Distributing Mobility Agents Hierarchically under Frequent Location Updates

D. Forsberg, J. T. Malinen, J. K. Malinen, T. Weckström, M. Tiusanen
TSE-Institute, Telecommunications and Software Engineering
Laboratory of Information Processing Science
Helsinki University of Technology, P.O. Box 9700
FIN-02015 HUT, Finland
E-mail: {dforsber,jtm,jkmaline,tweckstr,mikko}@cs.hut.fi

Abstract— **The proliferation of wireless LAN technologies and mobile terminals has prompted an increased need to support efficient and seamless roaming. Current mobility management protocols, such as the Mobile IP, as defined in RFC 2002, do not scale well into these requirements. Mobile IP employs mobility agents called home agent and foreign agent to support Internet-wide mobility. We present a distribution of the mobility agent functionalities into fully scalable, arbitrarily deep tree hierarchies of foreign agents.**

We show that by distributing part of the functionality of the home agent into the foreign network we increase the performance significantly and securely. Our performance measurements show that the reduction of network latency, due to signaling locality in the hierarchical mode, enables faster location update frequencies than in a RFC 2002-compliant non-hierarchical mode.

In our solution, the Dynamics - HUT Mobile IP version 0.6-pre4, we obtained an increase in the TCP stream cutoff location update frequency when using a hierarchical configuration. We also experienced similar increase when using fixed speed streamed UDP traffic with data speeds typical for some popular multimedia formats. Consequently, we claim that this kind of mobility agent architecture is suitable for scalable, fast-handoff networks where multimedia streaming is used.

Keywords— **frequent handoffs, handoff signaling, hierarchical tunneling, IP routing, Mobile IPv4, mobility management, signaling security, wireless mobility management.**

I. Introduction

Mobile users far from their home networks want the same services as they would get when attached to their office LANs and using the standard Internet protocols. Mobile internetworking will be an increasingly important element in future communication networks. Therefore, wireless network management, mobile agent technologies, protocol tunneling, and mobility management on the Internet are areas of growing interest.

Delivering data, like multimedia, to mobile computers sets high demands on the networking technologies in the physical, data link, and network layers. The current wireless LAN technology can deliver the high speed data, but lacks the ability to seamlessly scale mobility over the subnet boundary. To go beyond this motivates the use of network layer mobility management protocols such as the Mobile IP [1] with its Internet-wide mobility.

However, hard bandwidth requirements for multimedia data together with frequent changes in the location of the mobile user are not easily handled by the basic Mobile IP. Each location change requires signaling between the mobile computer and its home network and the latency of such an operation can be hundreds of milliseconds.

Additionally, the Mobile IP is not good at handling handoffs in a wireless network. It uses a timeout-based mechanism to detect the need to change the current point of access. This results in a noticeable pause, a glitch, in the connection, especially when the MN is moving and using multimedia streams on the network.

To meet the requirements of high mobility when using multimedia traffic on the Internet, we introduce a solution that re-distributes the tasks of the basic Mobile IP components. The mobility management in our solution is handled closer to the mobile computer.

In Section II, we consider the problem of distributing the mobility management functionality efficiently, and how this has been approached in some related work. Then, in Section III, we present our solution, the Dynamics – HUT Mobile IP version 0.6-pre4. It uses a mobility agent hierarchy in the foreign network. Therefore, it adds some protocol enhancements to the basic Mobile IP while being able to preserve downward compatibility with it. In Section IV, we present a set of measurements to show the difference in performance between a hierarchical and a non-hierarchical version of Mobile IP. Conclusions from this experiment and some suggestions for the future follow in Section V.

II. Distributing mobility agents under frequent location updates

In the Internet, routing is based on having the host location encoded into the IP address of the host. A host address is divided into a network prefix and the host number. The former is used to define the subnet where the host is permanently located. To enable Internet-wide mobility while preserving a permanent identity and uninter-

ruptible sessions, a mobility management protocol, such as the Mobile IP [1], is needed.

A. *Efficient location management*

There are several factors that affect the performance of location management in mobile internetworking. The mobile user must first *register* with the location management system to present her credentials. Then, the movement of the user causes *location management events* in the system.

Registration

The purpose of registering a user with the system is to establish that she is whom she claims to be and has the required privileges to use the system. In addition, an optional accounting mechanism to record the amount of used resources may be initialized. The registration should be secure. To achieve this, registration messages should be correct, confidentiality preserving, authenticated, and non-repudiated.

Registration implies that the mobile user's computer, the *Mobile Node* (MN) and the entity that provides the identity in the user's default location, the *Home Agent* (HA), have a mutual trust relationship. This is used in the registration to establish a *security association* between the MN and the HA. Additionally, since the location management system itself is distributed, the system may need to establish other security associations.

Location management event

The MN's arrival in a foreign network causes a location management event. Excluding the registration, a location management event is composed of two functions. First, the system must perform a search to establish knowledge of the user's location; then the system must update its state according to this information. The system must also periodically update its state when the location management maintains a lifetime for a connection.

Distributing the location management

Badrinath and Rajagopalan [2] consider a *working set* of hosts, a set of access points that an MN most frequently uses. In common mobility scenarios, this set is geographically clustered. Consequently, a location management system that supports optimizations based on the locality should have increased performance compared to those without such support.

The location management is supported by an interplay of tasks performed by the HA and the other mobility agent, an entity in the visited network called the *Foreign Agent* (FA). In a basic solution, the location information is stored mainly in the HA. To better support locality, we propose a distribution of the location information closer to the MN.

In this paper we consider the problem of distributing the functionality of mobility agents (FA and HA) to allow frequent and seamless location management operations while maintaining ongoing sessions and maximizing data throughput. Any other node, call them *Correspondent Nodes* (CNs), should be able to *transparently* communicate with MNs while the MNs are in the visited networks, that is, without any requirements on the CNs. Furthermore, the CN should not notice any difference between communicating with an MN at home or in the visited network. When an MN is moving and thus changing its location of attachment to the Internet it should maintain its sessions with CNs.

It can be expensive or even impossible for the MN to re-register with the HA. This could happen, for example, when the HA is far away and the MN is receiving real-time video data. If the MN changes its point of attachment while receiving the video, it could take a while for the HA to receive the registration of the new location, and the mobile user would perceive a glitch.

The frequency of location updates can be high, for example, in the wireless networks with a small cell size when the MN is crossing many cell boundaries. Signaling messages produce a traffic load on the path from the MN to the HA. When the number of MNs increases, the HA and the path between the visited network and the HA will encounter a high signaling load under frequent location updates.

Basic Mobile IP

In the basic Mobile IP, a set of mobility agents, the HA and the FAs, manage the routing between the mobile user's computer, the MN, and her default location in the home network when she is not there. The HA is a router on an MN's home network that tunnels datagrams for delivery to the MN. It thus maintains current location information for the MN. The location information is stored in a data structure called a *mobility binding*. The IP forwarding part of the router takes care of the data tunneling typically by using IP-IP encapsulation [3].

The FA is a router on an MN's visited network, or foreign network, which provides routing services to the MN while this is away from home, and registered to the mobility management system. Each MN uses its corresponding HA, while the FA just acts as an intermediate router. In this way, several MNs can use their corresponding HAs simultaneously over the same FA. The FA just delivers datagrams tunneled by the MN's HA to the MN. For datagrams sent by an MN, the FA may also serve as a default router for registered MNs [1].

The MN performs a registration with the HA. In advanced location management systems, the registration may include the use of authentication, authorization, and accounting (AAA) protocols, such as RADIUS [4], or DIAMETER [5]. The registration request – registration re-

ply message exchange is performed with the UDP protocol between the MN and the HA. The correctness is preserved by using a non-zero UDP checksum in the UDP header. Confidentiality is considered out of scope for the basic protocol, but authentication keys, and mechanisms such as link-level encryption are suggested, if confidentiality is wanted.

Once the MN has registered with the system, the location management is handled with a location update message. The search for location occurs through the interplay of the MN and the FA so that the FA broadcasts or multicasts advertisements that tell about existing access points. When the MN moves, it sends registration requests telling the system its current location. The update part is composed of upgrading the mobility bindings in the HA and the routing information in the system. Performance is affected by the delays involved in these tasks and signaling data transmissions.

B. Criteria for efficient location management

An ideal location management system should provide good performance in data throughput and signaling, reliable security characteristics, and transparency.

Handoff time is the time elapsed from the initiation of the location update to the retrieval of the registration reply signaling message in the MN. The smaller the handoff time is, the better the system can handle frequent location updates.

Data throughput in an ideal location management system would be constantly maximal. Under frequent location updates the comparison can be done to data throughput without location updates, and to data throughput without the Mobile IP scheme (i.e. without IP-IP encapsulation), to find out how the location updates decrease the throughput.

Average traveled distance of location update signaling messages, measured with the number of hops or traveling time in the network, should be minimal. This contributes to the location update cost that should be minimized.

Secure signaling is needed for session maintenance protection and MN authentication.

Transparency should be obtained so that the location management system is transparent to the MN and the CN. Additionally, the number and hierarchy of FAs between the MN and the CN should not be visible to any host other than the FAs in the current hierarchy.

C. Related work

There have been many suggestions for handling frequent location updates and minimizing packet loss in the Mobile IP scheme. These are basically of three kinds: implementational improvements involving buffering and different kinds of fine tuning, routing and other protocol-based solutions with, e.g., multicasting and broadcasting, and improvement suggestions to the Mobile IP protocol itself.

Woo and Leung [6] considered buffering and also proposed special extensions to Mobile IP and to the Internet Mobile Host Protocol (IMHP). Balakrishnan et al. [7] suggest intelligent buffering techniques and multicast-based routing to speed up the handoff, especially in a wireless Mobile IP environment.

In Foo's and Chua's Regional Aware Foreign Agent (RAFA) protocol [8], the distribution of agent responsibilities supports locality more than the basic Mobile IP. The FA is divided into two, the RAFA and the Local Foreign Agent (LFA). In this model, the RAFA is the proxy FA, known by the HA, while the LFA provides the point of access. The MN can register with the RAFA through the LFA instead of always registering with the HA. This brings an increase in the handoff efficiency when compared to the basic Mobile IP. El Malki, Fikouras, and Cvetkovic [9] describe a method to support fast handoff in a similar environment. It is based on the list of FAs in the agent advertisement and registration request.

McCann et al. [11] suggested a distribution of the FA functionality with a new type of specialized mobility agents called *surrogate agents*. They would create link layer tunnels to FAs before any Mobile IP signaling, thus making the distributed functionality and the hierarchy invisible to the MN. Additionally, THEMA draft discusses distributing the HA functionality. This works both in the foreign network where the surrogate agents reply to registration requests and in the home network where the HA may connect to a *home link* via surrogate agents. The home link then provides an access point for the MN and keeps the MN and the HA connected.

Caceres and Padmanabhan [12] describe fast and scalable handoffs for wireless internetworks. They implemented and tested a hierarchical solution based on Address Resolution Protocol (ARP) [14] and their own local mobility management protocol. With their solution, Caceres and Padmanabhan measured fast location update times with very small packet loss ratio. Their solution is based on two levels in the hierarchy. The security of Caceres's and Padmanabhan's solution is, as the authors admit, on the same level of security as when using ARP.

Perkins and Wang [13] proposed a solution that uses buffering with duplicate packet elimination to minimize the number of lost packets in the location updates. In addition they used hierarchical FA management to reduce the overhead of frequent handoffs. They simulated the proposed solution and were able to get substantial performance improvements.

III. Hierarchical mobility agent distribution

Some of the recent suggestions use hierarchy to improve the efficiency of the Mobile IP protocol. We propose a solution that does not restrict the number of levels in the hierarchy, builds only on the three elementary elements of Mobile IP: the MN, FA, and HA, and uses messages

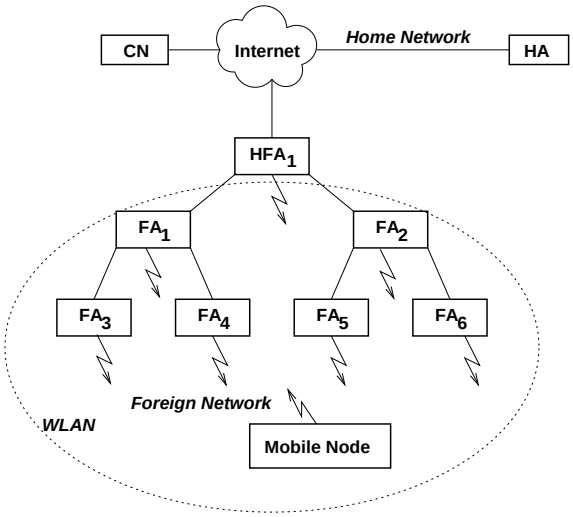

Fig. 1. Mobility agent hierarchy

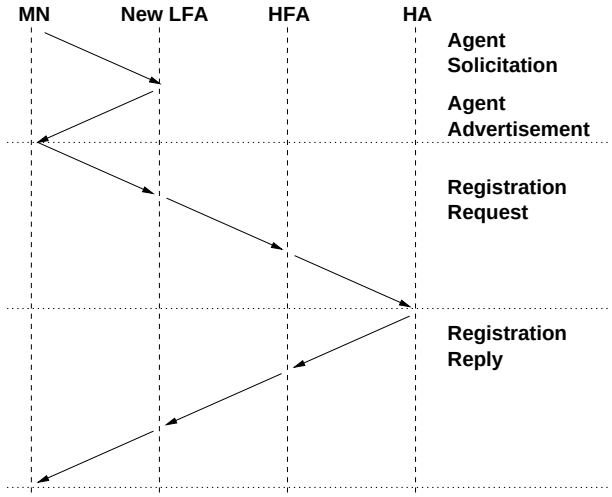

Fig. 2. Registration in a foreign network, full exchange

nearly as simple as the basic Mobile IP.

Our solution uses all the three fundamental improvement areas presented above. The emphasis is on routing related improvements and in protocol improvements. The implementation uses effectively the policy routing solutions available in the Linux operating system [15] kernels from the 2.2 versions upwards. The QoS features are also used so that the signaling messages have a higher priority than all other traffic. The solution complies with the basic Mobile IP. However, we have added functionality into the signaling protocol to enable localized frequent location updates.

We use a concept of FA hierarchies. Fig. 1 illustrates an example FA hierarchy. In our implementation it is possible to configure trees of FAs; these FA hierarchies within foreign networks can also be called *organizations*. This brings up new roles for FAs within the foreign networks.

Highest Foreign Agent (HFA) is the root of the FA hierarchy in an organization. *Lowest Foreign Agents* ($LFAs$) are those closest to the MN on a path from the MN to the HA. These are typically the leaves of the FA tree, though not necessarily, like in Fig. 1 where e.g. FA_1 or HFA_1 may also become an LFA. *Intermediate Foreign Agents* ($IFAs$) are those located between the LFA and the HFA on the path from the MN to the HA. As the MN moves, it changes its point of attachment, the LFA, thus changing also the path from the MN to the HA. The lowest FA that belongs to both the old and the new paths is called the *Switching Foreign Agent* (SFA).

The classification of FAs is only conceptual. The same software implementation is used for the FA functionality, no matter what the position of the FA is. This is important point for the implementation, as the hierarchy might collapse so that HFA, $IFAs$, and LFA are all in the same host, like in the original Mobile IP.

A. Signaling

In Fig. 2 and Fig. 3, we can see a typical registration procedure in a foreign network: first, when the registration reaches up to the HA, which is the case with every location update in a basic non-hierarchical Mobile IP system, and then a localized example. When the MN moves, it sends a registration request message to the new LFA, which creates an unconfirmed mobility binding and forwards the registration request upwards to the next higher IFA.

At some point, when the registration request is being forwarded up the hierarchy, it arrives at the SFA. This detects that the requesting MN already has a mobility binding, but the request is coming from a different lower FA. This means that the MN has changed its LFA. The SFA replies to the MN with a registration reply message

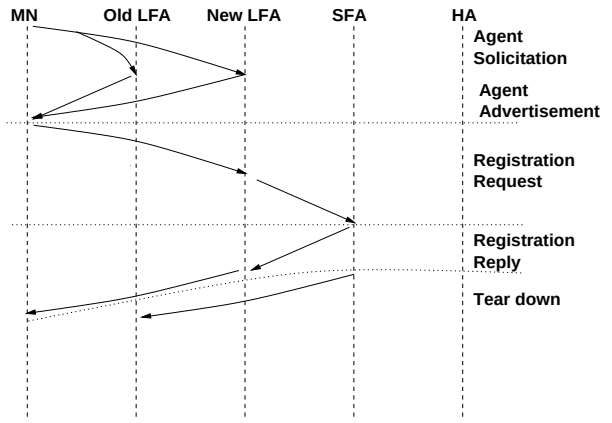

Fig. 3. Registration in a foreign network, localized exchange

[1, Ch. 3] containing the remaining lifetime of the mobility binding of that particular MN in the SFA. This creates an illusion that the HA has answered to the request. This functionality brings a part of the HA functionality down to the FAs.

The localized registration does not reach the HA and it cannot thus be used to update the binding lifetime at the HA. The MN is responsible for keeping the whole tunnel alive and it needs to force a registration up to the HA whenever the lifetime of the tunnel would be exceeded. The registration message can be send without the session key based authentication extension in order to force the FA organization to forward the request all the way to the HA. This HA binding update increases the signaling traffic between the home and visited networks. It is therefore reasonable to use long binding lifetimes at the HA so that the number of the needed messages can be limited. This way the location update messages need not be sent to the HA if the FA organization is not changed. However, the MN can still make sure that the binding will not expire at the HA or in any of the FAs in the path.

The SFA informs the lower part of the old path about the MN's movement by sending a tear-down message. This is a new functionality not included in the Mobile IP standard of RFC 2002. This is needed to prevent a situation in which an FA receiving a registration request from the MN is not aware of the changes made to the path higher in the hierarchy. This would result in the FA replying to the MN and not sending the registration request upwards even though it should, and the FA hierarchy would then be in an inconsistent state.

B. Security

Dynamics – HUT Mobile IP supports secure signaling through authentication, replay protection and IP address based configuration mechanisms. A separate session key management protocol and a specific extension are used to support secure localized location updates in the FA hierarchy.

MN and HA have a preconfigured shared secret that they use to authenticate each other. A *security parameter index* must be defined for every MN. It is used for indexing the security association at the HA [1]. The association includes a preconfigured shared secret. Keyed MD5 [16] is used as the authentication algorithm and it determines the *message authentication code* (MAC) by using a secure hash of the key and the message.

An FA may have optional security associations with other FAs, and the HFA similarly with the HA. If a security association exists, the session key can be encrypted with the help of the shared secret and thus man-in-the-middle style attacks can be prevented. If no security association is set for a certain $FA - FA$ or $HFA - HA$ pairs, public key encryption by RSA [17] is used. The use of public key encryption without key certificates makes it possible for active attackers to change the transmitted public key and try to capture the transmitted session keys. Our implementation has a feature to protect the key distribution from this kind of attacks. The FAs can advertise a MD5-based hash value of the HFA's public key so that the MN can add this hash code in the protected part of the registration request. This way the HA can check whether the public key has a valid hash code as it has a security association with the MN.

Timestamps or nonces are used for replay attack protection and message sequence detection between the MN and the HA. A specific extension is used to implement sequence number based replay protection between the FAs and the MN in localized location updates. Nonces and timestamps are protected by the Mobile-Home Authentication Extension [1], and sequence numbers by a session key based authentication extension.

Additionally, HAs and FAs may restrict the allowed signaling partners (MNs or FAs) with IP addresses or network addresses and network masks based configurations.

Session key management

The HA acts as a session key distributor for the FA hierarchy and the MN. Session key management protocol is used to distribute the session key into the MN's registered path that includes the HA, the FAs between, and the MN.

When the MN registers with the HA, each FA sends its RSA public key in the registration request to the next FA higher in the hierarchy, and the HFA correspondingly to the HA. If a security association exists between successive mobility agents, RSA is not used. The HA generates a session key from pseudo-random data. This key is sent in two copies in the registration reply: one for the HFA, encrypted with its RSA public key or its shared secret, and one for the MN encrypted with its shared secret. When the HFA receives the registration reply it decrypts the session key, and encodes it with the next lower level FA's public key or shared secret. This is repeated down the path to the MN. The MN in turn can decrypt its copy of the key with its own shared secret.

When a local location update occurs, the FAs in the old path already know the current session key and the SFA can authenticate the MN. MN authenticates itself to the FA hierarchy by computing a MAC for the signaling message using the session key. The MAC is then sent to the FA along with the registration request, and, as before, each FA along the new path sends its own public key to the next higher FA or uses a preconfigured shared secret. When the SFA receives the registration request, it checks the MAC with its own session key. It then sends a registration reply to the MN via the new path. The reply includes the session key encrypted with the next lower FA's public key or shared secret, and this way the FAs

in the new path will get the session key for the mobility binding.

IV. Performance analysis

We took a series of measurements to establish the latencies caused by the signaling protocol. We also measured how large location update frequencies could be obtained when transferring TCP or UDP data, with or without using a hierarchical system of FAs. A throughput suitable for a video streams of 1.4 Mbps, such as a near TV-quality MPEG-1 [18], was chosen as the speed for the paced UDP streams. Motivation for the latter measurements was to find out how frequently the location updates could be performed with our solution when using a representative multimedia application.

A. Testbed setup

A controllable closed-network testbed was constructed (Fig. 4) where we had two organizations and a mobile that could roam within and between them. Additionally, we configured a third, non-hierarchical, RFC 2002-compliant organization with four independent FAs. The same HA and MN were used in all the tests.

For the hierarchical case, we constructed a full 3-level deep binary tree of FAs into both organizations. The MN could connect in any preconfigured order to the FAs enabling handoffs between any pair of FAs in the figure. The effect of the Internet was emulated by a latency filter in the delay host (Fig. 4) through which all traffic from the home network, or the CN, to the FA networks passed.

All mobility agents resided in physically different hosts where the MN, $HFAs$, HA, and the CN were Pentium class hosts whereas all the other FAs were less efficient, custom-built, 486-based embedded access point hosts, called Martnodes [19], with a wired and a wireless network interface. The wired hierarchy in both networks was in private address-space subnets, like the network between the $HFAs$ and the delay host. We used Lucent's IEEE 802.11 [20] compatible 2 Mbps WLAN adapters in the FAs and the MN in ad hoc mode. The HA and the CN were in a network behind the delay host. All the networks were dedicated to the test.

The MN used a wireless network for all its communication with the FAs, and all the other data (i.e. between FAs and HA and between HA and CN) were transferred in the wired network. The clear bottleneck on the network was the wireless part. The maximum obtainable throughput without location updates was about 1.4 Mbps using TCP and 1.6 Mbps using UDP.

B. Measurements

We ran all the tests using FA decapsulation [1] and reverse tunneling [10] on the wireless environment. During a handoff, we did not use the wireless extensions of our implementation that can select the most appropriate FA for the MN using the monitored radio link quality. In these tests we merely forced the MN to follow a predefined FA path and handoff frequency. Therefore, the MN did not use the agent discovery part of the Mobile IP. In practice, the MN received the agent advertisements, but it completed a location update only when requested by a test script.

We configured security associations between the HA and the MN, and separately in each organization between its FAs. The $HFAs$ and the HA did not have a preconfigured shared secret. Therefore, they used RSA public key encryption with 768-bit public keys for session key distribution. This resulted in a noticeable latency when the organization was changed and a new session key was distributed for the new HFA. All the other key distribution operations could be done with the much faster method – using keyed MD5. This kind of configuration corresponds to the case where we do not have the complexity of managing shared secrets between the home networks and each foreign network. However, it is feasible to use shared secrets between FAs in one administrative organization as is the usual case with FA hierarchies. The third, non-hierarchical organization of four FAs did not use the session keys, so the RSA encryption was not needed. However, we configured a shared secret between these FAs and the HA to minimize the security overhead in our implementation with the non-hierarchical case.

Our implementation has optimizations that limit the number of lost packets by keeping the old and the new tunnel up simultaneously during location updates. In the test environment, this could affect the measurements as the number of FAs was limited. To see the worst case, where the MN would always arrive at a previously unvisited FA, we removed the old tunnel immediately when receiving the tear-down message thus canceling the tunnel optimization. The lower part of the tunnel was then rebuilt every time the MN moved.

We forced the location updates so that the MN started a new registration with given intervals. If the MN could not complete the previous registration before the new one began, our script added a firewall filter to drop the incoming packets from the CN until a registration succeeded. This corresponds to the situation in which an MN is too fast for the registration procedure to complete in time.

In the first test, we measured latencies in a large set of location updates. Typical latencies from three different types of hierarchy traversals can be seen in Fig. 5. Fig. 6 shows in detail the location update latencies between FA_1 and FA_2. These latencies had a peak at 24 ms. A similar peak for location updates between FA_3 and FA_5 was at 43 ms and between FA_{13} and FA_{15} at 62 ms. The SFA in all these cases was HFA_1 as it was the first FA that was in both the old and the new path to the HA. The new

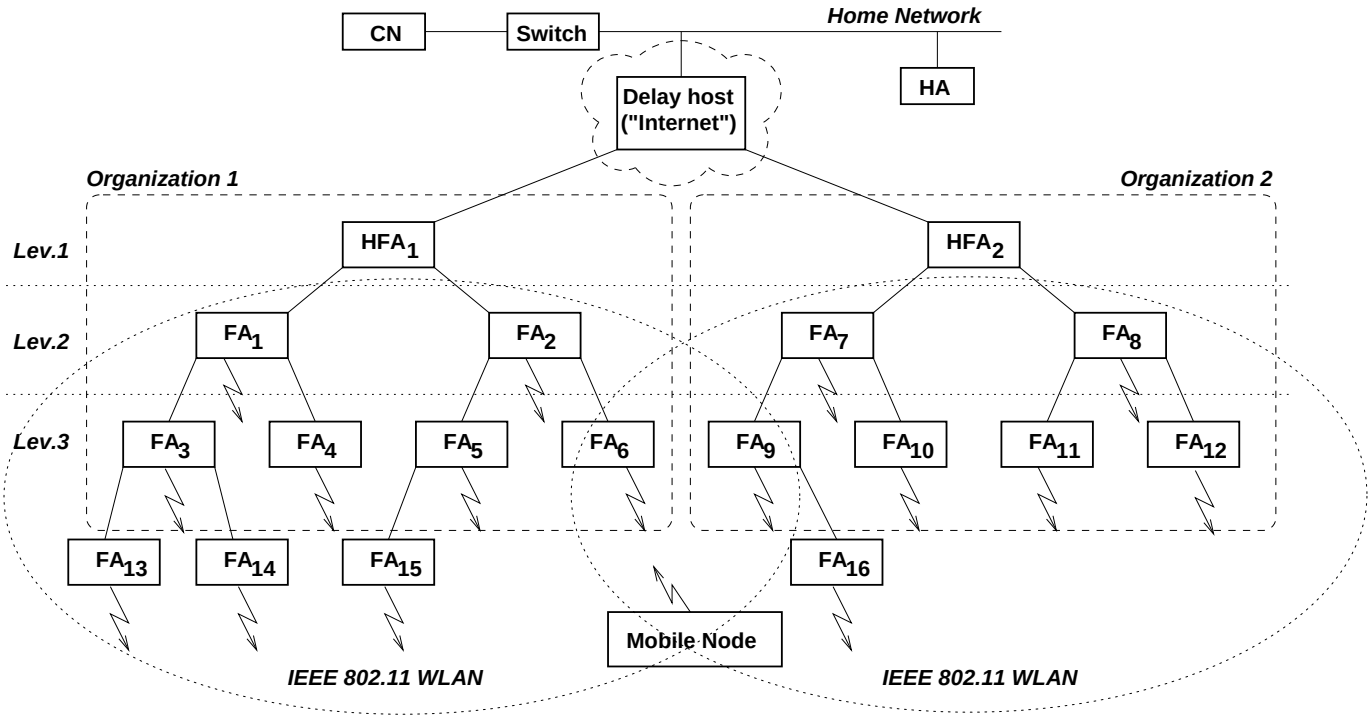

Fig. 4. A testbed setup with two organizations of FAs

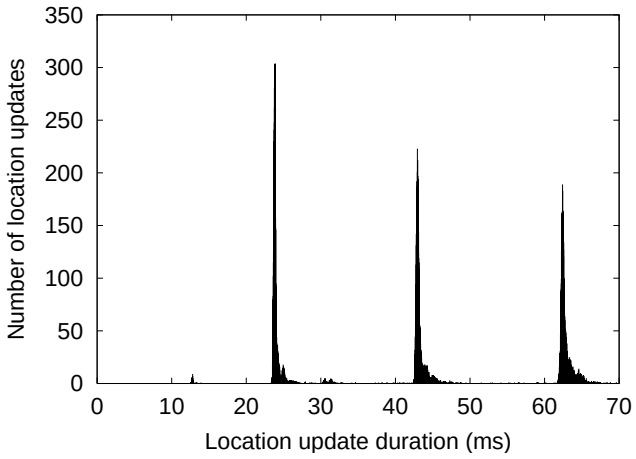

Fig. 5. Handoff latency $FA_1 - FA_2$, $FA_3 - FA_5$, and $FA_{13} - FA_{15}$

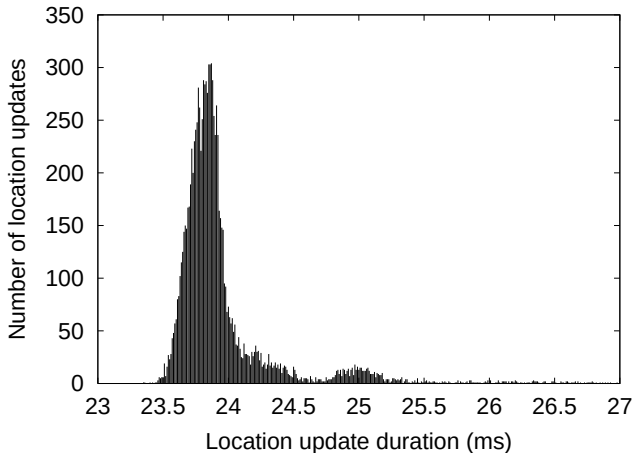

Fig. 6. Handoff latency $FA_1 - FA_2$

additional hierarchy level added about 19 ms of latency to the location update. In our tests, we could not observe non-linearities when scaling the system to multiple levels of hierarchy.

The average and standard deviation of latencies for 10000 samples of each handoff type (Table I) did not include any delay between the $HFAs$ and the HA. This would only have added a constant to the location update latency when the HFA is changed in the hierarchical case or when the FA is changed in the non-hierarchical case.

In the second test, we measured the TCP throughput as a function of location update frequency (Fig. 7). We varied the frequency from 0.1 to 25 location updates per second. We wanted to find out, how frequent location updates our implementation could handle without special link layer support. As a result of this some of the test runs used frequencies that are quite high for many environments. In the hierarchical case, the MN was moving between FA_3, FA_4, FA_5, and FA_6 (Fig. 4) with a given frequency. In the non-hierarchical case the MN moved between the four independent FAs.

TABLE I

LOCATION UPDATE LATENCIES FOR SOME TRANSITIONS

Handoff type	Org. changed	Average in ms	Std. dev. in ms
$FA_1 \rightarrow FA_2$	no	24.0	2.2
$FA_1 \rightarrow FA_3$	no	32.5	3.7
$FA_1 \rightarrow FA_5$	no	43.2	2.9
$FA_1 \rightarrow FA_{13}$	no	52.5	4.0
$FA_1 \rightarrow FA_{15}$	no	62.6	5.4
$FA_1 \rightarrow FA_7$	yes	97.8	0.7
$FA_1 \rightarrow FA_9$	yes	117.8	0.7
$FA_1 \rightarrow FA_{16}$	yes	137.7	1.1
$FA_3 \rightarrow FA_1$	no	13.2	3.2
$FA_3 \rightarrow FA_2$	no	24.0	2.4
$FA_3 \rightarrow FA_4$	no	29.8	2.3
$FA_3 \rightarrow FA_{13}$	no	30.5	3.0
$FA_3 \rightarrow FA_5$	no	43.3	2.5
$FA_3 \rightarrow FA_{15}$	no	62.9	3.3
$FA_3 \rightarrow FA_7$	yes	97.7	0.4
$FA_3 \rightarrow FA_9$	yes	118.1	1.3
$FA_3 \rightarrow FA_{16}$	yes	137.5	0.7
$FA_{13} \rightarrow FA_3$	no	13.0	2.3
$FA_{13} \rightarrow FA_1$	no	14.4	2.3
$FA_{13} \rightarrow FA_2$	no	24.1	2.6
$FA_{13} \rightarrow FA_{14}$	no	28.7	2.9
$FA_{13} \rightarrow FA_4$	no	29.5	2.4
$FA_{13} \rightarrow FA_5$	no	43.3	4.3
$FA_{13} \rightarrow FA_{15}$	no	62.6	3.9
$FA_{13} \rightarrow FA_7$	yes	97.7	0.5
$FA_{13} \rightarrow FA_9$	yes	117.8	1.8
$FA_{13} \rightarrow FA_{16}$	yes	137.5	0.7

The traffic was produced by `netperf(1L)` together with Linux 2.2.9 default settings for network protocol parameters. We changed the minimal routing cache flushing delay on the nodes in order to make the routing table changes immediately. The size of the sent messages was 4096 bytes and the data stream direction was from the CN to the MN. In Fig. 7 we have no simulated latency between the CN and the MN, while in Fig. 9 we have the same test with a latency of 100 ms. Latencies to distant hosts on the Internet can be even larger.

In the third test, we examined the packet loss with and without the hierarchy as a function of location update frequency when using 1.4 Mbps video-speed UDP streams (Fig. 8). Measurements for 128 kbps audio-speed UDP streams produced similar results. We transmitted UDP packets with a payload of 1000 bytes from the CN to the MN.

C. Discussion

Our results showed that the use of a hierarchical FA network provides considerable advantages over the basic RFC 2002 operation. It seems that the Mobile IP can be extended to work with frequent local location updates of even more than 10 times per second this way. When

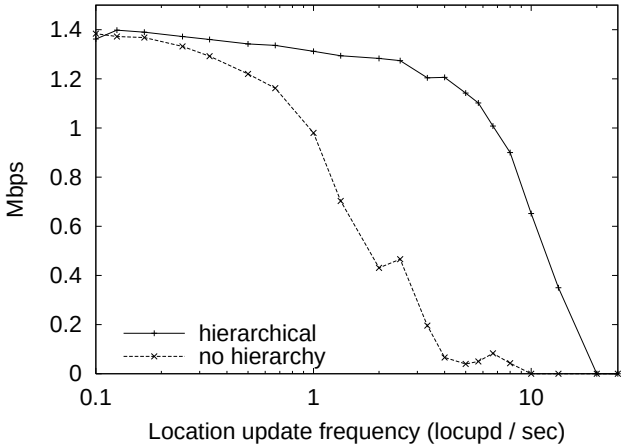

Fig. 7. TCP throughput, no latency

the FAs have the ability to reply to the registration requests, the possibly large latency between the HA and the visited network does not disrupt the communication as easily as with the non-hierarchical case. The hierarchical implementation also reduces the amount of signaling data passed between the HA and the visited network.

The measured latencies of the location updates did not reveal any non-linearity so there is no need to limit the number of levels in the hierarchy to two as is done in some designs for hierarchical Mobile IP [8]. The possibility to construct deeper hierarchies can be advantageous for organizations that like to hide the internal structure of the visited network routers. In addition it is a natural choice for many network topologies. Typically, each organization would have a HFA that could use efficient hardware-based routing and encapsulation to provide enough bandwidth. Each building could have the next level FA and each access point would be in yet another level of hierarchy. In this kind of a construction, the location updates

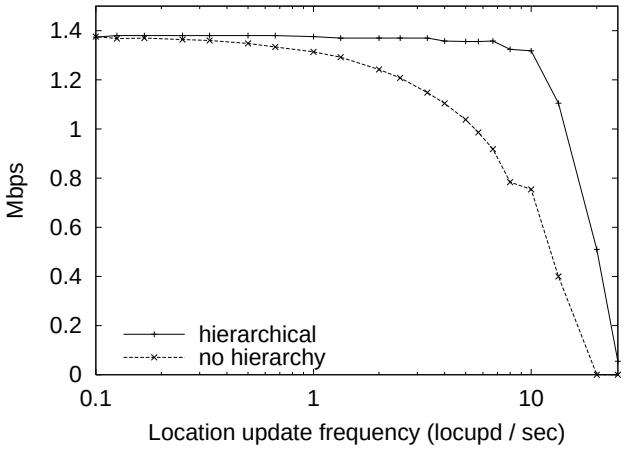

Fig. 8. Video-speed UDP streams, no latency

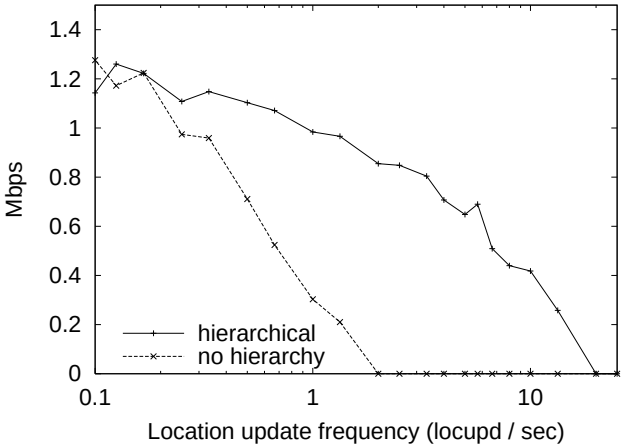

Fig. 9. TCP throughput, 100 ms latency

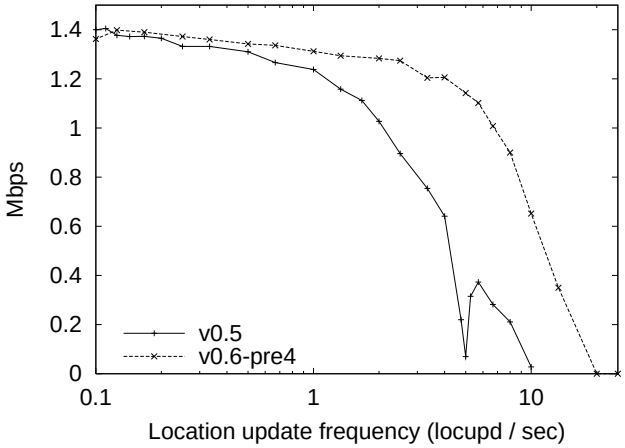

Fig. 11. TCP throughput, with hierarchy, no latency

are brought close to the MN. This will also lessen the signaling workload of the highest FA and its main job would be to provide enough routing capacity.

We did not tune the data transport protocols, though the measurements showed that the TCP throughput quickly degraded when packets were lost. The throughput dropped much faster than in the corresponding UDP test when we used a latency of 100 ms in the data path (Fig. 10). The simulated latency varied between 95 and 105 milliseconds. The loss of packets can be significant in wireless LANs so transport protocol improvements could be useful.

The performance was also sensitive to small changes in the design and use of the mobility agents. This could be observed e.g. from the TCP throughput when using a 200 ms location update frequency (Fig. 11). Earlier, with version 0.5 of the mobility agents and encapsulation up to the MN, a clear drop of the throughput was seen in the TCP tests without the delay when exactly five location updates were completed in a second. The corresponding UDP measurement did not show such a drop and the reason for this seems to be in a TCP timer that will halt the transmission momentarily when a significant amount of packets are lost on 200-ms intervals. With the same measurement and using version 0.6-pre4 of the mobility agents with FA decapsulation, not sufficiently many packets were lost to show such a drop in the throughput.

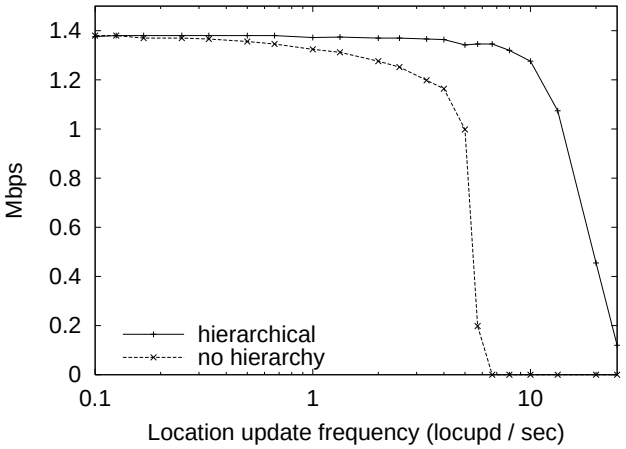

Fig. 10. Video-speed UDP streams, 100 ms latency

The processing speed of the agents seems to have a noticeable effect on throughput. Most of the lost packets were dropped while constructing the new path. The signaling messages need processing on each agent so the data packets can go a bit faster through a node. They were thus sometimes dropped, as a new tunnel was not ready when the data packets arrived.

Our testbed had a nearly ideal network environment since it did not drop packets significantly. The data throughput tests indicate that the hierarchical case has an advantage over the non-hierarchical one when the location update frequency is near the total cutoff point. In an ideal environment, the advantage is evident in the signaling latencies enabling operation, even if location updates occur more frequently than the latency between the visited network and the home network would allow. Furthermore, the reduced amount of signaling traffic between the HA and FAs can give significant advantages when multiple MNs of a HA are changing their location frequently.

V. Conclusions and future work

In this paper, we have presented a hierarchical version of Mobile IP that distributes the roles of the mobility agents. The mobility bindings are cached in the access network and the system protects their use with a session key protocol. This enables secure localized location updates with efficient signaling. The agent advertisement and the registration request contain only the address of the HFA, and

the registration request travels only to the closest FA that can handle the routing change. This results in a scalable system and enables the use of private addresses between intermediate and lowest FAs. The use of a short binding lifetime decreases the advantage of localized location updates and increases the signaling load on the HA and the FA hierarchy. When a binding expires, it forces the MN to send a binding update up to the HA. The binding lifetimes with this system need not be short, however.

The efficiency of the solution as compared to a non-hierarchical approach was shown to be significant when the handoff frequency was high. When we use signal-based handoffs in wireless environments, the system presented can provide the handoff speeds needed for glitchless multimedia streaming. The measurements indicated that in an ideal environment, which is not dropping noticeable amounts of packets, the hierarchy has a benefit on the data throughput when the location update frequency is near the cutoff limit. A similar testing should be done on a non-ideal system that drops, duplicates, and reorders packets.

Implementing the presented architecture revealed that it requires thoroughness to get a system that behaves correctly and maintains a consistent state in the FA hierarchy. In the version measured, loss of tear-down messages or certain other anomalies may cause problems. Therefore, a formal proof of correctness using the most recent version of this protocol would be useful. To further improve the performance, more prioritizing of traffic could be used, for example, to reduce the bandwidth based on user, paid fee, or type of service. Also, adaptivity could be introduced, e.g. the retry time for registration requests could be made adaptive to registration delays. This would suggest that lowering the minimum time between registration requests, as specified in the RFC 2002, is feasible in this architecture. Furthermore, a secondary fast agent discovery could also be practical.

In settings where latencies exist and fast mobility is required, Mobile IP [1] with our extensions can be useful. The increase in efficiency becomes the more relevant the more latency we introduce between the MN and the home network. If large latencies exist between the FAs in the foreign network, a similar advantage can be obtained from multiple levels of hierarchy.

Acknowledgments

We would like to thank Prof. Hannu Kari for comments and the initiative to build the system, as well as N. Asokan for his enlightening comments on security. Also, we would like to thank Björn Andersson, Jari Hautio, and Kimmo Mustonen for their roles in the implementation.

References

[1] C. Perkins, editor, "IP mobility support," IETF RFC 2002, IBM Watson Research Center, October 1996.

[2] S. Rajagopalan and B. Badrinath, "An adaptive location management strategy for Mobile IP," Dept. of Computer Science, Rutgers University, NJ, 1996.

[3] C. Perkins, "IP encapsulation within IP," IETF RFC 2003, IBM Watson Research Center, October 1996.

[4] B. Aboba, "Support for Mobile IP in RADIUS," IETF Internet Draft, work in progress, Microsoft, April 1998.

[5] P. Calhoun and C. Perkins, "DIAMETER Mobile IP extensions," IETF Internet Draft, work in progress, Sun Laboratories, November 1998.

[6] W. Woo and V.C.M. Leung, "Handoff enhancement in Mobile IP environment," in *Proceedings of the 5th IEEE International Conference on Universal Personal Communications*, vol. 2, pp. 760–764, 1996.

[7] H. Balakrishnan, R.H. Katz, and S. Seshan, "Handoffs in cellular wireless networks: the Daedalus implementation and experience," in *Wireless-Personal-Communications*, vol. 4, no. 2, pp. 141–62, March 1997.

[8] S. Foo and K. Chua, "Regional Aware Foreign Agent (RAFA) for fast local handoffs," IETF Internet Draft, work in progress, National University of Singapore, November 1998.

[9] K. El Malki, N.A. Fikouras, and S.R. Cvetkovic, "Fast handoff method for real-time traffic over scaleable Mobile IP networks," IETF Internet Draft, work in progress, University of Sheffield, June 1999.

[10] G. Montenegro, editor, "Reverse tunneling for Mobile IP," IETF RFC 2344, Sun Microsystems, Inc., May 1998.

[11] P. McCann, T. Hiller, J. Wang, A. Casati, C. Perkins, and P. Calhoun, "Transparent Hierarchical Mobility Agents (THEMA)," IETF Internet Draft, work in progress, Sun Laboratories, March 1999.

[12] R. Caceres and V. Padmanabhan, "Fast and scalable handoffs for wireless internetworks," in *Proceedings of the 2nd IEEE Annual International Conference on Mobile Computing and Networking (MOBICOM)*, Rye, NY, USA, November 1996.

[13] C. Perkins and K.-Y. Wang, "Optimized smooth handoffs in Mobile IP," in *Proceedings of the 4th IEEE Symposium on Computers and Communications*, 1999.

[14] D. Plummer, "An Ethernet Address Resolution Protocol," IETF RFC 826, Massachusetts Institute of Technology, November 1982.

[15] L. Torvalds, "Linux: a portable operating system," M.Sc. Thesis C–1997–12, University of Helsinki, Department of Computer Science, 1997.

[16] R. Rivest, "The MD5 message-digest algorithm," IETF RFC 1321, MIT Laboratory for Computer Science and RSA Data Security, Inc., April 1992.

[17] R. Rivest, A. Shamir, and L. Adleman, "A method for obtaining digital signatures and public key cryptosystems," *Communications of the ACM*, Vol. 21, pp. 120–126, 1978.

[18] "ISO/IEC 11172:1993 Information technology – Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s," ISO/IEC, 1993.

[19] "Mobile Ad hoc Routing Testbed (MART)," Helsinki University of Technology. TSE-Institute, Telecommunications and Software Engineering, URL:http://www.cs.hut.fi/Research/Mart, 1999.

[20] "Local and metropolitan area networks – IEEE 802.11-1997 standard for wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications," NY, USA: The Institute of Electrical and Electronics Engineers, Inc., 1997.