Project Final Report

A Data Mining Approach for Location Management in 4G Wireless Heterogeneous Networks

1. Introduction

The main objective of this research is to solve the problem of location management in the fourth generation (4G) of wireless mobile networks. The design of the location management technique aims to reduce the signaling overhead in the mobile networks and to deliver the calls correctly. The goal of this project is to design, implement, and evaluate a new location management technique for 4G wireless heterogeneous networks (WHN) using data mining technology. None of the current location management techniques have been designed for the new structure of 4G WHN. Multidimensional sequence mining techniques are used in this project to extract the mobility patterns of the mobile users. The new location management technique uses these patterns to predict the location of the mobile users inside the 4G WHN. This helps to reduce the communication overhead that is required to determine the location of the mobile users. Data mining tasks in the proposed technique are distributed between the mobile handsets (MH) in the 4G WHN. Simulation was conducted to study the performance of the posed location management technique and to compare its performance with other techniques.

2. Background and Literature Review

It is expected that the fourth generation (4G) mobile networks will support more multimedia communications and provide mobile services every time and everywhere. Location management enables mobile networks to track the locations of the mobile users between consecutive communications [18-20]. As mobile communications become very popular and even essential today, the radio signal traffic generated by location management is increasing rapidly due to the increases in the population and mobility of mobile users [46]. Also, the future mobile communication networks will be integrated with heterogeneous access methods and various kinds of cells. This can by accomplished by integrating several radio access networks (RANs) (such as cellular networks (2G, 3G, etc.) and Wireless LAN (WLAN)) in developing what is called "wireless heterogeneous networks" (WHN). Accordingly, effective location management is required to determine the location of mobile terminals in the new heterogeneous system.

Location management in mobile networks consists of two basic operations [2]: locations updating (or location registration) and paging (or search). Location updating procedure provides the network with initial information about an MS location (i.e. the location area where the MS is found). Since the mobile users are free to move within the coverage area, the network can only maintain the approximate location of each user [2]. When a connection needs to be established for a certain user, the network has to determine the exact location of the mobile user inside the coverage area. Paging procedure is used to determine the exact location information about the mobile user (i.e. the Base Station to which the subscriber is connected in a specified location area.) [1].

There is a trade-off between the costs for location updating and paging [1]. If the mobile terminals update its location whenever it crosses a cell boundary, the network can maintain its precise location, thus obviating the need for paging. But, the cost for location updating will be very high. On the other hand, if the mobile terminal does not perform location updating frequently, a large coverage area has to be paged when a call arrives to determine the location of the called user and the cost for paging will be very high. Thus, the basic problem of location management is to develop an algorithm that can be used to minimize the overall cost of location updating and paging.

The problem of location management will be more complex in WHN because different types of radio access networks (RANs) have different interfaces and different capabilities [15]. Manging the location of the mobile users in WHN is important to complete the communication between different RANs. This is also one of the important and hot research problems for the next generation of mobile networks. Most of the existing location management techniques that have been proposed by researchers in this area deal with a single RAN and none of these schemes was designed for multiple RANs as in the WHN.

The current location management techniques that can be applied in single RAN [17-20] are also inefficiently in the case of the mobile terminals are moving between the same location areas during day periods. For example, if the subscriber movements are done according to his life habits such as work, study, travel, visiting, etc., then using the current location management technique will produce unnecessary overhead due to the movement between the locations which could be known with less cost. The new researches are interested in introducing new methods for location management, which attempt to reduce the overhead traffic. In [3, 4], techniques which make use of location areas (LAs) and paging areas (PAs) of different sizes are introduced. In these techniques a LA is divided into several PAs. In [5, 6, 18] various database architectures are proposed with the aim of organizing the database. In [7] the multilayer concept is introduced. In this method, there are groups of MSs and each MS is assigned to a given group, and each group is assigned one or several layers of LAs. The method proposed in [8] uses a process which predicts the movements of the MS according to its direction of movement, velocity, and so on. Processing and prediction are made at both the MS and the HLR. When actual movements of the MS do not fit with those predicted, a registration is triggered by the mobile to inform the network of its actual location. Otherwise, no exchange is required, which allows savings in LU processing and signaling. In [9, 10, 20], the alternative strategy (AS) is introduced. Its main goal is to reduce the location updates by taking advantage of users' highly predictable patterns. In this AS, the system handles a profile recording the most probable LAs patterns of each user. The profile of the user which contains the visited LAs can be provided and updated manually. When the user receives a call the system pages him sequentially over the LAs until getting an acknowledgment from the mobile. The main savings allowed by this method are due to the non-triggered LUs when the user keeps moving inside his profile LAs. In [11], a technique similar to AS is defined. It is called Statistical Paging Area Selection (SPAS) and is based on location statistics collected by each MS, which periodically reports them to the network. These statistics consist of a list of the average duration the MS has been located in each LA. A priority rule is determined to settle the sequence of LAs visited by the mobile. If this sequence is different from the last one reported to the network, the MS transmits it; otherwise, nothing is done. The paging process is achieved in the same way as in AS. A variant of this method, called the Two-Location Algorithm (TLA), is proposed and studied in [12]. In this strategy, a mobile stores the two most recently visited LA addresses. The same is done at the HLR level. Obviously, the main advantage of this method relies on the reduction of LUs when a mobile goes back and forth between two LAs.

3. Project Description

This goal of this project is to develop a new predictive location management technique (called LM-WHN) for WHN that integrates cellular networks and WLANs. LM-WHN takes into account the different characteristics of the cellular networks and the WLANs. Data mining technology are used in this technique to predict the location of mobile users. This will be done by using the spatial, temporal, and usage information of the mobile users to predict the mobility profiles. A data mining technique called MobilePrefixSpan [16] is be used in the proposed WHN to build these mobility profiles.

3.1 Architecture of 4G Mobile Communication Systems Supported by LM-WHN

The architecture of the proposed heterogeneous wireless networks includes WLANs and cellular networks and assumes an All-IP based. All-IP wireless and mobile networks represent the convergence of two key technologies: Internet and wireless cellular systems [16]. The core IP network will serve as the backbone network with internet connectivity and packet data services [1, 2, 7].

The architecture of the proposed wireless heterogeneous network is shown in Fig. 1. As we can see, there are four basic components of the proposed architecture: 1) Base Stations, 2) Access Points, 3) Mobile Hosts, and 4) IP Core Network. The WLAN access points (APs) represent the fixed communication points for the WLAN while the base stations (BSs) represent the fixed communication points for the cellular networks. Mobile hosts (MHs) are supposed to be designed to work in two different modes: dual-ode and single-mode. Dual-mode mobile hosts will be able to support services provided by WLANs and cellular networks while single-mode mobile hosts will support only one type of mobile technology. The proposed WHN will be able to support dual mode mobile hosts as will as single-mode mobile hosts.

By design, cellular networks are aimed at users with high mobility and low connection rates while WLAN networks are aimed at users with low mobility and high connection rates. The configuration of the wireless heterogeneous networks will be based on the connectivity to IP networks. The advantages of using IP as a core network protocol are:

- Internet connectivity
- Efficient transmission of IP packets,
- Co-existence with other access systems
- Ease of system introduction and expandability.
- IP networks can also connect with or accommodate wireless access systems other than 4G systems.
- Providing better security requirements.

The MH will be able to communicate with one BS/AP according to the coverage area and its mode of communications. The connection can be handed off from one AP to another AP (or from BS to another BS), and this called *horizontal handoff*. The connection can also be handed off form WLAN to cellular networks or vise versa and this is called *vertical handoff*. In our system model, we can have several WLANs inside the same cell but they do not have overlapped coverage. So the mobile user at every moment has the following possibilities of coverage:

- Covered by cellular network only
- Covered by cellular network and WLAN.

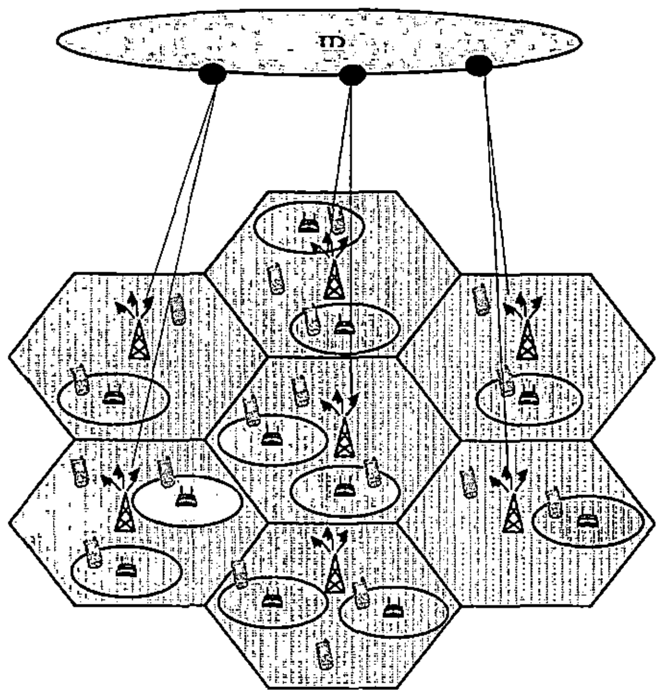

Fig. 1 Architecture of Proposed Wireless Heterogeneous Networks

3.2 Methodology

3.2.1 Data Mining Methodology

Data mining is the search for new, valuable, and nontrivial information in large volume of data. There goal of applying data mining techniques in the location management problem is to predict the mobility patterns. This will help to reduce the overall cost of location updating and paging. The key here is how we can effectively record and analyze the previous behavior of mobile users to generate these mobility profiles.

The proposed LM-WHN technique is mainly based on generating users' mobility profiles using a sequence mining technique called MobilePrefixSpan [16]. The mobility profiles are generated individually for each user by his/her MH. The MH will have the responsibility to collect the mobility data of its user. This data is used to build the mobility profile of the user and is continually updated based on its users' movements.

3.2.1.1 Data Collected by Mobile Host

The data collected by the MH is used to build the local mobility model that can describe the behavior of mobile users for a period of time. The data items collected for this model is shown in Table 1, where:

L_i: represents the ID of the current visited cell or access point as shown below.

VST_i: represents the time stamp when this MH enters L_i

VET_i: represents the time stamp when this MH exits L_i

W_i: a binary value describes if the information was recorded during the weekend or not. We will use the letter Y if it was on a weekend day, otherwise we use the letter N (this means $Wi \in \{Y, N\}$).

Table 1 Data Collected by the MH

ID of Visited BS/AP	Visit Start Time	VISIT END TIME	IS THIS ON A WEEKEND DAY ?
L_1	VST_1	VET_1	W_1
L_2	VST_2	VET_2	W_2
...	...	...	...
L_n	VST_n	VET_n	W_n

Note that a new record will start if the mobile user enters a new BS/AP or if the current type of service is changed. The type of service can be "idle" if the mobile phone was not communicating at the time of data recording. Also, note that the day at which this service was accessed (weekend or not) is included as an important factor. We believe that the behavior of the users is totally different during the weekends.

To generate the mobility paths, it is necessary to transform the collected data from Table 1 at the MH into a sequence of symbols, where each symbol represents BS/AP. Every path is composed of a sequence of BS/AP IDs for each recorded visit. In addition, each path also contains the information regarding the duration of MH stay in a cell. This is achieved by collecting mobility data of a MH at fixed time slots Δt. For example if $\Delta t =$ 2 minutes, this means that movements data are collected every tow minutes. In this case, the time between movements is included in the generated sequence of movements. This information is used to estimate the time at which the user will move from one location to another.

Any generated path P_{MH} can be represented as:

$P_{MH} = \langle L_1 \ L_2 \ \ldots \ L_n \rangle$

where $\qquad L_i \in \{ B \cup A \}$

$$B=\{C_1, C_2, \ldots, C_m\} \quad \text{and} \quad A=\{P_1, P_2, \ldots, P_k\}$$

and

C_i: is a cell ID.

B: is the set of all cells IDs in the system.

P_i: is an access point ID.

A: is the set of all access points IDs in the system.

3.2.1.1 Building Mobility Patterns

In this section we explain the details of the data mining technique, called MobilePrefixSpan, which we have developed to analyze the information collected from the mobile users. MobilePrefixSpan technique (which is shown in Fig. 2) is a modified version of the well-known Prefix-Span [21, 22] sequence mining technique and its modified version for multidimensional sequences UNISEQ [23]. The goal of MobilePrefixSpan technique is to extract the movement patterns of the mobile users using the collected information. There are two types of user information collected for generating the mobility patterns:

- **Spatial Information:** this indicates the location of the mobile user at every recording time. This information is collected by recording the ID of BS and AP.
- **Temporal Information:** indicates the time and the day information collected during the navigation of the mobile users. We divide the day into several intervals.

0. Prepare the multidimensional sequences and convert the problem to one-dimensional sequence mining.
1. Apply the first scan to find all of the single-item frequent sequences (prefix).
2. Find the projected dataset corresponding to the single-item frequent sequences.
3. We continue by finding out the single-item frequent sequences in each projected dataset.
4. Find the frequent sequences using this prefix:
- *If the items represent BS or AP in the network, then consider the consecutive order.*
- *Otherwise consider the order only.*

5. Record the frequent sequences that have been found using this prefix.
6. Use each of these recorded frequent sequences as a prefix to find its projected dataset.
7. Repeat steps from 4 to 7 until we find all sequential patterns.

Fig. 2 MobilePrefixSpan Algorithm

3.2.2 Location Management Technique (LM-WHN)

3.2.2.1 Location Updating Procedure

The location updating will be made into two levels. The first level is at the stored mobility profile in the MH. And the second is at the database of the BS or AP in the network. The updating of the stored data profile in the MH will be done by updating the collected data. The important thing here is to know the BS or AP that will be deleted from this profile, these nodes are called the expired nodes. The registered BS or AP will be considered as an expired node if the mobile user did not use connect through it during a certain period of operation. These expired nodes will be replaced with new nodes, which will be determined from the user mobility profile generated in the MH. The sequence mining software installed on the MH can analyze the collected data stored in MS and make its processes on it to build the user mobility profile. This profile will be sent to the database of the visitor location register (VLR) where the user is most probably to be found within nodes in that VLR region. These VLR will be in the BS or AP nodes. The address of this VLR is sent to home location register (HLR) in the IP core network. Here, the profiles of the users will be distributed in the several VLRs in the network and the corresponding addresses of these VLRs for the users will be stored in the HLR. Hence, the user profile was updated. This updating process can be initiated after a specified number of movements for each user or after a specified time interval. If the MH enters a new LA (where the location area will be a set of BS and Nodes), then the location updating will be done immediately and this new location will be added to the user profile.

3.2.2.2 Paging and Incoming Call Procedure

In the case of the incoming call, paging will be performed to needed to determine the exact location (BS or AP) of the mobile user. The incoming call procedure can be described in the following steps:

- The calling MH will send a message to its mobile switching center (MSC) through the current BS or AP to obtain the location of the dialed MH.
- The calling MSC will forward a message to the database at HLR at the interconnecting network, which contains the address of the VLR database where the user profile of the called user is stored.
- The control software at the VLR database will determine the current parameters of the time and the day according to the time of calling.
- Then the *location nodes (BS or AP)with highest probability* corresponding to this called user will be obtained from the stored mobility profile according to the current parameters.
- The paging (search) process will start by sending messages to these location nodes. If the MH is not found at *these location nodes* , then we page *the next set of location nodes according to the mobility profile.* We continue to do this until one of the paged nodes will answer. At this point, the connection can be started between the two MHs.

4. Evaluation

The evaluation of the proposed LM-WHN is based on the total cost of the location updating (LU) and paging (P). This evaluation is mainly depends on the cost formulation used in [13]. This cost formulation for the location management consists of two parts: *location updating cost* (C_{LU}) and *paging cost* (C_P). Each of these two parts will include the *network cost* (messages cost for signaling) and *database cost (required to access the database)* as shown in Fig. 3. The database cost consists of the following two types of costs:

D_h : The cost of accessing the Home Location Register (HLR).

D_v : The cost of accessing the Visitor Location Register (VLR).

The network cost consists of the following two types of costs:

N_f: The cost of sending a message through the fixed network(wired).

N_w : The cost of sending a message through the wireless network.

Using the cost of the location updating and the cost of paging, the total cost for LM-WHN ($C_{LM\text{-}WHN}$) technique can be calculated as:

$$C_{LM\text{-}WHN} = \lambda_{lu} C_{LU} + \lambda_c C_P$$

Where λ_{lu} is the location updating rate (location updates/sec) and λ_c is the call arrival rate in the WHN (calls/sec).

Simulation is used to evaluate the performance of the proposed LM-WHN technique. The call to mobility ratio (*CMR*) is used as an important parameter in our simulation. This parameter determines the ratio between the number of incoming calls and the number of movements between cells. This parameter will be in the range from 0.25 to 2. The performance of the proposed LM-WHN technique in terms of total cost will be compared with a Benchmark technique which will have the same location management technique with important assumption that the prediction is perfect. This means that for the Benchmark technique, we know exactly what will be current BS/AP for every MH in the network.

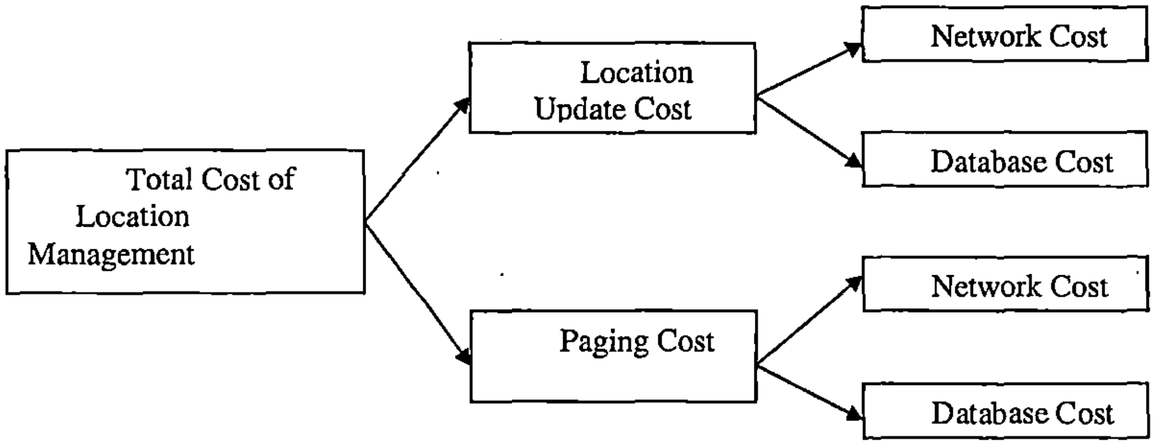

Fig. 3 Location Management Cost

5. Simulation and Experimental Results

5.1 Simulation Model

In order to evaluate the performance of the proposed LM-WHN technique, we have built a simulation model to study the performance of the proposed scheme and to compare it with a benchmarking technique. The simulation parameters are shown in Table 2. The CMR will be in the range from 0.25 to 2The call duration (call service time) is assumed to be exponentially distributed with mean of 180 seconds.

The simulation starts by building the topology of the cells. Some of these cells are marked as used cells and the others are marked as unused cells. The unused cells simulate the natural structure of the roads where there are some constraints to deploy base stations in locations (like river, mountain, etc…). The used cells are then divided into home cells, work cells or ordinary cells. These cells are associated for the users to travel through according to the moving statistics. The simulation has two phases. The first phase is the initialization phase. The main objective during this phase is to collect the movement data, which are used to generate the mobility profiles. The simulation parameters and statistics, described before, are used to generate the movements of the users between the different cells. In this initial period of simulation, we focus only on collecting the data of movements between the cells. Hence a simple LM technique based is used in the initial period by updating the location when the user moves from one location area to another location area.

Table 2 Simulation parameters

Parameter	Value
Number of cells	128 cells
Type of cells	Hexagonal cells
Diameter of each cell	1 mi
CMR	From 0.25 to 2
Call duration time	Exponentially distribution with mean μ=180 seconds
Maximum Speed	60 mi/h
D_h	2
D_v	1
N_f	2
N_w	4

5.2 Experimental Results

The simulation was conducted to evaluate the performance of the LM-WHN technique and compare its performance with the Benchmark technique. Benchmark technique is based on the same LM technique and it assumes that the prediction is perfect. For the Benchmark technique, we know exactly what will be the next path for every user, and what will be the exact handoff time. This will enable us to compare the performance of the proposed LM-WHN with the best approach. The results are shown in

Fig. 4. Here we can see that the LM-WHN have an acceptable performance compared to the performance of the benchmark technique. This means that using the LM-WHN we can predict the mobility of the users with high accuracy comparing to the Benchmark. This is because using the LM-WHN techniques allows us to predict the location of the BS/AP with high accuracy.

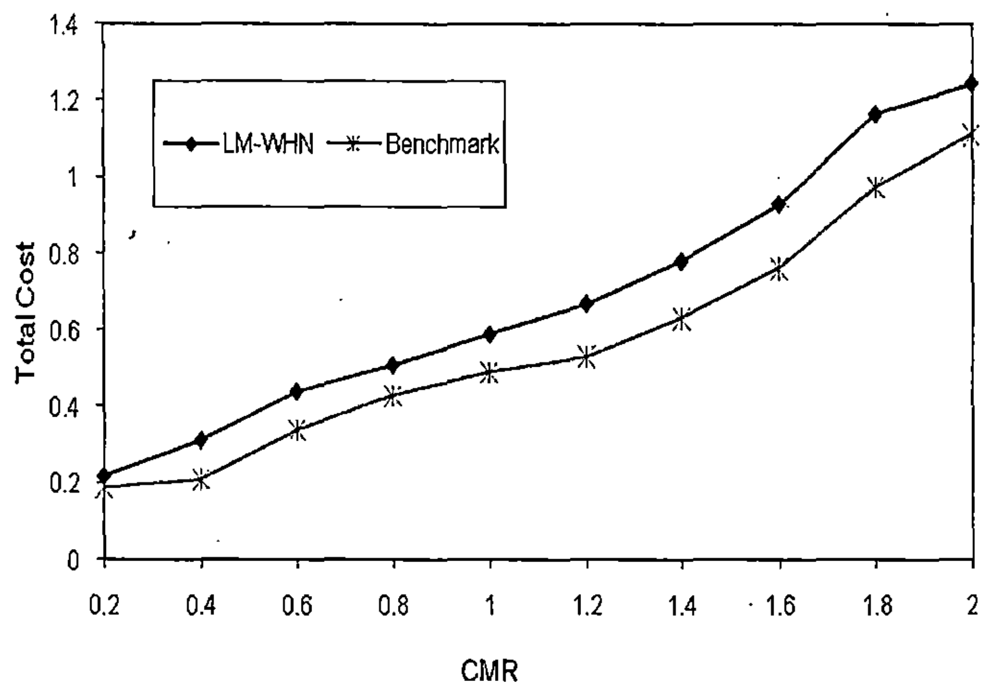

Fig. 4 Total cost for LM-WHN and Benchmark techniques

Also, the simulation was conducted to compare the performance of the LM-WHN with the performance of the traditional location area based location management technique (LM-LA) where we update the location with every movement between location areas. LM-LA is a non-predictive technique. This will help us to see the effect of the prediction in the proposed LM-WHN technique. The results are shown in Fig. 5. Here we can see that the LM-WHN have a significantly better performance compared to the performance of the LM-LA technique. This means that using LM-WHN reduces the communication signals that are required to update the location. The cost required for location updates and paging signals have been reduced significantly. The high accuracy prediction in LM-WHN techniques helps to determine the correct location of mobile users with less communication signals which reduces the overall cost.

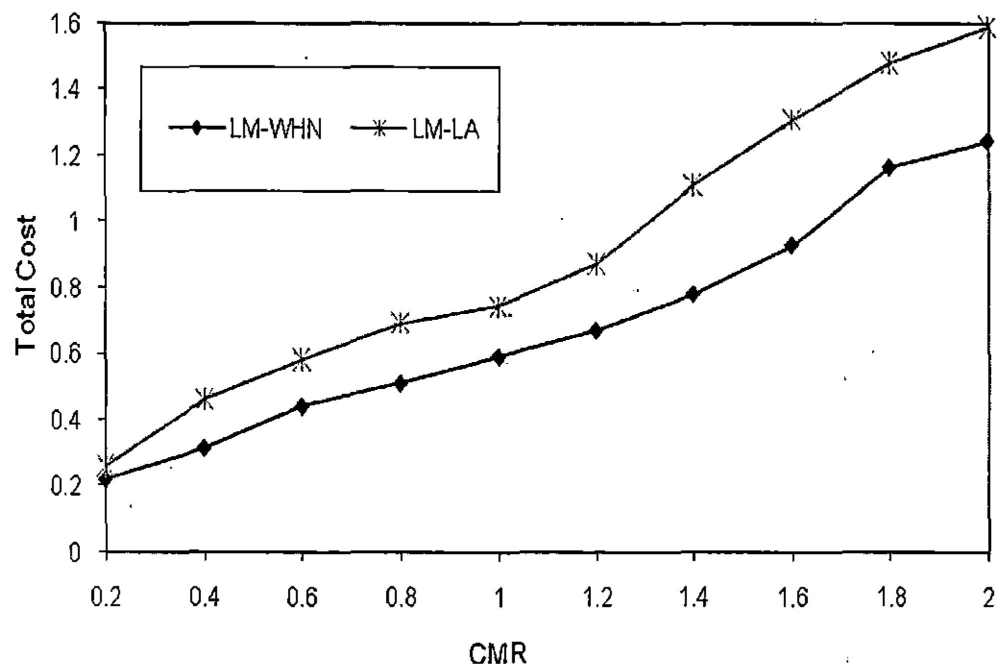

Fig. 5 Total cost for LM-WHN and LM-LA techniques

To terminate the incoming calls, we need to search the MS (by paging the BS/AP) to determine its location, so there will be a delay for incoming calls which is required to search the MS. We calculate it as a normalized value with respect to corresponding LM-LA value. Simulation was conducted to study the effect of the delay on the proposed technique. The results are shown in Fig. 6. The delay in the proposed technique is slightly higher than that of the delay in LM-LA technique as shown in Fig. 6. This is because we use a sequential paging for the BS/AP. This delay is acceptable and can be reduced in future by enhancing the prediction accuracy of the proposed technique.

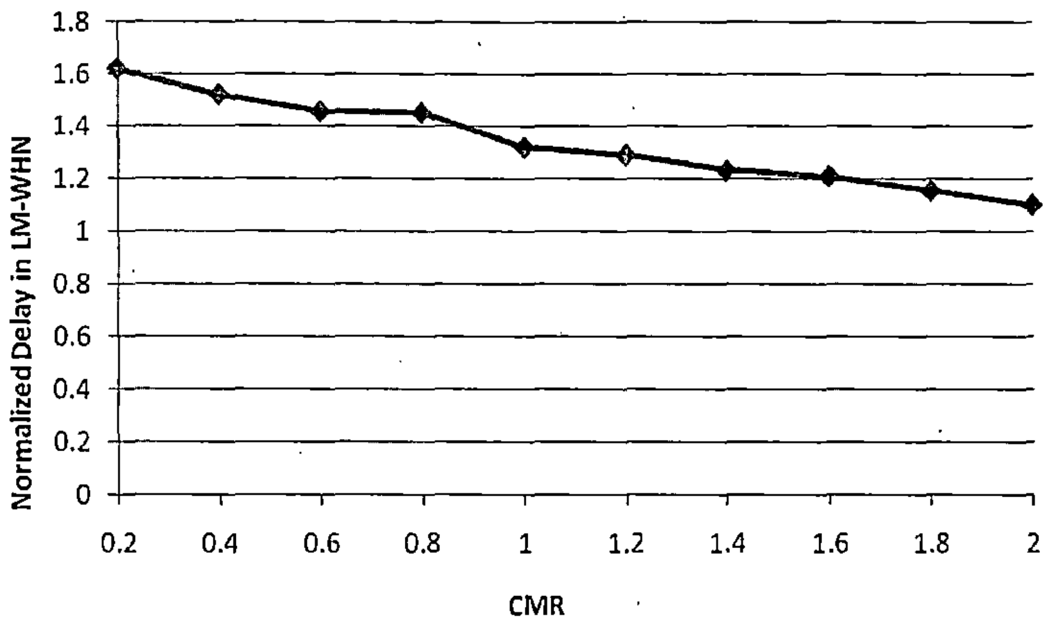

Fig. 6 Normalized Delay in LM-WHN

6. Conclusion and Future Work

Location management techniques aim to reduce the signaling overhead in the mobile networks and to deliver the calls correctly. The main objective of this research is to solve the problem of location management in the fourth generation (4G) of wireless mobile networks using predictive data mining techniques. In this project have designed, implemented, and evaluated a new location management technique for 4G wireless heterogeneous networks (LM-WHN) using data mining technology. The new location management technique utilizes the mobility patterns of mobile users to predict their location inside the 4G WHN. Data mining tasks in the proposed technique are distributed between mobile handsets. Simulation was used in this project to evaluate the performance of the proposed location management technique. Simulation results show that the proposed predictive technique (LM-WHN) technique has better performance compared to the non-predictive location management technique. Also the performance of the LM-WHN technique was compared to the Benchmark technique. Simulation results show that we can predict the user mobility with high accuracy using the proposed LM-WHN technique. Also simulation results show that the delay in the proposed technique is slightly higher than the delay in the non-predictive location management technique. This is because we use a sequential paging for the BS/AP. This delay is acceptable and can be reduced in future by enhancing the prediction accuracy of the proposed technique. Future work includes enhancing the prediction accuracy by adding new mobility parameters such as the speed and the direction of mobility. Also, similar location management techniques can be developed for integrated WLAN, Ad-hoc networks and cellular networks.

References

[1] Igor Brusic,Vesna Hassler and Wolfgang Lugmayr, "Deployment of Mobile Agents in the Mobile Telephone Network Management", Proceedings of the 33rd Hawaii International Conference on System Sciences - 2000.

[2] Tabbane, S., “Location Management Methods for Third-Generation Mobile Systems”, IEEE Communication Magazine, August 1997, pp. 72-84.

[3] D. Plassmann, “Location Management for MBS,” Proc. IEEE VTC, Stockholm, Sweden, June 8–10, 1994, pp. 649–53.

[4] D. Goodman, P. Krishnan, and B. Sugla, "Design and Evaluation of Paging Strategies for Personal Communications," Wksp. Multiaccess, Mobility and Teletraffic for Pers. Commun., Paris, France, May 1996.

[5] N. and S. Tabbane, "Database Architectures and Location Strategies for Mobility Management in Mobile Radio Systems," Proc. Wksp. Multiaccess, Mobility and Teletraffic for Pers. Commun., Paris, France, May 1996.

[6] D. C. C. Wang, "A Survey of Number Mobility Techniques for PCS," Proc. IEEE Int'l. Conf. Pers. Commun., Tokyo, Japan, Nov. 6–10, 1995.

[7] S. Okasaka et al., "A New Location Updating Method for Digital Cellular Systems," Proc. IEEE VTC '91, Saint Louis, MI, May 1991.

[8] C. H. Rokitansky, " Knowledge-Based Routing Strategies for Large Mobile Networks with Rapidly Changing Topology," Proc. ICCC '90, New Delhi, India, Nov. 1990, pp. 541–50.

[9] S. Tabbane, "Comparison between the Alternative Location Strategy (AS) and the Classical Location Strategy (CS)," WINLAB Tech. Rep. 37, Aug. 1992.

[10] S. Tabbane, "An Alternative Strategy for Location Tracking," IEEE JSAC,vol. 13, no. 5, June 1995.

[11] M. Shirota, Y. Yoshida, and F. Kubota, "Statistical Paging Area Selection Scheme (SPAS) for Cellular Mobile Communication Systems," Proc. IEEE VTC '94, Stockholm, Sweden, June 8–10, 1994.

[12] Y.-B. Lin, "Reducing Location Update Cost in a PCS Network,"IEEE/ACM Trans. Networking, vol. 5, no. 1, Feb. 1997, pp. 25–33.

[13] Kibom Kim, Kown Woo Yang, Joon-Ming Gil, and Chong-Sun Hwang, "Traking Mobile User Locality in Mobile Computing Systems", International Workshop on Parallel Processing, Wakamatsu, Japan, September 21-24, 1999.

[14] Narumi Umeda, Toru Otsu, and Tatsuro Masamura, "Overview of the Fourth-generation Mobile Communication System", NTT Technical Review, Vol.2 No.9, pp. 12-17, September 2004.

[15] Cavalcanti, D.; Agrawal, D.; Cordeiro, C.; Bin Xie; Kumar, A.; "Issues in integrating cellular networks, WLANs, and MANETs: a futuristic heterogeneous wireless network", IEEE Wireless Communications Magazine, Volume 12, Issue 3, pp. 30 - 41 , June 2005.

[16] Sherif Rashad, Mehmed Kantardzic, and Anup Kumar, "MSP-CACRR: Multidimensional Sequential Patterns Based Call Admission Control and Resource Reservation for Next-Generation Wireless Cellular Networks", the 2007 IEEE Symposium on Computational Intelligence and Data Mining (CIDM 2007), Honolulu, Hawaii, April 1-5, 2007.

[17] Ng, C.K.; Chan, H.W.;" Enhanced Distance-Based Location Management of Mobile Communication Systems Using a Cell Coordinates Approach", IEEE Transactions on Mobile Computing, Vol. 04 , Issue 1, pp. 41 - 55 , Jan.-Feb. 2005.

[18] Quintero, A.; Garcia, O., "A profile-based strategy for managing user mobility in third-generation mobile systems", IEEE Communications Magazine, Vo.l.42 Issue 9 , pp. 134 – 139, Sept 2004.

[19] Xiao, Y.; Pan, Y.; Li, J.;"Design and analysis of location management for 3G cellular networks", IEEE Transactions on Parallel and Distributed Systems, , Vol. 15 , Issue 4 , pp. 339 – 349, April 2004.

[20] Il Han; Dong-Ho Cho; "Group location management for mobile subscribers on transportation systems in mobile communication networks", IEEE Transactions on Vehicular Technology, Vol. 53 , Issue 1 , pp. 181 – 191, Jan. 2004.

[21] Jian Pei; Jiawei Han; Mortazavi-Asl, B.; Jianyong Wang; Pinto, H.; Qiming Chen; Dayal, U.; Mei-Chun Hsu, "Mining sequential patterns by pattern-growth: the PrefixSpan approach", IEEE Transactions on Knowledge and Data Engineering, Vol.16, Iss.11, pp. 1424- 1440, Nov. 2004.

[22] J. Pei, J. Han, B. Mortazavi-Asl, H. Pinto, Q. Chen, U. Dayal, and M.-C. Hsu, "PrefixSpan: Mining Sequential Patterns Efficiently by Prefix-Projected Pattern Growth," Proc. 2001 Int'l Conf. Data Eng. (ICDE '01), pp. 215-224, Apr. 2001.

[23] Helen Pinto, Jiawei Han, Jian Pei, Ke Wang, Qiming Chen, Umeshwar Dayal, "Multi-dimensional Sequential Pattern Mining" , Proceedings of the 2001 ACM CIKM International Conference on Information and Knowledge Management, Atlanta, Georgia, pp. 81-88, November 2001.