

Cache-enhanced dynamic movement- based essay



**ASSIGN
BUSTER**

1. Introduction

Millions of people within the United States and around the world use cellular phones. They are such great gadgets that with a cell phone one can talk to anyone on the planet from just anywhere!! No wonder they are so popular and commonly used!

These days, cell phones provide an incredible array of functions, and new ones are being added at a breakneck pace. Depending on the cell-phone model, we can store contact information, make task or to-do lists, keep track of appointments and set reminders, use the built-in calculator for simple math, send or receive electronic-mail, get information (news, entertainment, stock quotes) from the Internet, play simple games and integrate other devices such as PDA's, MP3 Players and GPS receivers. It's not exaggerating by saying that cell phones might become a necessity rather than a luxury in a few years to come.

Having heard about a cell phone for a long time, it natural of us to wonder how a cell phone actually works? This is precisely the question that has lead me to choose to do my thesis on a related topic. One of the most interesting things about a cell phone is that it is actually a radio — an extremely sophisticated radio, but a radio nonetheless. In the dark ages before cell phones, people who really needed mobile-communications ability installed radio telephones in their cars. In the radio-telephone system, there was one central antenna tower per city, and perhaps 25 channels available on that tower. This central antenna meant that the phone in your car needed a powerful transmitter — big enough to transmit 40 or 50 miles (about 70 km).

It also meant that not many people could use radio telephones — there just were not enough channels [1].

The genius of the cellular system is the division of a city into small cells. This allows extensive frequency reuse across a city, so that millions of people can use cell phones simultaneously. In a typical analog cell-phone system in the United States, the cell-phone carrier receives about 800 frequencies to use across the city. The carrier divides the entire city into cells. Each cell is typically sized at about 10 square miles (26 square kilometers) [1]. Cells are normally thought of as hexagons on a larger hexagonal grid, as shown in Figure 1.1:

Location Area

Base Station

Rings

FIGURE 1.1 THE CELL TOPOLOGY

Each cell has a base station that consists of a tower and a small building containing the radio equipment that is used to communicate with Mobile Terminals over preassigned radio frequencies.

Cell phones have low-power transmitters in them. Many cell phones have two signal strengths: 0.6 watts and 3 watts [1]. The base station also transmits at low power. Low-power transmitters have two advantages:

* The transmissions of a base station and the phones within its cell do not make it very far outside that cell. Therefore, in Figure 1.1, both of the cells in <https://assignbuster.com/cache-enhanced-dynamic-movement-based-essay/>

alternate rings and non-adjacent cells can reuse the same frequency. The same frequencies can be reused extensively across the city.

* The power consumption of a cell phone, which is normally battery-operated, is relatively low. Low power corresponds to small batteries, and this is what has made handheld cellular phones possible.

The cellular approach requires a large number of base stations in a city of any size. A typical large city can have hundreds of towers. But because so many people are using cell phones, costs remain low per user. Each carrier in each city also runs one central office called the Mobile Telephone Switching Office (MTSO). This office handles all of the phone connections to the normal land-based phone system, and controls all of the base stations in the region. Groups of several cells are connected to a Mobile Switching Center (MSC) through which the calls are then routed to the telephone networks. The area serviced by a MSC is called a Registration Area (RA) or Location Area (LA). A group of RA's composes a Service Area (SA). Each SA is serviced by a Home Location Register (HLR). A wireless network may include several SAs and thus several HLRs. More about the Location Registers is discussed in the Literature Review (Chapter2).

All cell phones have special codes associated with them. These codes are used to identify the phone, the phone's owner and the service provider. Electronic Serial Number (ESN) (a unique 32-bit number programmed into the phone when it is manufactured), Mobile Identification Number (MIN) (a 10-digit number derived from the owners phone's number), and a System Identification Code (SID) (a unique 5-digit number that is assigned to each

carrier by the FCC-Federal Communications Commission (A U. S. government agency charged with the task of regulating all forms of interstate and international communication)) are a few of the standard cell phone codes employed. While the ESN is considered a permanent part of the phone, both the MIN and SID codes are programmed into the phone when one purchases a service plan and has the phone activated.

The next most intriguing question would be what exactly happens when a person A is trying to make a call to a person B. Here is what happens to the call [1](in 1G systems):

- * When B first powers up his/her phone, it listens for an SID on the control channel. The control channel is a special frequency that the phone and base station use to communicate to one another about things like call set-up and channel changing. If the phone cannot find any control channels to listen to, it determines that it is out of range and displays a “ no service” message.

- * When it receives the SID, the phone compares it to the SID programmed into the phone. If the SIDs match, the phone knows that the cell it is communicating with is part of its home system.

- * Along with the SID, the phone also transmits a registration request, and the MTSO keeps track of B’s phone location in a database — this way, the MTSO knows which cell B is in when it wants to ring his/her phone.

- * The MTSO gets the call from user A, and it tries to find B. It looks in its database to see which cell B is in.

* The MTSO picks a frequency pair that the phone will use in that cell to take the call.

* The MTSO communicates with B's phone over the control channel to tell it which frequencies to use, and once B's phone and the tower switch on those frequencies, the call is connected. B is then able to talk by two-way radio to his friend A.

* As B moves toward the edge of the cell, B's cell's base station notes that the signal strength is diminishing. Meanwhile, the base station in the cell B is moving toward sees B's phone signal strength increasing. The two base stations coordinate with each other through the MTSO, and at some point, B's phone gets a signal on a control channel telling it to change frequencies. This hand off switches B's phone to the new cell. This is illustrated in Figure 1. 2 [1].

FIGURE. 1. 2 ROAMING-AS THE MT MOVES FROM CELL TO CELL THE SIGNAL IS PASSED ACCORDINGLY.

If the SID on the control channel does not match the SID programmed into one's phone, and then the phone knows it is roaming. The MTSO of the cell that the person is roaming in contacts the MTSO of his/her home system, which then checks its database to confirm that the SID of the phone he is using, is valid. His home system verifies his phone to the local MTSO, which then tracks his phone as he moves through its cells. And the amazing thing is that all of this happens within seconds.

Both of the above concepts constitute location management in cellular networks, which is the focus of this thesis. Location management is a key issue in the operation of PCS networks. Two kinds of activity are involved in the operation: location update and paging. The basic approaches in location update, and terminal paging and more about the network topology will be discussed in the second chapter. The technique that is used for location update in this thesis is LA-based i. e. location registration will be triggered when a MT moves into new LA [2].

There are generally two kinds of techniques for location update: static and dynamic. In the static location update, the size of a paging area is fixed and is equal to the LA. Each cell in the LA will page once when a call arrives. The problem with this however would be if the LA is large, the paging cost is very high and a MT located at the boundary of LAs may suffer from excessive location updates.

In dynamic schemes, the size of the location area is determined dynamically according to the changes of mobility and calling patterns of the mobile terminals. The different kinds of dynamic location management schemes in use and the concepts involved are discussed in Chapter 2.

Although commercial mobile telephone networks existed as early as the 1940's, many consider the analog networks of the late 1970's to be the first generation (1G) wireless networks. The details of 1G, 2G and 3G and their stages of evolution and the concepts involved are discussed in the Literature review of the second chapter.

The most interesting feature of this thesis is that not only is it implemented for 3G networks where an additional register called the Gateway location register [3] it utilized (more details in Chapter 3), but a caching strategy is also employed at the different registers to analyse how well the current structure can be optimised. The caching strategy [4] would basically deal with that set of users who make a number of calls to a particular region, as is usually the case in practise. The entire simulation is done for four cases:

- * The standard 3G network without any cache.
- * 3G with a cache at the GLR
- * 3G with a cache at the VLR
- * 3G with a cache at both GLR and VLR.

The improved scheme employing caching is discussed in more detail in the fourth Chapter.

A simulation model is designed to evaluate the performance of the proposed scheme. It is assumed that the cellular network is partitioned into hexagonal cells of the same size. The target network consists of two SAs, 4 G_LAs and eight LAs, two hundred and forty four BSs and nine hundred and seventy six MTs.

An exponential cell residence time distribution is assumed in the study. We assume each MT resides in a cell for a time period then moves to one of its neighbors with equal probability. The Poisson process and exponential

process are used to describe the incoming call arrivals and the service time of a phone call respectively.

A discrete event simulator is designed to implement the simulation model. The object-oriented design makes it easier to extend the model or make modifications to the existing model. The objects in the design represent the objects' natural behavior and structure. The design is implemented entirely with Java programming language, which is portable across platforms. The system design and location update and call delivery strategies employed in the improved scheme are discussed in Chapter 4. Chapter 5 has the detailed discussion about the simulation model and discrete event simulator.

The following experiments are conducted using discrete event simulator among the four schemes:

- 1) Relationship between Call Migrate Ratio (CMR- it is the ratio of an MT's call arrival rate to it's migration rate) and the total location management cost.
- 2) Total location management cost among four selected schemes.
- 3) The effect of movement threshold value and paging delay on our proposed scheme (with cache) as compared with the standard scheme.

The experimental results show that employing a cache at both the GLR and the VLR would be the most beneficial compared to any other case. However, there would always be a trade-off between the memory costs and the performance gain. But since the memory costs are coming down every day, employing a cache for most user populations is beneficial any day. More details about the experiments are given in Chapter 6.

In summary, this paper is organized as follows: Chapter 1 is the introduction. The literature review is provided in the second chapter. Location management in 3G in particular and the caching strategies employed are discussed in Chapter 3. The details of the improved location management scheme are in Chapter 4 with the details of the simulation model in Chapter 5. The experimental results are analyzed in Chapter 6. Chapter 7 draws the conclusion.

2. Literature Review

Location management is a key issue in the operation of wireless networks. The Mobile Terminals (MT's) are either the automobile or hand held telephones or portable computers that the users use to send and receive calls. In order to successfully deliver an incoming call, the wireless network must keep track of the location of each mobile terminal continuously [5]. This process of tracking and locating MTs so that calls arriving for them can be directed to their current location is called location management [6].

Location Management is different from the location services in routing. Routing is the act of moving information across networks from a source to a destination. Along the way, at least one intermediate node typically is encountered. Routing involves two basic activities: determining optimal routing paths and transporting information groups (typically called packets) through an internetwork. Location Management on the other hand deals with not only the tracking of the MTs but also the call arrival aspect of the delivery of calls.

2. 1. Evolution of Cellular networks and the concepts involved:

<https://assignbuster.com/cache-enhanced-dynamic-movement-based-essay/>

Although commercial mobile telephone networks existed as early as the 1940's many consider the analog networks of the late 1970's to be the first generation (1G) wireless networks [7]. These networks were designed as fixed size networks, where an analog image of the sound was transmitted over the air and through the networks. The receiver and transmitter were tuned to the same frequency, and the voice that was transmitted was varied within a small band to create a pattern that the receiver could reconstruct, amplify and send to a speaker.

Although this was quite a revolution at that time, it suffered from a number of serious shortcomings. For instance, there was only a certain amount of mobility, lack of efficiency (because very few callers could fit into the available spectrum) and since everything was analog, there was not much scope for optimisations like coding or compression and the components were big and expensive making the handsets seem quite huge.

The result of this and a number of other reasons led to digital transmission, which led to the 2G cellular networks. With 2G came lower power consumption, small equipment size, higher bandwidth requirement per call, and most important of all multiple access techniques were used, either TDMA or CDMA.

In a mobile system, different users need to use different channels in order to avoid traffic colliding. Frequency division multiple access (FDMA), Time division multiple access (TDMA) and Code division multiple access (CDMA) are three common technologies used by cell-phone networks for transmitting information.

FDMA:

It a technology used in analog systems and the first generation cellular networks. In this technology each user is given a different frequency for separation i. e., it separates the spectrum into distinct voice channels by splitting it into uniform chunks of bandwidth. This is the simplest technique, but the trouble is that nearby frequencies interfere with each other.

FIGURE 2. 1. IN FDMA, EACH PHONE USES A DIFFERENT FREQUENCY.

TDMA:

Most of the existing mobile data networks use Time Division Multiple Access in which each conversation only uses a frequency for part of the time i. e., it separates users by assigning different time slots for each channel. Using Time Division Multiple Access, a narrow band that is 30 kHz wide and 6.7 milliseconds long is split time-wise into three time slots. This is more clearly illustrated in Figure 2. 2.

FIGURE 2. 2. TDMA SPLITS A FREQUENCY INTO TIME SLOTS.

Narrow band means “ channels” in the traditional sense. Each conversation gets the radio for one-third of the time. This is possible because voice data that has been converted to digital information is compressed so that it takes up significantly less transmission space. Therefore, TDMA has three times the capacity of an analog system using the same number of channels. TDMA systems operate in either the 800-MHz (IS-54) or 1900-MHz (IS-136) frequency bands [7].

CDMA:

A second (and more sophisticated) approach is Code Division Multiple Access (CDMA) in which different users are separated by different codes i. e., CDMA encodes each transmission to provide, in theory, unlimited capacity. CDMA takes an entirely different approach from TDMA. CDMA, after digitizing data, spreads it out over the entire available bandwidth. Multiple calls are overlaid on each other on the channel, with each assigned a unique sequence code [1]. CDMA is a form of spread spectrum, which simply means that data is sent in small pieces over a number of the discrete frequencies available for use at any time in the specified range.

All of the users transmit in the same wide-band chunk of spectrum. Each user's signal is spread over the entire bandwidth by a unique spreading code. At the receiver, that same unique code is used to recover the signal. Because CDMA systems need to put an accurate time-stamp on each piece of a signal, it references the GPS system for this information. Between eight and 10 separate calls can be carried in the same channel space as one analog AMPS call. CDMA technology is the basis for Interim Standard 95 (IS-95) and operates in both the 800-MHz and 1900-MHz frequency bands. This is more clearly illustrated in Figure 2. 3.

FIGURE 2. 3. IN CDMA, EACH PHONE'S DATA HAS A UNIQUE CODE.

From the 2G of cellular networks onwards, either TDMA or CDMA are used and later WCDMA being the most popular. For second Generation: USA uses two standards IS-95 (CDMA) and IS-136 (D-AMPS). Europe consolidated these to one system called the GSM (Global System for Mobile communications).

<https://assignbuster.com/cache-enhanced-dynamic-movement-based-essay/>

Japan uses Personal Digital Cellular (PDC). Cellular radio is the fastest growing segment of the communications industry and with the advent of digital systems and the advances in digital chip technology, everything from smaller and slick handsets, voice mail, call waiting to advanced supplementary services such as SMS came to reality. Cellular companies reported a subscription base of more than 360 million people in Fall 2000 for GSM alone with a few more in other systems like cdmaOne, PDC and TDMA (IS-136) [7].

The UMTS proposed to achieve this ideal of PCS data, which includes services that are independent of Location, Network, Terminal and that led to the 3G Cellular networks. The major driving force behind the 3G wireless systems was not just the need for capacity and worldwide roaming, but also higher bit (data transmission) rates, a standardization of radio interfaces, use of common global frequencies for all cellular networks, and a higher QoS.

Third generation cellular networks had a number of features like integration of wired and wireless networks, single network infrastructure, encompass paging, cordless, phones, wireless LANs, enough bandwidth offered for multimedia services, the multiple access technique to be used are either TDMA or CDMA, or a combination of both and can be deployed and support multimedia applications. General Packet Radio Service (GPRS) and EDGE (Enhanced Data rates for Global Evolution) are 2.5 Generation packet based networks technology for GSM networks [7]. The two standards WCDMA and CDMA2000 are by far the main branches of the 3G standard. The evolution of the Cellular Networks right from analog over the years has been very

interesting and rapid. The different stages of this evolution [8] are shown in Figure. 2. 4.

Technology

Service

Data Capability

Expected Deployment

GSM

Circuit-switched data based on the standard GSM 07. 07

9. 6 Kbps or 14. 4 Kbps

Available worldwide now

High-speed circuit-switched data (HSCSD)

28. 8 to 56 Kbps service likely

Limited deployment 1999 and 2000 as many carriers will wait for GPRS

General Packet Radio Service (GPRS)

IP and X. 25 communications over Kbps

Trial deployments in 2000, rollout of service 2001

Enhanced Data Rates for GSM Evolution (EDGE)

IP communications to 384 Kbps. Roaming with IS-136 networks possible.

<https://assignbuster.com/cache-enhanced-dynamic-movement-based-essay/>

Trial deployment in 2001, rollout of service 2002

Wideband CDMA (WCDMA)

Similar to EDGE but adds 2Mbps indoor capability. Increased capacity for voice.

Initial deployment in 2002 or 2003

IS-136

Circuit-switched data based on the standard IS-135

9.6 Kbps

Some carriers may offer service, but not on widespread basis as key carriers already offer Cellular Digital Packet Data (CDPD)

EDGE

IP communications to 384 Kbps. Roaming with GSM networks possible.

Initial deployment 2002 or 2003

WCDMA/WTDMA

Similar to EDGE but adds 2Mbps indoor capability

No stated deployment plans

CDMA

Circuit-switched data based on the standard IS-707

9. 6 Kbps or 14. 4 Kbps

Available by some carriers now

IS-95B

IP communications to 64 Kbps

Expected in Japanese markets by early 2000

CDMA2000 – 1XRTT

IP communications to 144 Kbps

Trial deployment in 2001, rollout of service 2002

CDMA2000 – 3XRTT

IP communications to 384 Kbps outdoors and 2 Mbps indoors

Initial deployment in 2002 or 2003.

FIGURE 2. 4. SUMMARY OF FORTHCOMING CELLULAR-DATA SERVICES.

TIME ESTIMATES BY RYSAVY RESEARCH.

With its evolutionary approach, there are many reasons for expecting a long and steady expansion of the UMTS and, more generally, of 3G systems and services:

* The continuous trend towards integration of wireless cellular technologies with wired networks and services, as well as with other wireless short range

technologies, such as those under the suite of IEEE 802. 11, and Bluetooth series of standards;

* The expansion of the IP paradigm towards the wireless world, fuelled by the standardization of IPv6, with its expanded addressing capability, and the development of the mobile IP protocol;

All the above reasons imply the need for continuous upgrading of networks and services transparently to the user: even profound technological (r)evolutions must be hidden to the user. The most profitable networks and successful operators will be the most flexible and responsive, as competition will be based on the spectrum of delivered services rather than simply on tariffs as mostly was for the 2G era. We can therefore easily expect that a long series of advances in 3G wireless networks will set the scene and silently but deeply shape our future. The user will be unaware of most of them so that he/she will perceive service continuity while at the same time enjoying advantages in getting a faster, more ample, more ubiquitous, and more friendly service [9].

2. 2. Framework of a Wireless Network in 3G and Location Management in 3G

All of the popular existing Cellular /PCS Networks today employ the Home Location Registers and the Visitor Location Registers for location management. Location management is nothing but keeping track of the mobile terminals moving from one place to another and can be broadly set to consist of 2 major operations: Location update by which the system keeps track of the location of the mobile terminals that are not in conversation.

Mobile terminals are the subscribers that use either automobile hand held telephones or portable computers to send and receive calls. The MT reports its up-to-date location information dynamically.

The other operation is called paging, which is a search process by which the system searches for the MT by sending polling signals to cells in the paging area (which may include one or more cells). To perform either of these operations would incur a significant amount of cost, which should be minimized in the systems.

Almost all of the existing cellular/PCS networks have a cellular structure with a given service area divided into cells with each cell having a Base station for communication with mobile terminals over preassigned radio frequencies. Groups of several such cells are connected to a Mobile Switching center (MSC) through which the calls are then routed to the telephone networks. MSC is a telephone exchange specially assembled for mobile applications. It acts as an interface between the mobile phones (via base stations) and the PSTN or PSDN, which makes mobile services widely accessible to the public. The MSCs are where the location management system resides. In 2G cellular networks, the two-tier mobility databases, HLR and VLR are utilized to support this.

Each service is partitioned into Location areas (LA's). Each LA includes tens or hundreds of cells and is serviced by a VLR, which is in turn associated with an MSC in the networks. The VLR contains temporary records of the MT's profiles and location information. The VLR is used for retrieving information for handling calls to and from a visiting mobile terminal. The HLR is used to

record the permanent data (e. g., directory number, profile information, current location, and valid period) of the mobile terminals whose primary subscription is within the area.

Paging is a search process conducted in a Paging area (PA). Based on if the PA size is fixed, location management schemes can be broadly classified into static or dynamic location schemes. If the size of a paging area is fixed and is equal to the LA, each cell in the LA will page once when a call arrives for an MT currently registered in the LA. This is the static scheme but the problem would be if an MT is close to the boundary of an LA when there would be excessive location updates as it moves back and forth between two LA's.

Dynamic location management is what has been used for this thesis. In dynamic location management, the size of a location area is determined dynamically according to the changes of mobility and calling patterns of mobile terminals. Of the three kinds of dynamic location management existing (time-based, distance based and movement based) movement-based location management is the most practical and effective [2] and is hence employed in this thesis.

With increasing rates of international travel, the number of roaming users increases. In order to reduce the international/remote roaming signaling traffic, the Gateway Location Register (GLR) within the UMTS core network was proposed. (Specified by the Third Generation Partnership Project³ (3GPP); Technical Specification Group Services and Systems Aspects; Network Architecture (Release 5) 2000-12). This is the special feature that I would be implementing in my thesis to compare it with the analytical results

proposed in “ The Movement-based Location Management for 3G Cellular Networks” [2].

The GLR is the node between the VLR and or SGSN and the HLR. It handles location management of roaming subscribers in a visited network without involving the HLR in every change of LA's. Therefore, the signaling traffic between the visited and home mobile systems will be reduced and the location updating and handling of user profile data across network boundaries is optimized. The GLR is located in the visited network. It stores the roamer's information and handles location management within the network. Interface between the HLR and the GLR is the same as that between the HLR and VLR as the presence of the GLR is invisible from the home network.

2. 3. Teletraffic Modeling

Teletraffic models are very important tools for network analysis and design. Before discussing the model I would be using in the thesis, the various modeling techniques used in the performance analysis of location update and paging schemes in general are briefly discussed. Teletraffic models can be broadly divided into four categories [10]:

- * Network Topology Model

- * Mobile Residence Model

- * Mobility Model

- * Call Model

In the following section, each kind of model is discussed in detail.

2. 3. 1 Network Topology Model

The network topology model specifies the connectivity between base stations or cells. Regular cell topologies are commonly used to model the coverage area of a cellular network [11]. There are three kinds of regular topology models: Linear Model (for one-dimensional networks), and the Mesh Model and Hexagonal Model (for two-dimensional). Although these models simplify analytical computation, they do not give an accurate representation of a realistic cellular network topology, where the sizes of cells depend on transmit power, receiver sensitivity and antenna radiation pattern, propagation environment, and the number of neighboring cells varies from cell to cell [11].

A graph model is introduced to the cellular network in a paper [12] by S. K. Sen et al. Considering the case of a zone or LA-based cellular mobile system, the network can be represented by a bounded-degree, connected graph $G = (V, E)$ where the node-set V represents the LAs and the edge-set E represents the access paths between pairs of LAs.

2. 3. 2 The Mobile Residence Model

The residence time at a location represents the amount of time the mobile user stays in that location before moving somewhere else. Certain location update schemes such as selective LA and profile-based update algorithms depend on good estimation of the residence time at different LAs [11].

Generally, research on location management assumes a geometric cell

residence time distribution in their performance analysis. The distribution is assumed to be Independent and Identically Distributed (IID) for all cells. However the IID geometric residence time assumption does not capture an accurate representation of individual user mobility patterns, where a user may stay at certain locations, such as his home or office, for a relatively long period of time. In another interesting article [13] the authors presented a residence model that allows an IID general cell residence time distribution, but the model is only applied to a hexagonal cell configuration.

2. 3. 3 Mobility Models

Conceptually, the term mobility describes the ability of a person to move. Mobility is usually described by the average velocity, and distance covered by the mobile user [14]. Mobility models play an important role in examining different issues in wireless networks, including resource allocation, handoff, and location management. In general, the mobility models depend on the speed, direction, or movement history of the mobile user [15].

Mobility model can be divided into two categories: models based on individual movement behavior, i. e. the Gaussian Model and the Markov Model, and models based on aggregate movement behavior, i. e. the Fluid Flow Model and Gravity Model. In the next section, we will discuss those models one by one.

The Gaussian Mobility Model This model is established on the isotropic random user motion with drift, defined as the mean velocity in a given direction. The probability distribution function in a one-dimensional model is given by (Equation 2-1).

..... Equation 2-1

where x is the location variable, t is the time elapsed since last contact with the mobile terminal, v is the mean drift velocity, and D is a constant. When the time is partitioned into small intervals and $v = 0$. Gaussian user location distribution has been used to study the timer and state-based update schemes [11].

The Markov Model This model attempts to capture the direction of a user's movement pattern by assigning different probabilities to different neighboring cells. Suppose a mobile user is located in cell i . In each time interval, the user will either remain in the cell with probability $\Pr(i| i)$ or move to a neighboring cell j with probability $\Pr(j| i)$. Thus, the user's preference is characterized by the probability function. The cell residence time follows a geometric distribution. This model has been used for performance analysis of the selective LA and threshold-based update schemes. One of the limitations of this approach is that there is no concept of movement history or trip of a particular mobile user [11].

The Fluid Flow Model This model characterizes aggregate movement behavior as the flow of a fluid. Mobile users are assumed to move at an average velocity of v , and their direction of movement is uniformly distributed over $[0, 2\pi]$. Assuming that the mobile users are uniformly populated with a density of ρ and the location area boundary is of length L , the rate of users moving out of LA C is given by (Equation 2-2).

Equation 2-2

The above model is accurate for boundary crossing rate in a symmetric grid of streets (i. e., Manhattan-style). This model has been used to study the profile-based update scheme. One of the limitations of this model is that it describes aggregate traffic and is difficult to apply to scenarios where individual movement patterns are desired [11].

The Gravity Model This model has been used extensively in the area of transportation research to model human movement behavior. In this model, the amount of traffic T_{ij} moving from region i to region j is described by (Equation 2-3).

Equation 2-3

where P_i is the population in region i , and $\{K_i, j\}$ are parameters that have to be determined for all possible region pairs (i, j) . In few cases [15], variations of the gravity model have been used to describe the national and international mobility models. The national mobility model characterizes aggregate movement behavior between the ten largest metropolitan areas in the United States. The international mobility model characterizes aggregate movement behavior between the United States and ten other countries. Both models are constructed based on air passenger traffic data.

2. 3. 4 The Call Arrive Model

Poisson distribution is commonly used to model incoming call arrivals in wireless network; that is, the time between call arrivals follows an exponential distribution. Although the Poisson assumption is true for aggregate call arrivals in telephone networks, the call arrival rate for an

individual user may not be Poisson distributed. It may depend on the time of day and day of the week. Since the threshold values of different threshold-based update algorithms depend on the individual call arrival rate, a realistic time-varying call model should be used in order to achieve better results.

The time-varying call model can be constructed based on the call arrival data in the user's billing record [11].

2. 4. Basic Approaches of Location Management

Location management as a whole involves two basic activities: location update and paging. A mobile terminal can report its location as soon as it crosses a cell boundary. This reporting initiated by the MTs to limit the search space at a later point in time, is called a location update. The system on the other hand, can initiate the search for a MT, called paging, by simultaneously polling all the cells where the MT's can possibly be found [12].

It is well known that there is a trade-off between the costs of location update and paging. If the mobile terminal updates its location whenever it crosses a cell boundary, the network can maintain its precise location, thus obviating the need for paging. However, if the call arrival rate is low, the network wastes its resources by processing frequent update information, and the mobile terminal wastes its power transmitting the update signal. On the other hand, if the mobile terminal does not perform location update frequently, a large coverage area has to be paged when a call arrives, which wastes radio bandwidth. Thus, the central problem of location management

is to devise algorithms that minimize the overall cost of location update and paging [11].

Current wireless networks use a Location Area (LA)-based update algorithm and blanket polling paging strategy. The coverage area is partitioned into a number of LAs, each containing a group of cells. All base stations within the same LA broadcast the identifier (ID) of their LA periodically. Each mobile terminal compares its registered LA ID with the current broadcast LA ID. Location update is triggered if the two IDs are different. Upon a call arrival for a particular mobile terminal, all the cells within its current LA are polled simultaneously, ensuring success within a single step [11].

Although the LA-based update scheme is widely adopted by current cellular systems, there are a number of inefficiencies associated with this scheme like:

- * For an LA with a large number of cells, a significant amount of radio bandwidth is consumed in paging for each call arrival. This may not be scalable for future wireless broadband networks with a large number of mobile users.
- * MT's located close to an LA boundary may perform excessive location updates as they move back and forth between two LAs, thus increasing the signaling and processing load on the network database.
- * In addition, since each user has its own mobility pattern, it is difficult to choose an LA size that is optimal for all users [10].

Current research on location management focuses on per-user-based algorithms in which location update and terminal paging procedures can be adjusted dynamically based on a user's call and mobility patterns [11]. A per-user caching technique [4] is also employed in this thesis as explained in the Chapter 3.

In summary, location management has to address the following issues:

- * When and how MTs update its location to the network?
- * How the exact location of the called MTs are determined within a specific time constraint when a call arrives?

In following section, current researches about location update and paging are discussed separately.

2. 5. Current Approaches in Location Update

In order to reduce its location uncertainty, each mobile terminal has to report its location from time to time. The location update procedure begins with an update message sent by the mobile terminal over the uplink control channel, which is followed by some signaling procedures that update the database. Location update schemes can be divided into two categories: static and dynamic. In a static scheme, location update is triggered based on the topology of the network. In a dynamic scheme, location update is based on the user's call and mobility patterns [11]. Current research in location update can be divided into following categories [10]:

* Mobile initiated location update strategies. These can be divided into sub-areas:

- o Movement-based Update

- o Distance-based Update

- o Time-based Update

- o State-based Update

- * Prediction-based schemes

- * Profile-based location update schemes

- * Protocol-based schemes

The details of each of these strategies are explained in the following paragraphs.

2. 5. 1 Mobile Initiated Location Update Strategies

Mobile initiated update dynamic location update scheme wherein an MT performs location update after crossing a fixed threshold number of cells, at a fixed interval of time, whenever the distance MT moved exceed certain threshold, or at a certain state.

Movement-based location update. Ian F. Akyildiz et al. [10] present a movement-based location update scheme such that a mobile terminal performs a location update when the number of movements since the last location registration equals to a predefined value d . This value is called the

location update movement threshold. When an incoming call arrives, the network pages the cells within a distance d from the last registered location of the called mobile terminal. A selective paging mechanism is used in one of the papers [11] where, based on the maximum paging delay constraint, the paging process can take place in more than one step. In each step, the network selects a subset of the cells for paging. The paging process terminates as soon as MT is found.

Distance-based Location Update. In the distance-based location update scheme, MT needs to record the distance it has moved since the last location registration and performs an update when the distance exceed a certain threshold. The idea inside the strategy is similar to that in movement-based scheme, but threshold they compared is different. Madhow et al [16] consider the distance-update as an optimization problem. Under a one-dimensional linear model the optimal distance threshold is determined by dynamic programming. J. Ho et al [17] proposed an approach to compute the optimal threshold in a two-dimensional hexagonal model based on the assumption of symmetric random walk pattern.

Time-based Location Update. In the timer-based location update strategy [18], a MT updates its location every T time interval. This scheme does not require the MT to record or process location information during the time interval between updates. For implementation, the timer threshold can be programmed into the MT by a hardware or software timer. The authors introduce an analysis model for the time-based scheme in the same paper [18]. In this model, time-varying Gaussian user distribution a Poisson page-arrival model are used to formulate the paging/registration optimization

problem. Analytic results show that the timer-based scheme performs much better than the LA-based scheme.

State-based Location Update. In the state-based location update scheme [11], the MT determines whether to perform location update based on its current state. The state information can include the time elapsed or the number of cell crossings since the last update, the cell distance between the current and last registered locations, or some other criteria. Thus, maintaining different state information corresponds to different location update schemes.

C. Rose [19] proposed a state-based scheme where the system state includes the current location and the time elapsed since the last update. A time-varying Gaussian process is used to model the user's movement. The suboptimal solution for the average cost of location update and paging under no paging delay constraint is obtained by a greedy method. Results show that the state-based update scheme achieves 10 percent improvement in average cost compared to the timer-based scheme.

2. 5. 2 Prediction-based Location Update Schemes.

In prediction-based location update schemes, a list or some profiles are required to maintain to predict the current or next location of a MT. Thus radio bandwidth required for signaling can be saved. If the MT is not found by prediction, the call setup delay may actually increase.

C. Rose et al. [19] proposed a scheme that uses the probability distribution of user location to predict the sequential search order over many locations.

The probability distributions are provided either through measurement or analysis of motion models. The maximum delay constraint and weighted mean cases are solved via dynamic programming. However the minimization of mean locations polled could not be solved with dynamic programming and hence continuous formulation was used to solve the problem.

Y-B Lin proposed a “two location algorithm” in one of his papers [20]. In this algorithm, a mobile terminal has a small built-in memory to store the addresses for the two most recently visited LAs. The algorithm guarantees that the MT is in one of the two locations. If the new location is not found in the first memory location, the address for the LA just left is kept, and the other address is replaced by the address of the new LA.

Ben Liang et al presented a predictive distance-based mobility management Scheme [21], which is also quite interesting. The future location of a mobile is predicted based on the probability density function of the mobile's location, which is, in turn, given by the Gauss-Markov model based on its location and velocity at the time of the last location update. The prediction information is made available to both the network and the mobiles.

Therefore, a mobile is aware of the network's prediction of its location in time. The mobile checks its position periodically and performs location update whenever it reaches some threshold distance away from the predicted location. To locate a mobile, the network pages the mobile starting from the predicted location and outwards, in a shortest-distance-first order, and until the mobile is found.

S. K. Sen et al [12] proposed a near-optimal location update strategy. In this scheme, each MT only updates in certain pre-selected location areas based on its own mobility pattern. For example, a daily commuter may cross a number of LAs on his way from home to work. He may stay in some LAs for very short time. So update process at certain LAs can be skipped. To estimate the transition probabilities between LAs, MT's movements throughout the day can be observed over long period of time. Since LA based update scheme is used in current wireless network, the information about the mobility of the MT can be retrieved from the database. An analytical model is also introduced in [12]. The experimental results show that for low user location probability, low to moderately high call arrival rate and/or comparatively high update cost, skipping updating in several LAs leads to a minimization of the location management cost.

2. 5. 3 Profile-based Location Update Scheme

The rationale behind the profile-based update scheme [22] is to reduce the update cost by taking advantage of the user's mobility pattern. The network maintains a profile for each user, such as a sequential list of the LAs the user is most likely to be located at the different time period. This list is sorted from the most to least likely LA where a user can be found. When a call arrives, the LA on the list are paged sequentially. As long as the MT moves between LAs within the list, no location update is necessary. Location update is performed only when mobile terminal moves to a new LA not on the list. For implementation, each MT must maintain a valid sequential list corresponding to a particular time interval. This list has to be updated from time to time.

2. 5. 4 Protocol-based Location Update Scheme

Protocol-based Location Update Scheme is focus on the call setup protocol. The protocol defines the procedures for call setup, such as retrieving the user location information, notifying the user of incoming calls, and setting up the circuit connection through the wired and wireless network paths [23].

In one of his papers [24], Y. Cui et al presented a Lightweight location lookup protocol that can be used to reduce the network signaling cost. In this scheme, the location lookup stage in the call setup from X to Y retrieves only Y's VLR address instead of his entire routing information including the MSC and temporary address, thereby reducing the tasks of the location lookup. Reverse connection setup protocol has also been used to reduce the network cost of failed cost attempts by skipping the resource reservation when the user cannot answer the page.

Li et al [5] introduced a dynamic location update strategy for PCS network. The concept of the scheme is to provide dynamic copies of location information of MT's in the nearest HLR database, which allows mobile users to access efficiently and then reduces the signaling and database access traffic. The HLR/VLR database architecture in the network isn't changed. However, HLR is divided into master HLR and current HLR. The area the HLR served is called Service Area (SA). Correspondingly, there are master SA and current SA. If the MT moves out of the current SA and its master SA, the old current HLR sends its record indicating the current serving MSC of the MT to the new current HLR. The new current HLR sends a registration

acknowledgment message to the old current HLR. If the old HLR is the master HLR, it does nothing; otherwise, it deletes the record of the MT.

In [5], an analytical model is presented to study the performance of the location management method. The study shows that the proposed scheme can reduce the total costs of location registration significantly comparing to the existing location management scheme.

2. 6. Current approaches on Terminal Paging

Terminal paging is the process by which the network determines the exact location of a particular mobile terminal. In each polling cycle, polling signals are sent over the downlink control channel to all cells where the mobile terminal is likely to be present. All the mobile terminals listen to the page message, and only the target mobile terminal sends a response message back over the uplink control channel. In each polling cycle, there is a timeout period [10]. If the target mobile terminal replies before the timeout, the paging process is terminated. Otherwise, another group of cells is chosen in the next polling cycle [11].

To avoid call dropping, the mobile terminal must be located within an allowable time constraint. The maximum paging delay corresponds to the maximum number of polling cycles allowed to locate the mobile terminal. For example, if the maximum paging delay is equal to one, the mobile terminal has to be located within single search iteration. Since radio bandwidth is consumed during the paging process, the paging cost is proportional to the number of polling cycles, as well as the number of cells being polled in each

cycle. The paging area depends on the information provided by the location update function [11].

Current researches regarding the paging are mainly in following areas [10]:

- * Blanket Polling
- * Sequential Paging
- * Shortest-Distance-First Paging
- * Velocity Paging
- * Ensemble Polling

In the following paragraph, we will discuss the above listed paging strategies one by one.

Blanket Polling In blanket polling, all the cells within the LA in which the mobile terminal is located are polled simultaneously when a call arrives. Since the MT is located within the LA, its location can be determined within a single polling cycle. This paging strategy is currently deployed on top of the LA-based update scheme in existing PCS network. The major drawback of blanket polling is that since the number of cells within a typical LA is large, the paging cost is very high [11].

Sequential Paging In the sequential paging scheme, the current location of MT is predicted based on the user's location probability distribution. Polling signals are sent only to those cells in which the user might be. In [19], the authors state that, given the probability distribution on user location and

under no paging delay constraint, the paging cost is minimized by sequentially polling the cells in decreasing order of probability. The authors obtained the optimal paging sequence resulting in minimum paging strategy with average paging delay constraint. The sequential paging strategy has been used for performance analysis of the timer and state-based update schemes.

Shortest-Distance-First Paging In this paging strategy, the network pages the mobile terminal starting from the cell where the mobile terminal last updated its location, and moving outward in a shortest-distance-first order. The distance is measured in terms of the number of cells away from the last update location. If a threshold-based update scheme (e. g., distance or movement) is used, the paging or residing area of the mobile terminal is bounded. The mobile terminal can be located within a fixed number of polling cycles. The paging delay constraint can be incorporated by grouping cells at different distances for each polling cycle [11].

Velocity Paging The velocity-paging scheme [25] is dynamic paging algorithm. The scheme can predict the paging area for a MT based on its semi-real-time movement information. Mobile users are grouped into different velocity classes. When a call arrives, the system checks the MT's velocity class, estimates how far it can go and pages the candidate cells. The proposed strategy avoids maintaining complex user profiles for MT's but still reduces paging cost significantly. The velocity concept can be implemented with existing registration schemes, such as movement-based registration or distance-based registration with minimum additional cost.

Ensemble Polling In [26], authors considered a simple planned ensemble polling for reducing paging delay. This approach is based on the concept of effective paging load that depends not only request rates, but also on the average unit mobility and paging discipline employed.

2. 7. Qualitative comparisons of the location management schemes

A qualitative analysis of the various schemes discussed above is given in [27]. The overhead requirement of the various location update schemes can be evaluated in terms of number of signaling or update messages transmitted, number of paging messages, and the amount of processing and storage. A certain distribution of inter-arrival of calls can be assumed with average value dependent on user call activity factor. Then assuming exponential or other suitable distribution for interval between location update messages from a mobile user (with an average time being a function of user speed, movement patterns, the number of cells in a dynamic location area, or the distance between reporting cells), the probability distribution of number of updates before a call can be computed.

If a call arrives before exact location update, then number of paging messages required should be computed. In general, the static schemes such as always-update, always-page, static LA, static reporting cell etc. are simple to implement, but they do not take advantage of the user mobility patterns. The dynamic schemes are difficult to implement, and some times require more computational and storage requirements, but they take the user mobility patterns into account.

While there is no best or worst scheme, the choice of a location management scheme depends on several factors such as the ones listed below:

- * Network characteristics: Paging for location update and centralized database can be used in smaller wireless networks. However, for a large network, paging is not recommended. Also, distributed database architecture should be used for larger networks.
- * CMR: If the CMR is low, always-update should not be used. If it is too high, always-page scheme should not be used.
- * Signaling traffic: Local anchoring may help reduce this cost.
- * Reliability and consistency issues: Database failure recovery procedures should be well defined. Database replication can be used to increase reliability.
- * Call setup delay: This can be improved by maintaining local caches or by using distributed databases.

Other factors include the amount of storage and amount of processing requirements. Certain schemes like the prediction-based scheme may not be applicable if the amount of storage and processing power is limited.

3. A Location Management Scheme in 3G and Per-User Caching

In this chapter, we first introduce a recently proposed movement-based location scheme for 3G, which has been taken as the reference for the 3G scheme implemented in the thesis. All of the concepts of 3G are dealt with in this scheme. Caching is one very popular technique usually employed to

<https://assignbuster.com/cache-enhanced-dynamic-movement-based-essay/>

reduce the costs involved in location management. A technique called per-user caching is discussed.

3. 1. A Recently Proposed Movement-based Location Management Scheme in 3G

The following is the scheme on which the thesis would be based. A number of schemes have been proposed so far for the 2-tier architectures but this scheme deals with 3-tier architecture used in 3G cellular networks. The Figure shows the simplified network architecture for a UMTS system with the GLR deployment at the edge of the visited networks. A GLR contains roamers subscriber's profile and location information. At the first location update procedure under the GLR, the subscriber profile information is downloaded from the HLR to the GLR.

The following Figure (Figure. 3. 1) shows the basic architecture as far as the location management databases are concerned in the thesis. The entire service area would be handled by the HLR with the help of a number of GLR's under which in turn have a number of VLR's under them. Each VLR is under the control of an MSC, which would be covering the area of a number of BS's. Each Base station area is again divided into cells each having a number of MT's.

FIGURE 3. 1. THE STRUCTURE OF THE 3G NETWORK

The GLR handles Update Location messages from the VLR's as if they were the HLR of the subscribers at the subsequent location updates. This helps in the minimization of the costly inter-network signaling for location

management, as the entire procedure is invisible from the home network. The profile information is kept in the GLR until a Cancel Location message is received from the HLR. The relationship between the GLR and the HLR in 3G wireless systems is the same as that between the VLR and the HLR in the 2G wireless cellular systems (such as GSM) in terms of signaling traffic for location management. A GLR as shown in the Figure below (Figure 3. 2) can interact with multiple VLR's.

Home Network Home Network

Visited Network Visited Network

FIGURE. 3. 2 MOBILITY DATABASE ARCHITECTURE

There are 3 kinds of location updates in 3G cellular networks: HLR location updates, GLR location updates, and VLR location updates. All of these together with paging procedures will cause a significant amount of cost such as wireless bandwidth and processing power at the mobile terminals, the BSs and mobility databases.

The service area is divided into Gateway Location areas (G-LA's) with each G-LA consisting of a number of LA's, which in turn is a group of cells. A Paging Area includes a number of cells within an LA and its size is variable. To make things simple we assume that the cellular network is a homogeneous hexagonal cell configuration, in which all the cells have the same shape (hexagonal) and the same size. The ring concept for the movement based location management scheme is that the cell that is under consideration is treated as the center cell (Ring-0) and the Ring-1 includes all

6 around the center cell and so on. Hence it's easy to see that Ring- r has 6 cells except that Ring-0 has only 1 cell where $r = 0, 1, 2, \dots$

A HLR location update is performed when an MT crosses the boundary of a G-LA. A GLR location update, on the other hand, is performed when an MT crosses the boundary of an LA. A VLR location update is performed when an MT completes 'd' movements between cells, where d is the movement threshold. A HLR location update involves the updating of both the GLR and the VLR while a GLR location update involves only the VLR update. It is important to note that a PA is the area within the LA, where the last VLR location update is performed and the circle area with the diameter $d-1$ and the center where the last VLR location update happens.

3. 2. The Concept Of Caching In Cellular Networks

In call delivery, whenever incoming calls arrive, signaling messages are exchanged between the calling MT and the called MT. Furthermore, calls are routed through same links several times when a particular MT receives many calls from the same location. Avoiding extra links will not only decrease the signaling delay of establishing call connection but also improve system performance. To store the location of a frequently accessed MT by a remote MT at the local database (cache) of the remote MT can reduce the call connection delay between the two MTs. However, the cached information may be obsolete due to the mobility of MTs. In this case, the signaling delay and overhead of call delivery are larger rather than that of call delivery using cache.

To reduce the obsolescence of the cached information, the system must cache highly re-usable information only. Deciding which information is highly reusable is difficult. Most of the caching schemes available are static. The static based schemes have to maintain the mobile history of MTs during a long-term period in order to provide the mobility information of the MTs for the calling network. This however results in exhaustion of resources. A caching scheme on a dynamic location management might however answer the question of which data is highly reusable without extra computation overhead.

3.3. Per – user caching

The basic idea behind per-user location caching is that the volume of HLR message traffic and database accesses required in locating a called subscriber can be reduced by maintaining local storage, or cache, of user location information at an MSC (Mobile Switching Center). One paper [4] discusses a per-user caching technique for 2G cellular networks in which at any MSC, location caching for a given user should be employed only if a large number of calls originate for that user from that MSC, relative to the user's mobility. The cached information is kept at the MSC from which calls originate, which may or may not be the MSC where the called user is currently registered.

The per-user caching strategy, attempts to reduce the network signaling and database loads of the basic strategies in exchange for increased CPU processing and memory costs. Since technology trends are driving the latter costs down, deploying the caching strategy on a system-wide basis will

become increasingly attractive. Once deployed, whether the caching strategy should be invoked for a particular user is a function of the user's mobility and communications patterns.

Identifying the classes of users for which the caching strategy yields net reductions in signaling traffic and database loads is another issue. The notion of a call-to-mobility ratio (CMR) of a user is the average number of calls to a user per unit time, divided by the average number of times the user changes registration areas per unit time is used often. Another value called a local CMR (LCMR), which is the average number of calls to a user from a given originating switch per unit time, divided by the average number of times the user changes registration areas per unit time [4].

For each user, the amount of savings due to caching is a function of the probability that the cached pointer correctly points to the user's location, and increases with the user's LCMR. The minimum value of LCMR for caching to be worthwhile is also discussed in [4]. This caching threshold is parameterized with respect to costs of traversing signaling network elements and network databases according to the authors and can be used, as a guide to select the subset of users to whom caching should be applied. These values are employed for the caching strategy used in this thesis.

In general, schemes for estimating the LCMR range from static to dynamic, and distributed to centralized. Two simple distributed algorithms for estimating LCMR, based on a long-range and short-range running calculation are given in [4]. An alternative approach is to utilize some user-supplied information, by requesting profiles of user movements, and to integrate this

with the caching strategy [4]. A variation of this approach is to use some domain knowledge about user populations and their characteristics.

A related issue is that of cache size and management. In practice it is likely that the monetary cost of deploying a cache may limit its size. In that case, cache entries may not be maintained for some users; selecting these users carefully is important to maximize the benefits of caching. The c