# An ANN Based Approach
# to
# Location Management

A Dissertation submitted to the Jawaharlal Nehru University
in partial fulfillment of the requirements
for the award of the degree of

**MASTER OF TECHNOLOGY**
**IN**
**COMPUTER SCIENCE AND TECHNOLOGY**

**BY**
**ASHVTOSH GOEL**

**Under The Supervision of**
**Prof. P.C. Saxena**

**SCHOOL OF COMPUTER AND SYSTEMS SCIENCES**
**JAWAHARLAL NEHRU UNIVERSITY**
**NEW DELHI-110067, INDIA**
**July, 2006**

जवाहरलाल नेहरू विश्वविद्यालय

**JAWAHARLAL NEHRU UNIVERSITY**

**School of Computer & Systems Sciences**

**NEW DELHI- 110067, INDIA**

# CERTIFICATE

This is to certify that the dissertation entitled "**An ANN Based Approach to Location Management**" being submitted by Mr. **ASHVTOSH GOEL** to the School of Computer and Systems Sciences, **Jawaharlal Nehru University**, New Delhi, in partial fulfillment of the requirements for the award of the degree of **Master of Technology** in **Computer Science and Technology**, is a record of bonafide work carried out by him under the supervision of Prof. P.C. Saxena.

This work has not been submitted in part or full to any university or institution for the award of any degree or diploma.

Prof. P.C. Saxena 21/7/06
(Supervisor)
SC&SS, JNU, New Delhi-67

Prof. S. Balasundram
Dean, SC&SS,
JNU, New Delhi-67

जवाहरलाल नेहरू विश्वविद्यालय

**JAWAHARLAL NEHRU UNIVERSITY**

**School of Computer & Systems Sciences**

**NEW DELHI- 110067, INDIA**

# DECLARATION

This is to certify that the dissertation entitled "**An ANN Based Approach to Location Management**" is being submitted to the School of Computer and Systems Sciences, Jawaharlal Nehru University, New Delhi, in partial fulfillment of the requirements for the award of the degree of **Master of Technology** in **Computer Science & Technology**, is a record of bonafide work carried out by me.

This work has not been submitted in part or full to any university or institution for the award of any degree or diploma.

Ashutosh Goel. / 21-07-2006

July 2006
JNU, New Delhi

ASHVTOSH GOEL
M.Tech, Final Semester,
SC&SS, JNU, New Delhi.

Acknowledgement

I take this opportunity to express my sincere thanks to my guide Prof. P.C. Saxena for his support, encouragement, advice and motivation. Without his valuable thoughts, recommendation and patience, I would have never been able to complete this work.

I would especially thank Amitabh Chandra, Rajeev Kumar Singh and Avijit Nandi for their insightful comments on my research and many long hours of intellectually stimulating discussions. I would also like to thank all my fellow students in Data Communication and Distributed Computing Group Lab for their help and support. Dinesh Singh Gaira, Kuldeep Khokhar, Dilip Kumar, and Rajesh Singh Vats had inspired me greatly through many interesting discussions and feedback.

Finally I am thankful to all my friends here in Jawaharlal Nehru University for making this period unforgettable and precious for me. I am grateful to have Vinod, Shiva, Ashok, Amit, Sharad and Rahul as friends who did countless favors to me. Without their friendship, care, and support, my life in the last two years would have been unendurable.

(Ashvtosh Goel)

# Table of Contents

**List of Figures**

## Abstract

Over the last few years, the worldwide cellular communication service market has realized an exponential growth. Decreasing prices, improved radio coverage, lightweight and compact terminals with high processing capabilities are the major area of improvements. The number of subscribers has been increased enormously. In order to accommodate higher subscriber densities, the standard technique used is to reduce the radio cell size since the available bandwidth for these services is almost same as it was few years back. The power dissipated through the mobile terminals can also not be increased due to biological reasons. Due to the anticipated increase in the usage of the wireless services in the future, the next generation of mobile networks should be able to support a huge number of users and their bandwidth requirements. To support this enormous traffic, the available frequency bands should be reused larger number of times what leads to smaller cell areas. Network configuration with smaller cell areas results to a situation where with a small mobility only, the mobile terminal will need to update its location very frequently. Reduction in cell size increases signaling overhead for location management procedure. Location management, in general, incorporates procedures with which system can locate particular mobile subscriber at any given time to deliver a call. Location management comprises of two basic tasks, Location update and Paging. Both of the tasks consume the available frequency band. Since frequency band is a scars resource and can not be increased much because of various reasons, there comes a need of reducing this signaling overhead.

The movements of a mobile subscriber are not at all random, and driven by its own characteristics. A subscriber usually roams within the city where he lives, especially the areas of his interest, like his home, his workplace etc. This property of the user is called locality of movement. Through a proper vigilance, it can be seen that almost every subscriber follows a pattern of movements. Neural network with its learning and generalization ability may act as a suitable tool to predict the location of a terminal provided it is trained appropriately by the personal mobility profile of individual user.

In this work, we have employed an MLP network to learn about the places where the subscriber spends most of its time during different time zones through error back-propagation. These time zones reflect the normal routine of almost every kind of subscriber and the normal call pattern. This approach is free from all unrealistic assumptions about the movement and call pattern.

# Chapter 1

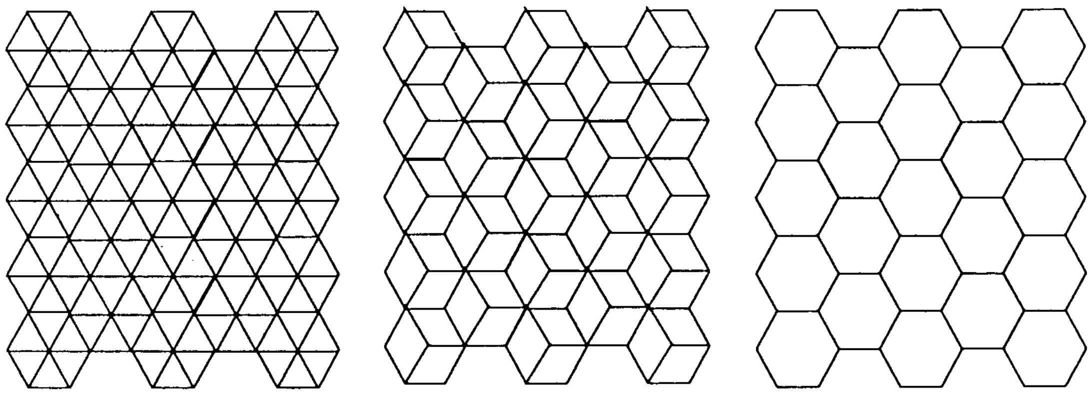
# Cellular networks and Personal Communication Services

The ability to communicate with people on the move has evolved remarkably since the demonstration of radio's ability to provide continuous wireless link with ships while sailing. The cellular radio network system is one amongst many communication systems which facilitate mobility in communication. Systems achieve mobility by making use of wireless network technology. Wireless networks use radio waves to transmit data. A cellular communication system provides a wireless connection to the Public switching Telephone Network for any user location within the coverage area of the cellular network.

## 1.1 Cellular Network Architecture

Cellular networks had been designed to provide communication link between two mobile stations or one mobile and one fixed station. Terminal mobility needs wireless access to the communication system. Obviously, the mobile stations participate in connection through wireless links. Mobile stations have the power limitation, so they use low power transmitters. One more limitation is of frequency, only a fixed band is provided for commercial public communication services. Since the mobile stations have low power transmitter, the geographical area is divided into numbers of smaller areas to make them capable to communicate with fixed wired network through wireless channels. Each of this area has its own antenna (what we call base station) to communicate with these mobile stations. These base stations are connected to the wired backbone communication network.

Now the question is that "what should be the strategy behind partitioning of this geographic area". The geographic area can be divided into triangular, parallelogram, or hexagonal areas as shown in figure 1.1. But since at base stations we have omni-directional antenna (Actually base stations consist of three directional antennas arranged at 120°, which gives an illusion of omni-directional antenna), which covers almost same distance in all directions, the shape of area should be circular (or similar to that). So the geographical area is divided into regular hexagonal areas (cells), though due to topographical limitations, precise hexagonal shape is not possible. It is of irregular shapes with overlapping in coverage areas. But considering cells as hexagons provide support for study of different concerns of cellular network.

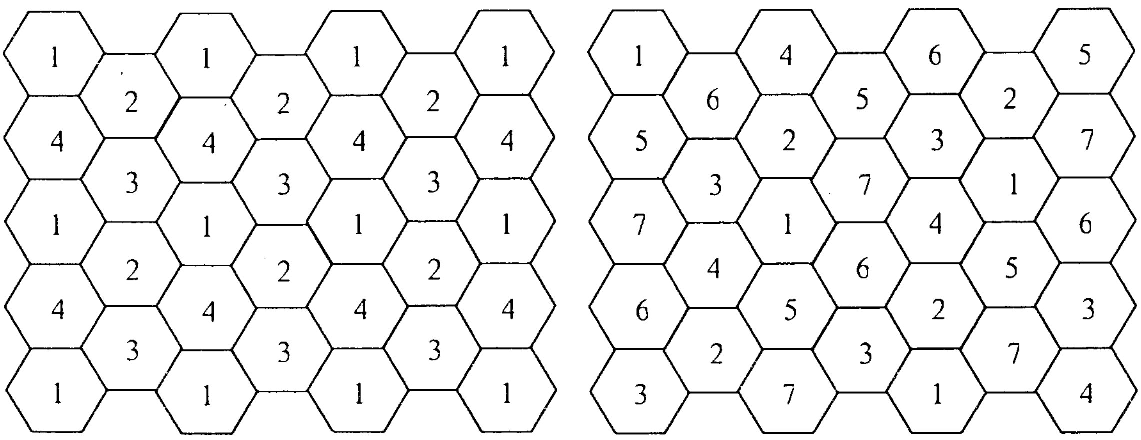


a. Triangular division b. parallelogram division c. hexagonal division

**Figure 1.1 Division of Geographical area**

### 1.2 Frequency reuse

Since the network has a fixed frequency bandwidth and number of users is enormously large, the reuse of frequency bands becomes necessary to provide connection to all the users simultaneously. The partitioned architecture of cellular network also provides support for frequency reuse. This reuse allows a cellular system to handle this huge number of users with a limited number of channels. Generally neighboring cells do not use the same frequency band for communication to avoid interference for the users located near cell boundaries. Available frequency is divided into different frequency bands and allocated to cells in a way that no two neighboring cell get the same frequency band. This can be done by dividing the total frequency into only four different bands (coloring of map theorem) but things are not so static. Some time the requirement of a cell can go high, so it borrows some frequency from some of its neighbor where the requirement is less. So this frequency partitioning should have one more constraint that no two neighbor of a cell should be allocated same frequency band. A frequency reuse pattern is a configuration of N cells, where N is the reuse factor, in which each cell uses a distinct frequency band. Figure 1.2 describes some reuse patterns. The frequency reuse factor is the rate at which the same frequency can be used in the network. Because of the increasing demand for wireless services, the available channels within the cells become insufficient to support the growing number of users. To increase the system capacity, techniques such as cell splitting and sectoring may be implemented. Using micro cells also improves cellular system capacity, and it is an attractive alternative to the two former mentioned techniques.



a. reuse factor = 4 b. reuse factor = 7

**Figure 1.2 Frequency Reuse Pattern**

### 1.3 Cellular Network Infrastructure

In cellular networks, the region of communication is tessellated into hexagonal cells with a BS, which is responsible to provide communication link at every point within the cell. The distance from the BS to any point on the cell boundary should not be longer than the range of the RF antenna used by the mobile stations and the base station. An MS needs to communicate with the base station of the cell where it is currently located and the BS acts as a gateway to the rest of the network. Each MS uses a separate, temporary radio channel available with the cell it is residing in, to talk to the BS. The BS talks to many Mobile stations at once, using one channel per mobile. Channels use a pair of frequencies for communication—one frequency (the forward link) for transmitting from the base station and one frequency (the reverse link) for the base station to receive calls from the users. Each BS uses digital techniques to enable a number of Mobile stations to be simultaneously connected to it, as well as simultaneously allowing a number of subscribers to make and receive calls. This sophisticated digital call-juggling ability is called *Multiplexing*. However, the combination of the tracking function and subscriber's unique digital signature, SIM number (Subscriber's Identification Module number) allows the MSC to route that call to the precise BS the MS happens to be connected to, and then exclusively to the MS - even if a number of other subscribers are simultaneously connected to that BS. A base station consists of a base transceiver system (BTS) and a base station controller (BSC). Tower and antenna are part of BTS and all associated electronics are part of BSC. These base stations are connected to Mobile telephone switching office (MTSO) or Mobile Switching Centre (MSC). These mobile switching centers are essentially end offices as in the telephone system and connected to telephone system end office. The MSC communicates with the base stations, other MSCs, and the *Public Switched Telephone Network* (PSTN) through a packet switching network. A usual cellular system is shown in figure 1.3.

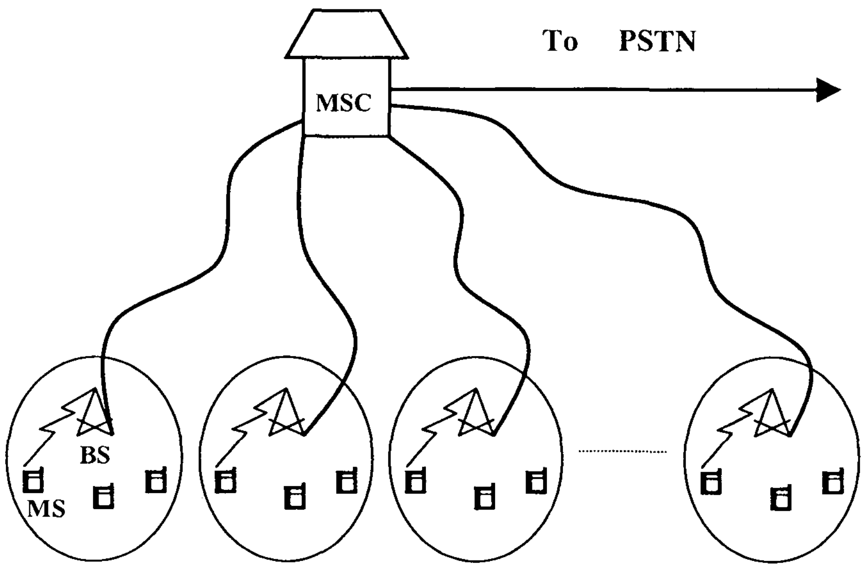



**Figure 1.3 Cellular System Architecture**

The MSC stores information about the subscribers located within the cluster and is responsible for routing, or switching, calls from the originator to the destination. It can be thought of managing the cell, being responsible for set-up, routing control and termination of the call and for collecting, charging and accounting information.

### 1.4 Call setup

When a user first asks for cellular services, she/he gets registered with some service provider. The service provider will provide a mobile identification number, what will be the user's identity to the network and with this number only the mobile station will be able to connect to the network. At the moment, a mobile station is turned on, before being able to make a call; it first scans the available forward control channels to determine the one with the strongest energy, and then monitors that control channel until the energy of that signal drops bellow a usable level. At this point it again scans the control channels in search of the strongest signal. Since the control channels are standardized and are identical throughout within the country or continent, every phone scans the same channels while idle.

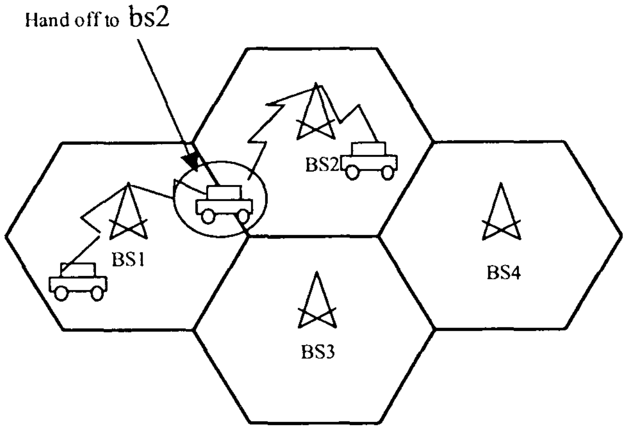
When an incoming call is placed to a mobile station, the MSC dispatches the request to all base stations in the cellular system. The mobile identification number is then broadcast as a paging message over all of the forward control channels throughout the cellular system. The mobile station receives the paging message sent by the base station that it monitors, and responds by identifying itself over the reverse control channel. The BS relays the acknowledgement sent by the MS and informs the subsequent MSC. Then the MSC instructs the base station to move the call to an unused communication channel within the cell. At the same moment, the base station signals the MS to change frequencies to an unused forward and reverse communication channel pair, and another data message called an alert is transmitted over the forward communication channel to inform the MS to answer the call. All of these actions occur within a few seconds and are not perceptible by the MS. Once a call is in progress, the MSC adjusts the transmitted energy of the MS and changes the channel of the MS and base stations in order to maintain call quality as the subscriber moves in and out of range of each base station, which is called *handoff.*

Now if a mobile station originates a call, it sends a call initiation request on the reverse control channel. This request consists of the caller's mobile identification number and the MIN of the user it want to call. The BS receives this request and sends it to the MSC. The MSC validates the request, makes connection to the called party through the PSTN, and instructs the BS and MS to move to an unused forward and reverse communication channel pair to allow the conversation to begin.

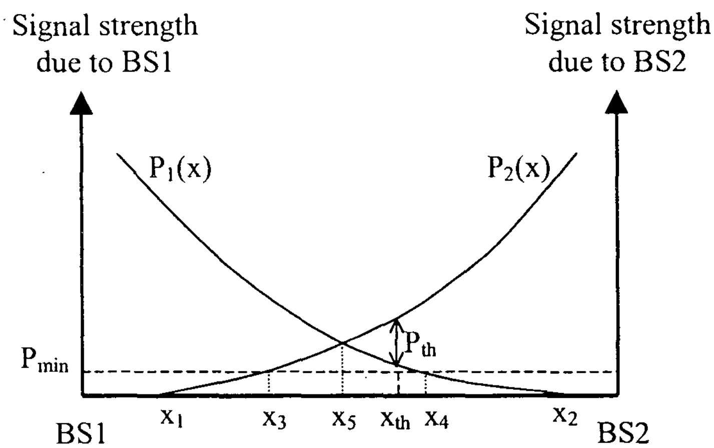
All cellular networks provide a service called roaming. It allows MS to operate in service areas other than the one from which it has subscribed with. When an MS enters an area that is different from its home service area, it get registered as a roamer in the new service area. Roaming is technically supported by mobility management, authentication and billing procedures. If a particular roamer has roaming authorization for billing purposes, roaming Mobile stations are allowed to receive and place calls from that area and billing is routed automatically to the subscriber's home service provider. If the visited network is in the same country as the home network, this is known as National Roaming. If the visited network is outside the home country, this is known as International Roaming or global roaming.

## 1.5 Handoff

Since a mobile station can move between the cells during a communication session, there comes a need of reconfiguration of the mobile host, wireless network and backbone wired network to support communication while an MS moves from one cell to another. This reconfiguration is termed as Handoff. Mobility is the most important feature of a wireless cellular communication system. Usually, continuous service is achieved by supporting handoff (or handover) from one cell to another. Handoff is the process of changing the channel associated with the current connection while a call is in progress. It is often initiated either by crossing a cell boundary or by deterioration in quality of the signal in the current channel. A handoff due to cell crossing phenomenon is shown in figure 1.4.




a.




b.

**Figure 1.4 Handoff in Cellular Network**

Here $P_i(x)$ shows the signal strength due to $BS_i$ at point x. After the point $x_5$ and before $x_{th}$, the signal strength due to $BS_1$ is weaker then signal strength due to $BS_2$, still it do not handed over to $BS_2$ because there is a fair chance that mobile subscriber return back towards $BS_1$ and that will cause excessive handoffs.

Handoff is divided into two broad categories—hard and soft handoffs. They are also characterized by "break before make" and "make before break", respectively. With hard handoff, the link to the prior base station is terminated before or as the user is transferred to the new cell's base station. That is to say that the mobile is linked to no more than one base station at a given time. Initiation of the handoff may begin when the signal strength at the mobile received from base station 2 is greater than that of base station 1. The signal strength measures are signal levels averaged over a chosen amount of time. A major problem with this approach to handoff decision is that the received signals of both base stations often fluctuate. When the mobile is between the base stations, the effect is to cause the mobile to wildly switch links with either base station. The base stations bounce the link with the mobile back and forth. Hence the phenomenon is called *ping- pong effect.* This ping-pong also occurs when a subscriber resides in a cell boundary area and receives fluctuating signals from both base stations. Besides ping pong this simple approach allows too many handoffs. It has been shown that much of the time the previous link was well adequate and that handoffs occurred unnecessarily. In hard handoffs, current resources are released before new resources are used; in soft handoffs, both existing and new resources are used during the handoff process. Soft handover has been introduced in the CDMA digital cellular standard. In networks following CDMA standard, base stations use same frequency band. Each MS has an identity based on a code. Since all base stations use same frequency band, no change in frequency or timing occurs as a mobile set passes from one base station to another, there are practically no dead zones. As a result, connections are almost never interrupted or dropped. This type of handoff is known as soft handoff.

### 1.6 Location management

Since Mobile stations are expected to move within the geographical area and should be provided with communication facilities everywhere within the coverage of cellular network, the network must know the exact position of mobile station for efficiently routing the call to particular MS. Delivering calls to the mobile station requires that the current location (point of attachment) of the mobile terminal be known in order for the network to route the calls to the mobile station.

Latency is the measure of the time taken to get the location information of mobile host. Location information of mobile host is to be obtained within certain time bounds. In order to effectively locate an MS when a call is initiated, location management schemes are used to keep track of the locations of the Mobile stations. It requires each MS to report its location to the network periodically. Update of location information causes data to be written at *location register*. Deregistration information causes location information to be removed from database. Both update and deregistration operations and lookup operations require access to the database. When a call is initiated, the network determines the current location of the called MS through a database lookup and paging procedure. This whole process of location register update (initiated by MS), database lookup and paging (initiated by network) is known as location management. Location management refers to accessing and maintaining user information for call routing purposes. Important per user information, such as current location, authentication information and billing information are stored in user profiles. Whenever a call needs to be delivered to the mobile station, the network uses the last known location of the mobile station to search for the mobile in the vicinity of that area. This may involve paging for the mobile terminal in certain neighborhood of the last known location of the mobile station. Strategy, which tries to reduce update cost, tends to increase search cost and strategy which tries to reduce search cost, tends to increase update cost. For example, if location of mobile host is updated frequently, then it can be searched more easily.

## 1.7 Personal Communication Services

Personal communication services (PCS) allow mobile users with wireless terminals to receive calls irrespective of their location in a seamless manner. In PCS network, MS communicates through portable handsets. Personal communication service networks provide its users a much more verity of services other than just voice cellular telephony. It provides a wide range of personal information services like personalized financial and stock market information, banking information, News clipping services, traveler information, as well as mobile shopping, video and multimedia, sales, inventory, and file access. Some of these services might involve only bursty network traffic, while others may require continuous connection-oriented network support. The assignment of a unique PCS telephone number to a customer will facilitate call initiation and completion across regional, national, and international borders. The network will do all the work of tracking a customer, knowing where he or she is, and facilitating a call through the nearest service site. There are many similarities in between cellular telephone service and PCS, but some advantages that PCS has over existing cellular telephone service are:

- better service quality through the use of digital technology
- more compact equipment using advanced antennas technology
- increased mobility support
- enhanced service features through the use of phone numbers not depending upon geographic location
- using digital technology makes price of PCS cheap

The main features of the PCS networks can be enumerated as follow:

- Multiple environments
- Multimedia services with high quality
- Multiple user types
- Global Roaming Capability
- Single personal telecommunication number
- Very high capacity
- Universal handset
- Security

PCS networks are supposed to provide these services to mobile stations anywhere, anytime in an uninterrupted and seamless way.

# Chapter 2

# Different approaches to Location management

## 2.1 Introduction

How to track mobile subscriber that moves from place to place in PCS network is one of the most important issues in personal communication networks. A communication system must track the location of its users in order to forward calls to the relevant cell within a network. A naive solution to this problem would involve users informing the network of their new location as they do transition between cells. Unfortunately such an approach is infeasible, with the subsequent update communication quickly overwhelming a network. Instead cells within a network are grouped into *Location Areas* (LAs). An LA is composed of one or several *cells* such that each cell is attached to at least one other cell in that location area. Users are free to move with a given location area without updating their location, informing the network only when transitioning to a new LA. If a call is to be forwarded to a user, the network must now page every cell within the location area to determine their precise location. Network cost is incurred on both location updates and paging, the balance of these defines the field of *Location Management.*

### 2.1.1 Location update Vs. Paging, the trade off

When a call arrives, the network needs to locate the cell in which the receiving mobile station is currently located; this process is known as terminal paging. In order to reduce paging costs, mobiles inform the network of their locations from time to time; this is known as location updating. There is a tradeoff between location update costs and paging costs. If a mobile expends power and bandwidth to update its location more often, it can reduce the area that needs to be paged when a call arrives. On the other hand, if updates are performed less frequently, more power and bandwidth will be expended on paging, since larger areas will need to be paged because of greater uncertainty of its position. The cost of mobility management over any given time period is the sum of the cost of the location updates and the cost of paging for the calls that arrive during that time period. Both paging and updates consume scarce resources like wireless network bandwidth and mobile equipment power, and each have a significant cost associated with it. The mechanism of paging involves the MSC sending a *page* message over a *forward control channel* in all the cells where the user is likely to be present. The mobile user listens to the page message and sends a response message back over a *reverse control channel*. The update mechanism involves the mobile user sending an update message over a reverse control channel, which may initiate a good amount of traffic and switching in the backbone network. The study of location management aims to reduce the overhead required in locating mobile devices in a cellular network.

### 2.1.2 Location Register

Location management strategies use a two tired database called *location register*, to store the information related to the location of each user information: the home location register (HLR), and the visitor location register (VLR). Every subscriber is registered with a home network, the HLR of which maintains the subscriber's current physical location. HLR maintains the profiles of all the users that are registered with the home network. When a mobile subscriber roams to another area, it has to register with the Visitor Location Register (VLR) of that area. VLR contains the information about all the users visiting that location area. The HLR maintains a pointer to the VLR which currently serves the mobile station. VLR supports registration, authentication, and call routing to/from a mobile while it is away from its home area. If the subscriber has roamed to another region then he/she has to register with the VLR that covers the new region. During registration, the VLR will contact the subscriber's HLR, and the HLR will update its database to reflect the new location of the subscriber. If the mobile has registered with some other VLR before, HLR will send a registration cancellation message to it.

### 2.1.3 Call delivery

An incoming call is routed to the called subscriber as follows. The dialed call is received by the MSC in the home system. This MSC is called the *originating MSC*. If the mobile host is currently being served by the originating MSC (i.e. the mobile host is not roaming), then this MSC queries the HLR to obtain the registration status and feature information of the mobile host. After receiving the response from the HLR, the originating MSC pages the mobile host. When the mobile host responds (i.e. subscriber accepts the call by pressing the proper button), the originating MSC sets up the circuit to continue with the call to the mobile host. When a mobile host is roaming in some other location area and a call to that mobile is dialed, as illustrated in figure 2.1, the call is first routed to the originating MSC. The originating MSC then sends a location request message to the HLR to find out the current location of the mobile. The HLR, in turn, sends a route request message to the VLR that is currently serving the mobile. The VLR then sends a route request message to the MSC that is currently serving the mobile. The serving MSC creates a Temporary Location Directory Number (TLDN) and returns it to the VLR. The TLDN is then passed back to the originating MSC through the HLR. The originating MSC then routes the call using this TLDN. When the serving MSC receives the call routed using the TLDN, it pages the mobile host. If the mobile responds, then the call is established at the mobile.

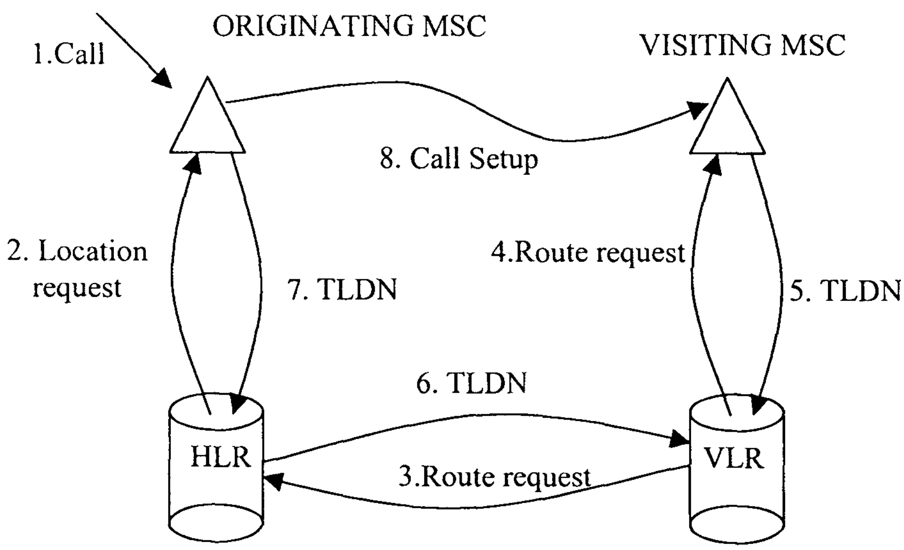



**Figure 2.1 Call delivery to an MS roaming in some other LA**

With mobile user subscription increasing rapidly and a continually restricted available spectrum, the problem of effective location management has become one of prime significance. The main approaches to location updating and paging, will be discussed here.

## 2.2 Location updating schemes

A location update is used to inform the network of a mobile device's location. This requires the device to register its new location with the current base station, to allow the forwarding of incoming calls. There are different location update strategies according to the maximum allowed latency. Each location update is a costly exercise, involving the use of cellular network bandwidth and core network communication; including the modification of location registers. Hence a wide variety of schemes have been proposed to reduce the number of location update messages required by a mobile station in a PCS network.

### 2.2.1 Always-update vs. Never-update

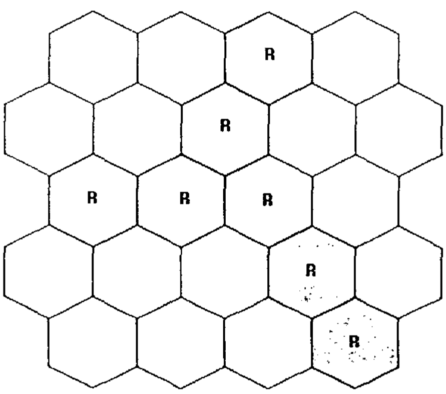
The always-update strategy is the simplest location update scheme. This can be used when the call is to be delivered with minimum latency. Under this scheme, the subscriber performs a location update every time the user moves into a new cell. The network always has complete knowledge of the user's location and requires to paging to locate the user when an incoming call arrives. This scheme reduces the paging cost but results in a large number of update signals. The always-update scheme performs well for users with a low mobility rate or high call arrival rate. It performs quite poorly for users with high mobility however, requiring many location updates and excessive use of resources. While not used in practice, this scheme forms the basis of many more complex location management mechanisms.

The never-update scheme is the logical counterpart to always-update, never requiring the mobile device to update its location with the network. This entails no location update cverhead but may result in excessive paging for a large network or for high call arrival rates. These two strategies represent the extremes of location management - always-update minimizing paging cost with never-update minimizing location update cost.

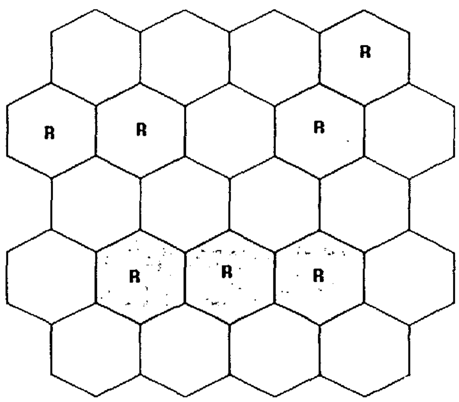
The two extremes are often combined to form a more comprehensive location management strategy, catering for differences in user and network characteristics.

### 2.2.2 Reporting Cells

Under this location update scheme, the mobile device updates its location only when visiting one of a set of predefined *reporting cells*. This scheme does not depend on the individual movement characteristics of the user, but simply the arrangement of these reporting cells in the network. When an incoming call arrives for the mobile device, a search must be conducted around the vicinity of the last reporting cell from which the user updated their location. The reporting cell topology may be bounded or unbounded. The unbounded approach requires a smaller number of reporting cells, in turn reducing the number of redundant location updates. Both of the approaches are shown in figure 2.2.




a. bounded approach




b. unbounded approach

**Figure 2.2 Reporting cell configuration**

The unbounded approach requires an intelligent paging scheme to handle the unbounded search space. The performance gain, achievable with a reporting cell topology is somewhat limited. Without direct consideration of the movements of users, it is not possible to assign reporting cells in an optimum arrangement. Even with the knowledge of the network, the selection of an optimum set of reporting centers is an NP-complete problem. Two extremes of poor performance are possible, owing to the unpredictability of user movement. When a mobile device continually passes over a reporting cell, excessive location update messages are generated, introducing a high level of overhead. Conversely, if a mobile device does not roam into any of the reporting cells, its location will not ever be updated and hence incur a high paging load.

### 2.2.3 Location Areas

The location area topology is widely used to control the frequency of location updates in personal communication networks. Here the network is partitioned via the amalgamation of groups of cells into larger meta-cells, or *location areas*. The scheme then functions very similarly to the always-update mechanism - mobile devices only update their location when leaving their current location area. This partitioning is shown in Figure 2.3, with three separate location areas. If the network knows the users current location area, the paging required to locate a user is confined to the meta-cell within which the user resides.

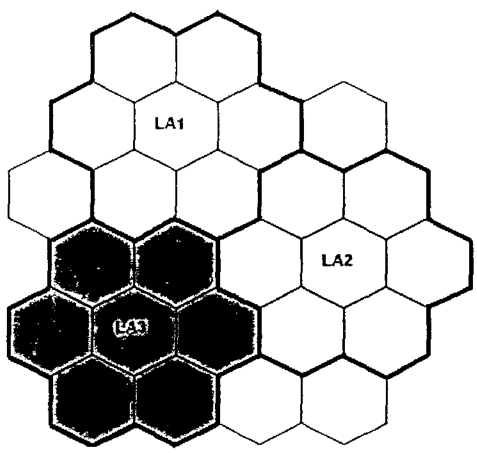



**Figure 2.3 Network partitioned into location areas**

The location update process may be instigated periodically or, more commonly, on location area boundary crossings. The periodic location updating scheme is the simplest to implement, merely requiring the mobile device to send a location update message containing its current location area at regular time intervals. The methodology used here captures none of the details of user movement however, and enforces an update frequency that may be highly unsuitable given the users current rate of movement. It also offers no guarantees that the network will have knowledge of the exact location area the user resides in, requiring sophisticated paging across multiple location areas. The boundary crossing method is more precise; updating the user location only when crossing to another location area. This method has its own inherent weaknesses however, particularly when a user moves repeatedly between the boundaries of two or more location areas, inducing a high location update rate with comparatively low physical mobility.

### 2.2.4 Time-based Update

The time-based strategy of location update requires that mobile stations update their location at predefined constant time intervals. This scheme only requires the mobile device to maintain a simple timer, allowing efficient implementation and low computational overhead. But if the user is moving rapidly, then even in small time duration it can move to a far distance. In between two updates, network will lose the location of that mobile subscriber. On the other hand, if the time duration between two location updates get reduced, and user is not highly mobile then it is wasting the network resources by these frequent update messages. Both of the situation is illustrated in Figure 2.4, with the updates (U1, U2, U3 and U4) performed at each time interval Δt, regardless of individual movements.

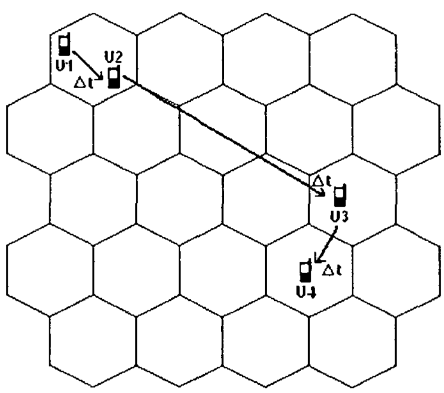



**Figure 2.4 Time-based location update**

### 2.2.5 Movement-based Update

The movement-based update scheme requires mobile station to count the number of boundary crossings to other cells in the network and to update their location when this number reaches a threshold value. In this case network does not know the exact cell in which the mobile station is, so it pages the cell where the mobile station has updated its location and surrounding cells according to that threshold value. This boundary-crossing threshold can be optimized for individual movements by assigning a per-user value, according to the call arrival rates. Figure 2.5 shows a movement-based scheme, with a movement threshold of two. Here the device updates its location every two crossings between cells.

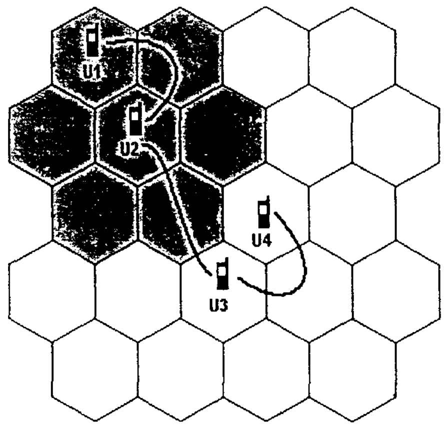



**Figure 2.5 Movement-based location update**

The required paging area is restricted to a neighborhood of radius equal to the distance threshold around the last updated location. The shaded cells shown in figure would be paged for a call after update 2. The paging area requirement is reduced through this scheme, although unnecessary updates may still be performed as a result of repeated crossings over the same cell boundary.

### 2.2.6 Distance-based Update

In a distance-based scheme the mobile device performs a location update when it has moved a certain distance from the cell where it last updated its location. It is an improved version of movement based location update scheme.

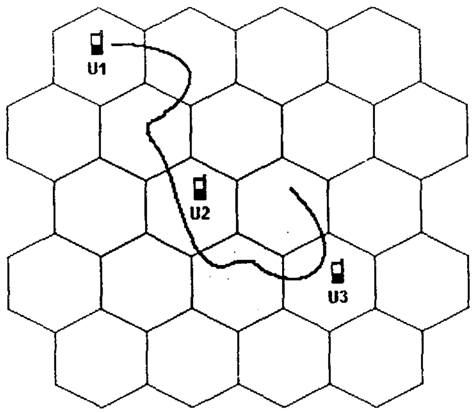



**Figure 2.6 Distance-based location update**

Under this scheme, the network knowledge about the location of mobile station is same (for same threshold value of distance) as movement based update scheme but the number of update messages are smaller then that. The situation is illustrated in figure 2.6. The shaded cells shown in figure would be paged for a call after update 2. Again, this distance threshold may be optimized per-user according to movement and call arrival rates. This scheme has the benefit of not requiring an update when a user repeatedly moves between a small subset of cells, provided these cells reside within the distance-threshold radius.

### 2.3 Paging Strategies

While mobile devices perform updates according to their location update scheme, the network needs to be able to precisely determine the current cell location of a user to be able to route an incoming call. This requires the network to send a paging query to all cells where the mobile device may be located, to inform it of the incoming transmission. It is desirable to minimize the size of this paging area, to reduce the cost incurred on the network with each successive paging message. Ideally the paging area will be restricted to a known group of cells, such as with the currently implemented location area scheme. An optimum paging area size calculation involves a trade-off between location update cost and paging cost. This technique is used in many location management schemes to reduce the location management costs incurred. The most commonly used paging schemes are summarized below. These have seen extensive use in real-world telecommunications networks

### 2.3.1 Simultaneous Paging

The simultaneous paging scheme, also known as blanket paging, is the most general scheme. Under this scheme, network does not keep any information other then the location register. All cells in the user's location area are paged simultaneously, to determine the location of the mobile device. This requires no additional knowledge of user location but may generate excessive amounts of paging traffic with high call rate and large number of users. Implementations of simultaneous paging favor networks with large cells and low user population and call rates. This scheme does not scale well to large networks with high numbers of users, necessitating the development of more advanced paging techniques.

### 2.3.2 Two-step Paging

Under this paging scheme, we assume that there is a high chance of user being in the same cell where it updated its location most recently. So the network keeps track of the last cell where the mobile station has updated its location.

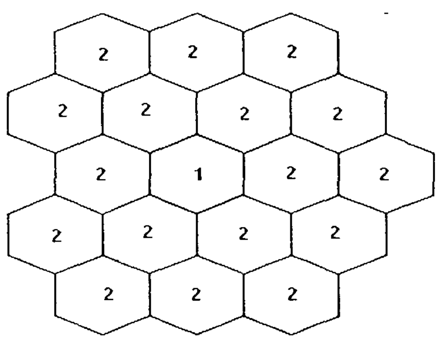
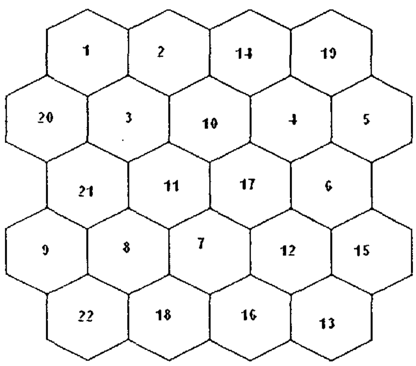



**Figure 2.7 Two-step Paging**

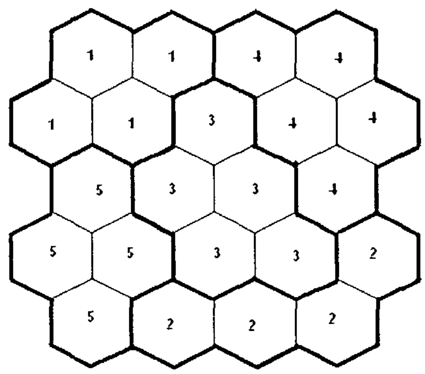
When a call needs to be delivered to a subscriber, network will page that only cell first. If subscriber is in that cell, it will respond within a time limit otherwise network will page all other cells in that location area. This scheme reduces the signaling overhead but with an increased average latency time for delivering a call.

### 2.3.3 Sequential Paging

Sequential paging avoids paging every cell within a location area by segmenting it into a number of paging areas, to be polled one-by-one. It is found in that the optimal paging mechanism, in terms of network utilization, is a sequential poll of every cell in the location area individually, in decreasing probability of user residence. The order by which each area is paged is central to the performance of the sequential paging scheme. This scheme results in a less paging overhead but with a noticeable increased latency. Several methods had been suggested to determine the ordering of paging areas in a sequential scheme. Some of them are shown in figure 2.7.

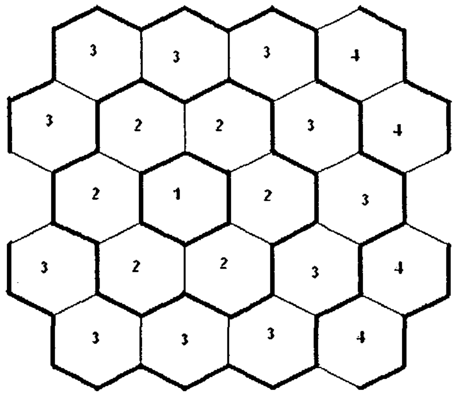



a. Random Order Paging




b. Sequential Group Paging






c. Sequential Ring Paging

**Figure 2.8 Sequential Paging Schemes**

The simplest ordering constraint is a purely random assignment, where each paging area is polled in a random order. While this reduces the total number of polling messages over a blanket scheme, it is far from optimal. The individual delays incurred in this scheme may be unacceptable however, and hence it is suggested that paging areas are formed from a larger number of cells, the scheme is called sequential group paging. The number of cells per paging area is a factor which needs to be optimized and may lead to excessive call delays, particularly in large networks. Sequential Ring Paging scheme favors paging the areas located geographically closer to the previously updated location. It has been found that it further reduces the total number of paging messages required. This scheme necessitates knowledge of the geographical structure of the network however, and may perform poorly for high movement rates.

## 2.4 Commonly used Mobility Models

Various mobility models have been discussed to model a user's motion patterns. Some of those are:

- Random walk: The random walk mobility model is regularly used to model the movements of users in a cellular network. It assumes that the direction of each user-movement is pure random, and hence each neighboring cell may be visited with equal probability. This model is easily implemented as it requires no state information to predict the next cell occupied by a user.
- Fluid flow: The fluid flow model views the system from a macroscopic perspective, representing the aggregate movement patterns of users. While this model can produce accurate representations of average boundary crossings per unit area, it does not consider the movement of individual users. This system-wide approach is hence ideal for optimizing total network utilization but not appropriate from a user perspective, considering only aggregate rates.
- Markovian Model: the Markovian mobility model defines distinct probabilities for movement from a given cell to each of its neighbors. This scheme is based on

the assumption that a user moving in a given direction will continue in same direction with greater probability than a divergence from the course.

## 2.5 Research Motivation

A wide variety of location management schemes have been proposed to reduce the cost of locating mobile devices in a cellular network. While reducing the cost involved in an idealized theoretical environment, they commonly make the assumption that patterns of call arrival and user's movement within the network is random. Such assumptions simplify the analysis and development of location management schemes significantly, yet are often not an accurate reflection of real-world cellular networks. In cellular communication networks, most of the subscribers follow regular routines during office hours, residing mostly at their place of work. These subscribers usually roam in some specific cells of their interest. If a proper vigilance is made for some amount of time then the location of a user can be predicted. This property is called *locality of movement.* For these subscribers, it is possible to predict their location at a particular time of day with a significant accuracy. Also the probability of initiation of a call is high in day time whether after the midnight and before early morning, it is very small.

In current era, since the mobile terminals are capable enough in terms of memory and processing, the main issue is reduction of the use of available channel capacity in location update tasks. Today mobile networks are characterized by high user density and high mobility and also the decreasing cell sizes which will increase the number of location updates and hand-off messages, thus limiting the available bandwidth.

Profile based schemes have been proposed to utilize the users profile and history for selective paging. A per user sequential paging has also been proposed where for an individual user, there will be a list of cells attached to its location register entry, which is according to the number of times user visited those cells and time spent there in last few days. All the above-discussed strategies reduce the location management cost but still lacking in terms of intelligence.

We will use an artificial neural network (ANN) to learn about the user's regular routines and mobility pattern. The main motivation behind the use of ANN is their ability to learn relationships between complex data sets that may not be perceived easily by humans. Deploying ANN for location prediction will give a relatively accurate information about the location of a mobile terminal with even a very few location update messages.

# Chapter 3

# ANN based approach to Location Management

## 3.1 Introduction

The human brain is subdivided into five major anatomical units called the cerebrum (or cortex), cerebellum, midbrain, pons and medulla. If brain tissue is properly stained and examined under a light microscope, it is possible to visualize the individual neurons, which compose the brain. The average human brain consists of $1.5 \times 10^{10}$ neurons of various types. A biological neuron has three types of components: dendrites, soma and axons. A biological neuron is shown in figure 3.1.

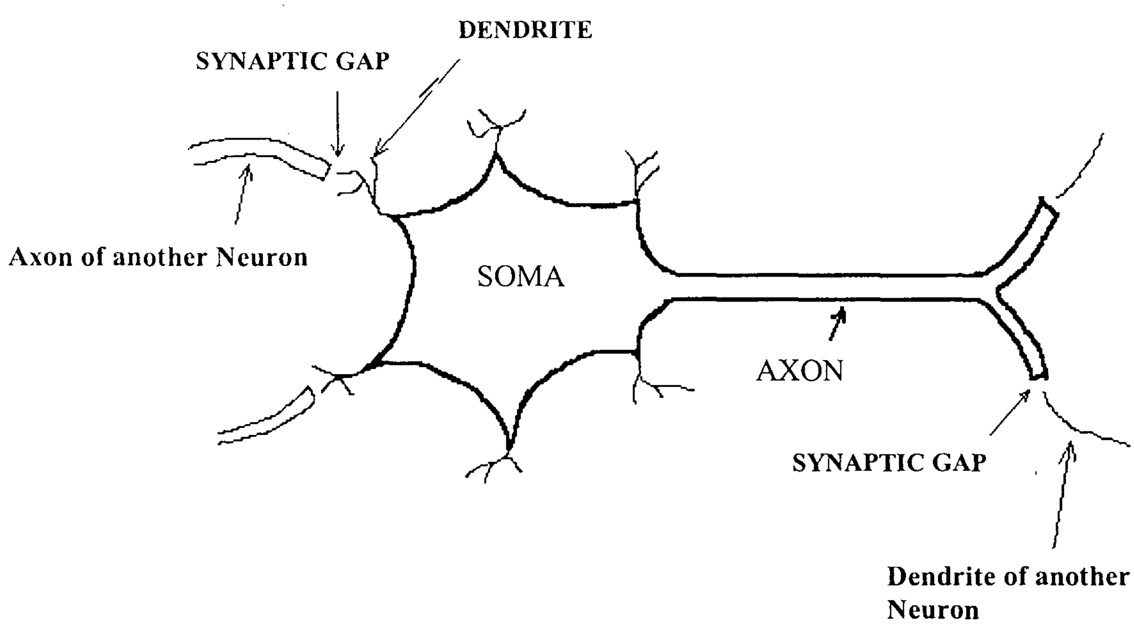



**Figure 3.1 Biological Neuron**

The many dendrites receive signals from other neurons through specialized structures called synapse, what is a connection between different neurons. The soma or cell body sums the incoming signals. When adequate input is received, the cell fires; that is, it transmits a signal over its axon to other cells. A single neuron can receive hundreds of thousands of input lines and may send its output to a similar number of other neurons. Synapses have a number of different forms, but have two basic varieties: inhibitory synapses-which make the neuron receiving them less likely to fire action potentials; and excitatory synapses-which make it more likely that the neuron receiving them will fire.

An Artificial Neural Network (ANN) is an information-processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information. The key element of this paradigm is the novel structure of the information processing system. It is composed of a large number of highly interconnected processing elements (artificial neurons) working in accord to solve specific problems. An artificial neuron has three components.

1. Synaptic weights over connection links, which tells the strength of the link. The signals connected to those links are multiplied by the synaptic weights.
2. A summation function for summing up the input signals.
3. An activation function to decide the behavior of neuron, whether it will fire or not.

A typical artificial neuron is shown in figure 3.2.

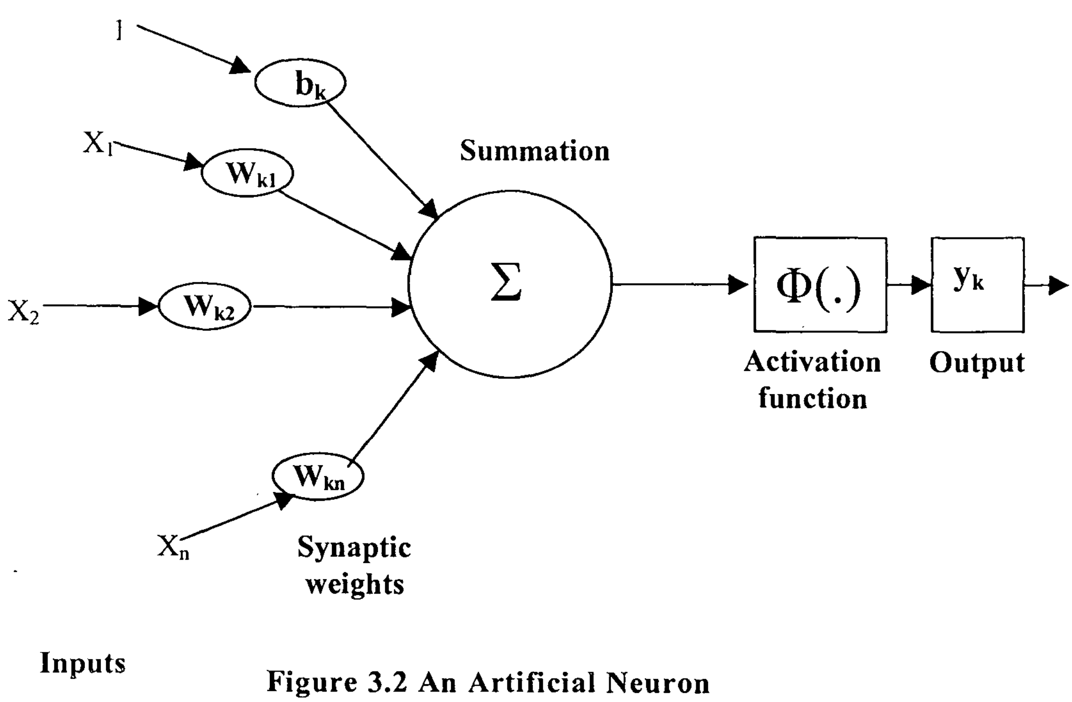



**Figure 3.2 An Artificial Neuron**

Here X is input vector to $K_{th}$ neuron, while $W_{ki}$ is weight vector. Mathematically for $K_{th}$ neuron

$$Y_k = \phi(V_k)$$

$$Where \quad V_k = \sum_{i=1 \to n} X_i \bullet W_{ki} + b_k$$

$b_k$ is the bias for neuron K.

ANNs, like people, learn by example. An ANN is configured for a specific application, such as pattern recognition or data classification, through a learning process. Learning in biological systems involves adjustments to the synaptic connections that exist between the neurons. ANN also learns in a similar way. Artificial neural networks have been developed as generalizations of mathematical models of human cognition or neural biology.

A neural net consists of a large number of simple processing elements called *neurons*, or nodes. Each neuron is connected to other neurons by means of directed communication links, each with an associated weight what represents the information being used by the net to solve any problem. Each neuron has an internal state; called its activation level, which is a function of the inputs, it has received. This function is called its activation function. Figure 3.2 represents a typical neuron. Neural nets can be applied to various problems, such as storing and recalling data or patterns, classifying patterns, performing general mapping from input patterns to output patterns or finding constrained optimization problems.

### 3.2 Activation functions

The basic operation of an artificial neuron involves summing its input signals and applying an output, or activation function. Typically, the same activation function is used for all neurons in any perticular layer of a neural net. Activation function defines the output of a neuron. Three basic types of activation functions are

1. Identity function

$$\phi(x) = x \forall x$$

2. binary step function(with threshold θ)

$$\phi(x) = \begin{cases} 1 & \text{if } x >= \theta \\ 0 & \text{if } x < \theta \end{cases}$$

3.sigmoid function

a. binary sigmoid

$$\phi(x) = \frac{1}{1 + \exp(-ax)}$$

b. bipolar sigmoid

$$\phi(x) = \frac{1 - \exp(-ax)}{1 + \exp(-ax)}$$

Where a is the parameter. With varying 'a' we can obtain sigmoid function with deffrent slopes.

### 3.3 Multi-layer Perceptron

Multi-layer perceptrons (MLPs) are feed forward neural networks trained with the standard back propagation algorithm. They are supervised networks so they require a desired response to be trained. They learn how to transform input data into a desired response. With one or two hidden layers, they can approximate virtually any input-output map. The Multi-layer perceptron is the most widely used type of neural network. It is both simple and based on solid mathematical grounds. Input quantities are processed through successive layers of neurons. There is always an input layer, with a number of neurons equal to the number of variables of the problem, and an output layer, where the perceptron response is made available, with a number of neurons equal to the desired number of quantities computed from the inputs. The layers in between are called "hidden" layers. With no hidden layer, the perceptron can only perform linear tasks (for example a linear discriminant analysis, which is also very useful). The neurons in the hidden layers do the actual processing, while the neurons in the input and output layers merely distribute and collect the signals. Although many hidden layers can be used, it has been shown that an MLP with one hidden layer can approximate any continuous function. Each neuron of a layer other than the input layer computes first a linear combination of the outputs of the neurons of the previous layer, plus a bias. The coefficients of the linear combinations plus the biases are called the weights. They are usually determined from examples to minimize, on the set of examples, the (Euclidean) norm of the desired output - net output vector. Neurons in the hidden layer then compute a non-linear function of their input. This nonlinear activation function should be continuos and its non-linearity should be smooth i.e. differentiable everywhere. The presence of non-linearity is important to map the existing non-linearity of the relations between input and output patterns. The MLP network is trained by adapting the synaptic weights through *Error Back-Propagation algorithm*. The network will be provided with trainig examples (input and output vectors). During training the network output is compared with desired output. The error between these two signals is used to adapt the weights. This rate of adaptation is controlled by the *learning rate*, denoted by $\eta$. A high learning rate will make the network adapt its weights quickly, but will make it potentially unstable, only getting a local vision. Setting the learning rate to zero makes the network keep its weights constant, i.e. network will not learn anything. If the learning rate is very small, then the algorithm proceeds slowly, but accurately follows the path of steepest descent in weight space. Training an MLP network by back propagation involves three stages.

- The feed forward of the example input pattern
- The back propagation of the associated error with respect to desired output
- The adjustment of the weights

### 3.4 Learning through error back-propagation

An MLP network with one hidden layer is illustrated in figure 3.3. The MLP net will be train with input training vector $X=(x_1, x_2, x_3, \ldots x_l)$, and the corresponding target output vector $T=(t_1, t_2, t_3, \ldots t_n)$. The network will calculate the output vector $Y=(y_1, y_2, y_3, \ldots y_n)$ according to its synaptic weights and activation function $\Phi$.

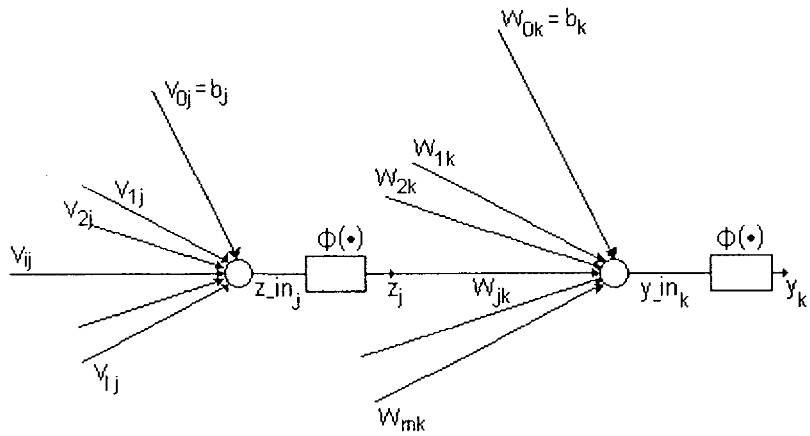



**Figure 3.3 MLP Network**

The error signal for output neuron k on providing pth training pattern is defined by

$$e_k(p) = t_k(p) - y_k(p) \qquad (3.1)$$

at this point, the instantaneous value of the error energy for neuron k is defined as ½ ek2(p), and total error energy over the output neurons ε(p) will be sum of error energies of all the output neurons.

$$\varepsilon(p) = \frac{1}{2}\sum_{k \in N} e_k^2(p) \qquad (3.2)$$

Where N is the set of all output neurons.

The instantaneous error energy ε(p) is a function of all free parameter like synaptic weights, bias of the network. The main purpose of training the network is to reduce this error energy by adjusting the free parameters of the network. Error back-propagation algorithm updates the weights on a pattern by pattern basis for a complete set of input trainig vector that is called one epoch. This adjustment of the weights will be calculated for each training input pattern and corresponding error.

According to delta rule, for an output neuron, this adjustment applied to weight matrix Wjk will be

$$\Delta w_{jk}(p) = -\eta \frac{\partial \varepsilon(p)}{\partial w_{jk}(p)} \qquad (3.3)$$

Where η is the learning rate of the network. Use of the minus sign is to seek a directionfor weight change that reduces the value of error energy. At an output neuron k, the induced local field $y\_in_k(p)$ produced at the input of activation function will be

$$y\_in_k(p) = \sum_{j=0\,to\,m} w_{jk}(p).z_j(p) \qquad (3.4)$$

Where m = no. of neurons in hidden layer.

And the output of neuron k will be

$$y_k(p) = \phi(y\_in_k(p)) \qquad (3.5)$$

now, to calculate the $\Delta w_{jk}(p)$, we need to calculate the sensitivity factor i.e. $\frac{\partial \varepsilon(p)}{\partial w_{jk}(p)}$.

According to chain rule of calculas

$$\frac{\partial \varepsilon(p)}{\partial w_{jk}(p)} = \frac{\partial \varepsilon(p)}{\partial e_k(p)} \bullet \frac{\partial e_k(p)}{\partial y_k(p)} \bullet \frac{\partial y_k(p)}{\partial y\_in_k(p)} \bullet \frac{\partial y\_in_k(p)}{\partial w_{jk}(p)} \qquad (3.6)$$

now partially differentiating $eq^n$ 3.2 w.r.t. $e_k(p)$ we get

$$\frac{\partial \varepsilon(p)}{\partial e_k(p)} = e_k(p) \qquad (3.7)$$

partially differentiating $eq^n$ 3.1 w.r.t. $y_k(p)$ we get

$$\frac{\partial e_k(p)}{\partial y_k(p)} = -1 \qquad (3.8)$$

and partially differentiating $eq^n$ 3.5 w.r.t. $y\_in_k(p)$ we get

$$\frac{\partial y_k(p)}{\partial y\_in_k(p)} = \phi'(y\_in_k(p)) \qquad (3.9)$$

And finally partially differentiating $eq^n$ 3.4 w.r.t. $w_{jk}(p)$ we get

$$\frac{\partial y\_in_k(p)}{\partial w_{jk}(p)} = z_j(p) \qquad (3.10)$$

Putting all these values from $eq^n$s 3.7 to 3.10 into 3.6 we get

$$\frac{\partial \varepsilon(p)}{\partial w_{jk}(p)} = -e_k(p) \bullet \phi'(y\_in_k(p)) \bullet z_j(p) \qquad (3.11)$$

So the value of update to the weight $w_{jk}$

$$\Delta w_{jk}(p) = \eta \bullet e_k(p) \bullet \phi'(y\_in_k(p)) \bullet z_j(p) \qquad (3.12)$$

or $$\Delta w_{jk}(p) = \eta \bullet \delta_k(p) \bullet z_j(p) \qquad (3.13)$$

Where $\delta_k(p) = e_k(p) \bullet \phi'(y\_in_k(p))$ (3.14)

Here $\delta_k(p)$ is called local gradient which is physically defined as

$$\delta_k(p) = -\frac{\partial \varepsilon(p)}{\partial y\_in_k(p)} \qquad (3.15)$$

this local gradient denotes the required change in synaptic weights which is equal to the product of error signal $e_k$ and the derivative of the activation function.

When we consider a neuron j in hidden layer, we do not know the value of desired output of that neuron. So the error for a hidden neuron can not be calculated directly, as for an output neuron. So it should be calculated in terms of the errors of all the neurons in output layer to which this hidden neuron is connected. According to $eq^n$ 3.15

$$\delta_j(p) = -\frac{\partial \varepsilon(p)}{\partial z\_in_j(p)} \qquad (3.16)$$

Since at hidden neuron j, the induced local field $z\_in_j(p)$ produced at the input of activation function will be

$$z\_in_j(p) = \sum_{i=0\,to\,l} v_{ij} . x_i(p) \qquad (3.17)$$

Where l = total number of neurons in input layer

By chain rule, eq$^n$ 3.16 can be written as

$$\delta_j(p) = -\frac{\partial \varepsilon(p)}{\partial z_j(p)} \bullet \frac{\partial z_j(p)}{\partial z\_in_j(p)} \tag{3.18}$$

or,
$$\delta_j(p) = -\frac{\partial \varepsilon(p)}{\partial z_j(p)} \bullet \phi'(z\_in_j(p)) \tag{3.19}$$

Since

$$z_j(p) = \phi(z\_in_j(p)) \tag{3.20}$$

Partially differentiating eq$^n$ 3.2 w. r. t. $z_j(p)$ we get

$$\frac{\partial \varepsilon(p)}{\partial z_j(p)} = \sum_k e_k(p) \bullet \frac{\partial e_k(p)}{\partial z_j(p)} \tag{3.21}$$

or,
$$\frac{\partial \varepsilon(p)}{\partial z_j(p)} = \sum_k e_k(p) \bullet \frac{\partial e_k(p)}{\partial y\_in_k(p)} \bullet \frac{\partial y\_in_k(p)}{\partial z_j(p)} \tag{3.22}$$

but from eq$^n$ 3.1 and 3.5

$$e_k(p) = t_k(p) - \phi(y\_in_k(p)) \tag{3.23}$$

so,
$$\frac{\partial e_k(p)}{\partial y\_in_k(p)} = -\phi'(y\_in_k(p)) \tag{3.24}$$

Now, partially differentiating eq$^n$ 3.4 w. r. t. $z_j(p)$, we get

$$\frac{\partial y\_in_k(p)}{\partial z_j(p)} = w_{jk}(p) \tag{3.25}$$

Putting the value of eq$^n$ 3.24 and 3.25 into eq$^n$ 3.22

$$\frac{\partial \varepsilon(p)}{\partial z_j(p)} = -\sum_k e_k(p) \bullet \phi'(y\_in_k(p)) \bullet w_{jk}(p) \tag{3.26}$$

or
$$\frac{\partial \varepsilon(p)}{\partial z_j(p)} = -\sum_k \delta_k(p) \bullet w_{jk}(p) \tag{3.27}$$

so putting the value of eq$^n$ 3.27 into eq$^n$ 3.19

$$\delta_j(p) = \phi'(z\_in_j(p)) \bullet \sum_k \delta_k(p) \bullet w_{jk}(p) \tag{3.28}$$

It is local gradient of hidden neuron j and will be used to correct the synaptic weights of inputs of hidden layer neurons. Since back-propagation learning relies upon local gradients to reduce the error energy, some times it results to local minima of that. To overcome this problem, a slow learning rate with repeated epochs is suggested while training the network.

## 3.5 Problem Mapping

The problem of location management using ANN is to predict the user's movement pattern and to learn the most likely areas where it spends maximum of its time, with help of the artificial neural network. MLP networks have been found helpful in learning how to transform input data into a desired response under a supervised training. In the present work, an MLP network is employed to predict the user's mobility patterns through a proper vigilence of user during training period. Thereafter the same network will be used to generate the profile of particular user. The number of entries in profile of a user will be decided considering the maximum allowed delay to setup a call.

### 3.5.1 Mapping considerations

While mapping the problem of location management, it is being considered that for the whole day, the user mobility and call pattern are not same. Normaly a person lives at one place, work at a fixed workplace, regularly goes to its workplace following a same path in morning on weekdays, follows a particular path to return to home, shop in some specific markets etc. The probability of movement of user is very less in late night and early morning hours, while it is very high in morning and evening time. A user spends most of his time at home or at his workplace. The mobilty pattern on weekdays will be different from that on weekends. Taking care of these considerations, we have divided the days as weekdays and weekends, and again a whole day into different time zones like early morning, morning, day time, evening and late night. If an MS has last updated its location at some particular time, probability of that MS being there is very high within a certain time limit $t_{th}$

from that time instance. The probability of getting a call in the morning and evening time is greater then the probability of same in day time which is again far greater then the probability of getting a call in late night and early morning hours. Considering this situation, we have given the preferences to the locations where he spend time according to the probability of getting a call in that time zone. Our input vector will provide information about these time zones with the knowledge of type of the day, whether weekday or weekend. During training period, network will monitor the movement pattern of the subscriber with help of user's mobile set. Network will be trained according to this pattern by error backpropagation learning algorithm. The MLP net will produce an output vector which should represent the cell id where the user may present. According to this output, the error will be calculated and fed back to previous layer to adjust the weights.

During training period, the system will employ a standard location management scheme (here we are taking distance based scheme for location update and sequential ring paging scheme). Since a user can change its mobilty pattern during a large time span, say one or two years (suppose he shifted his home or changed his workplace), so a user pattern should be learned again. For this we have taken a synchronized flag at both MS and BS which shows whether the user pattern should be learned or not. Learning computations will take place in late night hours, considering that in this time zone, there is a very less probability of initiation of a call.

### 3.5.2 Neural Model

Here we will be taking an MLP network with a single hidden layer with n input neurons, which will represent the time zones of a particular day. There will be as many neurons in output layer as numbers of cells in the coverage of an MSC. Number of neurons in hidden layer will be selected arbitrarily and decided later according to performance of the network.

When a subscriber is being learned, its history will be used to decide one epoch of training. Location of the subscriber in the time zones with high call arrival probability will be given more weight i.e. more number of patterns will be taken of those scenarios, while less number of patterns will be taken for time zones of early morning and late night. After making this epoch, it will be randomly divided into two subsets, the training subset and the testing subset. The training subset will be learned completely numerous numbers of times with randomized order of presentation of patterns from one epoch to the next till the prediction performance is not adequate. This performance will be tested against the testing subset. For input, the neuron representing the time zone of the pattern will be activated and the output will be calculated according to the current bias and synaptic weights. The output of the net should represent the current location of the subscriber. Errors will be calculated at output layer and back propagated to hidden layer and further used to correct the synaptic weights of the network. By this way the network will give output as all the areas where the subscriber usually spends time in the time zone corresponding to respective input through a vigilance of user for a time span. If for a particular input time zone, the activation value of an output neuron is higher then some threshold value, that neuron will be fired, i.e. the subscriber would be in the corresponding location. The neuron with maximum activation value will represent the most likely area to be paged. The other firing neurons will represent the less likely areas.

### 3.5.3 Pseudo-code

a. Learning

```
void main ( )  {
                initialiseWeight ( );
                adjustWeight ( );
}

void adjustWeight ( )  {
                        X= getInputTime ( );
                        Y= getOutputLocation ( );
```

```
//feed forward
for (j=1; j<=m; j++){
        HInducedField[j] = 0;
        for (i=0; i<=l; i++){
                HInducedField[j] += (x[i] * v[i][j]);
        }
        HActivationSignal[j] = activationFunc (HInducedField[j]);
}
for (k=1; k<=n; k++){
        PInducedField[k] = 0;
        for (j=0; j<=m; j++){
                PInducedField[k] += (HActivationSignal * w[j][k]);
        }
        PActivationSignal[k] = activationFunc (PInducedField[k]);
}
// error calculation
for (k=1; k<=n; k++){
        OutputLocalGradient = (y[k]-PActivationSignal[k]) *
                                dActivationFunc (PInducedField[k]);
        for (j=0; j<=m; j++){
                dw[j][k]= η * OutputLocalGradient[k] *
                                        HActivationSignal[j];
        }
}
for (j=1; j<=m; j++){
        LocalErrorInput[j] = 0;
        for (k=1; k<=n; k++){
                LocalErrorInput[j] += DW[j][k] * W[j][k];
        }
        HiddenLocalGradient[j] = LocalErrorInput[j] *
                                dActivationFunc (HInducedField[j]);
```

```
        }
        for (i=0; i<=l; i++){
                for (j=1; j<=m; j++){
                        DV[i][j] = η * HiddenLocalGradient[j] * x[i];
                }
        }
        //error back-propagation
        for (j=0; j<=m; j++){
                for (k=1; k<=n; k++){
                        W[j][k] += DW[j][k];
                }
        }
        for (i=0; i<=l; i++){
                for (j=1; j<=l; j++){
                        V[j][k] += DV[j][k];
                }
        }
        if (stop( ))
                Exit ( );
        else
                adjustWeight ( );
}
```

b. location update algorithm at MS

```
        When enter to a new cell
        If( mode = = prediction){
                If ( cell id is not in its own profile)
                        if(NeighbourFlag=1)
                                locationUpdate( );
                        else
                                NeighbourFlag=1;
```

```
        else
                NeighbourFlag=0;
}
else{
        record(current time,cell id);          //for the learning process
        distanceBasedLocationUpdate(lastUpdatedCellId, CurrentCellId);
}
```

c. Paging algorithm at MSC

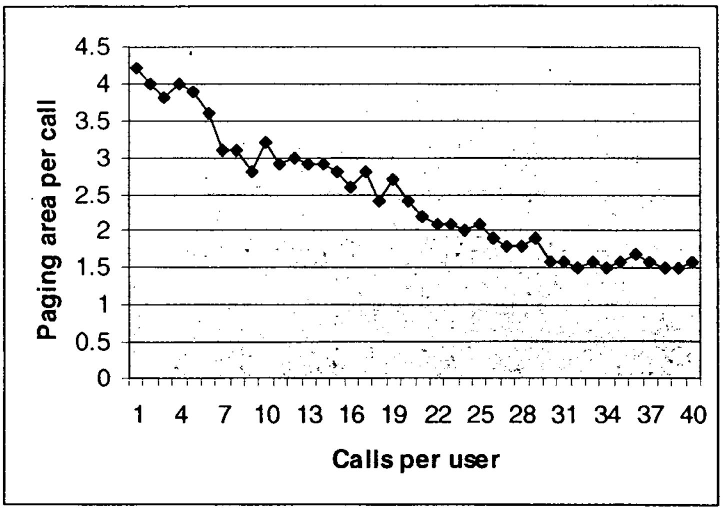
```
When a call is received for a user
{
Cycle 1:
if(TimeSinceLastUpdate <= Tth )
        page(LastUpdatedCell);
else
        page (MostLikelyArea);

If not found
Cycle 2:
if(TimeSinceLastUpdate <= Tth )
        page(MostLikelyArea);
        page(LessLikelyAreas);
else
        page(LessLikelyAreas);
        page(NeighborsMLArea);
        page(LastUpdatedCell);

If not found
Cycle 3:
Page(NeighborsLLAreas);
}
```

### 3.6 Simulation and Results

Sigmoid function has been taken as activation function for the learning process. For simulation of the above proposed algorithm, we have taken a hypothetical region divided in 31 cells. Every user has been assigned a home area, where it lives, few market areas of his interest, an office area etc. A user is being monitored for his movement in the region for a particular time (2 weeks) and accordingly his profile is being generated. These movements are motivated from above mentioned factors. A random call generator has been used to generate the calls for that user and user is paged accordingly. It had been seen that normally the user needs to update its locations for a large movement only, when it resides in a cell for very small time. The average no. of updates initiated per users per day has been found about 2.1 to 3.2 after watching ten users. Average number of paged area depends upon the call pattern for the user, since with each call; the network gets update about the location of the user. After watching ten users for ten consecutive days for the randomly generated independent calls we have drawn a graph between number of calls per user and average number of paging area per call.




Since with each call, mobile terminal also updates its location, with a high call arrival rate, the location of mobile subscriber can be determined more efficiently. All of the calls generated through random call generator and are independent of other calls. The only constraint on the call arrival rate is that probability of initiation a call is high in morning and evening, less in daytime, and very small in late night and early morning hours.

# Chapter 4
# Conclusion

The proposed ANN based location management scheme has been shown to satisfy the goal of reducing the signaling overhead for the task of location management and yields promising results. The scheme reduces location management cost over both of the basic task of location update and paging across a variety of users and network parameters. Importantly, no biased assumptions are made with regards to regularity of user behavior or uniformity of network topology. The proposed scheme is able to achieve high levels of performance for realistic user movement patterns and network topologies.

The main assumption behind the proposed scheme is that every subscriber does possess a mobility pattern, in general. By providing an appropriate training, a neural network can be trained to predict this pattern of the mobile subscriber. It may take a bit longer time to learn about the mobility pattern of the subscriber, but after learning, the algorithm is capable enough to predict the location of mobile subscriber instantly. Results show a good reduction in number of location update messages initiated by the subscriber after training. The number of average paged cell per call depends upon the call pattern of the subscriber as well as the number of paging cycles allowed to connect to a mobile terminal. At the maximum, only the cells within profile, the last updated cell and neighbors of cells in profile needs to be paged with a very rare chance whether most of the time it needs to page one or two cells only. The division of time into five zones reflects the gross user routine. With a larger number of paging cycles, the number of paging area can be reduced further, but it will lead to a greater average latency of call setup. During the training period, the mobile station is supposed to keep track of its daily movement pattern and transfer this information to MSC when the MSC asks to do so, normally in low traffic hours. This requires a data communication session between MSC and the mobile terminal. This also requires a limited memory and computational power at the mobile terminal which is feasible for a 3G mobile terminal.

An MLP network is trained in batch mode with the movement history of a subscriber, and during run time, only the last updated location with time of update is to be stored at the network controller at each update or call termination. The network is capable enough to page the subscriber within a limited number of paging cycles with minimal number of paging areas.

## Bibliography

[1] L. P. Araujo, and J.R.B. De Marca, "A comparative analysis of paging and location update strategies for pcs networks", In Communications, 1998. ICC 98. Conference Record.1998 IEEE International Conference on, volume 3, pages 1395–1399 vol3, 1998.

[2] I.F. Akyildiz, J.S.M. Ho, and Yi-Bing Lin, "Movement-based location update and selective paging for pcs networks", Networking, IEEE/ACM Transactions on, 4(4): 629–638, 1996.

[3] I.F. Akyildiz, Yi-Bing Lin, Wei-Ru Lai, and Rong-Jaye Chen, "A new random walk model for pcs networks. Selected Areas in Communications", IEEE Journal on, 18(7): 1254–1260, 2000.

[4] G. Chakraborty, B.B. Bista, D. Chakraborty, and N. Shiratori, "Location management in pcn by movement prediction of the mobile host", In Industrial Electronics, 2002. ISIE 2002. Proceedings of the 2002 IEEE International Symposium on, volume 1, pages 78–83 vol1, 2002.

[5] Kuo-Hsing Chiang and N. Shenoy, "A 2-d random-walk mobility model for location-management studies in wireless networks", Vehicular Technology, IEEE Transactions on, 53(2): 413–424, 2004.

[6] P.G. Escalle, V.C. Giner, and J.M. Oltra, "Reducing location update and paging costs in a pcs network", Wireless Communications, IEEE Transactions on, 1(1): 200–209, 2002.

[7] R. Subrata and A.Y. Zomaya, "Location management in mobile computing" In Computer Systems and Applications, ACS/IEEE International Conference on. 2001, pages 287–289, 2001.

[8] Yi-Bing Lin, "Reducing location update cost in a pcs network", Networking, IEEE/ACM Transactions on, 5(1): 25–33, 1997.

[9] V.W.-S. Wong and V.C.M. Leung, "Location management for next-generation personal communications networks", Network, IEEE, 14(5): 18–24, 2000.

[10] Simon Haykin, "Neural Networks", second edition, Pearson Education.

[11] William Stallings, "Wireless Communications and Networks", Pearson Education, 2004.

[12] G. Poliini and C.L.I., "A profile-based location strategy and its performance", IEEE Transactions on Selected Areas of communication vol. 15(8), Oct. 1997

[13] M. Pandey, A. Pal, "Location management and paging using mobility pattern", Communications, 2003. APCC 2003. The 9th Asia-Pacific Conference on Volume 2, 21-24 Sept. 2003 Page(s): 439 - 443 Vol.2.

[14] G. L. Lyberopoulos, J. G. Markoulidakis, D. V. Polymeros, D. F. Tsirkas, E. D. Sykas, "Intelligent paging strategies for third generation mobile telecommunication systems", Vehicular Technology, IEEE Transactions on Volume 44, Issue 3, Aug. 1995 Page(s): 543 - 554