Host Mobility for IP Networks: A Comparison

Thomas R. Henderson, Boeing Phantom Works

Abstract

The growth of wireless networking is enabling many Internet hosts to become mobile. This article describes and compares three alternatives for providing host mobility management in IP-based networks. We first summarize the operation and behavior of Mobile IP, on which the Internet Engineering Task Force has focused as a host mobility solution. We next describe two alternative architectures (Migrate and Host Identity Protocol) for providing mobility management. Our qualitative comparison focuses on contrasting the different performance, security, deployment, scalability, and robustness properties of each approach.

Advances in radio hardware are enabling the proliferation of wireless mobile communications networks, and trends in data networking point toward continued migration to an IP-centric architecture. However, mobility stresses the basic Internet architecture in several ways, requiring additional mechanisms to be implemented in both end hosts and the network; such mechanisms are generally called *mobility management* mechanisms. Mobility management can be handled at different protocol layers in the architecture, but network-level solutions are typically the most general since IP is ubiquitous in the Internet.

Development of host (or node) mobility management extensions to the Internet architecture and protocols began over 10 years ago in the Internet Engineering Task Force (IETF). These extensions are commonly referred to as *Mobile IP* [1]. Since that time, the Internet has continued to evolve at a rapid pace, in ways that have affected mobility management. Although Mobile IP development has adjusted to these changes, widespread deployment of Mobile IP has proceeded more slowly than originally expected. The slow growth may be attributed to two factors:

- Use of the Dynamic Host Configuration Protocol (DHCP) and virtual private network (VPN) tunneling has sufficed for many mobile users who to date have been primarily concerned with email and Web access while roaming.
- Heavy deployment of firewalls and Network Address Translation (NAT) devices has blocked the originally proposed operation of Mobile IP protocols.

Along the way, a number of additional mobility management techniques have been proposed and developed, motivated by perceived shortcomings in Mobile IP service capabilities or deployment. While it is generally accepted that Mobile IP, in some form, will be part of the Internet's future, it is still not clear which modes of Mobile IP, and which combination of complementary mechanisms, will evolve to become the most widely used.

This article provides a comparison of three host mobility management strategies for IP networks. We provide an overview of the current specification of Mobile IP, and describe two contrasting mobility management architectures: one in which mobility management is handled on an end-to-end basis (Migrate) [2], and one based on a new name space for Internet hosts (the Host Identity Protocol, HIP) [3]. Below, we first summarize the requirements for a general mobility management solution. We then provide a taxonomy of alternative mobility management solutions and list several examples. After summarizing the three considered approaches, our focus shifts to qualitatively comparing them from several angles, including performance, security, deployment, scalability, and robustness.

Requirements for Mobility Management

For small networks, dynamic routing protocols can handle mobility. Routing protocols for such mobile networks are being researched intensively and are summarized elsewhere (e.g., [4]); much of the emphasis in mobile ad hoc routing has been on routing for small subnets of tens to hundreds of nodes. However, as networks grow larger, full host-based routing makes routing table sizes unmanageable, and it becomes attractive to introduce addressing and routing hierarchy to reduce routing overhead. Once addressing becomes hierarchical and mobile hosts are capable of moving from one network prefix to another, it becomes necessary to support additional mobility management mechanisms. The most general requirements include:

- **Location-independent identifier.** Mobile nodes must have or obtain an identifier that remains static across location changes.
- **Compatibility with IP routing.** Mobility management must interwork successfully with IP routing, such as acquiring a new topologically correct IP address upon moving, since full host routes are not propagated in the Internet.
- **Location management.** If a mobile node offers services to other nodes, it must be able to be located by clients or peers as it changes its location.
- **Transparency.** Mobility management mechanisms should offer some level of transparency to higher-layer protocols and applications. For example, the act of readdressing should not normally cause a TCP connection to break.
- **Security.** Mobility management mechanisms should not introduce additional security vulnerabilities into the network.

0890-8044/03/$17.00 © 2003 IEEE

Name space	Examples	Paradigm
IP address	Mobile IP [1] MSM IP [16]	Assign a permanent IP address to the host. Allow the network to track the transient location of the mobile host and forward packets to the temporary destination. Optionally, allow the mobile host to directly notify peers upon movement.
Hostname	Migrate [2]	Allow the mobile host to change addresses. If needed, allow the mobile host to update DNS records upon movement. Allow the mobile host to directly notify peers of an address change upon movement.
New name space	Host Identity Payload (HIP) [3] Nimrod [12]	Separate host or process addresses from interface addresses. Allow the mobile host to update DNS records or network directories upon movement. Allow the mobile host to directly notify peers upon movement.

■ Table 1. *Three fundamental name space alternatives for network mobility management.*

Depending on the scenario of interest, a number of other (possibly conflicting) requirements may be present, including performance, scalability, deployability, and robustness. Later in this article we qualitatively compare the three approaches to mobility management along these lines.

Alternative Mobility Management Solutions

Before looking at Mobile IP, Migrate, and HIP more closely, it is instructive to first briefly summarize some other types of mobility management solutions. While these approaches are not generally applicable to all IP applications, they may have the benefit of more directly addressing the requirements for a particular protocol or service.

Transport-layer approaches. Transport protocols, TCP in particular, have not been compatible with IP readdressing since TCP protocol control blocks (PCBs) are named by the quadruple (source IP address, source port number, destination IP address, destination port number), and because the transport layer checksum covers the IP addresses. Extended TCP [5] proposed that a new TCP PCB identifier be used to name the TCP socket, thereby allowing underlying IP addresses to change. The MSOCKS proposal [6] resembles Mobile IP in use of agents in the network, but at the transport layer; split connection proxies allow TCP clients to move transparently. Finally, work is ongoing in the IETF to modify the Stream Control Transmission Protocol [7] to allow it to dynamically change endpoint addresses in the midst of a connection [8].

Application-level approaches. The Session Initiation Protocol (SIP) can be used to support terminal mobility [9]. Mobile nodes register new addresses with their SIP registrar. Midcall location updates are accomplished by sending a new INVITE to the correspondent node.

Session mobility (context transfer). The IETF has identified a need to allow edge mobile devices to transfer state information pertaining to access control, security context, QoS reservations, and so on, and has established the *seamoby* working group in this area. Session mobility may also occur on the end hosts; an interesting new approach in this area is the reliable sockets (*rocks*) project [10], which provides a modified user-level sockets library that facilitates transparent mobility and process migration by interposing an additional protocol above TCP while exporting a standard sockets application programming interface (API) to applications.

Personal mobility. Personal mobility allows a single user located at different terminals to be reachable by the same name or address. Examples of this type of mobility management include the Berkeley ICEBERG Universal Inbox [11] and SIP forking proxies [9].

Service mobility. Service mobility allows mobile users to access the same services while changing network providers or communication devices; examples are found in [9].

Alternative architectures. Researchers have proposed novel internetworking architectures that have implications on how mobility management is performed. Nimrod [12] is a proposed revamp of the Internet routing and addressing architecture. Nimrod separates node addresses from interface identifiers. However, RFC 2103 describes how Nimrod concepts could be aligned with Mobile IP [13]. IPNL [14] is a "NAT-extended" architecture that introduces a new routing layer above IPv4 that is routed by NATs based on fully qualified domain names (FQDNs). Mobile hosts in an IPNL architecture must register with an IPNL router in both the visited and home domains, and these routers in turn flood the information to all other IPNL routers in their respective realms. It is implied that IPNL router (NAT) connectivity is somewhat static.

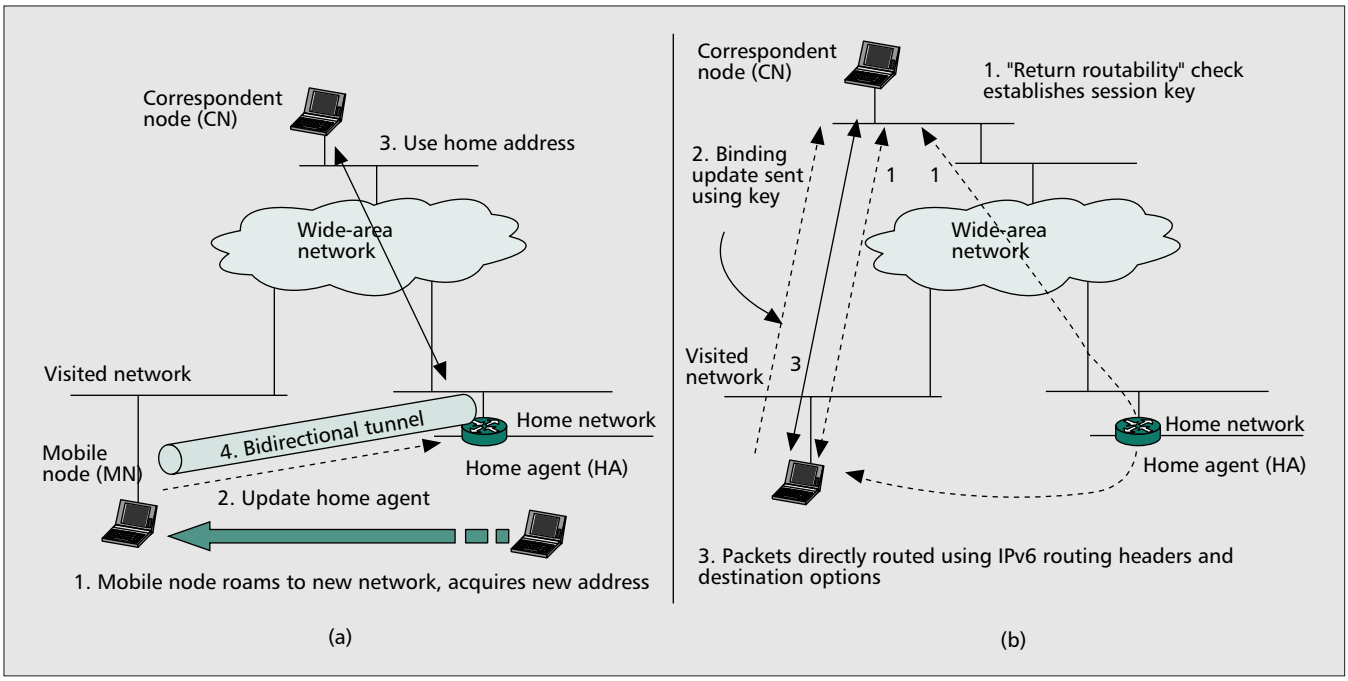
As mentioned above, a location-independent identifier is central to each mobility management approach. At the network level, two candidate name spaces are the two in use in today's Internet: the IP address space, and the set of FQDNs. A third possibility would be to create a new name space, either a fixed naming system or an agile one that implements a "late binding" option to track mobile services [15]. Table 1 summarizes these three possibilities, and we next summarize each approach.

Mobile IP

Mobile IP was originally developed as an extension to IPv4 protocols (MIPv4) [17]. A description of the original requirements and architecture development is found in [18]. In contrast, Mobile IPv6 has been developed as an integral part of IPv6 (MIPv6), which as of this writing has nearly been approved as a proposed standard [19]. It is the consensus of the IETF Mobile IP working group that MIPv6 is an improved version of MIPv4. Both MIPv4 and MIPv6 offer a mode of operation described below as "Mobile IP through a home agent." Additionally, MIPv6 specifies a second mode of operation, "Mobile IP with route optimization." We summarize both of these modes below. We also categorize additional work that has been performed in extending these basic approaches.

Mobile IP through a Home Agent

In Mobile IP, a mobile node is assigned a *home address* that serves as its unique endpoint identifier (EID). Packets from any correspondent nodes are routed to the home network prefix corresponding to the home address. If a mobile node moves away from its home network, packets that arrive for the mobile node must be delivered to its new location. This is accomplished by a *home agent*, which takes responsibility for intercepting packets destined to the mobile node and tunneling them (via packet encapsulation such as IP-in-IP encapsulation) to the mobile node's current destination. In MIPv4, a mobile node may make use of a *foreign agent*, if available in the visited network, for packet decapsulation, thereby avoiding the need to obtain a temporary IP address. Alternatively (and exclusive-



■ Figure 1. *Essential operation of Mobile IP, with route optimization: a) home-agent-based mobile IP; b) mobile IP route optimization extensions.*

ly in MIPv6), the mobile node acquires a topologically correct address, known as a *care-of address*, for the visited network, and takes responsibility itself for packet decapsulation. It is therefore necessary for the mobile node to register with and update its home agent when its point of attachment changes.

Packets originated from a mobile node while on its home network use the mobile node's home address as the IP source address. When a mobile node is on a visited network, it may attempt to send packets directly to a correspondent node using the home address as the source address. However, many networks do not permit the forwarding of packets with source addresses that are topologically incorrect for the network (a policy known as *ingress filtering*). In this case, the mobile node may "reverse-tunnel" its packets back to the home agent for decapsulation and forwarding on to the correspondent node by using its care-of address as the outer IP source address. Figure 1a provides an overview of reverse tunneling operation in which the mobile node obtains a care-of address from a DHCP server in the visited network, and for which authorization from an authentication, authorization, and accounting (AAA) server is possibly required. Internet RFC 3344 [17] contains many more details of operation and techniques to secure the various protocol transactions, dynamically discover agents, and intercept packets on the home network.

A fundamental strength of this mode of operation is that outside networks are not required to implement protocol extensions to either correspondent nodes or the Domain Name System (DNS). The cost of this backward compatibility is a need to deploy a home agent in the home network, and the need to route all packets through the home network when the correspondent node initiates the session. This suboptimal routing can be avoided if correspondent nodes implement route optimization extensions, as discussed next.

Mobile IP with Route Optimization

Although route optimization extensions were proposed for both MIPv4 and MIPv6, they have only been standardized for MIPv6. These extensions assume default use of home agent-based operation as described above, but also allow a mobile node to notify a correspondent node directly of the mobile node's current address, to permit the correspondent node to deliver packets directly to the mobile node. This optimization improves scalability and reliability and reduces network load [19]. The correspondent node maintains a *binding cache* that stores the current care-of address of the mobile node; the mobile node uses a binding update message to notify the correspondent node of an address change. When the correspondent node originates a packet, it includes a special type of IPv6 routing header to carry the home address of the mobile node, but uses the current care-of address as the destination address. Likewise, when a mobile node originates a packet, it uses an IPv6 home address destination option to identify its home address, while using the current care-of address as its source address, thereby passing through ingress-filtering devices. The use of routing headers and destination options constitutes partial or degenerate tunneling, since only one additional IPv6 address must be carried in each packet rather than two.

The major need for route optimization is to establish security parameters so that the mobile and correspondent nodes can authenticate and protect the integrity of signaling messages. This is crucial for preventing denial-of-service (or connection hijacking) attacks, since it is difficult for an arbitrary node to prove to another that it "owns" a home address and is authorized to change the packet routing for that address. In the absence of a cryptographic key infrastructure or preconfigured security associations, the current MIPv6 draft describes *return routability* procedures that allow for a security association between the mobile and correspondent nodes that is at least as trustworthy as the packet routing infrastructure between the correspondent nodes and the home network. The return routability procedure proves to a correspondent node that the mobile node is reachable at both its home address and its prospective care-of address. A sequence of four control packets are exchanged between the mobile and correspondent nodes, with two of these packets routed through the home network. The result of this exchange is a key that can be used to generate authentication data to secure the subsequent binding update. Currently, the security association is transient and must be reestablished for each new care-of address. Figure 1b provides an overview of the return routabil-

ity procedure; a detailed summary of its development and other recent MIPv6 security decisions can be found in [20].

Mobile IP Extensions

Development of Mobile IP extensions to the basic operating modes described above continues to be an active area of research and development. We briefly summarize a few general classes of extensions.

Micromobility. If a mobile node changes its IP address frequently (e.g., in an IP-based cellular system), the latency and overhead involved in repeatedly updating home agents and correspondent nodes may be too severe, and packet losses due to stale bindings in the interim may be unacceptable. There has been a significant amount of work on Mobile IP *micromobility* extensions, which are designed to localize the effects of mobility to a smaller region around the mobile terminal, thereby masking the rapid mobility from distant nodes. There are quite a few proposals summarized elsewhere [21, 22]. The proposals can be grouped into the following categories. *Host-based routing* schemes create distributed location databases in the visited network that dynamically route the packets to the right attachment point; upon mobility, the mobile node does not have to readdress, but instead has to update the local routing for its care-of address. *Hierarchical tunneling-based* schemes provide local *anchor points* to which packets from the outside are delivered; the binding updates for rapidly moving nodes are terminated at these anchor points, and packets received by these anchor points are subsequently tunneled again to the mobile node's current address. Finally, *smooth handover* schemes aim to reduce packet losses by using signaling to instruct previously used access routers to forward all packets received for an old care-of address to the new address, and by potentially using two access routers simultaneously ("make-before-break" handover) if the underlying link layer technology supports soft handover.

Access control. In addition to Mobile IP procedures to ensure that packets are routed correctly, mobile nodes must be able to obtain network access in networks under different administrative control. Presently in dialup networks, this access control is performed by AAA servers, for which RADIUS is an example protocol [23]. It would be desirable, then, to integrate Mobile IP binding updates with network access into a single procedure. Internet RFC 2977 identifies the requirements that must be met by AAA servers to aid in providing Mobile IP services [24]. A suggested architecture for integrating Mobile IP and AAA can be found in [25]. Interestingly, AAA registration is often based on a different name space, the Network Access Identifier (NAI), of the form `user@realm`, which means that an integrated MIP-AAA proposal requires coordination of two names for the mobile host.

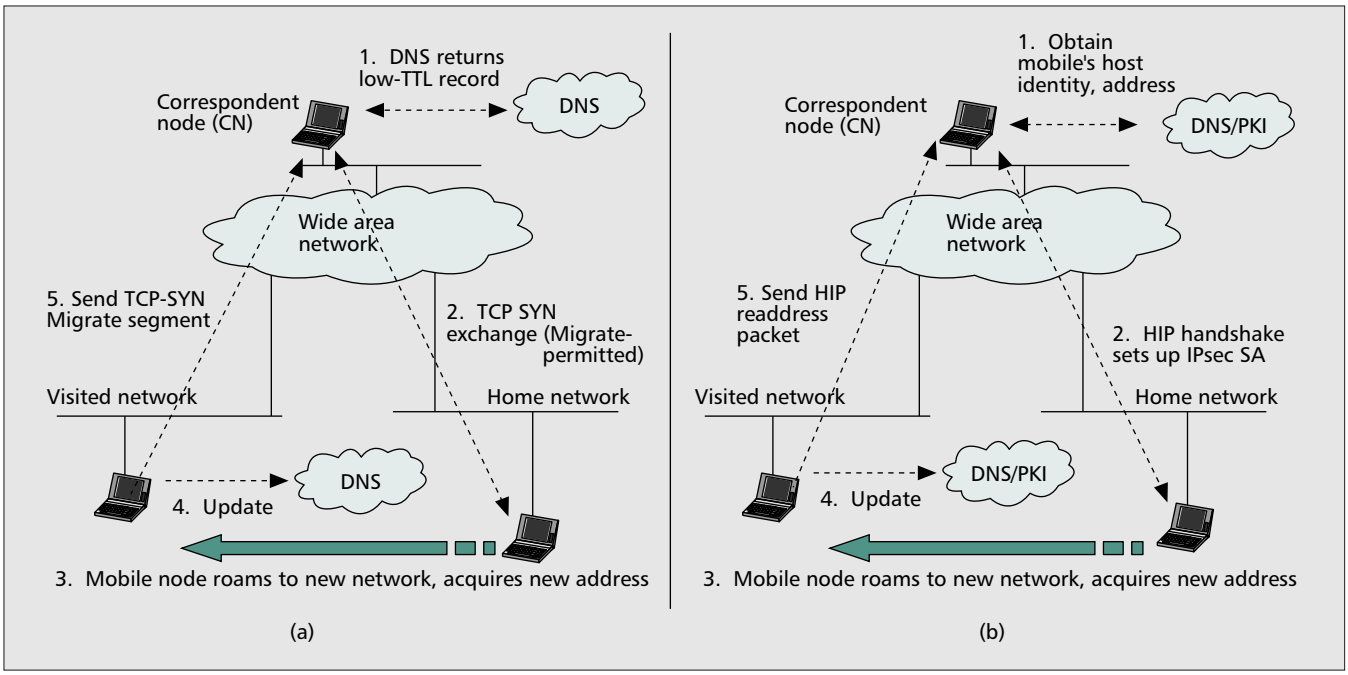
Avoidance of a home network. Even with route optimization, Mobile IP requires some level of operation through a home network, which can lead to robustness problems (single point of failure) and performance issues (latency, overhead) if the home network is far away. A number of proposals have concentrated on the possibility of less reliance on a fixed home network or home agent. In Mobile IP with location registers (MIP-LR), multiple geographically distributed *home location registers* (HLRs), by analogy with cellular systems, can be employed [26]. Mobile nodes update the HLR rather than a home agent, and correspondent nodes query the HLR for a current address of the mobile node. The cost of this optimization, aside from the need for additional infrastructure, is a loss of transparency with respect to correspondent nodes. A recent IETF proposal [27] introduced a "homeless" extension to MIPv6. A host supporting this extension is able to operate without a unique home address, and is always considered away from home. Specifically, it proposed that each host maintain a *host cache* of source (local) addresses valid for a connection, and a *foreign cache* of destination addresses valid for a connection. The main difference between this proposal and MIPv6 is that TCP or UDP PCBs and IP security (IPsec) associations are not bound to individual IP addresses but instead to host/foreign caches. At the protocol level, while neither MIPv6 routing headers nor home address destination options are needed when communicating with another homeless host, a number of new IPv6 destination options are needed, including a "homeless support" option for negotiating this option at connection initiation. While homeless MIPv6 still retains the concept of an address as the host's EID, it can be viewed as a step toward proposals, to be discussed next, that handle mobility management mainly on an end-to-end basis.

Migrate

When Mobile IP began to be standardized over seven years ago, an alternative candidate solution was to use a hierarchical directory service such as DNS to support location management (e.g., [28]). The fundamental difference with such an approach is that a hostname, rather than an IP address, serves as the invariant name. However, at the time, dynamic DNS updates were not part of the Internet standards, and there was concern that DNS approaches were not scalable. A further problem was that TCP uses IP addresses as part of its connection identifier, making IP address migration difficult.

A recent proposal by Snoeren and Balakrishnan [2] revisits this decision, invoking the classic end-to-end argument [29]. The authors argue that host mobility may best be provided for some applications on an end-to-end basis, without reliance on any new network mechanisms. The key to their overall architecture is the use of an FQDN as a host's invariant name. In their approach, as with Mobile IP, portability can be achieved through DHCP. However, location determination is done on the basis of DNS lookups on a per-session basis. That is, every time a host wants to initiate a new session with a server or another host, DNS is consulted. Each time the host moves, it updates the A and PTR records (mappings between hostnames and IP addresses) in the DNS server within the host's home domain. This feature leverages advances in secure DNS update protocols that are now available in DNS software distributions. Stale DNS bindings are avoided by making the binding uncacheable via zero time-to-live (TTL) values in the records. Since DNS is not involved in routing, it does not matter that hosts located within a particular subnet may have domain names from different domains. Also, uncacheable DNS records are only required if the mobile host is offering services to other hosts, which traditionally has been the less common case: mobile clients do not necessarily need DNS records, and static servers can continue to exploit DNS caching.

Session maintenance is the hardest challenge in this approach. It requires end-to-end participation between end hosts and, in particular, modifications to TCP. The authors propose a *Migrate* option to TCP that allows an existing TCP connection to be migrated by either host from an old IP address to a new IP address. This TCP connection migration can be accomplished by exchanging two TCP segments (SYN with Migrate option and ACK of that segment). To prevent connection hijacking, the exchange can be secured using IPsec or an optional Diffie-Hellman key exchange at connection onset. In a later article [30], the authors extend Migrate beyond TCP connection migration to also include an optional session layer to allow better handling of prolonged network disconnection. This session layer can survive changes in underlying protocol states by providing checkpointing, resource control, and resumption facilities to applications.



■ Figure 2. *Operation of a) TCP Migrate; b) HIP.*

Figure 2a summarizes the protocol transactions needed to support a mobile server with an active TCP connection.

Host Identity Protocol

For many years, several observers have advocated the separation of IP addresses and EIDs in the Internet architecture (e.g., [31, 32]). The Internet Research Task Force (IRTF) Name Space Research Group is presently studying the question of whether a new name space between the network and application layers would help solve architectural strains in the Internet introduced by the overloading of IP addresses to identify locations in the network, interfaces, host names, and TCP connection identifiers [33]. HIP by Moscowitz [3] is a recent formalization of such concepts. The HIP proposal suggests that a new cryptography-based name space may solve a number of problems in today's Internet, including routing table growth due to site multihoming, lightweight IPsec key establishment, and mobility management across multiple IP addressing realms. The fundamental idea is to assign a (statistically) globally unique name for any host with an IP stack. By making this name cryptography-based (a public key), this *host identity* can be used to authenticate transactions. A HIP protocol layer is effectively interposed between the IP and transport layers, allowing for decoupling of transport connections from IP addresses, and all packets carry a representation of the host identity, either implicitly or explicitly. The host identity could be stored in DNS or a public key infrastructure (PKI), or could be anonymous, in which case it can still be used to prevent connection hijacking.

HIP requires an initial four-packet "stateless" handshake [34] to set up keying material for a connection, although datagrams could be encrypted and piggybacked after the first two packets (Fig. 2b). In this respect, it operates like a simpler version of the Internet key exchange (IKE) protocol. When used with IPsec, subsequent packets do not require additional HIP overhead, since the identity can be implied by the security parameter index (SPI) carried in the IPsec-protected packets [35]. The host identity for a peer host could be obtained as part of the name resolution process. A compressed representation of the host identity, rather than the IP addresses, is used in the socket identifiers. Further details are found in [3, 36].

When using HIP, if a host changes its address during a connection, it can send a HIP Readdress packet to any HIP-enabled correspondent peer. The HIP Readdress packet contains the current ESP sequence number and SPI to provide denial-of-service and replay protection, and is authenticated with a HIP signature. Note that intermediate systems (e.g., NATs) monitoring the packet flow must also be informed of an address change, and must track the mapping of a host identity and IPsec SPI to the IP addresses. A consequence of this is that all stateful intermediate systems that use the address will have to inspect all packets for a HIP Readdress packet. With these changes, however, traversing multiple addressing realms becomes potentially simpler since the IP address is decoupled from the transport protocol in a secure manner.

Server mobility may be handled via secure DNS updates, just as in end-to-end mobility, but the HIP proposal suggests an optimization in the form of a small amount of infrastructure support called a *rendezvous server* [3]. The DNS record for the server registers the address of a rendezvous server, and mobile servers will send a HIP Readdress packet to the rendezvous server to keep it up to date with the current address (thereby removing DNS updates from the transaction). The function of a rendezvous server is to simply forward the initial HIP packet to the server's current location, after which the handshake proceeds with the current addresses. The protocol whereby this mapping is ensured is not yet developed; a generalized packet forwarding agent has also been suggested [36]. Note that this proposed rendezvous server bears resemblance to a Mobile IP home agent.

Comparison of Different Approaches

In this section we qualitatively compare the operation of Mobile IP through home agent (MIP-HA), Mobile IP with route optimization (MIP-RO), Migrate, and HIP. Our comparison looks at each approach in the context of *performance*, *security*, *deployment*, *scalability*, and *robustness*. We assume that IP address configuration for temporary addresses, as well as infrastructure authorization mechanisms for accessing a visited network are available (and are orthogonal to the comparison). We are interested in two fundamental operations:

Technique	Performance issues
MIP-HA	• Suffers from suboptimal routing (all packets travel through home agent) • Requires per-packet encapsulation overhead • Micromobility extensions that improve rapid handoff performance are being developed
MIP-RO	• Avoids suboptimal routing of MIP-HA but inherits other performance properties of MIP-HA • Incurs return routability latency/overhead/computation for each care-of address change
Migrate	• DNS records are less cacheable, leading to additional per-session packet overhead and initial latency • Non-TCP-based applications forced to use DNS queries or higher-layer approaches to detect mobility of a peer • Additional computation to generate keying material upon SYN exchange
HIP	• Requires operation with (and overhead of) IPsec Encapsulating Security Protocol (ESP) • Computational and packet overhead upon connection establishment to generate keying material (similar to IPsec) • Requires either dynamic DNS updates (like Migrate) or additional network infrastructure to allow hosts to find mobile servers.

■ Table 2. *Key performance issues.*

- Initial registration mechanisms needed to enable location by correspondent nodes
- Protocol necessary to coordinate a move to another address while active sessions may be ongoing

Performance

Some important performance considerations include per-session packet overhead and latency (e.g., cost to look up the mobile node's current location), per-packet bit overhead and latency, per-relocation packet overhead and latency, and computational load on end hosts and network infrastructure. Table 2 identifies some of the key performance issues for each approach.

Of the four approaches, MIP-HA offers the lowest performance in terms of packet route selection and additional per-packet overhead. The remaining three approaches are similar in that mobility management is performed on more of an end-to-end basis upon readdress, and some type of additional per-session overhead is required to establish security associations. However, they differ as follows:

• MIP-RO requires a return routability check through the home network for each care-of address change, in addition to the binding update sent to the home network. If this requirement is relaxed, and security associations (to set up a shared secret) can be reused across address changes, MIP-RO, Migrate, and HIP readdressing performance should be similar. If IPsec were to be used with each of the protocols, HIP and Migrate would have a small per-packet overhead advantage due to the avoidance of routing headers and destination options.

• Migrate requires a secure DNS update and a TCP Migrate handshake for each movement of a server. Although this is no more packet overhead than found in the other approaches, performance cost may come as a result of additional DNS traffic load and latency in the network for querying addresses of mobile servers.

• A *rendezvous server* has been proposed for HIP to reduce the need to dynamically update DNS upon readdress of a mobile server. The DNS record for a HIP host would then contain a pointer to the rendezvous server, which the mobile server would refresh with its current address, and the initial packet of a HIP handshake would flow through this rendezvous server. However, there is no implementation experience with or specification for this approach. Without such a server, HIP faces the same issues as Migrate with respect to DNS cacheability and updates.

• Mobile-IP-based approaches potentially offer micromobility extensions that could trump all other performance considerations for some environments. Although micromobility might be added to Migrate and HIP, the concept is fundamentally at odds with the end-to-end orientation of each proposal's design, unless packet forwarding agents can be integrated.

Security

Security and mobility management are inexorably intertwined, because mobility opens up the potential for a number of security problems, including attacks against the mechanisms of mobility (e.g., replay attacks, resource depletion attacks), host impersonation (man-in-the-middle attacks, connection hijacking), and privacy (disclosure of a node's whereabouts). Additional discussion of the various issues can be found in [19, 37]. Table 3 identifies some of the key security issues and features for each approach. All approaches fundamentally require some kind of trust relationship between the mobile node and some network infrastructure (the home agents in MIP, and DNS or an alternative key infrastructure for Migrate and HIP), but it is plausible to expect that such relationships could be provisioned. However, the challenge for all the approaches is to avoid needing a preconfigured trust relationship with all possible correspondent nodes.

Invariably, additional protocols open up the potential for additional attacks, but a stated goal for MIP and Migrate is to provide security assurances no worse than those found in the fixed networks. For example, while MIP and Migrate are susceptible to man-in-the-middle attacks, they are not more vulnerable than nonmobile transactions in the current Internet. HIP suggests that DNS or a PKI may be used as a directory to store host identities, in which case HIP exchanges would be robust to man-in-the-middle attacks. However, if a PKI were to be deployed, all approaches could make use of it, and in fact such a development could obviate certain mechanisms such as return routability.

Although all approaches are basically compatible with IPsec, current selectors for IPsec security associations assume that the packet destination addresses are fixed. In the Migrate and HIP proposals, the underlying addresses could change. In this case, either IPsec security associations would need to be renegotiated, or IPsec implementations could remove the requirement to use the destination address as an index to the security association database [35].

Deployment

Deployment of new services is increasingly complicated by backward compatibility concerns and the growth of functionality in the network that can inadvertently or intentionally block the operation of protocols. Deployment issues include required changes to end hosts, network infrastructure, and applications; ability to coexist with middleboxes (NAT and proxies); potential for auto-configuration; reliance on non-ubiquitous services (anycast, multicast); and IPv4 vs. IPv6 issues. All of the proposed mobility management approaches have deployment issues, which are summarized in Table 4.

MIP-HA has a significant deployment advantage over the other proposals in that it does not require changes to corre-

Technique	Security issues
MIP-HA	• Provides strong authentication and integrity of signaling packets via keyed MD5 (or similar algorithms) • Compatible with IP security protocols • Node mobility may be hidden from correspondent nodes (privacy)
MIP-RO	• Return routability is the current technique for securing the binding update to correspondent nodes • Otherwise inherits the security features of MIP-HA
Migrate	• Relies on security of dynamic DNS (thereby requiring shared secrets with DNS) • Uses elliptic curve Diffie-Hellman key generation to create per-session shared secrets, or otherwise IPsec • Introduces cookie mechanisms robust to TCP SYN flooding and connection hijacking
HIP	• As presently proposed, requires operation in conjunction with IPsec • Provides faster IPsec keying than Internet key exchange, albeit with loss of policy granularity • Implements "stateless" connection handshake for denial-of-service resilience • Allows for anonymous identities if privacy is desired

■ Table 3. *Key security issues.*

Technique	Deployment issues
MIP-HA	• Requires no changes to correspondent nodes or DNS • Requires deployment and administration of home agents • Operation is challenged by presence of firewalls and network address translation • Micromobility extensions require additional infrastructure deployment
MIP-RO	• Requires a return routability extension that is not yet mandatory for IPv6 nodes
Migrate	• Requires changes to both ends of TCP connections • Can be deployed as an incremental (pair-wise) capability • Deployment could be aided by proxies that support the extensions • Breaks applications that use IP addresses in the application data stream (e.g., FTP)
HIP	• Requires IPsec deployment, but otherwise can be deployed incrementally • For practical use, requires PKI or extensions to DNS, unless operated in anonymous mode • Requires changes to networking stacks and APIs at both ends of the connection • Unless applications are HIP-aware, has problems similar to Migrate if IP addresses are used in the application data stream

■ Table 4. *Key deployment issues.*

spondent nodes, and has been standardized for some time and is commercially available. However, as mentioned above, MIP deployment has been hindered by the presence of firewalls and NAT. NAT in particular poses a serious problem to the MIP architecture because many hosts may not have publicly routable home or care-of addresses. It has long been thought that a transition to IPv6 would solve this situation, but as IPv6 transition becomes less clear, and unless the IPv6 architecture has a compelling argument against the continued use of NAT, such problems may persist. Proposals that do not rely on publicly routed addresses as EIDs may have an advantage when multiple addressing realms are used, as discussed in [14].

Deployment issues are most severe for a proposal such as HIP, which requires widespread deployment of IPsec and more radical changes to the networking stacks of both connection endpoints. HIP does not strictly require a PKI to operate (HIP has an anonymous mode), but in the absence of a PKI or extensions to DNS, the security offered is similar to that of Migrate or MIP. Although HIP requires sockets to be bound to host identities and not IP addresses, it is likely that such changes can be made transparently to non-HIP-aware applications. Finally, neither the HIP nor Migrate proposals has initiated any significant IETF activity as of this writing.

Scalability

A key property of a globally deployed mobility management technique is that it scales to efficiently handle large numbers of nodes. Traditionally, scalability in the Internet is accomplished through hierarchy and by moving functionality out of the network to the end points. Mobile IP scales by widely distributing the infrastructure (home agents) needed to manage the mobility bindings, while route optimization improves scalability by reducing overhead due to suboptimal routing. Both Migrate and HIP scale by adopting an end-to-end solution, but they also leverage the DNS that achieves scalability by the hierarchy inherent in domain names. Micromobility techniques of MIP can also contribute to scalability in masking rapid node movements from the rest of the network. Table 5 summarizes some of the key scalability issues of each approach.

An open issue is whether it is wise to extend DNS to handle mobility. DNS scales through the use of hierarchy and caching, but DNS-centric approaches to mobility require a reduction in the TTL for records for mobile nodes, and hence a greater load on root nameservers. Research work is inconclusive on this point [38, 39]. Content delivery networks have demonstrated that DNS entries for large numbers of objects can successfully be updated on timescales of seconds. However, when DNS records change, a secure DNS server must cryptographically sign its zones before sending the updates to peer servers, an operation that can take substantial time (minutes) for large zones. Compounding this concern is the possible increase in the number of hostnames if peer-to-peer application usage increases and many more hosts require hostnames.

Presently, IP addresses are often aggregated for purposes of reduced administration; for example, an access control list may contain a network address and mask. If host identities are used to identify hosts, the loss of host aggregation based on network prefix could cause scaling problems for these types of access control lists. One possible solution would be to instead aggregate based on domain names [33].

Technique	Scalability issues
MIP-HA	• Uses wide distribution of home agents (in each home network) to handle the many mobility bindings • Per-packet operations could be implemented in hardware, reducing expensive overhead • Plans to develop micromobility techniques to reduce the scope of frequent binding updates
MIP-RO	• More scalable than MIP-HA in terms of bandwidth usage and avoiding home network bottlenecks
Migrate	• Impact of reduced TTL caching is unknown • Frequent updates of large DNS zones may lead to scalability issues
HIP	• Also has DNS scalability issues, unless proposed rendezvous server or additional dynamic infrastructure is developed

■ Table 5. *Key scalability issues.*

Technique	Robustness issues
MIP-HA	• Home agents and/or networks could be single points of failure • Handles simultaneous mobility through use of the home agent as an anchor point
MIP-RO	• Similar to MIP-HA
Migrate	• Requires extensions to support simultaneous node mobility, such as described in [40] • Inherits fault tolerance properties of DNS • End-to-end approaches less sensitive to asymmetric routing
HIP	• Similar to Migrate; thorough robustness comparison not possible until protocol approach is completed

■ Table 6. *Key robustness issues.*

Robustness

Robustness issues include fault tolerance, ability to handle simultaneous node mobility, ability to handle irregularities in routing such as unidirectional links and asymmetric routing, and multihoming capability. In some sense, all proposals are vulnerable to disconnection from a home network, whether through home agent-based routing (Mobile IP) or through home DNS servers (Migrate and HIP), and mobile nodes may be oblivious to failures in the infrastructure until a reregistration event occurs. Table 6 summarizes additional issues for the four approaches.

A related problem is host multihoming (an increasingly common situation in which a host can have multiple active network interfaces to improve connectivity). The use of a home address as an EID has interesting ramifications. On one hand, MIP admits a solution in that all interfaces other than the primary can be considered as visited addresses, but this implies that the home network may become a (suboptimally routed) conduit for packets sent to nonprimary interfaces. Alternatively, a mobile node may manage multiple home agents (one for each interface). Schemes not based on use of a particular IP address as an EID appear to have more natural solutions to host multihoming, but work in this area is still early, and other problems exist (e.g., how a client with only default routing information can select the appropriate destination address).

Finally, an important consideration is whether the approach accommodates the mobility of entire subnets. With end-to-end approaches such as HIP and Migrate, such mobility would seem to require either renumbering of the attached subnet, NAT for the subnet, or injection of a foreign routing prefix. In contrast, Mobile IP allows for a mobile subnet to mask the mobility from the attached hosts via tunneling [41].

Summary

In this article we have provided an overview and qualitative comparison of three potential approaches to host mobility for IP networks. Work in the IETF has focused on a solution known as Mobile IP, which has certain advantages and disadvantages described above and summarized in Table 7. In light of the slow deployment of Mobile IP, several alternative solutions have recently been proposed, and we have examined two such proposals (Migrate and HIP) in more detail herein.

We have drawn the following conclusions from our comparison:

- Mobile IP is a much more complete solution than either Migrate or HIP. While some of the completeness is due to the large head start Mobile IP has had, the ability to support micromobility performance enhancements and mobile subnets is fundamentally difficult to provide using end-to-end solutions. The cost of these enhancements is the addition of complexity to the network infrastructure, but this is a cost that the IETF has already accepted by developing and advocating Mobile IP. An interesting exercise would be to determine whether micromobility strategies could be extended to also complement the otherwise end-to-end solutions of Migrate and HIP.
- While Migrate and HIP offer clear performance advantages over MIP-HA, the operational differences between them and MIP-RO, on an end-to-end basis, do not appear to be significant. If competing against MIP-RO, these alternatives must therefore offer additional features, or be usable in a complementary fashion. HIP may, for example, offer a cleaner solution to binding update authentication than currently offered by return routability. HIP also could be of interest in networks that have higher degrees of host multihoming, a requirement for a distributed keying infrastructure not solely due to mobility management considerations, and hosts that may not naturally have a home network. More development of HIP is necessary to better assess its potential to complement Mobile IP in this way.
- Despite the above considerations, the success of the above proposals probably depends in large part on the evolutionary path of the Internet. If IPv6 receives widespread deployment in the next few years, the performance characteristics and feature set of MIPv6 will be hard to beat. If, however, IPv4 continues to persist, or if network address translation does not die out, the picture is much less clear, and alternatives that do not rely on use of a publicly routable address as an endpoint identifier may ultimately prevail.

Approach	Strengths	Weaknesses
Mobile IP (MIP-HA and/or MIP-RO)	• Does not require bilateral deployment of host modifications • Can support mobile subnetworks of nodes that individually cannot or do not desire to change addresses dynamically (e.g., a ship at sea) • More naturally supports simultaneous mobility of both communicating nodes • Micromobility support being developed • Longer history of research and development	• Tunneling (IPv4) and routing headers (IPv6) lead to additional per-packet overheads • Operation in networks with multiple addressing realms • Tunneling can conflict with firewall and IPsec security policies (IPv4) • Security relationships more complicated by third-party (e.g., home, foreign) agents in the network
Migrate	• Better path selection (over MIP-HA) • Potentially easier integration with NATs and firewalls • No tunneling or additional per-packet overhead is incurred • Does not require additional network infrastructure	• Requires changes to TCP implementation at both ends of the connection • Raises concern about DNS scalability due to loss of caching and increased DNS database distribution frequency • TCP-centric
HIP	• Better path selection (over MIP-HA) • No per-packet overhead beyond that of IPsec • More natural operation with multiple addressing realms • Tightly integrated with IP security protocols • More natural solution to multihoming	• Little implementation and operational experience with this approach • Significant deployment barriers, including widespread IPsec deployment • Lacks micromobility, mobile router, simultaneous node movement capabilities • High overhead (handshake) for short transactions

■ Table 7. *Strengths and weaknesses of different mobility management solutions.*

References

[1] C. Perkins, *Mobile IP Design Principals and Practices*, Addison Wesley, 1997.
[2] A. Snoeren and H. Balakrishnan, "An End-to-End Approach to Host Mobility," *Proc. ACM MOBICOM*, Aug. 2000, pp. 155–66.
[3] R. Moscowitz, "Host Identity Payload and Protocol," Internet draft, draft-moscowitz-hip-05.txt, expired; http://homebase.htt-consult.com/~hip/, Nov. 2001.
[4] C. Perkins, Ed., *Mobile Ad-Hoc Networking*, Addison Wesley, 2000.
[5] C. Huitema, "Multi-Homed TCP," Internet draft, expired; http://www.cs.ucla.edu/~bzhang/etcp/huitema-TCP.txt, May 1995.
[6] D. Maltz and P. Bhagwat, "MSOCKS: An Architecture for Transport Layer Mobility," *Proc. IEEE INFOCOM*, Mar. 1998, pp. 1037–45.
[7] R. Stewart *et al.*, "Stream Control Transmission Protocol," IETF RFC 2960, Oct. 2000.
[8] P. Conrad *et al.*, "SCTP in Battlefield Networks," *Proc. IEEE MILCOM*, Oct. 2001, pp. 289–95.
[9] H. Schulzrinne and E. Wedlund, "Application-Layer Mobility Using SIP," *Mobile Comp. Commun. Rev.*, vol. 4, no. 3, July 2000, pp. 47–57.
[10] V. Zandy and B. Miller, "Reliable Network Connections," *Proc. ACM MOBICOM*, Sept. 2002, pp. 95–106.
[11] B. Raman, R. Katz, and A. Joseph, "Universal Inbox: Providing Extensible Personal Mobility and Service Mobility in an Integrated Communication Network," *Wksp. Mobile Comp. Sys. and Apps.*, Dec. 2000.
[12] I. Castineyra, N. Chiappa, and M. Steenstrup, "The Nimrod Routing Architecture," IETF RFC 1992, Aug. 1996.
[13] R. Ramanathan, "Mobility Support for Nimrod: Challenges and Solution Approaches," IETF RFC 2103, Feb. 1997.
[14] P. Francis and R. Gummadi, "IPNL: A NAT-Extended Internet Architecture," *Proc. ACM SIGCOMM*, Aug. 2001, pp. 69–80.
[15] W. Adjie-Winoto *et al.*, "The Design and Implementation of an Intentional Naming System," *Proc. ACM SOSP*, Dec. 1999, pp. 186–201.
[16] J. Mysore and V. Bharghavan, "A New Multicasting-Based Architecture for Internet Host Mobility," *Proc. ACM MOBICOM*, Sept. 1997, pp. 161–72.
[17] C. Perkins, Ed., "IP Mobility Support for IPv4," RFC 3344, Aug. 2002.
[18] P. Bhagwat, C. Perkins, and S. Tripathi, "Network Layer Mobility: An Architecture and Survey," *IEEE Pers. Commun.*, vol. 3, no. 3, June 1996, pp. 54–64.
[19] D. Johnson *et al.*, "Mobility Support in IPv6," Internet draft, work in progress, draft-ietf-mobileip-ipv6-20.txt, Jan. 2003.
[20] C. Perkins, "Mobile IP and the IETF," *ACM Mobile Comp. Commun. Rev.*, vol. 6, no. 2, Apr. 2002, pp. 3–8.
[21] A. Campbell *et al.*, "Comparison of IP Micromobility Protocols," *IEEE Wireless Commun.*, vol. 1, no.2, Feb. 2002, pp. 72–82.
[22] F. Chiussi *et al.*, "Mobility Management in Third-Generation All-IP Networks," *IEEE Commun. Mag.*, vol. 40, no. 9, Sept. 2002, pp. 124–35.
[23] C. Rigney *et al.*, "Remote Authentication Dial In User Service (RADIUS)," IETF RFC 2058, Jan. 1997.
[24] S. Glass *et al.*, "Mobile IP Authentication, Authorization, and Accounting Requirements," IETF RFC 2977, Oct. 2000.
[25] M. Cappiello, A. Floris, and L. Veltri, "Mobility amongst Heterogeneous Networks with AAA Support," *Proc. IEEE ICC*, vol. 4, June 2002, pp. 2064–69.
[26] R. Jain *et al.*, "Enhancing Survivability of Mobile Internet Access using Mobile IP with Location Registers," *Proc. IEEE INFOCOM*, vol. 1, Mar. 1999, pp. 3–11.
[27] P. Nikander *et al.*, "Homeless Mobile IPv6," Internet draft, expired, draft-nikander-mobileip-homeless6-01.txt; http://www.tml.hut.fi/~pnr/publications/draft-nikander-mobileip-homelessv6-01.txt, Feb. 2001.
[28] B. Awerbuch and D. Peled, "Concurrent On-Line Tracking on Mobile Users," *Proc. ACM SIGCOMM*, Sept. 1991, pp. 221–33.
[29] J. Saltzer, D. Reed, and D. Clark, "End-to-End Arguments in System Design," *ACM Trans. Comp. Sys.*, vol. 2, no 4., Nov. 1984, pp. 277–88.
[30] A. Snoeren, H. Balakrishnan, and M. F. Kaashoek, "Reconsidering Internet Mobility," *Proc. 8th Wksp. Hot Topics in Op. Sys.*, May 2001.
[31] J. Saltzer, "On the Naming and Binding of Network Destinations," *Local Comp. Networks*, North Holland, 1982, pp. 311–17; also available as IETF RFC 1498, Aug. 1993.
[32] J. Chiappa, "Endpoints and Endpoint Names: A Proposed Enhancement to the Internet Architecture," unpublished note: http://users.exis.net/~jnc/tech/endpoints.txt, 1999.
[33] E. Lear and R. Droms, "What's In a Name: Thoughts from the NSRG," Internet draft, work in progress, draft-irtf-nsrg-report-08.txt, Dec. 2002.
[34] T. Aura and P. Nikander, "Stateless Connections," *Proc. ICICS*, Nov. 1997, pp. 87–97.
[35] S. Bellovin, "EIDs, IPsec, and HostNAT," Presentation at 41st IETF mtg., Los Angeles, CA, Mar. 1998.
[36] P. Nikander *et al.*, "Integrating Security, Mobility, and Multi-homing in a HIP Way," *Proc. Net. and Distrib. Sys. Security Symp.*, Feb. 2003.
[37] S. Mink *et al.*, "Towards Secure Mobility Support for IP Networks," *Proc. IFIP ICCT*, Aug. 2000, pp. 555–62.
[38] J. Jung *et al.*, "DNS Performance and the Effectiveness of Caching," *IEEE Trans. Net.*, vol. 10, no. 5, Oct. 2002, pp. 589–603.
[39] A. Shaikh *et al.*, "On the Effectiveness of DNS-based Server Selection," *Proc. IEEE INFOCOM*, Apr. 2001, pp. 1801–10.
[40] S. Tilak and N. Abu-Ghazaleh, "A Concurrent Migration Extension to an End-to-End Mobility Architecture," *ACM Mobile Comp. Commun. Revi.*, vol. 5, no. 3, 2002, pp. 26–31.
[41] J. Kim *et al.*, "Demonstration of Static Network Mobile Router for Mobile Platforms," *Proc. IEEE MILCOM*, Oct. 2001.

Biographies

THOMAS R. HENDERSON (thomas.r.henderson@boeing.com) is an associate technical fellow at Boeing Phantom Works. He holds a Ph.D. degree in electrical engineering and computer science from the University of California at Berkeley, and M.S. and B.S. degrees in electrical engineering from Stanford University. His present research interests are focused on efficient packet routing and transport over mobile wireless and satellite networks.