Optimized Smooth Handoffs in Mobile IP

Charles E. Perkins
Sun Microsystems, Inc.
901 San Antonio Rd.
Palo Alto, CA 94303, U.S.A.
Charles.Perkins@Sun.Com

Kuang-Yeh Wang
Department of Computer Science
University of Maryland
College Park, MD 20742, U.S.A.
kwang@cs.umd.edu

Abstract

Mobile IP provides mobility support for hosts connected to the Internet without changing their IP addresses. Route optimization strategies complement Mobile IP, to alleviate triangle routing, by informing correspondent hosts of the mobile host's care-of address. In this paper, we propose further strategies that are compatible with route optimization and its security model. First, foreign agents buffer packets for a mobile host and send them to its new location when it leaves. Second, hierarchical foreign agent management reduces the administrative overhead of frequent local handoffs, using an extension of the Mobile IP registration process so security can be maintained. Duplicate packets due to buffer handover are eliminated with the cooperation of mobile hosts. Simulation results show substantial performance improvements in terms of throughput, registration overhead, lost and duplicate packets during a handoff, without restrictions on physical placement of foreign agents.

1. Introduction

The emergence of the World Wide Web as a major driver for demand for the Internet access has proven the success of the internetworking technologies, specifically the Transmission Control Protocol (TCP) and the Internet Protocol (IP), which have integrated a wide range of different physical networks into a global Internet. At the same time, wireless communications technologies and various portable computing devices, ranging from laptop and palmtop computers to personal digital assistants (PDAs) and wearable computers, promise a new era of nomadic computing and seamless access to the global network of information.

Integrating wireless networks into the global Internet poses a new challenge. The main reason is that the TCP/IP based Internet technologies were designed for wired networks with mostly fixed hosts. Host mobility requires changes in the routing protocol so that packets for a moving host can be delivered to their correct destination. Mobile IP [10] provides a basic framework to solve this operability problem, with the assumption that there is enough infrastructure support so that a mobile host (MH) can communicate with a base station (BS), which is statically connected to the Internet.

However, several performance problems in Mobile IP need to be addressed. First, Mobile IP's tunneling scheme creates a *triangle routing* problem, causing packets to travel through sub-optimal routes. Second, packets in flight during a handoff are often lost because they are tunneled based on out-of-date location information. Third, base stations with small cells result in frequent handoffs, and requiring a registration with a distant home agent (HA) for each such local handoff causes higher overhead and further aggravates packet loss. Mobile IP route optimization [13] alleviates triangle routing and data loss during a handoff by informing the correspondent host and the previous foreign agent (FA) of the mobile host's care-of address.

In this paper, we propose to alleviate the frequent local handoff problem by using a hierarchical foreign agent management scheme (a previous version of which was submitted to the `mobile-ip` working group of the IETF [11]). With a hierarchy of FAs, small changes of location can be handled by one of the FAs in the hierarchy within whose covering range the MH remains. Our scheme extends the registration and authentication process in the base Mobile IP so that it is independent of the physical configuration of the foreign network and provides the same level of security as the base Mobile IP.

To reduce data loss during a handoff, we propose a buffering scheme at the FAs. The FA buffers any data packets it is forwarding to an MH. When a handoff occurs, the MH includes a handover request in its registration, and the new FA in turn requests that the previous FA hand over the buffered packets to the new location. To reduce duplicates, the MH buffers the identification and source address fields in the IP headers [14] of the packets it receives and includes them in the buffer handover request so that the previous FA does not need to transmit those packets that have already been received by the MH.

The rest of the paper is organized as follows. We first give a brief overview of Mobile IP and route optimization in Section 2. Descriptions of hierarchical FAs, the modified registration process, and FA buffering are in Sections 3 and 4. Related work is surveyed in Section 5. Our simulation results are presented in Section 6. Finally, we present our conclusions in Section 7.

2. Mobile IP and Route Optimization

The goal of Mobile IP is to provide mobility support for hosts connected to the Internet without changing their IP addresses. When an MH moves to a new location, it registers its current care-of address (e.g., the IP address of a base station, called a foreign agent (FA), to which it is currently connected) with its home agent, which is attached to its home network. When a packet for the MH arrives at its home network, the HA intercepts and forwards it to the care-of address by encapsulation [9]. The FA then decapsulates the packet and deliver it to the MH. A detailed description can be found in [10].

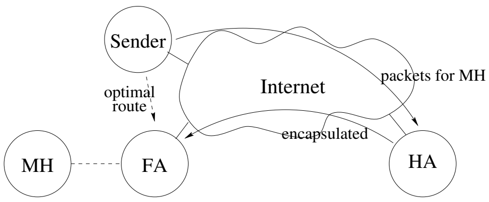

Figure 1. Triangle Routing Problem.

Mobile IP suffers from *triangle routing*: while the MH can send out packets through the FA and along an optimal path, incoming packets have to travel through the HA. If the current location of the MH is close to the sender's but the HA is far away, packets have to take a long detour (see Figure 1). Mobile IP route optimization [13] provides a mechanism to alleviate this problem. Any host which is willing to participate maintains a binding cache. When the HA intercepts a packet for an MH that is away, it may send a *binding update* message to the source of the packet, informing the source of the MH's current care-of address. The source then updates its binding cache, and tunnels any ensuing packets for the MH *directly* to its care-of address.

FAs can make use of binding updates to reduce packet loss during a handoff. As an extension of the registration process, the MH may ask the new FA to send a *Previous Foreign Agent Notification* message, which includes a binding update, to the previous FA. The previous FA then updates its binding cache and re-tunnels any packets for the MH to its new care-of address. This process is called *smooth handoff*.

These forwarding mechanisms, however, open a potential security hole for Internet hosts. Without Mobile IP, a malicious host may have to physically break into a network to grab packets for another host. To eliminate security exposures, Mobile IP requires that registration requests be authenticated. An *identification* field in registration messages provides replay protection. Details can be found in [10].

Authentication procedures are also required in route optimization for protection from similar attacks. A source host that accepts *binding update* [13] messages must authenticate them. During a smooth handoff, the previous FA must authenticate the binding update message using a registration key established with the MH [12].

3. Hierarchical FAs

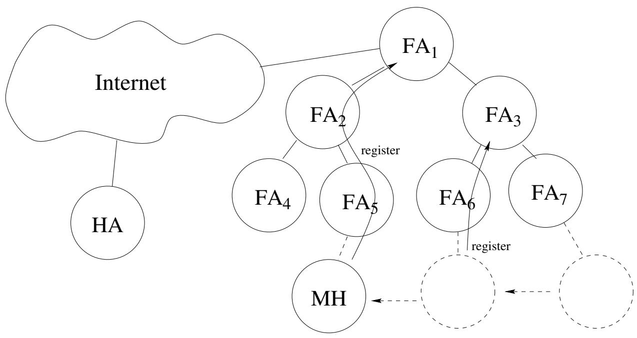

Figure 2. Hierarchical FAs.

As mobile devices are being made ever smaller and more convenient, they often require less power consumption to avoid carrying larger and heavier batteries. Less powerful wireless transceivers cannot reach very long distances. On the other hand, smaller and denser cells provide higher aggregate bandwidth and can locate a mobile device more accurately (cf. [3]).

The downside of a wireless network with small cells is that a moving host may cross cells very often, resulting in frequent handoffs. If the mobile node leaves one registration area before registering its next care-of address, packet losses may occur. Frequent handoffs aggravate this performance problem.

Our hierarchical FA management is aimed at alleviating this problem. The FAs in a domain are organized into a hierarchy (more precisely, a *tree* of FAs) to handle local movements of MHs within the domain (Figure 2). Such a tree of FAs can be chosen and configured in any way as the network administrator deems appropriate. One popular configuration might be to have a root FA associated with a domain firewall, and to have all other FAs at the second level of the hierarchy.

3.1. Registration and Data Forwarding Processes

An FA includes in its *Agent Advertisements* a *vector* of care-of addresses, which are the IP addresses of all its ancestors as well as its own. When an MH arrives at an FA, it registers with its HA not only that FA as the care-of address, but all its ancestors. A registration goes through and is processed by the FA, all its ancestors and the HA.

When a packet for the MH arrives at its home network, the HA tunnels it to the *root* of the FA hierarchy. When an FA receives such a tunneled packet, it re-tunnels it to its next lower-level FA. Finally the lowest-level FA delivers it directly to the MH. Therefore, any FA processing a registration should record the next lower-level FA as the other end of the forwarding tunnel (in other words, as the MH's care-of address).

As an example, in Figure 2 when MH first arrives at FA_7, it registers FA_7, FA_3, FA_1 as its care-of addresses. A registration request goes through this path to HA, and a registration reply the same path in the opposite direction. A packet for MH is intercepted by HA and tunneled to FA_1, which re-tunnels it to FA_3, which again re-tunnels it to FA_7, which delivers it directly to MH.

When a handoff occurs, MH compares the new vector of care-of addresses with the old one. It chooses the lowest-level FA that appears in both vectors, and sends a *Regional Registration Request* to that FA. Any higher-level agent need not be informed of this movement since the other end of its forwarding tunnel still points to the current location of the MH. In Figure 2, when MH moves from FA_7 to FA_6, FA_3 is the target of the regional registration, and FA_1, without knowledge of this movement, still correctly re-tunnels packets to FA_3. When MH moves from FA_6 to FA_5, the registration target is FA_1. In the meantime, HA has no knowledge of these local movements and none of these registrations reaches HA, and hence registration overhead is reduced.

3.2. Registration Authentication

Except for the leaf nodes in the FA hierarchy, any FA may now be the target of a regional registration request, and they must authenticate and reply to such a request. Therefore, they have to have security association with any MH. This can be achieved by establishing a registration key between such an FA and an MH. We note that before any FA in the hierarchy can be the target of a registration, the top-level FA in the hierarchy must have participated in an *initial registration* with the HA, during which a registration key can be established [12].

Intermediate FAs along a registration path do *not* necessarily authenticate the *initial* registration request and reply, just as an FA might not in base Mobile IP [10]. On the other hand, all registrations targeting an FA (i.e., the registrations occuring after the initial registration) are authenticated by the MH and the FAs in the hierarchy. A phony registration request would be rejected by the target and the intermediate FAs would not modify their forwarding tables requested by the registration. A phony registration reply would be detected by the MH, which may take appropriate actions.

4. FA Buffering and Smooth Handoff

4.1. Reducing Packet Loss During Smooth Handoff

Mobile IP route optimization [13] extends the use of binding cache and binding update messages to provide smooth handoff (*previous foreign agent notification*, cf. Section 2). However, tunneled packets that arrive at the previous FA *before* the previous FA notification does are still lost. Such data loss may be aggravated if the MH loses contact with *any* FAs for a relatively long period of time.

Our smooth handoff scheme includes an additional *foreign agent buffering* mechanism. Besides decapsulating tunneled packets and delivering them directly to an MH, the FA (point of attachment) also buffers these packets. When it receives a previous foreign agent notification, it *re-tunnels* the buffered packets, along with any future packets tunneled to it (original smooth handoff in [13]), to the MH's new FA. Packet loss during a handoff can be completely eliminated, unless the MH takes too long to find a new FA after it loses contact with its previous FA. In that case, the buffer at the previous FA may overflow.

Clearly, whether and how much packet loss can be avoided depends on how quickly an MH finds a new FA, and how many packets are buffered at the previous FA. This in turn depends on how frequently FAs send out beacons, or agent advertisements, and how long the MH stays out of range of *any* FA. A larger buffer at an FA can tolerate less frequent beacons and longer period of loss of contact. On the other hand, more frequent beacons take up more wireless bandwidth and denser coverage requires more FAs (i.e., more equipment). Balancing these factors is important for achieving optimal smooth handoff.

4.2. Eliminating Duplicate Packets

A major side effect of handing over buffered packets is packet duplication, especially when the buffer is relatively large and the handoff process completes quickly. Although duplicate packets are allowed in the Internet, they are rare under normal circumstances (e.g., in wired and fixed hosts). Consequently, modern implementations of TCP [16] assume that duplicated TCP acknowledgments are caused by lost data packets and will invoke congestion control mechanisms [6]. Duplicates due to buffer handover may therefore cause such upper layer protocols as TCP to slow down unnecessarily and degrade performance, as well as waste network resources.

We use a simple duplicate elimination mechanism. When the MH receives an IP datagram, it buffers the pair of the source address and the *identification* (originally used for IP fragmentation) of the datagram. Such a source-id pair occupies only 6 bytes [14]. When the MH requests a smooth handoff, it includes these buffered pairs (the most recent ones, since older pairs would be deleted if the buffer overflows) in the registration request, to be forwarded in the previous FA notification. The previous FA then uses these source-id pairs to drop those buffered packets that have been received by the MH. A similar duplicate elimination method has been proposed by Balakrishnan et al. [2] (cf. Section 5); a more detailed description of their scheme can be found in [15].

5. Related Work

Due to the limited space here, we can only briefly describe related work. The reader is referred to the cited papers for more details.

Balakrishnan et al. [2] use multicast [4] and intelligent buffering in nearby FAs [7] to alleviate the frequent local handoff problem and packet loss during a handoff. In this scheme, all packets destined for the MH are tunneled by the HA to all *nearby* FAs for buffering (through multicast) so that when a handoff occurs, any lost packets can be retransmitted by the new FA. One problem with this scheme is that it uses additional network resources and memory space at the buffering FAs.

It is a well-accepted idea that a HA (or forwarding agent) should not micro-manage every local movement of the mobile host in a foreign domain. Aziz [1] proposes an efficient intra-domain tunneling scheme, based on the Columbia scheme [5], a precursor of Mobile IP. A Mobility Support Border Router (MSBR) is provided in every domain, serving as a forwarding gateway. Any local movement of the MH within the same domain is managed by MSBR and transparent to the MH's home domain.

Cáceres and Padmanabhan [3] propose a different hierarchical FA management scheme with agent buffering. An MH registers only the highest level FA with its HA. The establishment and changes of the forwarding path in the FA hierarchy are achieved using the *Address Resolution Protocol (ARP) proxy* and *gratuitous ARP* messages. A major drawback of this method is the potential security concern. The use of proxy and gratuitous ARP opens up the possibility that a malicious host connected to the foreign network may pose as a proxy and steal packets from the MH.

6. Simulation Results

6.1. Simulated Environment

We have measured TCP and UDP performances with our smooth handoff scheme under several circumstances through simulations. Our simulation is based on the Network Simulator (*ns*) developed by the Lawrence Berkeley National Laboratory [8]. We have incorporated into the simulator the basic Mobile IP mechanism and our proposed hierarchical FAs and smooth handoff scheme. For comparison, we added an ARP-based smooth handoff and FA hierarchy mechanism similar to that proposed by Cáceres and Padmanabhan [3].

The simulated network topology is shown in Figure 3. The numbers beside a link indicate the bandwidth and delay of the link. We simulated an MH talking to a distant fixed host.

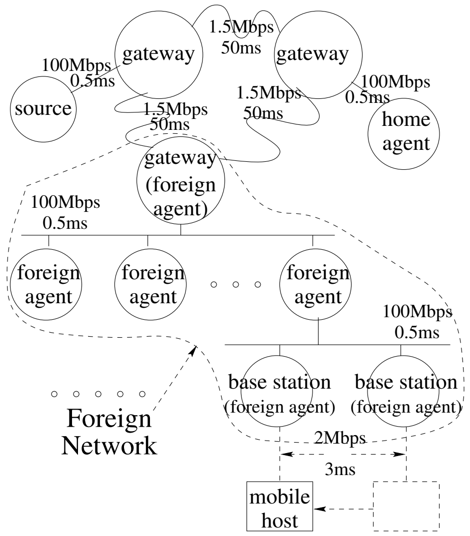

Figure 3. Simulated Network Topology

To measure the predicted performance improvements achieved by hierarchical FAs and FA buffering, we simulate a foreign network with a hierarchy of three levels. A domain FA sits at the root and is connected to four subnet FAs, through a shared medium such as Ethernet. Each of these subnet FAs is in turn connected to four lowest level FAs through a shared medium (a subnet). The FAs have wireless links to the visiting MHs.

The foreign network covers a 4×4 square room with no obstructions, as shown in Figure 4. Each square is covered by an FA; any MH in the square can communicate with that FA but none of the others. The four FAs in a quadrant are connected to a subnet FA, and the four subnet FAs are connected to the domain FA, as described above. The beacon period for an FA is 100 milliseconds, but the beacons from these FAs are *not* synchronized. Propagation delay is ignored.

An MH moves within the room according to the following pattern. It moves along a straight line for a certain period of time before it makes a turn. This moving period is a random number, normally distributed with average of 5 seconds and standard deviation of 0.1 second. When it makes a turn, the new direction (angle) in which it will move is a normally distributed random number with average equal to the previous direction and standard deviation of 30 degrees. Its speed is also a normally distributed random number, with a controlled average, ranging from 0.1 to 0.45 (unit/sec), and standard deviation of 0.01 (unit/sec). A new such random number is picked as its speed when it makes a turn. This pattern of mobility is intended to model node movement during which the nodes have momentum, and thus do not start, stop, or turn abruptly. When it hits a wall, it reflects off the wall at the same angle; in our simulated world, there is little other choice.

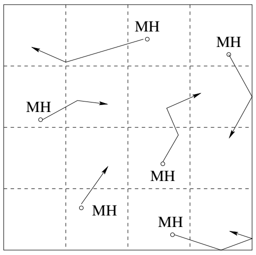

Figure 4. Simulated Foreign Network

In our scheme and the ARP-based smooth handoff, we vary the buffer size at the FAs from 0 to 18 Kbytes. In our duplicate reduction mechanism, the buffer size at an MH is $\frac{6}{100}\times$ the buffer size at the FAs. It is a nice feature that we can minimize the buffer cost at the MH, but at 6 bytes per buffer it is not a major consideration.

We incorporate into the simulation measurement of software overhead incurred by a host processing registration, encapsulation and decapsulation taken through experiments. The Mobile IP implementation used in these experiments is Mobile IP for Solaris [17], written by Vipul Gupta and Sunil Madhani. The machines used are shown in Table 1. The measured software costs of a registration, encapsulation and decapsulation are 1.8 ms, 270 μs and 160 μs, respectively.

In the simulation, when a registration request or reply is processed by a host, the registration overhead is used instead of the normal IP routing overhead. An encapsulation is more expensive than a decapsulation because an additional buffer has to be allocated for the outer IP header in the former, while the latter requires only shifting a pointer (the beginning of a packet) during a memory copy. The overhead caused by re-tunneling depends on implementation, but it would be faster to change the source and destination IP addresses of the outer header, and the cost of such an operation should be similar to that of a decapsulation instead of a combination of an encapsulation and a decapsulation.

agent	machine	operating system
HA	Sun SparcStation Ultra-1	Sun Solaris 2.5.1
FA	Sun SparcStation 20	Sun Solaris 2.6
MH	Toshiba Libretto 50CT (Pentium 75)	linux 2.0.30

Table 1. Machines used in experiments

In our TCP simulation, the source (fixed host) transfers a file of size 1 MB to the MH using a TCP connection. We measure the throughput after the file transfer is completed. In the UDP case, the source sends one 200-byte packet every 20 milliseconds. Such a traffic pattern is meant to mimic the traffic generated by an Internet audio application.

6.2. Results

We run 200 TCP file transfers of 1 MB for each set of parameters and take the average throughput and other measurements. To measure the impact of handoffs on the UDP streams described in the previous subsection, we run 200 handoffs for each set of parameters and take the average.

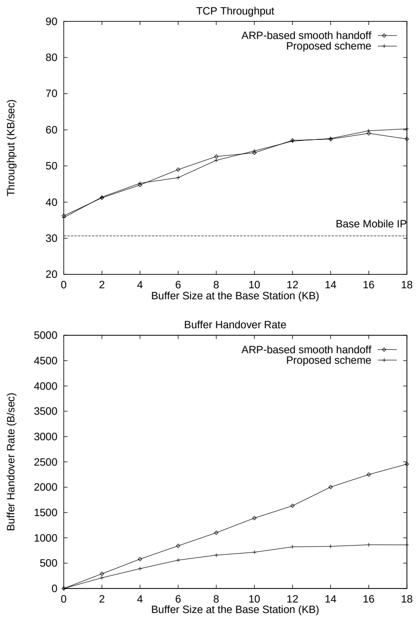

Figure 5. Throughput and buffer handover rate. Average moving speed of MHs = 0.3 (unit/sec).

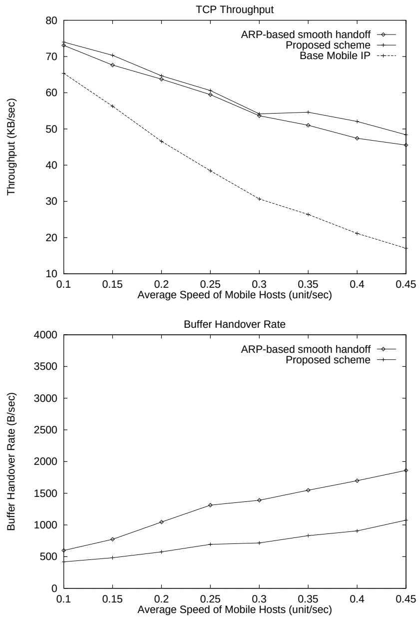

Figure 6. Throughput and buffer handover rate. Buffer size = 10 KB.

Figures 5 and 6 show the TCP throughput and buffer handover rate under the three different mobility management schemes. The buffer handover rate is defined as the average amount of buffered data all 16 FAs hand over to an MH's new FA *every second*. This measures how much network resources the smooth handoff schemes consume.

With higher costs of registration going through all FAs along a path and re-tunneling, we do not expect our scheme to outperform the ARP-based one. However, Figure 5 shows that these two schemes yield similar TCP throughput, and with the largest buffer size (18 KB), the performance of the ARP-based scheme drops slightly. One explanation is the benefit of the duplicate elimination mechanism in our design. As the buffer size grows, the simple buffer-and-handover method retransmits more and more data, while in our scheme the handover rate grows at a much slower pace. Clearly the duplicate elimination helps avoid retransmitting data already received by the MH but not yet flushed from the buffer. This also boosts TCP throughput by preventing the TCP sender from unnecessarily backing off due to duplicates, as mentioned in Section 4.2. We also note that even with no buffer space at the FAs, either of the smooth handoff schemes yields better throughput than the base Mobile IP because avoiding registration with a distant HA during a local handoff reduces overhead and the opening for lost packets.

As shown in Figure 6, faster moving MHs cause more frequent handoffs and poorer performance. As all three methods suffer, our scheme seems to maintains a small edge over the other two.

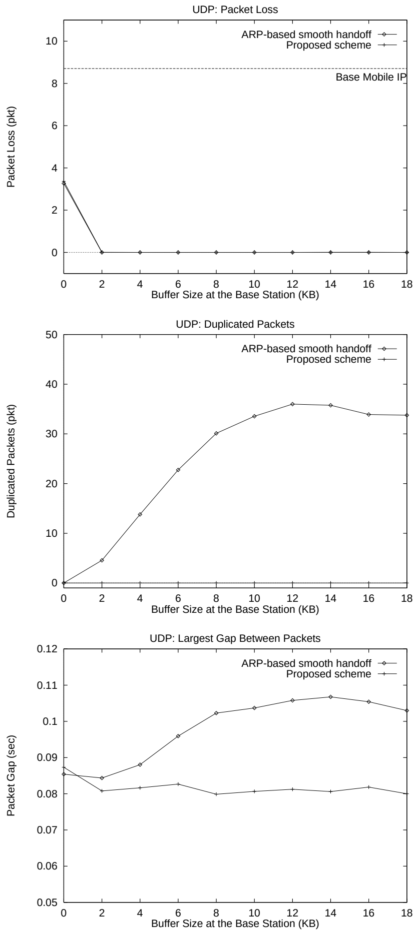

Figure 7. UDP: average loss, duplicates and gap during a handoff.

Our UDP simulation provides a closer look at the impact of duplicates. Figure 7 shows that while the two smooth handoff schemes are equally efficient in reducing packet loss, our design can eliminate duplicates (for the chosen data rate) by paying a price of higher registration overhead (Figure 8). The "gap" in Figure 7 is the longest time observed by the receiver between two consecutive, *previously unseen* packets. Reducing this gap is important for real-time applications sending steady data streams such as packet audio and video. Without duplicate elimination, the first few buffered packets handed over are likely to have already been received by the MH and therefore delay the arrival of the needed packets, as demonstrated in our results.

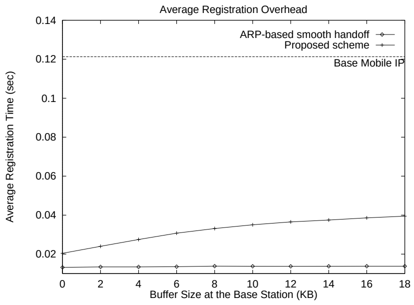

Figure 8. Registration overhead.

7. Conclusion and Future Work

In general, hierarchical FA management reduces the Mobile IP registration overhead of a local handoff and the transition period during which tunneled packets may be lost. FA buffering further alleviates loss by trying to recover those misplaced packets during a transition. Our design presented in this paper stands out in three aspects. First, the duplicate elimination improves performance by avoiding waste of network resources and misinterpretation by higher layer protocols, such as TCP. Second, we enable the local handoff and buffer handover to use the same security measures as in the base Mobile IP, unlike ARP-based schemes that inherit ARP's security problem. Third, our FA hierarchy is not tightly coupled with the physical network configuration, and therefore can be freely organized.

Our simulation results show the improvements obtained by using buffering at the FAs, and eliminating duplicates. Further implementation may better address other issues than our simulation, such as software overhead, overhead caused by re-tunneling, the performance impact of various security measures, and the appropriate buffer size. We would particularly like to measure the performance impact resulting from the extension of registration key handling to hierarchical FAs, briefly described in Section 3.2. As IETF security protocols achieve wider deployment, basic authentication mechanisms are likely to be the focus of hardware and software optimizations enabling their use for many applications besides just the registration operations described for Mobile IP and in this paper.

Acknowledgment

We would like to thank Dr. Vipul Gupta for his kind help with our base Mobile IP experiments and answers to our questions concerning the Solaris Mobile IP implementation [17].

References

[1] A. Aziz. A Scalable and Efficient Intra-Domain Tunneling Mobile-IP Scheme. *ACM Computer Communication Review*, 24(1), January 1994.

[2] H. Balakrishnan, S. Seshan, and R. H. Katz. Improving Reliable Transport and Handoff Performance in Cellular Wireless Networks. *ACM Wireless Networks*, 1(4), December 1995.

[3] R. Cáceres and V. N. Padmanabhan. Fast and Scalable Handoffs for Wireless Internetworks. In *The Second Annual International Conference on Mobile Computing and Networking*, pages 56–66, Rye, New York, November 1996. ACM and IEEE.

[4] S. E. Deering. *Multicast Routing in a Datagram Internetwork*. PhD thesis, Stanford University, December 1991.

[5] J. Ioannidis, D. Duchamp, and J. Gerald Q. Maguire. IP-based Protocols for Mobile Internetworking. In *ACM SIGCOMM '91*, pages 235–245. ACM, September 1991.

[6] V. Jacobson. Congestion Avoidance and Control. In *ACM SIGCOMM '88*. ACM, 1988.

[7] K. Keeton, B. A. Mah, S. Seshan, R. H. Katz, and D. Ferrari. Providing Connection-Oriented Service to Mobile Hosts. In *Proceedings of the First USENIX Symposium on Mobile and Location-Independent Computing*, August 1993.

[8] Network Research Group, Lawrence Berkeley National Laboratory. *ns—LBNL Network Simulator*. URL: `http://www-nrg.ee.lbl.gov/ns/`.

[9] C. E. Perkins. *IP Encapsulation within IP*. Network Working Group, Request for Comments 2003, October 1996.

[10] C. E. Perkins, editor. *IP Mobility Support*. Network Working Group, Request for Comments 2002, October 1996.

[11] C. E. Perkins. Mobile-IP Local Registration with Hierarchical Foreign Agents. Internet Engineering Task Force, Internet Draft—work in progress, February 1996.

[12] C. E. Perkins and D. B. Johnson. Registration Keys for Route Optimization. Mobile IP Working Group, Internet Draft—work in progress, November 1997.

[13] C. E. Perkins and D. B. Johnson. Route Optimization in Mobile IP. Mobile IP Working Group, Internet Draft—work in progress, November 1997.

[14] J. Postel. *Internet Protocol, DARPA Internet Program Protocol Specification*. Network Working Group, Request for Comments 791, September 1981.

[15] S. Seshan. *Low-Latency Handoff for Cellular Data Networks*. PhD thesis, University of California at Berkeley, December 1995.

[16] W. Stevens. *TCP Slow Start, Congestion Avoidance, Fast Retransmit, and Fast Recovery Algorithms*. Network Working Group, Request for Comments 2001, January 1997.

[17] Vipul Gupta and Sunil Madhani, Sun Microsystems, Inc. *Mobile IP for Solaris*. URL: `http://playground.sun.com/pub/mobile-ip/`.