Status of this Memo

This document is a submission by the Mobile IP Working Group of the Internet Engineering Task Force (IETF). Comments should be submitted to the mobile-ip@SmallWorks.COM mailing list.

Distribution of this memo is unlimited.

This document is an Internet-Draft. Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

To learn the current status of any Internet-Draft, please check the "id-abstracts.txt" listing contained in the Internet-Drafts Shadow Directories on ftp.is.co.za (Africa), nic.nordu.net (Europe), munnari.oz.au (Pacific Rim), ds.internic.net (US East Coast), or ftp.isi.edu (US West Coast).

Abstract

This document discusses problems arising out of the use of Mobile IP in a security conscious Internet and presents a list of solution requirements deemed important. These problems may need to be resolved in several stages. A priority order in which these problems should be solved is also proposed.

All firewalls are assumed to implement standard mechanisms [RFCs 1825,1826,1827] for authentication and encryption proposed by the IPSec working group of the IETF.
1. Introduction

The IETF Mobile IP protocol [Per96a] allows a mobile node (MN) to continue sending and receiving IP datagrams using a fixed address, its home address, even when it is no longer connected to its home subnet. When a mobile node visits a foreign subnet, it obtains a care-of address on that network and registers that address with its home agent (HA), a special entity residing on its home subnet. The home agent, in turn, intercepts datagrams meant for the mobile node and tunnels them to the registered care-of address. Tunneling refers to the process of enclosing the original datagram, as data, inside another datagram with a new IP header [Per96b, Per96c]. The new header carries the mobile node’s care-of address in the destination field. The care-of address may belong to a specially designated node -- the foreign agent (FA) -- or may be a temporary address assigned to one of the interfaces of the mobile node, e.g. through DHCP or PPP. In the latter case, the mobile node is said to have a co-located care-of address. This mode of operation obviates the need for explicit mobility support, in the form of a FA, on the foreign subnet. The recipient of a tunneled packet recovers the original datagram before processing it further.

Mobile IP assumes that routing of unicast traffic is based solely on the destination address. Many Internet routers now include other considerations in their forwarding decision, e.g. in an effort to guard against IP-spoofing attacks, source-filtering routers drop datagrams that arrive on an interface inconsistent with their source address [CA9621]. Such a router, if present on the foreign network, will block all datagrams generated by a visiting mobile node carrying its home address as source. A solution to this problem is to use a reverse tunnel directed from the mobile node’s care-of address to its home agent [Mont96]. Under this arrangement, datagrams sent from MN to a correspondent node (CN) now carry the care-of address (rather than the home address) as source in the outermost IP header and pass unchallenged through source-filtering routers to the mobile node’s home agent. The home agent strips the outer IP header and recovers the original packet. From then on, the packet is forwarded as if the mobile node were on its home subnet.

Additionally, many organizations use firewalls to protect their networks from the general Internet. These firewalls impose additional constraints, e.g. they may drop unsolicited datagrams from untrusted external hosts [ChZw95]. Such a policy is enforced by carefully screening the source and destination addresses, as well as ports, on all transport level packets. For TCP packets, the firewall may also monitor the ACK bit. In this situation, a mobile node’s registration requests are likely to be dropped by a firewall protecting its home network. Note that source-filtering is not an issue here because
registration requests arrive with a topologically correct source address - namely the current care-of address of the mobile node.

To complicate matters, organizations often hide the topology of their internal network by using private addresses. These addresses are not advertised to the general Internet. Such addresses include, but are not restricted to, those defined in RFC 1918. The Internet's routing fabric is unable to route packets to these addresses (resulting in ICMP unreachables). To allow connections from the internal network to the general Internet, application relays (also called application gateways or proxies) are used. In a typical configuration, the internal network is separated from the general Internet by a "perimeter network" on which the firewall and proxies are located [ChZw95] (see Figure 1). Hosts on the peripheral network use public addresses, i.e. their addresses are advertised to the general Internet. When a host on the internal network wishes to connect to the Internet, two separate connections are set up: one between the internal host and the proxy and another between the proxy and the outside host. To the external host, the user at the other end appears to be on the proxy host.

![Figure 1: Screened-subnet firewall architecture.](image)

The use of private addresses on firewall-protected networks poses an additional challenge. A mobile node belonging to such a network can not use its home address (a private address) to communicate directly with correspondent nodes when it is outside the protected domain as replies from correspondent nodes to the private address are likely to generate a "host unreachable" ICMP message. If, somehow, a reverse tunnel can be established from the mobile node to its home agent, the
mobile node can continue using its private home address. Datagrams generated by the mobile node using its home address will appear to emerge from its home network and connections to external hosts will still involve an intermediate proxy.

The presence of intermediate firewall(s) disrupts free flow of packets from a mobile node on the outside to its home agent on the inside. Appropriate security mechanisms are required to successfully traverse these firewalls to achieve required connectivity.

2. Proposed Solution Framework

The firewall traversal problem in its most general form (e.g. multiple firewalls under different administrative controls with limited mutual trust) does not lend itself directly to an easy solution. Therefore, it is reasonable to try solving this problem in several phases starting with the simplest (and still useful) variant described below.

Consider Mobile IP operation for a mobile node (e.g. a notebook computer belonging to a corporate employee) when it is moved outside its firewall protected home domain into the general internet (e.g. a university or ISP environment) which imposes no access restrictions other than network ingress filtering. In order to reach its home subnet, packets from a mobile node may need to traverse Multiple firewalls. However, they all belong to the mobile node’s protected domain and their existence and relative placement (with respect to a mobile node’s current location) is known apriori.

The simplified problem does not address the case when some of the intermediate firewalls belong to administrative domains other than the mobile node’s, nor does it address the case when a correspondent node (outside the mobile node’s protected domain) is itself protected by a firewall. This latter issue, however, is orthogonal to mobility since the problem is unchanged even when the mobile node is at home.

A reasonable aim is to provide a mobile node with the same connectivity (modulo performance penalties) outside its protected network that it enjoys at home. Additionally, it should be possible to encrypt all traffic to/from the mobile node in an effort to preserve the level of privacy that a mobile node enjoys at home.

The problem of dynamically discovering intermediate firewalls should be treated separately at a later time (see item (e) in Section 3).

3. Issues to Consider
This section discusses several important issues that should be considered while solving the firewall traversal problem for Mobile IP.

(a) Security processing of registrations requests sent on behalf of (or by) a Mobile node must not depend on its reachability at its home address. A violation of this assumption will create a circular dependency since a Mobile Node’s reachability at its home address is at best uncertain until the registration request gets past the firewall.

Solutions that involve a separate phase in which the mobile node negotiates access parameters with the firewall should attempt to minimize the number of round-trip delays.

(b) Colocating Home Agent and Firewall functionality on the same host considerably simplifies the firewall traversal problem. However, doing so effectively restricts the number of networks with mobility support. For example, in Figure 1, hosts on the perimeter network would enjoy mobility support but none on the internal network. Therefore, a solution must not require Home Agent and Firewall functionality to be colocated.

(c) It is possible to configure intermediate firewalls such that UDP packets sent to port 434 of "well known" Home Agents are allowed to pass through without any IPSEC processing. This solves the firewall traversal problem for registration requests but not for reverse tunneled traffic. This approach also involves administrative maintenance of the list of "well known" home agents and assumes that all processes running on port 434 of these hosts are trusted. Ideally, a firewall should decide to let in a packet based solely on its IPSEC characteristics (e.g. is it authenticated, encrypted or both) rather than any higher level information. A solution that does not require Firewalls to understand Mobile IP is preferred.

(d) If firewalls are unable to parse Mobile IP registrations, complications arise when the mobile node uses a care-of address belonging to a Foreign Agent (rather than a co-located address). The main problem is that even if the Foreign Agent (FA) implements IPSEC and is able to establish a security association with the Mobile Node’s firewall (FW), it can not win the Firewall’s trust. A valid authentication header on a packet generated by a FA only tells a FW that the packet was indeed generated by the FA. However, the FW has no way of knowing whether it was generated on behalf of a trusted mobile node. The FA may have generated this packet on its own and the FW may not trust the FA. (Obviously, a mobile node must not divulge the secret it shares with the FW to its FA just so registrations can go through.)
Consider the case of a FA relaying a mobile node’s (MN) registration request. For this request to pass through FW, it must have an authenticator based on a security association between FW and MN. We could have MN precompute an "IP Authentication Header (AH)" [RFC 1826] authenticator which, if inserted by the FA in the relayed datagram, will pass authentication at the firewall. The tricky part here is that the AH authenticator includes the Identifier field of the immediately preceding IP header and there is no easy way for MN to predetermine its value generated by FA for the relayed request.

There is a questionable solution. The MN could generate its own Identifier and convey it to the FA with the understanding that the FA must use this value in the IP header of the relayed request. This value, the AH authenticator and the firewall address etc could all be bundled in a new MN to FA registration extension. With this mechanism, the FA need not even be an IPSEC node. For the case we are considering (FA belongs to a trusting university/ISP environment), it is quite likely that none of the mobility agents are capable of IPSEC processing. However, this approach gets more convoluted when a mobile node must separately authenticate itself to multiple firewalls (see item (e) below).

In contrast, Mobile nodes operating with a co-located care-of address are a lot simpler to deal with. In this case, requests seen by the firewall are generated directly by a mobile node for which a trusted relationship already exists.

Solving the firewall traversal problem for mobile nodes registered through foreign agents may perhaps be postponed to a later phase.

(e) When multiple firewalls within the home domain must be traversed, each firewall may require a different level of trust, e.g. a large corporation may have one firewall guarding its connection to the Internet and another guarding the finance department network from the rest of the company (Figure 2). Amongst all the employees that are allowed access through the main firewall, only a small subset may be allowed in through the finance firewall.
Figure 2: Nested firewalls requiring different levels of trust.

In this scenario, a mobile node on the outside that wishes to communicate with its home agent in the finance network must be able to explicitly (separately) authenticate itself to each intervening firewall.

(f) Many organizations use firewalls along with private addresses. If private addresses are used by the Mobile node’s home domain, these addresses must not be exposed to routers on the outside. This may require additional tunneling e.g. a tunnel from the care-of address to at least the (outer most) Firewall may be needed to hide/bury the Home Agent’s ‘private’ address as destination on reverse tunneled traffic.

The problem is complicated when routers within the home domain are (by design) kept unaware of external destinations. Analogously, extra tunneling may be needed even on the inside, e.g. in Figure 2, a tunnel may be needed between a home agent on the finance network and (the proxy) FW1 to hide the ‘external’ care-of destination address on encapsulated packets meant for a mobile node roaming outside.

So far we’ve only considered the problem of exposing addresses as destinations to routers that are unaware of them. To make things worse, routers are likely to disallow unknown addresses even in the source field of the datagrams they forward. For example, internal routers may disallow an external care-of address as source on reverse tunneled traffic. This may necessitate additional tunneling within the protected domain even for incoming packets e.g. an external care-of address may need to be hidden from R (Figure 2) on reverse-tunneled traffic. Similarly, extra tunneling may be required to hide internal addresses (e.g. home agent’s) from appearing as source on outgoing packets routed by external routers.

Due to the widespread use of private addresses, solutions to the firewall traversal problem must accommodate private addresses.
(g) Proposals for dynamic discovery of intervening firewalls must not rely on firewalls generating ICMP "administratively unreachable" messages since many firewalls are designed to hide as much of their presence as possible.

According to Chapman and Zwicky [ChZw95, Chap 6, Section titled "Returning ICMP Error Codes"]:  

"Returning the new ‘host administratively unreachable’ or the fact that there is a packet filtering system at your site, which you may or may not want to do. These codes may also cause excessive reactions in faulty IP implementations."

"If your router returns an ICMP error code for every packet that violates your filtering policy, you’re also giving an attacker a way to probe your filtering system. By observing which packets evoke an ICMP error response, attackers can discover what types of packets do and don’t violate your policy. You should not give this information away, because it greatly simplifies the attacker’s job. The attacker knows that packets that don’t get the ICMP error are going somewhere, and can concentrate on those protocols, where you actually have vulnerabilities. You’d rather that the attacker spent plenty of time sending you packets you happily throw away."

"All in all, the safest thing to do seems to be to drop packets without returning any ICMP error codes. If your router offers flexibility, it might make sense to configure it to return ICMP error codes to internal systems (which would like to know immediately that something is going to fail, rather than wait for a timeout), but not to external systems ... "

(h) The Authenticated Firewall Traversal (AFT) working group of the IETF has developed the SOCKS protocol [LGLKKJ96] as a general mechanism for firewall traversal. The protocol is conceptually a "shim-layer" between the application layer and the transport layer, and as such does not provide network layer gateway services, such as forwarding of ICMP (or even IPinIP) messages. IPinIP messages (from HA to MN) would have to be encapsulated within UDP or TCP in order to use SOCKS. Beside the resulting inefficiency, there are other concerns regarding its appropriateness for the problem at hand [MoGu96].

The SOCKS solution requires that the mobile node -- or another node on its behalf -- establish a TCP session to exchange UDP traffic with each firewall. SOCKS incurs a minimum delay of four round-trips (six with GSS-API) before a client is able to pass data through the firewall. It seems unlikely that this negotiation will be efficient enough to allow a mobile node to change its point of attachment as
often as once per second, as specified in [Per96a].

Given the likelihood that most firewalls in the future will be based on IPSEC mechanisms, it is imperative that a solution be based on standard IPSEC protocols, e.g. ISAKMP/Oakley.

4. Security Considerations

Mobile IP assumes that routing of unicast datagrams is based solely on the destination address. Packet filtering routers and Firewalls include additional considerations in their routing decisions. Special mechanisms are needed to ensure Mobile IP operation in the presence of these and related (e.g. use of private addresses) security measures currently becoming popular on the Internet.

Acknowledgments

Ideas in this document have benefited from discussions with at least the following people: Gabriel Montenegro, Rich Skrenta and Tsutomo Shimomura.

References


Author’s Address

Vipul Gupta
Sun Microsystems, Inc.
2550 Garcia Avenue
Mailstop UMPK 15-214
Mountain View, CA 94043-1100

Tel: (415) 786 3614
Fax: (415) 786 6445

EMail: vipul.gupta@eng.sun.com

Steven M. Glass
FTP Software
2 High Street
North Andover, MA 01949

Tel: (508) 685 4000
Fax: (508) 684 6105

EMail: glass@ftp.com