Mobile IP as seen by the Internet Engineering Task Force

Charles E. Perkins
perk@watson.ibm.com
IBM, T.J. Watson Research Center
Hawthorne, NY 10562

Abstract

Due to advances in wireless communication technology there is a growing demand for providing continuous network access to the users of portable computers, regardless of their location. Existing network protocols cannot meet this requirement since they were designed with the assumption of a static network topology where hosts do not change their location over time. The IETF has developed a model which fits naturally into the usual framework for routing over Internet Protocol (IP) networks. The requirements which shaped the evolution of the IETF model are outlined. The entities characterized by model are defined, and their interactions described. A suggested way in which these interactions occur within the existing infrastructure is mentioned, and proposed extensions are briefly outlined. A proposed interface to Layer 2 is also shown.

1 Introduction

The Internet Engineering Task Force is an internationally recognized group of network engineering experts which meets three times a year. There are about 70 "working groups", which convene at the IETF meeting site, but which also conduct a significant amount of their business by mailing lists. Each group focuses on a different area of investigation; there are numerous groups looking at various routing protocols, ATM operations, Domain Name Service extensions, and many others. In particular, the "mobile-IP" working group has been organized to investigate the protocol requirements and techniques for correct operation of mobile computers from the Layer 3 perspective. It is by no means sufficient for a Working Group to achieve consensus at one of the thrice annual meetings, although that is highly desirable. Consensus requires also that all the mailing list participants from around the world have a chance to review any proposed agreements, and make technical comments.

As time has passed, the mobile-IP working group of the IETF (which will be called the mobile-IP WG) has studied numerous proposals for enabling mobile networking. Each proposal has had advantages and disadvantages. Each proposal has effectively advanced the "state of the art", so that by now we feel that we (collectively) have a pretty good handle on the overall problem area. My intention in this paper is to discuss the goals, problems, and solutions for mobile networking. I will explain the mobile-IP WG model for mobility, and hopefully put forth a convincing case that this model provides a fairly complete solution which naturally fits into Layer 3 (the network layer) of the commonly accepted protocol stack. We have framed the problem essentially as a packet forwarding problem – in other words, we have been trying to solve the problems presented by mobile computers, by finding ways to forward packets to wherever they may currently be located.

2 Goals

Users of mobile computers will basically desire to use their mobile units as easily as today's office dwellers use desktop computers. We all know the benefits of having large networks of computer resources. People with mobile computers will naturally expect to have the same resources available to them. This will include the current methods for interpersonal communications. Ideally, and practically, this extra facility for mobility must be available without requiring user intervention. Clearly, we want to engineer our solutions so that network efficiency is not affected; this means, for instance, that a good solution will not use features that cause router performance to deteriorate markedly. Just as clearly, we cannot institute a protocol which requires undue host processing. The most important requirement is that the solution must provide the same network connectivity to network resources as is available today. That implies compatibility with existing computer equipment, and application transparency. The solution cannot require expensive equipment, compared to the computer itself.

At the mobile-IP WG, we decided on three "hard" requirements:

- Interoperability with present computer equipment and applications
- Continuous access across multiple networks
- Security as good as with existing networks

The first hard requirement is obvious. We can't expect people to use a solution that does not let them access their existing resources. Meeting the second criterion allows people to use mobile computers in about the same way as they use stationary computers, within obvious physical limits. And, while the mobile-IP WG recognizes that mobile computers seem to create additional security problems compared to stationary computers, nevertheless we have, tentatively, decided to avoid making security the main focus of this work. We intend to analyze fully the security exposures of the protocol, and to make it easy to insert any reasonable security mechanism when it becomes available. There is another IETF working group charged with the job of specifying a good security procedure for IP, and we will await the results of their efforts. There are people who are involved with both groups.

3 Problems

Historically, a "network" corresponded nicely to a long piece of wire. Maybe the network wasn't really a piece of wire, but the equipment manufacturers were nice enough to make it seem so. One way of expanding a network to accommodate more computers is to add more equipment, and more wires connected together with things like repeaters and bridges. However, eventually this strategy runs out of steam, and different networks must be connected together. The different networks occupy different sequences of addresses within the network layer addressing scheme, and packets between networks are handled by network layer routers. Such has been the model used very successfully by the Internet Protocol. However, it is not immediately clear what happens when no direct relationship can be made between a Mobile Host's network address and a piece of wire connected to a router.

Since hosts were thought of as being hooked up to a network, it was natural enough to encode network identification information within the network layer address. This network layer address now serves at least two distinct roles. One, it serves as an "endpoint identifier" – a way to specify what entity resides at the endpoint of a communication path between two hosts. The other use for the network address is as a roadmap for locating the computer within the Internet. This roadmap is followed by isolating out the network identification part of the address, and subsequently forwarding the packet to whatever agent is currently routing packets to hosts on that network. This is usually done using a simple hop-by-hop algorithm whereby each intermediate router keeps track of what the next hop is along the way, for each particular destination network.

Then, once the particular network (wire, perhaps?) is reached, the remaining network layer address bits specify which host on the network should receive the packet. Once the packet reaches the appropriate network, it is assumed that the particular host involved will receive the packet. It is this last assumption which breaks the worst in the realm of wireless networks. Even given a consistent model of a wireless network which has a router advertising reachability to that network, it is mostly not true that the packet will get to the particular destination without the active and special intervention of the router. When computers are mobile, the network address no longer specifies the location.

Mobile computers introduce another new problem that has never been dealt with by IP and most other network layer protocols – that of tracking the movements of the computers. It is not a surprise, then, that IP doesn't have any built in facilities for handling this, unless one counts the ability to change routes dynamically. Issuing location updates in response to movements of the mobile computers will be a central feature of any acceptable mobile networking protocol. Controlling the dissemination of this data turns out to be of central importance, and the matter has yet to be fully resolved. Clearly, one would like to make sure that current location information is available wherever it is needed, but nowhere else. And, the level of need is not a simple binary decision. The wider dissemination of location data is advantageous in allowing better routing, but disadvantageous when considering how to eliminate stale information.

Of course, one of the prime requirements for any workable system is compatibility with existing systems. This translates directly to the requirement that mobile computers using any protocol we specify must be able to maintain network connections with existing hosts, using the routing services of existing routers. We can add equipment (e.g, mobile computers, wireless transceivers) but we can't design a system that doesn't work with today's computers. This same requirement applies to application software, which is the only reason most people have computers. Mobile computers using a new protocol cannot expect to recompile or relink existing application software, nor run only software which is newly designed to enable mobility. The best we can expect is that necessary protocol changes will be installed into the protocol service itself, and only there, on the additional equipment.

One consequence of application transparency, as mentioned above, is that the network address of the mobile computer cannot change when the computer moves. Otherwise, movement of the computer might require rebooting or restarting applications, which seems to be unacceptable from the perspective of user convenience.

It may be quite important for the new protocol and procedures to allow the determination of optimal paths to the mobile computer, even as the user moves from place to place. This ability to dynamically determine optimal routes in the face of changing information is a very hard problem in general, but fortunately for use there is a big restriction we can place on the general problem. Namely, we can assume that the mobile computers get access to the rest of the network through fixed attachment points. The attachments points are only a single hop away from the mobile computers, so finding the mobile computer is about the same as finding its current attachment point. Nevertheless, choosing a strategy to allow the optimization of routing paths to mobile computers turns what might be a pretty simple problem into a matter requiring great consideration.

Last but not least, it must be emphasized that security in mobile networking is a matter that must be addressed, especially for wireless mobile networking. This has implications for every aspect the design, and in some cases has been the determining factor for the relevant design decisions.

4 TCP/IP

IP stands for Internet Protocol, and is a widely used, connectionless or "best-effort" layer 3 protocol [10, 2]. The initial development of IP was funded by DARPA, and it gained widespread popularity with the development and distribution of BSD Unix from the University of California at Berkeley. IP "internetworks" separate networks together, to form a cohesive routing infrastructure in which (usually) any computer with an IP address can exchange packets with any other computer possessing an IP address. IP has grown to the point of offering worldwide connections to over a million computers, a feature no other computer network can offer. Moreover, IP has continued to approximately double in size for several years now, and does not seem to show much sign of slowing down.

Routers are responsible for delivering packets from one network to another within this infrastructure. However, IP does not *guarantee* delivery of a packet to its eventual destination. It merely acts upon the best information available to deliver the packet hop by hop to the destination. If, at some stage, there is inaccurate or stale routing information, packets will possibly be lost. If a particular destination becomes momentarily uncommunicative, IP will not take any additional measures to retry the transmission later. In fact, each packet may be delivered along different routes, which then may require different times to accomplish the end-to-end delivery. This may result in packets being delivered out of order. Correcting these errors and possibly others is the job of higher-level protocols, in particular the Transmission Control Protocol (TCP) [11].

TCP sequences packets and presents the appearance to its applications of a reliable data stream. TCP accomplishes this by (among other things) retransmitting packets when they appear to have gotten lost. In addition to this, TCP takes certain measures to avoid network congestion as well as host application overruns, so that an application using TCP can count on several flow control features as well as the basic reliable data service. TCP even performs source route reversal on behalf of its applications.

We have chosen to frame the requirement of enabling computers to move freely, as a routing problem. That is, when a computer moves from one place to another, we characterize the design problem as one of ensuring that packets can be forwarded to the current location of the computer. Thus, we choose to make computers mobile by discovering what needs to be done to the Internet Protocol. This will have the effect of making all IP applications transparently mobile, without requiring any changes to them. The applications most likely will not be able to tell when the computer has moved.

5 Definitions

In this section, we will define terms, and describe a basic model. We will describe the basic relationship between the entities populating this model system.

Host Any computer, not considered to be performing routing or bridging functions.

Mobile Host A Host which moves from place to place, invalidating historical Internet design assumptions about static placement of computers.

Correspondent Host A Host communicating with another Host; particularly in most relevant discussions, communicating with a Mobile Host.

Home Address An address used to identify a Mobile Host no matter where it may currently be located (cf: discussions about"endpoints").

Foreign Address An address used to locate a Mobile Host at some particular instant of time.

Foreign Agent A specialized forwarding agent which offers a Foreign Address, and maintains and performs a mapping between that address and the Home Address of a Mobile Host in its care.

Home Agent An agent that redirects or tunnels packets from a Home Network to a the Foreign Address of a Mobile Host.

Home Network The (logical) network on which a Mobile Host's Home Address resides.

Ad-Hoc Networking Networking between Mobile Hosts in the absence of any other agents.

Triangle Routing A situation in which a Correspondent Host's packets to a Mobile Host follow a path which is longer than the optimal path because the packets must be forwarded to the Mobile Host via a Home Agent.

Weak Security The same level of security provided by today's Internet. A weakly secure system does not protect against snoopers populating the normal path along which a packet would traverse.

5.1 General description of model entities

Our system involves participation of three types of entities, viz., *Mobile Host*, *Foreign Agent* and *Home Agent*. The networking architecture that we assume is that of a set of Foreign Agents connected through a wired backbone. A Foreign Agent supports at least one interface which is being made available to Mobile Hosts, and makes available a kind of forwarding address (the Foreign Address) to any Mobile Host accepting its services.

Within one campus or administrative domain there could be multiple Mobile Subnets. Each Mobile Subnet can be handled by a separate Home Agent, or the same Home Agent can handle several otherwise unrelated Mobile Subnets. Unlike other routers, a Home Agent is not required to have an interface corresponding to the Mobile Subnet it serves. The association between each Mobile Host and its current Foreign Address is kept at least by its Home Agent.

A Mobile Host retains its Home Address regardless of its current Foreign Address. It can start sessions with other hosts (both mobile and stationary) and obtain new Foreign Addresses without disrupting any active sessions. The movement of a Mobile Host is completely transparent to the running applications, except possibly for a momentary pause which may occur during the transition. A Mobile Host has only one Foreign Address at any given time. Even if two different Foreign Agents are serving the same physical area, a Mobile Host accepts service from only one of them. A Foreign Agent can offer service to multiple Mobile Hosts. There isn't any relationship in general between the Foreign Address of a Foreign Agent and the Home Network which the Mobile Host belongs to, and similarly there is no relationship between the addresses of the Foreign Agents and the Home Agents. The relationship between the various mobile networking entities is illustrated in figure 1.

We use the term *Correspondent Host* to refer to the host communicating with an Mobile Host. In the following discussion, a stationary correspondent host is also referred to as *Stationary Host*.

5.2 General operation of the IETF model

As a result of the way we have framed the problem, solutions occur more naturally within the visible design space. We basically propose that the movement of Mobile Hosts can be enabled by solving a simply stated routing problem. Namely, we can achieve our goals by finding a way to route packets between the Home Agent and the Foreign Agent. Packets destined for a Mobile Host will, unless special measures are taken, be forwarded to the Home Agent by normal IP forwarding methods, because the Home Agent advertises connectivity

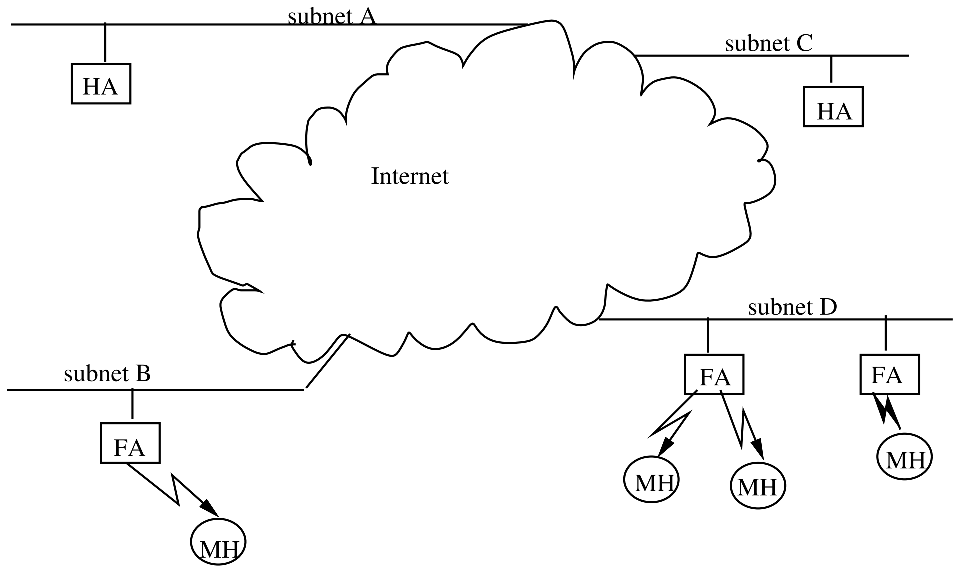

Figure 1: Mobile Networking Entities in the Internet (HA=Home Agent, FA=Foreign Agent, MH=Mobile Host)

to the Home Network specified by the address of the Mobile Host. Once packets get to the correct Foreign Address (i.e, the Foreign Agent), they will be delivered correctly by the natural action of the Foreign Agent. And, we can get packets from the Home Agent to the Foreign Agent by the simple expedient of encapsulating them and inserting the Foreign Address as the destination IP address of the encapsulated packet.

A location update function is also needed, to allow the Home Agent to know where all of its Mobile Hosts are. And, lastly, the Mobile Host needs to complete a transaction with the Foreign Agent before the Foreign Agent will agree to make its advertised services available to the Mobile Host. But these functions, important as they are, should not obscure the simplicity of the model and the approach of providing a wide-area solution based on inter-network routing. Internetworking is IP's strength, and we expect to use that strength to good advantage by providing this natural model of operation for Mobile Hosts. Even with no further protocol operation it should be clear that this basic model, using encapsulation, can provide the basic routing services needed.

6 Previous Proposals

6.1 IBM Loose Source Routing

This proposal, by Rekhter and Perkins [12, 9], took advantage of the existing IP option which allows hosts to specify intermediate forwarding nodes for return traffic. If each Mobile Host specifies that its current Foreign Agent (access point) is in the source route, then its Correspondent Hosts should be able to return packets to it by the natural route reversal action specified as part of the Loose Source Route option handling. Ideally, this would allow a fairly simple solution within the existing infrastructure, and in addition the routing paths taken would usually be close to optimal. Unfortunately, not many hosts implement Loose Source Routing correctly even though it has been part of the IP specification for a long time. Worse yet, hosts correctly processing Loose Source Route options are susceptible to a fairly serious breach of security. Moreover, existing routers are optimized to handle packets quickly if they don't have any IP options. That means, the existing routers will perform badly if they have to forward a lot of packets with IP options; there have already been instances where significant amounts of source routed traffic have caused major difficulties within the Internet. Thus, any solution involving IP options is unlikely to be popular with the administrators involved with the maintenance of high-traffic backbone routers. Lastly, in the particular case of the Loose Source Route option, no known UDP applications do the route reversal correctly, so that popular services such as NSF and DNS cannot make use of the desirable route optimizations offered by the Loose Source Route proposal.

However, aside from the particular means of route optimization and delivery of packets from the Home Agent to the current Foreign Agent serving the Mobile Host, the overall appearance of the model used with the Loose Source Route proposal resembles the current proposal.

6.2 Sony

Fumio Teraoka at Sony developed an approach [13, 14] which modeled the Mobile Host as possessing a Physical Address and a Virtual Address. The Sony approach relies on a generalized "Virtual Network", which when implemented for IP subdivides Layer 3 protocol processing into two parts, one for handling each of the addresses associated with the Mobile Host. This handling can be specified either by a new IP option, or by a new protocol number which triggers the sublayer processing within IP. Either way, the IP implementation is known as VIP (Virtual IP). The Physical Address of the Mobile Host is obtained by whatever means when the Mobile Host moves to a new location, and the Virtual Address corresponds to a network (real or virtual) which is served by a specialized router. The network served by this router is called the Home Network.

To handle the case of existing hosts, the VIP approach suggests that these specialized routers snoop packets while forwarding them. If the packets emanate from a Mobile Host, and thus contain both a Physical Address and a Virtual Address, Sony routers can glean this information for possible future use. The information gleaned is put into an Address Mapping Table (AMT), whose entries have timeouts and must be managed by deletion or update according to the timestamps of new packets passing by. The next time such a router forwards a packet destined for the Virtual Address of the Mobile Host, the Sony router can modify the IP header so that it shows the Physical Address of the Mobile Host as its new destination. Thus, existing hosts will send out packets destined for the Home Network of the Mobile Host, but that will quickly be corrected so that the packets are destined for the current location of the Mobile Host.

6.3 Carlberg's Host Route

Since existing routers can support host routes (i.e, routes to a specific host that do not imply reachability to other hosts on the same "network"), one approach might be to have the Foreign Agents be routers. Then they could advertise reachability to just those Mobile Hosts which were currently located within their range. This is the basic approach advocated by Ken Carlberg [1] in an earlier paper, although in the context of a ISO/GOSIP Mobile End Systems.

While this approach has the advantage of simplicity ease of understanding, and lack of reliance on new protocols, it seems likely to fall prey to problems of scale. In an organization with hundreds of mobile hosts, and dozens of routers, it is easy to imagine that the constantly required updates flowing through the organizational networks would start to cause difficulties. In the larger Internet, it seems that supporting arbitrary mobility is not feasible.

6.4 Columbia (JI) MSSs

One popular approach was specified and implemented at Columbia University as part of the Ph.D. requirements of John (JI) Ioannides. JI's approach [4, 3] modeled mobility between networks of a campus and solved the problems by the use of Mobile Support Systems MSSs (also called Mobile Support Routers (MSRs)). The MSSs share information about the current whereabouts of Mobile Hosts, and when a Mobile Host moves to a new network, the new MSS and the old MSS coordinate the forwarding of packets to the Mobile Host. The Mobile Host always knows the address of its MSS, so this is feasible. The MSSs conspire to provide the appearance of a virtual mobile subnet spread over a campus-sized number of real subnets, and use a Mobile Internetworking Control Protocol (MICP) for communicating among themselves the location information needed for the Mobile Hosts.

JI's approach specified the use of an encapsulation protocol, IPIP, for delivering packets from one MSS to another. The MSSs cooperatively act as a distributed router for the virtual subnet, and are known to the routers for the other subnets to have reachability to that virtual subnet. JI also specified a beaconing protocol so that the Mobile Hosts could discover which MSS was serving their region.

For mobility outside the campus, JI added a "popup" feature to this basic protocol setup. The Mobile Host would acquire a local address outside the campus, and report that address to one of its home MSSs as its new location. Thus the Mobile Host would effectively act as its own MSS. The Home MSS tunnels packets to the popup, and advertises to the other campus MSSs that it is currently serving the popup Mobile Host.

6.5 Matsushita

Another approach, by Tatsuya Ohnishi, Hiromi Wada, and Brian Marsh of Matsushita[15], specifies the use of Packet Forwarding Servers (PFSs). Packet Forwarding Servers operate somewhat like Sony Routers by intercepting packets destined for Mobile Hosts and forwarding them to the current location of the Mobile Host. A Mobile Host has a Home Address, and when it reconnects, it gets a new temporary IP address. Modified stationary hosts cache the binding between the Mobile Host's two addresses, but unmodified hosts just transmit the packet to the Mobile Host's Home Address. If there aren't any PFSs which know the current binding, the packet will eventually arrive at the home PFS, which will be able to encapsulate the packet (using the Internet Packet Transmission Protocol, IPTP) and deliver it to the Mobile Host.

When a Mobile Host gets an temporary address on a new network, it tries to find a local PFS by using a ping operation. If a PFS replies, its existence will be made known to the Home PFS. Then other Stationary Hosts on that network will be notified by the local PFS if they send packets to the temporary address of the Mobile Host. The local PFS will be notified by the Home PFS when the Mobile Host moves to a new network.

6.6 IBM Readdressing

This proposal (by Perkins and Rekhter) [8] is an attempt to abstract and generalize previous approaches. Except for Carlberg's "host route" idea, all the other proposals assume the existence of a special "Mobile Network", and a specialized router or cooperating set of routers advertisting reachability to that network. The techniques of encapsulation or source routing can be both used to deliver packets from the Mobile Network to wherever the particular Mobile Host is currently located. In the description of this proposal, Mobile Hosts are the clients of Re-addressing Servers, which can readdress packets destined for one of their clients by either stripping the encapsulation or managing the source routes according to IP specification. In addition to remembering which Mobile Hosts are its clients, a Re-addressing Server also may maintain a cache of useful information about the Correspondent Hosts with which its clients may wish to communicate. When the Correspondent Host is itself a Mobile Host, this cache would contain the current location information about that Correspondent Host. In this way, a Readdressing Server can help its clients obtain optimal routing whenever possible. Correspondent Hosts which can perform their own Readdressing can also maintain similar caches. These hosts are characterized as being served internally by their own Readdressing servers. Thus optimal routing can be achieved whenever a pair of Readdressing servers are communicating with each other. One advantage of this approach is that multi-level mobility can be achieved by creating nested Readdressing servers.

6.7 CMU MHRP

Dave Johnson from CMU proposed a variant using Loose Source Routing to effect the management of packet delivery to Mobile Hosts. This was subsequently modified to use a new encapsulation technique which allowed lazy updates to the location information cached at the previous connection points of the Mobile Hosts. The new proposal [5] was called the Mobile Host Routing Protocol (MHRP). The Mobile Hosts are modeled, again, as having addresses on a Mobile Network upon which resides a Location Server which is capable of delivering packets to the Base Stations which are at the current locations of the Mobile Hosts. The Base Stations themselves maintain Location Caches for traveling Correspondent Hosts and previous clients (i.e, Mobile Hosts which have moved). Any agent which wishes to deliver a packet to a Mobile Host, and knows about the current location of that Mobile Host can encapsulate the packet so that the current location becomes the destination. In this way, optimal routing can be achieved. If the Location Cache is stale, and thus a packet is delivered to the wrong current location of a Mobile Host, the agent at the end of the tunnel changes the destination to reflect the more current information and adds its own address as part of the data encapsulated by the newly modified IP packet header. When the packet finally arrives at the correct current location of the Mobile Host, all stale cache entries in the intermediate Location Caches can be updated to contain the current information. Of course, there will rarely be any times when more than one intermediate Location Cache has a stale route entry.

One feature of MHRP is its attempt to specify a minimal encapsulation protocol for the needed tunnelling. In many cases, the encapsulation only requires 8 additional bytes per packet destined for a Mobile Host, in contrast to earlier approaches which used 20 or more bytes per encapsulated packets. In addition, MHRP tried to attend to the necessary details to allow Mobile Hosts to interact correctly with other hosts even when directly attached to its Home Network. In this way, MHRP could be used to effect wired mobility, if for instance a Mobile Host was to be moved from one Ethernet network to another. This could, conceivably, be done without losing any active network sessions, since the physical operation still makes sense. Such wired mobility might not make sense on other kinds of physical networks.

6.8 Myles/Perkins MIP

Building on ideas found in IBM's proposals, MHRP, and Sony's VIP, Myles and Perkins [6] tried to build an improved version of MHRP using a new IP option, the MIP option. The basic model of a Mobile Network served by a Home Agent, and Internet Access Points (IAPs) offering service to Mobile Hosts on arbitrary networks, is preserved. But the Location Caches at the IAPs have timeouts associated with them and can be refreshed by MIP packets, which now have timestamps in them. MIP explicitly offered equivalent operation whether the IP option, or MIP encapsulation, was selected, because we wanted the protocol selection to depend mainly on the specific mechanisms offered, not on the variety of readdressing used to effect those mechanisms. Although technically there may be some slight advantages to using IP options, the disadvantages detailed above for Loose Source Routing made it clear that encapsulation had fewer objections from a practical standpoint.

MIP also specified a method for intermediate routers to participate in optimization of routes between unimproved existing hosts and Mobile Hosts. This operates similarly to the Sony router approach. Much attention was spent trying to improve the reliability of protocol operation in the face of various network or MIP entity failures.

6.9 SMIP (CDPD-like)

Simple MIP (Mobile Internet Protocol), or SMIP, was released by Penners and Rekhter [7] as a simplified approach to enabling mobility without worrying about achieving optimal routing. The concern was that previous attempts to optimize the routing of packets between Mobile Hosts and their Correspondent Hosts left unsolved certain problems of security. For one thing, it was not clear how the integrity of Location Caches would be maintained in the face of a determined opponent wishing to usurp communications to a Mobile Host. For another thing, the various kinds of management information flowing through the network, especially the location updates, could be used by an unfriendly snooper to discover private information about the current location of a Mobile Host. SMIP also was based on the premise that no one could really quantify the extent to which optimal routing was needed, or indeed if it were even likely to be provided by the other proposals. Most of the other existing proposals certainly work best in the hypothetical future when the existing infrastructure has finally been upgraded to fully support mobility.

Another distinguishing characteristic of SMIP was its relatively new approach to cell discovery (left unspecified in MIP and MHRP and others). The SMIP approach mimics the CDPD specification in its overall aspect, and for the first time proposed that the Visiting Register (providing the temporary or "Care-of" address for the Mobile Host as it moves) intercede for the Mobile Host in negotiating the transmission of the location update to the Home Register which plays the part of the Home Agent in SMIP. Either the Visiting Register or the Home Register in SMIP can refuse the new connection if necessary. The same method of approving new connections by multiple registration steps is found in CDPD.

The recognition by Penners and Rekhter of the need to consider the security aspects of transmitting location updates has had a large effect on the recent thinking of the IETF. In addition, the multi-step registration model has been adopted, presumably for the additional control it makes available as well as to leave open the option for possible convergence or simplifying multi-targeted implementations for both IETF and CDPD.

7 Current IETF direction

Having developed a generally agreed-upon model of the system, the next task is to make certain design decisions, going from the abstract to the specific, until a concrete protocol results. This section will attempt to list and justify various design decisions that have been made up until now by the IETF Working Group. Some of the decisions have been made more firmly than others, and where appropriate the degree of consensus will be indicated.

7.1 Encapsulation

In order to deliver a packet from the Home Agent to the current location of the Mobile Host (i.e, its Foreign Agent), the packet has to be "readdressed". Otherwise, as soon as any intermediate router tried to forward the packet, the packet would be delivered back to the Home Agent just as it was the first time. This readdressing can occur by using a source-routing technique, or by actually changing the destination of the packet to be that of the Foreign Agent. Although architecturally elegant, source routing techniques have been found through hard experience to have certain grave drawbacks in practice, given the structure of our current Internet. Among the several problems, we have found that loose source routing:

- Slows down intermediate routers
- Exacerbates existing security problems
- Is implemented incorrectly by TCP in most computers, and
- Is not handled correctly by any known UDP applications

Since it is unlikely that existing computers will repair their implementation of Loose Source Routing any time soon, we have to accept the current situation as a characteristic of the existing infrastructure regardless of any protocol specification. The additional security exposure is presented by using Loose Source Routing because the common schemes propose to include the Foreign Agent's address as a component of the source route. Without additional precautions, this would allow any computer within the Internet to pose as the Foreign Agent for a Mobile Host, and thus be able to inspect a stream of data between that Mobile Host and any Correspondent Host. Effectively, the interloper would thus be able to pose as any desired Correspondent Host.

Thus, encapsulation has been selected as the preferred method of readdressing. Encapsulation means that data from the original IP header is prepended to the IP data field, and the original header is then modified to contain different destination and protocol fields. Other fields may have to be changed too, depending upon the exact form of encapsulation chosen by the IETF. Nothing in the protocol depends directly on the style of encapsulation which will be chosen, but it is likely that one style will be required to be supported by all mobile entities involved. Other styles of encapsulation might be supported after optional negotiation steps.

7.2 Solicitation

When a Mobile Host cannot detect any advertisement of available services from a Foreign Agent (see the next subsection), it may decide to solicit service. This might happen, for instance, if a Mobile Host was moving from one wired network to another. In that case, there might be a Foreign Agent making a Foreign Address available on the wired network, but which was not wasting any wired bandwidth with periodic beacons. The device driver for the wired Mobile Host might then have to detect the presence of carrier or some other condition on the wired network. When the condition was satisfied, then the Mobile Host would perform the necessary solicitation for service.

The following fields are good candidates for inclusion in the solicitation packet transmitted by a Mobile Host:

- type, code, checksum
- MH's IP address
- MH's MAC address

Note that the last two fields are found in the packet headers for layers 3 and 2 respectively, and so do not necessarily have to be included in the data part of the packet.

The solicitation would to be either broadcast or multicast, because the Mobile Host would not necessarily have, a priori, any indication of the local Foreign Agent's MAC address. Similar considerations might apply to wireless connections also. That is, a Mobile Host might decide to send out a solicitation for service whenever the current Foreign Agent seemed unavailable. Solicitations are useful whenever there is no periodic beacon (service advertisement), and perhaps even when the periodic beacon has been delayed or omitted for whatever reason.

7.3 Advertisement

As indicated in the last subsection, a Mobile Host expects to receive service advertisements from a local Foreign Agent. The exact format of this advertisement remains incomplete, but certain fields are very good candidates for inclusion. for instance:

- type, code, checksum
- Foreign Address
- Foreign Agent incarnation number
- advertisement interval

Also, just as a natural part of the transmission of the advertisement packet, all potential clients will discover the base station's MAC address and its IP address. This IP address may be distinct from the Foreign Address, however, because the base station may be distinct from the Foreign Agent; for example, there might be a wired network with a single Foreign Agent and multiple base stations. The exact method for transmitting packets between a Foreign Agent and the base stations in this situation is not going to be specified by the IETF Working Group, but we have to be aware of this possibility and make sure the protocol allows for correct operation in this case.

The incarnation number is included for use by the Mobile Hosts when a Foreign Agent crashes. Whenever the Foreign Agent crashes, it will have to increment its incarnation number. When any Mobile Hosts, which remain clients of the Foreign Agent, discover that the incarnation number has changed, they will know that the Foreign Agent may have forgotten about them. That Foreign Agent may have to be reminded to provide service for the Mobile Hosts. This would presumably look much like the process of entering the service area of a new Foreign Agent.

If the Foreign Agent sends out a service advertisement in response to a solicitation made by a Mobile Host, then the response can be sent out as a unicast message instead of a broadcast or multicast (beacon) message.

7.4 Mobile Host $\longleftrightarrow$ Foreign Agent Registration

The Registration protocol will probably consist of 4 messages. First, is a registration message between the Mobile Host and the Foreign Agent, covered in this subsection. In the next subsection the registration between the Foreign Agent and the Home Agent on behalf of the Mobile Host will be discussed. Finally, there are two acknowledgment messages corresponding to these two registration messages. It is possible to imagine that the same packet type might perform both positive acknowledgment functions, as well as negative acknowledgments.

Once the Mobile Host discovers that it can get service from a Foreign Agent, it must register with that Foreign Agent. Likely candidates for inclusion in this registration packet are as follows:

- type, code, checksum
- Home Agent's address
- Sequence number
- previous Foreign Address
- Mobile Host authenticator to Home Agent
- Mobile Host authenticator to Foreign Agent
- Mobile Host's IP address
- Mobile Host's MAC address

Notice again that the Mobile Host's layer 2 and 3 addresses may be found in the registration packet's headers.

The first authentication field is present so that the Home Agent can be somewhat assured, as described in the next subsection, that the information about the Mobile Host is indeed coming from that Mobile Host and not an impostor. The Mobile Host also passes an different authentication to the Foreign Agent, because the Foreign Agent may receive update information regarding the Mobile Host in the future, and it would be valuable for the Foreign Agent to discover whether the update information was authentic or not. Note that these are not particularly strong measures for ensuring authenticity, and when more security is required different mechanisms will have to be be built. We expect that these different mechanisms will use registration packets with a different (type, code) pair, even if the rest of the registration protocol is the same.

The sequence number field plays the same role as a timestamp, so that delayed and out-of-order packets do not cause any interruption of service. Otherwise, a stale packet from a nearly-recent cell switch might cause confusion upon its eventual arrival at a Home Agent. This sequence number is the one included in all management packets related to this particular location information for the Mobile Host.

7.5 Foreign Agent $\longleftrightarrow$ Home Agent Registration

When the Foreign Agent is satisfied that it has been contacted by a Mobile Host which is a good citizen, then the Foreign Agent can initiate the next registration phase, by registering its new client with the client's Home Agent. The registration packet will be likely to contain the following information:

- type, code, checksum
- Foreign Address
- Sequence number
- Mobile Host authenticator to Home Agent

There may also be included the address of the base station if that IP address is ever distinct from the Foreign Address; this issue has not been resolved. The sequence number is the same one transmitted by the Mobile Host to the Foreign Agent, and similarly for the authenticator. Notice that the packet can usually be delivered to the Home Agent simply by sending it out with the destination address equal to the IP address of the Mobile Host. Such packets arrive at the Home Agent by the normal operation of the existing Internet, since the Home Agent is well-known as the router for the Home Network. This would relieve the Mobile Host of the administrative requirement to be configured with the address of its Home Agent.

7.6 Registration Acknowledgments

The Home Agent will send a packet back to the Foreign Agent either accepting or rejecting the Mobile Host's request. A positive acknowledgment will be likely to contain the following information:

- type, code, checksum
- Mobile Host's IP address
- sequence number
- Foreign Address
- Cache timeout

The cache timeout specifies how long the Mobile Host can trust the Home Agent to keep track of the Mobile Host's whereabouts. If it is not infinite or zero, then the Mobile Host should attempt to re-register as often as is indicated by the timeout. This feature was included to avoid problems caused by crashing Home Agents. If the Home Agent crashes, and the Mobile Host never re-registers, then it would be possible for the Mobile Host to be unaware that the Home Agent is dropping all packets which otherwise would have been delivered to the Mobile Host.

The Foreign Agent also needs to deliver an acknowledgment to the Mobile Host. This acknowledgment contains the following:

- type, code, checksum
- Mobile Host's IP address

- sequence number
- cookie value
- Cache timeout
- Cell expiration

The cache timeout is the one received from the Home Agent, and the cell expiration value specifies how often the Mobile Host should re-register with the Foreign Agent. The cookie value can be used by the Mobile Host to authenticate future control messages to this Foreign Agent. For instance, when the Mobile Host moves to a new Foreign Agent, the Mobile Host may wish to send a control message to its previous Foreign Agent, to notify the previous Agent of the change.

See figure 2 for an schematic picture of a typical registration sequence.

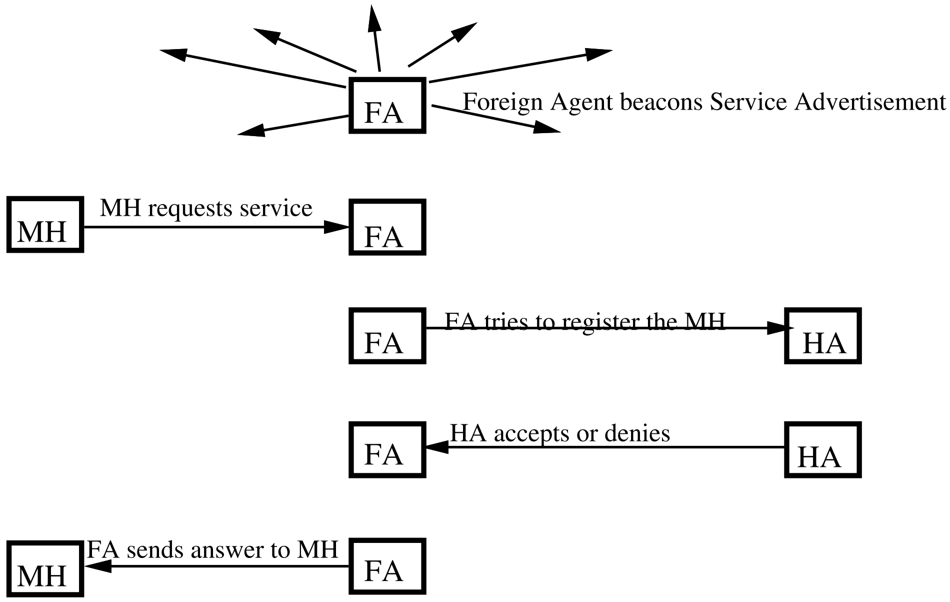

Figure 2: A typical sequence of events for registering a MH

7.7 Remote Redirect

So far, all of the protocol provisions have had the effect of making sure that the Home Agent is able to deliver packets to the Mobile Host. This will satisfy a minimal requirement to get packets delivered to the Mobile Host, because packets addressed to the Mobile Host will find their way to the Home Agent, and from there will be encapsulated and delivered to the appropriate Foreign Agent. However, as discussed previously, requiring the Home Agent to be involved in the delivery of every packet represents a substantial routing inefficiency in many cases. This inefficiency, known as "Triangle Routing" (see 3), can be eliminated by notifying the source of packets destined to the Mobile Host about the current whereabouts of the Mobile Host; in other words, the Correspondent Host could receive and act upon what has been called a "remote redirect". In the figure, the link between the Home Agent and the Foreign Agent is shown as a bold line to indicate that the packet is encapsulated before delivery.

Remote redirect is different than the usual redirect, because it specifies that the Correspondent Host should readdress the packet to the Foreign Address of the Mobile Host,

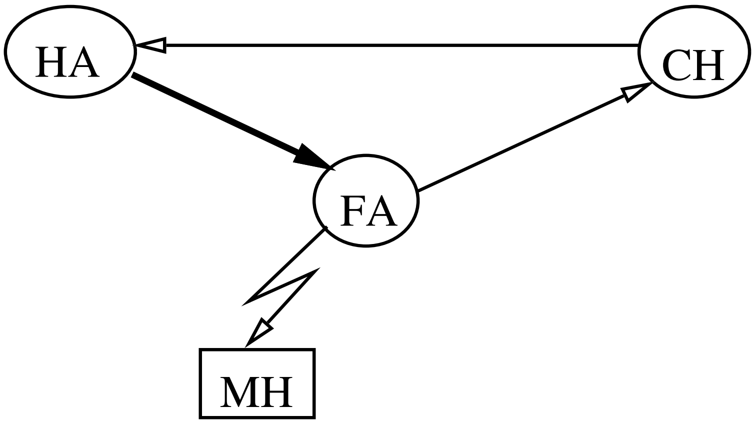

Figure 3: Triangle Routing

instead of the Mobile Host itself. This style of redirection usually specifies a Foreign Address not present on the local network.

Allowing such a feature introduces a new requirement for security, too. When no such redirects occur, the requirement for weak security is automatically met just because all packets to the Mobile Host appear as if they were sent from the Home Agent. And, any Correspondent Host would automatically send packets to the Home Agent, so that no interloper outside the normal path of packets to the Home Agent would be able to intercept these packets. However, with remote redirect, just as with Loose Source Routing, any interloper computer could cause a correctly functioning Correspondent Host to mistakenly route packets through the interloper instead of through the correct Foreign Agent. Thus, such an interloper could pretend to be any particular Mobile Host, unless additional measures are taken to restore weak security.

The additional measures provide that any Correspondent Host may choose to disregard remote redirects until they are validated by the Home Agent for the target Mobile Host. The Home Agent does indeed have the authentic information, and if the interloper is on the path between the Correspondent Host and the Home Agent, then weak security cannot protect the Correspondent Host in any event. Thus, the Correspondent Host will be as assured as it would otherwise be, that it has gotten good routing information about Mobile Host. This procedure is illustrated in 4.

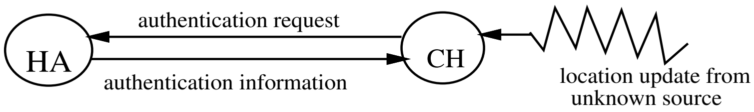

Figure 4: Validating an unsolicited Location Update

The Correspondent Host does not have to have the IP address of the Home Agent to receive this validation; it can just address the validation request directly to the Mobile Host. The Home Agent is specified to intercept validation requests which are addressed to the Mobile Host. These packets will be delivered to the Home Agent anyway, just as any packet to a Mobile Host would be in the absence of remote redirects. The Home Agent then answers the request by consulting its current list of associations between Mobile Hosts and Foreign Agents.

The remote redirect protocol for eliminating Triangle Routing also provides a method for a Mobile Host to notify its previous Foreign Agent of its new Foreign Address whenever the Mobile Host moves. The Mobile Host can send the same remote redirect message to that Foreign Agent. In cases where the Mobile Host does not want to transmit information about its current whereabouts (say, for privacy reasons), it would just notify the previous Foreign Agent that it had moved, without giving a new Foreign Address. When the new Foreign Address is made available, however, the previous Foreign Agent could forward packets to the new Foreign Address instead of causing the packets to be routed back through the Home Agent. In this situation, the magic cookie provided by the Foreign Agent to its clients (the Mobile Hosts) during registration time would be useful to prevent unauthorized remote redirects from effectively disconnecting an unwilling Mobile Host from its current Foreign Agent. The use of such a mechanism can make the forwarding process faster since some Foreign Agents will be satisfied that no additional authentication is necessary.

7.8 No help for "Ignorant" Hosts

At this time, there is no specified means for allowing unmodified hosts to avoid Triangle Routing. There are ideas, which have not achieved consensus, about how to allow additional cache agents to be placed in strategic locations. These agents could help ignorant hosts by readdressing packets according to cached location information gleaned from snooping into passing packets.

8 Layer 2 Interface

There are three possible pieces of information that we would like to receive from the protocol layers below the Network Layer (IP in our case). In each case, the availability of this information may offer substantial performance benefits to the overall operation. We expect to provide correct operation even where no additional information is available, but to obtain the necessary information at the Network Layer may incur a performance penalty that would be wholly unnecessary if the information were already available via other means, say through an interface to lower level protocols.

The first useful piece of information would be to discover when a Mobile Host has entered a new cell, or correspondingly when a Mobile Host has exited the effective range of a previous cell. If this information is not available otherwise, the Network Layer in the Foreign Agent might be required to issue periodic beacons to recover the same information. We would like to avoid the need for sending extraneous beacons.

Secondly, it might be very useful at times to be able to discover when there is a momentary loss of the communications channel between the Mobile Host and its Foreign Agent (or, in alternate terms, between the Mobile Station and the Base Station). Armed with this information, we might in the future be able to do some special buffering until the channel is reconstituted. For instance, in the case of an Infrared data channel, the channel might be momentarily interrupted as someone walks behind an obstacle, and then very soon the channel would be usable again.

Lastly, in some situations we could make good use of the MAC address of the currently serving Base Station. This address would not always be useful, but it might be in many circumstances. For instance, if a cell association event occurs but the MAC address does not change, then in many systems the network layer software might determine that there was no need to perform extraneous registrations between the Home Agent and Foreign Agent, or Foreign Agent and Mobile Host.

It is our express intention to design all of our network layer protocols to be independent of the characteristics of lower level protocols, thus none of the suggested information is required. It is only requested for those cases where it makes sense. The working group does not in any way intend to place burdensome requirements upon the protocol specifications of the IEEE 802.11 committee, which is currently undertaking the huge task of specifying a layer 2 protocol for wireless data communications devices such as radio-frequency and infrared transceivers.

9 Summary

In summary, the IETF has proposed a basic model of operation which naturally provides the services needed to enable network access for Mobile Hosts. The Mobile Hosts can move freely from one network to another, and their applications will continue to operate just as they would from any existing computer, because the packets to and from the Mobile Host are routed invisibly to the application. It would be substantially more difficult to provide the same level of service between different networks by making modifications at the MAC layer or Logical Link layer of the protocol suite.

We expect that the network operation can be made more efficient if certain parameters and conditions are made known from Layer 2 whenever they are available.

Although we have presented the solution in the framework of providing network access for Mobile Hosts using TCP/IP, there is little within this protocol which is really dependent upon IP, and nothing which depends on TCP. Thus, we believe that the basic solution can be adapted for use with other network-layer protocols. And, there is nothing in the protocol which is machine or operating system dependent.

As this document was being prepared, there has been a small group of people that has been working to get a draft proposal out for comment by mid-October. That draft proposal incorporates most of the design elements discussed within this paper. The draft document has been made available for public comment. My hope is that eventually it will have been implemented by enough different people that it achieves the status of Internet RFC, and finally becomes a standard.

10 Acknowledgment

Although it is inappropriate to single out individuals for particular mention in a group effort such as the IETF, I should like to emphasize that the work presented in this paper is largely the product of the whole group, and any individual contributions have benefitted greatly from interactions with the rest of the group. I must also emphasize that, while I attempt to fairly represent the ideas and considerations of the IETF Working Group, I do not speak with any special authority. I am not an official representative of the IETF, nor even the mobile-IP Working Group, even though I have been chosen to provide a liaison between the Working Group and the IEEE 802.11 committee. Other members of the Working Group may well interpret the technical content of past conversations differently than I have here.

References

[1] Kenneth G. Carlberg. A Routing Architecture That Supports Mobile End Systems. In *Proceedings of IEEE MILCOM*, 1992.

[2] Douglas E. Comer. *Internetworking with TCP/IP*, volume 1. Prentice Hall, 1991.

[3] John Ioannidis, Dan Duchamp, and Gerald Q. Maguire Jr. IP-based Protocols for Mobile Internetworking. In *Proceedings of ACM SIGCOMM*, pages 235–245, 1991.

[4] John Ioannidis and Gerald Q. Maguire Jr. The Design and Implementation of a Mobile Internetworking Architecture. In *Proceedings of Winter USENIX*, pages 491–502, San Diego, CA, Jan 1993.

[5] Dave Johnson. Transparent Internet Routing for IP Mobile Hosts. Draft RFC – work in progress, July 1993.

[6] Andrew Myles and Charles Perkins. Mobile IP. Draft RFC – work in progress, August 1993.

[7] John Penners and Yakov Rekhter. Simple Mobile IP. Draft RFC – work in progress, September 1993.

[8] C. Perkins and Y. Rekhter. Support for Mobility with Connectionless Network Layer Protocols. Draft RFC – work in progress, November 1992.

[9] Charles Perkins. Providing Continuous Network Access to Mobile Hosts Using TCP/IP. *Computer Networks and ISDN Systems*, pages 357–370, Nov 1993.

[10] J. Postel. Internet Protocol. RFC 791, Sep 1981.

[11] J. Postel. Transmission Control Protocol. RFC 793, Sep 1981.

[12] Y. Rekhter and C. Perkins. Loose Source Routing for Mobile Hosts. Draft RFC – work in progress, July 1992.

[13] Fumio Teraoka and Mario Tokoro. Host Migration Transparency in IP Networks. *Computer Communication Review*, pages 45–65, Jan 1993.

[14] Fumio Teraoka, Yasuhiko Yokote, and Mario Tokoro. A Network Architecture Providing Host Migration Transparency. In *Proceedings of ACM SIGCOMM*, pages 209–220, Sept 1991.

[15] Hiromi Wada, Takashi Yozawa, Tatsuya Ohnishi, and Yasunori Tanaka. Mobile Computing Environment Based on Internet Packet Forwarding. In *Proceedings of Winter USENIX*, pages 503–517, San Diego, CA, Jan 1993.