

Mobile Internet Access and QoS Guarantees Using Mobile IP and RSVP with Location Registers

Ravi Jain* and Thomas Raleigh
Applied Research, Bellcore

Charles Graff and Michael Bereschinsky
U. S. Army CECOM

Abstract. The Mobile IP (MIP) protocol for IP version 4 as being standardized by the Internet community provides continuous Internet connectivity to mobile hosts, without requiring any changes to existing routers and higher-layer applications. We propose an alternative protocol, Mobile IP with Location Registers (MIP-LR) which is closer to the “service node” database approach used in the Public Switched Telephone Network (PSTN): before launching a packet to the mobile host, the sender first queries a database to obtain the recipient’s current location. MIP-LR is designed for operation in enterprise environments or within logical administrative domains, as it requires a sending host to be aware which hosts are potentially mobile and implement the MIP-LR protocol. The benefits of MIP-LR are that potentially long routes, called “triangle routes”, from the sender to the mobile host are avoided, encapsulation of packets sent to a mobile host is not required, the load on the home network as well as the home and foreign agents is reduced, and there is substantially improved interoperability with protocols such as RSVP for providing QoS guarantees. We carry out a simplified average-case analysis of the costs and benefits of MIP-LR and show it can result in significant reductions in mean network costs compared to MIP.

1. Introduction

There has been tremendous interest in the last few years in the areas of mobile and wireless communications. To provide these advanced services PCS and cellular systems (and the PSTN in general) tend to use a “service node” architecture, where databases store the critical signaling information and intelligence, and switches are optimized for simplicity and high speed. In contrast, mobility and Quality of Service (QoS) support in the Internet are typically provided by means of enhancements or additions to the Internet Protocol (IP) [1] routers in the Internet fabric.

In this paper we consider the situation where a “service node” type of approach is used to provide continuous Internet connectivity to mobile hosts in a controlled enterprise environment. The Mobile IP protocol (MIP) [2] supports continuous Internet connectivity for mobile hosts (MH). An MH is always identified by the IP address it has when in its home network, called its home address. When a mobile host moves away from its home network to a foreign network, it

obtains a temporary Care-Of-Address (COA). The MH registers with a Home Agent (HA), which is typically a router, in its home network, informing the latter of its COA. Any Correspondent Host (CH) wishing to communicate with the MH need not be aware the mobile host has moved; it simply sends IP packets addressed to the mobile host’s home address. These packets are routed via normal IP routing to the mobile host’s home network, where they are intercepted by the HA. The latter encapsulates each such packet in another IP packet which contains the mobile host’s COA as the destination address, and these packets are thus delivered to the mobile host’s new location (a process called tunneling.) Note that packets from the mobile host to the correspondent host need not necessarily be tunneled; the mobile host can simply send them directly to the correspondent host.

A well-known problem with MIP is that it uses “triangle routing”, i.e., packets from the correspondent host to the mobile host must in general travel via three (sub)networks: the correspondent host’s subnet, the home agent’s subnet, and the subnet where the mobile host is currently located. Triangle routing incurs potentially significant overheads in the delay and network resources consumed for communication with mobile hosts. An extension to the basic MIP protocol called Route Optimization (MIP-RO) [3] avoids triangle routing as follows. When a mobile host’s HA intercepts an IP packet it informs the correspondent host of the mobile host’s current COA (this is called a binding update message); the correspondent host can cache this information and send subsequent packets by tunneling them directly to the mobile host’s COA.

A problem with MIP-RO is that packets sent by the correspondent host still use the triangle route until the correspondent host receives the binding update message. A more important implication arises when using protocols such as RSVP [4] for providing Quality-of-Service guarantees to communications between correspondent and mobile hosts. The RSVP protocol provides a mechanism for reserving resources along the path from a source host to a destination host so that subsequent data packets are guaranteed to have certain bandwidth available and meet certain delay bounds. The operation of RSVP can be summarized as follows. The source host sends an initial signaling packet (called a *Path* message) to record the route taken to the destination. The destination node determines the network resources needed to meet the desired QoS, and replies with a resource reservation packet which travels, in reverse, exactly the route taken by

*Address correspondence to: Ravi Jain, Bellcore, 331 Newman Springs Rd, Red Bank, NJ 07701. Phone: (732) 758-2844. Fax: (732) 758-4372. Email: rjain@bellcore.com

the Path message, and as it does so, reserves bandwidth and processing resources at the routers along the reverse path. Subsequent data packets sent by the correspondent host thus enjoy guaranteed bandwidth and resources.

Problems arise when RSVP is used in conjunction with MIP or MIP-RO. Firstly, routers will not be able to recognize a Path message encapsulated while tunneled from the HA to the mobile host, and thus will not record the information required for reservations to be effected by the resource reservation message. The second problem is that the resources will only be reserved along the triangle route from the correspondent host to the mobile host. Since RSVP issues Path messages periodically (in order to overcome the effects of routing changes, etc., that may take place in a fixed network), eventually resources will be reserved along the direct route to the mobile host, but unnecessary delay and resource consumption will still result, and the QoS guarantees desired may not be achieved since packets sent along the triangle route receive different treatment than those sent directly.

We propose a method, called Mobile IP with Location Registers (MIP-LR), to use a set of databases, called Location Registers, to maintain the current COA of the mobile host in a manner similar to that used for maintaining the location of a mobile telephone in a PCS or cellular system [5], and recently also considered for ATM networks with mobile nodes [6]. When a mobile host moves from one subnet to another, it registers its current COA with a database called a Home Location Register (HLR). When a correspondent host has a packet to send, it first queries the HLR to obtain the mobile host's COA, and then sends packets directly to the mobile host; the correspondent host caches the mobile host's COA to avoid querying the HLR for every subsequent packet destined for the mobile host. MIP-LR not only eliminates the inefficiency of triangle routing in MIP, and the inefficiency of triangle routing of initial packets with MIP-RO, but also generally avoids tunneling and allows resource reservation using RSVP to provide QoS guarantees.

In sec. 2 we describe MIP-LR and in sec. 3 we use a simplified analytical model to investigate its costs and benefits of using MIP-LR. The analysis shows the mean network costs can be substantially reduced by MIP-LR. We observe that one of the features of MIP-LR is that, in a manner similar to the Advanced Intelligent Network (AIN) architecture of the PSTN, it provides a separation of concerns between the database functions performed by the home agent (i.e., maintaining the location of the mobile host) and its routing functions (intercepting packets destined to the mobile host and tunneling them to the mobile host's COA). Finally, in sec. 4 we end with some concluding remarks.

2. Mobile IP with Location Registers

In cellular and PCS systems standards like IS-41 and GSM, a two-level system of databases, called the Home Location Register (HLR) and Visitor Location Register (VLR) is used to keep updated location information for mobile terminals (see [5] for a tutorial.) We call such schemes Location Register (LR) schemes.

We note that given the basic framework of using location registers for MIP, the ideas behind the numerous variations of PCS and cellular LR schemes which aim to make them more efficient (by forwarding [7], profile replication [8], local anchoring [9], hierarchical organization [10,11] and other methods) can be applied and leveraged for IP networks, although obviously some modifications may be required in the details.

In MIP-LR each subnet contains a host which functions as a VLR, and a host (possibly the same one) which functions as an HLR. Each mobile host is served by a single HLR, specifically the HLR located in its Home Network. Each VLR and HLR advertises its presence on its local subnet using periodic broadcasts (similar to Agent Advertisement messages for MIP.)

When a mobile host is located at its local subnet it is not registered at either the HLR or the VLR, and originates and receives IP packets using normal IP routing. When the mobile host moves to a Foreign Network it obtains a COA. This can be done by either: (1) Each VLR owns a pool of IP addresses which it can assign visiting mobile hosts as COAs, and broadcasts the currently available list of COAs periodically (or implements some alternate means for mobile hosts to contend for COAs), or (2) The mobile host obtains a COA from a local DHCP server [1].

The mobile host chooses and registers its COA with the foreign VLR, which in turn relays the registration to the mobile host's HLR. The HLR returns a registration reply containing the allowed Lifetime for this registration (similar to MIP); the VLR records the mobile host's COA and the Lifetime and forwards the reply to the mobile host.

A Correspondent Host wishing to send a packet (see Figure 1) to the mobile host for the first time must first discover the IP address of the mobile host's HLR (we will describe this process below, but for the moment note that it needs to be carried out infrequently.) The correspondent host then issues a query to the HLR, which returns the mobile host's COA as well as the remaining registration Lifetime. The correspondent host then directly sends the packet to the mobile host's COA. The IP layer at the correspondent host thus hides the mapping from the mobile host's IP address to its COA from higher layers protocols (e.g. TCP [1]), and the IP layer at the mobile host does the same for the reverse mapping.

The correspondent host caches the mobile host's COA and uses the cached binding for subsequent packets destined to the mobile host. The correspondent host must refresh its binding cache by querying the HLR again before the mobile host's remaining registration Lifetime expires.

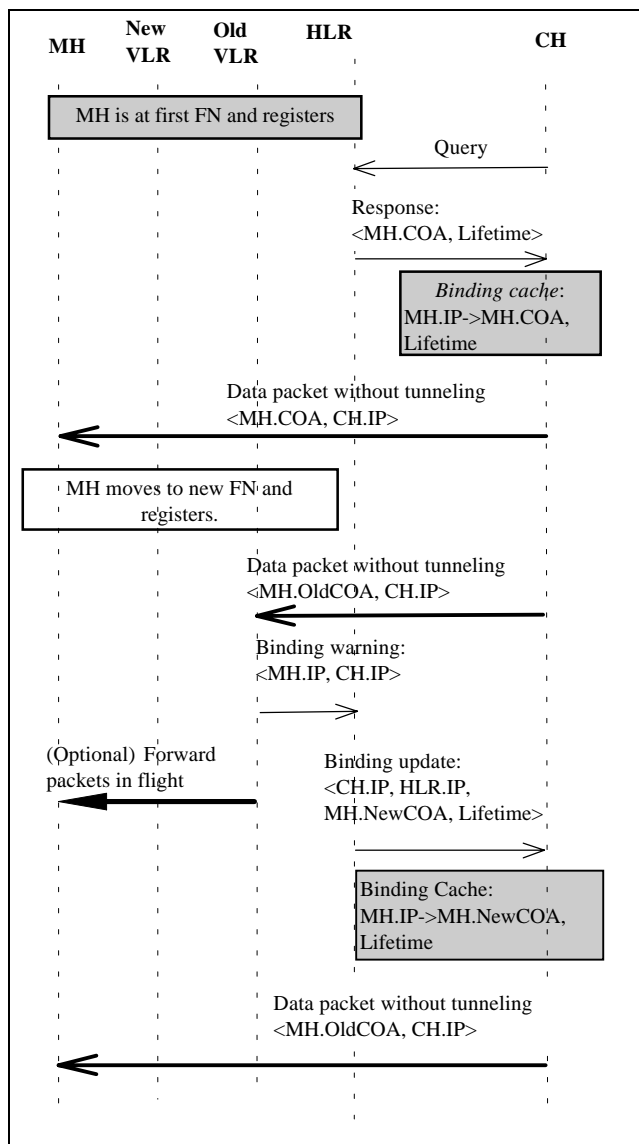


Figure 1: Packet delivery in MIP-LR scheme

After a mobile host moves, if the mobile host was previously registered at some other foreign VLR, the new VLR deregisters the mobile host at the old VLR. If the old VLR manages COAs, it deallocates the COA lately used by the mobile host for eventual reuse. Otherwise, it eventually informs the local DHCP server that the COA can be deallocated.

After a mobile host moves a mechanism is required to update the cache at the correspondent host. In *lazy caching* the

mobile host informs the old VLR (via the new VLR), which traps any packets destined to the old COA and sends a binding warning message to the HLR. The HLR sends a binding update message containing the mobile host's new COA to the correspondent host (as for MIP-RO). In *eager caching* [11] the mobile host maintains a list of all the correspondent hosts it has active connections with, and issues a binding update to each such host. (A form of eager caching has been suggested for Mobile IPv6 also [12].) Note that eager and lazy caching need not be mutually exclusive. Also, the mobile host can, as an option (in case the application being run cannot tolerate packet loss and also cannot tolerate delay), request the old VLR to tunnel packets which it has intercepted to the new VLR.

One of the issues that arises with MIP-LR is how the correspondent host obtains the address of the mobile host's HLR. There are two possible approaches:

1. *Trap query at home subnet.* The correspondent host uses the mobile host's permanent IP address to issue the query. If the mobile host is away from home and has registered, the query will be trapped by the HLR, which will respond with the mobile host's COA.
2. *Database lookup.* Introduce databases, called *Translation Servers* (TS), which store the mapping from a host's IP address to the IP address of the HLR which serves that host. Since this information does not change frequently, a correspondent host can cache the response for relatively long periods of time. The address(es) of the TS must be fixed and well-known to all hosts. (The use of TS has been previously proposed for providing advanced services in PCS systems [13].)

Although the first approach is simpler, the second approach can provide load balancing and better survivability by allowing the TS to contain a list of HLR addresses, and introducing appropriate protocols.

In MIP and MIP-RO the correspondent host need not be aware in advance, which, if any, hosts are mobile. MIP-LR sacrifices this transparency for improved performance and interoperability with RSVP. Different possible approaches to doing this are:

1. *Assume all hosts are mobile.* This approach is simple but the correspondent host then has to issue a query even for hosts which are fixed or mobile hosts which are at home. However, for certain specialized enterprise environments, this may not be unreasonable. Hosts outside this environment use MIP-RO or MIP, and the HLR functions as a Home Agent, tunneling packets to the mobile host.
2. *Address space partitioning.* Certain address ranges in the enterprise system are known to belong to mobile hosts, and correspondent hosts issue location queries only for destination hosts in those ranges. An example of address space partitioning is in the public telephone system (e.g. 800 numbers.)

- Use a directory or off-line discovery mechanism. Correspondent hosts within the enterprise system are periodically notified (or discover) which hosts are mobility-capable and also capable of participating in MIP-LR.

3. Cost Analysis

We develop a simplified “first cut” analytical model of the average costs of using MIP and MIP-LR. The model is then evaluated for specific scenarios by assigning appropriate parameter values, some of which are estimated using empirical measurements.

Assume that a correspondent host generates data packets destined for a mobile host at a mean rate λ , and mobile hosts move from one subnet to another at a mean rate μ . We define *Packet to Mobility Ratio (PMR)* as the mean number of packets received by a mobile host from a corresponding host per move. Assuming the movement and packet generation processes are independent, stationary and ergodic, the PMR is given by $p = \lambda / \mu$.

The distance between two hosts is given by the number of hops between them; thus the distance between two hosts on the same subnetwork is 1, the distance between two hosts which are separated by a single IP router is 2, etc. The distances between the various entities involved in our protocols is shown in Figure 2. We assume that over a sufficiently long period of time these distances actually represent averages. Let the average length of a control packet (e.g. an ICMP packet, a Mobile IP registration packet, etc.) be l_c and a data packet be l_d and define their ratio $l = l_d / l_c$. (In our evaluations we use $l_c = 100$ bytes and $l_d = 1024$ bytes.)

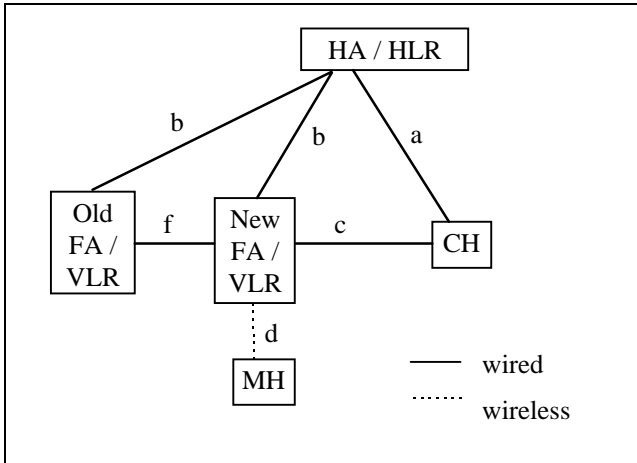


Figure 2: System model for performance analysis

Let the cost of transmitting a control packet be given by the distance between the sender and receiver, and the cost of transmitting a data packet be l times greater. Let the average

cost of processing control packets at any host (and forwarding data packets at a home agent) be r .

Consider the time interval when a mobile host moves to a new subnet, until the instant just before it moves to the next subnet. For MIP, during this time the network cost incurred is $C_{MIP} = m + C_{OldFA} + C_{NewFA}$, where m is the cost of the mobile host registering at the new foreign network, C_{OldFA} is the cost of data packets lost by being delivered to the old foreign agent during the registration delay, i.e., before the home agent is informed, and C_{NewFA} is the cost of data packets delivered to the mobile host via the new foreign agent. Counting the packets transmitted and the processing at the various MIP entities during a registration, $m = 2(b + d) + 5r$. The cost of a single data packet delivered from the correspondent host to the mobile host via tunneling at the home agent is $C_{dr} = l(a + b + d) + 2r$.

Let t_{rd-MIP} be the registration delay for MIP. Then $C_{OldFA} = \lambda t_{rd-MIP} C_{dr}$. Let the average time required for a control message to traverse distance i in the network be denoted t_i and the average time required to process a control or data message at a host be denoted t_r . Then $t_{rd-MIP} = t_b + t_d + 3t_r$.

We make the simplifying assumption that the only cost due to packets being incorrectly delivered to the old foreign agent is that they must be sent by the correspondent host again, and that this retransmission does not affect the PMR. Then $C_{NewFA} = p C_{dr}$.

For MIP-LR, recall that since the correspondent host has a cached mobility binding for the mobile host, when the latter moves data packets are delivered to the old VLR and lost, for a time period t_{rd-LR} until the old VLR is updated with the mobile host’s new COA. Assuming lazy caching, the next packet from the correspondent host which reaches the old VLR causes the latter to issue a binding update to the correspondent host (via the HLR) informing it of the mobile host’s new COA; let the delay for this process be denoted t_{bd-LR} . During this binding update delay, we assume any additional packets generated by the correspondent host are sent to the old VLR and forwarded to the mobile host via the new VLR.

Let C_{rd-LR} be the cost of updating the old VLR with the mobile host’s COA, C_{dd} be the cost of a data packet sent via the direct route from the correspondent host to the old (or the new) location of the MH, C_{bd-LR} be the cost of the binding update issued by the old VLR to the correspondent host via the HLR, and C_{df} the cost of forwarding a data packet from the old VLR to the new VLR. It can be shown that the decrease in total average cost due to using MIP-LR is given by

$$\frac{C_{LR}}{C_{MIP}} = \frac{m + p(1 + s_{rd-LR})C_{dd} + p^s s_{bd-LR}(lf + 2r) + C_{rd-LR} + C_{bd-LR}}{m + p(1 + s_{rd-MIP})C_{dt}}$$

where, between the two successive moves, s_{rd-MIP} = fraction of packets lost in MIP due to registration delay = μt_{rd-MIP} , s_{rd-LR} = fraction of packets lost in MIP-LR due to delay in updating old VLR = μt_{rd-LR} , and s_{bd-LR} = fraction of packets forwarded in MIP-LR from old VLR to new VLR = μt_{bd-LR} .

3.1 Communication and mobility model

We derived an empirical communication delay model suitable for the scenario where the correspondent host, home agent, and foreign agent are connected to a wired enterprise network consisting of switched 10 Mbps Ethernet LAN segments and IP routers, by carrying out experiments on the Bellcore network backbone. Regression analysis of the collected data yields $t_{RT}(h, k) = 3.63k + 3.21(h - 1)$, where k is the length of the packet in kB, h is the number of hops, and t_{RT} is the round-trip time in milliseconds. The R^2 value is 0.89. We assume that the one-way time is half the round-trip time.

Similar experiments over a good-quality one-hop 2 Mbps WaveLANTM wireless link show $w_{RT}(k) = 17.1k$, where k is the packet length in kB and the round-trip delay for a single wireless hop is given in ms; $R^2 = 0.94$.

To model mobility we use the admittedly somewhat simplistic, but well-known, uniform fluid flow model [14]. We assume that on average a subnet takes up a square of size $s = 150 m$. Then, at a pedestrian speed of 3 mph, $\mu = 0.01$. At vehicle speeds of 60 mph, $\mu = 0.2$.

3.2 Cost evaluation

We examine the scenario where the home agent is quite close to the correspondent host and the foreign agent, and the correspondent host, mobile host and foreign agent are on the same subnet. We assume that the cost of processing a message is equivalent to the cost of communication over a single hop ($r = 1$). Figure 3 shows the variation of the ratio $C = C_{LR} / C_{MIP}$ with the PMR, p .

MIP-LR results in substantial cost savings except where the mobile host has a very low PMR with respect to a particular correspondent host, i.e., receives packets from the correspondent host infrequently relative to the rate at which it changes foreign agents. This is to be expected, since at low PMR, the costs incurred during registration in MIP-LR outweigh the benefits of avoiding the triangle route taken by MIP. At high values of PMR ($p > 40$), the cost ratio approaches an asymptotic value because the cost incurred during registration in MIP-LR becomes insignificant compared to the cost savings obtained by avoiding the

triangle route of MIP. The asymptotic value of the cost ratio

$$\text{is given by } C_a = \lim_{p \rightarrow \infty} C = \frac{l(c + d)}{l(a + b + d) + 2r}.$$

At medium to low values of PMR ($p < 20$), the length of data packets (relative to the length of control packets) becomes a significant factor. This is because the benefits of MIP-LR are due to data packets avoiding the triangle route, and the costs are largely due to the extra control packets exchanged during registration. If the user application has short data packets (e.g., $l_d = 100$ B, i.e., $l = 1$), the benefits of MIP-LR at low to medium PMR are reduced since the relative impact of the extra control packets during registration is higher.

Most connection-oriented applications use long data packets, or data packets that are close to the MTU. We define the *Triangle to Direct Ratio* as $TDR = (a + b + d) / (c + d)$. For large values of l (i.e., relatively long data packet lengths) or small values of r (i.e., relatively low packet processing costs), $C_a \approx 1 / TDR$.

Thus a user application with long data packets has greater benefits with MIP-LR than an application with short data packets at low PMR; however, at high PMR, this advantage is reversed.

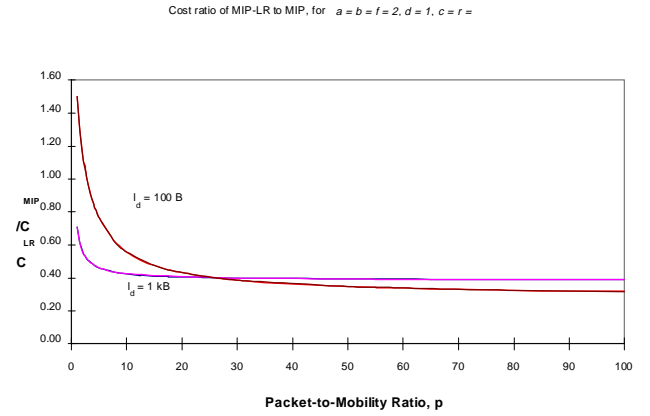


Figure 3: Cost ratio (C) of MIP-LR to MIP, for $a = b = f = 2$, $d = 1$, $c = r = 1$

3.3 Related work

There has been substantial previous work on mobility management schemes for PCS as well as IP networks. There has also been an analysis of a scheme similar to MIP-RO [15]; however it ignores the effects of processing delays and also of the communication delays which result in binding update delays and other race conditions. More importantly, this analysis assumes that all messages are of the same

length; thus an expression derived for the minimum PMR for the scheme to be worthwhile is inaccurate by a factor equal to l , i.e., for many applications, it is inaccurate by an order of magnitude or more.

For locating mobile users in ATM networks, a scheme somewhat similar to MIP-LR, called simply LR, has recently been compared with a modified PNNI routing protocol, called Mobile-PNNI (MPNNI) [6]. However, these results are specifically relevant to the hierarchical routing architecture of PNNI and the particular user location strategies studied in [6] and are not applicable to the MIP-LR scheme we have described.

4. Discussion and Conclusions

In this paper we have argued for the use of databases—similar to service nodes or service control points in the Public Switched Telephone Network—to store the information required to provide advanced service in an Intranet or Internet environment. As a case in point we have developed the MIP-LR protocol for continuous mobile Internet access.

MIP-LR is designed for operation in enterprise environments or within logical administrative domains, as it requires a sending host to be aware which hosts are potentially mobile and implement the MIP-LR protocol. The benefits of MIP-LR are that MIP-LR provides improved performance by avoiding triangle routing and encapsulation of data packets, and also better interoperability with protocols such as RSVP which attempt to provide QoS guarantees. Our simple analytical model indicates that MIP-LR can also result in significant reductions in the mean total network costs compared to MIP. This benefit depends upon two key parameters: (1) the *Packet to Mobility Ratio (PMR)* which is the ratio of the rate of packets received by a mobile from a correspondent host to the rate at which it moves between subnets, and (2) the *Triangle to Direct-distance Ratio (TDR)*, which is the ratio of the distance along the triangle route to the distance via the direct route.

MIP-LR will result in longer initial latency for packet delivery than MIP and MIP-RO in the special case where a correspondent host is initiating packets to a mobile host for the first time. This disadvantage is not likely to be an issue for most applications unless the PMR is very low indeed, and may be offset by the fact that the difference in packet delays (i.e., delay jitter) for the first few packets will, in general, be much lower for MIP-LR than MIP-RO.

MIP-LR is similar to MIP-RO in that it requires correspondent hosts to be aware of host mobility, which MIP does not. However, unlike MIP-RO, MIP-LR allows interoperability with RSVP and avoids packet encapsulation, and can reduce the load on foreign agents. We also observe that compared to both MIP, MIP-LR can serve to reduce the

load on the home network and the home agent, thus avoiding their becoming a bottleneck. MIP-LR essentially separates out the database functionality of Home Agents and allows it to be accessible via a query from correspondent hosts, thus providing more flexibility for offering advanced services.

Acknowledgments. We thank Mitesh Patel of U. S. Army CECOM and Li Fung Chang of Bellcore for their comments on a previous draft of this paper and many useful discussions.

References

- [1] Stevens, W. R., *TCP/IP Illustrated, Vol. 1*, Addison-Wesley, 1994.
- [2] Perkins, C., (ed.), "IP Mobility Support", *Internet RfC 2002*, Oct. 1996.
- [3] Johnson, D. B., and C. Perkins (eds.), "Route Optimization in Mobile IP", *Internet Draft*, draft-ietf-mobileip-optim-06.txt, Nov. 1997.
- [4] Braden, R., et al, "Resource ReSerVation Protocol (RSVP) - Version 1 Functional Specification," *Internet draft*, draft-ietf-rsvp-spec-16.txt, June 1997.
- [5] Jain, R., Y. B. Lin and S. Mohan, "Location strategies for Personal Communications Services". In J. Gibson (ed.), *Mobile Communications Handbook*, CRC Press, 1996.
- [6] Dommety, G. and M. Veeraraghavan, "Location management in wireless ATM networks", *Proc. WINLAB Workshop*, Rutgers Univ., Mar. 1997.
- [7] Jain, R., and Y. -B. Lin, "An auxiliary user location strategy employing forwarding pointers to reduce network impacts of PCS", *ACM Wireless Networks*, 1995.
- [8] Jannink, J., D. Lam, N. Shivakumar, J. Widom, D. C. Cox, "Efficient and flexible location management techniques for wireless communication systems", *Proc. ACM MobiCom 96*, Nov. 1996.
- [9] Ho, J. S. M., and I. F. Akyildiz, "Local anchor scheme for reducing location tracking costs in PCNs", *Proc. ACM MobiCom 95*, Nov. 1995.
- [10] Wang, J. Z., "A fully distributed location registration strategy for universal personal communications", *IEEE J. Sel. Areas. Comm.*, Aug. 1993.
- [11] Jain, R., "Reducing network impacts of impacts of PCS using hierarchical location registers," *Proc. IEEE ICC 96*, June 1996.
- [12] Perkins, C. and D. Johnson, "Mobility support in IPv6," *Proc. Mobicom '96*, 27-37, Nov. 1996.
- [13] Jain, R., S. Rajagopalan, and L. F. Chang, "Phone Number Translation for PCS Systems with ATM Backbones Using Distributed Dynamic Hashing," *IEEE J. Sel. Areas. Comm.*, 15, 1, 96-105, Jan. 1997.
- [14] Thomas, R., H. Gilbert and G. Mazzioto, "Influence of the mobile station on the performance of a radio mobile cellular network," *Proc. 3rd Nordic Sem.*, paper 9.4, Copenhagen, Denmark, Sep. 1988.
- [15] Yates, R., C. Rose, S. Rajagopalan and B. Badrinath, "Analysis of a mobile-assisted adaptive location management strategy", *ACM Mobile Networks and Applications (MONET)*, vol. 1, no. 2, 105-112, 1996.