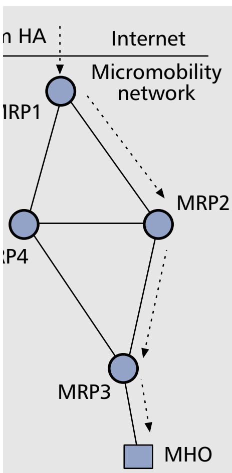
COMPARISON OF IP MICROMOBILITY PROTOCOLS

ANDREW T. CAMPBELL, JAVIER GOMEZ, SANGHYO KIM, AND CHIEH-YIH WAN,
COLUMBIA UNIVERSITY
ZOLTAN R. TURANYI AND ANDRAS G. VALKO, ERICSSON RESEARCH



Over the past several years a number of IP micro-mobility protocols have been proposed, designed and implemented that complement the base Mobile IP protocol. The development of these protocols has generated considerable interest in industry and academia.

ABSTRACT

We present a performance comparison of a number of key micromobility protocols that have been discussed in the IETF Mobile IP Working Group over the past several years. IP micromobility protocols complement Mobile IP by offering fast and seamless handoff control in limited geographical areas, and IP paging in support of scalability and power conservation. We show that despite the apparent differences between IP micromobility protocols, the operational principles that govern them are largely similar. We use this observation to establish a generic micromobility model to better understand design and performance trade offs. A number of key design choices are identified within the context of the generic model related to handoff quality and route control messaging. We present simulation results for Cellular IP, Hawaii, and Hierarchical Mobile IP, and evaluate the handoff performance of these protocols. Simulation results presented in this article are based on the Columbia IP Micromobility Software (CIMS), which is freely available from the Web (comet.columbia.edu/micromobility) for experimentation.

INTRODUCTION

Over the past several years a number of IP micromobility protocols [1] have been proposed, designed, and implemented that complement the base Mobile IP protocol [2] by providing fast, seamless, and local handoff control. The development of these protocols has generated considerable interest in industry and academia, and resulted in the ongoing standardization efforts within the Internet Engineering Task Force (IETF) Mobile IP [3] and Seamoby [4] Working Groups on low-latency handoff [5] and IP paging [6], respectively.

IP micromobility protocols are designed for environments where mobile hosts change their point of attachment to the network so frequently that the base Mobile IP mechanism introduces significant network overhead in terms of increased delay, packet loss, and signaling. For example, many real-time wireless applications (e.g., voice over IP) would experience noticeable degradation of service with frequent handoff. Establishment of new tunnels can introduce additional delays in the handoff process, causing packet loss and delayed delivery of data to applications. This delay is inherent in the round-trip incurred by Mobile IP as the registration request is sent to the home agent (HA) and the response sent back to the foreign agent (FA). Route optimization [7] can improve service quality but cannot eliminate poor performance when a host moves while communicating with a distant correspondent host. Micromobility protocols aim to handle local movement (e.g., within a domain) of mobile hosts without interaction with the Mobile IP enabled Internet. This has the benefit of reducing delay and packet loss during handoff and eliminating registration between mobile hosts and possibly distant HAs when mobile hosts remain inside their local coverage areas. Eliminating registration in this manner reduces the signaling load experienced by the network in support of mobility. To minimize poor performance during handoff, micromobility protocols support *fast*, *seamless*, *local mobility*.

As the numbers of wireless users grow so will the signaling overhead associated with mobility management. In cellular networks registration and paging techniques are used to minimize the signaling overhead and optimize mobility management performance. Currently, Mobile IP supports registration but not paging. An important characteristic of micromobility protocols is their ability to reduce the signaling overhead related to frequent mobile migrations taking into account a mobile host's operational mode (i.e., active or idle). When wireless access to the Internet becomes the norm, IP mobility solutions will have to provide efficient and scalable location tracking in support of idle users, and paging in support of active communications. Support for "passive connectivity" to the wireless Internet balances a number of important design considerations. For example, only keeping the approximate location information of idle users requires significantly less signaling and thus reduces the load over the air interface and in the network. Reducing signaling over the air interfaces in this manner also has the benefit of preserving the power reserves of mobile hosts. To minimize signaling overhead and optimize mobility management performance, micromobility protocols support *IP paging*.

There is a growing need to better understand the differences between many of the micromobility proposals found in the literature [1] in terms of their design and performance. In this

1070-9916/02/$17.00 © 2002 IEEE

article we present a performance comparison of Cellular IP (CIP) [8], Hawaii [9], and Hierarchical Mobile IP (HMIP) [10] based on the Columbia IP Micromobility Software (CIMS) [11] ns-2 extension. Our analysis focuses on the user's perceived performance during handoff. We leave protocol complexity, processing requirements, and paging issues for future work. In this article we show that despite the apparent differences between these three protocols, the operational principles that govern them are largely similar. We use this observation to establish a *generic micromobility model* to better understand the similarities and differences between these protocols. A number of key design choices are identified within the context of the proposed generic model for route control messaging and handoff performance. We show that the difference in handoff quality observed during simulation is related to the design of these protocols. Finally, we discuss a number of other implementation issues that may influence the future deployment of micromobility protocols.

This article is structured as follows. We provide an overview of the protocols under study, and present a simple taxonomy of the mobility management approaches used by micromobility protocols. A generic model for micromobility is then presented and discussed. Following this, we present our simulation model and performance results for handoff and route control messaging. We discuss a number of practical aspects that influence the deployment of micromobility protocols. Finally, we discuss a number of open issues and present some concluding remarks.

Micromobility Architectures

The primary role of micromobility protocols is to ensure that packets arriving from the Internet and addressed to mobile hosts are forwarded to the appropriate wireless access point in an efficient manner. To do this, micromobility protocols maintain a *location database* that maps mobile host identifiers to location information. In what follows we provide an overview of Cellular IP, Hawaii, and Hierarchical Mobile IP, and then present a simple taxonomy to distinguish the different mobility management approaches used to design IP micromobility protocols. This leads us to establish a generic model for micromobility.

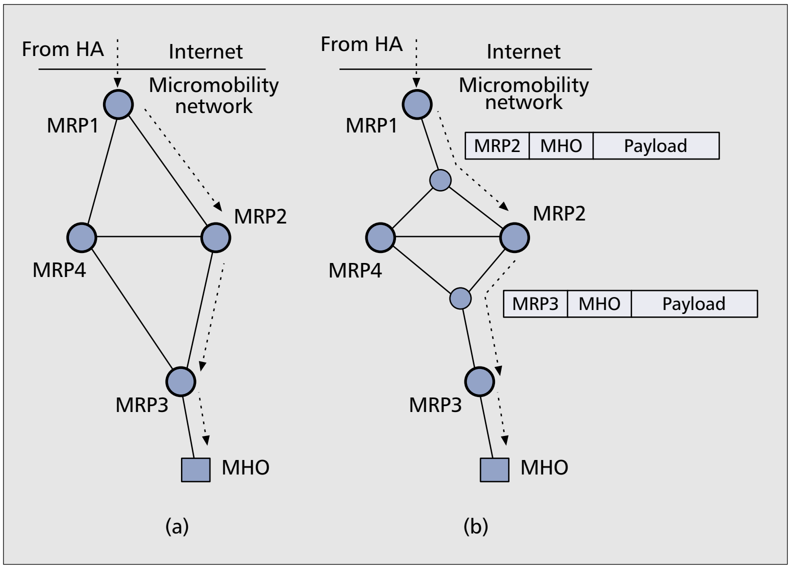
Protocols

Cellular IP — The Cellular IP protocol [8] from Columbia University and Ericsson Research supports paging and a number of handoff techniques. Location management and handoff support are integrated with routing in Cellular IP access networks. To minimize control messaging, regular data packets transmitted by mobile hosts are used to refresh host location information. Cellular IP uses mobile-originated data packets to maintain reverse path routes. Nodes in a Cellular IP access network monitor (i.e., "snoop") mobile originated packets and maintain a distributed, hop-by-hop location data base that is used to route packets to mobile hosts. Cellular IP uses IP addresses to identify mobile hosts. The loss of downlink packets when a mobile host moves between access points is reduced by a set of customized handoff procedures. Cellular IP supports two types of handoff scheme. Cellular IP hard handoff is based on a simple approach that trades off some packet loss in exchange for minimizing handoff signaling rather than trying to guarantee zero packet loss. Cellular IP semisoft handoff prepares handoff by proactively notifying the new access point before actual handoff. Semisoft handoff minimizes packet loss, providing improved TCP and UDP performance over hard handoff. Cellular IP also supports IP paging, and is capable of distinguishing active and idle mobile hosts. Paging systems help minimize signaling in support of better scalability and reduce the power consumption of mobile hosts. Cellular IP tracks the location of idle hosts in an approximate and efficient manner. Therefore, mobile hosts do not have to update their location after each handoff. This extends battery life and reduces air interface traffic. When packets need to be sent to an idle mobile host, the host is paged using a limited scope broadcast and in-band signaling. A mobile host becomes active upon reception of a paging packet and starts updating its location until it moves to an idle state again. Cellular IP also supports a fast security model that is suitable for micromobility environments based on fast session key management. Rather than defining new signaling, Cellular IP access networks use special session keys where base stations independently calculate keys. This eliminates the need for signaling in support of session key management, which would inevitably add additional delay to the handoff process.

Hawaii — The Hawaii protocol [9] from Lucent Technologies proposes a separate routing protocol to handle intradomain mobility. Hawaii relies on Mobile IP to provide wide-area interdomain mobility. A mobile host entering a new FA domain is assigned a collocated care-of address. The mobile node retains its care-of address unchanged while moving within the foreign domain; thus, the HA does not need to be involved unless the mobile node moves to a new domain. Nodes in a Hawaii network execute a generic IP routing protocol and maintain mobility-specific routing information as per host routes added to legacy routing tables. In this sense Hawaii nodes can be considered enhanced IP routers, where the existing packet forwarding function is reused. Location information (i.e., mobile-specific routing entries) is created, updated, and modified by explicit signaling messages sent by mobile hosts. Hawaii defines four alternative path setup schemes that control handoff between access points. The appropriate path setup scheme is selected depending on the operator's priorities between eliminating packet loss, minimizing handoff latency, and maintaining packet ordering. Hawaii also uses IP multicasting to page idle mobile hosts when incoming data packets arrive at an access network and no recent routing information is available.

Hierarchical Mobile IP — The Hierarchical Mobile IP protocol [10] from Ericsson and Nokia employs a hierarchy of FAs to locally handle Mobile IP registration. In this protocol mobile

The primary role of micromobility protocols is to ensure that packets arriving from the Internet and addressed to mobile hosts are forwarded to the appropriate wireless access point in an efficient manner. To do this, micromobility protocols maintain a "location database" that maps mobile host identifiers to location information.



■ **Figure 1.** *a) A network of mobile routing points; b) a full network with intermediate nodes.*

hosts send Mobile IP registration messages (with appropriate extensions) to update their respective location information. Registration messages establish tunnels between neighboring FAs along the path from the mobile host to a gateway FA (GFA). Packets addressed to the mobile host travel in this network of tunnels, which can be viewed as a separate routing network overlay on top of IP. The use of tunnels makes it possible to employ the protocol in an IP network that carries non-mobile traffic as well. Typically one level of hierarchy is considered where all FAs are connected to the GFA. In this case, direct tunnels connect the GFA to FAs that are located at access points. Paging extensions for Hierarchical Mobile IP are presented in [12] allowing idle mobile nodes to operate in a power saving mode while located within a paging area. The location of mobile hosts is known by HAs and is represented by paging areas. After receiving a packet addressed to a mobile host located in a foreign network, the HA tunnels the packet to the paging FA, which then pages the mobile host to reestablish a path toward the current point of attachment. The paging system uses specific communication time slots in a paging area. This is similar to the paging channel found concept found in second-generation cellular systems.

Mobility Management Taxonomy

Hierarchical Tunneling — In hierarchical tunneling approaches the location database is maintained in a distributed form by a set of FAs in the access network. Each fa reads the incoming packet's original destination address and searches its visitor list for a corresponding entry. If the entry exists, it contains the address of the next lower-level FA. The sequence of visitor list entries corresponding to a particular mobile host constitutes the mobile host's location information and determines the route taken by downlink packets. Entries are created and maintained by registration messages transmitted by mobile hosts. Hierarchical tunneling schemes rely on a tree-like structure of FAs. Encapsulated traffic from an HA is delivered to the root FA. Each FA on the tree decapsulates and then reencapsulates data packets as they are forwarded down the tree of FAs toward the mobile host's point of attachment. As a mobile host moves between different access points, location updates are made at the optimal point on the tree, tunneling traffic to the new access point. Micromobility protocols based on hierarchical tunneling techniques sometimes require that mobile hosts either send new types of control messages or need to be aware that a hierarchical tunneling protocol is in use. Examples of micromobility protocols that use hierarchical tunneling include regional tunneling management [10] used by a number of hierarchical Mobile IP proposals.

Mobile-Specific Routing — Mobile-specific routing approaches avoid the overhead introduced by decapsulation and reencapsulation schemes, as is the case with hierarchical tunneling approaches. These proposals use routing to forward packets toward a mobile host's point of attachment using mobile specific routes. These schemes introduce implicit (e.g., based on snooping data) or explicit signaling to update mobile-specific routes or they are aware that a routing protocol is in use. In the case of Cellular IP, mobile hosts attached to an access network use the IP address of the gateway as their Mobile IP care-of address. The gateway decapsulates packets and forwards them toward a base station. Inside the access network, mobile hosts are identified by their home address and data packets are routed using mobile-specific routing without tunneling or address conversion. The routing protocol ensures that packets are delivered to a mobile host's actual location. Examples of micromobility protocols that use mobile-specific routing include Cellular IP and Hawaii.

Although this article focuses on IP layer mobility management protocols, for completeness we mention simple layer 2 mobility solutions. The "iwander" wireless micromobility network discussed in [13], for example, consists of interconnected off-the-shelf Ethernet switches. As an autoconfiguration function, Ethernet switches maintain bindings between host MAC addresses and the ports through which packets are received from a particular host [14]. Packets addressed to a host are forwarded through the port indicated by the binding. In [13] this function is used to provide layer 2 mobility management. When a mobile host moves, its "uplink" data packets transmitted from the new location implicitly update bindings and ensure proper delivery of data packets associated with the mobile host on the "downlink." This simple micromobility network identifies mobile hosts using MAC addresses and stores location information as a sequence of switch port bindings instead of a sequence of next FA addresses, as is the case in Hierarchical Mobile IP. The network of Ethernet switches relies on mobile-host-originated data packets rather than explicit signaling messages to update location information. Another commonly used layer 2 mobility management protocol is the Inter-Access Point Protocol (IAPP), which is currently standardized as IEEE 802.11F [15].

	Ethernet switch	Cellular IP	Hawaii	Hierarchical MIP
MRP layer	L2	L3	L3	"L3.5"
MRP	All switches	All CIP nodes	All routers	FAs
MH identifier	MAC address	Home address	C/o address	Home address
Intermediate nodes	-	L2 switches	L2 switches	L3 routers
Next MRP field	Port (L1)	MAC addr (L2)	MAC addr (L2)	IP addr (L3)
Means of update	Data packet	Data packet	Signaling message	Signaling message

■ **Table 1.** *Mapping of micromobility protocols to the generic micromobility model based on MRPs.*

Micromobility protocols based on hierarchical tunneling techniques sometimes require that mobile hosts either send new types of control messages or need to be aware that a hierarchical tunneling protocol is in use.

A Generic Model

There are a number of similarities in the operation of micromobility protocols despite the differences in the type of host identifier used, structure of location database, and means of updating the database. For example, both iwander [13] and Hierarchical Mobile IP require that some network nodes maintain a list of host entries and search this list for each downlink packet. List entries in both protocols are assigned timers and removed after a prespecified time unless the list entry is refreshed. Each entry contains a pointer to the next node toward the mobile host's actual point of attachment. The series of next-hop entries constitutes the downlink route. To forward a downlink packet, nodes must read the original destination address, find the corresponding entry in the list, and forward the packet to the next node.

We will use the term *mobile routing point* (MRP) to refer to nodes that participate in the procedure described above (i.e., Ethernet switches in [13] or FAs in Hierarchical Mobile IP). We argue that despite differences in design approach, most micromobility protocols can be regarded as networks of MRPs that implement the procedure described above. Furthermore, most micromobility protocols can be mapped to a generic micromobility model consisting of MRPs. Each MRP contains a list of hosts whose data path traverses the MRP. When a downlink packet arrives at an MRP, the list is searched for the destination address and the packet forwarded to the next MRP, as indicated by the list entry. The series of entries associated with a particular mobile host constitutes the host's location (i.e., routing) information.

In Fig. 1a, data packets addressed to a mobile host MH0 are forwarded through three MRPs. The HA associated with mobile host MH0 (not shown in the figure) tunnels packets to MRP1. Upon reception of a data packet, MRP1 decapsulates the packet, checks its list of host entries, and determines that the packet should be forwarded to MRP2. MRP2, in turn, forwards the packet to MRP3, which finally forwards the packet over the air to the mobile host.

Neighboring MRPs may be physically interconnected, but do not need to be. In a Cellular IP network, for example, MRPs correspond to Cellular IP nodes. Neighboring Cellular IP nodes can be separated by any number of layer 2 switches. In Fig. 1b, the network illustrated in Fig. 1a is shown with a number of physical nodes that are not MRPs. These nodes are represented by smaller circles in Fig. 1b. The figure also shows the destination address fields contained in a downlink packet as it traverses a number of MRPs en route toward the destination mobile host. Each packet carries the mobile host's identifier because all MRPs along the path use this identifier to find the next MRP. In addition, each packet must carry the address of the next MRP because nodes (i.e., non-MRP nodes) that lie between two MRPs, by definition, cannot route packets based on the mobile host identifier. As a packet traverses the access network from MRP to MRP, its internal address (i.e., the mobile host's identifier) remains unchanged while its external address (i.e., the next MRP's address) is replaced by each MRP the packet encounters en route toward the destination mobile host. Carrying both addresses in the packet is achieved by using protocol encapsulation. In Cellular IP and Hawaii, IP packets are encapsulated in L2 frames, while Hierarchical Mobile IP uses "IP in IP" encapsulation. The iwander Ethernet switch network is a special case because it does not support any intermediate nodes and therefore does not need to carry MRP addresses due to the lack of encapsulation below layer 2.

Most micromobility protocols can be regarded as implementations of the MRP model described above. In Table 1, we map four micromobility protocols to the generic model. One can observe that the most important difference between these protocols is the choice of protocol layer at which per-host location information is stored and maintained. The type of the mobile host (MH) identifier, the protocol layer associated with intermediate nodes, and the type of "next MRP" information all depend on the choice of the MRP protocol layer. If, for example, MRPs are implemented at layer 3, mobile hosts are most logically identified by IP addresses. Therefore, all L3 devices in the micromobility network must be mobility-aware. This, in turn, implies that intermediate (non-MRP) nodes must operate at layer 2. In contrast, if MRPs are tunneling endpoints, denoted layer 3.5 in the table, intermediate devices can be layer 3 routers as well. In the case of an Ethernet switch network, MRPs operate at L2 and hence are directly interconnected. This prohibits any other intermediate nodes. Following on from this, the properties shown in rows 1–5 of Table 1 are limited and tightly coupled to the operations of the MRP protocol layer. In contrast, the means of updating MRP state, as shown in the last row of Table 1, is largely independent of the choice of

	Layer 2	Layer 3	Layer 3.5
Data packet	Ethernet switch	Cellular IP	
Signaling message		Hawaii	HMIP

Table 2. *Micromobility protocols grouped by the MRP protocol layer.*

MRP protocol layer. To illustrate the importance of these two independent design decisions we group the micromobility protocols according to these properties, as shown in Table 2.

While Tables 1 and 2 are focused on the differences between micromobility protocols, they also highlight a fundamental similarity (i.e., the MRP principle). Regardless of the layer that realizes MRPs, the type of host lists and the task performed by MRPs are basically identical in each case.

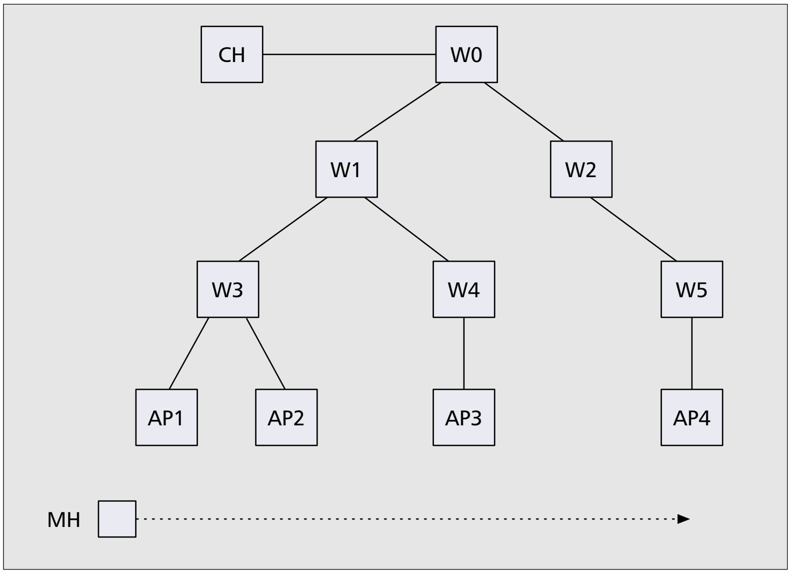


Figure 2. *The simulated network topology.*

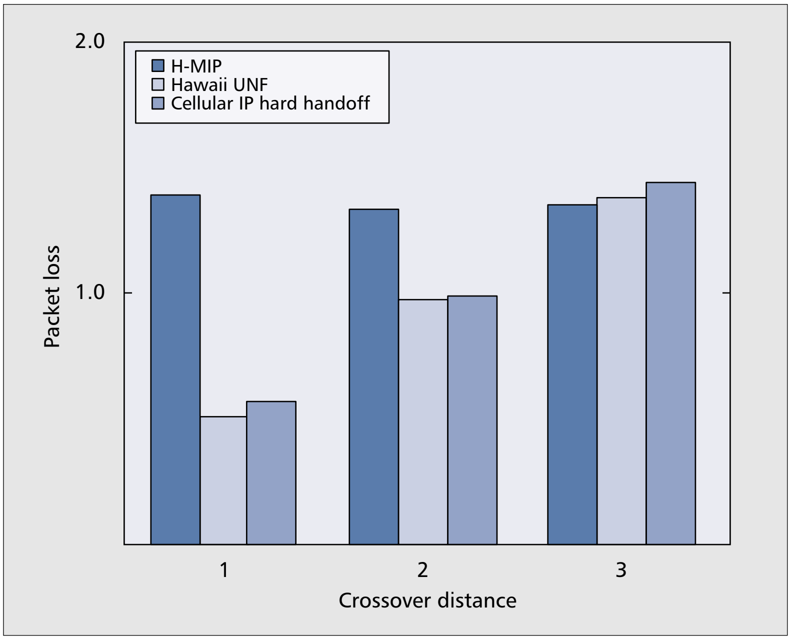


Figure 3. *UDP packet loss at handoff.*

Performance Analysis

In what follows, we present our simulation model and examine the performance of Cellular IP, Hawaii, and Hierarchical Mobile IP with respect to handoff quality, routing control messaging, and enhanced handoff control.

Simulation Model

The simulation study presented in this article uses the Columbia IP Micromobility Software (CIMS) [11], which represents a micromobility extension for the *ns-2* network simulator based on version 2.1b6 [16]. CIMS supports separate models for Cellular IP, Hawaii, and Hierarchical Mobile IP. In what follows, we briefly describe these simulation models. For a detailed description the reader is referred to the CIMS online source code and documentation [11].

The Cellular IP simulation model is based on the latest description of the protocol [17]. We implemented both hard and semisoft handoff algorithms. Paging and security functions are not used in the simulations but are available in CIMS. The Hawaii simulation model is based on the description provided in [18]. We used the unicast nonforwarding (UNF) and multiple stream forwarding (MSF) handoff schemes. Because Hawaii access points need to implement Mobile IP FA functionality without decapsulation capability and are responsible for generating Hawaii update messages, we modified the `BaseStationNode` object to include these features. In addition, we extended the mobile host object to include the PFANE [7] functions required by Hawaii. Hawaii routers are implemented in special `HawaiiAgent` objects that can process Hawaii messages and perform protocol-specific operations. The Hierarchical Mobile IP simulation model implements the two-level version of the protocol where there is a single GFA and FAs in each access point. To simulate this protocol we added a `GFAAgent` object to the existing simulation model. This object is responsible for setting up tunnels to FAs and encapsulating downlink packets based on the appropriate visitor list entry.

All simulations are performed using the network topology shown in Fig. 2. In Cellular IP simulations each `Wi` and `APi` corresponds to Cellular IP nodes where `W0` acts as a gateway to the Internet. In Hawaii simulations all `Wi`s and `APi`s are Hawaii-enabled routers, and `W0` is the domain root router. When simulating Hierarchical Mobile IP, the GFA function is implemented by `W0`, while `W1–W5` represent mobility-unaware routers with collocated FAs `APi`. We assume that this network is the mobile host's (`MH`) home network and hence packets arrive from a corresponding host (`CH`) without encapsulation. In this network each wired connection is modeled as a 10 Mb/s duplex link with 2 ms delay. Mobile hosts connect to access points (APs) using the ns-2 carrier sense multiple access with collision avoidance (CSMA/CA) wireless link model where each AP operates on a different frequency band. Simulation results were obtained using a single mobile host, continuously moving between APs at a speed that could be varied during simulation. Such a movement pattern ensures that mobile hosts always go through the maximum overlapping region between two radio cells. Nodes are modeled without constraints on switching capacity or message processing speed.

The simulation network accommodates using both UDP and TCP traffic. UDP probing traffic is directed from CH to MH and consists of 210 byte packets transmitted at 10 ms intervals. TCP sessions represent greedy downloads from the corresponding host to the mobile host using Reno congestion control, except where stated otherwise.

HANDOFF QUALITY

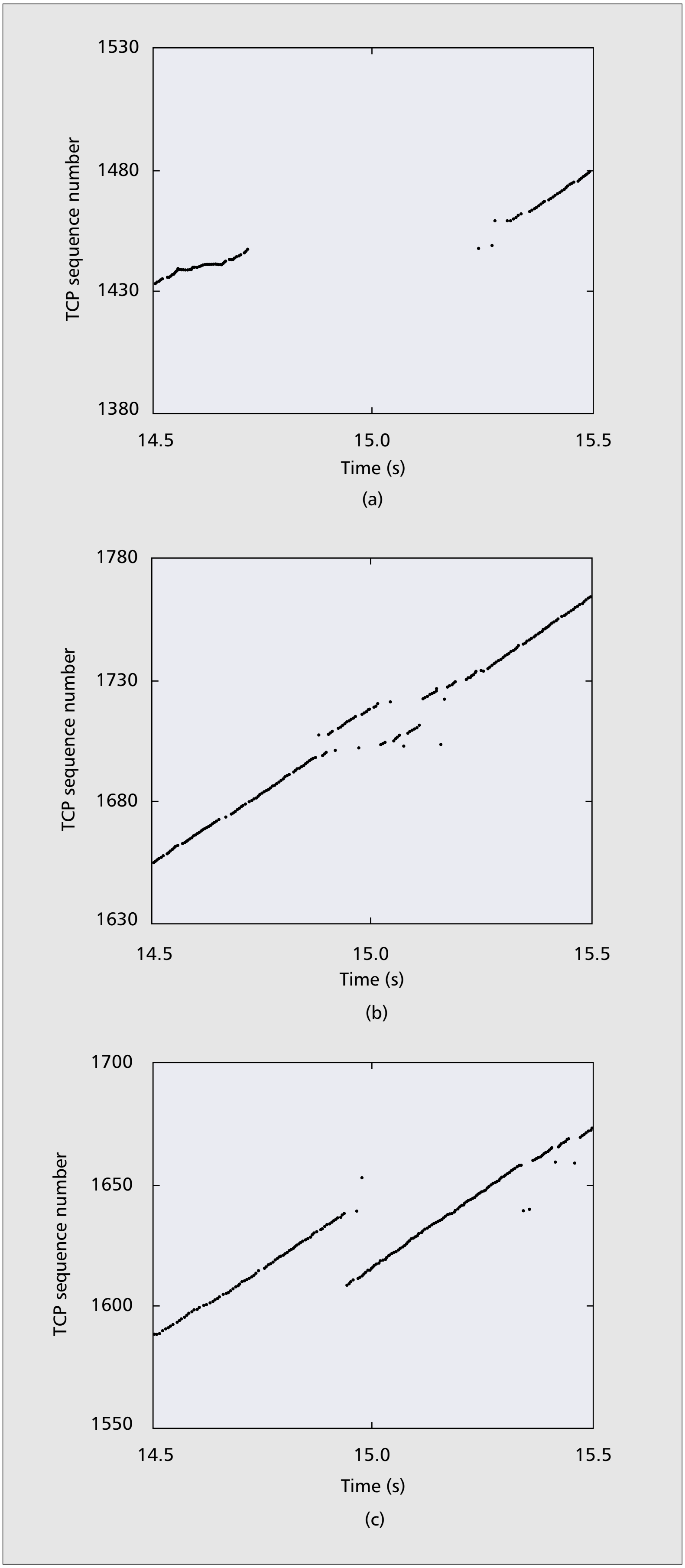
We first present simulation results for the basic (hard) handoff performance of each micromobility protocol. During simulation, a mobile host moves periodically between neighboring access points at a speed of 20 m/s. The circular areas covered by neighboring access points have an overlap region of 30 m. We use UDP probing traffic between the corresponding host and mobile hosts, and count the average number of packets lost during handoff for each protocol. Using this approach we measure handoff delay (i.e., the time it took for routing to converge). We performed simulations for three different scenarios with various crossover distances (i.e., the number of hops between the crossover node and AP). The crossover distance is 1, 2, or 3 hops when the mobile host moves between AP1-AP2, AP2-AP3, and AP3-AP4, respectively. Figure 3 shows the average number of packets lost for each of the three cases. Each data point corresponds to the average of more than 100 independent handoff events.

Our first observation is that results for Cellular IP hard handoff and Hawaii UNF are very similar. In both cases handoff delay is related to the packet delay between the APs and the cross-over node. When the mobile host moves between AP1 and AP2 the delay is small. If the crossover distance is larger, the handoff delay increases with an extra packet delay of 2 ms for each additional hop. The results are a direct consequence of the similarity between these two protocols, particularly in the way in which the protocols build up the route between a crossover node and new AP.

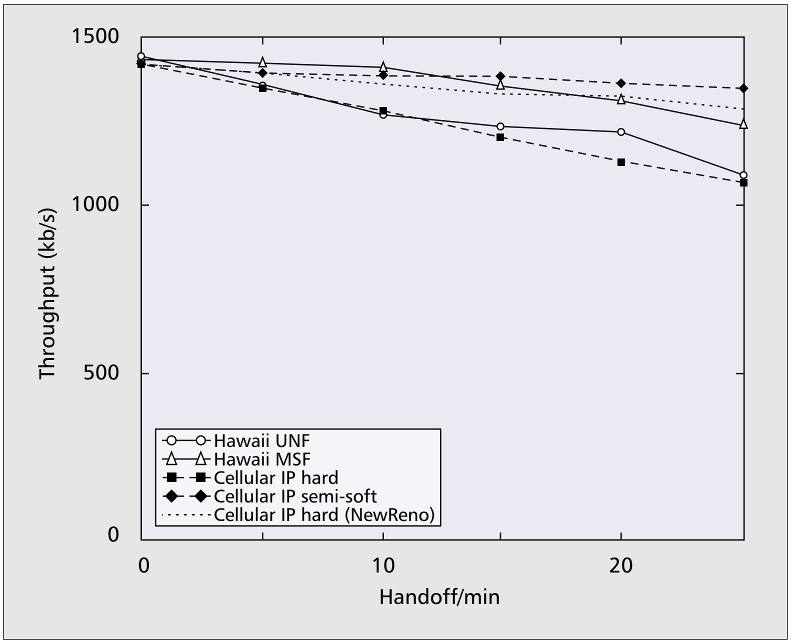
In contrast to Cellular IP and Hawaii, Hierarchical Mobile IP updates routing only when registration messages reach the GFA. Therefore, Hierarchical Mobile IP cannot benefit from the fact that a crossover node is topologically close to the APs. This phenomenon is illustrated in the results where the handoff delay for Hierarchical Mobile IP is shown to be independent of the crossover distance, and is equal to the handoff performance in the case of the maximum crossover distance for Cellular IP and Hawaii.

Next, we show simulation results for a TCP download during handoff. The dots shown in Fig. 4a correspond to sequence number of data packets associated with a single TCP connection, as seen by the mobile host. At 14.75 s into the simulation a Cellular IP hard handoff occurs. The figure shows that the packet loss caused by the handoff results in a TCP timeout. No data is transmitted during the timeout period, and the performance of the TCP connection is seriously degraded.

The degradation caused by packet loss increases with the increasing handoff frequency. This phenomenon is illustrated in Fig. 5 where



■ Figure 4. *TCP sequence numbers at the time of a Cellular IP: a) hard handoff; b) semi-soft handoff (T_{ss} = 50 ms); and c) semi-soft handoff (T_{ss} = 300 ms).*



■ Figure 5. *Application-level TCP throughput in periodic handoffs.*

we plot the long-term throughput of bulk TCP connections while the mobile host periodically performs handoff between AP3 and AP4. The squares with the dashed line shown in the plot correspond to Cellular IP hard handoff and indicate that, in comparison to the stationary case, application level throughput decreases by 25 percent when the mobile host moves between APs every 2 s. This degradation would be more severe if we considered the potential processing delays that would be anticipated in a real system.

Route Control Messaging

In the previous section we compared the handoff performance of the three protocols. Results for Cellular IP and Hawaii are similar given that the protocols operate in the same manner for tree topologies. After the mobile host moves to a new access point, it generates a control message that propagates toward the crossover node and creates downlink routing information along the new path. The operation is also similar in Hierarchical Mobile IP, but the crossover node is always at the GFA (node W0 in the simulation network shown in Fig. 2), which accounts for the additional delay.

The operation of Cellular IP and Hawaii is different when the network topology is not a tree, however. In Hawaii path setup messages are directed toward the old access point, while Cellular IP route update packets are sent toward the gateway. For non-tree topologies this difference will often result in different nodes being used as the crossover point. In Hawaii the crossover node lies at the intersection of the old downlink path and the shortest path between the old and new access points. As a result, the new downlink path will not necessarily be the shortest path between the domain root router (i.e., gateway) and the new access point. We illustrate this problem using the simulation network shown in Fig. 6. If a mobile host, initially attached to the network at AP1, moves between access points AP2 and AP8, the resulting downlink path between the domain root router W0 and the new access point AP8 will be suboptimal, as illustrated in the figure. In the case of Cellular IP, the crossover node is always at the intersection of the old downlink path and the shortest path between the gateway and the new access point. This guarantees optimal downlink paths.

This suboptimal routing problem represents a generic trade-off associated with handoff control signaling in micromobility protocols. If handoff control messages reach the gateway, MRPs higher up in the hierarchy will have to deal with a potentially large number of messages causing performance bottlenecks. Keeping routing update messaging close to access points seems reasonable because in most cases the old and new downlink paths overlap, and routing entries do not have to be updated along the common section of the paths. By discarding update messages at the crossover MRP, MRPs higher up the hierarchy do not have to process these messages, hence minimizing the signaling load at those nodes.

In order for a crossover MRP to be capable of discarding route update messages, the node must be aware that it is a crossover MRP with respect to the particular handoff in progress. In Hawaii, for example, a node that receives an update message referring to a mobile host that already has a valid entry assumes it is the crossover MRP. This relies on the protocol's property that at any given time a mobile host has only a single chain of route entries from the gateway to the current access point. In Hawaii, this is ensured by carefully removing old entries after handoff. Guaranteeing that all old entries are successfully removed in the network is problematic, however. For example, lost update messages or radio blackout periods during handoff may jeopardize such consistency. This imposes additional requirements on protocols, such as persistent retransmissions or message numbering to resolve any race conditions. Consistency problems can be avoided if crossover MRPs are explicitly determined. For example, one could design a protocol where mobile hosts are aware of their downlink route, and after handoff they include this information in the update message. This would allow a topology-aware new AP to explicitly determine the crossover MRP.

Protocols that do not identify the crossover MRP by either of the previously techniques have no ability to safely discard update messages before the gateway. Cellular IP is one such protocol that cannot support such behavior. Based on this discussion, we observe that micromobility protocols have the following design options with regard to route control:

- Send all handoff update messages to the gateway.
- Ensure that old entries are always removed in the network and let MRPs identify themselves as crossover nodes based on this property.
- Explicitly determine the crossover MRP at handoff.

Improved Handoff Schemes

In the previous sections we focused on the basic hard handoff schemes provided by each protocol. We found that differences in performance can mostly be attributed to a couple of design decisions. The first is that the base Hierarchical Mobile IP protocol employs a single-level MRP hierarchy. This design decision is motivated by the desire to reduce the number of mobility-aware nodes in the network. However, it results in slightly higher protocol delay in the case of handoffs between topologically close access points. The other design decision relates to routing updates and identifying crossover MRPs.

Several enhancements to the basic hard handoff schemes have been developed for each micromobility protocol under study in order to reduce or eliminate the packet loss during handoff. Cellular IP *semi-soft handoff* [19] allows a mobile host to set up routing to the new access point prior to handoff. In this case, packets are duplicated at the crossover node and delivered to both the new and old access points for a short period of time. By the time the mobile host attaches to the new access point, its downlink packets are already flowing along the new path. In this case, no time is lost in updating routes in the access network. However, if the path between the crossover node and the new access point is shorter than path between the crossover node and the old access point, packets may still be lost. To overcome this problem, Cellular IP crossover nodes delay packet duplicates for a fixed period amount of time (T_{ss}) before forwarding them toward the new access point. This compensates for a shorter new path. While this solution may completely eliminate loss, it may cause packet duplicates to be delivered to mobile hosts.

Another loss reduction technique is supported by the MSF path setup scheme in Hawaii [9]. Instead of setting up routing in advance of handoff, as is case with Cellular IP semi-soft handoff, MSF operates after handoff. Packets that arrive at the old access point after a mobile host has lost its air channel to the old access point are buffered and forwarded to the mobile host at its new point of attachment using the access network. Routing state is also updated at the same time so new downlink packets are directly forwarded to the new access point. Packets that are buffered and forwarded from the old access point may arrive at the new access point interleaved with new packets. This results in misordered packets being delivered to mobile hosts. The MSF scheme works best if the link layer at the old access point can determine which packets were not received by the mobile host. In such a case, MSF can efficiently forward packets using IP. If this cannot be accommodated, the IP layer at the access points must store all packets received for a certain period (T_{msf}) and forward them to the new access point. This may result in the delivery of duplicate packets at the mobile hosts, as is the case with Cellular IP semi-soft handoff.

Cellular IP and Hawaii use two different approaches to improve handoff performance:
- Bi-casting techniques
- Buffering and forwarding techniques

The former prevent packet loss by taking proactive steps that require knowledge of the new access point prior to handoff. The latter do not rely on any such knowledge, but attempt to recover packets from the old access point after handoff. The proposed seamless handoff extensions for Hierarchical Mobile IP operate along similar lines, advocating bi-casting [20, 21] and buffering and packet forwarding [22, 23] generalized.

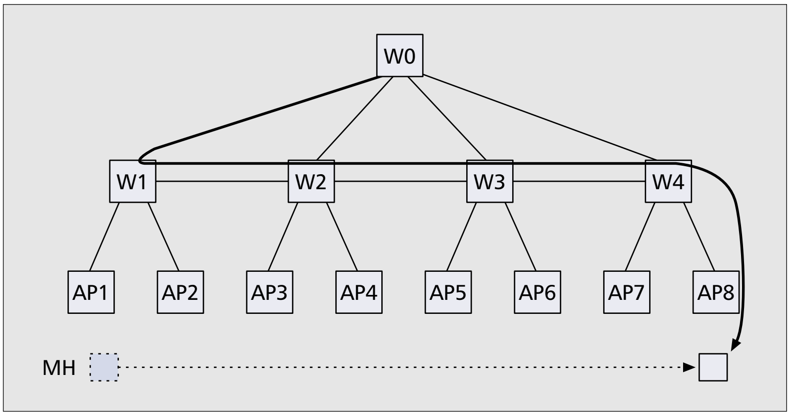


Figure 6. *Suboptimal routes after Hawaii handoffs.*

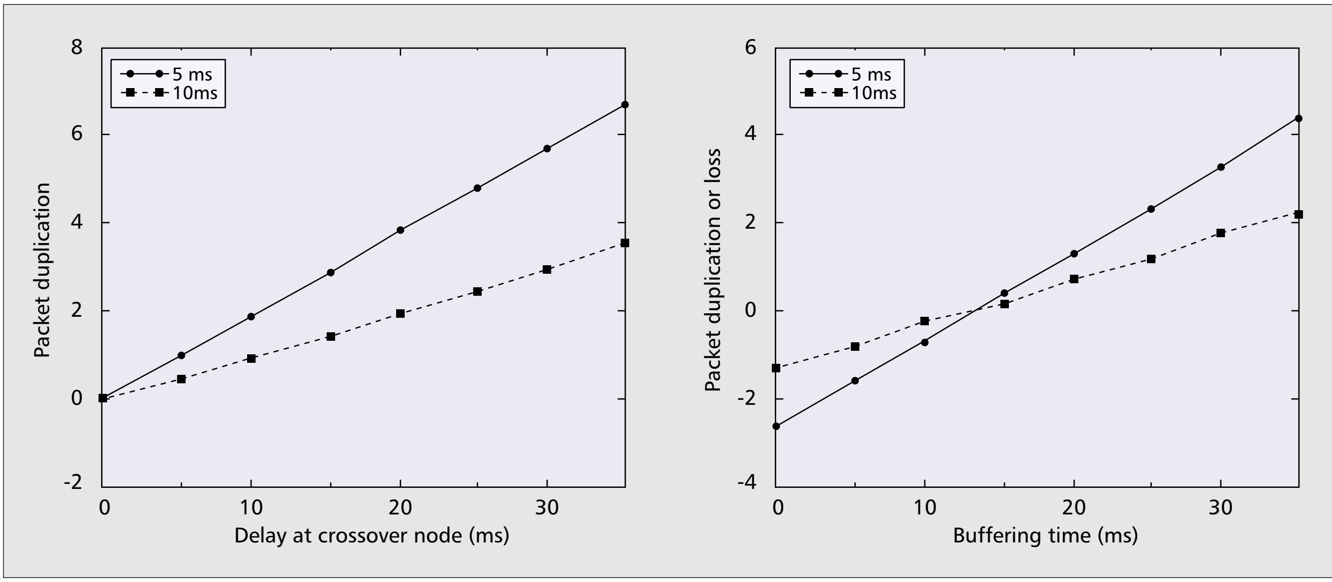
Figure 7 shows the effect of these handoff improvements for Cellular IP and Hawaii. We have plotted the average packet loss (negative values) or duplication (positive values) as functions of Cellular IP crossover delay and Hawaii buffering delay, respectively. In this case, UDP probing traffic is sent from the corresponding host to the mobile host while the mobile host performs Cellular IP semi-soft handoff or Hawaii MSF handoff. The solid and dashed lines correspond to probe traffic with packet interarrival times of 5 and 10 ms, respectively. In both cases the crossover distance is three hops.

We can observe that in the case of Hawaii MSF handoff (Fig. 7b) the lack of buffering (T_{msf} = 0) causes approximately 12 ms worth of data to be lost. This is similar to the performance of Hawaii UNF handoff. Increasing the buffering time results in increasing the number of packets being buffered and forwarded (i.e., recovered) until loss is eliminated at T_{msf} = 14. If we keep increasing T_{msf}, some packets successfully transmitted to the mobile host will also be forwarded from the old access point. This results in packet duplication. The figure shows that the actual number of lost and duplicated packets is dependent on the arrival process. However, the optimal T_{msf} value is independent of traffic. The buffering time that leads to zero packet loss and no duplication is topology-dependent and is equal to the layer 2 handoff time, plus the time it takes for the path setup message to reach the old access point.

The same observations can be made in case of the Cellular IP semi-soft handoff, as shown in Fig. 7a. One difference is that Cellular IP uses bi-casting instead of forwarding to recover packets; hence, semi-soft handoff results in zero packet loss in symmetrical topologies.

This explains the fact that the Cellular IP semi-soft loss/duplicate curve never incurs negative values (i.e., we did not observe packet loss). If the transmission time between the crossover node and the new and old access points differ, the two curves would be shifted up or down showing loss or more duplication, respectively, depending on which path is longer.

We can observe this phenomenon in the following simulation results, as shown in Fig. 4b



■ Figure 7. *UDP packet loss and duplication in a) Cellular IP semi-soft handoff; b) Hawaii MSF handoff. This is for packet interarrival times of 5 ms and 10 ms.*

and 4c. In this experiment we use TCP traffic to test the impact of semi-soft handoffs. The TCP download causes congestion at the bottleneck air interface, which has the effect of increasing the transmission time between the crossover node and the old access point. Even if the crossover node delays packet duplicates by 50 ms as shown in Fig. 4b, the packet stream at the new access point is still seen to be "ahead" of the old access point. This condition manifests itself at the mobile host during handoff as a sudden increase in the observed transport sequence numbers triggering TCP's retransmit and recovery mechanisms. On the other hand, if T_{ss} is much larger, as shown in Fig. 4c, the packet stream at the new access point will be "behind" the old access point. Packet loss is eliminated at the expense of duplication in this case. The value of T_{ss} that leads to zero packet loss and duplication is equal to the layer 2 handoff time, plus the difference between the transmission time to the old and new access points. If the latter is larger, packet duplication cannot be avoided.

We observe a number of similarities between the performances of these two enhanced handoff schemes. Both enhancements buffer packets for some time. In both cases, the amount of time data packets are buffered influences handoff performance. Both are capable of totally eliminating packet loss at the expense of packet duplication. The only performance difference is that Hawaii's forwarding scheme introduces packet reordering in addition to duplication. The effect of reordering is also visible in Fig. 5. The performance of Hawaii MSF handoff, as seen by the application, is somewhat lower than that of Cellular IP semi-soft handoff. This difference is because the TCP protocol reacts adversely to the level of packet reordering introduced by the Hawaii MSF scheme. Note that the parameters for these simulations were T_{ss} = 120 ms and T_{msf} = 50 ms.

Figure 5 also plots the throughput obtained by Cellular IP hard handoff using NewReno congestion control instead of Reno. The results demonstrate that NewReno can effectively improve performance in the presence of frequent handoff. This is attributed to the fact that batch loss events, which are the main cause of the drop in throughput experienced by Reno TCP flows, have less impact on NewReno flows. Applying NewReno congestion control represents a different approach to improving handoff performance in relation to the micromobility protocol enhancements, as previously discussed. Rather than eliminating packet loss, NewReno makes the end system more robust to packet loss. NewReno is not designed to compensate for loss that is specific to handoff behavior. However, it can be advantageous, for example, in the case of batch losses due to radio fading. NewReno is designed for TCP flows, while micromobility protocols can reduce disruption experienced by other transport protocols (e.g., UDP, RTP) too.

Deployment Considerations

We have studied, compared, and evaluated the architectural aspects and performance properties of Cellular IP, Hawaii, and Hierarchical Mobile IP. A number of other design considerations, however, can influence the performance and suitability of micromobility protocols.

Protocol Layers

The choice of the protocol layer that supports MRPs has important implications related to network management. MRPs will typically reuse the network management, traffic engineering, and quality of service (QoS) features supported by a particular protocol layer. For example, a micromobility access network built from Ethernet switches [13] can take advantage of IEEE 802.1p priorities. MRPs operating at L3, on the other hand, can employ differentiated services per-hop behaviors. The choice of the MRP protocol layer also impacts the ability to mix mobile and nonmobile traffic in a single access network. Tunneling-based micromobility protocols are easily deployed on top of existing mobility-unaware wide-area IP networks. Solutions operating within the IP layer [17, 18] can mix mobile and nonmobile traffic at L2 and are therefore more suited to LANs or networks dedicated to mobile traffic. In addition, the choice of protocol layer also influences device availability and cost, the type of encapsulation or tunneling machinery required by MRPs, and the means to integrate the access network with a global mobility solution.

LOCATION UPDATING

The means of updating per-host location information has implications on the design of MRPs. Because Cellular IP is based on implicit signaling and uses data packets to drive location updates, MRP nodes need packet snooping capabilities and per-packet logic to update the location database. In contrast, explicit signaling approaches place additional load on an MRP's general-purpose processors, which would be the case for Hawaii. In this case, signaling load needs to be carefully considered, especially in the case of protocols that have a single-level MRP hierarchy, as in the case of base Hierarchical Mobile IP. The extent to which signaling influences scalability limits of a protocol also depends on the complexity of generating and interpreting messages.

AAA AND SECURITY

The level of security support required in a micromobility protocol is determined by the operational networking scenario in which the protocol operates. While authenticating location update messages seems necessary in many cases, data encryption over the air interface or in the fixed network is not always needed. User authentication for authorization or accounting may be required in some cases, while anonymous free access is sufficient in others. The extent to which various micromobility protocols support security and authentication, authorization, and account (AAA) functions has a large impact on the practical applicability of micromobility protocols. The security model employed by micromobility protocols also influences network and device performance, QoS, manageability, handoff performance, and interoperation with other (possibly global) AAA systems.

CONCLUSION

In this article we present a comparison of a number of IP micromobility protocols that have been designed and implemented over the past several years. Micromobility protocols complement Mobile IP with fast, seamless, local handoff control. We introduce the notion of mobile routing points and established a generic micromobility model to help best understand the performance and design issues of Cellular IP, Hawaii and Hierarchical Mobile IP. Despite the different design approaches, the fundamental operating principles that underpin these protocols are similar. We consider Cellular IP, Hawaii, and Hierarchical Mobile IP as realizations of this generic model.

We developed the CIMS [11] ns-2 extension that supports separate programming models for Cellular IP, Hawaii, and Hierarchical Mobile IP. We present a set of simulation results to illustrate the performance of these protocols. The results demonstrate that the basic handoff performance depends only on the position of the crossover MRP. We identify three fundamental design choices for selecting the crossover MRP.

We compare the performance of Cellular IP semi-soft handoff and Hawaii MSF handoff, identifying a number of similarities and differences. We also discuss a number of differences not directly related to handoff quality. We argue that in selecting the appropriate micromobility protocol for a given network environment, these issues may be more important than the small differences we found in terms of user-perceived handoff quality.

A number of open issues remain. Micromobility protocols will have to support the delivery of a variety of traffic including best effort and real-time traffic. There has been very little work on a suitable QoS model for micromobility. Extending the differentiated services model to micromobility seems like a logical starting point. However, differentiated services concepts such as aggregation, per-hop behavior, service level agreement, and slow timescale resource management may be impractical in wireless IP networks. For example, it may be impractical to allocate resources at every base station in a wireless access network in support of a service level agreement that offers assured service, or use traffic engineering techniques that promote underutilization of wireless links in support of some per-hop behavior characteristic. Work on QoS support for micromobility is predicated on differentiated services first being resolved in the wired network.

Finally, there has been considerable debate in the IETF on suitable fast and seamless handoff extensions for Mobile IPv4 and Mobile IPv6. For a summary of the various proposals discussed over the last several years see [1]. The development of Cellular IP, Hawaii, and Hierarchical Mobile IP has led to significant discussion in the community and has helped shape the ongoing standardization efforts within the IETF on low-latency handoff, context transfer, QoS, and IP paging.

ACKNOWLEDGMENT

This work is sponsored by the NSF WIRELESS TECHNOLOGY Award ANI-9979439 with support from Ericsson, Intel, and Nortel Networks.

REFERENCES

[1] A. T. Campbell and J. Gomez, "IP Micro-Mobility Protocols," *ACM SIGMOBILE, Mobile Comp. and Commun. Rev.*, vol. 4, no. 4, Oct. 2001, pp. 45–54.
[2] C. Perkins, Ed., "IP Mobility Support," Internet RFC 2002, Oct. 1996.
[3] IETF Mobile IP Working Group, http://www.ietf.org/html.charters/mobileip-charter.html
[4] IETF Seamoby Working Group, http://www.ietf.org/html.charters/seamoby-charter.html
[5] K. El Malki *et al.*, "Low Latency Handoffs in Mobile IPv4," Internet draft, draft-ietf-mobileip-lowlatency-handoffs-v4-03, work in progress, Nov. 2001.
[6] J. Kempf, "Dormant Host Mode Alerting ('IP Paging') Problem Statement," Internet RFC 3132, June 2001.
[7] D. B. Johnson and C. Perkins, "Route Optimization in Mobile IP," Internet draft, draft-ietf-mobileip-optim-09, work in progress, Feb. 2000.
[8] A. Valkó, "Cellular IP: A New Approach to Internet Host Mobility," *ACM SIGCOMM Comp. Commun. Rev.*, vol. 29, no. 1, Jan. 1999, pp. 50–65.
[9] R. Ramjee *et al.*, "HAWAII: A Domain-Based Approach for Supporting Mobility in Wide-area Wireless Networks," *Proc. IEEE Int'l. Conf. Network Protocols*, 1999.

The choice of protocol layer also influences device availability and cost, the type of encapsulation or tunneling machinery required by MRPs, and the means to integrate the access network with a global mobility solution.

The development of Cellular IP, Hawaii and Hierarchical Mobile IP has lead to significant discussion in the community and has helped shape the on-going standardization efforts within the IETF on low-latency handoff, context transfer, QoS, and IP paging.

[10] E. Gustafsson, A. Jonsson, and C. Perkins, "Mobile IP Regional Registration," Internet draft, draft-ietf-mobileip-reg-tunnel-03, work in progress, July 2000.

[11] Micromobility home page, http://comet.columbia.edu/micromobility

[12] H. Haverinen and J. Malinen, "Mobile IP Regional Paging," Internet draft, draft-haverinen-mobileip-reg-paging-00.txt, work in progress, June 2000.

[13] J. W. Lockwood, "Implementation of Campus-wide Wireless Network Services using ATM, Virtual LANs and Wireless Basestations," *Wireless Commun. and Net. Conf.*, New Orleans, LA, Sept. 1999.

[14] IEEE 802.1d standard, MAC Bridges, 1998 edition.

[15] IEEE 802.11 working group, http://grouper.ieee.org/groups/802/11

[16] ns2 home page, http://www.isi.edu/nsnam/ns

[17] A. Campbell *et al.*, "Cellular IP," Internet draft, draft-ietf-mobileip-cellularip-00, work in progress, Dec. 1999.

[18] R. Ramjee *et al.*, "IP micro-mobility Support using HAWAII," Internet draft, draft-ietf-mobileip-hawaii-00, work in progress, June 1999.

[19] A. Valkó *et al.*, "Design and Performance of Cellular IP Access Networks," *IEEE Pers. Commun.*, special issue on IP-Based Mobile Telecommunications Networks, 2000.

[20] K. El-Malki and H. Soliman, "Fast Handoffs in Mobile IPv4," Internet draft, draft-elmalki-mobileip-fast-handoffs-03, work in progress, Sept. 2000.

[21] P. Calhoun *et al.*, "Foreign Agent Assisted Handoff," Internet draft, draft-calhoun-mobileip-proactive-fa-03, work in progress, Nov. 2000.

[22] C. Perkins, G. Krishnamurthi, and R. Chalmers, "Buffer Management for Smooth Handovers in Mobile IPv6," Internet draft, draft-krishnamurthi-mobileip-buffer6-00, work in progress, July 2000.

[23] A. O'Neill, S. Corson, and G. Tsirtsis, "Generalized IP Handoff," Internet draft, draft-oneill-craps-handoff-00, work in progress, Aug. 2000.

Biographies

ANDREW T. CAMPBELL (campbell@comet.columbia.edu) is an associate professor in the Department of Electrical Engineering, and a member of the COMET Group and the Columbia Networking Research Center, Columbia University, New York. His areas of interest encompass mobile networking, programmable networks, and QoS research. He currently serves as technical program co-chair for the 8th ACM International Conference on Mobile Computing and Networking (ACM MobiCom 2002), and technical chair of the special track on networking technologies, services, and protocols for IFIP Networking 2002. He received his Ph.D. in computer science in 1996 and the NSF CAREER Award for his research in programmable mobile networking in 1999.

ZOLTAN RICHARD TURANYI received his M.Sc. degree in computer science in 1996 from the Technical University of Budapest. At the same year he started his Ph.D. studies focusing on packet network performance, QoS, and simulation issues. Currently he is a staff member at Ericsson Research, TrafficLab in Hungary. His research interests include mobile and ad hoc packet networks, QoS, network simulation, and routing.

ANDRAS G. VALKO received his M.Sc. (EE) and Ph.D. (CS) degrees in 1994 and 1999, respectively, both from the Technical University of Budapest (TUB), Hungary. In 1994 he joined the High Speed Networks Laboratory at TUB where he worked mainly with ATM performance issues. Since 1996 he is with Ericsson Research, Traffic Analysis and Network Performance Laboratory. His research interest is focused on Internet mobility, wireless/mobile Internet access, mobile ad hoc networking, and mobile systems performance analysis. The work presented in this article was started while he was a visiting researcher at CTR, Columbia University, New York.

JAVIER GOMEZ CASTELLANOS [StM] obtained a B.S. degree with honors in electrical engineering in 1993 from National Autonomous University of Mexico (UNAM), Mexico, and an M.S. degree in electrical engineering in 1996 from Columbia University, New York. Since 1996 he has been a Ph.D. student in the COMET Group at the Center for Telecommunications Research, Columbia University, New York. His research interests cover routing, QoS, and power-aware design for cellular and ad hoc wireless networks.

CHIEH-YIH WAN is a Ph.D. candidate in the COMET Group and the Department of Electrical Engineering, Columbia University. His areas of interest are IP micromobility and pervasive networks.

SANGHYO KIM is a Ph.D. candidate in the COMET Group and Department of Electrical Engineering, Columbia University. His areas of interest are IP micromobility and programmable mobile networks.