An Adaptive Location Management Strategy for Mobile IP

Subhashini Rajagopalan*
Bellcore,
331, Newman Springs Rd.
Red Bank, NJ-07701-7040
subha@nyquist.bellcore.com

B.R.Badrinath
Dept. of Computer Science
Rutgers University,
New Brunswick, NJ-08903
badri@cs.rutgers.edu

ABSTRACT

Mobile internetworking revolves around the premise that a mobile host (MH) will frequently attach to and detach from the Internet, thus changing its address frequently. Location management in mobile internetworking refers to the process of keeping track of the current addresses of MHs in the Internet. Protocols proposed for mobile internetworking rely on a combination of searches and updates for location management. A drawback of these protocols is that they are not uniformly efficient over all possible call-to-mobility ratios (i.e., the relative frequency of searches as compared to updates). An important issue, therefore, is how the overall costs of location management can be reduced, regardless of the call-to-mobility ratio. The aim of this work is to explore a new and practical scheme for location management, with emphasis on network cost reduction.

We make the key observation that while the potential set of sources for the MH may be large, the set of sources that a given MH communicates most frequently with is very small. Based on this, we develop the concept of a *working set of hosts* for the MH (similar to the working set concept in operating systems). An adaptive scheme of location management is proposed, that enables an MH to dynamically determine its working set, and trade-off routing and update costs in order to reduce the total cost. Comparative analysis of the proposed scheme, using simulation, shows its efficiency over a wide range of call-to-mobility ratios.

1. INTRODUCTION

1.1 Basic Concepts of Mobile Internetworking Protocols (IP)

Mobile internetworking revolves around the premise that a mobile host will frequently attach to and detach from the Internet, thus changing its address frequently. Mobile internetworking protocols that have been proposed rely on a combination of search and update strategies to locate an MH in the Internet. Under these protocols, each MH is assigned a permanent Internet address, called the *home address,* which is based on it's *home network* ID. As long as the MH resides at its home address, packets are routed to it as if it were a fixed host. When the MH moves to a new network - *foreign network,* it detects a change in location by receiving one of the periodic beacon messages sent by a local router, called the *foreign agent (FA),* that indicates the current network ID. When the MH makes its presence known to the router, it is assigned a temporary local address, called the *foreign address.* The MH then updates a designated router in its home network, called the *home agent (HA),* of its current point of attachment in the network. The HA always has correct and up-to-date information about the mobile host.

MOBICOM 95 Berkeley CA USA

Now, when a remote source sends a packet to the mobile at its home address, an attempt is made to deliver the packet to the MH at that address. If the MH is not present in its home network, the home agent receives the packet and forwards it on to the current location of the MH. This process is called *triangle routing*. There are two forms of triangle routing, basic and enhanced. In basic triangle routing, proposed in [12], all packets from the source to the MH, travel the indirect path through the HA. Since much of the Internet traffic is connection-oriented (using TCP), an enhancement over the basic triangle routing scheme can be proposed ([5]). With enhanced triangle routing, the connection set up procedure relies on a three-way handshake and during this procedure the current address of the MH is conveyed to the source. Therefore, only the initial few connection set-up packets from the source take the sub-optimal path through the HA. Once the connection is established, the data packets may flow directly to the current location of the MH.

We can formally define a quantity called the *call-to-mobility ratio*, proposed in [9], that indicates the rate at which a source sets up connections to an MH relative to the rate at which the MH changes its location. To be precise, we define the call-to-mobility ratio, with respect to a given source, as the number of connection requests generated per unit time by the source, divided by the number of times the location of the MH changes per unit time.

The disadvantage of the basic triangle routing procedure is that much overhead is incurred by redundant searches when a source sets up connections at a faster rate relative to the mobility of the MH. Under the basic triangle routing scheme, the overheads due to redundant searches increase as the call-to-mobility ratio increases. On the other hand, when the call-to-mobility ratio is low (i.e., relatively high mobility), triangle routing is very efficient, since it incurs the minimal update overhead and there are not many redundant searches.

Compared to the basic triangle routing scheme, an improved strategy can be proposed for high call-to-mobility ratios, where the mobile actively informs remote sources of its current address, whenever it moves. This approach, which we call the *static update* scheme, however, has problems associated with it. First, the MH may be communicating with a large number of sources, thereby giving rise to a large update set. Second, it is not possible to correctly determine the static set of sources, apriori. Third, at low call-to-mobility ratios, the update overheads may increase more than the corresponding gains in the search overhead. Therefore, this is a hypothetical scheme. However, the performance of the static update scheme at high call-to-mobility ratios serves as a benchmark to compare the performance of the various location management schemes against.

1.2 Outline

Proposed mobile internetworking protocols, while dealing

* This work was done while the author was a graduate student in the ECE department at Rutgers University, NJ

with the mechanisms for search and update, have not considered an important issue in mobile internetworking: how the overall costs (routing and update) of location management can be reduced, regardless of the call-to-mobility ratio. Our approach is to first recognize that the location management schemes at the two ends of the spectrum, the basic triangle routing scheme and the static update scheme, incur reduced overheads when the call-to-mobility ratio is low and high, respectively. Next, we make the key observation that while the potential set of sources that can communicate with a mobile host may be large, the set of sources that actually communicate *most frequently* with a given MH is quite small. The basis of this work is the empirically established locality properties of the Internet traffic [6]. We not only present some evidence of these properties, but also a way to exploit these properties effectively to reduce the costs incurred in location management. To this end, we develop the concept of a *working set* of hosts for an MH, that is conceptually similar to the working set model proposed by [7]. This observation facilitates the design of an *adaptive* scheme for location management that emulates the basic triangle routing scheme when the call-to-mobility ratio is low and the static update scheme when the ratio is high. It will be shown that our scheme performs well regardless of the call-to-mobility ratio.

Section 2 of the paper explores existing research in the area of mobile internetworking. The motivation for our research is given in section 3 along with a description of the working set concept and its correspondence with Internet user behavior. The details of the proposed adaptive algorithm is given in the following section. Descriptions of the simulation environment, traffic and mobility models used and performance metrics are given in section 5, followed by analysis of the research results and comparison of performance of the proposed algorithm with basic and enhanced triangle routing schemes in section 6. Some implementation issues related to the proposed adaptive algorithm will be discussed in section 7. Finally, section 8 gives conclusions and ideas for future work.

2. REVIEW OF RELATED WORK

Several protocols have been proposed in the mobile internetworking community. These protocols are meant to work within the existing TCP/IP protocol framework to route packets to MH's in the Internet. As mentioned earlier, it is assumed that the MH is capable of attaching to the Internet through radio-equipped base stations called Foreign Agents or FA's. The MH is assigned a constant IP address (that does not change with location) called it's home address, by its home agent (HA) and a temporary foreign address by its FA.

The IETF Mobile IP proposal [15] is similar to the Columbia proposal ([11]), in the sense that both embrace the sub-optimal basic triangle routing method of routing packets to MHs located away from their home networks. The Mobile IP proposal from Harvard University [5] proposes implementation of a secure short-cut routing method, or enhanced triangle routing, whereby location information is distributed to enhanced sources by the HA. Except for the first few packets, all packets between an MH and enhanced source avoid the sub-optimal triangular route via the HA. This scheme is a definite improvement over the IETF proposal. However, in [5], location information is propagated to the sources only as long as an active connection is maintained between the source and the MH. We will show in our work that overheads in location management can be further reduced if sources frequently communicating with an MH are supplied the location information and allowed to keep this information beyond the time frame of an active connection. A separate set of extensions to the basic IETF Mobile IP scheme has been proposed in [12]. Mobile location caches, maintained at the correspondent hosts or FAs, store the mobility bindings for the MHs. This cache can be updated either by the home agent or by other FAs that have a more recent binding. [12] also extends the registration protocol in [15], in order to ensure that previous FAs can update their cache with the new foreign address after the MH has moved to a new location. Our approach differs from [5] and [12] in that information will not be arbitrarily given to all sources that communicate with the MH.

Another approach to the problem is the concept of a *propagating cache* introduced in the Sony system ([17]). However, the implicit assumption in this proposal is that all routers deployed in the network are enhanced routers. This requirement makes the Sony proposal only a long-term solution for wide-area mobility. The IBM proposal [16] is similar to all the other proposals in that the MH registers and it informs its HA about changes in its location. However, it proposes incorporating the *Loose Source Routing* (LSR) option in the IP header, in packets sent out by the MH. Although the mechanics of the scheme ensure that optimal routing can be achieved using existing IP protocol features, the lack of correct loose source routing implementations makes this scheme difficult to implement ([13]).

In summary, some mobile internetworking proposals have attempted to address search cost reduction using feedback from the MH to sources with active connections ([5]) and cacheing location information at the FAs [12], but there has not been an effort to systematically reduce the network costs of both updating and searching. Also, there has been no attempt to reduce search costs across multiple connections. A theoretical approach to reducing the costs of tracking a mobile in a distributed network, has been proposed in [3]. Distributed data structure is constructed and this hierarchy of structures contains pointers to the location of the mobile. The costs are reduced by varying the granularity of location information at each level in the hierarchy. Some pertinent work has been done in the cellular environment in the area of hierarchical location management ([2], [8], [14], [18], [19]). The concept of hierarchical updating in the internetworking environment has also been proposed in [4]. However, hierarchical approach alone is not a solution to the problem. It is understood that with this approach, update and tracking messages are localized in the network. Search costs are also minimized for terminals that are located close to each other. However, assuming that a mobile terminal frequently communicates with a fixed terminal, and they are located at opposite ends of the hierarchy, the hierarchical approach in cellular networks, does not result in much savings.

None of the strategies, both in cellular and in mobile IP, have explored the existence of a working set of sources for a mobile user. Our work will show that not only does such a set exist for a mobile user but also that location management costs in the fixed network can be effectively reduced by propagating location information to selected sources in the network.

3. MOTIVATION

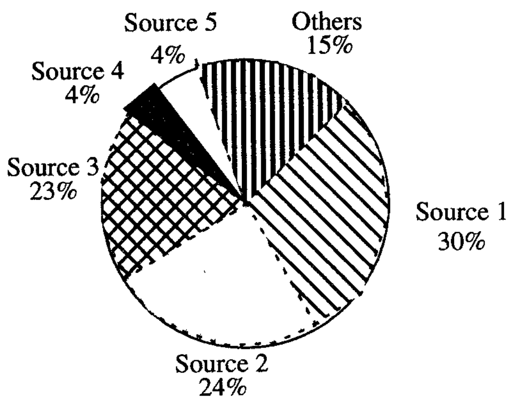
3.1 The Working Set Model for Mobile Hosts

The working set model, proposed in [7], defines a working set of pages associated with a process as the collection of its most recently used pages in the main memory. The purpose of the working set model is to help the operating system make a decision about what pages in the main memory are in use on the basis of page reference patterns and thus determine memory demand. The working set concept proposed for computer operating systems can be applied to mobile computing.

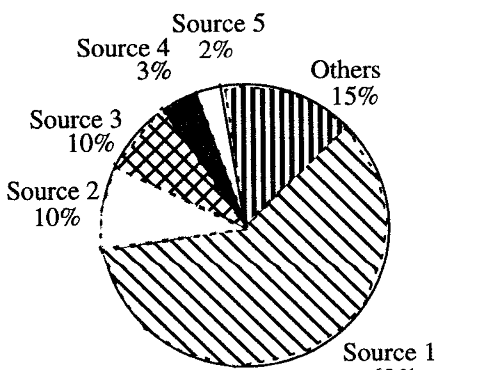
It has been observed in studies of TCP/IP wide-area internetwork traffic that there exists a strong sense of locality in the way hosts and networks are accessed over the Internet. [6] presents conversation level analysis of wide-area TCP traces collected in the University of Southern California, University of California, Berkeley and Bellcore. One of the observations in the study of wide-area network locality was that 50% of the *telnet* conversations in UCB are directed to just 10 sites. It shows that locality of reference exists between host pairs or network pairs in wide-area internetworks. Traffic locality in local area networks has also been similarly observed [1]. Although, these observations have been made at the network level, from experience we can observe such locality of reference at the user level too.

3.2 E-mail Traffic Experiment

In order to highlight the point that hosts or networks tend to correspond more with certain hosts or networks than with others, we conducted our own study of email traffic received by users. Email connections received by 4 different users over a long time period was studied. We present results for two users, plotted as pie graphs (Graphs 3.1, 3.2).



Graph 3.1: Connections over a 10 day period



Graph 3.2: Connections over a 9 day period

5 sources (as defined in section 1) that contribute the most to the traffic to a user were considered for plotting graphs. Each source is represented by a wedge in the graph. In Graph 3.1, user 1 received 487 email messages over a 10 day period from 27 unique sources. Of the total number of messages received, 30% were received from just source 1 and 77% were received from sources 1, 2 and 3, combined. In Graph 3.2, user 2 received fewer messages than user 1, but it can still be seen that of the 63 messages received, 60% (or 38 messages) were from a single source, source 1. It is understood from the results that even though user behavior is influenced by user environment, locality in the hosts that send email messages to a user, can be observed.

3.3 Motivation

Although limited in scope, the above experiment shows that even if the potential set of sources that could communicate with a user is large, the set of sources that actually do correspond on a frequent basis is small. This set of sources is also relatively *stationary*, i.e., any change to this set will come very slowly and only after a period of time. This property of locality in the hosts and networks a user communicates with, is not exclusive to electronic mail but can be observed in other Internet related activities (as discussed in [6]) and it is closely related to the environment of the user.

We define the working set of a user to be the set of sources the user communicates most frequently with. Such a set would exist regardless of the mobility of the user, since it is merely a reflection of the user's activities (professional or personal). The working set of hosts might change with time, but the locality property underlying the communication behavior would not. In other words, a user might stop communicating with certain sources and start communicating with others (for example, a user graduates from school and takes up a job in a remote location). Our work is that this locality property in communication should and can be exploited to reduce costs of location management for mobile users. For example, the cost of locating user 1 (Graph 3.1), could be reduced if its location information is provided to sources that correspond with it frequently, namely sources 1, 2 and 3. Our approach to reducing location management costs, therefore, is to trade off routing and update costs in such a manner that those sources which benefit from location updates are indeed kept informed and those that originate connections to the MH relatively rarely are allowed to search. This adaptive approach is neither like the "always search" approach typified by triangle routing, nor like the "always update" approach of the static update scheme. But it is shown in section 6, that the adaptive scheme emulates the best of these two, depending on the call-to-mobility ratio.

4. DESCRIPTION OF ADAPTIVE LOCATION MANAGEMENT SCHEME

The purpose of this section is to describe the derivation of the adaptive scheme and to set the basis for discussing its performance in section 6.

4.1 The Adaptive Location Management Scheme

The issues facing the adaptive scheme are, how to determine the current working set of the MH and how to determine which sources in the working set must be updated as the MH moves. The adaptive scheme uses a simple *on-line* algorithm that resolves both these issues together. Briefly, for each source that communicates with the MH, the adaptive scheme evaluates whether overall costs could be reduced by allowing the source to do a search, each time it sets up a connection to the MH, or by updating the source each time the MH moves. This evaluation is done by computing the following quantities for each source, s, that actively communicates with the MH:

- f_s: The estimated frequency with which connection set-up requests are received from the source s.

- Δ_s : The additional routing cost incurred by s, in setting up a connection to the MH. This cost is defined as:
  cost(s, HA) + cost(HA, MH) - cost(s, MH) .... Eq. 4.1
  where these terms denote the network cost of sending a message from the source to the HA, the cost of sending the message from the HA to the MH and the cost of sending the message directly from the source to the MH, respectively. In other words, the cost measure is the additional overhead incurred when the source does not know the current location of the mobile. The individual cost values may be statically provisioned by the network administrators, based on distance.
- f_{update} : The estimated frequency with which the MH changes its location, and
- U_S : The cost incurred in both announcing the location of the MH to the source s, and also invalidating the location information at the source. Invalidation means that the information at the source is made obsolete and all future connection requests from the source directed to the MH have to be routed through the HA.

Having determined these parameters, the adaptive scheme evaluates in an on-line manner whether the following simple inequality holds for the source, s:

$$f_s * \Delta_s > f_{update} * U_S \quad \text{...... Eq. 4.2}$$

The left and right hand sides of the inequality denote the routing cost and the update cost components, respectively. The above parameters are updated and the inequality is evaluated at the MH each time the source sets up a connection or when the MH moves. The basic idea is to reduce the total network costs, by updating or invalidating the location information at each source, based on the estimated cost over the long run of the routing costs versus the cost of updates. The total network cost for all sources is minimized by reducing the total cost on a per source basis.

The adaptive scheme has the potential to be efficient over a wide range of call-to-mobility ratios. For instance, when the ratio is low (i.e., mobility is high compared to connection set-up rates), the estimated update cost (the RHS of Eq. 4.2) tends to be high and sources (other than the HA) are typically not given the location information. Thus, the scheme behaves like the basic triangle routing scheme at low call-to-mobility ratios. On the other hand, when the call-to-mobility ratio is high, the estimated routing costs (the LHS of Eq. 4.2) tend to be high and sources that communicate most frequently with the MH are put in the update set. Thus, at high call-to-mobility ratios, the scheme behaves like the static update scheme. Indeed, the simulation results presented in section 6 verify this behavior. A simple theoretical analysis of the adaptive location management scheme, presented in [20], further underlines the effectiveness of this scheme.

4.2 Cost Measure

As mentioned previously, the cost of the links between the network nodes can be based on the distance between the nodes. This is a reasonable assumption since distance is a good indicator of the transmission costs and router processing and hence the network cost. When more than one source needs to be updated by the MH, the update cost, f_{update}, can be minimized if update messages are forwarded along the branches of a *minimum spanning tree* rooted at the current location of the MH. All sources that need to be updated are the nodes in the tree. The cost of updating a source will then be the cost of sending the update message to the source from its parent in the tree. From previous discussion we know that the working set of sources for a user is small and relatively static. Since the working set of sources form the nodes of the minimum spanning tree, the cost of computing the minimum spanning tree is not very high. Implementation of the adaptive algorithm at the MH requires the mobile terminal to do frequency estimation and computation of the inequality, Eq. 4.2, for each source, indicating that the adaptive algorithm is not computation intensive.

Besides the minimum spanning tree-based measure of cost, it is also possible to propose a hop count-based cost measure. In the internet, it is desirable to minimize the overall number of hops traversed by packets to mobile hosts, since this minimizes the cost of router processing and link bandwidth consumed. Therefore to minimize resource consumption, the cost measures, Δ_s and U_S, may be computed based on hop count. The number of hops to a given node can be determined using the IP *time-to-live* (TTL) field. When the MH receives the packet, it can determine the number of hops to the source. In other words, *cost(s, HA) + cost(HA, MH)* is known when a packet is routed to the MH through the HA and *cost(s, MH)* is known when packet is routed directly to the MH. Therefore the update cost, *cost(MH, s)*, is known. From Eq. 4.2, for two sources with same f_s, U_S being equal and for a given f_{update}, the source with a higher hop count through the triangle route will be updated.

The problem with using hop count detection to determine costs is that it may not be possible for the MH to construct the minimum spanning tree to send the update and invalidate messages. This is because spanning tree construction requires knowledge of the cost between every pair of sources in the update set. In the absence of the spanning tree, the MH must send updates or invalidates separately to each source. While adding to the update cost, hop count based cost measure may be the most easy measure to derive and use. However, we will see that the correctness of the adaptive algorithm is not affected by the cost measure used. Cost measure based on the distance between nodes has been assumed in our simulations. We use the minimum spanning tree method for computing the update costs.

4.3 Implementation of the Adaptive Scheme

The adaptive scheme is implemented at the MH, in cooperation with the sources and the HA. The frequency measures, f_s and f_{update}, are determined using a simple averaging procedure. The basic idea is to determine if it is more efficient to supply the source with the location information or to allow the source to route the connection request using triangle routing. For computing the average arrival rate, the last n connection requests that arrived at the MH from source s are observed. Similarly, the last n moves made by the MH are used to compute the average update rate. f_s is computed by dividing the number of connection requests received by the time between the last and the first connection request. f_{update}, is similarly computed.

The algorithm for adaptive location management scheme has two parts, as explained below.

- *MH actions after a connection request is received from a source, s:*
  If the source s is already a member of the update set, i.e. it has location information of the MH, the only action that needs to be taken at the MH is to record the time at which this connection request arrived in the search queue for the source s. This will be used to compute f_s. On the other hand, if s is currently not a part of the update set, the MH needs to make a decision about whether or not to include the source in its update set, by

evaluating Eq. 4.2. The parameters of the inequality, f_s and an estimated value of f_{update}, are computed using the frequency estimation procedure explained earlier. Δ_s is computed using Eq. 4.1. To determine, U_s, the MH constructs a temporary spanning tree, which includes s and all the current update set members. This gives an estimate of the update cost that will be incurred if s is included in the update set. If the inequality holds, s is included in the update set and location information is supplied over the already established connection.

- *MH action when it changes location:*
  When the MH changes its location in the network, it evaluates f_{update}, and modifies the update queue. The MH must determine if the sources currently in the update set should continue to be updated from the new location. It also determines if there are other sources currently not in the update set, but which may have qualified to be put in the update set since the last move. It thus reevaluates the inequality, Eq. 4.2, for each source in its working set. If the inequality holds, the source will be included in the update set. Otherwise, the location information at the source must be invalidated. Invalidation forces the information at the source to be made obsolete, causing the source to go through the home agent for future connection set-ups. Once the new update set has been determined, a temporary minimum spanning tree is computed by the MH. This tree is rooted at the new MH location, and it includes the sources in the previous tree plus new sources added since the last move. Update and invalidate messages to the sources are sent along this tree. Sources that are invalidated will then be removed from the update set.

It can be seen that the adaptive scheme dynamically estimates future behavior of the system in terms of call rate and mobility rate with the help of past searches and updates. Although the precise estimation of the frequencies will require long term samples, it is shown later that good results can be obtained even with this simple short-term averaging scheme.

5. DESCRIPTION OF THE SIMULATION MODEL

The simulations implement triangle routing, static update and the adaptive location management schemes described earlier, along with different mobility models. The simulation program was written using *CSIM*, a process-oriented simulation package.

5.1 The Network Model

The aim of the simulations is to determine the routing and update overheads incurred by various schemes in a wide area internet. For the purposes of simulation, a 68 node national backbone network is considered. The network is represented as an undirected graph. The nodes are assumed to be geographically distributed and the cost of each link is proportional to the distance between the nodes it connects. On the average, each node in this network is connected to approximately five other nodes. A given pair of nodes are directly connected by at most one link. The cost of sending a message between any pair of nodes in the network is taken to be cost of the shortest path between the nodes, as computed using Djikstra's algorithm. The network nodes can be visualized as the epicenter of several small *administrative domains* (AD) or a single large AD. Thus, each node consists of several smaller networks and represents a large area of movement for an MH.

All the results obtained were for one MH in the network. Connection requests were generated by fixed sources only, whose location is indicated by a node in the network.

5.2 The Source Model

The source model encodes the locality properties described in section 3. Accordingly, we consider three sets of sources. The first set, called the *main set*, consists of sources that communicate most frequently with the MH. Sources in this set are assumed to contribute the most traffic to the MH and this set is assumed to be static. The home agent is part of the main set for the MH. The second set, called the *transient set* of sources, consists of sources that communicate with the MH but not on a very frequent basis. This set is assumed to be of a transient nature, since sources may be removed from this set or new sources may be added. The main set and the transient set are, of course, disjoint. Finally, the third set consists of sources that communicate with the MH on a "one-time basis", i.e. these sources may or may not continue to communicate with the MH after the first time. We call this set of sources as the *external set*. If these sources continue to contact the MH, they may become part of the transient set of sources. This set tries to capture the stray connection requests (such as one-time email messages) an Internet user may receive.

Under the adaptive scheme, regardless of which set a source belongs to, it may be put in the set of nodes the MH updates, depending on the frequency with which the source generates connections and the location of the MH in the network with respect to the source. The probability of a call being generated by a source is pre-determined by probabilities assigned to each set of sources.

5.3 Non-hierarchical and Hierarchical Update Strategies

Each of the 68 network nodes, called *big nodes*, can be thought of as a single AD, or a part of a large AD. Each big node controls many *sub-nodes* spread across a sizable geographic area. The MH can move across sub-nodes (i.e., within a big node) or move between big nodes. Once the current big node location of the MH has been selected, the amount of time the MH stays within any sub node is given by a negative exponential distribution with mean *meansmallstay*. Similarly, the amount of time the MH stays within a big node is given by a negative exponential distribution with mean *meanbigstay*. The parameter *meanbigstay* is always greater than *meansmallstay*.

The notion of a location hierarchy in the network is introduced to show that location update costs for the MH can be reduced with hierarchy. A two-level geographic hierarchy is implemented. The location hierarchy corresponds to a system in which there are *regional location servers* that track MHs within their region. We can think of each big node as such a region and it is assumed to be the top level of the hierarchy.. In our simulations we study the performance of location management schemes both with and without regional servers. When there are no regional servers, the MH must update its HA and other sources (in case of static update and adaptive schemes), even for moves within a region. This scheme is the non-hierarchical approach to location updating. In a hierarchical location updating scheme, the home agent of an MH keeps track of only the region the MH is in, and not the specific location (i.e., sub node) within the region. Thus, in this approach, when the MH moves within a region, it updates only the regional server. When it moves into another region, it updates both the new regional server and the home agent. For simplicity, in this approach, the cost of updating the regional server is assumed to be negligible and only the cost of updating the HA is kept track of.

While protocols for implementation of hierarchical updates have been proposed in research papers [4], the implementation of them is far into the future. The current mobile internetworking protocols (based on triangle routing) under consideration for standardization do not have the notion of location hierarchy. The adaptive scheme proposed in this work can be implemented both with and without regional servers. In the latter case, the scheme can in fact be integrated into the current mobile internetworking framework being considered.

5.4 Mobility Models

When the MH decides to move out of a big node, the next node it visits is determined by the mobility model. Three mobility models, which derive their characteristics from different user profiles, have been selected to evaluate the performance of the location management strategies. They are the *Traveling Salesman Model*, the *Pop-up Model* and the *Boring Professor Model.*

The traveling salesman mobility model characterizes users who have a *epicenter* in their movement pattern, i.e. they exhibit zonal characteristics. The zone of movement is specified by the number of hops or *numhops*, from the epicenter. When the MH moves between big nodes, this parameter ensures that the MH will move to only those nodes which are *numhops* hops away from the epicenter.

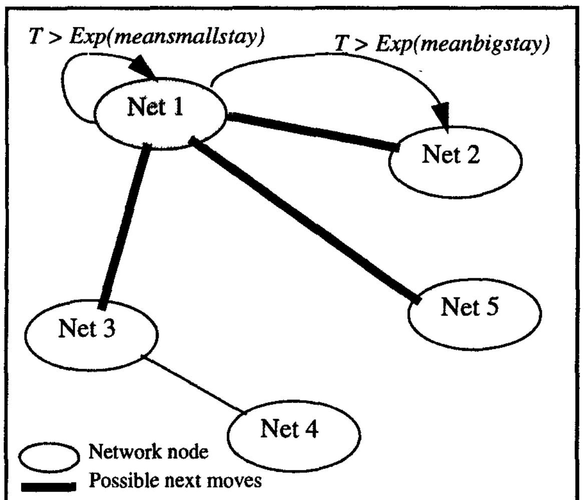


Figure 5.1: Traveling Salesman Mobility Model

In figure 5.1, the current location of the MH is indicated as a shaded node (Net1). Let Exp(*meansmallstay*) be the exponential time with mean given by *meansmallstay* and Exp(*meanbigstay*) be the exponential time generated with mean given by *meanbigstay.* The MH stays within the same sub-node until Exp(*meansmallstay*) expires. Then it moves to a different sub-node within the same big node. On expiry of Exp(*meanbigstay*), the MH moves to a new big node. Assuming that Net1 is the epicenter for the MH and *numhops* = 1, the MH can choose to move to Net2 or Net3 or Net5 (Net4 is not a possibility since it is 2 hops, not 1 hop, away from the epicenter). The choice of the next node is made at random.

It has been pointed out in previous work, [10], that most Internet users tend to have a relatively *unchanging* mobility behavior, which is closely related to the environment of the Internet user. Once this characteristic is recognized in a user profile, location management not only becomes easier but also more efficient. The boring professor model makes use of this characteristic. Without loss of generality, we can say that a professor has a chosen few locations that he visits more often than others, such as his campus node and his home node (which could also be the same node). For the purposes of simulation, we assumed that the nodes are one hop away from each other. In figure 5.2, the amount of time the MH stays in the home and campus nodes is taken to be normally distributed with mean given by *meanhomestay* and *meancampstay*, respectively, and standard deviation (s.d.) of 2 hours. To imitate departure from the schedule of home and work, the MH decides to

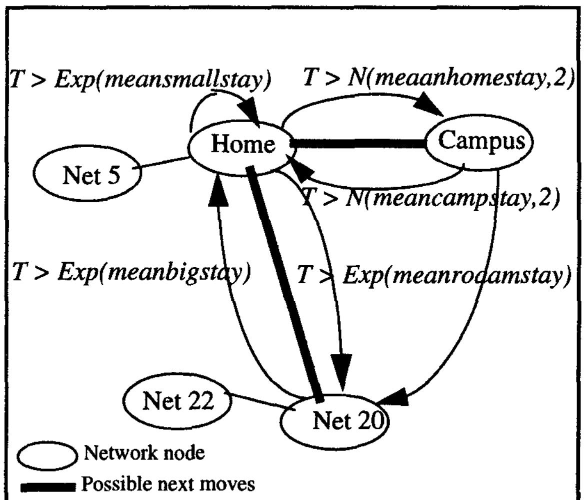


Figure 5.2: Boring Professor Mobility Model

roam to a different node when time given by Exp(*meanroamstay*) expires. This models the conferences or meetings the professor may attend. Although the model uses a professor as an example, it can be seen that most Internet users follow a specific mobility pattern, in terms of the locations they visit and the amount of time they spend at each location.

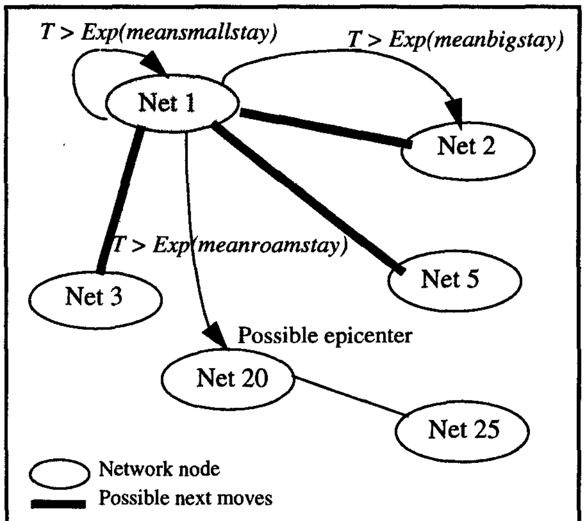


Figure 5.3: Pop-up Mobility Model

The pop-up mobility model corresponds to the case when the MH has a tendency to pop-up at different points in the network, such that the new location has no relation to the previous movements of the MH. This mobility model can be considered to be an extension of the traveling salesman model, where the salesman after traveling in the neighborhood of one region (say, New York), flies to a new region (say, Florida) and travels there. That is, in our model illustrated in Figure 5.3, the MH after moving in one zone, decides to pop-up in a different node in the network after an exponentially distributed time, Exp(*meanpopstay*), with mean *meanpopstay*. (*meanpopstay* is greater than *meanbigstay*.)

6. RESULTS AND ANALYSIS

The aim of the simulations is to analyze the effectiveness of the adaptive scheme in the presence of different mobility models and different numbers of sources, over a range of call-to-mobility (C/M) ratios. The performance metrics for comparison of the various schemes are the routing, update and total costs. We will discuss these metrics for triangle, static update and adaptive location management schemes, for both hierarchical and non-hierarchical update strategies. Although simulations were performed for many scenarios, only a representative sample is reported here to illustrate the general behavior of the various schemes. We have chosen to illustrate the results obtained for the traveling salesman model, since the general property of the algorithm was found to remain the same. Simulations have been performed using distance based cost measure and with minimum spanning tree for updates. We will discuss the results for non-hierarchical updating in greater detail than the hierarchical update strategy.

6.1 The Simulation Setting

As described in section 4, the adaptive scheme reduces the overall cost of location management of an MH, by reducing the costs on a per source basis. Therefore, it is appropriate to study the behavior of this scheme in response to the changes in the C/M ratio as observed for a single source (recall that in section 1, the C/M ratio was defined in terms of a single source).

For the purposes of comparison, we assume that for the static update scheme, the working set of the MH is known with complete certainty apriori and this set never changes. Also, the sources in static update scheme are assumed to have the location information at all times, i.e. they are not invalidated by the MH.

All sources generating connection requests are considered to be part of the working set of the MH. The probability that a connection is generated by sources in the main and transient set is taken to be 0.7 and 0.2, respectively, leaving 1% of the connections to be generated by the external set. However, when the transient set size is zero, sources in the main set generate 100% of the connection requests.

6.2 Behavior in a Uniform Cost Network

In order to isolate the effects of just the call-to-mobility ratio on routing and update costs for various schemes, a uniform cost network is considered. In this network, the cost of the path between every pair of nodes is set to 100. This ensures that the cost results obtained will reflect precisely the number of searches and updates performed by each scheme, because each of these have the same cost, regardless of the location of the MH and the sources relative to each other.

Graphs 6.1, 6.2 and 6.3 show the average routing, update and total costs, respectively, incurred in the uniform cost network, using the traveling salesman mobility model, when a total of 5 sources (all in the main set) generate connection requests for the MH. In these graphs, the X-axis indicates the C/M ratio per source and the Y-axis indicates the average cost. The significance of the trade-offs in location management can be clearly seen. In Graph 6.1, the routing costs are highest for the triangle routing scheme and zero for the static update scheme. The average routing cost is constant for the triangle scheme, since every connection request results in a search directed to the HA, and all incur the same cost. For the adaptive scheme, the average routing cost decreases with the C/M ratio, indicating that each source gets updated quickly of the MH's location, avoiding redundant routing costs. A point to mention is that the routing costs for triangle routing is 80 and not 100. This is because the HA of the MH is chosen to be one of the sources in the main set (and routing costs for connections originating at the HA is zero).

Now, looking at the average update costs (Graph 6.2), the triangle and static schemes incur constant overheads for every update, since the number of sources updated is 1 and 5, respectively, for each update. The adaptive scheme, on the other hand, behaves like the triangle routing scheme at low C/M ratios and like the static scheme at high C/M ratios, as expected. The average update costs for the adaptive scheme are low at low C/M ratios because the scheme tends to update only the HA and not other sources. At high C/M ratios, however, many other sources get into the update set of the MH and hence the cost per update increases.

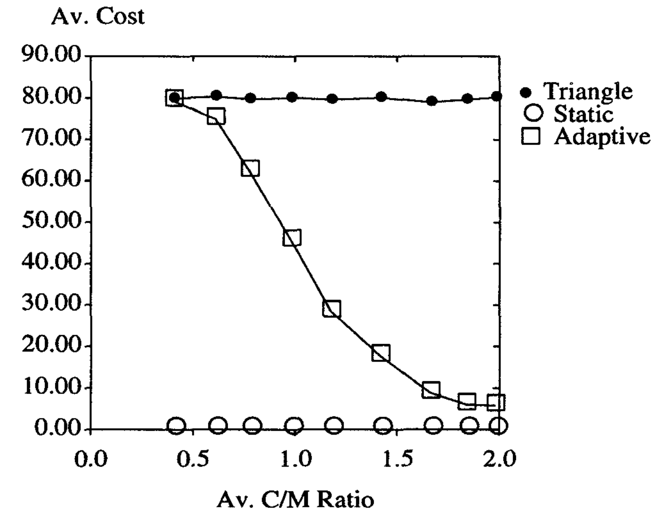
Finally, the average total cost of location management, amortized over all the updates and searches, is shown in Graph 6.3. Here, the adaptivity of the proposed scheme is clearly seen. In the region where the update costs tend to dominate (i.e., at low C/M ratios), the adaptive scheme performs much like the triangle routing scheme, and where the routing costs are high, the scheme mimics the static update scheme. It is also clear that neither the triangle scheme nor the static update scheme are efficient over the range of C/M ratios.

6.3 Behavior in an Actual Cost Network

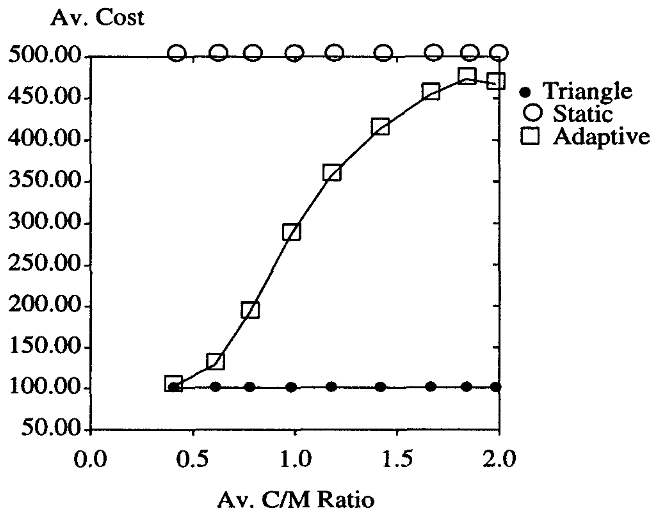
In the actual cost network scenario, the links of the 68-node network are assigned costs that represent the geographical distance between the nodes. One difficulty with the actual cost network is the wide disparity in costs between nodes. This disparity tends to blur the effect of C/M ratio on the performance of the various schemes. For instance, a source may be close to an MH but far away from the HA. In this case, the routing costs are high and update costs are low. Therefore, even at low C/M ratios, it becomes profitable to include the source in the update set of the MH. Placing the MH and the sources in different nodes usually only results in different values for costs, without change in the behavior of the algorithm. This effect is seen in Graph 6.4 where the average total cost for the triangle routing scheme is greater than the adaptive scheme at low C/M. Leaving aside this factor, it can be seen that the essential performance of each scheme is similar to that in the uniform cost network. It is clear that the adaptive scheme continues to adapt to varying C/M.

Graph 6.5 illustrates the average total cost for the boring professor mobility model with 5 nodes in the main set. To recall, this model allows the MH to move larger distances in the long run than the salesman model, which had a hop limit of 1. As such, it somewhat mimics the salesman model with a hop limit of 1. The behavior of the location schemes as seen in Graph 6.5 is identical to that observed previously. The adaptive scheme sticks to the best of static and triangle schemes. In Graphs 6.4 and 6.5, at higher C/M ratios, the adaptive and static schemes show a significant reduction in cost as compared to the triangle routing scheme.

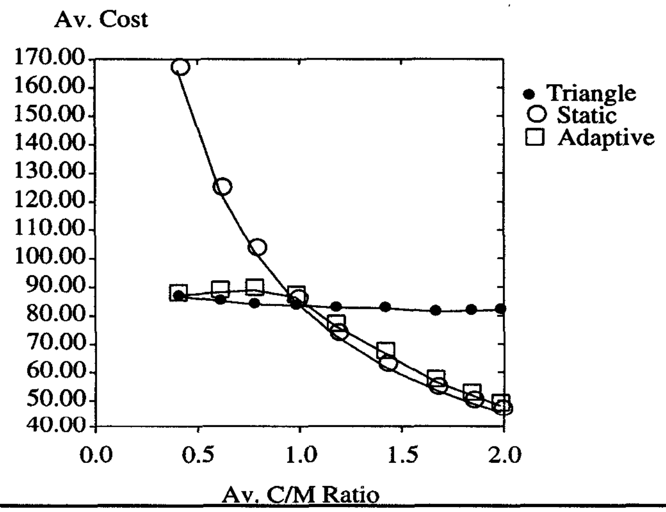
Graph 6.1: Av. Search Cost (Salesman, No hierarchy, MS=5, TS=0)



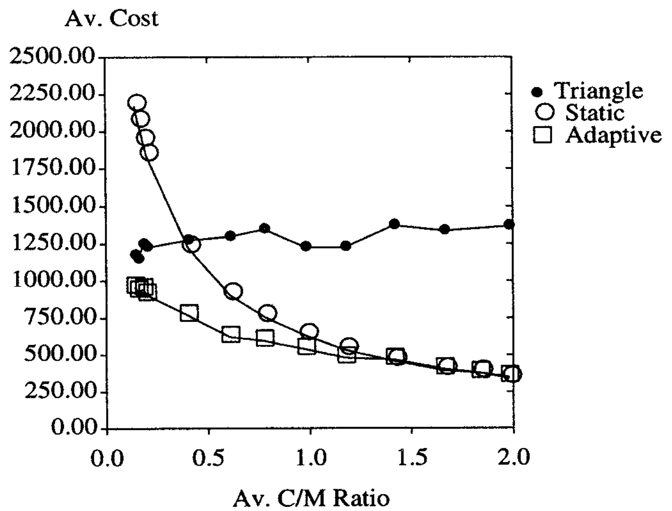
Graph 6.2: Av. Update Cost (Salesman, No hierarchy, MS=5, TS=0)



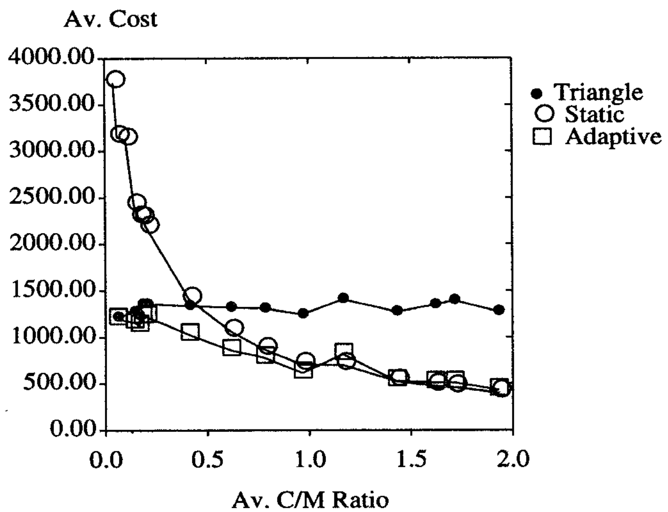
Graph 6.3: Av. Total Cost (Salesman, No hierarchy, MS=5, TS=0)



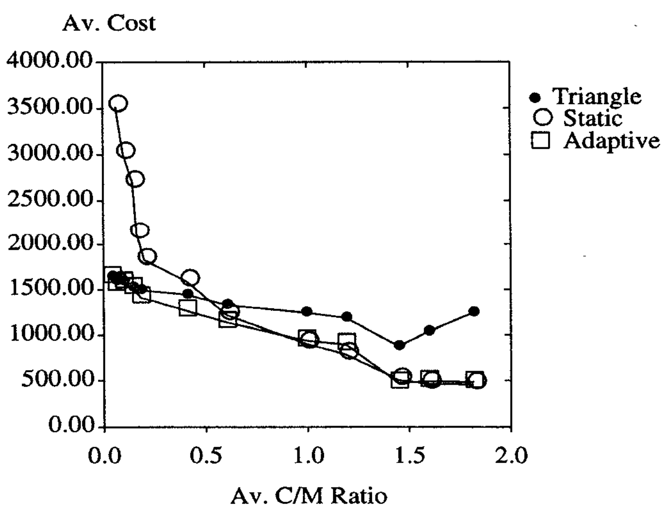
Graph 6.4: Av. Total Cost (Salesman, No hierarchy, MS=5, TS=0)



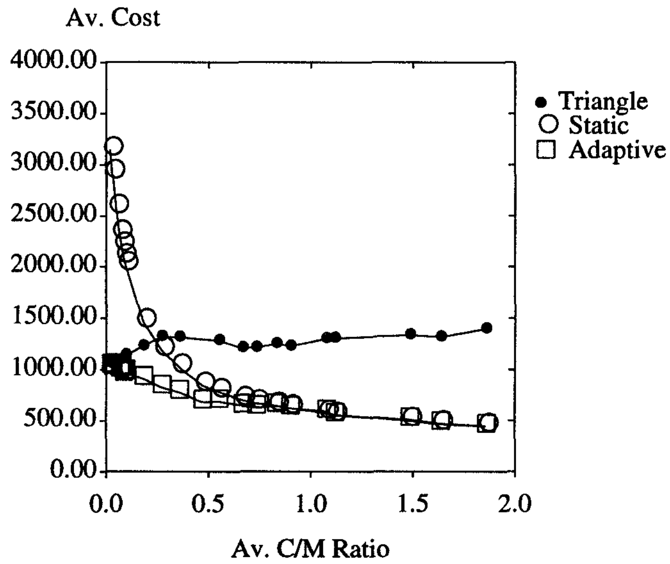
Graph 6.5: Av. Total Cost (Prof., No hierarchy, MS=5, TS=0)



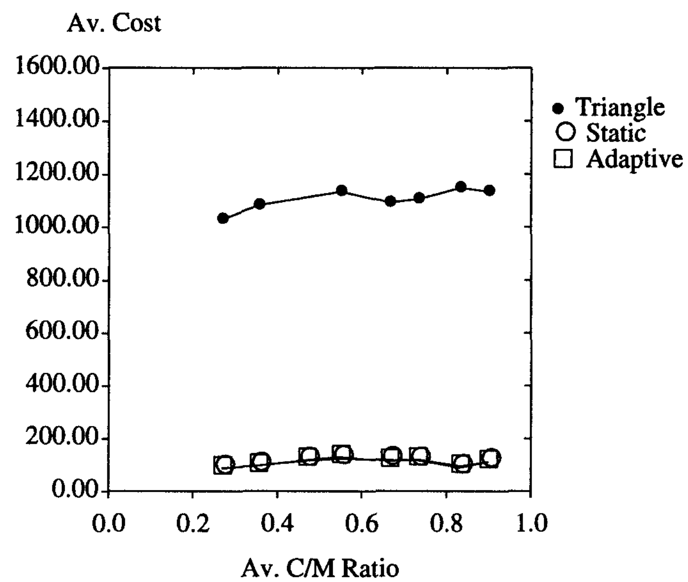
Graph 6.6: Av. Total Cost (Pop-up, No hierarchy, MS=5, TS=0)



Graph 6.7: Av. Total Cost (Salesman, No hierarchy, MS=5, TS=10)



Graph 6.8: Av. Total Cost (Salesman, Hierarchy, MS=5, TS=10)



The pop-up mobility model turns out to be the one that causes the three location management schemes to perform unpredictably. Graph 6.6 gives the average total cost for this model plotted against the average C/M ratio. While on the surface the performance of the three schemes are similar to that seen before, there are noticeable discontinuities in the manner in which the cost changes with the C/M ratio. Under the pop-up model, the MH leaves one big node and pops up in a remote node, thus completely changing the cost profile of searches and updates. This seems to affect the overall costs of the various schemes for the pop-up mobility model.

The effect of the number of sources on the behavior of the various schemes is shown in Graph 6.7. This graph depicts the average total cost as a function of the average C/M ratio. Of the total number of sources generating connection requests, 5 belong to the main set and 10 belong to the transient set. Sources belonging to the external set also generate requests in order to simulate the stray connection requests that a user might receive. It is clear from the result that regardless of the number of sources, the behavior of the adaptive algorithm is governed primarily by the C/M ratio of an individual source. It should be noted that in Graph 6.7, the C/M ratio shown has been approximated from the overall observed C/M ratio using weighted averaging. Recalling that 70% of the connections are generated from the 5 nodes in the main set, 20% of the connections are generated from the 10 nodes in the transient set and the remaining 10% are generated by the rest of the nodes, the overall C/M ratio is divided by 10.8 to get the average C/M ratio.

6.4 Hierarchical Location Management

When a location hierarchy is introduced, it benefits both static and adaptive schemes tremendously. For these schemes, it is as if the effective call-to-mobility has been lowered significantly. This is a result of slow movement across regions, i.e., the MH stays for a longer time within a big node, as opposed to a sub-node. So, with a hierarchy, these schemes tend to update sources in their update sets less frequently. Also, when the MH is within a big node, the connection set-up frequency need not be too high for a source to be included in the MH's update set. Thus, it was seen that all the sources in the working set get included in the MH's update set, for the adaptive scheme, resulting in low routing costs. Graph 6.8 illustrates exactly this phenomenon for the traveling salesman mobility model. Here, while the actual C/M ratio varies from 0 to 1, a majority of the moves are intra-region (i.e., within a big node) and result in no updates. This is true for all schemes, including triangle routing, but sources under triangle routing incur a lot of overhead due to redundant searches. Similar performance was observed for other mobility models.

8. CONCLUSIONS

The concept of a working set of hosts is developed for mobile hosts. Based on this concept, an *adaptive* scheme for location management is proposed, that enables an MH to dynamically generate its working set and trade-off routing and update costs in order to reduce the total costs of location management. Comparative analysis of triangle routing, static update and adaptive schemes shows that neither the triangle routing scheme nor the static update scheme performs well over a range of C/M ratios. The adaptive scheme, on the other hand, is shown to adapt better to varying call-to-mobility ratios and varying mobility patterns of the user. The behavior of this scheme is governed mainly by the C/M ratio for individual sources and not by the size of the working set. The uniform cost network clearly highlighted the adaptivity of this scheme, while in the actual cost network, the overheads were influenced by the location of the sources and the MH in the network and hence the path costs. It is also seen that the performance of the adaptive scheme is not affected by the number of sources in the working set. Two update schemes, based on hierarchical structure of the network are also considered. It is concluded that the static update and the adaptive scheme benefit a lot from a hierarchical organization of the location servers, especially when the cost of updating a regional server is insignificant.

9. ACKNOWLEDGEMENTS

We would like to thank David J. Goodman for his support and comments.

References

[1] P.D.Amer, R.N.Kumar, R.Kao, J.T.Phillips, L.N.Cassel, *Local Area Broadcast Network Measurement: Traffic Characterization*, IEEE Jan. 1987, pp 64-70.

[2] V.Anantharam, M.L.Honig, U.Madhow, V.K.Wei, *Optimization of a Database Hierarchy for Mobility Tracking in Personal Communications Networks*, Proc. of Performance '93 (16th Intnl. Symp. on Computer Performance, Modelling, Measurement and Evaluation), pp. 293-304, Spet. 1993, Rome, Italy

[3] B.Awerbuch, D.Peleg, *Concurrent Online Tracking of Mobile Users*, Proc. of the ACM, SIGCOMM, Symposium of Communication, Architecture and Protocols, Sept. 1991.

[4] A.Aziz, *A Scalable and Efficient Intra-Domain Tunneling Mobile-IP Scheme*, ACM SIGCOMM, Comp. Comm. Review, Jan. 1994.

[5] T.Blackwell *et al, Secure Short-Cut Routing for Mobile IP*, USENIX Summer 1994 Technical Conference, June 1994

[6] R.Caceres, P.B.Danzig, S.Jamin, D.J.Mitzel, *Characteristics of Wide-Area TCP/IP Conversations*, ACM SIGCOMM, 1991, pp 101-112.

[7] P.J.Denning, *The Working Set Model for Program Behavior*, Comm. of the ACM, Vol II, Number 5, May 1968, pp 323-333.

[8] M.Fujioka, S.Sakai, H.Yagi, *Hierarchical and Distributed Information Handling for UPT*, IEEE Network magazine, Nov. 1990, pp 50-60.

[9] T.Imielinski, B.R.Badrinath, *Querying in Highly Mobile Distributed Environments*, Proc. of the 18th VLDB, Aug. 1992, pp 41-52.

[10] T.Imielinski, B.R.Badrinath, *Wireless Computing: Challenges in Data Management*, Comm. of the ACM, pp 19-28, Oct. 1994.

[11] J.Ioannidis, G.Q.Maguire Jr, *The Design and Implementation of a Mobile Internetworking Architecture*, Winter USENIX, Jan. 1993, pp 489-500.

[12] D.B.Johnson, C.Perkins, A.Myles, *Route Optimization in Mobile IP*, draft-ietf-mobileip-optim.02.txt, Network WG, Internet Draft, July 1995.

[13] A.Myles, D.Skellern, *Comparison of Mobile Host Protocols for IP*, Internetworking - Research and Experience, 4(4), Dec. 1993, pp 173-194

[14] L.J.Ng, R.W. Donaldson, A.D. Malyan, *Distributed Architectures and Databases for Intelligent Personal Communication Networks*, IEEE, 1992, pp 300-304.

[15] C.Perkins (ed.), *IP Mobility Support*, draft-ietf-mobileip-protocol-11.txt, Mobile IP WG, Internet Draft, July 1995.

[16] C.Perkins, Y.Rekhter, *Optimal Routing for Mobile Hosts Using IPs Loose Source Route Option*, IETF Internet Draft, Oct. 1992.

[17] F.Teraoka, M.Tokoro, *Host Migration Transparency in IP Networks: The VIP Approach*, Computer Comm. Review, ACM, Jan. 1993, pp 45-65.

[18] J.Z.Wang, *The Hierarchical Structure of Tracing Strategy for Universal Personal Communication Systems*, April, 1992, IEEE.

[19] H.Xie, D.J.Goodman, *Signaling System Architecture Based on Metropolitan Area Networks*, WINLAB Tech. Report - 39, August 1992.

[20] R.Yates, C.Rose, S.Rajagopalan, B.R.Badrinath, *Analysis of a Mobile Assisted Adaptive Location Management Strategy*, Submitted to Intl. Journal for Wireless Information Networks.