Scalable support for transparent mobile host internetworking *

David B. Johnson

Computer Science Department, Carnegie Mellon University, Pittsburgh, PA, USA

Abstract. This paper considers the problem of providing transparent support for very large numbers of mobile hosts within a large internetwork such as the Internet. The availability of powerful mobile computing devices and wireless networking products and services is increasing dramatically, but internetworking protocols such as IP used in the Internet do not currently support host movement. To address this need, the Internet Engineering Task Force (IETF) is currently developing protocols for mobile hosts in the Internet. This paper analyzes the problem to be solved, reviews the current state of that effort, and discusses its scalability to very large numbers of mobile hosts in a large internetwork.

1. Introduction

The global Internet is growing at a tremendous rate. There are now about 5 million hosts connected to the Internet, and this number is doubling approximately every year. The average time between new networks connecting to the Internet is about 10 minutes. Initiatives such as the National Information Infrastructure and the increasing commercial uses of the Internet are likely to create even faster growth in the future.

At the same time, portable computing devices such as laptop and palmtop computers are becoming widely available at very affordable prices, and many new wireless networking products and services are becoming available based on technologies such as spread-spectrum radio, infrared, cellular, and satellite. Mobile computers today often are as capable as many home or office desktop computers and workstations, featuring powerful CPUs, large main memories, hundreds of megabytes of disk space, multimedia sound capabilities, and color displays. High-speed local area wireless networks are commonly available with speeds up to 2 megabits per second, and wide-area wireless networks are available that provide metropolitan or even nationwide service.

With these dramatic increases in portability and ease of network access, it becomes natural for users to expect to be able to access the Internet at any time and from anywhere, and to transparently remain connected and continue to use the network as they move about. However, internetworking protocols such as IP [23] used in the Internet do not currently support host mobility. A mobile user, today, must generally change IP addresses when connecting to the Internet at a different point or through a different network; the user must modify a number of configuration files and restart all network connections, making host movement difficult, time consuming, and error prone.

To address this need in the Internet, the Mobile IP Working Group of the Internet Engineering Task Force (IETF) has been working over the past few years to develop standard protocols to support mobile hosts operating in the Internet [6,7,8,9,10,13,14,17,18,19,20, 21,22,25,29,30,31,32,34]. This work represents the contributions of many people within the Working Group, and development of these protocols is still underway. This paper analyzes the problem to be solved, reviews the current state of that effort, and discusses its scalability to very large numbers of mobile hosts in a large internetwork.

Section 2 of this paper describes the general problem of mobility management and packet routing to mobile hosts in a large internetwork. Section 3 gives a summary of the current state of the basic IETF Mobile IP protocol, and section 4 describes extensions to this protocol also being developed within the IETF for optimizing packet routing to mobile hosts. Section 5 discusses the scalability of this work to very large numbers of mobile hosts, and section 6 presents conclusions.

2. Problem analysis

2.1. Internetwork routing

In order to provide scalable routing support, internetworking protocols such as IP [23], ISO CLNP [27], NetWare IPX [33], and AppleTalk [28], use *hierarchical* addressing and routing schemes. For example, in IP, the network address of a host is divided into two levels of hierarchy: a *network number* identifying the network to which the host is connected, and a *host number* identifying the particular host within that network. Routers within the Internet know (and care) only how to route packets based on the network number of the destination address in each packet; once the packet reaches that net-

* This research was supported in part by the Wireless Initiative of the Information Networking Institute at Carnegie Mellon University.

work, it is then delivered to the correct individual host on that network.

Aggregating the routing decision at each level of the hierarchy in this way reduces the size of the routing tables that each router must maintain, reduces the size of the routing updates that routers must exchange, and simplifies the decisions at each router. Hierarchical addressing and routing has proven to be essential to keep up with the exponential growth of the Internet, in particular. The original two-level hierarchy of Internet addressing in IP has already been transparently extended at the bottom with *subnetting* [16] and at the top through use of *CIDR* [5]. In the IETF's "IPng" effort to develop the next generation of the IP protocol [2], support for many more levels of hierarchy than in the present version of IP is an explicit design goal [15].

However, this hierarchy in addressing and routing prevents packets from being routed correctly to a mobile host while it is away from its home network. Since a host's address logically encodes its location, without special handling for mobility, packets addressed to a mobile host will be routed by the Internet only to the mobile host's home network. This problem exists with any protocol using a hierarchical addressing and routing scheme, whether the hierarchy is provider-based or geographical.

2.2. Location registry

It is important to be able to support packet routing to mobile hosts from existing correspondent hosts that have not been modified to support mobility. Given the very large number of hosts already deployed within the Internet, it seems quite likely that some will not be upgraded to support mobility for some time. Furthermore, some existing hosts may never be upgraded, for example because the organizations owning some hosts may lack the interest or resources to upgrade, or because the original vendor no longer offers support for particular products owned by some customers. The ability to support unmodified correspondent hosts also allows any correspondent host to communicate with any other host without being concerned whether or not it is currently mobile and away from its home network.

It therefore becomes logical to provide basic mobility support for a mobile host through a location registry recording the mobile host's current location, that can be accessed through the mobile host's home network. An unmodified correspondent host (or one that simply does not know that a particular mobile host is in fact mobile) will send IP packets for that mobile host in the same way as all IP packets are sent today. Such packets will thus reach the mobile host's home network, where they may be intercepted by some mobility support agent and forwarded to the mobile host's current location.

Requiring the sender to instead explicitly query the location registry before sending a packet is incompatible with the goals of supporting existing unmodified correspondent hosts and of not requiring the sender to be aware of whether a particular destination host is currently mobile. Accessing the location registry through the mobile host's home network also avoids any requirement for changes to the basic routing algorithms of the Internet, and allows each organization owning some network to manage this functionality for all of its own mobile hosts with this home network, improving scalability and easing manageability.

In addition, requiring the location registry to be explicitly queried in this way, either requires this overhead to be added for *all* destination addresses or requires restrictions on the assignment of IP addresses. If a host's address encodes information as to whether it is a *mobile* or a *stationary* host, then only packets destined for mobile host's need to cause the location registry to be queried. However, this encoding would require permanently designating each host into one of these two classes, greatly reducing flexibility and complicating host and network administration.

2.3. Packet tunneling

Some mechanism is needed to cause a packet addressed to a mobile host to be routed to that host's current location rather than (only) to its home network. In order to avoid distributing routing information for a mobile host throughout the Internet so that the new routing decision could be made at each hop, it must be possible to modify each packet for a mobile host in such a way that the routing infrastructure of the Internet will route the modified packet to a location identified in the packet. This type of packet forwarding is known as *tunneling*. For IP, tunneling may be done using an encapsulation protocol, or through an IP option such as loose source routing [23].

In tunneling a packet from one node to another, only these two nodes (the two endpoints of the tunnel) need know that tunneling is taking place. Routers between the node tunneling the packet and the new destination node to which the packet is tunneled simply route the packet at each hop in the same way as any ordinary IP packet. There is thus no need to modify existing routers, such as within the Internet backbone, nor to modify existing Internet routing algorithms.

2.4. Caching and consistency

The mechanisms suggested above allow packets for a mobile host to be sent to it at its current location, but support forwarding only through an agent on the mobile host's home network. For example, if a mobile host, say MH1, is visiting some network, even packets from a correspondent host on this same network must be routed through the Internet to this agent on MH1's home network, only to then be tunneled back to the original net-

work for delivery to MH1. If the correspondent host in this example is actually another mobile host, say MH2, then packets from MH1 to MH2 must likewise be routed through some agent on MH2's home network and back to the original network for delivery to MH2. This indirect routing places unnecessary overhead on the Internet, on each mobile host's home network, and on the agent providing forwarding service from each home network. Such indirect routing may also significantly increase the latency in packet delivery to a mobile host.

Correspondent hosts that have been modified to support mobility should be able to learn the current location of a mobile host with which they are communicating, and to then use this location to tunnel their own future packets directly to the mobile host. By caching this location, the expense of discovering this location can be avoided on each individual packet sent to the mobile host. However, this caching creates the problem of *cache consistency* when the mobile host then moves to a new location, since the correspondent host's cache will still point to the old location. In order to support smooth handoff from one location to another, the protocol must be able to update correspondent host's caches, and should provide some support for packets that may be tunneled based on a temporarily out-of-date cache.

3. The basic IETF Mobile IP protocol

This section provides an overview of the current state of the basic IETF Mobile IP protocol [20]. The protocol provides transparent routing of packets to a mobile host and requires no modification to existing routers or correspondent hosts. No support is provided, however, for caching a mobile host's location at correspondent hosts or for allowing correspondent hosts to tunnel packets directly to a mobile host's current location. These features are being developed within the IETF as a separate set of extensions to this basic protocol, and are discussed in section 4.

3.1. Infrastructure

Each mobile host is assigned a unique *home address* in the same way as any other Internet host, within its *home network*. Hosts communicating with a mobile host are known as *correspondent hosts* and may, themselves, be either mobile or stationary. In sending an IP packet to a mobile host, a correspondent host always addresses the packet to the mobile host's home address, regardless of the mobile host's current location.

Each mobile host must have a *home agent* on its home network that maintains a registry of the mobile host's current location. This location is identified as a *care-of address*, and the association between a mobile host's home address and its current care-of address is called a *mobility binding*, or simply a *binding*. Each time the mobile host establishes a new care-of address, it must *register* the new binding with its home agent so that the home agent always knows the current binding of each mobile host that it serves. A home agent may handle any number of mobile hosts that share a common home network.

A mobile host, when connecting to a network away from its home network, may be assigned a care-of address in one of two ways. Normally, the mobile host will attempt to discover a *foreign agent* within the network being visited, using an *agent discovery* protocol. The mobile host then *registers* with the foreign agent, and the IP address of the foreign agent is used as the mobile host's care-of address. The foreign agent acts as a local forwarder for packets arriving for the mobile host and for all other locally visiting mobile hosts registered with this foreign agent. Alternatively, if the mobile host can obtain a temporary local address within the network being visited (such as through DHCP [4]), the mobile host may use this temporary address as its care-of address.

While a mobile host is away from its home network, a mobile host's home agent acts to forward all packets for the mobile host to its current location for delivery locally to the mobile host. Packets addressed to the mobile host that appear on the mobile host's home network must be intercepted by the mobile host's home agent, for example by using "proxy" ARP [24] or through cooperation with the local routing protocol in use on the home network.

For each such packet intercepted, the home agent tunnels the packet to the mobile host's current care-of address. If the care-of address is provided by a foreign agent, the foreign agent removes any tunneling headers from the packet and delivers the packet locally to the mobile host by transmitting it over the local network on which the mobile host is registered. If the mobile host is using a locally obtained temporary address as a care-of address, the tunneled packet is delivered directly to the mobile host.

Home agents and foreign agents may be provided by separate nodes on a network, or a single node may implement the functionality of both a home agent (for its own mobile hosts) and a foreign agent (for other visiting mobile hosts). Similarly, either function or both may be provided by any of the existing IP routers on a network, or they may be provided by separate support hosts on that network.

3.2. Agent discovery

The *agent discovery* protocol operates as a compatible extension of the existing ICMP *router discovery* protocol [3]. It provides a means for a mobile host to detect when it has moved from one network to another, and for it to detect when it has returned home. When moving into a new foreign network, the agent discovery

protocol also provides a means for a mobile host to discover a suitable foreign agent in this new network with which to register.

On some networks, depending on the particular type of network, additional link-layer support may be available to assist in some or all of the purposes of the agent discovery protocol. A standard protocol must be defined for agent discovery, however, at least for use on networks for which no link-layer support is available. By defining a standard protocol, mobile hosts are also provided with a a common method for agent discovery that can operate in the same way over all types of networks. If additional link-layer support is available, it can optionally be used by mobile hosts that support it to assist in agent discovery.

Home agents and foreign agents periodically advertise their presence by multicasting an *agent advertisement* message on each network to which they are connected and for which they are configured to provide service. Mobile hosts listen for agent advertisement messages to determine which home agents or foreign agents are on the network to which they are currently connected. If a mobile host receives an advertisement from its own home agent, it deduces that it has returned home and registers directly with its home agent. Otherwise, the mobile host chooses whether to retain its current registration or to register with a new foreign agent from among those it knows of.

While at home or registered with a foreign agent, a mobile host expects to continue to receive periodic advertisements from its home agent or from its current foreign agent, respectively. If it fails to receive a number of consecutive expected advertisements, the mobile host may deduce either that it has moved or that its home agent or current foreign agent has failed. If the mobile host has recently received other advertisements, it may attempt registration with one of those foreign agents. Otherwise, the mobile host may multicast an *agent solicitation* message onto its current network, which should be answered by an agent advertisement message from each home agent or foreign agent on this network that receives the solicitation message.

3.3. Registration

Much of the basic IETF Mobile IP protocol deals with the issue of registration with a foreign agent and with a mobile host's home agent. When establishing service with a new foreign agent, a mobile host must register with that foreign agent, and must also register with its home agent to inform it of its new care-of address. When instead establishing a new temporarily assigned local IP address as a care-of address, a mobile host must likewise register with its home agent to inform it of this new address. Finally, when a mobile host returns to its home network, it must register with its home agent to inform it that it is no longer using a care-of address.

To register with a foreign agent, a mobile host sends a *registration request* message to the foreign agent. The registration request includes the address of the mobile host and the address of its home agent. The foreign agent forwards the request to the home agent, which returns a *registration reply* message to the foreign agent. Finally, the foreign agent forwards the registration reply message to the mobile host. When registering directly with its home agent, either when the mobile host has returned home or when using a temporarily assigned local IP address as its care-of address, the mobile host exchanges the registration request and reply messages directly to its home agent.

Each registration with a home agent or foreign agent has associated with it a *lifetime* period, negotiated during the registration. After this lifetime period expires, the mobile host's registration is deleted. In order to maintain continued service from its home agent or foreign agent, the mobile host must re-register within this period. The lifetime period may be set to infinity, in which case no re-registration is necessary.

3.4. Registration authentication

All registrations with a mobile host's home agent must be authenticated in order to guard against malicious forged registrations that could arbitrarily redirect future packets destined to a mobile host. In particular, without authentication, an attacker could register a false care-of address for a mobile host, causing the mobile host's home agent to misroute packets destined for the mobile host. An attacker could, for example, reroute the mobile host's packets in order to eavesdrop on its traffic, alter any packets destined for the mobile host, or deny service to the mobile host by misdirecting its packets. Registration authentication must verify that the registration request legitimately originated with the mobile host, that it has not been altered in transit to the home agent, and that an old registration request is not being replayed (perhaps long after the mobile host was at that care-of address).

Although any authentication algorithm shared by a mobile host and its home agent may be used, the IETF protocol defines a standard authentication algorithm based on the MD5 message-digest function [26], using a secret key shared between these two nodes. MD5 is a *one-way* hash function, in that it is considered to be computationally infeasible to discover the input to the hash function given its output, or to find another sequence of input that produces the same output. A "keyed MD5" algorithm is used, in which the MD5 hash over the bytes of the shared secret key and the important fields of the message is included in each registration message or reply; the secret key itself is not included in the message sent over the network. This authentication value allows the receiver to verify the source of the message and the fact that none of the important fields in the message have

been changed since the message was sent. If the hash matches at the receiver, the registration message must have been generated by a node knowing the secret key and must not have been modified in transit; without knowledge of the secret key included in the MD5 hash, no other node can modify or forge a registration message.

Administration of the shared secret key is fairly simple, since both the mobile host and its home agent are owned by the same organization (both are assigned IP addresses in the home network owned by that organization). Manual configuration of the shared key may be performed, for example, any time the mobile host is at home, while other administration of these nodes is being performed.

Replay protection for registration messages may be provided under the IETF Mobile IP protocol using either *nonces* or *timestamps*. Using nonces, the home agent generates a random value and returns it to the mobile host (in cleartext) in its registration reply message, and the mobile host must include this same value in its next registration request message. If the value in the message does not match on the next registration attempt, for example because the mobile host has lost its saved state containing this value, the home agent returns a registration error and includes the correct new value in the registration reply. The next registration attempt by the mobile host should then succeed, and no other node can use this value in the message to forge a registration message, since it does not know the share secret key used in the message authentication that must be computed and included in each registration message. The use of timestamps for replay protection is similar, except that the timestamp included in the registration message must closely match the current time at the receiver.

3.5. Tunneling

The Mobile IP protocol allows the use of any tunneling method shared between a mobile host's home agent and its current foreign agent (or the mobile host itself when a temporary local IP address is being used). During registration with its home agent, a list of supported tunneling methods is communicated to the home agent. For each packet later tunneled to the mobile host, the home agent may use any of these supported methods.

The protocol requires support for "IP in IP" encapsulation for tunneling, as illustrated in Fig. 1. In this method, to tunnel an IP packet, a new IP header is wrapped around the existing packet; the source address in the new IP header is set to the address of the node tunneling the packet (the home agent), and the destination address is set to the mobile host's care-of address. The new header added to the packet is shaded in gray in Fig. 1. This type of encapsulation may be used for tunneling any packet, but the overhead for this method is

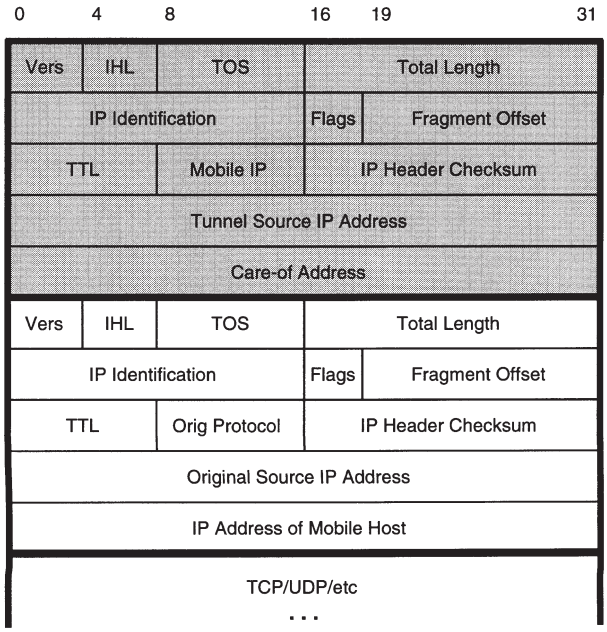

Fig. 1. Mobile IP tunneling using "IP in IP" encapsulation.

the addition of an entire new IP header (20 bytes) to the packet.

Support is also recommended for a more efficient "minimal" tunneling protocol [10,12], which adds only 8 or 12 bytes to each packet. This type of tunneling protocol is illustrated in Fig. 2, with the new header added to the packet shaded in gray. Here, only the modified fields of the original IP header are copied into a new forwarding header added to the packet between the original IP header and any transport-level header such as TCP or UDP. The fields in the original IP header are then replaced such that the source address is set to the address of the node tunneling the packet (only if the packet is

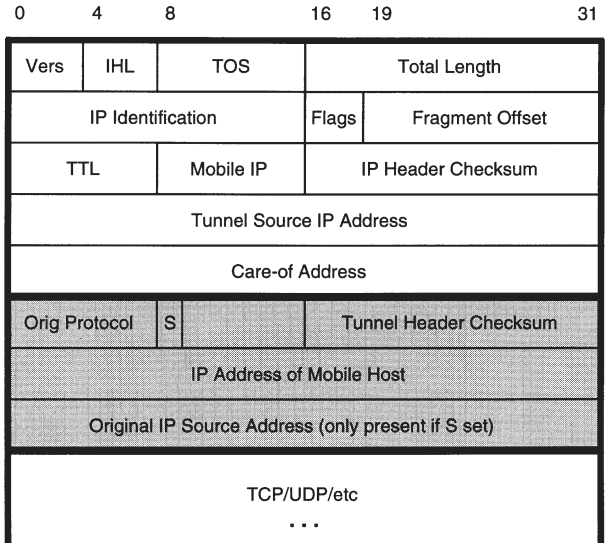

Fig. 2. Mobile IP tunneling using the minimal tunneling protocol.

being tunneled by a node other than the original sender), and the destination address is set to the mobile host's care-of address. This type of encapsulation adds less overhead to each packet, but it cannot be used with packets that have already been fragmented by IP, since the small forwarding header does not include the fields needed to represent that the original packet is a fragment rather than a whole IP packet.

4. Route optimization

The basic IETF Mobile IP protocol fulfills its primary goal of providing transparent packet routing to mobile hosts operating in the Internet. However, *all* packets for a mobile host away from home must be routed through the mobile host's home network and home agent, severely limiting the performance transparency of the protocol and creating a significant bottleneck to potential scalability.

As suggested in section 2, what is needed is the ability for correspondent hosts to be able to cache the location of a mobile host and to then tunnel packets directly to the mobile host at its current location. This functionality has become known within the IETF as *route optimization*, and a group consisting of Andrew Myles of Macquarie University, Charles Perkins of IBM, and the author have been working particularly to develop this functionality within the IETF protocol [14]. This section provides an overview of the current state of the protocol extensions for route optimization.

4.1. Location caching

Any node may optimize its own communication with mobile hosts by maintaining a *location cache* in which it caches the binding of one or more mobile hosts. When sending a packet to a mobile host, if the sender has a location cache entry for this mobile host, it may tunnel its own packet directly to the care-of address indicated in the cached binding. Likewise, a router when forwarding a packet may tunnel the packet directly to the destination mobile host's care-of address if the router has an entry in its location cache for the destination IP address of the packet; such a router may thus optimize the mobile host communication for a group of nodes not supporting the route optimization extensions.

In the absence of any location cache entry, packets destined for a mobile host will be routed to the mobile host's home network in the same way as any other IP packet, and are then tunneled to the mobile host's current care-of address by the mobile host's home agent. This is the only routing mechanism supported by the basic Mobile IP protocol. With route optimization, though, as a side effect of this indirect routing of a packet to a mobile host, the original sender of the packet is informed of the mobile host's current mobility binding (section 4.3), giving the sender an opportunity to cache the binding.

A node may create a location cache entry for a mobile host only when it has received and authenticated the mobile host's binding. Likewise, a node may update an existing location cache entry for a mobile host, such as after the mobile host has moved to a new foreign agent, only when it has received and authenticated the mobile host's new binding.

A location cache will, by necessity, have a finite size. Any node implementing a location cache may manage the space in its cache using any local cache replacement policy such as LRU. If a packet is sent to a destination address for which the cache entry has been dropped from the cache, the packet will be routed normally to the mobile host's home network and will be tunneled to the mobile host's care-of address by its home agent. As when a location cache entry is initially created, this indirect routing to the mobile host will result in the original sender of the packet being informed of the mobile host's current binding, allowing it to add this entry again to its location cache.

Optimal routing of packets from a correspondent host can be achieved if the correspondent host implements a location cache. A router implementing a location cache can also provide routing assistance for packets that it forwards from correspondent hosts that do not implement the Mobile IP route optimization extensions. For example, a local network of nodes that do not implement route optimization could be supported by a common first-hop router that maintains a location cache. Router software should be configurable, however, to allow disabling the maintenance of a location cache, such as within backbone routers, where little or no benefit of caching could be obtained.

4.2. Foreign agent handoff

When a mobile host moves and registers with a new foreign agent, the basic Mobile IP protocol does not notify the mobile host's previous foreign agent that the host has moved. After the mobile host's new registration at its home agent, IP packets intercepted by the home agent are tunneled to the mobile host's new care-of address, but any packets in flight that had already been tunneled to the old care-of address are lost and are assumed to be retransmitted by higher-level protocols if needed. The old foreign agent eventually deletes the mobile host's registration after the expiration of the lifetime period established when the mobile host registered with that foreign agent.

Route optimization extends the registration protocol to provide a means for a mobile host's previous foreign agent to be reliably notified that the mobile host has moved, and optionally to inform it of the mobile host's new binding. When registering with a foreign agent, a mobile host may establish a registration "session key"

for the duration of its registration with this foreign agent. When the mobile host later moves and registers a different care-of address, it may notify this previous foreign agent by sending it a *binding update* message; this binding update message is authenticated in the same way as registration messages between a mobile host and its home agent, but in this case, using the registration session key established when it registered with that foreign agent as the shared secret key for the authentication. Such a registration session key could also optionally be used to encrypt packets sent between the two, in order to improve privacy in the common case in which they are connected by a wireless link, but such use has not yet been considered within the IETF.

Notifying the previous foreign agent that the mobile host has moved allows packets in flight to this foreign agent, as well as packets tunneled from correspondent hosts with out-of-date location cache entries for the mobile host (they have not yet learned that the mobile host has moved), to be forwarded to the mobile host's new care-of address. This notification also allows any resources consumed by the mobile host's registration at the previous foreign agent (such as radio channel reservations) to be released immediately, rather than waiting for the mobile host's registration to expire.

By notifying the previous foreign agent of the mobile host's new binding, the previous foreign agent may create a location cache entry for the mobile host, acting as a "forwarding pointer" to its new location. Such a location cache entry at a mobile host's previous foreign agent is treated in the same way as any other location cache entry. In particular, this location cache entry may be deleted from the cache at any time. Suppose a node (such as this previous foreign agent) receives some packet that has been tunneled to this node, but this node is unable to deliver the packet locally to the destination mobile host (it is not the mobile host itself, and it does not believe that it is currently serving as a foreign agent for this mobile host). In this case, the node tunnels the packet to the mobile host's home agent, from which the packet will be re-tunneled to the mobile host's current location.

4.3. Location cache updates

When a mobile host's home agent intercepts a packet from the home network and tunnels it to the mobile host, the home agent may deduce that the original sender of the packet has no location cache entry for the destination mobile host. In this case, the home agent sends a *binding update* message to the sender, informing it of the mobile host's current binding. No acknowledgement for this binding update is needed, since any future packets intercepted by the home agent from this sender for the mobile host will serve to cause a retransmission of the update.

When a router receives a normal IP packet (not tunneled) for forwarding, if the router has a location cache entry for the destination IP address of the packet, the router may deduce that the original sender of the packet has no location cache entry for this destination mobile host. Similarly, when a node receives a packet that was tunneled to this node, if the node has a location cache entry for the destination IP address of the packet carried within the tunnel, the node may deduce that the original sender of the packet has an out-of-date location cache entry for this destination mobile host (pointing to this node). In these two cases, the node sends a *binding advice* message to the original sender of the packet, advising it to request the mobile host's current binding as a binding update from the mobile host's home agent. As with the binding update message from the home agent, no acknowledgement for this binding advice message is needed.

With the exception of the notification to a mobile host's previous foreign agent (which is sent by the mobile host itself), all binding update messages are sent by a mobile host's home agent, which is in complete control of which correspondent hosts it allows to learn the mobile host's binding. If, for any local administrative reasons, the home agent wants to keep a particular mobile host's current binding private (from all or only some correspondent hosts), it is not required to send a binding update that would otherwise be sent by the protocol.

Included in each binding update message sent by the home agent is an indication of the time remaining in the lifetime associated with the mobile host's current registration. Any location cache entry established or updated in response to this binding update must be marked to be deleted after the expiration of this period. A node wanting to provide continued service with a particular location cache entry may attempt to reconfirm that binding before the expiration of this lifetime period. Location cache entry reconfirmation may be appropriate when the node has indications (such as an open transport-level connection to the mobile host) that the location cache entry is still needed. This reconfirmation is performed by the node actively requesting the mobile host's home agent to send a new binding update message to the node.

Each node must provide some mechanism to limit the rate at which it sends binding update or binding advice messages to the same node about any given binding. Some nodes will not implement the route optimization extensions of the Mobile IP protocol, and those that do may be limited in the number of bindings they can cache or the speed with which they can process these messages. A new binding update or binding advice message should not be sent for each individual packet described above that is received over a short period of time; rather, some minimum interval should be maintained between binding update or binding advice messages, and after a small number of these messages have

been sent to the same node about some binding, the sending node must quickly increase the interval between new binding update or binding advice messages.

4.4. Location update authentication

All messages that add or change an entry in a location cache must be authenticated using the same type of authentication algorithm as is used in the basic Mobile IP protocol for registration with a mobile host's home agent (section 3.4). This authentication verifies the source of the message and ensures that none of the important fields of the message have been changed since the message was sent.

In particular, a node receiving a binding update message must verify the message's authentication before altering the contents of its location cache in response to the message. This requirement for authentication covers all binding update messages: those sent to build or update a location cache entry in response to a packet routed indirectly to a mobile host, as well as those sent to notify a mobile host's previous foreign agent that it has moved. Without such authentication, a malicious node anywhere in the Internet could forge a binding update message, allowing it to arbitrarily intercept or redirect packets destined for any other node in the Internet.

In the basic Mobile IP protocol, only a mobile host's registration with its home agent must be authenticated, allowing the simple solution of a manually configured secret key shared between the mobile host and its home agent. For route optimization, a home agent must in general be able to send an authenticated binding update message to any other node in the Internet, since any node may want to maintain a location cache containing entries for one or more mobile hosts served by that home agent. This form of general authentication is currently complicated by the lack of a standard key management or authentication protocol in the Internet, and by the lack of any generally available key distribution infrastructure; patent restrictions and export controls on the necessary cryptographic algorithms have slowed development and deployment of such facilities in the Internet.

A number of restricted authentication schemes for route optimization are possible in the short term, however, before the necessary protocols and infrastructure are available. The route optimization extensions within the IETF [14] have currently been designed to utilize manually configured shared secret keys in the same way as the authentication used in registration in the basic Mobile IP protocol, but the required shared keys may be configured to reduce the number of pairwise keys that must be maintained. In particular, by manually establishing a shared secret key with a particular home agent, a node is able to receive authenticated binding updates (and thus to maintain location cache entries) for all mobile hosts served by this home agent; if no shared secret key is available for some node, no binding update messages are sent by the home agent to that node, and only the basic Mobile IP protocol is used for packets sent to mobile hosts from that node.

This configuration of manually established shared secret keys is fairly natural, since the mobile hosts served by any particular home agent, in general, all belong to a single organization (that also owns the home agent). If the user of this node often collaborates with any number of people from this organization, establishing the shared secret key may be worthwhile. The route optimization procedures described in sections 4.2 and 4.3 have been designed with this restricted style of authentication in mind, and may be modified when more general authentication mechanisms become available.

This type of authentication is secure as long as the shared secret key remains secret, and it is not subject to export restrictions since it does not use encryption. A simpler style of authentication that also does not use encryption was proposed within the IETF for the IMHP protocol [13,17,22], and was also used in recent mobile routing work done at Harvard University [1]. This scheme relies on a general property of routing in the Internet in which nodes not connected to the normal routing path of a packet cannot eavesdrop on or reroute that packet. By including a randomly generated authenticator value in a packet sent to another node, the original sender can authenticate the reply from that node, by requiring that the same random value is returned in the reply. Although this simpler scheme requires no configuration of shared secret keys, it is less secure, since this general property of Internet routing security has been severely weakened by increasing attacks in recent years; in addition, this scheme is further weakened, since any of the links over which such an authentication may take place may be wireless, enhancing the ability of any attacker to eavesdrop on the exchange containing the authenticator value.

5. Protocol scalability

The combination of the basic IETF Mobile IP protocol described in section 3 and the extensions for route optimization described in section 4 can provide highly scalable support for packet routing to large numbers of mobile hosts in the Internet. This section considers the different factors affecting the scalability of the protocol.

5.1. The home network

Each organization owning an IP network supports all mobile hosts for which this is the home network. As new networks are added to the Internet, each deploys its own home agent to support its own mobile hosts. This arrangement allows mobility support within the home

network to scale as new organizations and new networks connect to the Internet, avoiding any centralized support bottleneck. Since a home agent maintains the location registry and tunnels packets only for the mobile hosts for which this is the home network, this approach allows these functions to scale with the number of networks containing mobile hosts.

Each organization may also control the level of expense or effort which they expend to support their own mobile hosts, and their own mobile hosts directly benefit from these expenditures. For example, an organization wanting to provide higher performance or more reliable access to the home agent for any of its mobile hosts may install higher bandwidth or additional links connecting their own home network to the Internet. The functionality of the home agent may also be replicated or distributed on multiple nodes on the home network; as long as a consistent view of the bindings of this home network's mobile hosts is maintained, such arrangements are entirely at the option of the organization owning the network and need not affect other nodes within the Internet. The home agent functionality and the home network may be scaled to support any number of mobile hosts owned by this organization.

While a mobile host is at home, it is treated in the same way as any ordinary IP host, and no overhead is added to packets sent to it while at home. When the mobile host leaves home and registers a care-of address, its home agent begins tunneling packets for it, location cache entries are gradually created at different correspondent hosts or routers, and they then begin tunneling packets for the mobile host directly to the mobile host's current location. As the mobile host moves from one care-of address to another, the location caches are updated as needed. When the mobile host later returns home, this same mechanism causes these location cache entries to be deleted; packets destined to this mobile host are then sent in the same way as any IP packets sent to an ordinary stationary host that has never been mobile.

It thus becomes feasible to upgrade all hosts, at any convenient time, to be "mobile capable," with no performance penalty to the network or to the host for the extra capability of being mobile [11]. Any mobile capable host could then become mobile at any future time as needed simply by leaving its home network and registering elsewhere. This property simplifies the installation of new hosts, since no decision need be made as to whether each host will need to be mobile at any future time.

5.2. The foreign network

Each organization owning an IP network that allows mobile hosts to visit deploys its own foreign agent to support mobile hosts visiting that network. This arrangement allows mobility support within the foreign network to scale as new organizations and new networks connect to the Internet. Since a foreign agent maintains a list of only those mobile hosts currently registered with it, and only locally delivers packets for these mobile hosts, this approach allows these functions to scale with the number of networks that allow mobile hosts to visit.

In addition, each organization owning an IP network allowing mobile hosts to visit may control its own resource allocation within that network as needed by any local policies of that organization. For example, a foreign agent may be configured to limit the number of simultaneous visitors that it allows to register; if additional mobile hosts request registration, the foreign agent may return an error to each indicating that registration has been denied due to local resource allocation limits. Any organization may install additional or more powerful foreign agents or higher bandwidth local networks in order to provide any desired level of support for visiting users. Each organization may also impose any administrative policies on the provision of service to visiting mobile hosts. For example, they may only allow mobile hosts for which prior billing arrangements have been established to register.

By deploying one or more foreign agents, the protocol places no new demands on IP address space allocation, avoiding the limits to scalability that would otherwise be imposed by the current limits on available IP address space. Any organization wanting to provide service for visiting mobile hosts but not willing to deploy a foreign agent may support any number of visitors by reserving a portion of their local IP address space for dynamic allocation as care-of addresses for visiting mobile hosts.

5.3. Location caches

The deployment and operation of a location cache in any node is only an optimization to the protocol, and no location caches are required, although the use of location caches is highly desirable. Each location cache may scale to any size as needed by any local administrative policies, but no specific location cache size is imposed by the protocol. Similarly, any local cache replacement policy may be used to manage the space within the cache.

If the location cache at some node is too small to be able to store a cached binding for each mobile host with which this node is actively communicating, the local cache replacement policy determines which entries are retained in the cache. For example, the use of LRU replacement will keep the most recently used entries in the cache. Other possible cache replacement policies might weight each entry by the number of times it had been recently accessed, or by some administratively assigned priority based on a list of preferred hosts for which bindings should be cached. Such decisions are entirely local to the node (and organization) implementing the location cache.

The use of location caches improves the scalability of the protocol by avoiding the need to send most packets

through the Internet to and from the mobile host's home network, and by avoiding the need for the home agent in the mobile host's home network to handle each packet. The location cache in a correspondent host maintains cache entries only for the individual mobile hosts with which that correspondent host is communicating. This approach scales well, as each individual correspondent host will at any time only be communicating with a limited number of mobile hosts. Furthermore, since in general the set of mobile hosts with which a correspondent host is communicating will change only slowly over time, any reasonable cache replacement policy such as LRU should work well.

5.4. Impact on the network

No changes to the routing infrastructure of the Internet are required to support Mobile IP. By tunneling packets to a mobile host, all routers through which the tunneled packet must pass treat the packet exactly as any ordinary IP packet, using existing Internet routing algorithms. The routing scalability of the Internet is thus maintained, since each router need not know the location of any individual mobile hosts, even though it may forward packets to them; only the two endpoints of the tunnel need know that tunneling is taking place or need care that mobility is the purpose of the tunneling. The Mobile IP protocol can thus be deployed incrementally, with each organization adding home agents or foreign agents as the need arises. Any or all hosts and routers can be upgraded at any time, if desired, to support location caches.

By using route optimization, the overall overhead on the Internet can be minimized. Routing packets indirectly to a mobile host through the mobile host's home network and home agent places unnecessary overhead on all links and nodes along this path, but route optimization allows this longer, indirect path to be avoided. Route optimization also reduces the resource demands on each home network, and avoids any possible performance bottleneck at the home network or at the home agent.

6. Conclusion

Recent increases in the availability of mobile computers and wireless networks provides the opportunity to integrate these technologies seamlessly into the Internet. Mobile users should be able to move about, transparently remaining connecting to the Internet, utilizing the best available network connection at any time, whether wired or wireless. For example, a mobile host in its owner's office may be connected to an Ethernet, but when disconnected and carried away, it could transparently switch to a connection through a high-speed local area wireless network. While moving around within the building, the host could switch transparently from one wireless subnet to another, and when leaving the building, could again switch transparently to a wide-area wireless data service.

The current work in the IETF Mobile IP Working Group provides a good approach to reaching this vision of seamless transparent mobility. These protocols can efficiently scale to very large numbers of mobile hosts operating in a large internetwork. Such scalability will become crucial as the Internet continues its exponential growth, and as mobile users begin to account for a growing fraction of this population.

Acknowledgements

This paper has benefited greatly from discussions with many other participants in the Mobile IP Working Group of the Internet Engineering Task Force (IETF). I would particularly like to thank Andrew Myles and Charlie Perkins for their collaboration in our work within the IETF. I would also like to thank the anonymous referees for their comments and suggestions which have helped to improve the clarity of the paper. The protocols described in this paper are a product of the Mobile IP Working Group of the IETF, but the views and conclusions expressed here are those of the author.

An earlier version of this paper was presented at the Ninth Annual IEEE Workshop on Computer Communications, Duck Key, Marathon, Fl, USA, 1994.

References

[1] T. Blackwell, K. Chan, K. Chan, T. Charuhas, J. Gwertzman, B. Karp, H.T. Kung, W.D. Li, D. Lin, R. Morris, R. Polansky, D. Tang, C. Young and J. Zao, Secure short-cut routing for Mobile IP, *Proc. USENIX Summer 1994 Technical Conference* (1994).

[2] S. Bradner and A. Mankin, The recommendation for the IP Next Generation protocol, Internet Request For Comments RFC 1752 (1995).

[3] S.E. Deering, ICMP router discovery messages, Internet Request For Comments RFC 1256 (1991).

[4] R. Droms, Dynamic Host Configuration Protocol, Internet Request For Comments RFC 1541 (1993).

[5] V. Fuller, T. Li, J. Yu and K. Varadhan, Classless Inter-Domain Routing (CIDR): an address assignment and aggregation strategy, Internet Request For Comments RFC 1519 (1993).

[6] J. Ioannidis, D. Duchamp and G.Q. Maguire Jr., IP-based protocols for mobile internetworking, *Proc. SIGCOMM '91 Conference: Communications Architectures & Protocols* (1991) pp. 235–245.

[7] J. Ioannidis, G.Q. Maquire Jr. and S. Deering, Protocols for supporting mobile IP hosts, Internet Draft (work in progress) (1992).

[8] J. Ioannidis and G.Q. Maguire Jr., The design and implementation of a mobile internetworking architecture, *Proc. Winter 1993 USENIX Conference* (1993) pp. 491–502.

[9] D.B. Johnson, Mobile host internetworking using IP loose source routing, Technical Report CMU-CS-93-128, School of Computer Science, Carnegie Mellon University, Pittsburgh, Pennsylvania (1993).

[10] D.B. Johnson, Transparent Internet routing for IP mobile hosts, Internet Draft (work in progress) (1993).
[11] D.B. Johnson, Ubiquitous mobile host internetworking, *Proc. Fourth Workshop on Workstation Operating Systems* (1993) pp. 85–90.
[12] D.B. Johnson, Scalable and robust internetwork routing for mobile hosts, *Proc. 14th Int. Conf. on Distributed Computing Systems* (1994) pp. 2–11.
[13] D.B. Johnson, A. Myles and C. Perkins, The Internet Mobile Host Protocol (IMHP), Internet Draft (work in progress) (1994).
[14] D.B. Johnson, C. Perkins and A. Myles, Route optimization in Mobile IP, Internet Draft (work in progress) (1995).
[15] F. Kastenholz and C. Partridge, Technical criteria for choosing IP: the next generation (IPng), Internet Draft (work in progress) (1994).
[16] J. Mogul and J. Postel, Internet standard subnetting procedure, Internet Request For Comments RFC 950 (1985).
[17] A. Myles, D.B. Johnson and C. Perkins, A mobile host protocol supporting route optimization and authentication, IEEE J. Select. Areas Commun. 13 (1995) 839–849.
[18] A. Myles and C. Perkins, Mobile IP (MIP), Internet Draft (work in progress) (1993).
[19] J. Penners and Y. Rekhter, Simple Mobile IP (SMIP), Internet Draft (work in progress) (1993).
[20] C. Perkins (ed.), IP mobility support, Internet Draft (work in progress) (1995).
[21] C. Perkins and Y. Rekhter, Support for mobility with connectionless network layer protocols (transport layer transparency), Internet Draft (work in progress) (1993).
[22] C.E. Perkins, A. Myles and D.B. Johnson, The Internet Mobile Host Protocol (IMHP), *Proc. INET'94/JENC5: The Annual Conference of the Internet Society*, held in conjunction with 5th Joint European Networking Conference (1994) pp. 642-1–642-9.
[23] J.B. Postel (ed.), Internet Protocol, Internet Request For Comments RFC 791 (1981).
[24] J.B. Postel, Multi-LAN address resolution, Internet Request For Comments RFC 925 (1984).
[25] Y. Rekhter and C. Perkins, Short-cut routing for mobile hosts, Internet Draft (work in progress) (1992).
[26] R.L. Rivest, The MD5 message-digest algorithm, Internet Request For Comments RFC 1321 (1992).
[27] M.T. Rose, *The Open Book: A Practical Perspective on OSI* (Prentice-Hall, Englewood Cliffs, NJ, 1990).
[28] G.S. Sidhu, R.F. Andrews and A.B. Oppenheimer, *Inside Apple Talk* (Addison-Wesley, Reading, MA, 1990).
[29] C. Sunshine and J. Postel, Addressing mobile hosts in the ARPA Internet environment, Internet Engineering Note IEN 135 (1980).
[30] F. Teraoka, K. Claffy and M. Tokoro, Design, implementation, and evaluation of Virtual Internet Protocol, *Proc. 12th Int. Conf. on Distributed Computing Systems* (1992) pp. 170–177.
[31] F. Teraoka and K. Uehara, The virtual network protocol for host mobility, Internet Draft (work in progress) (1993).
[32] F. Teraoka, Y. Yokote and M. Tokoro, A network architecture providing host migration transparency, *Proc. SIGCOMM '91 Conference: Communications Architectures & Protocols* (1991) pp. 209–220.
[33] P. Turner, NetWare communications processes, *NetWare Application Notes*, Novell Research (1990) pp. 25–81.
[34] H. Wada, T. Ohnishi and B. Marsh, Packet forwarding for mobile hosts, Internet Draft (work in progress) (1992).

David B. Johnson is currently an Assistant Professor of Computer Science at Carnegie Mellon University in Pittsburgh, Pennsylvania, where he has been since 1992. Prior to joining the faculty at Carnegie Mellon, he was a Research Scientist and Lecturer at Rice University in Houston, Texas, for three years. He received the B.A. degree in computer science and mathematical sciences in 1982, and the M.S. and Ph.D. degrees in computer science in 1985 and 1990, respectively, all from Rice University. Dr. Johnson's research interests include network protocols, distributed systems, and operating systems. For the past three years, he has been working extensively in wireless and mobile networking protocols, including internetwork routing for mobile hosts. He has been an active participant in the Internet Engineering Task Force and is among the primary designers of the IETF Mobile IP protocol. He is a member of the IEEE Computer Society, IEEE Communications Society, ACM, USENIX, Sigma Xi, and the Internet Society.
E-mail address: dbj@cs.cmu.edu