Mobile IP: A Solution for Transparent, Seamless Mobile Computer Communications

J. Redi [a], P. Bahl [b]
[a] Dept.of Electrical and Computer Engineering, Boston University, *redi@acm.org*
[b] Microsoft Research, Redmond, WA, *bahl@microsoft.com*

There is hardly a more pervasive trend in the world today than communications. The momentum towards anytime, anywhere and any-type of communication is fueled by advances in radio frequency technology, near-exponential sales of laptop computers and personal digital assistants (PDAs), as well as new models of business that expect and rely on instantaneous information availability. The commonly romanced vision of a traveling executive being able to have uninterrupted, uninhibited access to his office mail, office databases, personal files, and video conferencing while moving between geographically separated customer locations, airports and hotels is a reality that will happen in the next few years.

Fundamentally, the hardware solutions necessary for supporting ubiquitous mobile communications already exist: Laptop and notebook computers can be equipped with cellular and landline modems, cellular and public service telephone network (PSTN) service is everywhere, and Internet service providers (ISPs) exist everywhere from downtown office buildings to the basements of houses belonging to entrepreneurs in rural areas. However, though key to end-user satisfaction, the protocols and software required for seamless mobile communications over these diverse pieces of hardware and technologies are not yet wide-spread.

The ability of the Internet to scale from it's humble beginnings 20 years ago to today's estimated 30 million connected nodes is primarily because of the robustness, scalability, and interoperability of Internet Protocol version 4 (IPv4). The ability of the Internet to expand so dramatically without compromising its ability to route packets effectively is one of it's greatest assets and can be directly attributed to how the Internet Protocol (IP) functions. Generally speaking, part of the node's IP address provides its physical location on the network, much in the way the area code does for a phone number. In IP, the subnetwork (or subnet) a host resides on can be determined by its IP address and therefore routing consists of sending packets in the general direction of the subnet instead of requiring explicit knowledge of the location of all nodes. This mix of addressing and routing makes the whole network scale well, but is based on the assumption that nodes are stationary. However, with an increasing number of the workforce becoming mobile, and the need for these mobile warriors to stay connected to their "home network" becoming more critical than ever before, the mapping of the bits in the IP address field to a physical location presents a problem for mobile communications. When a mobile computer is not physically connected to it's home subnet, IP is incapable of routing packets to it correctly. This then provides the motivation and the fundamental objective for Mobile IP: to allow a computer to maintain normal communications with its home network and all other nodes on the Internet regardless of its point of attachment and while it is moving.

A fully deployed wide-area Mobile IP system will allow the nomadic user to plug her palmtop computer into a network in a conference room or at a coffee house without a need for her to reconfigure her machine. She would be able to create and maintain a video-conferencing session even while moving between the building's internal wireless local area network (WLAN) and an external wide-area wireless data network. When visiting another company she would be able to plug her laptop into a "foreign" LAN and using cryptographically secure communications, mount file systems from her own company to get complete access to all her personal files, databases, email, and other similar resources. It is useful to note here that although Mobile IP has been designed with the future in mind, the scenarios just described can already be achieved today using hardware that is already commercially available. Looking towards the future, we can expect IP to be the pervasive communication protocol across a diverse set of devices, not just limited to computers. For example, a person's wristwatch could have a miniature network transceiver and an IP address so that no matter where she goes she will be able to access her email by connecting to whichever IP network exists in her vicinity. Similarly sophisticated instrumentation such as moving robots will be able to communicate wirelessly with their home networks without the need for configuration as they move to different geographic areas and between different subnets, automatically receiving software updates and command messages from their control servers. In summary, the demand for Mobile IP will be limited only by the need to support seamless mobile data communication between homogeneous and heterogeneous networks. Fortunately, due to the medium-independent design philosophy, Mobile IP does not require specialized hardware and several software implementations are already available freely . Thus the cost of deployment is mostly from constructing the infra-structure of wireless and wired networking hardware. Anytime, anywhere, to anyone communications can be made transparent, robust and seamless with Mobile IP.

In this article, we first discuss in detail the need for Mobile IP in today's mobile world and point out reasons why other solutions such as *Cellular Digital Packet Data (CDPD)* are geographically constrained and do not support seamless heterogeneous mobility. We follow this with a section describing how Mobile IP works, including the entities involved, and how packets are routed to mobile nodes within the Internet. Next, we discuss some of the challenges Mobile IP has faced over its evolution and the availability of some implementations. We conclude with our perspectives on the future of Mobile IP and its relation to the next generation of IP protocols.

1 Routing in the Internet

To appreciate Mobile IP, we have to first consider how the Internet works normally. The Internet provides a means of letting geographically separated users exchange messages in the form of voice, data, and video packets. Occasionally, the devices ("nodes") employed by the users to communicate over the Internet are physically linked to each other and messages can be exchanged directly. However, due to serious scalability limitations arising out of directly connecting millions of nodes, a majority of the time there are no such direct links and packets have to traverse several intermediate nodes before reaching their final destination. These message packets are able to reach their destination because the intermediate nodes cooperate with them and with each other by "routing" the packets appropriately and towards the final destination node. These intermediate nodes are thus called *routers*. Since routing in the Internet is based on the address of the destination node, it is intimately influenced by how addresses are assigned to the communicating nodes.

The Internet is made up of millions of subnets. A subnet is a network of nodes that may be common to an organization such as a university campus or a corporation. If there are more nodes in the organization than can be accommodated within a single subnet additional subnets can be created for the same organization. Every node on the Internet has a unique 4 byte IP address. Out of these 4 bytes, as many as 3 bytes may be used to identify the node's subnet. The node's unique address within the subnet is assigned from a pool of addresses known to the maintainer of the subnet. Address assignment and subsequent configuration within the node is done either manually or automatically. Furthermore, addresses are given out either as static IP addresses, that is, nodes are identified by the same IP address always, or dynamically via a protocol such as the Dynamic Host Configuration Protocol (DHCP), in which case the IP address is leased to the node for a finite period of time with the provision that the node can renew its lease every time it needs to.

Each router on an IP network can contain three different types of routing table entries. These are described as either host-specific, network-specific, or default routing entries. Host-specific routes are typically for delivering packets to nodes that are connected to the router directly via a particular network interface. Host-specific route entries therefore provide a one-to-one mapping between the destination IP address and the specific interface used to forward the packet. Network-specific route entries are used for routing packets according to the destination node's subnet location. Thus the bits in the IP destination address which represent the subnet which the destination node is on, are used to determine which network interface the packet is to be forwarded with. Default-routing table entries are for addresses whose appropriate network interface cannot be resolved either via host-specific routes or via network-specific routes. In this case the router simply forwards the packet to the "next-hop" router with the hope that the next router will know how to further route the packet properly. Network-specific routing tables entries and the default-routing table entries are the key to the scalability of the Internet. By determining the subnet address in the header field of the packet, routers between the destination and source nodes select the next node, or next-hop that should receive the packet. Thus even though intermediate nodes do not necessarily know how to get data to a particular destination node, they do know how to get data to a node that does. With only a few properly selected routers, all nodes on the Internet can reach all other nodes. In general routers are specially-constructed hardware devices dedicated to the task of routing several million packets every second, we will often use the term router to refer to any general purpose machine which can forward packets, not destined for them, across the network. Thus our definition of routers can often apply to machines which appear to be simply "hosts" or "nodes" (i.e. PCs or workstations) on the network.

As mentioned earlier, the addressing described is reminiscent of the area code and prefix numbers of the public service telephone network which allow the phone company to instantly know how to route our calls. Consider a potential alternative in the case of a random address assignment to nodes. Routing in this case would require an extremely inefficient or unscalable scheme such as a few central routing hubs with full knowledge of the locations of all nodes, or alternatively all routers would require an entry in their routing tables for every node in the world they could potentially communicate with.

2 The need for Mobile IP

From the preceding description it should be apparent that the fundamental need for Mobile IP arises when a node connected to the Internet changes its point of attachment. This is typically due to a change in its physical location, which then necessitates a change in its IP address. If during the course of communication the mobile node moves to a different subnet, for example if it moves between a wired and a wireless network, other nodes will no longer able to communicate with it. Packets will arrive at the mobile node's original subnet, identified by its original IP address, but do not reach the node since it is now connected to a different subnet. IP was not designed with the mobile computer in mind.

Although it is possible to communicate with a mobile node which changes its IP address as it moves, it is expensive and comes with great difficulty. For instance, every time the node acquires a new IP address, all connections that were open with the previous IP address have to be shutdown and then re-initiated with the new address. In several operating systems, the IP stack is such an integral part of the kernel that a complete reboot of the machine is required to change the IP address. If the mobile node is quickly moving in a wireless environment with small cell sizes, the IP subnet can change very frequently as the mobile moves between base stations or service providers, such that the task of performing these changes quickly enough is not even possible.

The motivation for adapting IP to support mobility instead of starting over with a brand new protocol includes several compelling reasons: First, as IP has expanded across the world, so has the installed base of applications designed for it. From database and file connectivity to electronic mail messaging, IP applications have become the de facto standard for data networking. End-users have come to expect the same computational power, application availability, and communications capabilities from a mobile computer as they get from their desktop systems. Thus, in order to maintain the usefulness of the installed base of applications there is compelling need to keep IP as the networking protocol. Second, mobile communications not only includes communications between mobile nodes only, but also between mobile nodes and stationary nodes. To maintain seamless connectivity between the mobile nodes and the existing millions of fixed Internet nodes, without adding substantial complexity, there is compelling need to keep enough of the IP protocol stack intact so that non-mobile nodes can communicate with the mobile ones without requiring any changes at the application or kernel level.

Through these reasons, we can distill the goals of Mobile IP to be the following: First, the mobile node must be able to communicate even when its point of attachment changes. Second it must be able to communicate using its original (home) address. This requirement is fundamental to the ability of nodes, connected to the Internet, to initiate connections with a mobile node and maintain connectivity with it even as it moves. This is also important from the point of view of keeping the Internet scalable since this requirement removes the need for propagating address changes all over the network. Furthermore, this requirement implicitly states that all nodes implementing IP should be able to communicate with mobile nodes even if they themselves do not implement the specialized Mobile IP functions. The advantages of this goal cannot be understated as it causes users to be unaware of whether the node they are communicating with is a mobile or non-mobile computer. This then maintains utility of the already existing software designed for the IP stack, and avoids the unreasonable expectation of requiring protocol changes throughout many million of nodes connected to the Internet. The third and final goal of Mobile IP is that it must provide at least as much security as standard IP. This issue is non-trivial since there is a need to reroute packets when a node moves throughout the network. This opens the possibility for *Denial of Service Attacks* to be launched against mobile nodes, rerouting their traffic somewhere else without their consent. Since Mobile IP would not be implemented unless all security concerns were met, this has been a priority for the designers.

Mobile IP is a network layer routing protocol. It makes no lower-level assumptions about the link characteristics such as bit-rates, error-rates or delay and thus is a device and communication medium independent solution. Mobile IP can be as easily used when moving between different high-speed Ethernet LANs in an office environment as it can when moving over a wide-area wireless data service in the field. It supports both homogeneous mobility, that is when moving between similar medium networks and heterogeneous mobility, that is when moving from one medium to another, as is the case when the node moves between a wired network and a wireless one.

2.1 Alternatives to Mobile IP

Two examples of other wireless systems which offer IP connectivity without Mobile IP are CDPD and IEEE's 802.11 Wireless LAN standard. These two systems are fairly well known, so in this section we compare their capabilities with Mobile IP.

CDPD was developed by IBM and backed by a consortium of nine major US cellular service providers who were primarily interested in utilizing the unused channel capacity of the exiting first generation analog cellular system called *Advanced Mobile Phone System (AMPS)*. Together, these providers claim a coverage of about 95% of the United States with voice-oriented analog AMPS service. CDPD layered on top of AMPS in a manner so as not to disturb the original system. Data connections utilize unused analog voice channels for digital packet data transmission. AMPS connections take precedence over CDPD connections, thus a CDPD connection may be "bumped-off" whenever a new voice connection starts up on the same channel. Most of the time, data users do not notice this automatic bumping-off as data transmissions quickly jump to a different available frequency when they the sense that the one they are currently occupying has been allocated to a analog voice connection. This process is often referred to as *channel hopping*. CDPD's compatibility with AMPS allows it to be installed in existing analog cellular services in the U.S.

The CDPD system makes mobility transparent to the IP layer by adapting to it at the link and physical layers. Consequently, mobile users can run their standard IP applications without a need for any modifications. The CDPD network spans the entire country linking all of its base stations together. Each node in the CDPD network is assigned a permanent IP address. Packets sent to the mobile node by landline or other non-CDPD node are routed to the CDPD network. Using a propriety mobility management protocol, the mobile node's location is determined. If the node is on the CDPD network and is available, the packets are routed to it through the network of base stations. The CDPD network can therefore be considered a single, country-wide IP subnet. Mobile nodes can get seamless IP connectivity while they are on this network. However, if the mobile users connect their computer directly to a landline network, or change their data service providers, a change in their IP address is required and this link-layer solution stop working. The CDPD solution is therefore both geographically and media constrained. Although currently, it is a considered a favorable solution to providing IP connectivity to mobile users over wide-area packet data network, its limitations will become more pronounced in the future as users realize that it restricts them to staying with the designated CDPD network service providers only. This in turns limits mobility support coverage to areas where this network exists and prevents competitive pricing and service models. Furthermore, CDPD provides only homogeneous mobility support, not supporting mobility as nodes move between different media types.

In contrast to CDPD's low data rate, wide-area service, IEEE's 802.11 Wireless LAN standard specifies a high 1-2Mbps data rate in an office-sized microcells. The 802.11 standard is a comprehensive standard that specifies the physical layer along with the medium access control protocol that nodes must implement in order to be able to communication with each other. The standard does not specify routing, addressing, or handover issues. When a node moves between cells on different subnets it must terminate all its connections with the Internet and re-establish communications using its new IP address assigned in the new cell. This is effectively the same situation that occurs in the case of wireline networks and which motivated the development of Mobile IP. For this reason, many wireless LAN companies are including Mobile IP implementations for use with the WLAN adapters and base stations.

In conclusion, link-layer and physical-layer solutions are both media limited and coverage limited. In contrast, Mobile IP is media-independent and is not limited by the coverage of any one underlying hardware technology. Mobile IP can reroute packets seamlessly even when the underlying link changes or when the user moves between different service areas without requiring her to stop and restart communications. Mobile IP can thus run over both CDPD or 802.11 hardware, while allowing users to roaming between them freely.

3 How Mobile IP Works

Mobile IP was designed by the "IP Routing for Wireless/Mobile Hosts" working group (mobileip WG) of the Internet Engineering Task Force (IETF) and published as a proposed standard in November 1996. A list of publicly-available request for comments (RFCs) that define Mobile IP are described at the end of this article.

Before explaining how Mobile IP works, it is useful to become familiar with the terminology used in the rest of this article. We will use these terms extensively in our subsequent description of Mobile IP operation and in describing the exchange of messages between the mobile nodes and the other key entities within the network.

Mobile Node: A node running the Mobile IP protocol stack which moves between different IP subnets. This node is assigned a (permanent) IP address which defines where all its packets should be sent. When other nodes send packets to the mobile node, they only specify this home IP address in the packet, regardless of where the mobile node is physically located.

Home Network: The subnet which corresponds to the home address of the mobile node as well as that of the home agent. It is considered the mobile node's "home" point of attachment.

Home Agent: A router on the home network that is responsible for intercepting packets destined for the mobile node when the mobile node is attached to a foreign network. The home agent is responsible for forwarding these packets to the mobile node.

Foreign Network: A network, other than the mobile node's home network, that a mobile node attaches itself to.

Foreign Agent: A router on the foreign network configured for Mobile IP. When the mobile node has a foreign agent care-of address all packets are relayed through this node. When using a collocated care-of address, the mobile node may still use a foreign agent for its default router or for registration with the foreign network.

Care-of Address: The address that the mobile node uses for communication when it is away from its home network. This address can either be a *foreign agent care-of address*, when the mobile node uses the foreign agent's IP address as its care-of address or a *collocated care-of address*, where the network interface of the mobile node is temporarily assigned an IP number on the foreign network.

Correspondent Node: Any host which is communicating with the mobile node. This node could be located on the home network, foreign network, or any other place which is able to route packets to the mobile node's home network.

Tunneling: The process of *encapsulating* an IP packet within another IP packet for the purpose of routing it to a location other than the one specified in the original destination field. Specifically, when a packet is received by the home agent, it *encapsulates* the original packet inside a new packet, placing the mobile node's care-of address in the new destination address field before forwarding it to the appropriate router. The path that is followed by this new packet is called the *tunnel*.

From the above descriptions it should be clear that only three entities have to be modified in order to support Mobile IP over the Internet: the mobile node, the home agent, and the foreign agent. When a collocated care-of address is used by the mobile node, a foreign agent is often not even required.

Home agents and foreign agents periodically broadcast their willingness to act as Mobile IP routers through agent advertisements. If a mobile node needs to immediately know the address of a potential agent without waiting for the next advertisement, it can broadcast an agent solicitation message. The mobile node uses these advertisements to determine if it has

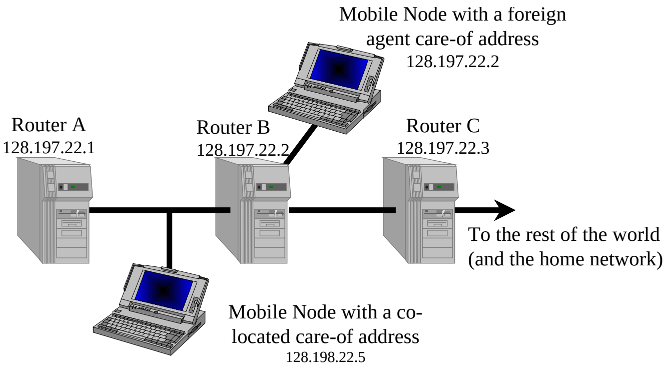

Figure 1: A mobile node with a foreign agent care-of address

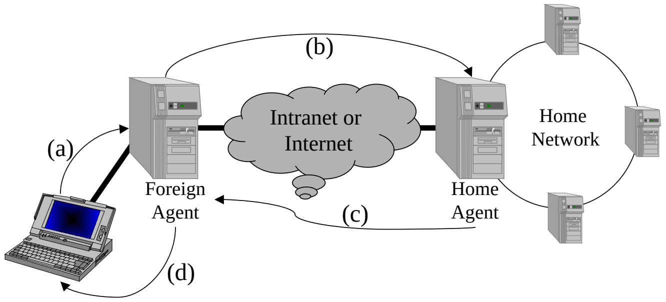

Figure 2: Flow of registration messages between a mobile node with a foreign agent care-of-address and its home agent

moved to a new location and where it can attach to a network if it has. If it is on the home network, no changes take place and no modifications to the IP protocol operation are required for communications. If the node determines that it is on a foreign network, it obtains a care-of address from the foreign agent (for a foreign agent care-of address), or through another protocol such as DHCP (for a collocated care-of address). An example of how these two types of care-of addresses differ is depicted in Figure 1. When using a foreign agent care-of address the mobile node registers the IP address of the foreign agent with its home agent (in the figure, this is 128.197.22.2). The foreign agent is responsible for unmarshalling the tunneled packets sent to it by the mobile node's home agent and relaying these to the mobile node. It is also responsible for relaying packets from the mobile node to correspondent nodes and to the home agent. Alternatively, the mobile node can be directly connected to the foreign network itself and therefore directly communicate with the home agent (in the figure, this machine is given IP address 128.197.22.5). The foreign agent type of care-of address is preferable in IPv4 due to its limited address space.

Once a node has obtained a new care-of-address, it registers this address with its home agent. Registration with the home agent is performed for the following reasons: a node has moved to a new foreign network and therefore needs to register a new care-of-address, a node needs to deregister an old care-of-address with its home agent when it returns to the home network, or when a node needs to reregister with the home agent because its previous registration is about to expire. Understand that many care-of-addresses can be registered for a mobile node at once. This allows for potential situations such as a node quickly alternating between two adjacent wireless cells, and multiple points of attachment are preferable to rapid deregistration and registration between cells.

The transmission of registration messages is shown in Figure 2 for the case of a foreign agent care-of-address. The mobile node submits a registration request to the foreign agent (a). The care-of-address which the mobile node requests will be determined by the foreign agent's agent advertisement messages. The foreign agent does some validity checks on the registration request and then relays the packet to the home agent. The foreign agent makes sure that parameters such as the registration lifetime, and the tunneling method are supported. It also verifies security authentification information. The foreign agent will maintain information such as the link-layer address and the IP source address that the mobile node is using before it resends the registration request with it's own address as the IP source address to the home agent as labeled in (b).

The home agent receives the registration request, checks the options of requested service and authentification from the foreign agent. If the request is considered to be valid and serviceable the agent updates its bindings to record the new care-of-address for the mobile node. The home agent then forms an authenticated registration reply to send back to the foreign agent informing it of a successful or unsuccessful registration (c). Finally the foreign agent receives the reply, and sends a new authenticated message back to the mobile node. If all these replies indicate a successful registration, the mobile node will begin to receive its packets tunneled from the home agent through the foreign agent.

All registration messages include strong authentification to eliminate the possibility of *Denial of Service* attacks to the nodes on the network. Denial of Service can occur if a rogue machine sends bogus registration requests to a home agent which results in a mobile node's packets being rerouted to the wrong place. Mobile IP keeps this from occurring by requiring all registration messages contain a 16-bit *message digest*. This is created by using a well-known cryptographic algorithm to create a short unique sequence of bits from the packet's data, a random number to prevent *Replay attacks*, and a secret key which is shared between the destination and source of the message. The secret key is a 16-bit number known only between two parties. A discussion on how to distribute keys is beyond the scope of this article, but is a topic which has been considered heavily by the IETF.

In order for the home agent to get packets to the mobile node when it is located away from the home network, the home agent uses *encapsulation* to forward the mobile node's packets across the network. This is depicted in Figure 3. When a correspondent node wants to send a packet to a mobile node, it simply uses its permanent home address to send the packet to the home network (a). The home agent intercepts this packet and performs an encapsulation which consists of creating a new packet which has the entire original packet as its payload. This new packet has the mobile node's care-of address as its destination so it can be sent to the foreign network (b). The

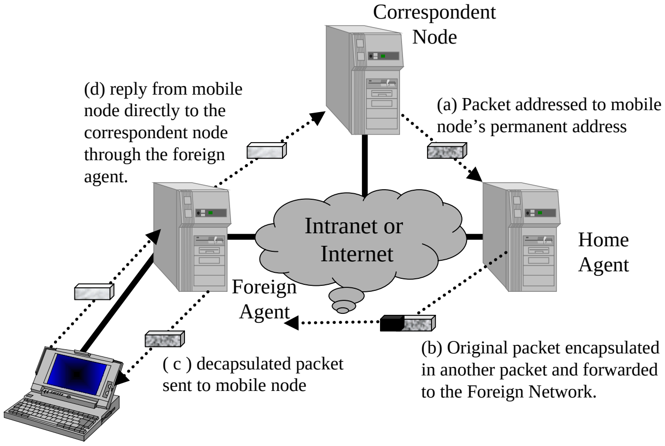
Correspondent Node
(d) reply from mobile node directly to the correspondent node through the foreign agent.
(a) Packet addressed to mobile node's permanent address
Intranet or Internet
Home Agent
Foreign Agent
(c) decapsulated packet sent to mobile node
(b) Original packet encapsulated in another packet and forwarded to the Foreign Network.

Figure 3: Communication between a correspondent node and a mobile node on a foreign network

foreign agent accepts this packet since it is identified as the destination, and decapsulates the original packet to forward it to the mobile node (c). Finally, if the mobile node wants to communicate with the corresponding node it can route its packets directly to that node through the foreign agent, using its home address as the source IP address of the packet (d). Note that this sequence of routing events results in what is called "triangle routing". Many of the packets from a corresponding node may have to travel sub-optimal paths in order to be routed through the home network. This is the small price to be paid for the advantage of the corresponding node not needing any sort of Mobile IP extensions. In a later section, we will discuss optimal Mobile IP routing extensions which require small changes at corresponding nodes.

There are three types of encapsulation which a home agent may use for Mobile IP: IP-in-IP encapsulation, minimal encapsulation, and generic routing encapsulation. IP-in-IP encapsulation creates a new packet with the payload section of the packet containing the original packet. This adds 20 bytes of overhead for the tunneling, but has the advantage that the resulting packet is exactly like any other IP packet and can handle intermediate network issues such as fragmentation. Minimal encapsulation takes the approach that since much of the information in an IP-in-IP encapsulated packet is redundant, it only adds the new header information which is different from what was included in the original IP header. Consequently, minimal encapsulation essentially just changes the source and destination addresses and includes the original addresses elsewhere in the packet. This requires an only 8 to 12 bytes of overhead to perform the tunneling, but is less robust since the packet's Time To Live (TTL) timer is not protected in this scheme, and the packet cannot survive fragmentation across the network. These two side-effects limit the usefulness of using minimal encapsulation for the purposes of tunneling. Finally Generic Routing Encapsulation was designed such that any protocol can be used inside or outside the encapsulated packet. It suffers from even greater amount of overhead than IP-in-IP, but unlike the others can support any protocol. It has explicit protection support for situations such as recursive encapsulation.

To recap, in this section we described the basic operation of Mobile IP. We showed how agent advertisements and solicitation allow the mobile node to determine its point of attachment on the network. We then described how the registration process uses authenticated messages to inform the foreign agent and the home agent of its new IP address. This is a necessary step for equipping the home agent with enough data so it can tunnel packets to the mobile node's new location. In the next few sections we describe some of the challenges Mobile IP has faced during its course of development and how these have been addressed by its designers. We also provide an example of a commercially viable application that uses Mobile IP and the current state of availability in terms of freely available prototype implementations.

4 Challenges to Mobile IP

Mobile IP is an elegant and relatively simple solution to seamless geographically-unconstrained roaming. It is built on top of the current version of IP in a way that is transparent to the stationary nodes which want to communicate with the mobile nodes. However, as the Internet evolves and new challenges are met, Mobile IP must often include improvements to overcome changes in today's networks. In this section we discuss three recent challenges which have faced the Mobile IP community and describe the solutions which have been drafted. One of these issues, ingress filtering, is related to the changes in security policies that subnetworks are implementing, while the second, route optimization, could create a potentially substantial performance increase for Mobile IP at a cost of a slight increase in complexity. The third issue is not a protocol level issue but one that is useful to address as more and more people start to use Mobile IP.

4.1 Ingress filtering

Recently a number of attacks on networked computers have involved forging the source IP address of an IP packet. In particular the TCP SYN Flood attack is performed by merely overwhelming a host with TCP setup connection requests. Each of these connections require memory in the host machine, such that an enormous number of them will at best not allow enough memory or address space for legitimate connections, and at worst will crash the host. Forging an IP source addresses (also called *spoofing*), allows an attacker to mount a flood attack which is difficult to trace. One of the few assumptions of IP routing is that only the destination IP address is used for the routing decisions, so packets with spoofed IP source addresses will still reach their destination.

In response to this the Internet Architecture Board (IAB) and the Computer Emergency Response Team (CERT) has suggested that all administrators of IP domains perform *ingress filtering* on all packets they pass into the Internet routing fabric. Only packets with topologically correct IP source addresses would be allowed to pass out of a subnet. It is easy to see that this type of filtering would also force routers to drop packets from a mobile node which was communicating through a foreign network. Recall that packets sent from a mobile node which is using a home agent to forward its packets, would include its home address as the IP source despite the location it was actually communicating from. Ingress filtering therefore makes this communication impossible.

The solution which has been proposed is called *Reverse Tunneling*. As we saw before, the tunneling procedure creates a triangle route between the correspondent node, the mobile node and the home agent. Correspondent node to mobile node communications are through the home agent, while the mobile node to correspondent node packets travel directly to the correspondent node. Reverse Tunneling would instead create a reverse path through the home agent for all the mobile node to correspondent node communications. The IP source address for the first leg of the reverse tunnel would then be the care-of address and, the IP source address for the second leg of the reverse tunnel would be the home agent. In this way, we have topologically correct addresses at all times, though we potentially increase the delay and congestion of the packets following the reverse path. The reverse tunneling specification also provides for circumvention of the reverse tunnel, so as to instead revert to using the standard Mobile IP triangle routing. This allows for situations such as a mobile node communicating with another node on the local network without requiring a round trip of the packet to the home agent.

4.2 Route Optimization

As we discussed in a previous section, when a mobile node is communicating from a foreign network all its packets must be forwarded through its home agent, even though it can directly send packets to the correspondent node. This triangle routing can result in significant performance degradation. For example, even if the mobile node is connected to the same foreign network as the correspondent node, all packets from the correspondent node must leave the subnet, be routed all the way back to the mobile node's home network, only to return in a tunnel to the original subnet.

Route optimization extensions to Mobile IP which have recently been proposed would allow a correspondent node to be informed of the mobile node's care-of address so that it can send packets directly to it. When a correspondent node sends packets to a home agent for tunneling to the mobile node, the home agent can assume that the correspondent node is unaware of the mobile node's current care-of address. After tunneling this packet for the correspondent node, the home agent sends an authenticated *Binding Update* to the correspondent node, advising it to update its *Binding Cache* with the care-of address of the mobile node it is sending packets to. This binding cache contains mappings from home addresses to the temporary care-of addresses. Each entry is specified to be valid for an amount of time which is equal to the time the node is registered with the home agent. Once a correspondent node updates its binding cache it can tunnel packets directly to the care-of address, which solves the triangle routing issue. These route optimization extensions also improve on the routing of packets which are "in flight" when a mobile node moves to a new foreign agent. Since the foreign agents themselves will keep binding caches of the new care-of addresses, they can forward packets which are intended for mobile nodes which have left their subnet. The forwarding foreign agent will then also send a *Binding Warning* message to the home agent specifying that the corresponding node has an out-of-date binding cache entry and should be sent an new Binding Update.

4.3 Fault Tolerance, Load Balancing and Congestion Control

In order for routing to work correctly, the forwarding node has to know the link-layer address (e.g Ethernet address) of the next hop router. This is necessary since the IP packet is forwarded within a link-layer frame which contains a link-layer address needed for identify the destination hardware. The process of determining the link-layer address corresponding to a given IP address is called address resolution and the most common method of accomplishing this on the Internet is though the *Address Resolution Protocol (ARP)*. The sender broadcasts a *ARP Request* message on the link it is connected to and if the target IP address contained within this message matches the IP address of one of the nodes connected to the link, then that particular node responds by sending a *ARP Reply* message to the requester, specifying its link-address.

In Mobile IP, foreign agents support the mobile node in the foreign network, while the home agent maintains the mobility binding and forwards packets to the mobile node when the mobile node has roamed out of its home network. While these agents can be implemented in the router, the protocol processing required by these agents are much more than the packet forwarding functions performed by a typical router. For example, when the mobile node moves to a new link from its home network, it can no longer receive ARP Request messages sent by the nodes on the previous link. The home agent then has to respond in place of the mobile agent with a *Proxy ARP*, so that messages intended for the mobile node are still sent to the home network so the home agent can forward them to the appropriate router.

When a very large number of mobile hosts from the same home network, are all visiting foreign networks, then the servers that perform the agent functions can become the bottlenecks. Furthermore, a single home agent or a single foreign agent is not a robust architecture, as it constitutes a single point of failure. In such a situation, it is useful to employ multiple foreign agents and multiple home agents in an IP subnet. In order to effectively utilize the processing capacity of these mobility agents while minimizing delay, it is necessary to distribute the workload among the several agents according to their processing capabilities and the traffic characteristics. A first home agent may transfer the control of a mobile host to another home agent in the same network as the system dynamically adjusts to the changing load conditions. The methods for triggering the transfer and the policy for selecting the next home agent define the various load balancing schemes which exhibit different performance characteristics.

5 Example of a Commercial Application that uses Mobile IP

Today, mobile warriors can dial into their home network via the *Point to Point Protocol* or other similar remote access protocols. However, to be able to do this the corporation to which these mobile warriors belong has to have an infrastructure of modems, phone lines, and remote-access servers. The cost of purchasing and maintaining such equipment, along with the administrative costs can become prohibitive. Consequently, there is a strong desire within the Internet community to develop protocols that allow sharing this cost with entities external to the corporations while still providing this important capability to the workforce. Corporations which have already made the investment in building this infrastructure would be happy to share their expenses by charging the mobile warriors of other corporations to gain access to their own home networks, if the integrity of the service providers internal networks is not compromised. Similarly, corporations that have not yet build this infrastructure would be happy to use the infrastructure of other corporation if they can feel secure in the knowledge that their data will not be compromised and that the foreign network will provide their mobile workforce with secure connectivity to their home networks.

Mobile IP can provide a solution to lowering the cost of providing Internet access to mobile warriors by enabling *Virtual Private Network (VPN)*. A Virtual Private Network is one in which the mobile node can send and receive messages exactly as if it was connected to its home network directly. The Internet can thus be used for virtual dial-up. The way this is done is to have the mobile node acquire a IP care-of-address from the local network and then have it pass this on to the home agent at its home network. This way the roaming mobile node can be contacted by any node on the Internet as easily as if it was on its home network.

An example of a proposal for a protocol that uses Mobile IP to establish such virtual private networks is the *Virtual Tunneling Protocol (VTP)* built into the "BaySteam Dial VPN Service" supported by Bay Networks. VTP uses Mobile IP to dynamically establish bi-directional tunnels through the Internet. Interestingly, in case of VTP, the mobile user node does not have to have Mobile IP support built within its networking stack. Examples of other similar protocols that achieve the functionality described above but without using Mobile IP are the *Point-to-Point Tunneling Protocol (PPTP), Layer Two Forwarding (LTF),* and the *Layer Two Tunneling Protocol (L2TP)*. While all of these protocols provide support for nomadicity they do not provide support for mobility. In other words, the mobiles are unable to retain their IP addresses and consequently cannot move within the subnets of the foreign network and over different mediums, while continuing to stay connected to their home network.

6 Availability of Mobile IP Implementations

The goal of the Mobile IP Working group is to move the Mobile IP specification all the way to establishment as a standard by the IETF. This requires extensive testing between different Mobile IP implementations to iron out uncertain portions of the draft standard. Since the Mobile IP draft standard document is freely available from the Internet Engineering Task Force web location, there are many freely available implementations of Mobile IP.

The following are only a few of the more popular ones:

- The MosquitoNet Research Group at Stanford University has a Mobile IP implementation for the Linux operating system that uses the collocated care-of address scheme. Their code is downloadable from the groups web site at: *http://mosquitonet.stanford.edu/software/mip.html*

- The Monarch Research Project at Carnegie-Mellon provides a free implementation of both Mobile IP for IPv4 and for IPv6. Their implementation of IPv4 includes code for foreign agent based care-of address scheme along with the route optimization extensions. The software has been written for the Linux operating system and is available for downloading from their web site: *http://www.monarch.cs.cmu.edu/*

- The Portland State Secure Mobile Networking Project puts emphasis on security issues in Mobile IP and its integration into the National Information Infrastructure. (*http://www.cs.pdx.edu/research/SMN/index.html*)

- The State University of New York, Binghamton's Linux Mobile IP software has been regularly demonstrated at conferences dedicated to mobile communications and is one of the premier implementations for the Linux operating system. (*http://anchor.cs.binghamton.edu/ mobileip/*)

Additionally, FTP software includes a full commercial implementation of Mobile IP as well as IPv6 in their "OnNet32", "SecureClient" and "OnNet Host Suite" protocol stack products. These products are designed for Windows 95/NT and feature completely customizable and integrated applications for communications with Unix and Windows TCP/IP servers (*http://www.ftpsoftware.com*).

7 The Future of Mobile IP

7.1 IPv6 and Mobility

As the use of IP version 4 has expanded, many lessons have been learned from its use. Improvements that are necessary for IP to remain the dominant protocol in the coming years have been solidified into a proposal called IP version 6 (IPv6). The most significant change to the protocol is in the size of the address space in each packet. Where version 4 had only 32 bits for each IP addresses, version 6 has 128 bits. This expansion solves the problem of the lack of available address space already plaguing the Internet. Additionally, recall that the reason that Mobile IPv4 foreign agent care-of addressing exists is to reduce the number of extra IP numbers needed to support Mobile IP. With foreign agent care-of addressing, every mobile node connected to a single foreign agent uses the foreign agent's address for it's own care-of address. With Mobile IPv6 address space is no longer an issue, so the use of the foreign agent is eliminated and packets are always directly tunneled to the mobile node itself.

IP version 4 packet options can present performance problems for routers since every router needs to examine each packet to determine if there are router-specific options in the packet. IPv6 solves this problem by identifying a router-specific option space. There is also a unified system in IPv6 for authentication support. When Mobile IP was being developed, generic IP security extensions were not yet standardized. Mobile IPv4 then had to make up its own authentification method and require all nodes to adhere to them. With IPv6's standardized authentification, improvements such as route optimization are much simpler as the authentification they require is already built into the protocol stack. IP version 6 solves and simplifies a large number of problems with the current Internet and in doing so greatly simplifies Mobile IP's implementation.

7.2 Areas for Further Investigation

As of this writing, there are several issues that are currently being discussed both within the IETF and the research communities. Most of these have to do with the interaction between Mobile IP and other existing or proposed Internet protocols. Examples include: (1) *Quality of Service Issues*: Providing quality-of-service (QoS) over the Internet has proved to be a daunting task. Adding support for node mobility in the Internet makes it even harder for the network to provide guaranteed QoS to connections. One of the proposals currently being discussed within the IETF community that addresses the problem of QoS over the Internet is the flow-based *Resource Reservation Protocol (RSVP)*. Like other similar proposals from the research community, RSVP makes an implicit assumption that nodes on the Internet are stationary and the protocol relies on being able to make explicit reservations across a predetermined packet route for the lifetime of the connection. The interaction of Mobile IP with such protocols is unclear and is being examined currently (2) *Security Issues:* As discussed previously, security is still one of the major issues that is preventing wide scale deployment of Mobile IP. A combination of techniques that incorporate elements of authentication, packet filtering, and cryptography are currently under investigation within the IP and research communities. (3) *Billing Issues:* What is the right model for charging customers who use Mobile IP to connect to their home networks through a foreign network. (4) *Mobile Networks*: Mobile IP for IPv6 does not support Foreign Agents, consequently mobile nodes that are moving together while being connected to the Internet require new care-of addresses simultaneously. This can cause problems in terms of delay and congestion. For example consider the case of hundreds of business travelers sitting in a moving train and connected to the Internet with their laptop machines and wireless modem. As the train moves rapidly between different subnets all nodes inside the train have to deregister and acquire new care-of addresses in the different subnets. Contrast this with the case when all the nodes on the train use the foreign agent's address as their forwarding address, where the foreign agent may be on a server machine on the train. As the train moves, only the foreign agent will have to release and acquire a new care-of address in the new subnetwork. The question is then how should one fix this problem for the case of IPv6.

Additional Reading

Standards and draft standards of protocols for the Internet are issued by the Internet Engineering Task Force. The IETF web site - *http://www.ietf.org* provides these standards, draft standards, proceedings, and committee meeting minutes. Specifically the Mobile IP WG pages containing the various RFCs relevant to this standard are located at - *http://www.ietf.org/html.charters/mobileip-charter.html*

The following Internet RFCs define the Mobile IP standard:

- RFC 1883 describes version 6 of IP.
- RFC 2002 is titled "IP Mobility Support" and is the document that describes the complete details of Mobile IP.
- RFC 2003 and 2004 describe two methods of tunneling. IP in IP tunneling and Minimal Encapsulation. The protocol-independent tunneling called Generic Routing Encapsulation is explained in RFC 1701.
- RFC 2005 describes the applicability of Mobile IP. It summarizes the operation of the protocol in order to show how the goals for host mobility were met by RFC 2002.
- RFC 2344 documents the proposed backwards-compatible extensions to Mobile IP for supporting reverse tunneling.
- The draft document titled "draft-ietf-mobileip-ipv6" describes the extensions to IPv6 to support host mobility.
- "draft-ietf-mobileip-optim" specifies the methods for supporting route optimization extensions in Mobile IPv4.

Additionally, two books have been released recently by leaders in the development of Mobile IP:

- C. Perkins, *Mobile IP: Design Principles and Practice*, Addison-Wesley Longman, Reading, Massachusetts, USA, 1998 (ISBN: 0-201-63469-4)
- J. Solomon, *Mobile IP: The Internet Unplugged*, Prentice Hall, Englewood Cliffs, New Jersey, USA, 1998 (ISBN: 0-13-856246-6)

The first of these two contains detailed description of the protocols and is geared towards those implementing their own Mobile IP stack. The second is a detailed overview which includes many deployment details and discussions of potential applications.

Additional information about the other wireless data networking standards discussed in this article can be obtained from:

- CDPD Consortium, *Cellular Digital packet Data Specification*, July 1993, *http://www.cdpd.org*
- IEEE 802.11 Committee Draft Standard, *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification*, Rev. D5, IEEE Catalog Number DS2972, May 1996

Biographies

Jason Redi received a B.S. in computer engineering from Lehigh University in 1992, and the M.S. and Ph.D. degrees in computer engineering from Boston University in 1994, and 1998. He is author of nearly twenty papers and patents in the area of mobile computing and communications. Dr. Redi has worked as a principal investigator in wireless network protocols for Boston Communications Networks, and as a consultant in the area of MAC-level protocols for Motorola ISG. He is a member of the IEEE, ACM, Tau Beta Pi, Sigma Xi and is associate editor of *ACM SIGMOBILE Mobile Computing and Communications Review (MC^{2}R)*. He can be reached via email at *redi@acm.org*.

Paramvir Bahl holds a Ph.D. in Computer Systems Engineering from the University of Massachusetts Amherst. He is currently with Microsoft Corporation where he is carrying out research in peripatetic computing, multi-hop multi-service adhoc networks, mobility aware self-configuring operating systems and real-time audio-visual communications. Prior to Microsoft, Dr. Bahl spent nine years at Digital Equipment Corporation where he initiated, led, and contributed to several seminal multimedia hardware and software projects.

Dr. Bahl is the Vice-Chairman of *ACM's special interest group on mobility (SIGMOBILE)*. He is the founding editor-in-chief of *ACM Mobile Computing and Communications Review* and has served as a guest editor for the *IEEE Journal on Selected Areas in Communications*, the *ACM Journal on Mobile Networking and Applications* and the *IEEE Communications Magazine*. He serves on a number of organizing and technical program committees of the IEEE and ACM conferences including MobiCom, INFOCOM, PIMRC, MoMuC and MMT and has organized sessions and panels on the topic of wireless LANs, mobile computing and multimedia communications. He is the author of more than two dozen scientific papers and more than one dozen patents issued and filed in the areas of wireless networks, computer communications and digital signal processing. Dr. Bahl is a senior member of the IEEE, a member of ACM, and a past president of the electrical engineering honor society Eta Kappa Nu (Zeta Pi Chapter).