Journal of High Speed Networks 11 (2002) 177–198
IOS Press

Internet micromobility

Andrew T. Campbell, Javier Gomez, Sanghyo Kim, Zoltán R. Turányi, András G. Valkó and Chieh-Yih Wan
COMET Group, Columbia University, New York, USA

Abstract. This paper presents results from the Cellular IP Project at Columbia University on Internet micromobility. Cellular IP complement Mobile IP with support for fast, seamless and local handoff control, and IP paging. We discuss the design, implementation and evaluation of the Cellular IP protocol using simulation, analysis and experimentation. We report on the ability of Cellular IP to offer seamless mobility for TCP and UDP applications operating in highly mobile environments. We present a comparison of a number of IP micromobility protocols using the Columbia IP Micromobility Software (CIMS) ns-2 extension that supports separate programming models for Cellular IP, Hawaii and Hierarchical Mobile IP. We discuss simulation results to illustrate the performance of these micromobility protocols. The source code for CIMS and the Cellular IP experimental testbed are freely available from the Web (comet.columbia.edu/cellularip).

1. Introduction

Recent initiatives to add mobility to the Internet have mostly focused on the issue of address translation [1] through the introduction of location directories and address translation agents. The problem of address translation is fundamental to global Internet mobility and comes from the hierarchical nature of IP addressing. In Mobile IP [4] packets addressed to a mobile host are delivered using regular IP routing to a temporary address assigned to a mobile host at its actual point of attachment. This approach results in a simple and scalable scheme that offers global mobility. Mobile IP is not appropriate, however, for seamless mobility because after each migration a local address must be obtained and communicated to a possibly distant location directory or home agent (HA). We define the term seamless mobility as the ability of the network to support fast handoff between base stations with low delay and minimum or zero packet loss. We believe that support for seamless mobility will be needed in order to provide good service quality to mobile users particularly in pico-cellular environments where the rate of handoff and associated signaling load grows rapidly.

Network support for seamless mobility was not a primary design consideration when Mobile IP was first defined in the early 90s. More recently the Mobile IP Working Group has been addressing this issue. In the case of frequent handoff, *micromobility protocols* have been proposed [2,10,11,14] to handle local movement 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 distant home agents when mobile hosts remain inside their local coverage areas. Eliminating registration in this manner reduces the signaling load experienced by the core network in support of mobility. Reducing signaling in this manner is necessary for the wireless Internet to scale to support very large volumes of wireless subscribers.

We envision a wireless Internet with many hundreds of millions of wireless subscribers. As in the case of the cellular phone we imagine that wireless IP communicators will be “turned on” around the clock ready to receive or initiate services. In fact the vast majority of subscribers will not be actively communicating most of the time. Rather, wireless IP communicators will be switched on ready for service constantly reachable by the wireless Internet. In essence, mobile hosts will be in an idle state, passively connected to the network infrastructure. As in the case of the mobile telephony network, it will be sufficient for the wireless Internet only to know the approximate location of its population of idle users. The exact location of idle mobile hosts only becomes important when data needs to be forwarded to them in which case the network needs to be able to efficiently search and find these users in a scalable and timely manner. In cellular telephony systems this process is called *paging*.

0926-6801/02/$8.00

As the number of mobile subscribers grows so does the need to provide efficient location tracking in support of idle users and paging in support of active communications. In order to achieve scalable location management the wireless Internet needs to handle the location tracking of active and idle mobile hosts independently. Support for passive connectivity 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 core network. Reducing signaling over the air interface also has the benefit of preserving the power reserves of mobile hosts.

Currently, Mobile IP does not support the notion of seamless mobility, passive connectivity or paging. We argue that the future wireless Internet will need to support these requirements in order to deliver service quality, minimize signaling and scale to support hundreds of millions of subscribers. In this paper, we present results from Cellular IP Project [2] at Columbia University, which is investigating new approaches to Internet host mobility. Cellular IP [2,3] is a micromobility protocol that is optimized to provide local access to a Mobile IP enabled Internet in support of fast moving wireless hosts. The protocol incorporates a number of important design features present in cellular networks but remains firmly based on IP design principles. The protocol is specifically designed to support seamless mobility and paging. Cellular IP access networks can be constructed in a plug and play manner scaling from pico-cellular to metropolitan area networks. Distributed location management, routing and handoff algorithms lend themselves to a simple, efficient and low cost software implementation for host mobility requiring no new packet formats, encapsulation or address space allocation beyond what is present in IP.

This paper is structured as follows. Section 2 presents related work. Section 3 provides an overview of the Cellular IP protocol and discusses per-host routing, seamless handoff control, paging and security. Section 4 presents results from an implementation of Cellular IP in an experimental wireless testbed. We analyze the performance of Cellular IP in support of UDP and TCP applications. Section 5 presents a detailed quantitative comparison of Cellular IP and other micromobility protocols discussed in the literature [10,11]. Finally, in Section 6, we present some concluding remarks.

2. Related work

A number of micromobility solutions have been discussed in the literature. In [14] a hierarchical mobility model is described where independent wireless access networks interwork with a global mobility protocol. Address translation and security are functions of the global mobility solution. In contrast, wireless access networks provide mechanisms for the support of local location management and mobility. In [10] and [12] Mobile IP is extended by arranging foreign agents in a hierarchy. The top of the hierarchy is rooted at the edge of the access network and is defined by a care-of address registered with home agents. Upon reception of a packet, the foreign agent at the top of the hierarchy interacts with a local database to determine which lower level foreign agent (located in the access network) to forward the packet to. This procedure may be repeated, depending on the depth of the routing hierarchy. Similar ideas are adopted in the case of campus and domain foreign agents [14] and local registration schemes [16].

We observe that Cellular IP and the protocols discussed above employ per mobile host state and hop-by-hop routing to achieve fast handoff control. These hierarchical mobility proposals do not, however, support the notion of passive connectivity with its separation of active and idle users, as is the case with Cellular IP. In these proposals, a foreign agent maintains database entries for each mobile host in its region having to search a potentially large database in order to route each packet. In contrast, Cellular IP routing cache only contains entries for mobile hosts that have recently transmitted packets. This reduces the search time and increases protocol scalability. Other differences exist. Hierarchical foreign agent schemes operate on top of IP whereas Cellular IP is itself a layer three routing protocol; that is, Cellular IP replaces IP routing in the wireless access network but without modifying the IP packet format and forwarding mechanism. To increase efficiency, location management is integrated with routing in Cellular IP access networks. The per host location information stored in Cellular IP nodes is not a network address. Rather, per host location state represents the next hop route to forward packets toward a given mobile host. Such an integrated approach simplifies both routing and location management in wireless access networks.

In [18] off-the-shelf Ethernet switches and wireless LAN cards are used to build wireless access networks. The learning feature of Ethernet switches is used for location management. Data frames transmitted by mobile hosts are used to establish and refresh location information inside the access network. Although this approach of using Ethernet switches for location management results in simple, cheap and efficient access networks, the concept is hard to extend with desirable features, such as smooth and secure handoff or paging. Cellular IP uses data packets to refresh location management state and can operate at layer two or three. However, mobility support is built into Cellular IP nodes.

Support for seamless mobility, passive connectivity and paging is fundamental to improving scalability, minimizing power consumption and delivering suitable service quality to mobile hosts. Few solutions, however, support these features [9]. One protocol that supports seamless mobility, passive connectivity and paging is Hawaii [11]. In contrast to Cellular IP nodes, which preserve the simplicity of the Ethernet switch solution discussed above, Hawaii nodes are IP routers. It is interesting to note that low-cost layer two switches can be used to build Cellular IP access networks supporting tens of thousands of mobile hosts [21]. We believe that this approach becomes increasingly important when constructing low-cost pico-cellular infrastructure. The use of an all IP-based router solution for pico-cellular networks may become prohibitively expensive. This motivates the need to have a layer two and three solution for micromobility. Hawaii assumes that an intra-domain routing protocol is operational in the access network allowing each node to have routes to other nodes. This routing information is used to exchange explicit signaling messages and to forward packets between old and new access points during handoff. The use of explicit signaling messages is limited in Cellular IP, which uses the IP data packets to convey location and paging information.

Different proposals have different scaling properties. The base stations associated with the original Columbia protocol [20] represent radio-enabled routers operating in campus area networks. Base stations broadcast search messages among each other in order to determine the location of mobile hosts. By tunneling packets between base stations, the Columbia scheme effectively creates a mobile overlay network on top of the wired campus network. This protocol works well for small numbers of mobile hosts but encounters scalability problems due to the nature of the broadcast search algorithm used. The local mobility protocol proposed by [14] uses workstations as base stations and hence is more appropriate in networks with small cells. However, this protocol is similar to commercially available solutions [8] in the respect that it only provides mobility within the area covered by a local area network. A key design requirement of Cellular IP is its capability to scale from local to metropolitan area networks. Cellular IP can be deployed across a number of different installations including office, campus and wireless ISP coverage areas [21].

Many of the proposals discussed above are capable of minimizing service disruption during handoff. In [19] an IP multicasting technique is used to support fast handoff. Here mobile hosts are identified by multicast IP addresses. Base stations are capable of joining a mobile host's multicast group. This includes the current base station that the mobile host is connected to as well as others, which it may move to after handoff. In the latter case, packets are delivered to the new base station even before the host has migrated. In Hawaii, seamless handoff is achieved by exchanging a series of signaling messages between the old and new base stations. This facilitates the forwarding of packets from the old base station to the new one during handoff. Both of these approaches require nodes either to be multicast capable routers or process signaling messages. Cellular IP handoff aims at simplicity, eliminating the reliance on multicast and minimizing explicit signaling.

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 that have been discussed over the last several years see [1]. 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.

3. Protocol description

As the name suggests Cellular IP inherits cellular principles for mobility management such as passive connectivity, paging and fast handoff control but implements them around an IP paradigm. Cellular IP access networks require minimal configuration (e.g., similar to switched Ethernet LANs) thereby easing the deployment and management of wireless access networks. An important concept in Cellular IP design is simplicity and the minimal use of explicit signaling enabling low cost implementation of the protocol. In what follows, we present an overview of the Cellular IP protocol and discuss support for routing, handoff, paging and security. For a full specification of the protocol see [2,3].

3.1. Network model

The universal component of Cellular IP access networks is the *base station* which serves as a wireless access point and router of IP packets while performing all mobility related functions. Base stations are built on the regular IP forwarding engine with the exception that IP routing is replaced by Cellular IP routing and location management. Cellular IP access networks are connected to the Internet via *gateway* routers. Mobile hosts attached to an access network use the IP address of the gateway as their Mobile IP care-of address. Figure 1 illustrates the path taken by packets addressed to a mobile host. Assuming Mobile IPv4 [4] and no route optimization [5], packets first will be routed to the host's home agent and then tunneled to the gateway. The gateway "detunnels" packets and forwards them toward a base station. Inside a Cellular IP network, mobile hosts are identified by their home address and data packets are routed without tunneling or address conversion. The Cellular IP routing protocol ensures that packets are delivered to the host's actual location. Packets transmitted by mobile hosts are first routed toward the gateway and from there on to the Internet.

In Cellular IP, location management and handoff support are integrated with routing. To minimize control messaging, regular data packets transmitted by mobile hosts are used to refresh host location information. *Uplink* packets are routed from a mobile host to the gateway on a hop-by-hop basis. The path taken by these packets is cached by all intermediate base stations. To route *downlink* packets addressed to a mobile host the path used by recently transmitted packets from the mobile host is reversed. When the mobile host has no data to transmit then it sends small, special IP packets toward the gateway to maintain its downlink routing state. Following the principle of passive connectivity mobile hosts that have not received packets for some period of time allow their downlink routes to be cleared from the cache as dictated by soft state timers. *Paging* is used to route packets to idle mobile hosts in a Cellular IP access network.

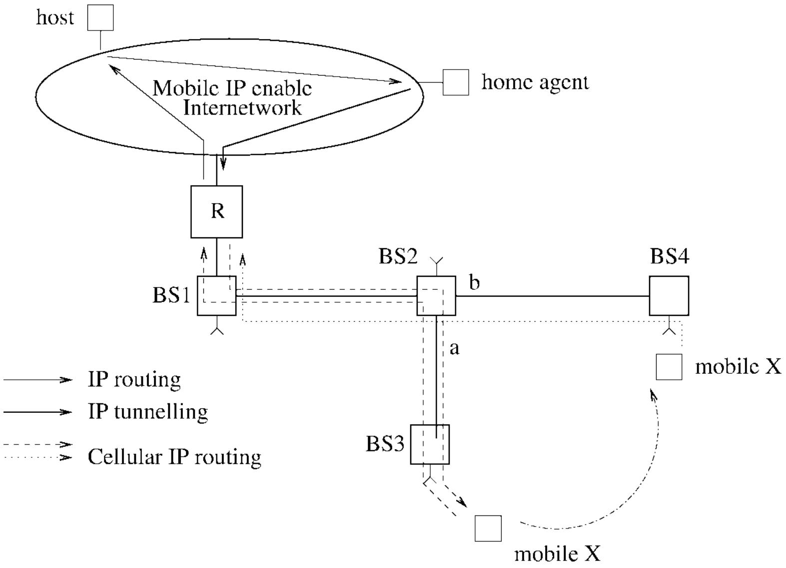


Fig. 1. Cellular IP access network.

3.2. Routing

A Cellular IP gateway periodically broadcasts a beacon packet that is flooded in the access network. Base stations record the neighbor they last received this beacon from and use it to route packets toward the gateway. All packets transmitted by mobile hosts regardless of their destination address are routed toward the gateway using these routes.

As these packets pass each node on route to the gateway their route information is recorded as follows. Each base station maintains a *routing cache*. When a data packet originated by a mobile host enters a base station the local routing cache stores the IP address of the source mobile host and the neighbor from which the packet entered the node. In the scenario illustrated in Fig. 1 data packets are transmitted by a mobile host with source IP address **X** and reach base station **BS3** via **BS2**. In the routing cache of **BS2** this is indicated by a *mapping* **(X,BS3)**. This soft-state mapping remains valid for a system specific time called *route-timeout*. Data packets are used to maintain and refresh mappings. As long as mobile host **X** is regularly sending data packets then base stations along the path between the mobile's actual point of attachment and the gateway will maintain valid routing cache mappings forming a soft-state path between the mobile host and gateway node. Packets addressed to the mobile host **X** are routed on a hop-by-hop basis using this established routing cache.

A mobile host may sometimes wish to maintain its routing cache mappings even though it is not regularly transmitting data packets. A typical example of this is when a mobile host receives a UDP stream of packets on the downlink but has no data to transmit on the uplink. To keep its routing cache mappings valid mobile hosts transmit *route-update packets* on the uplink at regular intervals called *route-update time*. These packets are special ICMP packets addressed to the gateway. Route-update packets update routing cache mappings as is the case with normal data packets. However, route-update messages do not leave a Cellular IP access network.

3.3. Handoff

Cellular IP supports two types of handoff scheme. Cellular IP *hard handoff* is based on 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* minimizes packet loss providing improved TCP and UDP performance over hard handoff.

3.3.1. Hard handoff

Mobile hosts listen to beacons transmitted by base stations and initiate handoff based on signal strength measurements. To perform a handoff a mobile host tunes its radio to a new base station and sends a route-update packet. The route-update message creates routing cache mappings on route toward the gateway configuring the downlink route cache to point toward the new base station. Handoff latency is the time that elapses between handoff initiation and the arrival of the first packet along the new route. In the case of hard handoff this duration is equal to the round-trip time between the mobile host and the *cross-over base station*, as illustrated in Fig. 2. We define the cross-over base station as the common branch node between the old and new base stations, an example of which is illustrated in the figure. In the worst case the cross-over point is the gateway. During this interval downlink packets may be lost. Mappings associated with the old base station are not cleared when handoff is initiated. Rather, mappings between the cross-over node and the old base station timeout and are removed. No packets are transmitted along the old path once the route-update message has created a new mapping at the cross-over base station that points toward the new base station.

Although packets may get lost during a hard handoff, the time taken to redirect packets to the new point of attachment is shorter than that of Mobile IP. This is due to the fact that only a local node has to be notified rather than a possibly distant home agent as in the case of Mobile IP.

There are several ways to reduce packet loss during handoff. One approach relies on interaction between the old and new base stations [11] during handoff. In this case the new base station notifies the old base station of the pending handoff. Packets that arrive at the old base station after notification of handoff are forwarded to the new base station and onto the mobile host. In contrast, packets that arrive at the old base station before notification is

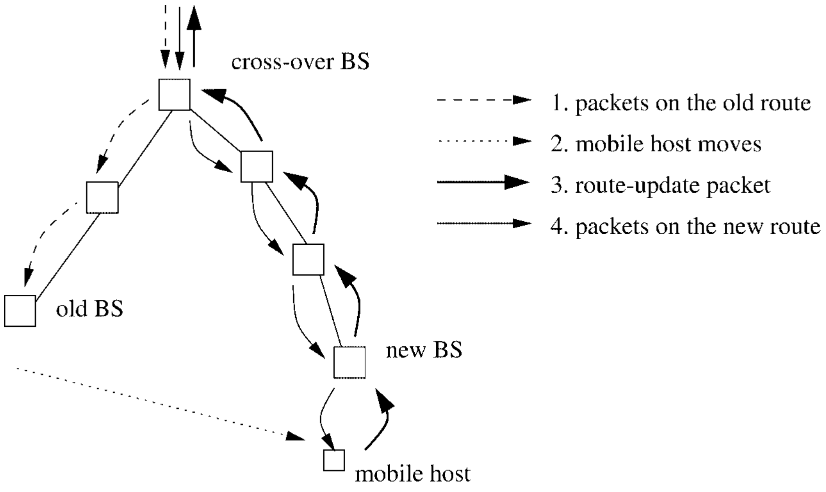


Fig. 2. Cellular IP handoff.

complete will be lost. If the notification time (i.e., the round-trip time between the new and the old base stations) is not smaller than handoff duration (i.e., the round-trip time between the new and cross-over base stations) then this approach does not significantly improve handoff. An additional cost of these schemes is that communications, signaling and information state exchange required between base stations for this approach to work. To preserve the simplicity of hard handoff, Cellular IP employs a different approach to counter the problem of packet loss.

3.3.2. Semisoft handoff

After hard handoff, the path to the old base station remains in place until the soft-state cache mappings time out. We leverage this feature to support a new handoff service called semisoft handoff that improves handoff performance while maintaining the lightweight nature of the base Cellular IP protocol. Semisoft handoff calls for one temporary state variable to be added to the protocol running in the mobile hosts and base stations.

Semi-soft handoff scales well for large numbers of mobile hosts and frequent handoff and comprises two architectural components. First, in order to reduce handoff latency, the routing cache mappings associated with the new base station must be created before the actual handoff takes place. Before a mobile host hands off to a new access point it sends a *semisoft packet* to the new base station and immediately returns to “listening” to the old base station.

The purpose of the semisoft packet is to establish new routing cache mappings between the cross-over base station and the new base station. During this route establishment phase the mobile host is still “connected” to the old base station. After a *semisoft delay*, the mobile host performs a regular handoff. The semisoft delay can be an arbitrary value that is proportional to the mobile to gateway round-trip delay. This delay ensures that by the time the mobile host finally tunes its radio to the new base station, its downlink packets are being delivered through both the old and new base stations. We observe that downlink packets consume twice the amount of resources during this period. However, this period represents a short duration when one considers the complete semisoft handoff process.

While the semisoft packet ensures that mobile hosts continue to receive packets immediately after handoff, it does not however, assure smooth handoff between base stations. Depending on the network topology and traffic conditions, the time to transmit packets from the cross-over point to the old and new base stations may differ and the packet streams transmitted through the two base stations will typically be unsynchronized. If the new base station is “behind” the old one, the mobile host will receive duplicate packets, which does not disrupt many applications. For example, TCP will not be forced into slow start due to the arrival of duplicate acknowledgments. If the new base station is “ahead” then packets will be missing from the stream received by at the mobile host.

The second architectural component of semisoft handoff resolves this issue of the new base station getting ahead. The solution to this problem is based on the observation that perfect synchronization of packet streams is unnecessary. This condition can be eliminated by temporarily introducing a constant delay along the new path between the cross-over base station and the new base station using a simple “delay device” mechanism. The device needs to provide sufficient enough a delay to compensate, with high probability, for the time difference between the two streams traveling on the old and new paths. Optimally, the device delay should be located at the cross-over base station. The cross-over base station is aware that a semisoft handoff is in progress from the fact that a semisoft packet arrives from a mobile host that has mapping to another interface. Mappings created at cross-over points by the reception of semisoft packets include a flag to indicate that downlink packets must pass through a delay device before being forwarded for transmission along the new path. After handoff is complete, the mobile host sends a data or route-update packet along the new path. These packets have the impact of clearing the flag causing all packets in the delay device to be forwarded to the mobile host. Base stations only need a small pool of delay buffers to resolve this issue. Packets that cannot sustain additional delay can be forwarded without passing through the delay device. This differentiation can be made on a per packet basis, using e.g., differentiated service or transport (e.g., TCP, UDP or RTP) codepoints.

3.4. Paging

Typically, fixed hosts connected to the Internet (e.g., desktop computers) remain on-line for extended periods of time even though most of the time they do not communicate. Being “always connected” in this manner results in being reachable around the clock with instant access to Internet resources. Mobile subscribers connected to the wireless Internet will expect similar service. However, in the case of mobile hosts maintaining location information in support of being continuously reachable would require frequent location updates which would consume precious bandwidth and battery power.

Cellular systems employ the notion of passive connectivity to reduce the power consumption of idle mobile hosts. Base stations are geographically grouped into *paging areas*. When there is no call ongoing, mobile hosts only need to report their position to the network if they move between paging areas. This makes location update and handoff support for idle hosts unnecessary. When an incoming call is detected at the gateway a paging message is transmitted to the mobile host’s current paging area to establish the call. The mobile node informs the infrastructure of its location as a result of the paging process and transitions to active mode to take the call.

While the definition of an idle mobile device is well understood in the context of cellular systems, which are connection oriented in nature, its meaning in IP-based mobile networks is unclear. Cellular IP defines an idle mobile host as one that has not transmitted packets for a system specific time *active-state-timeout*. Due to lack of updates, the soft-state routing cache mappings of idle mobile hosts will time out in a fully distributed manner. In order to remain “reachable” mobile hosts transmit *paging-update* packets at regular intervals defined by a *paging-update-time*. A paging-update packet is an ICMP packet, which is addressed to the gateway and is distinguished from route-update packets by its type parameter value. Mobile hosts send paging-update packets to base stations that have better signal quality. As in the case of data and route-update packets, paging-update packets are routed toward the gateway on a hop-by-hop basis. Base stations may optionally maintain *paging cache*. Paging cache has the same format and operation as routing cache with the following exceptions. Paging cache mappings have a longer timeout period called *paging-timeout* hence a longer interval exists between consecutive paging-update packets. In addition, any packet sent by mobile hosts including route-update packets can update paging cache. However, paging-update packets cannot update routing cache. This results in idle mobile hosts having mappings in the paging cache but not in the routing cache. In contrast, active mobile hosts will have mappings in both routing and paging cache.

Packets addressed to a mobile host are normally routed by routing cache mappings. Paging occurs when a packet is addressed to an idle mobile host and the gateway or base stations find no valid routing cache mapping for the destination. If the base station has no paging cache, it will forward the packet to all of its interfaces except the one the packet came through. Cellular IP has no explicit paging control message. Rather, the first data packet that

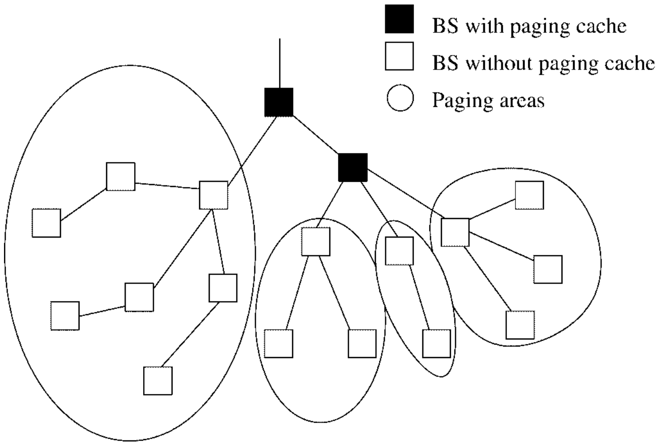


Fig. 3. Paging areas.

arrives at the gateway forms an implicit "paging message" that is forwarded in the access network. Paging cache is used to avoid broadcast search procedures. Base stations that have paging cache will only forward a paging packet if the destination has a valid paging cache mapping. In this case the paging message is only forwarded to the mapped interface. If there is no paging cache in an access network then the first packet addressed to an idle mobile will be broadcast, increasing the load on the access network.

The network operator can limit paging load in exchange for memory and processing cost by using paging cache in the access network. By placing paging cache in base stations, paging areas can be defined as required. An operator can construct paging areas and determine what nodes in the access network should support paging cache and which should not. For example, paging cache could be located at the gateway only or at the majority of the base stations in the access network. The construction of paging areas (i.e., the number of base stations that comprise a paging area) and the distribution of paging cache within a paging area (i.e., which nodes do and do not have paging cache) is a configuration issue, some examples of which are illustrated in Fig. 3.

In the case of Cellular IP a paging area identifier is broadcast as part of beacon messages. Idle mobile hosts will only transmit paging-update packets when they move between paging areas. An idle mobile host that receives a paging packet transitions from idle to active state and immediately transmits a route-update packet towards the gateway. This ensures that routing cache mappings are quickly established limiting any further paging in the location area.

3.5. *Security*

Cellular IP has been designed to support seamless and secure handoff. Mobile systems are open to a number of security problems that do not exist in their stationary counterparts. In a fixed network, the prefix of a subnet is usually configured manually and the location of the prefix is communicated between routers that have either some form of inherent trust model or use a secure protocol. This makes it hard to impersonate someone. Mobile hosts, on the other hand, must update their location while moving. These location messages make impersonation possible unless properly secured. Wireless access networks compound these security problems because packets can be snooped over the air interface. Cellular IP faces impersonation and snooping attacks because it is wireless and mobile.

Cellular IP addresses these security issues, as follows. First, only authenticated packets can establish or change cache mappings in a Cellular IP access network. By authenticating paging and routing update control messages malicious users are prevented from capturing traffic destined for mobile hosts. In Cellular IP access networks control packets are authenticated because they establish and change existing mappings. In contrast, data packets can only refresh existing mappings. Active mobile hosts transmit route-update packets during handoff to create a new chain of soft-state cache mappings that point to the new point of attachment.

In case of Cellular IP seamless handoff is of primary importance. Therefore, session keys used by mobile hosts to perform authentication must be promptly available at the new base station during handoff. Timeliness of the authentication process is critical in the case of micromobility due to the requirement of fast handoff control. In contrast, global mobility solutions may have broader requirements such as user identification, bilateral billing and service provisioning agreements. These boarder requirements out weigh the need to support fast handoff control where the scalability of the global Authentication, Authorization and Account (AAA) [17] system is of more importance than seamless handoff. One can envision, however, micromobility protocols that build on global AAA preferences by offering enhanced services (e.g., fast session key management) to aid seamless handoff.

During handoff, the new base station could hypothetically acquire a session key by contacting the old base station, the cross-over base station or some central key management server. Cellular IP, however, uses a fast session key management scheme. Rather than defining new signaling, a special session key is used in Cellular IP access networks. Base stations can independently calculate session keys. This eliminates the need for signaling in support of session key management, which would inevitably add additional delays to the handoff process. The session key is a secure hash, which combines:

1. The IP address of a mobile host (IP_{MH}).
2. A random value (R_{MH}) assigned to a mobile host when it first registers with an access network.
3. A network secret ($K_{network}$) known by all base stations within an access network.

The session key is calculated using an MD5 hash function:

$$K_{session} = \mathrm{MD5}(IP_{MH}, R_{MH}, K_{network}).$$

A session key is first calculated and transmitted to a mobile host when it first contacts the Cellular IP network during global mobility authentication and authorization. The random value R_{MH} is assigned to the mobile host at this point.

Control packets carry this random value (R_{MH}) together with authentication information. A timestamp is used for replay protection. The session key is used to perform authentication. Base stations can quickly calculate the session key by combining the IP address and the random value found in the control packet with the “network secret”. Base stations can validate the authentication easily with the session key. The base stations perform the validation process without any further communication or pre-distributed subscription databases. This results in fast and secure handoff. To enhance security, the network key could be periodically replaced thereby triggering session key changes making brute force attacks more difficult.

4. Experimental micromobility testbed

To evaluate Cellular IP performance in an experimental setting we have built a wireless testbed and designed a set of experiments to analyze the protocol. In what follows, we describe our Cellular IP testbed and experimental results. The goal of the experiments is to evaluate the performance and scalability of the protocol. Cellular IP has been implemented and evaluated on a FreeBSD 2.2.6 software platform. Note that other operating systems are supported including Windows and Linux. In this section, we report and evaluate the FreeBSD implementation of the protocol. The Cellular IP base station and mobile software modules execute in user space and use the Berkeley Packet Filter’s Packet Capture library (PCAP) [7] for processing and forwarding of IP packets. The Cellular IP testbed source code is freely available [33].

The experimental results reported in this section are based on measurements taken from the Cellular IP testbed illustrated in Fig. 4. The access network consists of three base stations based on multihomed 300 MHz Pentium PCs. One of the base stations also serves as a gateway router to the Mobile IP enabled Internet. Base stations

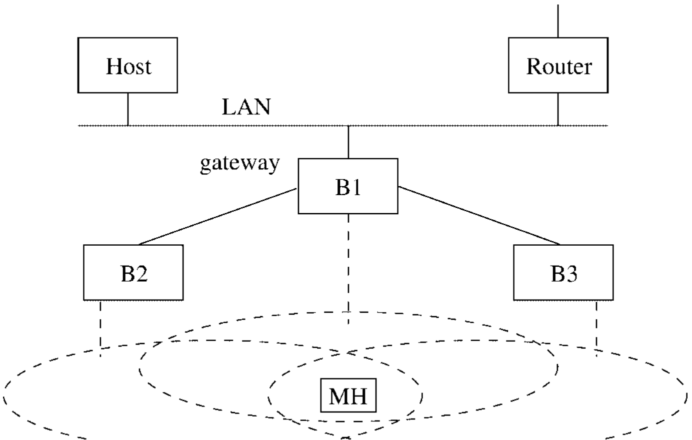


Fig. 4. Cellular IP testbed.

are interconnected using 100 Mbps full duplex links. Mobile hosts are 300 MHz Pentium PC notebooks. Mobile hosts and base stations are equipped with 2 Mbps WaveLAN 2.4 GHz radio interfaces. Note that the current software release [33] of the protocol supports device drivers for a number of 11 Mbps radios including IEEE 802.11 WaveLAN and Aironet radios. The 2.4 GHz WaveLAN radios can operate at eight different frequencies to avoid interference between adjacent cells. In the testbed the base stations are statically assigned frequencies while mobile hosts can dynamically change frequency to perform handoff. Throughout the experiments the mobile host shown in Fig. 4 is in an overlapping region of cells. For experimentation purposes a utility tool located on the mobile host is capable of periodically triggering handoff regardless of the measured signal quality. Handoff initiated by the utility tool is, however, identical to the Cellular IP mobile initiated handoff.

4.1. Handoff

An important objective of this experiment is to analyze the performance of hard and semisoft handoff and investigate the impact of handoff on UDP and TCP performance. Here we measure the packet loss for hard and semisoft handoff, respectively.

4.1.1. UDP performance

During this experiment the mobile host shown in Fig. 4 receives 100 byte UDP packets at rates of 25 and 50 packets per second (pps) while making periodic handoffs (driven by the utility tool) between base stations **B2** and **B3** every 5 seconds.

The measurement results are plotted in Fig. 5. Each point on the graph was obtained by averaging loss measurements over 50 consecutive handoffs. The solid lines in Fig. 5 show that hard handoff causes packet losses proportional to the round-trip time and the downlink packet rate. Under these experimental conditions hard handoff results in at least one packet loss for small mobile to gateway round-trip delays and up to four packet losses for delays of 80 ms.

The dashed line in Fig. 5 represents the packet loss results from Cellular IP semisoft handoff. The experimental conditions for semisoft and hard handoff are identical. In this experiment, a delay device buffers packets before they are forwarded along the new downlink path. Each downlink packet is inserted into the delay device at the cross-over base station **B1** until the arrival of the next downlink packet at which point the first packet is dequeued and forwarded toward the new base station. When the semisoft handoff is complete, the last packet is cleared from the buffer and is sent to the mobile host. Figure 5 illustrates that semisoft handoff eliminates packet loss. Note that buffering a single packet in the delay device is sufficient to eliminate loss even in the case of a large round-trip time where hard handoff results in the loss of up to four packets.

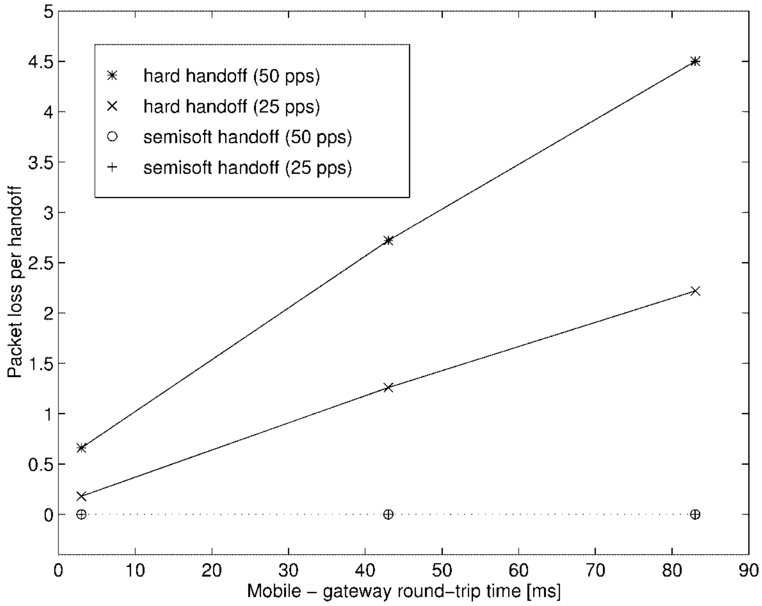


Fig. 5. UDP packet loss with handoff.

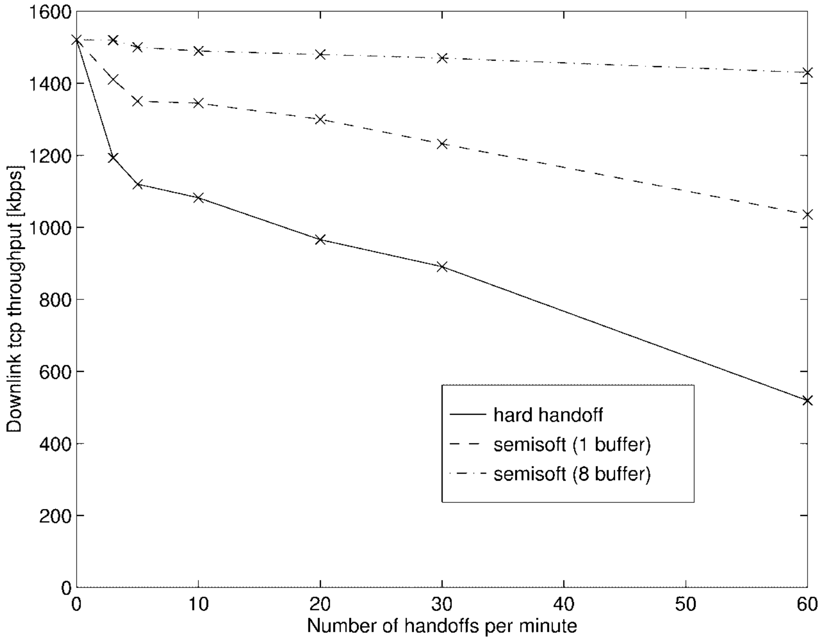


Fig. 6. TCP throughput with handoff.

4.1.2. TCP performance

In the next experiment, we study the impact of handoff performance on TCP Reno throughput. The mobile host performs handoff between **B2** and **B3** at fixed time intervals. We measure TCP throughput using `ttcp` by downloading 16 Mbytes of data from a correspondent host to a mobile host. Each data point is an average of 6 independent measurements.

The TCP throughput to the mobile host performing hard handoff is shown by the solid curve in Fig. 6. The throughput measured at zero handoff frequency (i.e., no handoffs) is marginally lower than the 1.6 Mbps achieved using standard IP routing in the same configuration. The difference between IP and Cellular IP forwarding is attributed to the fact that IP is implemented in the kernel and Cellular IP in user space. In addition, Cellular IP uses PCAP to forward packets which is not optimized for IP forwarding. We observe that the performance of TCP degrades as the hard handoff frequency increases due to packet loss. As the handoff rate increases TCP has less time to recover from loss. This forces TCP to operate below its optimal operational point resulting in a significant drop in transport performance as the handoff rate approaches once per second. Note that a mobile handing off every second is an aggressive handoff rate.

The next experiment investigates TCP improvement gains using semisoft handoff. The experimental conditions for the semisoft and hard handoff experiments are identical. The dashed curve for a 1-packet delay device shown in Fig. 6 shows the TCP throughput achieved by a mobile host that performs semisoft handoff at an increasing rate. From the figure we can observe that semisoft handoff reduces packet loss and significantly improves the transport throughput in relation to the hard handoff scheme.

Unlike the semisoft handoff experiment for UDP traffic, packet loss is not entirely eliminated with TCP. This can be observed in the decline in the measured throughput as the handoff frequency increases. We attribute the lack of synchronization and subsequent loss to the single buffer delay device used. Buffering packets is tied to the packet inter-arrival time, which is both shorter and more irregular in TCP streams than in the case of the UDP experiment. To introduce sufficient delay, we configure the semisoft delay device to support an 8-packet circular buffer. In Fig. 6 the dash curve for the 8-packet delay device shows performance results associated with using a larger buffer. We observe from the graph that packet loss is eliminated at higher handoff rates. A slight disturbance remains at handoff rates approaching one handoff per second due to the transmission delay variations encountered during handoff. The semisoft handoff results look very promising. Even at the highest handoff rate TCP throughput is almost identical to that of a stationary host, as shown in Fig. 6.

4.2. Scalability

The use of per mobile host routes in Cellular IP access networks naturally raises concerns about the ability of the protocol to scale to support higher throughput with very large numbers of mobiles hosts. As the number of active mobile hosts grows so will the routing tables in the access network. Routing cache needs to be efficiently searched for each data packet that gets forwarded by a base station. In the case of routing cache misses, the paging cache will be searched for the delivery of downlink packets. The routing cache will maintain mappings for packets that have been recently forwarded. The paging cache is therefore rarely accessed for these packets. Per host route lookup time in Cellular IP does not limit the number of users "connected" to the Cellular IP network. Rather the number of active users is limited. In this case Cellular IP networks can support an order of magnitude more users than other micromobility protocols that do not implement passive connectivity and paging.

To estimate the impact of different routing cache sizes on our user space Cellular IP implementation, we create random cache mappings and place them permanently into the routing cache. The solid line in Fig. 7 shows the base station throughput measured for a multihomed 300 MHz Pentium PC base station using `ttcp` and 1500 byte packets for different routing cache sizes. In this experiment we substitute a 100 Mbps Ethernet connection for a radio interface. The fact that the throughput curve hardly decreases with increasing routing cache size suggests that the performance bottleneck is not the cache lookup time. As shown in Fig. 7, the Cellular IP base station throughput is somewhat lower than the standard IP throughput. This is due to the additional packet processing involved with PCAP and additional packet copies that take place across kernel and user space domains. We note that the operation of routing cache is very similar to the self-learning operation of Ethernet switches, which can maintain tables of tens of thousands of entries at gigabit speeds. Our results indicate that Cellular IP software base stations are capable of supporting large numbers of mobile hosts and high aggregate throughput. We observe that per host routes can be supported without diminishing the performance of the base station implementation.

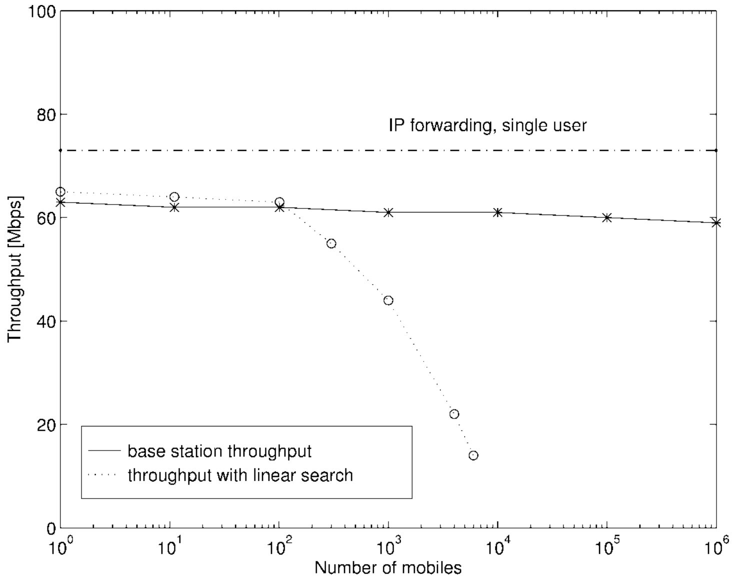


Fig. 7. Base station throughput.

5. Comparison of micromobility protocols

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 differences. 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 data base" that maps mobile host identifiers to location information. In this section, we present a performance comparison of Cellular IP (CIP) [23], Hawaii [11] and Hierarchical Mobile IP (HMIP) [27] based on the *Columbia IP Micromobility Software (CIMS)* [32] *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. We show that despite the apparent differences between these three protocols, the operational principles that govern them are largely similar. We show that the difference in handoff quality observed during simulation is related to the design of these protocols.

5.1. Protocols

The Hawaii protocol [11] from Lucent Technologies proposes a separate routing protocol to handle intra-domain mobility. Hawaii relies on Mobile IP to provide wide-area inter-domain mobility. A mobile host entering a new foreign agent domain is assigned a co-located care-of address. The mobile node retains its care-of address unchanged while moving within the foreign domain, thus the home agent 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 as 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.

The Hierarchical Mobile IP protocol [27] from Ericsson and Nokia employs a hierarchy of foreign agents to locally handle Mobile IP registration. In this protocol mobile hosts send Mobile IP registration messages (with appropriate extensions) to update their respective location information. Registration messages establish tunnels between neighboring foreign agents along the path from the mobile host to a gateway foreign agent. 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 foreign agents are connected to the gateway foreign agent (GFA). In this case, direct tunnels connect the gateway foreign agent to foreign agents that are located at access points. Paging extensions for Hierarchical Mobile IP are presented in [34] 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 home agents and is represented by paging areas. After receiving a packet addressed to a mobile host located in a foreign network, the home agent tunnels the packet to the paging foreign agent, which then pages the mobile host to re-establish 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 concept found in second generation cellular systems.

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.

5.2. Simulation model

The simulation study presented in this section uses the CIMS, which represents a micromobility extension for the *ns-2* network simulator based on version 2.1b6 [22]. 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 [32].

The Cellular IP simulation model is based on the latest description of the protocol [24]. 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 [26]. We used the unicast non-forwarding (UNF) and multiple stream forwarding (MSF) handoff schemes. Because Hawaii access points need to implement Mobile IP foreign agent 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 [5] functions that are 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 gateway foreign agent and foreign agents 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 foreign agents and encapsulating downlink packets based on the appropriate visitor list entry.

All simulations are performed using the network topology shown in Fig. 8. For Cellular IP simulations each `Wi` and `APi` corresponds to Cellular IP nodes where `W0` acts as a gateway to Internet. For Hawaii simulations all `Wi` and `APi` are Hawaii enabled routers and `W0` is the domain root router. When simulating Hierarchical Mobile IP, the gateway foreign agent function is implemented by `W0`. Nodes `W1` through `W5` represent mobility unaware routers, and `APi` nodes are routers with collocated foreign agents.

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 Mbps duplex link with 2 ms delay. Mobile hosts connect to access points (AP) using the ns-2 CSMA/CA wireless link model where each access point operates on a different frequency band. Simulation results were obtained using a single mobile host, continuously moving between access points 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.

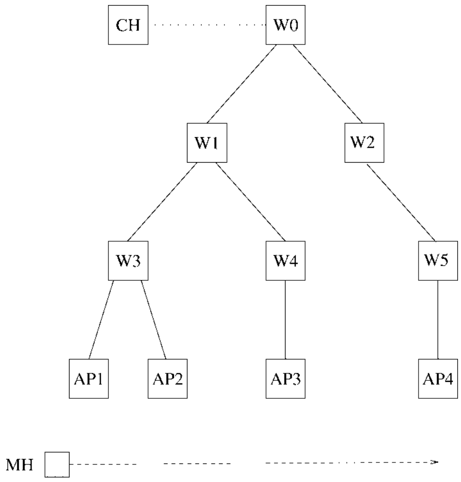


Fig. 8. Simulated network topology.

The simulation network accommodates 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 TCP Reno congestion control, except where stated otherwise.

5.3. *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 meters. 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 cross-over distances (i.e., the number of hops between the cross-over node and access point). The cross-over distance is 1, 2 or 3 hops when the mobile host moves between `AP1`-`AP2`, `AP2`-`AP3` and `AP3`-`AP4`, respectively. Figure 9 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 access points and the cross-over node. When the mobile host moves between access points `AP1` and `AP2` the delay is small. If the cross-over distance is larger then 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 cross-over node and new access point.

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

Next, we show simulation results for a TCP download during handoff. The dots shown in Fig. 10(a) correspond to sequence number of data packets associated with a single TCP connection, as seen by the mobile host. At 14.75 seconds 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.

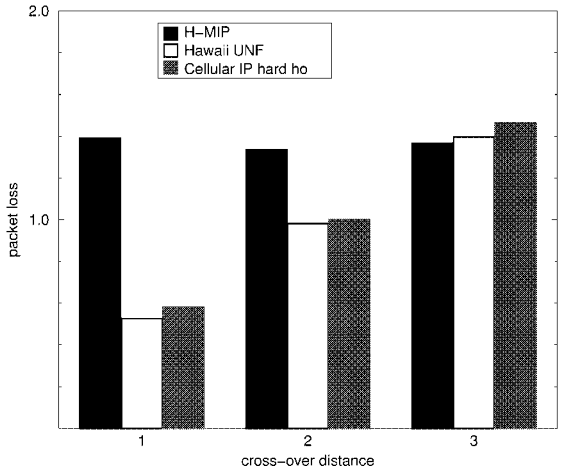
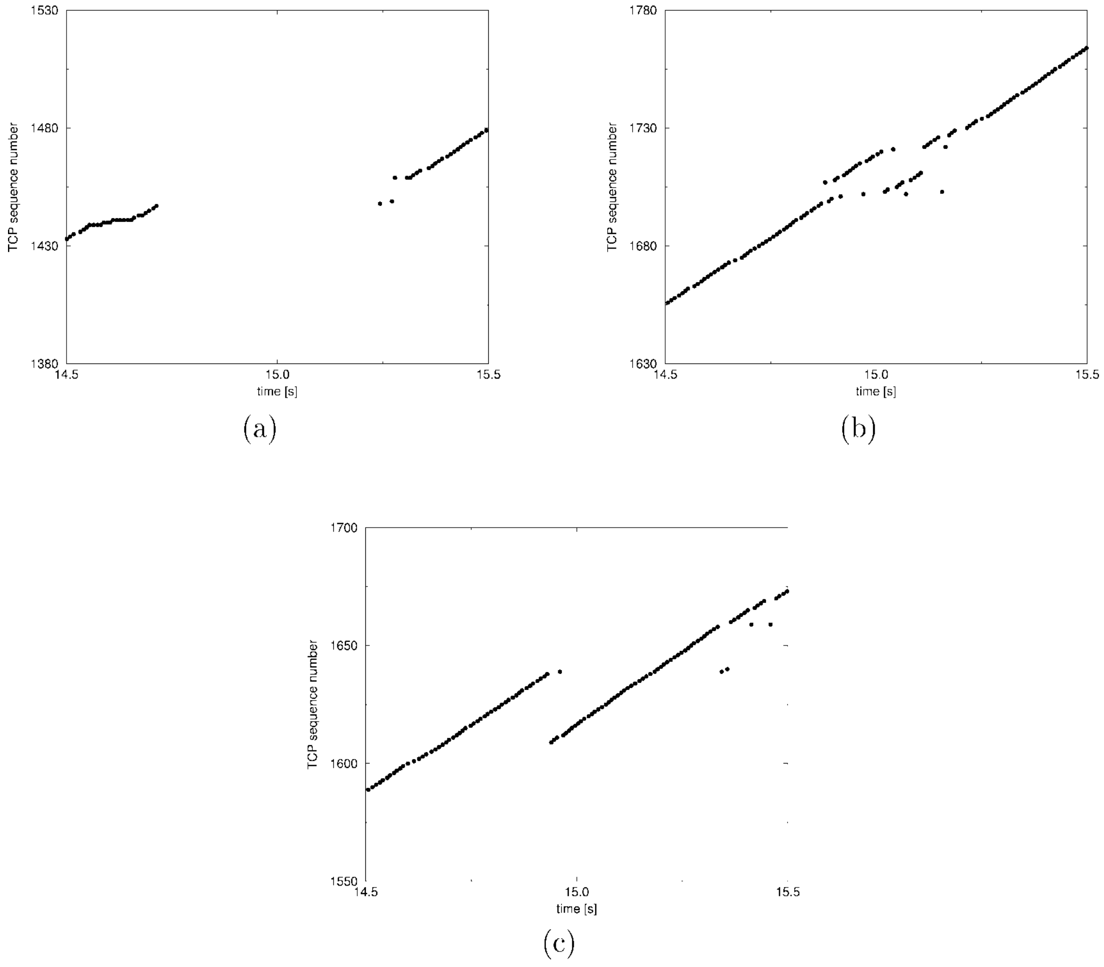


Fig. 9. UDP packet loss at handoff.



(a) (b)

(c)

Fig. 10. 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).

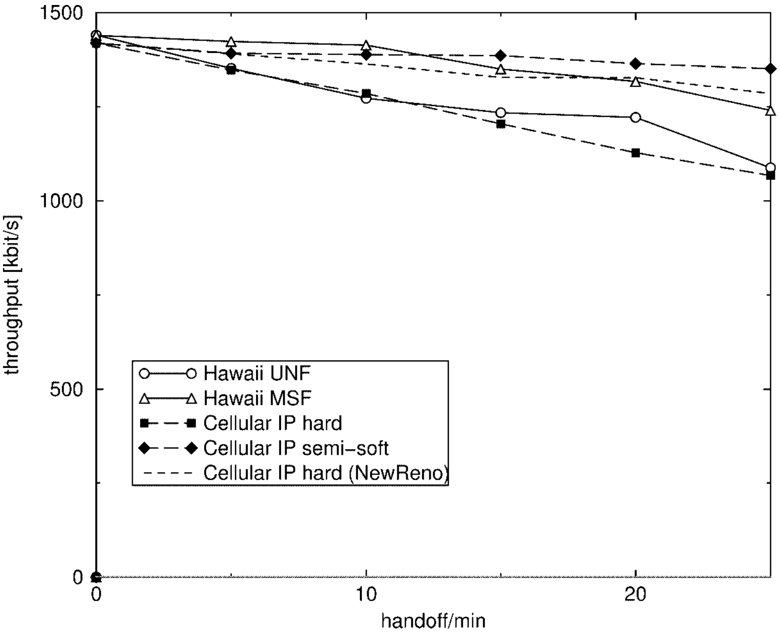


Fig. 11. Application level TCP throughput in case of periodic handoffs.

The degradation caused by packet loss increases with the increasing handoff frequency. This phenomenon is illustrated in Fig. 11 where 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% when the mobile host moves between access points every 2 seconds. This degradation would be more severe if we considered the potential processing delays that would be anticipated in a real system.

5.4. 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 cross-over node and creates downlink routing information along the new path. The operation is also similar in Hierarchical Mobile IP, but the cross-over node is always at the gateway foreign agent (node `W0` in the simulation network shown in Fig. 8), 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 cross-over point. In Hawaii the cross-over 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 Fig. 12. If a mobile host, initially attached to the network at `AP1`, moves between access points `AP2` to `AP8` then 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 cross-over 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.

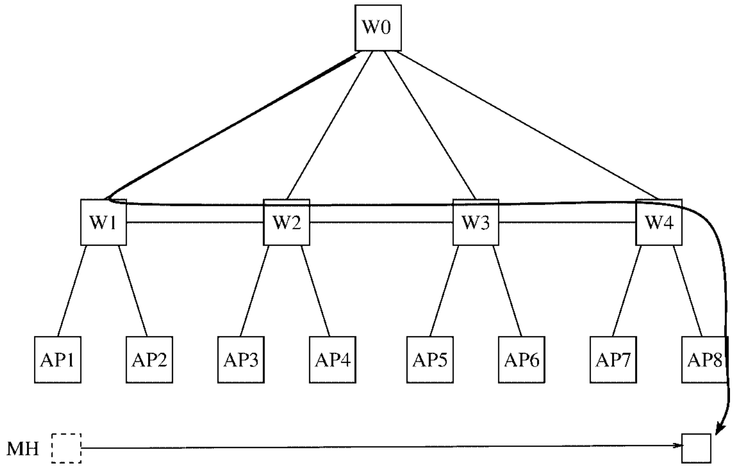


Fig. 12. Suboptimal routes after Hawaii handoffs.

This suboptimal routing problem represents a generic trade-off associated with handoff control signalling in micromobility protocols. If handoff control messages reach the gateway then nodes higher up in the hierarchy will have to deal with a potentially large numbers 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 cross-over node, nodes higher up the hierarchy do not have to process these messages hence minimizing the signalling load at those nodes.

In order for a cross-over node to be capable of discarding route update messages, the node must be aware that it is a cross-over node with respect to the particular handoff in progress. In Hawaii, for example, a node, which receives an update message referring to a mobile host that already has a valid entry, assumes it is the cross-over node. 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 assured 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 black-out 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 cross-over nodes 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 access point to explicitly determine the cross-over node.

Protocols that do not identify the cross-over node by either of the previously techniques have no ability to safely discard update messages before the gateway. Cellular IP represents 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 nodes identify themselves as cross-over nodes based on this property; or
- explicitly determine the cross-over node at handoff.

5.5. *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 be mostly attributed to a couple of design decisions. The first one is that the base Hierarchical Mobile IP protocol employs a single level node 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 cross-over nodes.

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. The Cellular IP semi-soft handoff [25] allows a mobile host to set up routing to the new access point prior to handoff. In this case, packets are duplicated at the cross-over 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 cross-over node and the new access point is shorter than the path between the cross-over node and the old access point then packets may still be lost. To overcome this problem, Cellular IP cross-over nodes delay packet duplicates for a fixed period amount of time (T_{ss}) before forwarding them toward the new access point, as discussed in Section 3.3.2. 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 [11]. Instead of setting up routing in advance of handoff, as is case with the 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 achieved, 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 the Cellular IP semi-soft handoff.

Cellular IP and Hawaii use two different approaches to improve handoff performance:

- bi-casting techniques and
- buffering and forwarding techniques.

The former prevents packet loss by taking pro-active steps that requires knowledge of the new access point prior to handoff. The latter does not rely on any such knowledge, but attempts 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 [28,29], and buffering and packet forwarding [30,31].

Figure 13 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 cross-over 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 performed Cellular IP semi-soft handoff or Hawaii MSF handoff. The solid and dashed lines correspond to probe traffic with packet inter-arrival times of 5 and 10 ms, respectively. In both cases the cross-over distance is 3 hops.

We can observe that in the case of the Hawaii MSF handoff (Fig. 13(b)) the lack of buffering ($T_{msf} = 0$) causes approximately 12 ms worth of data to be lost. This is similar to the performance of the Hawaii UNF handoff. Increasing the buffering time has the result of 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} then 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 the 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.

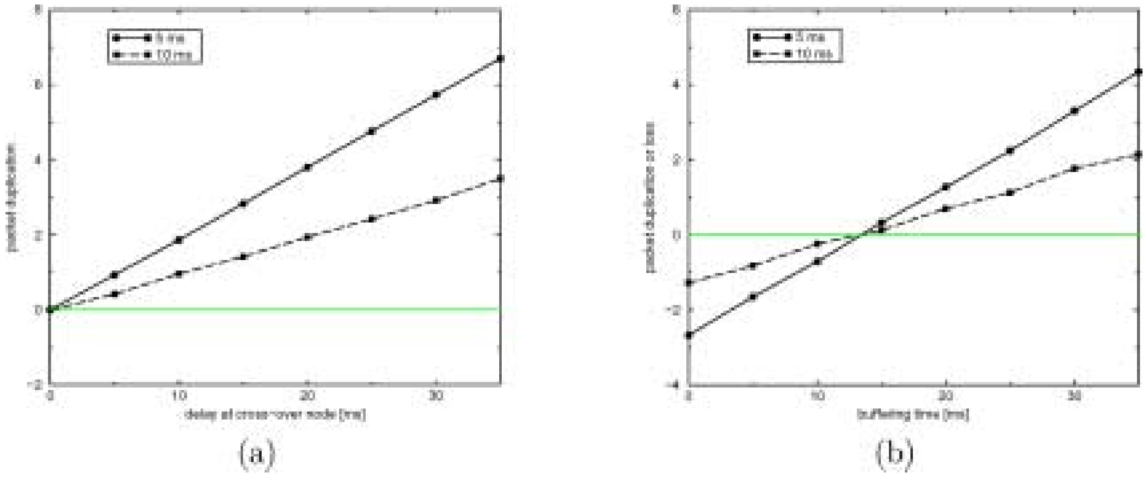


Fig. 13. UDP packet loss and duplication in case of (a) Cellular IP semi-soft handoff and (b) Hawaii MSF handoff. This is for packet interarrival times of 5 ms and 10 ms.

The same observations can be made in case of the Cellular IP semi-soft handoff, as shown in Fig. 13(a). One difference is that Cellular IP uses bi-casting instead of forwarding to recover packets and 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 cross-over node and the new and old access points differ then 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. 10(b) and 10(c). 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 cross-over node and the old access point. Even if the cross-over node delays packet duplicates by 50 ms as shown in Fig. 10(b), 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. 10(c) then 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 the larger one then 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 being that Hawaii’s forwarding scheme introduces packet re-ordering in addition to duplication. The effect of re-ordering is also visible in Fig. 11. The performance of the Hawaii MSF handoff, as seen by the application, is somewhat lower than that of the Cellular IP semi-soft handoff. This difference is because the TCP protocol reacts adversely to the level of packet re-ordering introduced by the Hawaii MSF scheme. Note that the parameters for these simulations were $T_{ss} = 120$ ms and $T_{msf} = 50$ ms.

Figure 11 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 specifically designed for TCP flows while micromobility protocols can reduce disruption experienced by widely used transport protocols (e.g., UDP, RTP, etc.) including TCP.

6. Conclusion

This paper presents the main results from the Cellular IP Project [33] at Columbia University. Cellular IP represents a new approach to IP host mobility that incorporates a number of important cellular system features such as passive connectivity, paging and seamless handoff. The Cellular IP routing, handoff, paging and security algorithms are simple and scalable resulting in the development of a highly scalable software base stations built using off-the-shelf PC hardware, operating systems and radios. Cellular IP complement Mobile IP with fast, seamless and local handoff control, and IP paging.

We have presented the design, implementation and evaluation of the Cellular IP protocol using simulation, analysis and experimentation. We have reported on the ability of Cellular IP to offer seamless mobility for TCP and UDP applications operating in highly mobile environments. The experimental testbed reported in this paper has shown that stronger control and management features can be built into commodity IP-based mobile networks without the need for costly and complex circuits. The source code for the Cellular IP testbed is available from the Web [33].

We also presented a comparison of a number of IP micromobility protocols that have been designed and implemented over the past several years. We developed the CIMS [32] ns-2 extensions that supports separate programming models for Cellular IP, Hawaii and Hierarchical Mobile IP. We presented a set of simulation results to illustrate the performance of these protocols. We compared the performance of the Cellular IP semi-soft handoff and the Hawaii MSF handoff identifying a number of similarities and differences. We also discussed a number of differences that are not directly related to handoff quality. The CIMS source code is also publicly available [32].

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 [23]. However, the differentiated services concepts such as aggregation, per-hop behavior, service level agreement and slow time scale 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 to use traffic engineering techniques that promote under utilization 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 that have been discussed over the last several years see [1]. 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.

References

[1] A. Campbell and J. Gomez, IP micromobility protocols, *ACM SIAMOBILE Mobile Computer and Communications Review (MC2R)* **4**(4) (2001), 45–54.

[2] A.G. Valkó, Cellular IP: a new approach to Internet host mobility, *ACM Computer Communication Review*, January 1999.

[3] A. Campbell, J. Gomez, C.-Y. Wan, S. Kim, Z. Turányi and A. Valkó, Cellular IP, Internet Draft, draft-ietf-mobileip-cellularip-00.txt, December 1999, Work in Progress.

[4] C. Perkins, ed., IP mobility support, Internet RFC 2002, October 1996.

[5] D.B. Johnson and C. Perkins, Route optimization in mobile IP, Internet Draft, draft-ietf-mobileip-optim-07.txt, November 1998, Work in Progress.

[6] S. Blake, D. Black, M. Carlson, E. Davis, Z. Wang and W. Weiss, An architecture for differentiated services, Internet RFC 2475, December 1998.

[7] S. McCanne and V. Jacobson, The BSD packet filter: a new architecture for user-level packet capture, *USENIX 93*, San Diego, 1993.

[8] *WaveLAN Air Interface*, Data Manual, AT&T Corporation, Doc. No. 407-0024785 Rev. 2 (draft), July 11, 1995.

[9] M. Mouly and M.-B. Pautet, The GSM system for mobile communications, published by the authors, ISBN 2-9507190-0-7, 1992.

[10] E. Gustafsson, A. Jonsson and C. Perkins, Mobile IP regional tunnel management, Internet Draft, draft-ietf-mobileip-reg-tunnel-01.txt, Internet Draft, August 1999, Work in Progress.

[11] R. Ramjee, T. La Porta, S. Thuel, K. Varadhan and S. Y. Wang, HAWAII: a domain-based approach for supporting mobility in wide-area wireless networks, in: *Proc. IEEE International Conference on Network Protocols*, 1999.

[12] S.F. Foo and K.C. Chua, Regional Aware Foreign Agent (RAFA) for fast local handoffs, Internet Draft, draft-chuafoo-mobileip-rafa-00.txt, November 1998, Work in Progress.

[13] R. Ramjee, T. La Porta and L. Li, Paging support for IP mobility using HAWAII, Internet Draft, draft-ietf-mobileip-paging-hawaii-00, June 1999, Work in Progress.

[14] R. Caceres and V.N. Padmanabhan, Fast and scalable handoffs for wireless internetworks, in: *Proc. ACM Mobicom*, 1996.

[15] C. Castelluccia and L. Bellier, Toward a unified hierarchical mobility management framework, Internet Draft, draft-castelluccia-uhmm-framework-00.txt, June 1999, Work in Progress.

[16] M.C. Chuah and Y. Li, Distributed registration extension to mobile IP, Internet Draft, draft-chuahli-mobileip-dremip-00.txt, October 1997, Work in Progress.

[17] Mobile IP Authentication, Authorization, and Accounting Requirements, Internet Draft, draft-ietf-mobileip-aaa-reqs-03.txt, March 2000, Work in Progress.

[18] J.W. Lockwood, Implementation of campus-wide wireless network services using ATM, virtual LANs and wireless basestations, *Wireless Communications and Networking Conference (WCNC '99)*, New Orleans, LA, September 1999.

[19] H. Balakrishnan, S. Seshan and R. Katz, Improving reliable transport and handoff performance in cellular wireless networks, *ACM Wireless Networks* **1**(4) (1995).

[20] J. Ioannidis, D. Duchamp and G.Q. Maguire, Jr., IP-based protocols for mobile Internetworking, in: *Proc. ACM Sigcomm '91*, 1991, pp. 234–245.

[21] A. Campbell, J. Gomez, C.-Y. Wan, S. Kim, Z. Turányi and A. Valkó, Performance of Cellular IP access networks, Technical Report, Center for Telecommunications Research, Columbia University, January 1999.

[22] Network Simulator, http://www-mash.cs.berkeley.edu/ns/ns.html.

[23] A.T. Campbell, M. Barry and L.-H. Sun, Supporting service differentiation in wireless packet networks using distributed control, *IEEE Journal of Selected Areas in Communications (JSAC)*, Special Issue on Mobility and Resource Management in Next-Generation Wireless Systems **19**(10) (2001), 2094–2104.

[24] A. Campbell, J. Gomez, C.-Y. Wan, S. Kim, Z. Turányi and A. Valkó, Cellular IP, Internet Draft, draft-ietf-mobileip-cellularip-00, December 1999, Work in Progress.

[25] A. Campbell, J. Gomez, C.-Y. Wan, S. Kim, Z. Turányi and A. Valkó, Design and performance of Cellular IP access networks, *IEEE Personal Communications Magazine*, special issue on IP-based Mobile Telecommunications Networks, 2000.

[26] R. Ramjee, T. La Porta, S. Thuel, K. Varadhan and L. Salgarelli, IP micromobility support using HAWAII, Internet Draft, draft-ietf-mobileip-hawaii-00, June 1999, Work in Progress.

[27] E. Gustafsson, A. Jonsson and C. Perkins, Mobile IP regional registration, Internet Draft, draft-ietf-mobileip-reg-tunnel-03, July 2000, Work in Progress.

[28] K. El-Malki and H. Soliman, Fast handoffs in mobile IPv4, Internet Draft, draft-elmalki-mobileip-fast-handoffs-03, September 2000, Work in Progress.

[29] P. Calhoun et al., Foreign agent assisted hand-off, Internet Draft, draft-calhoun-mobileip-proactive-fa-03, November 2000, Work in Progress.

[30] C. Perkins, G. Krishnamurthi and R. Chalmers, Buffer management for smooth handovers in mobile IPv6, Internet Draft, draft-krishnamurthi-mobileip-buffer6-00, July 2000, Work in Progress.

[31] A. O'Neill, S. Corson and G. Tsirtsis, Generalized IP handoff, Internet Draft, draft-oneill-craps-handoff-00, August 2000, Work in Progress.

[32] Columbia IP Micromobility Software (CIMS) home page, http://comet.columbia.edu/micromobility.

[33] Cellular IP home page, http://comet.columbia.edu/cellularip.

[34] H. Haverinen and J. Malinen, Mobile IP regional paging, Internet Draft, draft-haverinen-mobileip-reg-paging-00.txt, June 2000, Work in Progress.