MOBILE NETWORKING THROUGH MOBILE IP

CHARLES E. PERKINS
Sun Microsystems

Mobile IP is a proposed standard protocol that builds on the Internet Protocol by making mobility transparent to applications and higher level protocols like TCP.



Although the Internet offers access to information sources worldwide, typically we do not expect to benefit from that access until we arrive at some familiar point—whether home, office, or school. However, the increasing variety of wireless devices offering IP connectivity, such as PDAs, handhelds, and digital cellular phones, is beginning to change our perceptions of the Internet.

To understand the contrast between the current realities of IP connectivity and future possibilities, consider the transition toward mobility that has occurred in telephony over the past 20 years. An analogous transition in the domain of networking, from dependence on fixed points of attachment to the flexibility afforded by mobility, has just begun.

Mobile computing and networking should not be confused with the portable computing and networking we have today. In mobile networking, computing activities are not disrupted when the user changes the computer's point of attachment to the Internet. Instead, all the needed reconnection occurs automatically and noninteractively.

Truly mobile computing offers many advantages. Confident access to the Internet anytime, anywhere will help free us from the ties that bind us to our desktops. Consider how cellular phones have given people new freedom in carrying out their work. Taking along an entire computing environment has the potential not just to extend that flexibility but to fundamentally change the existing work ethic. Having the Internet available to us as we move will give us the tools to build new computing environments wherever we go. Those who have little interest in mobility per se will still benefit from the ability to resume previous applications when they reconnect. This is especially convenient in a wireless LAN office environment, where the boundaries between attachment points are not sharp and are often invisible.

The evolution of mobile networking will differ from that of telephony in some important respects. The endpoints of a telephone connection are typically human; computer applications are likely to involve interactions between machines without human intervention. Obvious examples of this are mobile computing devices on airplanes, ships, and automobiles. Mobile networking may well also come to depend on position-finding devices, such as a satellite global positioning system, to work in tandem with wireless access to the Internet.

Another difference may well be rate of adoption. It took many years for mobile phones to become cheap and lightweight enough to be perceived as convenient. Because wireless mobile computing devices such as PDAs and pocket organizers have already found user acceptance, mobile computing may become popular much more quickly.

However, there are still some technical obstacles that must be overcome before mobile networking can become widespread. The most fundamental is the way the Internet Protocol, the protocol that connects the networks of today's Internet, routes packets to their destinations according to IP

1089-7801/98/$10.00 ©1998 IEEE

addresses. These addresses are associated with a fixed network location much as a nonmobile phone number is associated with a physical jack in a wall. When the packet's destination is a mobile node, this means that each new point of attachment made by the node is associated with a new network number and, hence, a new IP address, making transparent mobility impossible.

Mobile IP (RFC 2002),[1] a standard proposed by a working group within the Internet Engineering Task Force, was designed to solve this problem by allowing the mobile node to use two IP addresses: a fixed home address and a care-of address that changes at each new point of attachment. This article will present the Mobile IP standard in moderate technical detail and point the reader toward a wealth of further information.[2,3] In addition, readers can go to the sidebar Mobile IP Web Resources in this issue's *IC Online* at http://computer.org/internet/ for a convenient set of hyperlinked resources.

I also describe how Mobile IP will change with IP version 6,[4,5] the product of a major effort within the IETF to engineer an eventual replacement for the current version of IP.[6] Although IPv6 will support mobility to a greater degree than IPv4, it will still need Mobile IP to make mobility transparent to applications and higher level protocols such as TCP.

There is a great deal of interest in mobile computing and apparently in Mobile IP as a way to provide for it. A quick Web search for items related to Mobile IP returned over 60,000 hits—impressive even given the notorious lack of selectivity for such procedures. Mobile IP forms the basis either directly or indirectly of many current research efforts and products. The Cellular Digital Packet Data (CDPD),[7] for example, has created a widely deployed communications infrastructure based on a previous draft specification of the protocol. In addition, most major router vendors have developed implementations for Mobile IP.

The outlook for Mobile IP in the complex Internet marketplace is far from clear, and some technical problems remain, security being the most important. However, once the security solutions are solid, nomadic users may finally begin to enjoy the convenience of seamless untethered roaming and effective application transparency that is the promise of Mobile IP.

HOW MOBILE IP WORKS

IP routes packets from a source endpoint to a destination by allowing routers to forward packets from incoming network interfaces to outbound interfaces according to routing tables. The routing tables typically maintain the next-hop (outbound interface) information for each destination IP address, according to the number of networks to which that IP address is connected. The network number is derived from the IP address by masking off some of the low-order bits. Thus, the IP address typically carries with it information that specifies the IP node's point of attachment.

To maintain existing transport-layer connections (see the sidebar "Nomadicity: How Mobility Will Affect the Protocol Stack" on the next pages) as the mobile node moves from place to place, it must keep its IP address the same. In TCP (which accounts for the overwhelming majority of Internet connections), connections are indexed by a quadruplet that contains the IP addresses and port numbers of both connection endpoints. Changing any of these four numbers will cause the connection to be disrupted and lost. On the other hand, correct delivery of packets to the mobile node's current point of attachment depends on the network number contained within the mobile node's IP address, which changes at new points of attachment. To change the routing requires a new IP address associated with the new point of attachment.

Mobile IP uses two IP addresses: a fixed home address and a care-of address that changes at each new point of attachment.

Mobile IP has been designed to solve this problem by allowing the *mobile node* to use two IP addresses (see the sidebar "Mobile Networking Terminology" for definitions of italicized terms). In Mobile IP, the *home address* is static and is used, for instance, to identify TCP connections. The *care-of address* changes at each new point of attachment and can be thought of as the mobile node's topologically significant address; it indicates the network number and thus identifies the mobile node's point of attachment with respect to the network topology. The home address makes it appear that the mobile node is continually able to receive data on its *home network*, where Mobile IP requires the existence of a network node known as the *home agent*. Whenever the mobile node is not attached to its home network (and is therefore attached to what is termed a *foreign network*), the home agent gets all the packets destined for the mobile node and arranges to deliver them to the mobile node's current point of attachment.

Whenever the mobile node moves, it *registers* its new care-of address with its home agent. To get a packet to a mobile node from its home network, the home agent delivers the packet from the home network to the care-of address. The further delivery requires that the packet be modified so that the care-of address appears as the destination IP address. This modification can be understood as a packet transformation or, more specifically, a *redirection*. When the packet arrives at the care-of address, the reverse transformation is applied so that the packet once again appears to have the mobile node's home address as the destination IP address. When the packet arrives at the mobile node, addressed to the home address, it will be processed properly by TCP or whatever higher level protocol logically receives it from the mobile node's IP (that

NOMADICITY: HOW MOBILITY WILL AFFECT THE PROTOCOL STACK

Mobile IP is a large part, but by no means the only part, of the story of mobile computing and networking. To see Mobile IP in its true place requires an understanding of the relationships between the various layers of network protocols. Each layer should present a clear model of operation to the architect. Once the model is identified, the effects of mobility can be studied in relation to it. *Nomadicity* is the name used by the Cross-Industry Working Team (XIWT) at the Corporation for National Research Initiatives (CNRI) to denote an architecture for the entire mobile computing environment.[1]

Figure A is a simplified view of the International Standards Organization's protocol stack as it applies to Internet networking. The major goal of Mobile IP protocol design was to handle mobility at the network layer and to leave transport and other higher layers unaffected, so that the existing routing infrastructure, nonmobile hosts, and current applications would not be required to change.

Protocol layer two, the *data link* layer, is responsible for link establishment and maintenance. Thus, physical effects from mobility are likely to require changes in the layer-two protocols. Changes in position affect the signal-to-interference ratio (SIR). Link layers that adapt forward error correction to SIR can exhibit variable bandwidth but far fewer lost packets. Wireless media typically introduce many other design requirements at layer two. In particular, the desire for confidentiality leads to the incorporation of encryption techniques, especially for wireless links. Often, lower bandwidth (compared with wired media) suggests the use of compression techniques. And, typically, transmitting a signal causes the local receiver to lose detection of any other signal because of the great difference in effective power levels between local and remote transmitters. Thus, collision-detection techniques, such as those used with Ethernet, must be replaced by less reliable collision-avoidance measures and careful etiquette.

Networking Layers	Standard Protocols
Applications	HTTP, NFS, SNMP, DNS, Telnet, FTP, ...
Window Mgr	
Sockets	
Transport	TCP, UDP, RTP
Network	IP, ICMP, IGMP, IPSec, Mobile IP ... (IPX, Appletalk)
Data Link	IEEE 802.*, PPP
Physical	Network adapter

Figure A. The Internet networking stack showing common protocols associated with each layer.

Other distinguishing characteristics of wireless communications media include the difficulty of establishing a precise range (cell size) for connectivity to the medium, and the ability for separate stations to use the media without interference. This latter property of reuse depends upon avoiding interference between neighboring transmitters, and a great engineering discipline has been built up to understand optimal placement of such wireless equipment as base stations. To reuse the physical wireless medium to the fullest extent, the cell size should be as small as possible. This means that as demand for wireless communications increases, cell sizes will decrease, and the frequency with which mobile computers will switch cells (change their point of attachment to the Internet) will grow correspondingly.

The Internet Protocol is at layer three, the *network* layer. IP selects routes (determines paths) through a loosely confederated association of independent network links. IP offers routing from one network to another, in addition to some minor services such as fragmentation and reassembly, and checksumming. Moving from one place to another can be modeled as changing the network node's point of attachment to the Internet. Supporting mobility at this layer is therefore naturally modeled as changing the routing of datagrams destined for the mobile node so that they arrive at the new point of attachment. This turns out to be a very convenient choice, and was the option chosen by the Mobile IP working group.

At the *transport* layer, TCP (RFC 793)[2] and other transport protocols attempt to offer a more convenient abstraction for data services than the characteristically chaotic stream of data

is, layer 3) processing layer. More information on the abstract modeling as a way to perform layer 3 redirection on packets can be found in Bhagwat, Perkins, and Tripathi.[8]

In Mobile IP the home agent redirects packets from the home network to the care-of address by constructing a new IP header that contains the mobile node's care-of address as the destination IP address. This new header then shields or encapsulates the original packet, causing the mobile node's home address to have no effect on the encapsulated packet's routing until it arrives at the care-of address. Such *encapsulation* is also called *tunneling*, which suggests that the packet burrows through the Internet, bypassing the usual effects of IP routing.

Mobile IP, then, is best understood as the cooperation of three separable mechanisms:

- Discovering the care-of address;
- Registering the care-of address;
- Tunneling to the care-of address.

emanating from IP. The vagaries and time dependencies of routers and Internet congestion often cause datagrams to be delivered out of order, duplicated, or even dropped entirely before reaching their destination. TCP attempts to solve those problems, but offers little help in supplying a steady (constant bandwidth) stream of data, or in delivering data within specified time bounds. Over time, TCP has been modified to treat dropped packets as an indication of network congestion, and therefore to throttle transmissions as soon as a lost packet is detected (by managing sequence numbers).[3] This is the wrong strategy when packets are corrupted by transmission over a noisy wireless channel, because for such packets immediate retransmission is much better than delayed retransmission. Ways to change this behavior are still under debate.

At the top layer are the application protocols. Depending on the transport model employed, application protocols are largely freed from much of the drudgery of error correction, retransmission, flow control, and the like. However, mobility creates new needs at the application layer, which require additional protocol support: automatic configuration, service discovery, link awareness, and environment awareness.

These protocol support mechanisms form a set of middleware services. For example, a mobile computer might need to be reconfigured differently at each different point of attachment. Among other things, a new DNS server, IP address, link MTU, and default router may be required. These configuration items are usually thought of as being worked out at setup time for desktop systems, but for mobile computers no single answer can be sufficient. Recent deployment of the Dynamic Host Configuration Protocol (RFCs 2131, 2132)[4,5] goes some way toward resolving configuration difficulties, but is not the whole answer. Discovering services can be modeled as a requirement for automatic configuration, but is more naturally useful when services are located upon demand and according to the needs of application protocols. This need is just now being met by the Service Location Protocol (RFC 2165).[6]

One of the more challenging aspects of architecting such middleware lies in offering applications the opportunity to detect the state of the physical link, which changes dynamically and can easily affect the application's desired operation. The simplest example is the need for Web applications to adjust their presentation of graphical data depending on the available end-to-end bandwidth. Today that bandwidth is largely constrained by the link conditions at the endpoints and the congestion status of infrastructure connectivity. Mobile computers introduce more variability into this mix and thus exacerbate the growing need for multimedia applications to detect and act on dynamic connection parameters, such as link bandwidth, error rate, and round-trip times. Other logical parameters, such as cost and security, may eventually exhibit similar dynamic behavior and further complicate application response to connection status information.

Lastly, a word should be said about the granularity of protocol response to node movement. Today's typical user must be content with portable computing, which requires reinitializing and reestablishing connections at each new point of attachment to the Internet. However, acceptance of this mode of operation may well evaporate if the reinitialization process has to be performed a lot more frequently. Left unchecked, the expected decreases in cell sizes will require exactly that in the future. The existing methods typify portable network computing, which means establishing the availability of network computing when one arrives at a new point of attachment but being unable to continue previous computing activities. The point of Mobile IP, DHCP and similar protocols is to provide completely automatic, noninteractive reconnection to those activities.

REFERENCES

1. Corporation for National Research Initiatives. XIWT: Cross-Industry Working Team, 1994, http://www.cnri.reston.va.us:3000/XIWT/public.html.
2. "Transmission Control Protocol," J. B. Postel, ed., RFC 793, Sept. 1981.
3. W. Stevens, "TCP Sow Start, Congestion Avoidance, Fast Retransmit, and Fast Recovery Algorithms," RFC 2001, Jan. 1997.
4. R. Droms, "Dynamic Host Configuration Protocol," RFC 2131, Mar. 1997, ftp://ds.internic.net/rfc/rfc2131.txt.
5. S. Alexander and R. Droms, "DHCP Options and BOOTP Vendor Extensions," RFC 2132, Mar. 1997.
6. J. Veizades, et al., "Service Location Protocol," RFC 2165, July 1997.

Discovering the Care-of Address

The Mobile IP *discovery* process has been built on top of an existing standard protocol, Router Advertisement, specified in RFC 1256.[9] Mobile IP discovery does not modify the original fields of existing router advertisements but simply extends them to associate mobility functions. Thus, a router advertisement can carry information about default routers, just as before, and in addition carry further information about one or more care-of addresses. When the router advertisements are extended to also contain the needed care-of address, they are known as *agent advertisements*. Home agents and foreign agents typically broadcast agent advertisements at regular intervals (for example, once a second or once every few seconds). If a mobile node needs to get a care-of address and does not wish to wait for the periodic advertisement, the mobile node can broadcast or multicast a solicitation that will be answered by any foreign agent or home agent that receives it.

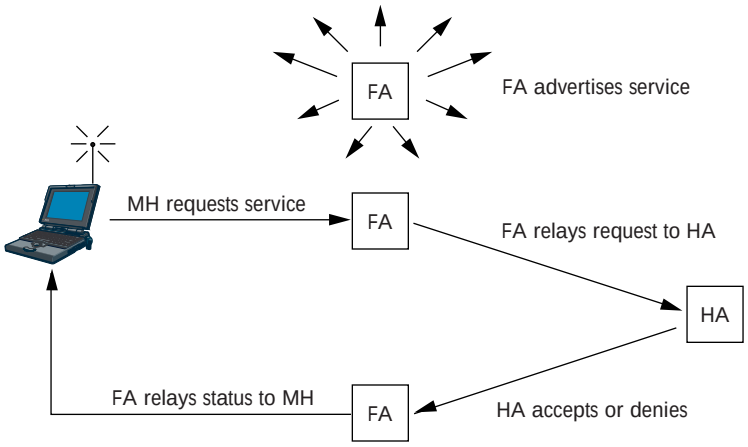

Figure 1. Registration operations in Mobile IP. FA is foreign agent, HA is home address, and MH is mobile host.

Home agents use agent advertisements to make themselves known, even if they do not offer any care-of addresses. However, it is not possible to associate preferences to the various care-of addresses in the router advertisement, as is the case with default routers. The IETF working group was concerned that dynamic preference values might destabilize the operation of Mobile IP. Because no one could defend static preference assignments except for backup mobility agents, which do not help distribute the routing load, the group eventually decided not to use the preference assignments with the care-of address list.

Thus, an agent advertisement performs the following functions:

- allows for the detection of mobility agents;
- lists one or more available care-of addresses;
- informs the mobile node about special features provided by foreign agents, for example, alternative encapsulation techniques;
- lets mobile nodes determine the network number and status of their link to the Internet; and
- lets the mobile node know whether the agent is a home agent, a foreign agent, or both, and therefore whether it is on its home network or a foreign network.

Mobile nodes use router solicitations as defined in RFC 1256 to detect any change in the set of mobility agents available at the current point of attachment. (In Mobile IP this is then termed *agent solicitation*.) If advertisements are no longer detectable from a foreign agent that previously had offered a care-of address to the mobile node, the mobile node should presume that foreign agent is no longer within range of the mobile node's network interface. In this situation, the mobile node should begin to hunt for a new care-of address, or possibly use a care-of address known from advertisements it is still receiving. The mobile node may choose to wait for another advertisement if it has not received any recently advertised care-of addresses, or it may send an agent solicitation.

Registering the Care-of Address

Once a mobile node has a care-of address, its home agent must find out about it. Figure 1 shows the *registration* process defined by Mobile IP for this purpose. The process begins when the mobile node, possibly with the assistance of a foreign agent, sends a registration request with the care-of address information. When the home agent receives this request, it (typically) adds the necessary information to its routing table, approves the request, and sends a registration reply back to the mobile node. Although the home agent is not required by the Mobile IP protocol to handle registration requests by updating entries in its routing table, doing so offers a natural implementation strategy, and all implementations I know of take this approach.

Authentication. Registration requests contain parameters and flags that characterize the tunnel through which the home agent will deliver packets to the care-of address. Tunnels can be constructed in various ways, described briefly in the next section.[10,11] When a home agent accepts the request, it begins to associate the home address of the mobile node with the care-of address, and maintains this association until the *registration lifetime* expires. The triplet that contains the home address, care-of address, and registration lifetime is called a *binding* for the mobile node. A registration request can be considered a *binding update* sent by the mobile node.

A binding update is an example of a *remote redirect*, because it is sent remotely to the home agent to affect the home agent's routing table. This view of registration makes the need for authentication very clear.[12] The home agent must be certain registration was originated by the mobile node and not by some other malicious node pretending to be the mobile node. A malicious node could cause the home agent to alter its routing table with erroneous care-of address information, and the mobile node would be unreachable to all incoming communications from the Internet.

The need to authenticate registration information has played a major role in determining the acceptable design parameters for Mobile IP. Each mobile node and home agent

must share a security association and be able to use Message Digest 5 (RFC 1321) with 128-bit keys to create unforgeable digital signatures for registration requests.[13] The signature is computed by performing MD5's one-way hash algorithm over all the data within the registration message header and the extensions that precede the signature.

To secure the registration request, each request must contain unique data so that two different registrations will in practical terms never have the same MD5 hash. Otherwise, the protocol would be susceptible to *replay attacks*, in which a malicious node could record valid registrations for later replay, effectively disrupting the ability of the home agent to tunnel to the current care-of address of the mobile node at that later time. To ensure this does not happen, Mobile IP includes within the registration message a special identification field that changes with every new registration. The exact semantics of the identification field depend on several details, which are described at greater length in the protocol specification.[1] Briefly, there are two main ways to make the identification field unique.

One is to use a timestamp; then each new registration will have a later timestamp and thus differ from previous registrations. The other is to cause the identification to be a pseudorandom number; with enough bits of randomness, it is highly unlikely that two independently chosen values for the identification field will be the same. When randomness is used, Mobile IP defines a method that protects both the registration request and reply from replay, and calls for 32 bits of randomness in the identification field. If the mobile node and the home agent get too far out of synchronization for the use of timestamps, or if they lose track of the expected random numbers, the home agent will reject the registration request and include information to allow resynchronization within the reply. Using random numbers instead of timestamps avoids problems stemming from attacks on the NTP protocol that might cause the mobile node to lose time synchronization with the home agent or to issue authenticated registration requests for some future time that could be used by a malicious node to subvert a future registration.

The identification field is also used by the foreign agent to match pending registration requests to registration replies when they arrive at the home agent and to subsequently be able to relay the reply to the mobile node. The foreign agent also stores other information for pending registrations, including the mobile node's home address, the mobile node's Media Access Layer (MAC) address, the source port number for the registration request from the mobile node, the registration lifetime proposed by the mobile node, and the home agent's address. The foreign agent can limit registration lifetimes to a configurable value that it puts into its agent advertisements. The home agent can reduce the registration lifetime, which it includes as part of the registration reply, but it can never increase it.

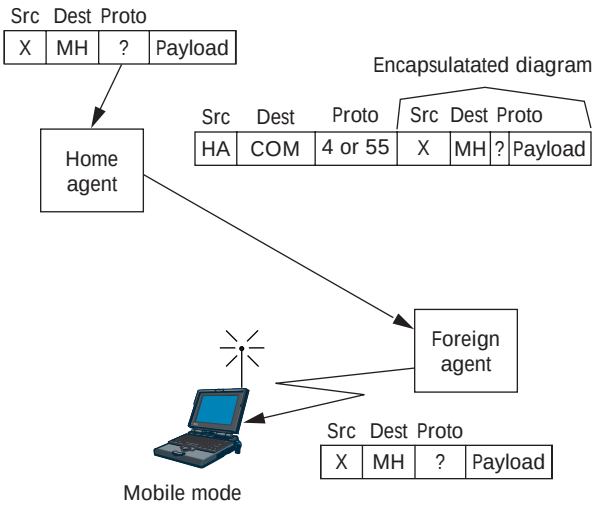

Figure 2. Tunneling operations in Mobile IP.

As Figure 1 shows, in Mobile IP foreign agents are mostly passive, relaying registration requests and replies back and forth between the home agent and the mobile node, doing mostly what they are told. The foreign agent also decapsulates traffic from the home agent and forwards it to the mobile node. Note that foreign agents do not have to authenticate themselves to the mobile node or home agent. A bogus foreign agent could impersonate a real foreign agent simply by following protocol and offering agent advertisements to the mobile node. The bogus agent could, for instance, then refuse to forward decapsulated packets to the mobile node when they were received. However, the result is no worse than if any node were tricked into using the wrong default router, which is possible using unauthenticated router advertisements as specified in RFC 1256.[9]

Automatic home agent discovery. When the mobile node cannot contact its home agent, Mobile IP has a mechanism that lets the mobile node try to register with another unknown home agent on its home network. This method of *automatic home agent discovery* works by using a broadcast IP address instead of the home agent's IP address as the target for the registration request. When the broadcast packet gets to the home network, other home agents on the network will send a rejection to the mobile node; however, their rejection notice will contain their address for the mobile node to use in a freshly attempted registration message. Note that the broadcast is not an Internet-wide broadcast, but a *directed* broadcast that reaches only IP nodes on the home network.

Tunneling to the Care-of Address

Figure 2 shows the tunneling operations in Mobile IP. The default encapsulation mechanism that must be supported

by all mobility agents using Mobile IP is IP-within-IP.[10] Using IP-within-IP, the home agent, the *tunnel source*, inserts a new IP header, or *tunnel header*, in front of the IP header of any datagram addressed to the mobile node's home address. The new tunnel header uses the mobile node's care-of address as the destination IP address, or *tunnel destination*. The tunnel source IP address is the home agent, and the tunnel header uses 4 as the higher level protocol number, indicating that the next protocol header is again an IP header. In IP-within-IP the entire original IP header is preserved as the first part of the payload of the tunnel header. Therefore, to recover the original packet, the foreign agent merely has to eliminate the tunnel header and deliver the rest to the mobile node.

Figure 2 shows that sometimes the tunnel header uses protocol number 55 as the inner header. This happens when the home agent uses *minimal encapsulation*[11] instead of IP-within-IP. Processing for the minimal encapsulation header is slightly more complicated than that for IP-within-IP, because some of the information from the tunnel header is combined with the information in the inner minimal encapsulation header to reconstitute the original IP header. On the other hand, header overhead is reduced.

CHANGES WITH IP VERSION 6

How will Mobile IP change when IP version 6 is adopted? IPv6 includes many features for streamlining mobility support that are missing in IP version 4 (current version), including Stateless Address Autoconfiguration[14] and Neighbor Discovery.[15] IPv6 also attempts to drastically simplify the process of renumbering, which could be critical to the future routability of the Internet.[16] Because the number of mobile computers accessing the Internet will likely increase, efficient support for mobility will make a decisive difference in the Internet's future performance. This, along with the growing importance of the Internet and the Web, indicates the need to pay attention to supporting mobility.[17]

As proposed by the Mobile IP working group, mobility support in IPv6 follows the design for Mobile IPv4, using encapsulation to deliver packets from the home network to the mobile point of attachment.

Mobility Support in IPv6,[18] as proposed by the Mobile IP working group, follows the design for Mobile IPv4. It retains the ideas of a home network, home agent, and the use of encapsulation to deliver packets from the home network to the mobile node's current point of attachment. While discovery of a care-of address is still required, a mobile node can configure its a care-of address by using Stateless Address Autoconfiguration and Neighbor Discovery. Thus, foreign agents are not required to support mobility in IPv6. IPv6-within-IPv6 tunneling is also already specified.[19]

Route Optimization

IPv6 mobility borrows heavily from the route optimization ideas specified for IPv4,[20] particularly the idea of delivering binding updates directly to correspondent nodes. When it knows the mobile node's current care-of address, a correspondent node can deliver packets directly to the mobile node's home address without any assistance from the home agent. Route optimization is likely to dramatically improve performance for IPv6 mobile nodes. It is realistic to require this extra functionality of all IPv6 nodes for two reasons. First, on a practical level, IPv6 standards documents are still at an early stage of standardization, so it is possible to place additional requirements on IPv6 nodes. Second, processing binding updates can be implemented as a fairly simple modification to IPv6's use of the destination cache.[15]

Security

One of the biggest differences between IPv6 and IPv4 is that all IPv6 nodes are expected to implement strong authentication and encryption features[21,22] to improve Internet security. This affords a major simplification for IPv6 mobility support, since all authentication procedures can be assumed to exist when needed and do not have to be specified in the Mobile IPv6 protocol. Even with the security features in IPv6, however, the current working group draft for IPv6 mobility support specifies the use of authentication procedures as infrequently as possible. The reasons for this are twofold. First, good authentication comes at the cost of performance and so should be required only occasionally. Second, questions about the availability of Internet-wide key management are far from resolved at this time.

Source Routing

In contrast to the way in which route optimization is specified in IPv4, in IPv6 correspondent nodes do not tunnel packets to mobile nodes. Instead, they use IPv6 routing headers, which implement a variation of IPv4's *source routing* option. A number of early proposals for supporting mobility in IPv4 specified a similar use of source routing options,[23,24] but two main problems precluded their use:

- IPv4 source routing options require the receiver of source-routed packets to follow the reversed path to the sender back along the indicated intermediate nodes. This means that malicious nodes using source routes from remote locations within the Internet could impersonate other nodes, a problem exacerbated by the lack of authentication protocols.

MOBILE NETWORKING TERMINOLOGY

Agent advertisement. The procedure by which a mobility agent becomes known to the *mobile node*.

Agent discovery. The process by which a *mobile node* can obtain the IP address of a *home agent* or *foreign agent*, depending upon whether the mobile node is home or away from home. Agent discovery occurs when a mobile node receives an *agent advertisement*, either as a result of periodic broadcast or in response to a solicitation.

Automatic home agent discovery. The process by which a *mobile node* can obtain the IP address of a *home agent* on its *home network*, involving the transmission of a *registration* request to the subnet broadcast address of its home network.

Binding. The triplet of numbers that contains the *mobile node's* home address, its *care-of address*, and the **registration lifetime**—how long the mobility agents may use the binding.

Binding update. The message that supplies a new *binding* to an entity that needs to know the new *care-of address* for a *mobile node*. The binding update contains the mobile node's home address, new care-of address, and a new *registration lifetime*.

Care-of address. An IP address at the *mobile node's* current point of attachment to the Internet, when the mobile node is not attached to the *home network*. A **collocated care-of address** is a care-of address assigned to one of the mobile node's network interfaces, instead of one being offered by a *foreign agent*.

Correspondent node. A node that sends or receives a packet to a *mobile node*; the *correspondent node* may be another mobile node or a nonmobile Internet node.

Discovery. In this article, short for *agent discovery*.

Encapsulation. The process of incorporating an original IP packet (less any preceding fields such as a MAC header) inside another IP packet, making the fields within the original IP header temporarily lose their effect.

Foreign agent. A *mobility agent* on the *foreign network* that can assist the *mobile node* in receiving datagrams delivered to the *care-of address*.

Foreign network. The network to which the *mobile node* is attached when it is not attached to its *home network*, and on which the *care-of address* is reachable from the rest of the Internet.

Fully qualified domain name (FQDN). An Internet node's FQDN is its complete domain name as defined by the Domain Name System (DNS). A node can be known locally by a relative domain name that is a substring of its FQDN, but such a relative name cannot be resolved correctly by Internet nodes outside of the part of the domain name hierarchy indicated by the relative name. The fully qualified name can be resolved from anywhere in the Internet, subject to access control and routability of the resolution request.

Home address. The IP address assigned to the *mobile node*, making it logically appear attached to its *home network*.

Home agent. A node on the *home network* that effectively causes the *mobile node* to be reachable at its *home address* even when the mobile node is not attached to its home network.

Home network. The network at which the *mobile node* seems reachable, to the rest of the Internet, by virtue of its assigned IP address.

Minimal encapsulation. A variant *encapsulation* technique specified in RFC 2003 that temporarily alters the structure of the original IP header, but uses fewer bytes for tunneling packets to the *care-of address* than the default method (IP-within-IP) uses.

Mobile node. A node that, as part of normal use, changes its point of attachment to the Internet.

Mobility agent. A node (typically, a router) that offers support services to *mobile nodes*. A mobility agent can be either a *home agent* or a *foreign agent*.

Nomadicity. The full range of network technology being designed to come to the assistance of the mobile (or nomadic) computer user, not limited to network-layer protocols.

Redirection. A message that is intended to cause a change in the routing behavior of the node receiving it.

Registration. The process by which the *mobile node* informs the *home agent* about its current *care-of address*.

Remote redirection. A redirect sent from a source not present on the local network. The source can be located anywhere in the global Internet and may have malicious intent and be untraceable.

Replay attacks. A security violation whereby a malicious entity attempts to imitate a transaction recorded during a previous and valid transaction between two protocol entities. Both protocol entities have to be aware that the subsequent identical traffic streams may no longer be valid. Since the previous transaction was valid, the algorithms for detecting replay attacks need to incorporate data that can never be reproduced in any correct subsequent transaction.

Route optimization. A process that enables the delivery of packets directly to the *care-of address* from a *correspondent node* without having to detour through the *home network*.

Source routing. A routing technique that causes some or all intermediate routing points to be represented directly in the data packet to be forwarded. This is in contrast to the typical situation in which intermediate routers rely on acquired routing state information to forward incoming packets.

Tunneling. The same as *encapsulation*, but with additional connotations about changing the effects of Internet routing on the original IP packet.

MOBILE IP WEB RESOURCES

You can view the Mobile IP working group's charter and all Internet drafts and RFC documents at http://www.ietf.org/html.charters/mobileip-charter.html. You can also join the general mail list for the Mobile IP working group by sending mail to majordomo@smallworks.com and including the line "subscribe mobile-ip" in the body of the message. Archives of the mail list are available at ftp://ftp.smallworks.com/mobile-ip.archive. New members of the general discussion list should read the materials found at http://www.ietf.org/overview.html and http://www.ietf.org/tao.html.

Further major Web resources for Mobile IP, including various freeware implementations, can be found at the following sites:

The CMU Monarch Project
Protocols for Adaptive Mobile and Wireless Networking •
http://www.monarch.cs.cmu.edu/

Portland State Secure Mobile Networking Project •
http://www.cs.pdx.edu/research/SMN/

Mobile IP at the National University of Singapore •
http://mip.ee.nus.sg/

State University of New York, Binghamton
Linux-Mobile IP •
http://anchor.cs.binghamton.edu/~mobileip/

Stanford's Operating Systems and Networking Group
MosquitoNet Mobile IP •
http://mosquitonet.stanford.edu/software/mip.html

BBN Technologies Mobile IP Security page •
http://www.net-tech.bbn.com/moips/moips-index.html

- Existing routers exhibit terrible performance when handling source routes. Consequently, the results of deploying other protocols that use source routes have not been favorable.

However, the objections to the use of source routes do not apply to IPv6, because IPv6's more careful specification eliminates the need for source-route reversal and lets routers ignore options that do not need their attention. Consequently, correspondent nodes can use routing headers without penalty. This allows the mobile node to easily determine when a correspondent node does not have the right care-of address. Packets delivered by encapsulation instead of by source routes in a routing header must have been sent by correspondent nodes that need to receive binding updates from the mobile node. It is a further point of contrast to route optimization in IPv4 that, in IPv6 mobility support, the mobile node delivers binding updates to correspondent nodes instead of to the home agent. In IPv6, key management between the mobile node and correspondent node is more likely to be available.

Other features supported by IPv6 mobility include

- coexistence with Internet ingress filtering;[25]
- smooth handoffs, which in Mobile IPv4 is specified for foreign agents as part of route optimization;
- renumbering of home networks; and
- automatic home agent discovery.

ONGOING WORK AND OPEN QUESTIONS

Problems Facing Mobile IP

The most pressing outstanding problem facing Mobile IP is that of security, but other technical as well as practical obstacles to deployment exist.[26] Work is also continuing to refine and extend the protocol within the academic and commercial communities and within the IETF. This section surveys the state of implementation of Mobile IP and speculates on a possible timetable for deployment.

Routing inefficiencies. The base Mobile IP specification has the effect of introducing a tunnel into the routing path followed by packets sent by the correspondent node to the mobile node. Packets from the mobile node, on the other hand, can go directly to the correspondent node with no tunneling required. This asymmetry is captured by the term triangle routing, where a single leg of the triangle goes from the mobile node to the correspondent node, and the home agent forms the third vertex controlling the path taken by data from the correspondent node to the mobile node. Triangle routing is alleviated by use of techniques in the route optimization draft,[20] but doing so requires changes in the correspondent nodes that will take a long time to deploy for IPv4. It is hoped that triangle routing will not be a factor for IPv6 mobility.

Security issues. A great deal of attention is being focused on making Mobile IP coexist with the security features coming into use within the Internet. Firewalls,[27] in particular, cause difficulty for Mobile IP because they block all classes of incoming packets that do not meet specified criteria. Enterprise firewalls are typically configured to block packets from entering via the Internet that appear to emanate from internal computers. Although this permits management of internal Internet nodes without great attention to security, it presents difficulties for mobile nodes wishing to communicate with other nodes within their home enterprise networks. Such communications, originating from the mobile node, carry the mobile node's home address, and would thus be blocked by the firewall.

Mobile IP can be viewed as a protocol for establishing secure tunnels. Gupta and Glass have proposed a firewall traversal solution.[28] Efforts along these lines are also being made at BBN as part of the MOIPS (Managed Objects for IP Mobility Support)[29] project to extend Mobile IP operation across firewalls, even when multiple security domains are involved.

Ingress filtering. Complications are also presented by ingress filtering[25] operations. Many border routers discard packets coming from within the enterprise if the packets do not contain a source IP address configured for one of the enterprise's internal networks. Because mobile nodes would otherwise use their home address as the source IP address of the packets they transmit, this presents difficulty. Solutions to this problem in Mobile IPv4 typically involve tunneling outgoing packets from the care-of address, but then the difficulty is how to find a suitable target for the tunneled packet from the mobile node. The only universally agreed on possibility is the home agent, but that target introduces yet another serious routing anomaly for communications between the mobile node and the rest of the Internet. Montenegro has proposed the use of reverse tunnels to the home agent to counter the restriction imposed by ingress filtering.[30] Mobile IPv6 also offers a solution in the home address destination option.[18]

User perceptions of reliability. The design of Mobile IP is founded on the premise that connections based on TCP should survive cell changes. However, opinion is not unanimous on the need for this feature. Many people believe that computer communications to laptop computers are sufficiently bursty that there is no need to increase the reliability of the connections supporting the communications. The analogy is made to fetching Web pages by selecting the appropriate URLs. If a transfer fails, people are used to trying again. This is tantamount to making the user responsible for the retransmission protocol and depends for its acceptability on a widespread perception that computers and the Internet cannot be trusted to do things right the first time. Naturally, such assumptions are strongly distasteful to many Internet protocol engineers, myself included. Nevertheless, the fact that products exhibiting this model are currently economically viable cannot be denied. Hopefully in the near future better engineering will counter this perception and increase the demand for Internet reliability.

Issues in IP addressing. Mobile IP creates the perception that the mobile node is always attached to its home network. This forms the basis for the reachability of the mobile node at an IP address that can be conventionally associated with its *fully qualified domain name* (FQDN).[31] If the FQDN is associated with one or more other IP addresses, perhaps dynamically, then those alternative IP addresses may deserve equal standing with the mobile node's home address. Moreover, it is possible that such an alternative IP address would offer a shorter routing path if, for instance, the address were apparently located on a physical link nearer to the mobile node's care-of address, or if the alternative address were the care-of address itself. Finally, many communications are short-lived and depend on neither the actual identity of the mobile node nor its FQDN, and thus do not take advantage of the simplicity afforded by use of the mobile node's home address. These issues surrounding the mobile node's selection of an appropriate long-term (or not-so-long-term) address for use in establishing connections are complex and are far from being resolved.

Slow growth in the wireless LAN market. Mobile IP has been engineered as a solution for wireless LAN location management and communications, but the wireless LAN market has been slow to develop. It is difficult to make general statements about the reasons for this slow development, but with the recent ratification of the IEEE 802.11 MAC protocol,[32] wireless LANs may become more popular. Moreover, the bandwidth for wireless devices has been constantly improving, so that radio and infrared devices on the market today offer multimegabyte-per-second data rates. Faster wireless access over standardized MAC layers could be a major catalyst for growth of this market.

Competition from other protocols. Mobile IP may well face competition from alternative tunneling protocols such as PPTP[33] and L2TP.[34] These other protocols, based on PPP, offer at least portability to mobile computers. Although I believe portable operation will ultimately not be a long-term solution, it may look quite attractive in the short term in the absence of full Mobile IP deployment. If these alternative methods are made widely available, it is unclear if the use of Mobile IP will be displaced or instead made more immediately desirable as people experience the convenience of mobile computing. In the future, it is also possible that

HISTORY OF THE MOBILE IP WORKING GROUP

The Mobile IP Working Group of the Internet Engineering task Force (IETF) had its origin in BOF (Birds of a Feather) sessions held at the Atlanta (July 1991), Santa Fe (November 1991), and San Diego (March 1992) IETF meetings. In June 1992 Steve Deering, chair of the working group, submitted a proposed charter for a formal Working Group to the IETF, and, following a revision of the charter, the Working Group was officially formed in June 30, 1992. An IPv4 Mobile Host Protocol was submitted to the Internet Engineering Steering Group (IESG) as a proposed standard in 1996. An IPv6 protocol will be submitted to the IESG later this year.

Mobile IP could specify use of such alternative tunneling protocols to capitalize on their deployment on platforms that do not support IP-within-IP encapsulation.

Current Development Efforts

Mobile IP has been studied in a number of wireless communication research projects. At the University of California at Berkeley,[35] Mobile IP is being used to construct vertical handoffs between dissimilar media (for example, infrared, radio LANs, wide-area cellular, and satellite), depending upon error rates and bandwidth availability. Other factors such as cost and predictive service might also be taken into account. CMU's Monarch project[36] has been the focus of investigation into campus wireless networks, Mobile IP, Mobile IPv6, and ad-hoc networking.[37] Other academic efforts have been proceeding at the University of Portland, University of Alabama, University of Texas, UCLA, Macquarie University, SUNY Binghampton, University of Singapore, Swedish Royal Institute of Technology, and many others. Two books about Mobile IP have recently been published.[2,38]

Current IETF drafts that employ Mobile IP include the Tunnel Establishment Protocol[39] and Mobile IP Local Registration with Hierarchical Foreign Agents.[40] The latter uses the ability to advertise multiple foreign agents to arrange hierarchies of mobility agents. This may help cut the number of registrations that must transit the global Internet between the home and foreign networks. DHCP for Mobile Networking with TCP/IP[41] investigates the suitability of using the Dynamic Host Configuration Protocol[42,43] to provide care-of addresses to mobile nodes.[41] Mobile IPv4 Configuration Option for PPP IPCP[44] is a new extension to PPP[45] that will enable dial-up users to more efficiently employ their dynamic IP addresses as care-of addresses.

CONCLUSION

As this brief introduction to mobile networking has shown, Mobile IP has great potential. Security needs are getting active attention and will benefit from the deployment efforts underway. Within the IETF, Mobile IP is likely to move from a proposed standard[1] to a draft standard[46] in the near future.

The IETF standardization process requires the working group to rigorously demonstrate interoperability among various independent implementations before the protocol can advance. FTP Software has hosted two interoperability testing sessions, and many vendors have taken advantage of the opportunity. Test results have given added confidence that the Mobile IP specification is sound, implementable, and of diverse interest throughout the Internet community. Only a few minor revisions have been needed to ensure the specification can be interpreted in only one way by the network protocol engineers and programmers who must implement it.

It is possible that the deployment pace of Mobile IP will track that of IPv6, or that the requirements for supporting mobility in IPv6 nodes will give additional impetus to the deployment of both IPv6 and mobile networking. The increased user convenience and the reduced need for application awareness of mobility can be a major driving force for adoption. Since both IPv6 and Mobile IP have little direct effect on the operating systems of mobile computers outside of the network layer of the protocol stack, application designers should find this to be an acceptable programming environment. Of course, everything depends heavily on the willingness of platform and router vendors to implement Mobile IP and/or IPv6, but indications are strong that most major vendors already have implementations either finished or underway. ■

Interoperability testing sessions have indicated that the Mobile IP specification is sound, implementable, and of diverse interest throughout the Internet community.

REFERENCES

1. "IP Mobility Support," C. Perkins, ed., IETF RFC 2002, Oct. 1996.
2. C. Perkins, *Mobile IP: Design Principles and Practice*, Addison-Wesley Longman, Reading, Mass., 1998.
3. C. Perkins, "Mobile IP," *IEEE Comm.*, Vol. 35, No. 5, 1997, pp. 84–99.
4. S. Deering and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification," IETF RFC 1883, Dec. 1995.
5. R. Hinden and S. Deering, "IP Version 6 Addressing Architecture," IETF RFC 1884, Dec. 1995.
6. "Internet Protocol," J.B. Postel, ed., IETF RFC 791, Sept. 1981.
7. CDPD Consortium, "Cellular Digital Packet Data Specification," PO Box 809320, Chicago, Ill., July 1993, http://www.cdpd.org/public/specification/index.html.
8. P. Bhagwat, C. Perkins, and S.K. Tripathi, "Network Layer Mobility: An Architecture and Survey," *IEEE Personal Comm.*, Vol. 3, No. 3, June 1996, pp. 54–64.
9. "ICMP Router Discovery Messages," S.E. Deering, ed., IETF RFC 1256, Sept. 1991.
10. C. Perkins, "IP Encapsulation Within IP," IETF RFC 2003, May 1996.
11. C. Perkins, "Minimal Encapsulation Within I," IETF RFC 2004, May 1996.
12. V.L. Voydock and S.T. Kent, "Security Mechanisms in High-Level Networks," *ACM Computer Surveys*, Vol. 15, No. 2, June 1983, pp. 135–171.
13. R.L. Rivest, "The MD5 Message-Digest Algorithm," IETF RFC 1321, Apr. 1992.
14. S. Thomson and T. Narten, "IPv6 Stateless Address Autoconfiguration," IETF RFC 1971, Aug. 1996.
15. T. Narten, E. Nordmark, and W. Simpson, "Neighbor Discovery for

IP Version 6 (IPv6)," IETF RFC 1970, Aug. 1996.
16. I. Castineyra, J. Chiappa, and M. Steenstrup, "The Nimrod Routing Architecture," IETF RFC 1992, Aug. 1996.
17. S. Bradner and A. Mankin, "The Recommendation for the IP Next Generation Protocol," IETF RFC 1752, Jan. 1995.
18. D. Johnson and C. Perkins, "Mobility Support in IPv6," *ACM Mobicom 96*, ACM, Nov. 1996, pp. 27–37.
19. A. Conta and S. Deering, "Generic Packet Tunneling in IPv6," ftp://ftp.ietf.org/internet-drafts/draft-ietf-ipngwg-ipv6-tunnel-07.txt, July 1996 (work in progress).
20. C.E. Perkins and D.B. Johnson, "Route Optimization in Mobile-I," ftp://ftp.ietf.org/internet-drafts/draft-ietf-mobileip-optim-07.txt, Nov. 1997 (work in progress).
21. S. Kent and R. Atkinson, "IP Authentication Header," ftp://ftp.ietf.org/internet-drafts/draft-ietf-ipsec-auth-header-03.txt, Nov. 1997 (work in progress).
22. S. Kent and R. Atkinson, "IP Encapsulating Security Payload (ESP)," ftp://ftp.ietf.org/internet-drafts/draft-ietf-ipsec-esp-v2-02.txt, Nov. 1997 (work in progress).
23. C. Perkins and P. Bhagwat, "A Mobile Networking System Based on Internet Protocol(IP)," *Proc. USENIX Symp. Mobile and Location-Independent Computing*, Aug. 1993, USENIX Assoc., pp. 69–82.
24. D.B. Johnson, Scalable and Robust Internetwork Routing for Mobile Hosts, *Proc. 14th Intl. Conf. Distributed Computing Systems*, June 1994, pp. 2–11.
25. P. Ferguson and D. Senie, "Ingress Filtering in the Internet," ftp://ftp.ietf.org/internet-drafts/draft-ferguson-ingress-filtering-03.txt, Oct. 1997 (work in progress).
26. S. Cheshire and M. Baker, "Internet Mobility 4x4," *Proc. ACM SIGCOMM Conf. Applications, Technologies, Architectures, and Protocols for Computer Comm.*, Vol. 26, No. 4, *ACM SIGCOMM Computer Comm. Rev.*, ACM Press, New York, 1996, pp. 318–329.
27. W.R. Cheswick and S. Bellovin, *Firewalls and Internet Security*, Addison-Wesley, Reading, Mass., 1994.
28. V. Gupta and S. Glass, "Firewall Traversal for Mobile IP: Guidelines for Firewalls and Mobile IP Entities," ftp://ftp.ietf.org/internet-drafts/draft-ietf-mobileip-firewall-trav-00.txt, Mar. 1997 (work in progress).
29. J. Zao et al., "A Public-Key Based Secure Mobile IP," *Proc. ACM Mobicom 97*, ACM, New York, Oct. 1997, pp. 173–184.
30. G. Montenegro, "Reverse Tunneling for Mobile I," ftp://ftp.ietf.org/internet-drafts/draft-ietf-mobileip-tunnel-reverse-04.txt, Aug. 1997 (work in progress).
31. P. Mockapetris, "Domain Names—Concepts and Facilities," STD 13, IETF, Nov. 1987.
32. "Wireless LAN, MAC, and PHY Specifications," IEEE Document P802.11/D6.1.97/5, June 1997, AlphaGraphics #35, 10201 N. 35th Ave., Phoenix, AZ 85051.
33. G. Pall et al., "Point-to-Point Tunneling Protocol—PPT," ftp://ftp.ietf.org/internet-drafts/draft-ietf-pppext-pptp-02.txt, July 1997 (work in progress).
34. W. Palter et al., "Layer Two Tunneling Protocol 'L2TP,'" ftp://ftp.ietf.org/internet-drafts/draft-ietf-pppext-l2tp-08.txt, Nov. 1997 (work in progress).
35. R.H. Katz, "Adaptation and Mobility in Wireless Information Systems," *IEEE Personal Comm.*, Vol. 1, No. 1, 1994, pp. 6–17.
36. D.B. Johnson and D.A. Maltz, "Protocols for Adaptive Wireless and Mobile Networking," *IEEE Personal Comm.*, Vol. 3, No. 1, Feb. 1996, pp. 34–42.
37. D.B. Johnson and D.A. Maltz, "Dynamic Source Routing in Ad Hoc Wireless Networks," in *Mobile Computing*, T. Imielinski and H. Korth, eds., Kluwer Academic Publishers, 1996, pp. 153–181.
38. J. Solomon, *Mobile IP: The Internet Unplugged*, Prentice Hall, Englewood Cliffs, N.J., 1998.
39. P. Calhoun and C. Perkins, "Tunnel Establishment Protocol (TEP)," ftp://ftp.ietf.org/internet-drafts/draft-ietf-mobileip-calhoun-tep-00.txt, Aug. 1997 (work in progress).
40. C. Perkins, "Mobile-IP Local Registration with Hierarchical Foreign Agents," ftp://ftp.ietf.org/internet-drafts/draft-perkins-mobileip-hierfa-00.txt, Feb. 1996 (work in progress).
41. C. Perkins and J. Tangirala, "DHCP for Mobile Networking with TCP/IP," *Proc. IEEE Int'l Symp. Systems and Comm.*, June 1995, pp. 255–261.
42. R. Droms, "Dynamic Host Configuration Protocol," IETF RFC 2131, Mar. 1997.
43. S. Alexander and R. Droms, "DHCP Options and BOOTP Vendor Extensions," IETF RFC 2132, Mar. 1997.
44. J. Solomon and S. Glass, "Mobile-IPv4 Configuration Option for PPP IPC," ftp://ftp.ietf.org/internet-drafts/draft-ietf-pppext-ipcp-mip-02.txt, July 1997 (work in progress).
45. "The Point-to-Point Protocol (PPP)," W.A. Simpson, ed., IETF RFC 1661, July 1994.
46. "IP Mobility Support Version 2," C. Perkins, ed., draft-ietf-mobileip-v2-00.txt, Nov. 1997 (work in progress).

Charles E. Perkins is a senior staff engineer at Sun Microsystems, where he is developing Service Location Protocol and investigating dynamic configuration protocols for mobile networking. He is document editor for the Mobile IP working group of the IETF, where he also serves on the Internet Architecture Board. He has worked on various committees for the National Research Council and is currently chair of the Nomadicity Working Team of the Cross-Industry Working Team (XIWT). Perkins holds a BA in mathematics and an MEE from Rice University,, and an MA in mathematics from Columbia University. He is on the editorial boards of *ACM/IEEE Transactions on Networking, Wireless Networks*, and *Mobile Communications and Computing Review*. He is a member of IEEE, ISOC, ACM, and IETF. .

Contact Perkins at Technology Development Group, Mail Stop MPK 15-214, Sun Microsystems, Mountain View, CA 94303; email charles.perkins@sun.com; http://.www.srvloc.org/charliep.