cellular systems are being developed. As can be seen in the review of different location management algorithms in Chapter 3, most authors have used simple random mobility models to evaluate the performance of their proposals. To reflect the day to day movements of actual mobile users, a mobility model may be needed which is more realistic than a random mobility model. This chapter presents a realistic mobility model designed to simulate realistic mobility behavior patterns of the mobile users. A random mobility model is also presented as most of the previous work is based on simplistic models that incompletely characterize human mobility patterns. The two mobility models will be used for the performance evaluation of selected proposals from Chapter 3. The procedures used for computing the optimal parameter values for selected proposals are also mentioned.

A mobility model, in the context of location management, is a model of the daily movements of mobile subscribers, or more precisely, the daily movements of registered mobile stations. Such a model is of paramount importance in mobility management studies. The number of paging and location update messages, required for comparisons of the efficiency of different location management schemes, depends fundamentally on user mobility patterns. There have been several proposals for location management algorithms in the recent past mentioned in Chapter 3, to improve upon the current method of using statically defined location areas described in Chapter 2. In attempting to quantify the performance of their proposals, most authors have used mobility models that are overly simplified and unrealistic, completely random, or dependent upon many assumed ‘typical’ values for key variables.

Different location management proposals react differently to the randomness of users behavior. Random mobility models have been frequently used in the past due to their simplicity, and the fact that early users of cellular communications had typically been very mobile subscribers whose behavior could be relatively well modeled as random. With the tremendous growth of cellular communications, random mobility models are no longer accurate for the majority of subscribers, whose behavior is more routine. A random mobility model would cause very biased results for proposals that attempt to utilize the behavioral aspects of subscriber mobility, for example, the importance of home and workplace as trip origins and destinations.

The mobility model that formed the basis of the simulation study is the same as presented in [7]. A brief overview of the mobility model is given here. Interested readers are referred to [7] for a detailed description of the mobility model.

The mobility model description is followed by a description of our random mobility model. As mentioned in the previous chapter, random mobility models fail to take into account certain observed aspect of travel such as trip purpose, and places where people spend a large proportion of their day, such as home, office, etc. The selected proposals will also be simulated under a random mobility model and a detailed comparison is done for the selected proposals under both mobility models. A description of selected algorithms and their implementation under both mobility models follows next.

4.1 Realistic Mobility Model

The goal of the mobility model is to provide, at the individual subscriber level, a set of paths traversed on a daily basis. Using this information, location management algorithms can be compared according to the relative number of location updates and paging messages they generate. The mobility model used here is the activity-based mobility model [7]. The input data for this mobility model was derived from the trip survey conducted by the Regional Municipality of Waterloo in 1987. The driving parameter in the model is the concept of activity, which is equivalent to trip purpose. Each activity has an associated time of the day, duration, and location. An activity is selected based on the previous activity and the current time period. Once the next activity is selected, its duration is determined from chronologically sorted trip survey records, by calculating the difference between the arrival time of one trip record, and the departure time of the following trip record. Finally the location of the activity is selected, based on certain heuristics and type of activity. Since the current location of the subscriber is already known, once the location for the next activity is selected, the intermediate route (in terms of cells crossed) is determined from a geographical lookup table. The subscriber stays in the destination cell for the duration of activity, and the sequence is repeated. As opposed to a random mobility model, where there is no concept of subscribers with different mobility patterns, four different user groups are defined with different mobility behavior. As the model will be used to evaluate location management algorithms, a period of several days will be simulated to allow the system to reach steady state. The final output from the model will therefore be a list of cells traversed by an individual subscriber over a period of several days, together with the amount of time spent in each cell.

4.2 Random Mobility Model

In a random mobility model, a user moves to a neighboring cell with random and equal probability. Two cells are called neighboring cells if a mobile user can move from one of them to another, without crossing any other cell. As opposed to the realistic mobility model where four categories of person type were defined, the analysis presented for the random mobility model in the sequel will refer to a particular user (arbitrarily chosen). According to the cell layout considered in this study, a cell can have at the most six neighboring cells. In a realistic mobility model a number of cells may be crossed between the origin and destination of a trip; the subscriber will spend some amount of time at the destination, which will likely be longer than the time spent in the intermediate cells. In a random mobility model there is no concept of destination cell, thus every next neighboring cell is selected with equal probability. The duration for which a user stays in a cell is selected randomly (ranging from 0 – 8 hours at the most).

4.3 Simulation Procedure

A discrete event simulation schedules events in a queue to occur at certain times based on a global clock, and these events are dequeued in chronological order and acted upon based on the event type. The occurrence of a particular type of event triggers a corresponding function. The simulation is started with an initial trip-generation event, whose parameters are entered by the user. The trip-generation function creates and queues a set of cell-change events. A cell-change function performs any location updating functions that may be triggered when a change of cell occurs. The trip-generation function also creates and queues another trip-generation event.

Call arrivals are not modeled using the typical Poisson arrival process, since the model is not very realistic for individual subscribers receiving calls. A subscriber is much more likely to receive a call at 2 PM than at 2 AM, a fact not accounted for by a Poisson model. A call-arrival function generates and queues call arrivals for the next 24 hours based on a heuristic probability distribution for call arrival time. The call-arrival function also queues an event that will trigger the generation of the next set of call arrivals.

For the realistic mobility model, four independent variables are used as inputs to the simulation. Person type and work location variability parameterizes the mobility aspects, while the number of daily call arrivals affect the location management algorithms. For each combination of the independent variables, a number of repetitions of the simulation are executed, each with a randomly selected set of internal parameters (such as home location).

For the random mobility model, as there is no concept of user groups, home cell etc, the only input is the number of incoming calls. The trip-generation event, cell-change event and call-arrival event are the same for the random mobility model as for the realistic mobility model.

The simulation continues running until the master clock crosses a user-defined threshold, nominally set at 15 simulated days. Output detailing location update and paging messages generated by the different location management algorithms is logged to external files, which are analyzed by separate software.

4.4 Goals and Constraints of the Algorithms

In Chapter 3, different proposals were compared against the general goals of location management. These same goals form the basic requirements for the selection of the selected algorithms. Certain trade-offs are necessary since many of the goals are contradictory in practice. Priority was given to minimizing the bandwidth required for location updates and paging. Providing adequate quality of service in terms of call setup delay (due to paging delay), as well as preserving battery life, were also given high priority in the selection procedure. The description of the selected algorithms along with the experimental setup of both models for the algorithms follows.

Anna Hac and Xian Zhou

The proposal by Anna Hac [5] is based on minimizing the total signaling cost for the entire network rather than for each individual user or user group. Since each user group has its own characteristics and mobility patterns, each cell has a different frequency with which the mobile users enter that cell. The frequency with which the mobile users enter a cell corresponds to the updating cost for the cell. The higher the frequency, the higher the updating cost. The problem is how to choose the cells as reporting centers to minimize the total location management cost, i.e., both updating and searching cost. The concept of total location management cost is a way to observe the overall effects and behavior of different algorithms and is defined in detail in the next chapter. Here it can be thought of as linear combination of the number of location updating and paging messages. To cope with the problem of reporting centers selection, the following approach is used.

To calculate the frequency with which the mobile users enter the cell in the realistic mobility model, simulations were run for different sets of input parameters and the output results from the simulations were used to evaluate the movement of simulated subscribers on a per cell basis. The movements and messages generated by each subscriber along with the path taken were logged for 15 simulated days. For each combination of the two independent variables (person type, fixed work location) that parameterize the mobility aspects, one hundred repetitions were run each with a randomly selected set of internal parameters (initial start time, location of home, work, and school). Output detailing the path traversed by the subscribers along with the frequency of subscriber entering in each cell is logged to an external file. The frequencies with which users enter each cell were added together, which gives the total frequency at which all mobile users enter the cell. The cells with the highest frequency were selected as non-reporting centers by making use of the reporting center selection algorithm mentioned in [5].

Simulations were run with different values of vicinity Z for different sets of input parameters and the output is logged to an external file for total location management cost calculation. For each set of controlled variables (person type, fixed work location, maximal vicinity size *(Z=3,4,5,6,7)*, and number of daily call arrivals), ten repetitions were run, each with a different set of internal parameters (initial start time, location of home, work, and school). The number of location updates and pages were tabulated, for each different set of controlled variables and the total cost was calculated. From *Figure 4.1*, it can be seen that the costs are relatively independent of Z as different values of Z give very similar performance. Thus the one giving the minimum location management cost, $Z = 5$, is selected as the largest vicinity size for final simulations.

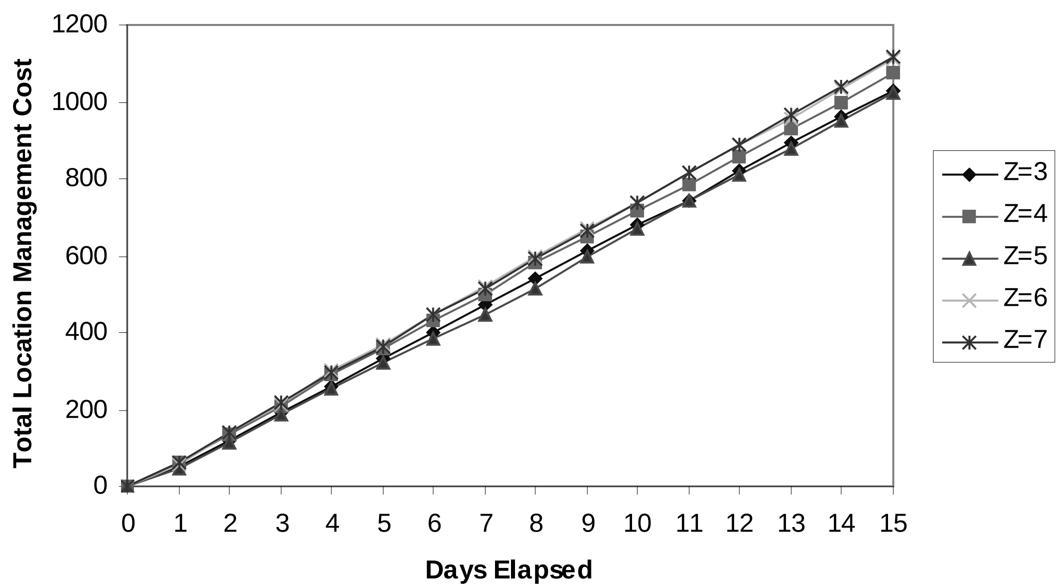


Figure 4.1: Total Location Management Cost under the Realistic Mobility Model for Z = 3,4,5,6,7

In case of our random mobility model, as there is no concept of user groups and destination cells, the concept of calculating the frequency of mobile users in the cells differs from the realistic mobility model. As in the realistic mobility model, a user traversed the shortest path in travelling from cell *a* to cell *b* and the number of times he crosses each cell is counted. In the random mobility model, the destination cell can only be one of the neighboring cells of the cell in which the subscriber currently resides in.

The frequency with which mobile users enter the cells depends upon the number of times the cell is selected as the next cell. Thus for calculating the frequency, one thousand repetitions were simulated five times. Out of five thousand simulation results, the in-cell frequency was calculated.

After the frequency calculation, the algorithm in [5] is used again for reporting center selection. Simulations were run for different values of vicinity *(Z = 1,2,3,4,5,6,7)*. After a careful review of all the results, Z = 2 was selected as the maximum vicinity size for the random mobility model *(Figure 4.2)*. The cost difference between realistic and random mobility model *(Figure 4.1 vs. Figure 4.2)* results is due to the fact that under a realistic mobility model a user has to cross several cells to reach to the destination. Under the random mobility model, the same cells can be reselected each time as the destination cell so the user can keep moving between two cells which can be in the vicinity of the same reporting cell.

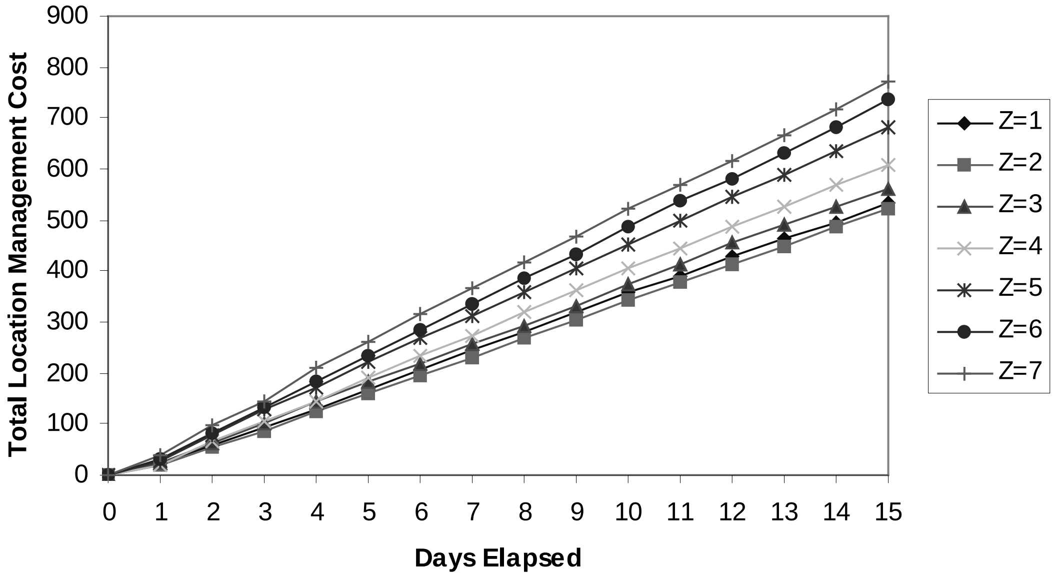


Figure 4.2: Total Location Management Cost under the Random Mobility Model for Z = 1,2,3,4,5,6,7

Fixed Location Area Strategy (as implemented in Current Cellular Networks):

In the Fixed Location Area strategy [25], a *N*-cell system is partitioned into static location areas. The mapping from cell to location area is fixed. Location areas can have different numbers of cells, depending on the number of mobile users and geographical proximity. A mobile initiates a location update when it moves from one LA to another. In [7] four different locations area groups were compared using the realistic mobility model. After a careful review of the results in [7], the location area mapping giving the best results is selected. The selected scheme partitions the 45 radio cells into total of eleven location areas, with three to five cells in each.

<u>Y-Bing Lin:</u>

The LA mapping for the TLA scheme [4] is the same as that of the fixed location area strategy. TLA is implemented by modifying the fixed location area strategy. The modifications were done in two places. First, one more field is added to the subscriber's location record in the main database to hold the address of the second location area. Second, a few changes were made to the MT record to exercise TLA.

<u>Distance-Based Algorithm:</u>

In the Distance-Based Method [9], a MT performs a location update when its distance from the cell where it performed the last location update exceeds a predefined value T (this distance value is referred to as the distance threshold). To select the threshold distance for the realistic mobility model, simulations were run with different threshold T values and the output is logged to an external file for total cost calculation. For each set of controlled variables (person type, fixed work location, threshold distance *(T = 1,2,3,4,5,8,10)*, and number of daily call arrivals), ten repetitions were run, each with a different set of internal parameters (initial start time, location of home, work, and school).

The number of location updates and pages were tabulated for each set of controlled variables. From Figure 4.3, it can be seen that both T = 2 and T = 3 have very similar performance, with T = 3 performing slightly better. So a threshold value of T = 3 is selected that gives the minimum location management cost.

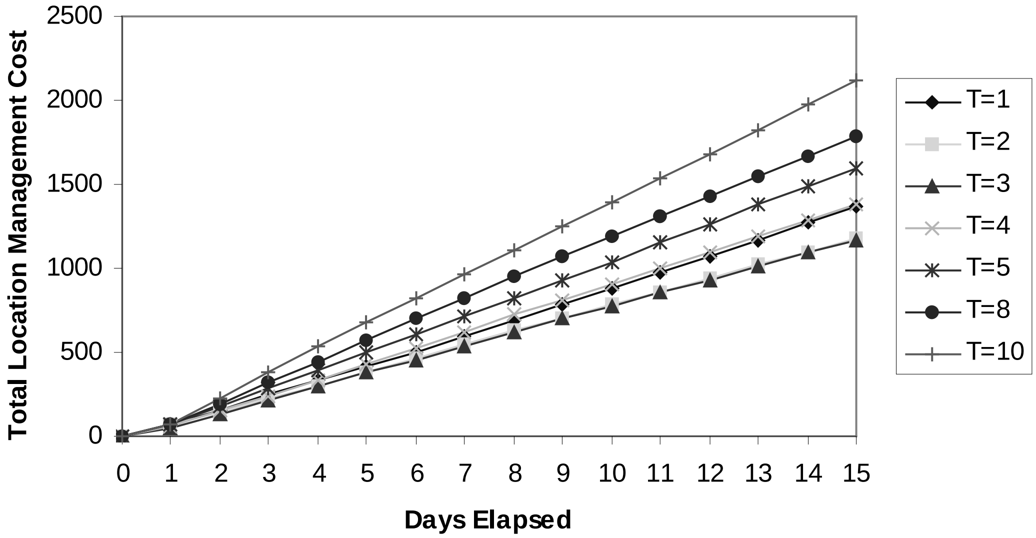


Figure 4.3: Total Location Management Cost under the Realistic Mobility Model for T = 1,2,3,4,5,6,7

For the random mobility model, fifteen simulations were run with the different values of threshold distance *(T = 1,2,3,4,5,8,10)* and the total location management cost is calculated. The results show that T = 1 and T = 2 results in the minimum values of the total location management cost showing very similar results *(Figure 4.4)*. So T = 2 is selected for final simulations.

Dynamic Location Management Algorithm:

The Dynamic Location Management Algorithm [7] attempts to utilize the mobility history of the subscriber to dynamically create location areas for individual subscribers, and to dynamically determine the most probable paging area. The system starts with an empty history record and develops a mobility history as the subscribers traversed between cells. The mobility history consists of the number of transitions the subscriber has made

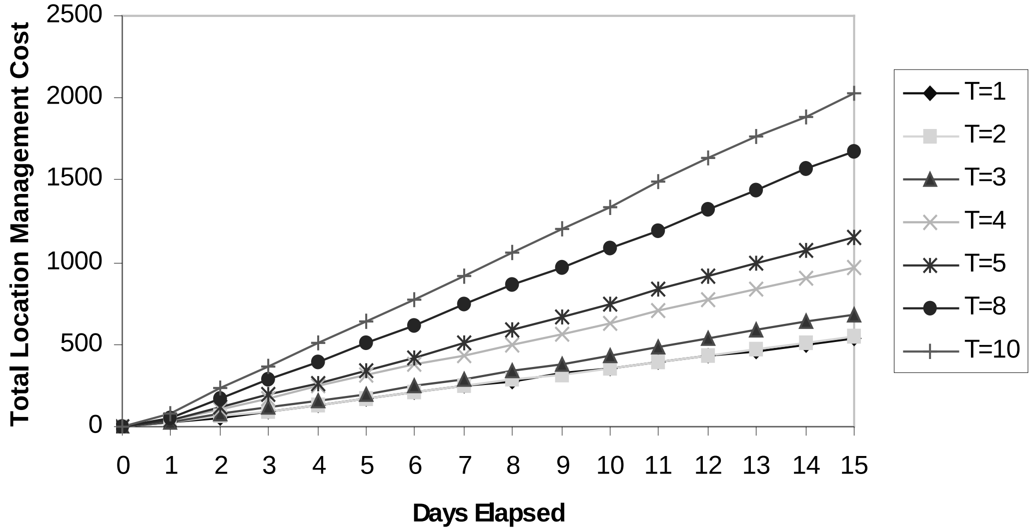


Figure 4.4: Total Location Management Cost under the Random Mobility Model for T = 1,2,3,4,5,6,7

from cell to cell, and the average duration of visits to each cell. A set of cells is selected for the new location area, such that the probability of the subscriber subsequently roaming within those cells is sufficiently high. For an incoming call, the subset of the location area where the subscriber is most likely to be found is paged first, based on the subscriber's average visit duration in each cell.

4.5 Summary

An efficient mobile location tracking strategy should minimize the combined cost of paging and location update. This combined cost depends on a mobile's mobility as well as its incoming call arrival rate. Two mobility models, realistic mobility model and random mobility model, are presented that are used to simulate the mobility patterns of the mobile users. The realistic mobility model is implemented to simulate the daily movements of mobile subscribers (in terms of a sequence of radio cells), incorporating realistic, individualized activity patterns and geographical focal points. The model is used as a tool for comparison of the selected proposals. A random mobility model is also implemented as most authors have used some version of a random mobility model for performance evaluation of their proposal. Simulations were performed to analyze the key performance parameters of selected algorithms under both mobility models. The two mobility models will be used for the performance evaluation of the selected location management algorithms.

Chapter 5

Performance Analysis

The selected location management algorithms are compared in this chapter, primarily in terms of radio bandwidth efficiency. As discussed in the previous chapters, one of the most important goals of location management algorithms is efficient utilization of radio spectrum, which is a fixed and limited resource. Maintaining the number of location updating and paging messages that must be transmitted across the radio interface can satisfy this requirement. Simulations are performed to analyze the performance of the selected location management strategies. Results will be obtained for both realistic and random mobility models. Statistical analysis is performed to compare the performance of selected algorithms under both mobility models. The need for statistical output analysis is based on the observation that the output data from the simulation exhibits random variability when random number generators are used to produce the values of the input variables; that is, two different streams of random numbers will produce two sets of outputs which (probably) will differ.

5.1 Description of Data and Procedures

The analysis involved performance evaluation of location management algorithms under both realistic and random mobility models. The original output from the mobility models provided details about paging and location update messages generated over time by a simulated subscriber, for each selected algorithm. For the realistic mobility model, the movements and messages generated by each subscriber was logged for 15 simulated days, an interval assumed at the outset to be long enough for the model to reach steady state. For each set of controlled variables (person type, fixed work location, and number of daily call arrivals), thirty repetitions were run, each with a different set of internal parameters (initial start time, location of home, work, and school). The number of location updates and pages were tabulated, per simulated day, for each different algorithm and for each different set of controlled variables. Overall, there were 120 parameter combinations (4 person types x 2 fixed work location states x 3 daily call arrival levels x 5 algorithms) with 30 repetitions for each combination, for a total of 3600 individual runs.

As in the realistic mobility model, the movements and messages generated by a subscriber in the random mobility model were also logged for 15 days. Thirty repetitions were run for a different value of daily call arrivals and a different initial start time. The number of location updates and pages were tabulated, per simulated day, for each different algorithm and for different call arrival rates. Overall, there were 15 parameter combinations (3 daily call arrival levels x 5 algorithms) with 30 repetitions for each combination, for a total of 450 individual runs.

Due to the number of independent and dependent variables and the complexity of the data, statistical analysis on the data was deliberately kept simple. Various means, standard deviations, and confidence intervals at a 99% confidence level are calculated using

Microsoft Excel. Calculation of confidence intervals helped evaluating the performance of the selected algorithms. If the mean cost matrix of one algorithm is better than that of others and their respective confidence intervals do not overlap than, based on the data at hand, there is a strong statistical evidence that this algorithm is better than the other. If there is an overlapping of the confidence intervals, then there is no strong statistical evidence that one system design is better than the other and the two algorithms show very similar performance.

5.2 Comparison of Location Management-Related Signaling

The average number of location update and paging messages generated by the different algorithms were compared. The total location management cost a rather abstract but very significant concept, was also calculated and used to compare the overall performance of the algorithms under both mobility models.

5.2.1 Overall Average Performance

Location Updating:

The graph in *Figure 5.1* shows the cumulative number of location update messages for different algorithms averaged over all relevant variables (person type and fixed work location) for the realistic mobility model. It shows the overall relationship between selected algorithms. The reporting center strategy performed the worse. This is evident from the fact that the reporting centers are fixed based on the reporting center selection

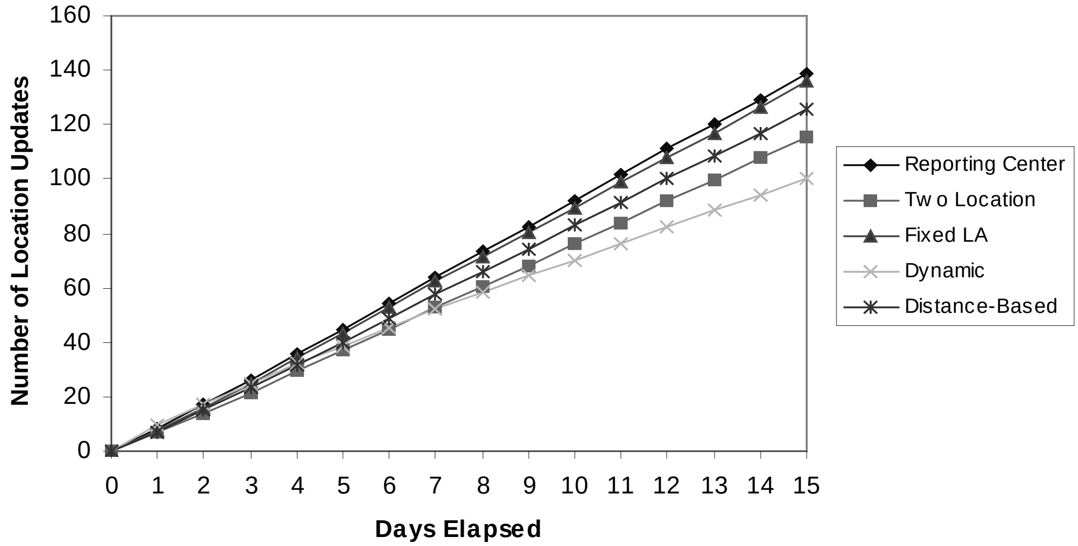


Figure 5.1: Global Average Number of Location Updates versus Elapsed Time under the Realistic Mobility Model

algorithm mentioned in Chapter 4, and not created dynamically based on the mobility behavior of the subscribers. According to the network topology used in this study, some of the reporting centers have a vicinity size of one, thus generating a lot of update messages if the subscriber traverses along a path containing these cells. The fixed location area strategy performs better than the reporting center strategy but worse than distance-based strategy. As expected, the TLA strategy performed better than the fixed location area strategy which shows that the mobile users tend to move back to the previously visited location area.

The dynamic location area strategy gives the best result. The ‘learning curve’ associated with the dynamic algorithm is clearly visible on the graph. As time progresses, the user profile is able to provide better information for cell selection. At first the results of the TLA algorithm were better than the dynamic algorithm, but the dynamic strategy starts performing better after six elapsed days, and the performance gap increased as time increased.

Table 5.1 further demonstrates the results in *Figure 5.1*, showing the means, standard deviations and 99% confidence intervals of all the five selected algorithms for the total number of location updates sent under the realistic mobility model. It is clear from the table that there is strong statistical evidence that dynamic algorithm is better than the rest of the algorithms.

	Mean	**StdDev**	**99% Confidence Interval**
Reporting Center	138.48	44.95	[131.01 – 145.95]
Two-Location	118.76	39.20	[112.24 – 125.28]
Fixed LA	135.88	38.28	[129.51 – 142.25]
Dynamic Algorithm	100.08	26.88	[95.61 – 104.55]
Distance-Based	125.45	34.45	[119.72 – 131.18]

Table 5.1: 99% Confidence Interval Estimation of Location Updates under the Realistic Mobility Model

The graph in *Figure 5.2* shows that the two-location algorithm performed better than the dynamic algorithm under the random mobility model. It was expected, since the randomness of cells visited does not allow useful user profiles to be constructed for the dynamic algorithm. For the two-location strategy, the next selected cell can be in the same location area as the previous cell, or the address of its location area is already present in the main memory as the last visited LA, thus reducing the number of location updates.

The reporting center algorithm performed the worse again for the same reason as explained above. The fixed location area strategy performed better under the random mobility model than under the realistic mobility model. This can be explained by

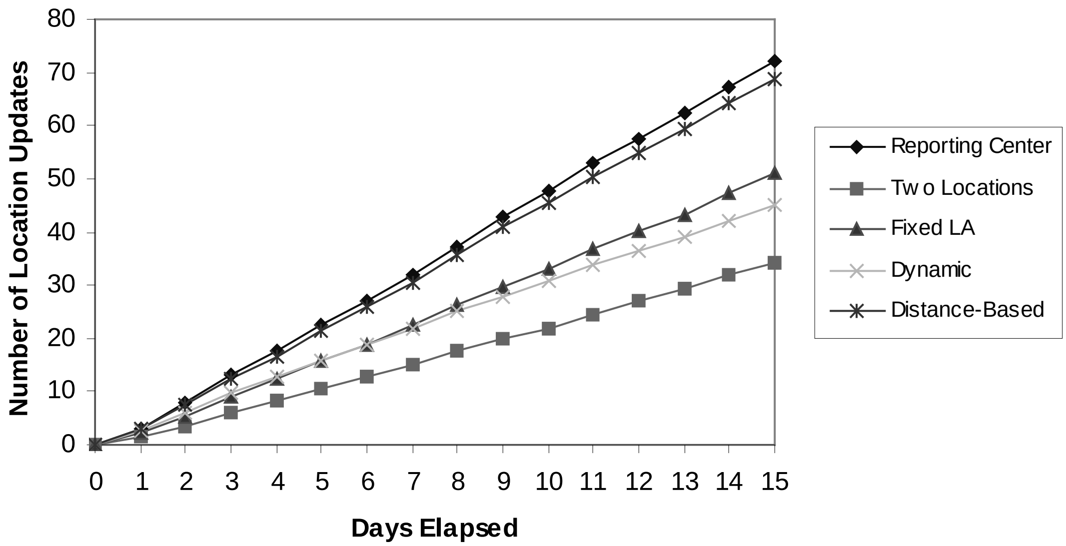


Figure 5.2: Global Average Number of Location Updates versus Elapsed Time under the Random Mobility Model

reasoning similar to the two-location algorithm case.

The number of location updates generated by every algorithm under the random mobility model are also analyzed statistically by calculating the means, standard deviations and 99% confidence intervals *(Table 5.2)*. The results show strong statistical evidence that the two-location algorithm performed extremely well, while the reporting center algorithm performed the worse.

	Mean	StdDev	99% Confidence Interval
Reporting Center	72.2	6.41	[69.19 – 75.21]
Two-Location	34.03	4.77	[31.79 – 36.27]
Fixed LA	51.07	6.63	[47.95 – 54.19]
Dynamic Algorithm	44.97	5.34	[39.63 – 50.31]
Distance-Based	68.7	5.58	[66.07 – 71.33]

Table 5.2: 99% Confidence Interval Estimation of Location Updates under the Random Mobility Model

Paging

The overall average number of cells paged is shown in *Figures 5.3, 5.4, 5.5, and 5.6* for levels of 6 and 12 incoming calls per day, respectively under both mobility models. Under the realistic mobility model, the results scale almost linearly with the different incoming

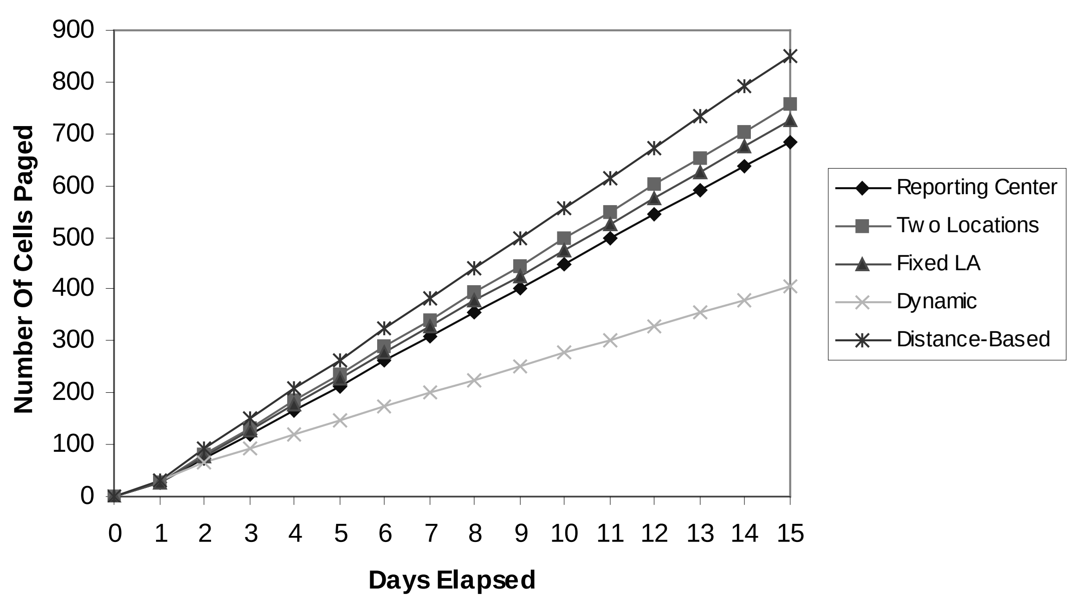


Figure 5.4: Global Average Number of Cells Paged for 12 Incoming Calls versus Elapsed Time under the Realistic Mobility Model

call levels, indicating that the dynamic paging algorithm performs better relative to other algorithms regardless of paging load. During the first few days, while the user profile is being compiled, the performance of the dynamic algorithm was somewhat worse than other selected algorithms, as the reporting center algorithm performed the best.

This gap, however, typically disappeared after 2 elapsed days, after which the performance of the dynamic algorithm is the best. The distance-based algorithm performs the worst for both incoming call levels. This is evident from the fact that if the mobile user is not located in the neighboring cells in the first paging try, then extra penalty has to

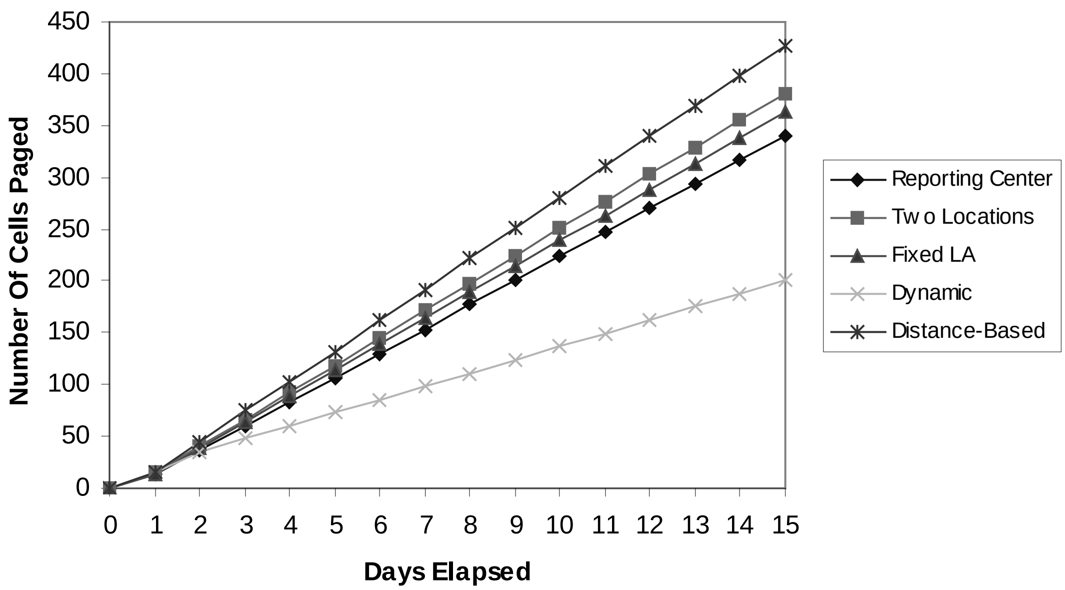


Figure 5.3: Global Average Number of Cells Paged for 6 Incoming Calls versus Elapsed Time under the Realistic Mobility Model

be paid by paging all the cells under the threshold distance T = 3.

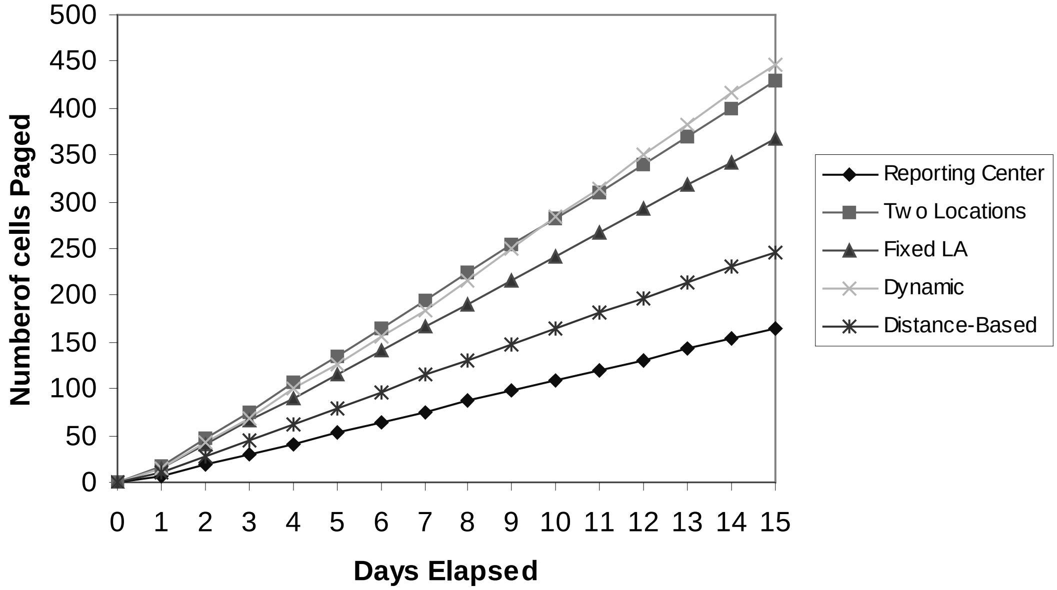


Figure 5.5: Global Average Number of Cells Paged for 6 Incoming Calls versus Elapsed Time under the Random Mobility Model

Under the random mobility model, the reporting center strategy performs the best and the two-location algorithm and dynamic algorithms perform the worst. For the dynamic strategy, it was expected, since the randomness of cells visited does not allow useful user profiles to be constructed and most of the time a large location area (whose maximum size is bound by the algorithm) is selected as paging area. For the two-location algorithm, its poor performance was also expected since if the user is not located in the first try, extra penalty incurs to locate the user for the second try, that is, the paging messages are sent to the second LA in the memory.

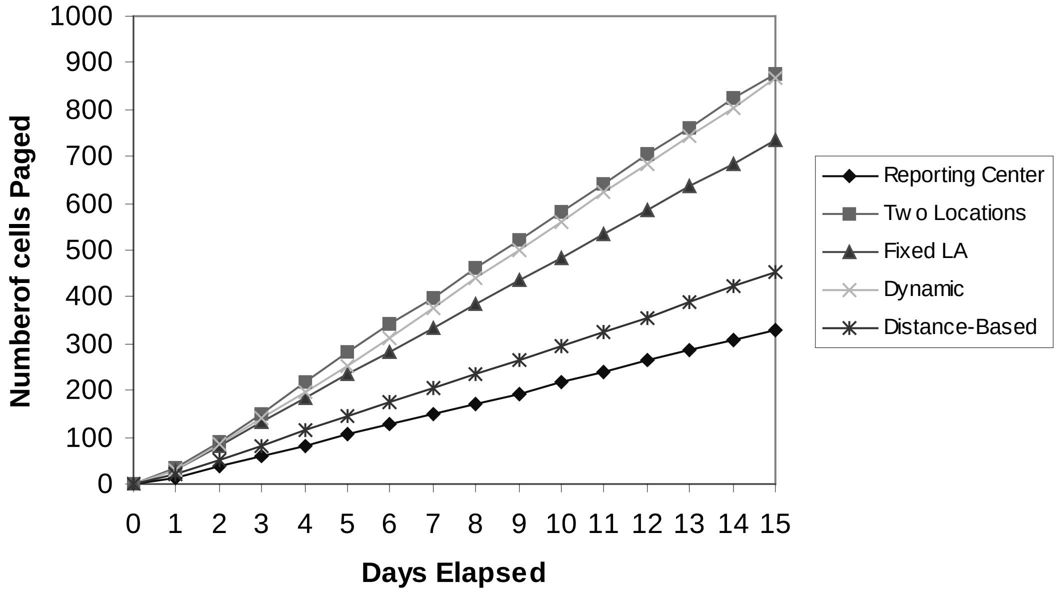


Figure 5.6: Global Average Number of Cells Paged for 12 Incoming Calls versus Elapsed Time under the Random Mobility Model

Despite the poor performance of the reporting center algorithm with respect to the number of location update messages sent under the random mobility model, its paging cost is the minimum for both levels of incoming calls. This was somewhat expected, as the number of cells paged in case of an incoming call are atmost two.

The statistical analysis for the number of cells paged for levels of 6 to 12 incoming calls per day under both mobility models is shown in *Tables 5.3, 5.4, 5.5 and 5.6*. For both levels of incoming calls under the realistic mobility model, there is strong evidence that the dynamic algorithm performs significantly better while the distance-based algorithm performs the worse *(Table: 5.3,5.4)*.

	Mean	StdDev	99% Confidence Interval
Reporting Center	340.39	85.6	[326.16 – 354.62]
Two-Location	381.25	47.42	[373.37 – 389.13]
Fixed LA	363.23	29.22	[358.37 – 368.09]
Dynamic Algorithm	200.48	45.23	[192.96 – 208]
Distance-Based	426.45	91.21	[411.29 – 441.61]

Table 5.3: 99% Confidence Interval Estimation of Number of Cells Paged for 6 Incoming Calls under the Realistic Mobility Model

	Mean	StdDev	99% Confidence Interval
Reporting Center	683.85	149.67	[658.97 – 708.73]
Two-Location	756.6	75.87	[743.98 – 769.22]
Fixed LA	725.16	58.10	[715.5 – 734.82]
Dynamic Algorithm	403.91	91.96	[388.62 – 419.2]
Distance-Based	849.75	197.48	[816.92 – 882.58]

Table 5.4: 99% Confidence Interval Estimation of Number of Cells Paged for 12 Incoming Calls under the Realistic Mobility Model

Under the random mobility model the statistical analysis provides strong evidence that the reporting center algorithm perform better than the other algorithms as its mean number of cells paged is smaller for both levels of 6 *(Table 5.5)* and 12 *(Table 5.6)* incoming calls. Since there is overlapping between the confidence intervals of the dynamic algorithm and the two-location algorithm, there is no significant difference between the results of these two, but they are clearly the worst performers.

	Mean	StdDev	99% Confidence Interval
Reporting Center	165.17	6.79	[161.98 – 168.36]
Two-Location	429.1	26.49	[416.64 – 441.56]
Fixed LA	367.87	11.37	[362.53 – 373.21]
Dynamic Algorithm	446.7	34.6	[430.45 – 462.95]
Distance-Based	246.07	32.5	[230.8 – 261.34]

Table 5.5: 99% Confidence Interval Estimation of Number of Cells Paged for 6 Incoming Calls under the Random Mobility Model

	Mean	StdDev	99% Confidence Interval
Reporting Center	330.17	15.72	[322.78 – 337.56]
Two-Location	878.13	60.94	[849.47 – 906.79]
Fixed LA	734.07	26.76	[721.49 – 746.65]
Dynamic Algorithm	869.2	65.4	[838.45 – 899.95]
Distance-Based	452.5	58.74	[424.88 – 480.12]

Table 5.6: 99% Confidence Interval Estimation of Number of Cells Paged for 12 Incoming Calls under the Random Mobility Model

5.2.2 Total Location Management Cost

The dynamic scheme performs best for location updates and paging under the realistic mobility model. Under the random mobility model, the two-location strategy performs best for location updates, but is nearly worse for pages along with the dynamic strategy, due to trade-off between efficient location updating and paging, while the reporting center strategy performs the best because of small paging area sizes. The total location management cost defined in the Chapter 4 as the linear combination of the number of location updating and paging messages, is a way to observe the overall effects and behavior of different algorithms. It can be defined in several ways, such as the total number of signaling messages crossing a specific interface, or the total size in bytes of all exchanged signaling messages. Some authors simply represent cost as an abstract constant. In this analysis, an intermediate approach has been taken, which is quantitative, yet not limited to a specific technology. The location management cost is given by:

$$Cost = cN_{lu} + N_p$$

N_{lu} is the number of location update messages generated, N_{p} is the number of cells paged as a result of an incoming call, and c is a constant representing the relative cost of updates to pages. Since a location update is a much more expensive procedure in terms of signaling than pages, the two representative values of c are taken as 5 and 10. These values are somewhat arbitrary, but represent the approximate size and number of signaling messages required by a location update compared to a paging message.

Significant differences in the relative cost of different algorithms arise as the number of incoming call increases under the realistic mobility model. For relative cost c = 5, and 6 incoming calls per day *(Figure 5.7)*, the distance-based algorithm, reporting center algorithm and the fixed location area algorithm have very similar performance and incur the highest costs. This is due to the excessive updating and paging costs required by these algorithms.

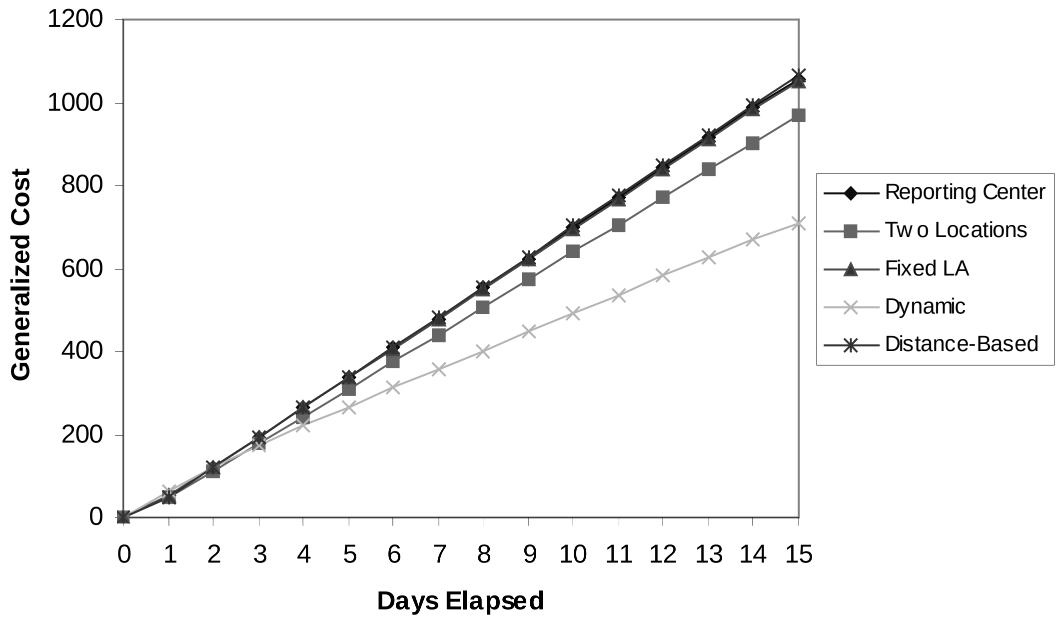


Figure 5.7: Global Total Location Management Cost (c = 5) for 6 Incoming Calls versus Elapsed Time under the Realistic Mobility Model

The dynamic algorithm outperforms all other algorithms after a brief 'learning curve' of about three elapsed days. This was expected since the dynamic algorithm produced the least location update and paging messages. As the number of incoming calls increases *(Figure 5.8)*, the gap between the dynamic algorithm and the other algorithms increases.

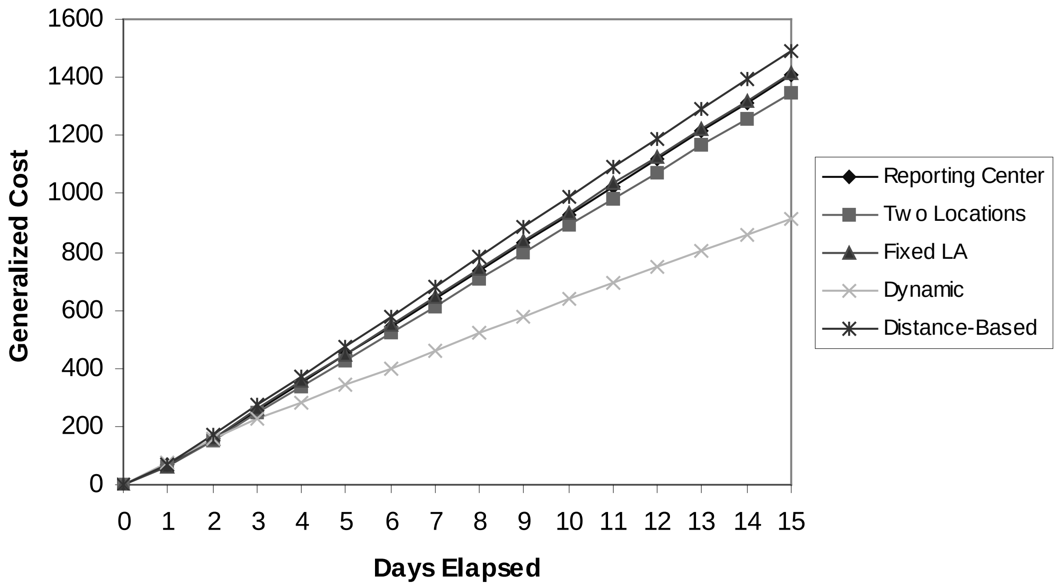


Figure 5.8: Global Total Location Management Cost (c = 5) for 12 Incoming Calls versus Elapsed Time under the Realistic Mobility Model

For a relatively larger cost of updates (c = 10), the worst performers are still the distance-based, reporting centers and fixed location algorithms *(Figures 5.9 and 5.10)* for both 6 and 12 levels of incoming calls. This was not surprising, since updates are relatively more expensive, and these algorithms generate by far the largest number of updates. As before,

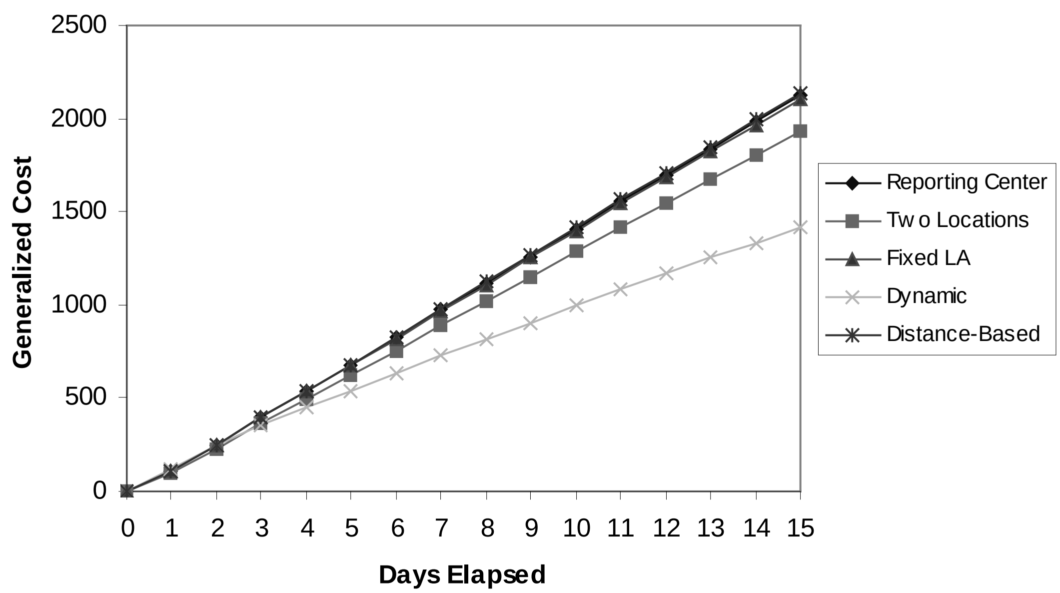


Figure 5.10: Global Total Location Management Cost (c = 10) for 12 Incoming Calls versus Elapsed Time under the Realistic Mobility Model

the two-location algorithm performs better than the other three but worse than the dynamic algorithm. The two-location algorithm with corresponding low rates of updates, starts out the best of all algorithms, but its large paging cost causes total cost to rise significantly with a higher level of incoming calls.

The dynamic algorithm again outperforms the other algorithms, after a ‘learning curve’ of two to three days. The gap between the total cost of the dynamic and the other algorithms

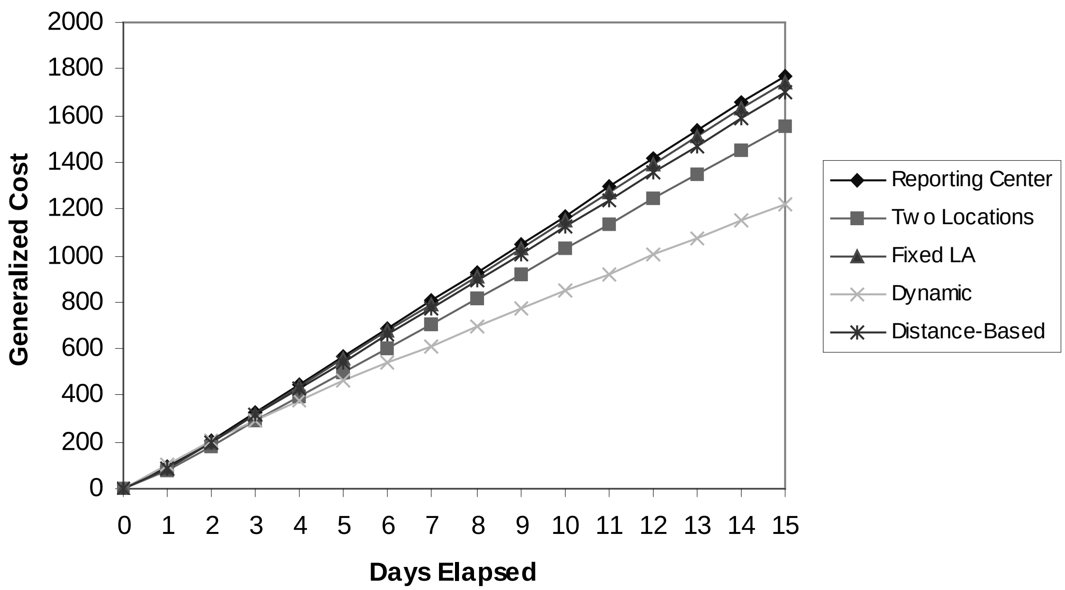


Figure 5.9: Global Total Location Management Cost (c=10) for 6 Incoming Calls versus Elapsed Time under the Realistic Mobility Model

is even greater for c = 10, since the updating efficiency of the dynamic algorithm

becomes even more important.

The statistical analysis of the total location management cost for c = 5 and c = 10, for 6 and 12 incoming calls per day under the realistic mobility model is shown in *Tables 5.7*, *5.8, 5.9 and 5.10.*

	Mean	StdDev	99% Confidence Interval
Reporting Center	1055.91	225.47	[1018.42 – 1093.4]
Two-Location	966.62	182.23	[936.33 – 996.91]
Fixed LA	1051.87	204.39	[1017.89 – 1085.85]
Dynamic Algorithm	710.75	150.75	[685.69 – 735.81]
Distance-Based	1062.75	168.64	[1034.71 – 1090.79]

Table 5.7: 99% Confidence Interval Estimation of Total Location Management Cost for c = 5 and 6 Incoming Calls under the Realistic Mobility Model

	Mean	StdDev	99% Confidence Interval
Reporting Center	1404.42	245.78	[1363.56 – 1445.28]
Two-Location	1344.63	188.15	[1313.35 – 1375.91]
Fixed LA	1412.12	206.77	[1377.75 – 1455.49]
Dynamic Algorithm	911.48	181.44	[881.32 – 941.64]
Distance-Based	1492.85	232.20	[1454.25 – 1531.45]

Table 5.8: 99% Confidence Interval Estimation of Total Location Management Cost for c = 5 and 12 Incoming Calls under the Realistic Mobility Model

	Mean	StdDev	99% Confidence Interval
Reporting Center	1771.43	450.6	[1696.51 – 1846.35]
Two-Location	1551.99	367.84	[1490.83 – 1613.15]
Fixed LA	1740.52	410.32	[1672.3 – 1808.74]
Dynamic Algorithm	1221.02	284.84	[1173.66 – 1268.38]
Distance-Based	1699.04	331.46	[1643.93 – 1754.15]

Table 5.9: 99% Confidence Interval Estimation of Total Location Management Cost for c = 10 and 6 Incoming Calls under the Realistic Mobility Model

	Mean	StdDev	99% Confidence Interval
Reporting Center	2124.98	453.42	[2049.6 – 2200.36]
Two-Location	1932.65	369.45	[1871.23 – 1994.07]
Fixed LA	2099.08	410.39	[2030.85 – 2167.31]
Dynamic Algorithm	1419.04	310.53	[1367.41 – 1470.67]
Distance-Based	2135.96	374.13	[2073.76 – 2198.16]

Table 5.10: 99% Confidence Interval Estimation of Total Location Management Cost for c = 10 and 12 Incoming Calls under the Realistic Mobility Model

It can be seen that there is strong statistical evidence that the dynamic algorithm performs significantly better than other algorithms while reporting center, fixed LA and distance-based strategies perform worst for both levels of incoming calls.

Under the random mobility model, the relative performance of all algorithms is nearly

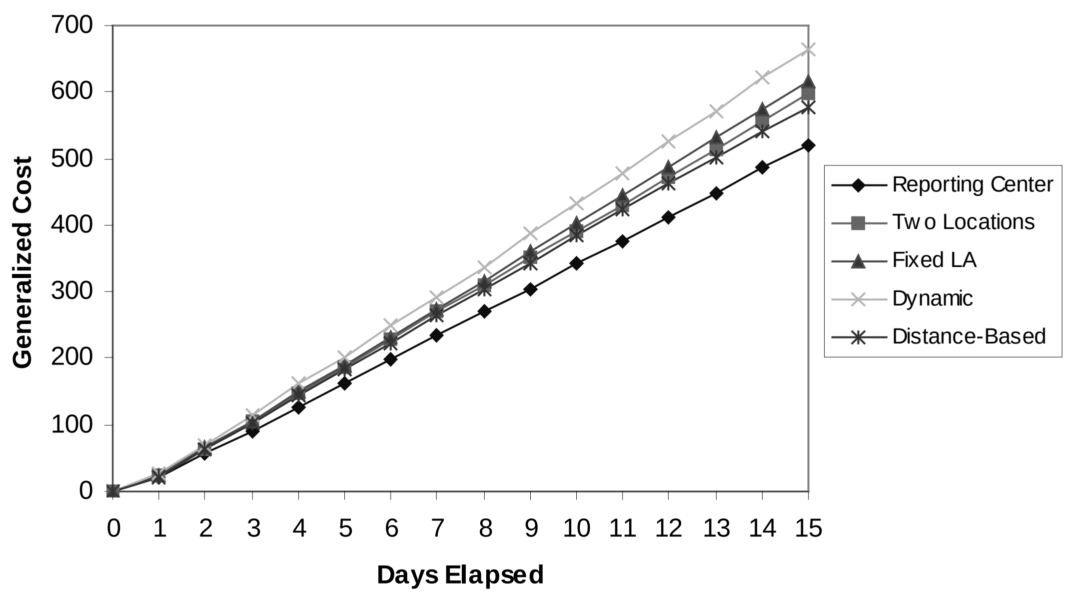
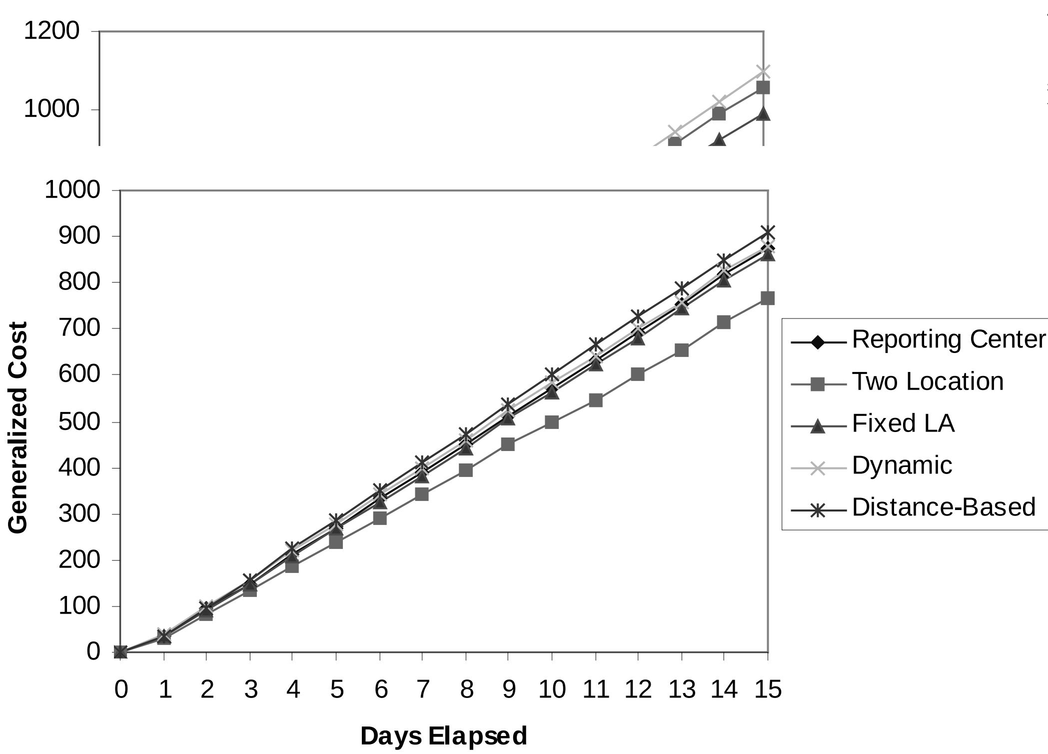


Figure 5.11: Global Total Location Management Cost (c = 5) for 6 Incoming Calls versus Elapsed Time under the Random Mobility Model

opposite to their relative performance under the realistic mobility model. For relative cost c = 5 and 6 incoming calls per day *(Figure 5.11)*, the dynamic algorithm that outperformed all the other algorithms under the realistic mobility model incurs the highest cost. This is due to the excessive paging cost incurred by this algorithm, which is not offset by its inherently low updating cost. This is expected, since the dynamic algorithm requires construction of the user profile that depends on the movement of the mobile user. But in the random mobility model, the randomness of cells visited did not allow useful user profiles to be constructed. The reporting center strategy that performed the worst under the realistic mobility model gives the best results under the random mobility model. This results from the relatively low paging cost, and the gap increases as more incoming calls arrive *(Figure 5.12)*.



than
ivals,
rms
or a
was
tion
cost

Figure 5.13: Global Total Location Management Cost (c = 10) for 6 Incoming Calls versus Elapsed Time under the Random Mobility Model

For relatively more expensive updates (c = 10), the two-location algorithm is the least expensive algorithm for 6 incoming calls *(Figure 5.13)*, but is overtaken by the reporting center strategy as the number of incoming calls increases *(Figure 5.14)*. This is because the two location algorithm incurs low rates of updates and shows the best result for less incoming calls, but its large paging cost causes total cost to rise significantly with higher levels of incoming calls.

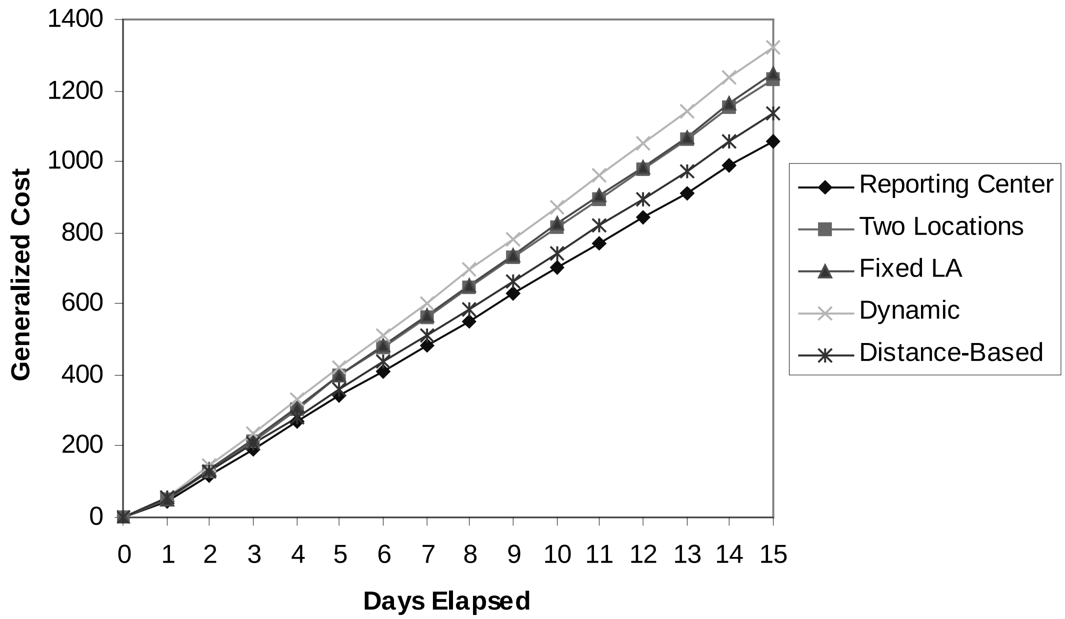


Figure 5.14: Global Total Location Management Cost (c = 10) for 12 Incoming Calls versus Elapsed Time under the Random Mobility Model

The statistical analysis results for c = 5 and c = 10, and 6 and 12 incoming calls per day under the random mobility model are shown in *Tables 5.11, 5.12, 5.13 and 5.14*. It can be seen that for c = 5 and both 6 and 12 incoming calls per day, there is a strong statistical evidence that the reporting center strategy is better than the other algorithms. On the same statistical grounds it is evident that the dynamic algorithm is the worst performer under the random mobility model for the levels of 6 and 12 incoming calls per day and a relatively small update cost multiplier (c = 5).

	Mean	StdDev	99% Confidence Interval
Reporting Center	519	28.81	[505.5 – 532.54]
Two-Location	597.27	33.34	[581.59 – 612.95]
Fixed LA	615.03	34.75	[598.69 – 631.37]
Dynamic Algorithm	662.53	43.44	[642.1 – 682.96]
Distance-Based	578.07	46.56	[556.18 – 599.96]

Table 5.11: 99% Confidence Interval Estimation of Total Location Management Cost for c = 5 and 6 Incoming Calls under the Random Mobility Model

For a larger relative cost of updates (c = 10), the statistical analysis shows that the two-location algorithm is better than the rest of algorithm for 6 incoming calls per day while for 12 incoming calls per day the reporting center strategy proves to be the best.

	Mean	StdDev	99% Confidence Interval
Reporting Center	693.5	29.81	[679.48 – 707.52]
Two-Location	1053.47	60.44	[1025.05 – 1081.89]
Fixed LA	990.9	39.59	[972.28 – 1009.52]
Dynamic Algorithm	1096.53	74.79	[1061.36 – 1131.7]
Distance-Based	793.5	64.35	[763.24 – 823.76]

Table 5.12: 99% Confidence Interval Estimation of Total Location Management Cost for c = 5 and 12 Incoming Calls under the Random Mobility Model

	Mean	StdDev	99% Confidence Interval
Reporting Center	872.83	57.32	[845.88 – 899.78]
Two-Location	765.43	47.36	[743.16 – 787.7]
Fixed LA	862.2	64.32	[831.95 – 892.45]
Dynamic Algorithm	878.37	60.19	[850.07 – 906.67]
Distance-Based	910.07	76.49	[874.1 – 946.04]

Table 5.13: 99% Confidence Interval Estimation of Total Location Management Cost for c = 10 and 6 Incoming Calls under the Random Mobility Model

	Mean	StdDev	99% Confidence Interval
Reporting Center	1056.83	53.78	[1031.55 – 1082.11]
Two-Location	1228.8	68.57	[1196.56 – 1261.04]
Fixed LA	1247.73	67.42	[1216.03 – 1279.43]
Dynamic Algorithm	1323.87	94.57	[1279.4 – 1368.34]
Distance-Based	1134.5	78.61	[1097.44 – 1171.46]

Table 5.14: 99% Confidence Interval Estimation of Total Location Management Cost for c = 10 and 12 Incoming Calls under the Random Mobility Model

Chapter 6

Conclusions and Future Work

The increased demand for wireless mobile communications, coupled with the finite available spectrum, has motivated investigation into alternative methods of tracking users and delivering calls. A mobility management scheme that reduces signaling traffic load and connection setup time is a pivotal issue in designing future Personal Communication Service (PCS) networks to satisfy Quality of Service requirements and to use network resources efficiently. Particularly required is scalable mobility management, to meet the explosive growth in number of users for the current second-generation wireless communication systems, and to materialize PCS concepts such as terminal, personal, and service mobility. Many mobility management schemes have been proposed for the reduction of signaling traffic. However, these schemes have not been sufficiently compared using a unified performance measure that is free of assumptions as to mobility model or database architecture.

6.1 Conclusions

In this thesis, we categorized the various mobility management schemes and clarify the appropriate domain for each type. Some proposals were selected for further evaluations that are based on reducing the signaling traffic over the wireless networks.

Two mobility models form the basis for comparative performance analysis in this thesis. The first was a realistic, measurement-based, place and time-dependent individual mobility model to simulate the mobility behavior of actual people. The second was a random mobility model based on simplified assumption, such as a uniform distribution of the direction of travel for each mobile subscriber, as being used in a number of proposed schemes. We analyzed the performance of each one of the five selected strategies under both mobility models, and showed the performance differences between the strategies. In the realistic mobility model, overall, the dynamic algorithm significantly lowers the location management costs, in terms of signaling messages generated, for all parameters examined and the distance-based strategy is the worst one. In the random mobility model, no single proposal shows consistent superior performance. In general, the reporting center strategy seems the best, while the dynamic strategy performed the worst.

From the results it is clear that the mobility model has a significant impact and the results described in various schemes using a random mobility model may not reflect the relative performance when deploying schemes in actual systems.

6.2 Future Work

In this thesis we have examined the impact of location update and paging messages on the radio interface. These evaluations have been made with a future wireless network framework in mind, using realistic and random mobility models for user movements. The selection of algorithms was based on the assumptions that the resource usage on the air interface is of primary importance. The signaling consumes scarce fixed radio bandwidth, and thus must be kept to a minimum. Another impact of the location updates is noticed at the VLR/HLR side, where the transaction rates can be very high and may be a cause of either saturation or reduction of the processing power dedicated to providing sophisticated services to the subscribers. One area of future research is to analyze the performance of the different proposals under both mobility models on the basis of database transactions, the processing overheads associated with them, the amount of memory required, etc.

For the analysis of the selected proposals it was assumed that the location update messages are the major contributor to the total location management cost while the number of incoming calls are considered fixed for each set of readings. As can be seen in Chapter 5, call arrivals also influence the performance of the location management schemes especially when the subscriber is not very mobile or receives a lot of calls during roaming. Another future set of experiments is to use realistic call arrival data along with realistic mobility patterns in order to thoroughly evaluate the selected algorithms.

The model used here in this thesis tries to incorporate the structure of daily travel patterns of the mobile users without making use of much simplified assumptions. Some of the assumptions considered in this model are the notion of fixed home and school cells. But in reality, these cells may be different if the subscriber moves to a home (in a different cell). The question is that how detailed do mobility models have to be? How to extend or modify the mobility model to answer “WHAT-IF” questions? For example, what will be the effect on the system performance if we add another user group, such as school going children? If we do not want to make any simplifying assumptions, we can alternatively collect traces of the actual movement of the mobile users from a real life system. In this way, we do not have to categorize mobile users into user groups and therefore we avoid simplifying assumptions that result from the somewhat arbitrary nature of such groupings. However using traces, we have to collect data for each user separately and this will result in higher overheads. It also becomes more difficult to experiment with alternative scenarios to answer "WHAT-IF" questions if the user mobility is directly based on traces.

One area of future research is to perform sensitivity analysis of the selected proposals and analyze how they behave during peak hours. For this we have to make some changes in the existing realistic mobility model to identify and analyze the proposal in peak hours only.

Possible future uses and extensions of the realistic mobility model include modification of the model to compare the performance of different handoff algorithms. This would require enhancement of the simulation to maintain knowledge of the duration of calls (incoming and outgoing) as the subscriber moves from cell to cell. Also, rather than simulating the movement of individual subscribers, the simulator has to deal with the simultaneous movement of many subscribers to create hotspots and capacity congestion.

Bibliography

[1] Sami Tabbane. *An alternative strategy for location tracking*. IEEE Journal on Selected Areas in Communications, vol.13, no.5, June 1995. pp. 880-892.

[2] Gregory P. Pollini and Chih-Lin I. *A profile-based location strategy and its performance.* IEEE Journal on Selected Areas in Communications, Vol. 15, No. 8, October 1997. pp. 1415-1423.

[3] Ozan K. Tonguz and Sumita Mishra. *A Simple Analytical Framework for Location Management in personal Communication Systems.* IEEE Transactions on Vehicular Technology, vol. 47, No. 2, May 1998. pp. 428-439.

[4] Yi-Bing Lin. *Reducing location update cost in a PCS network.* IEEE/ACM Transactions on Networking, vol. 5, No. 1, February 1997. pp. 25-33.

[5] Anna Hac and Xian Zhou. *Locating Strategies for Personal Communication Networks: A Novel Tracking Strategy*. IEEE Journal on Selected Areas in Communications, Vol. 15, No. 8, October 1997. pp. 1425-1436.

[6] L. R. Hu and S.S.Rappaport. *Adaptive location management scheme for global personal communications.* IEEE Proceedings on Communications., Vol. 144, No. 1, February 1997. pp.54-60.

[7] John Scourias and Thomas Kunz. *A Dynamic Individualized Location Management Algorithm.* In Proceedings of the 8th IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications, Helsinki, Finland. September 1997. pp. 1004-1008.

[8] A. Bar-Noy, I. Kessler and M. Sidi. *Mobile users: To update or not to update?.* Wireless Networks, Vol. 1, pages 175-185, 1995.

[9] U. Madhow, M. l. Honig and K. Steiglitz. *Optimization of Wireless Resources for Personal Communications Mobility Tracking.* IEEE/ACM Transactions on Networking, Vol. 3, December 1995. pp. 577-584.

[10] Naoki Matsumaru, Kshirasagar Naik and David S.L Wei. *Comparing three location management strategies for tracking mobile systems.* PDPTA '98 International conference. pp. 771-777.

[11] Jan Jannink, Derek Lam, Narayanan Shivakumar, Jennifer Widom and Donald C. Cox. *Efficient and flexible location management techniques for wireless communication systems.* Wireless Networks 3. 1997. pp. 361-374.

[12] Derek Lam, Yingwei Cui,, Jennifer Widom and Donald C. Cox.. *A location Management technique to support lifelong numbering in Personal communications services.* IEEE Global Telecommunications conference. November 1997.

[13] Shlomi Dolev, Dhiraj K. Pradhan and Jennifer L. Welch. *Modified tree structure for location management in mobile environments.* Computer Communications. 1996. pp. 335-345.

[14] P. Krishna, Natin H. Vaidya, and Dhiraj K. Pradhan. *Static and adaptive location management in mobile wireless networks.* Computer Communications 19. 1996. pp. 321-334.

[15] Narayanan Shivakumar and Jennifer Widom. *User profile replication for faster location lookup in Mobile environment.* In 1st ACM International Conference on Mobile Computing and Networking (MOBICOM), 1995. pp. 161-169.

[16] Chih-Lin, Gregory P.Pollini and Richard D.Gitlin. *PCS mobility management using the reverse virtual call setup algorithm.* IEEE/ACM Transactions on Networking, vol. 5, No. 1, February 1997. pp. 13-24.

[17] Joseph S. M. Ho and Ian F. Akyildiz. *Dynamic Hierarchical database Architecture for Location management in PCS Networks.* IEEE/ACM Transactions on Networking, vol. 5, No. 5, October 1997. pp. 646-660.

[18] R. Jain and Yi-Bing Lin. *An auxiliary user location strategy employing forwarding pointers to reduce network impacts of PCS.* ACM-JCM Wireless Networks, pp. 197-210. 1995.

[19] Yi-Bing Lin and Wen-Nung Tsai. *Location tracking with distributed HLR's and pointer forwarding.* IEEE Transactions on Vehicular Technology, vol. 47, No. 1, February 1998. pp. 58-64.

[20] Kuen-Liang Sue and Chien-Chao Tseng. *One-Step pointer forwarding strategy for location tracking in distributed HLR environment.* IEEE Journal on Selected Areas of

Communications. Vol. 15, No. 8, October 1997. pp. 1455-1466.

[21] Kin K. Leung and Yonatan Levy. *Global Mobility Management by Replicated Databases in Personal Communication Networks*. IEEE Journal on Selected Areas in Communications, Vol. 15, No. 8, October 1997. pp. 1582-1596.

[22] John G. Markoulidakis, George L. Lyberopoulos, Dimitrios F. Tsirkas, and Efstathios D. Sykas. *Mobility Modeling in Third-Generation Mobile Telecommunications Systems.* IEEE Personal Communications. August 1997. pp. 41-56.

[23] Hideaki Yoshino, Hisao Yamamoto and Hideaki Matsue. *Mobility Management Schemes and their Characteristics for Advanced Personal Communication Services in Distributed Environments*. IEICE Transactions on Communication., Vol. E81-B, No. 6, June 1998.

[24] Ravi Jain. *A caching strategy to reduce network impacts of PCS*. IEEE Journal on Selected Areas in Communication, 12 (8), November 1994.

[25] Moe Rahnema. *Overview of the GSM system and protocol architecture.* IEEE Communications Magazine. Vol.43. April 1993.

[26] EIA/TIA. *Cellular Radio-Telecommunications Intersystem Operations*. Technical Report. IS-41 Revision C, 1995.

[27] Tranplan Associates. *Waterloo Region Travel Survey 1987: An Overview of the Survey Findings.* Regional Municipality of Waterloo, Department of Planning and Development, October 1989.