A Multilayered Mobility Management Scheme for Auto-Configured Wireless IP Networks

K. Daniel Wong, Ashutosh Dutta, Jim Burns, Ravi Jain, Kenneth Young, Henning Schulzrinne[1]

Abstract

The convergence of wireless and Internet Protocol (IP) has led to the need for IP to handle mobility. The Mobile IP protocol was developed to facilitate IP mobility. However, it has a number of shortcomings for dynamically auto-configured networks. Mobility protocols like Mobile IP with Location Registers (MIP-LR), and Session Initiation Protocol (SIP), have been developed to address some of its shortcomings. Micro-mobility protocols like Cellular IP have been developed to address other shortcomings of Mobile IP. In this paper, we present a new integrated mobility management scheme that advantageously combines the strengths of SIP and MIP-LR with the benefits of a micro-mobility management protocol similar to Cellular IP. A prototype implementation of our scheme is explained, and lessons learnt in the prototyping process are presented.

1 Introduction

Internet routing was designed based on the assumption that nodes don't move. As the need for merging wireless and IP arose, a mechanism was needed to handle mobile nodes. Mobile IP (MIP [1]) was created to provide this mobility support. However, there is a need to add new features to the mobility handling for auto-configured wireless networks. In this paper, we (a) introduce a new multi-layered mobility management scheme for auto-configured wireless IP networks; (b) explain the design issues; and (c) document a laboratory prototype implementation of our scheme and share lessons learnt in the prototyping process.

We distinguish between real-time and non-real-time traffic. Real-time traffic is streaming traffic where the time relation between successive data packets must be preserved. In other words, only small deviations can be tolerated between the packet arrival times. Real-time traffic is often carried over RTP/UDP [12]. Non-real-time traffic includes all other kinds of traffic that do not have such strict delay, jitter and loss requirements and is mostly carried over TCP. Real-time traffic is especially important because it is the kind of traffic used by voice and video conferencing including IP telephony. We are interested in mobility management for wireless IP networks that handle real-time traffic while supporting traditional non-real-time traffic as well.

The mobility management scheme we present is flexible enough to apply to both (a) traditional infrastructure-backed networks; and (b) quasi-static ad hoc networks. In the traditional infrastructure-backed networks on the one hand, the network infrastructure topology changes only infrequently. There are also mobile hosts that could move around fairly frequently. In traditional ad hoc networks on the other hand, the topology is constantly changing as every node moves and can be a router. Our concept of quasi-static networks is somewhere in between a traditional infrastructure-backed network and a traditional ad hoc network in terms of network constancy. For example, a quasi-static network may have whole subnets that are mobile. However, the subnet moves together and maintains a relatively stable internal topology. An example of a quasi-static network might be a makeshift communication network in a disaster area. It needs to be rapidly deployed, but once deployed, has a relatively stable central core, alongside mobile hosts and mobile subnets.

Since the underlying components in a quasi-static network offer less stability than in an infrastructure-backed network, the mobility protocols must be designed with greater care for robustness and survivability. For our purposes, we maintain that the design considerations for quasi-static networks are in this sense more stringent than for infrastructure-backed networks. Therefore our schemes designed to meet the tighter survivability requirements work well in both types of networks. Many mobility management schemes have been developed for IP networks to support both inter-domain and

---

[1] All authors are with Telcordia Technologies except Henning Schulzrinne, who is with Columbia University. Ravi Jain was with Telcordia Technologies and is now with NTT DoCoMo Labs.

intra-domain mobility of mobile hosts supporting both real-time and non-real-time traffic. There are significant challenges however with regard to the robustness, management overhead requirements and latency of some of these existing approaches and hence none of these traditional mobility management schemes alone sufficiently support our requirements. This will be elaborated in Section 2.

1.1 System Assumptions

In this paper, we make the following assumptions:

- All real-time sessions are managed by Session Initiation Protocol (SIP [2]), i.e. SIP is used to initiate real-time sessions, modify them and terminate them. This is a reasonable assumption to make, given the momentum that SIP has gained in recent years.
- When a mobile host moves into a foreign network, it needs to (a) obtain an IP address in that network; and (b) arrange for traffic to be routed to that IP address. While Mobile IP provides a means to do both, we assume that the former is performed as part of a possibly independent auto-configuration protocol and the latter is the concern of the mobility management scheme.
- The inter-domain Authentication, Authorization and Accounting (AAA) is handled together with auto-configuration.

Furthermore, we use the terms “home network”, “HA”, and so on, with implicit reference to the mobile host being discussed, unless otherwise noted.

2 Motivations and Requirements

Mobile IP (MIP) is the standard scheme for IP mobility management.

MIP has several strengths, including:

- Transparency to upper layers. MIP is designed as an overlay over the IP layer in the protocol stack. Therefore, its operation is transparent to upper layers.
- No modifications are needed in the Correspondent Host (CH). Therefore, existing IP nodes can be CHs without modification.

However, basic MIP has some shortcomings, including:

- Routing efficiency problems. Having to route through the Home Agent is an inefficiency known as triangular routing. However, a route optimization enhancement [3] has been introduced to fix this problem (unfortunately, this requires CH modification to understand binding updates).
- Overhead problems. Encapsulated packets are at least 8 to 12 bytes larger than the original packets. There is also signaling overhead from the MIP registration requests and replies.
- Handoff latency problems. In addition to the handoff latency related to the handoff at the physical and link layers of the MH links, MIP signaling could add significant latency. There is on-going work to enhance MIP to reduce the handoff latency.
- Survivability problems. The Home Agent is a “single point of failure” in MIP routing. If a single node, the Home Agent, is unavailable, packets will not be routed correctly to a roaming MH.

The first major requirement for our mobility management scheme is that it handles mobility without the shortcomings of MIP. A second requirement is that real-time traffic must be handled with special care. For example, handoff latency is especially disruptive to real-time traffic, even if most of the packets in transit during the handoff are not lost but buffered and eventually delivered to the MH. A third requirement is that it must be survivable and robust in a quasi-static, dynamically auto-configured network.

3 Architecture

Our two-layer integrated mobility management scheme is designed while keeping in mind the requirements discussed in Section 2. There is currently no single protocol that handles global (macro-) mobility as well as micro-mobility, but that both are important and necessary. We designed Micro-mobility Management Protocol (MMP), to handle micro-mobility for our integrated mobility management scheme. For macro-mobility, we chose SIP to handle the mobility for the real-time traffic [5] and MIP with Location Registers (MIP-LR [6]) to handle the mobility for the non-real-time traffic. In either case, MMP [7] handles the micro-mobility.

Our architecture introduces several novel features:

- Policy-based usage of SIP for macro-mobility signaling for real-time traffic, and MIP-LR for

macro-mobility signaling for non-real-time traffic.
- Integration of SIP for macro-mobility with MMP for micro-mobility.
- Integration of MIP-LR for macro-mobility with MMP for micro-mobility.

The reader is referred to References [5] (SIP), [6] (MIP-LR) and [7] (MMP) for details on the "base" schemes, beyond the brief introduction we provide next.

3.1 Brief Introduction to Component Mobility Protocols

SIP uses INVITE messages to initiate sessions. SIP-based terminal mobility uses re-INVITE messages to provide fast handoff for real-time multimedia traffic as the MH moves from one cell to another. SIP user agents (UA) and SIP servers interact with other entities such as AAA servers, and DHCP/DRCP or PPP servers to provide macro-mobility support.

MIP-LR provides an efficient approach compared to MIP by taking care of forwarding and profile replication. In MIP-LR, the database mapping of the MH's IP address to its care-of address is maintained by a Home Location Register (HLR) that is queried similar to how the HLR is queried in cellular systems for MH location. Unlike the Home Agent, it need not necessarily be located in the home network, and it can be replicated for survivability.

Micro-mobility Management Protocol (MMP) is an extension of Cellular IP suitable for adhoc networks, where the nodes and gateways are auto-configured using protocols such as DRCP. It provides mobility support when the client moves within a domain, by using host-based routing internal to the domain. After auto-configuration, the MH sends an update to the gateway, including its new IP address. This IP address is stored in host-based route caches along the path to the gateway. When handoffs occur, route caches are updated, so handoff latency is at most the time it takes for the update to reach the gateway within the domain. MMP also optionally provides survivability features by adding multi-path and multi-gateway features for the same domain.

3.2 Policy-Based Usage of SIP and MIP-LR

For macro-mobility, we use both SIP and MIP-LR. Although MIP-LR alone can handle the macro-mobility for both real-time and non-real-time traffic, we let it handle only non-real-time traffic, and use SIP for macro-mobility for real-time-traffic because: (a) SIP is already used for session control signaling for real-time applications, and mobility for these applications can be handled using the same signaling mechanisms; (b) SIP handling of terminal mobility integrates well with SIP personal mobility (employing a unique URI for the user and obtaining the assistance of SIP proxies); and (c) a SIP-based solution exists for smooth handoffs of real-time traffic streams [4].

On the other hand, could SIP be used the handle the macro-mobility for *both* real-time and non-real-time traffic? There are at least two proposals that describe how SIP can be used to handle the macro-mobility for non-real-time traffic as well [8] [9]. This attractively provides a uniform macro-mobility protocol using an application layer signaling protocol. However, we believe this work is still under development and thus it has not been used in the presently proposed scheme. Therefore, we chose an IP-layer macro-mobility solution to handle non-real-time traffic. We chose MIP-LR for its advantages over MIP, especially for its survivability, reduced overhead, and routing efficiency.

In order to use both SIP and MIP-LR for macro-mobility management, we use a policy table. Abstractly, between the IP-level processing and the link layer processing, there is an entity that examines each IP packet and dispatches it to the appropriate handler. The decision is based on the policy table. For example, the MIP-LR software could capture every IP packet and process every packet that is not related to real-time traffic (i.e. RTP packets or SIP signaling). The real-time traffic passes through untouched, and is re-directed by the SIP application when the IP address changes.

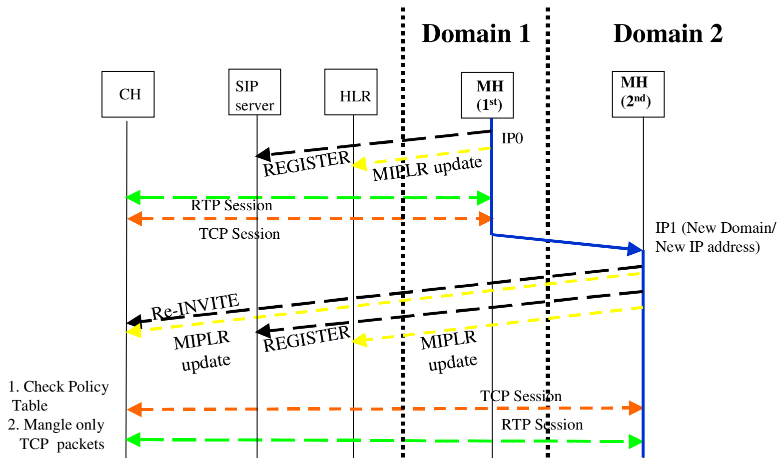

Figure 1: Integration of SIP and MIP-LR

Figure 1 shows how SIP and MIP-LR can both be managing the mobility at the same time, for RTP and TCP packets, respectively. Suppose a voice or video session is in progress (carried by RTP), and a file transfer (e.g. using ftp over TCP) is also in progress at the same time. The MH starts in Domain 1, where it is labeled "MH (1^{st})", referring to the 1^{st} phase of its movement. The MH then moves to Domain 2, where it is labeled as "MH (2^{nd})", referring to the 2^{nd} phase of its movement. The thick solid line shows the current position of the MH at all times, including how it moves between domains. When the MH detects that it is in a new domain (after arriving in Domain 2), it performs auto-configuration. MIP-LR then updates the CH and the HLR(s) with this new address, so the CH can update the destination IP address of the TCP packets. At the same time, SIP (on the MH) issues a re-INVITE and also updates the SIP registrar for location management. The SIP User Agent (UA) on the CH then informs the real-time applications that the address of the MH has changed.

Additionally, for the real-time traffic, a fast handoff scheme could be deployed [4] without affecting the MIP-LR mobility management.

3.3 SIP and MIP-LR Integration with MMP

Global update signaling time in SIP, as in MIP, can result in significant handoff latency. It has been previously suggested at a high-level [5] that micro-mobility schemes could be used together with SIP to improve its performance for micro-mobility situations. Here we provide the details of how these two can co-exist. We have also developed a prototype of the integrated approach.

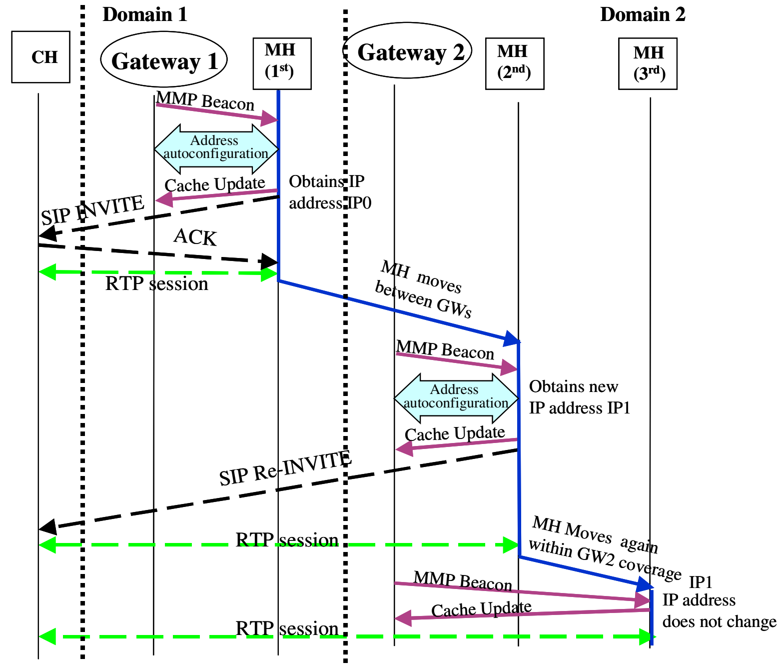

Figure 2: SIP/MMP integration call flow

We consider an example scenario in which a mobile host (MH) moves from one domain to a second. While in the first domain, it initiates a SIP session with a CH. The MH then moves into the second domain (macro-mobility), continuing the session. Within the second domain, the MH moves again (micro-mobility), and the session continues. Figure 2 shows the signaling that takes place. The thick solid line shows how the MH moves from locations "MH (1^{st})" to "MH (2^{nd})" to "MH (3^{rd})". In general, there might be a number of intermediate route caches between the MH and the gateway in each domain. These are not shown to reduce the clutter.

The scenario starts with the MH entering Domain 1. From the MMP beacon, it knows it's in a new domain. It auto-configures. There are several ways to do this, and we illustrate an example in Section 4, where Dynamic Rapid Configuration Protocol (DRCP [11]) is used for auto-configuration. Having obtained a local address in Domain 1, IP0, it updates the MMP gateway. It should then send one or more SIP REGISTER messages to appropriate SIP servers (not shown in the figure to reduce the clutter). Some time later, it initiates a SIP session with a CH. After a subsequent move into Domain 2, the MH hears the gateway beacon and realizes that it's in a new domain. It auto-configures and sets itself up for micro-mobility management with its new local address. It then sends a SIP re-INVITE to the CH with its new address, so the SIP handoff can be completed with the CH changing the destination address of the packets it sends to the MH. The MH also sends one or more SIP REGISTER messages to appropriate SIP servers, which are not shown for brevity. When the MH moves again, it is within Domain 2. Hence, it hears the MMP beacon and knows the move is only a local move. Therefore, it only updates the MMP gateway. SIP is completely not involved in the process because the IP address is unchanged. Compared with the inter-domain handoff, this intra-domain handoff occurs with very low handoff delay.

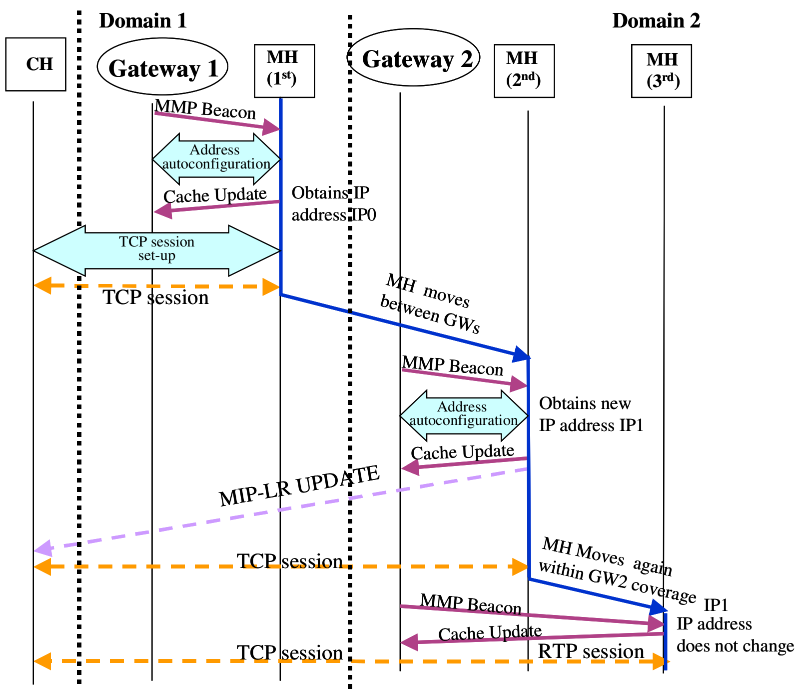

Figure 3: MIP-LR/MMP integration flow

We illustrate the integration of MIP-LR and MMP in Figure 3. While in the first domain, the MH initiates a TCP session, e.g. a file transfer, with a CH. In Domain 1, the MH sends MIP-LR update messages to appropriate HLRs (not shown for brevity). Then it initiates a file transfer session. After moving into Domain 2, the MH hears the gateway beacon, auto-configures and performs micro-mobility setup signaling. It then sends a MIP-LR UPDATE to the CH with its new address. The MH should also send MIP-LR UPDATE messages to appropriate HLRs.

3.4 Comparison with Related Work

Columbia University work on integrating MIP and Cellular IP considered an FA co-located with the gateway, and the MH used its home address in the micro-mobility domain. The MH cannot use its foreign network address in MIP co-located mode in the foreign network, unless Cellular IP is modified. Therefore, it is known that Cellular IP integrates best with MIP with FAs at the gateway, and not using a temporary foreign network address in the micro-mobility domain [10].

SIP and MMP integration, as well as MIP-LR and MMP integration, avoids the complications of using Cellular IP with co-located care-of-address. This is because at the MH, packets are sent and received using the foreign network address, whereas with MIP, packets are sent using the MH home address.

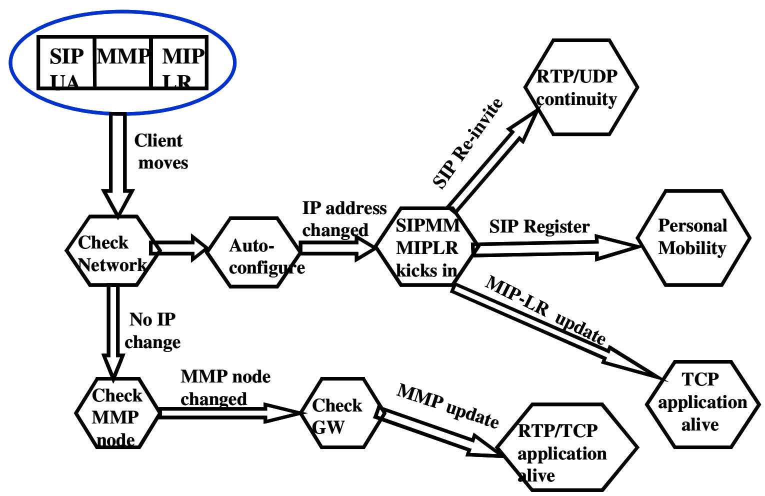

Figure 4: IMM Protocol Flow

Putting all three components (SIP, MIP-LR and MMP) together, we have a protocol flow that is illustrated in Figure 4. Whenever the MH moves, it checks the network to discover if it is an inter-domain or intra-domain move it has made (based on MMP beacons or other means). If it is an inter-domain move, it auto-configures with a new IP address and establishes MMP connectivity, if needed. Then SIP and MIP-LR kick in, as shown in Figure 1 (whether or not there is a current SIP session, at least the SIP Registrar is informed of the move; similarly, whether or not there are current non-real-time sessions, at least the MIP-LR Location Register is informed of the move). Otherwise, if it is an intra-domain move, only MMP updates happen, and the move is transparent to both SIP and MIP-LR.

4 Implementation and Performance Issues

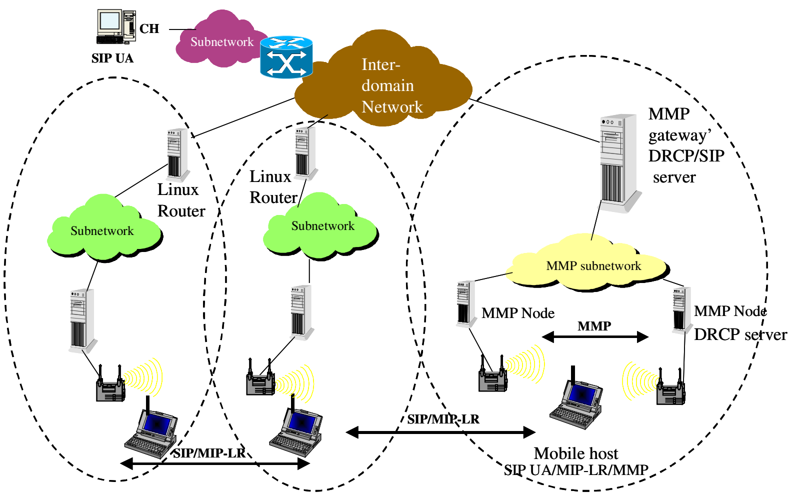

Figure 5: Laboratory Prototype of Integrated Mobility Management

Figure 5 shows the set-up of our Linux-based laboratory prototype using 802.11 Wireless LAN for the wireless links. IP address management (including auto-configuration) is provided by DRCP/DCDP servers. Dynamic Rapid Configuration Protocol (DRCP) is a version of Dynamic Host Configuration Protocol (DHCP) optimized for wireless environments. DRCP server configures a node's interface with an IP address, and provides the addresses of DNS server, SIP server etc. Dynamic Configuration and Distribution Protocol (DCDP) is a new protocol that distributes pools of IP addresses to the nodes in a quasi-static ad hoc network so that they become DRCP servers. Our implementation of MIP-LR eliminates the tunneling function (and its encapsulation overhead) by using Linux's new libipq and iptables utilities to "mangle" packets (change IP header fields) appropriately at the endpoints [6].

The MH obtains a new IP address once it moves to a new domain, and it keeps this IP address as long as it remains within this domain. This is handled "automatically" by DRCP. As shown in Figure 4, as the mobile node moves between the domains it uses SIP or MIP-LR depending upon the type of application being supported. But while moving within a domain, mobility management is handled by MMP, where the gateway would act as a DRCP/DCDP server, and one of the MMP nodes acts as a DRCP server. For convenience in this test-bed, all access points within a domain use the same wireless LAN frequency, whereas access points in different domains use different frequencies. It is reasonable for all access points within a domain to use the same frequency, and the micro-mobility handoff is optimized in this manner, as will be discussed in the next section.

4.1 Performance

SIP, MIP-LR and MIP all provide binding update mechanism that updates a mapping between a permanent address and a temporary one. With SIP, this is done with REGISTER (for pre-session mobility) and re-INVITE (for mid-session mobility). With MIP and MIP-LR, this is done with registration (with home agents and home location registers, respectively). MIP (with route optimization), SIP and MIP-LR all allow binding updates for CHs to route packets directly to the MHs after mid-session mobility. SIP servers and MIP-LR Location Registers can be replicated for survivability.

How well does the new mobility management scheme meet the requirements stipulated in Section 2? By virtue of the use of macro-mobility protocols like SIP and MIP-LR, the triangle routing problem of Mobile IP is eliminated. We have found that this significantly increases routing efficiency when the home network of the MH is far from the visited network and the CH is closer to the MH. Our scheme has much less overhead than Mobile IP because encapsulation is not used by any of the components protocols, and because the use of micro-mobility significantly reduces the global signaling overhead. Avoidance of triangular routing and absence of encapsulation has contributed to low latency in both real-time and non-real-time communication. The scheme is survivable by having SIP proxies and multiple HLRs that act like dynamic home agents. In general, the MH maintains a current list of SIP proxies or HLRs that can be contacted prior to a session or during communication.

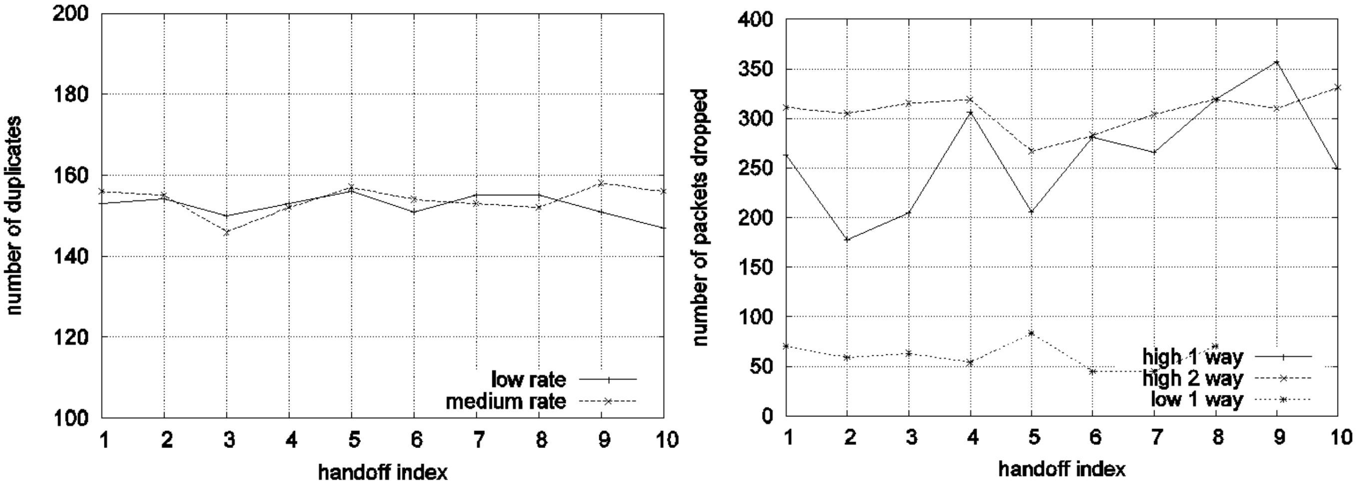

Figure 6: (a) duplicate packets arriving at MH during micro-mobility handoff; (b) packets dropped during macro-mobility handoff

We investigated the performance of the multi-layer mobility management scheme using the laboratory test-bed. We used SIP to initiate a video session between the MH and CH. During the movement of the MH, both micro-mobility and macro-mobility handoffs occurred.

For micro-mobility handoffs (within a domain), since the two access points are on the same frequency, the handoff does not require a change in frequency. The IP address also remains unchanged. The only difference (for the MH transmitting) is that the default gateway and MAC address of the outgoing packets is changed to the new access point. This results in practically no disruption in outgoing packets. For incoming packets, it can receive packets from both access points (same frequency), so we measured no dropped packets. However, there was a short time during the handoff when the same packets were transmitted through both access points, resulting in duplicate packets. Figure 6(a) illustrates the number of duplicate packets measured at different handoffs. The variation is low (less than 5%), and the number doesn't change significantly when the video bit rate doubles from 10 kbit/s, "low rate", to 20 kbit/s, "medium rate" (we suspect this is because the packet size changes, so the packet rate is roughly the same). Duplicate RTP packets should not pose a problem to most streaming video receivers. However, duplicates could be eliminated by performing a hard switch in the MMP gateway between sending the packets to the old access point and the new access point. The tradeoff is that there may be a couple of dropped packets if this is done.

Figure 6 (b), on the other hand, shows handoff behavior when the same MH moves across domains. It acquires a new IP address using DRCP, which triggers SIP handoffs. However, it takes time to change frequencies and resume the physical layer connectivity and then to auto-configure with a new IP address. Furthermore, more packets are lost in the longer "pipeline". Therefore, dropped packets are observed (the measurements of dropped packets are made at the MH. The high rate video traffic is 200 kbit/s (whether 1 way, "high 1 way" or in both directions, "high 2 way"), and the "low 1 way" (low rate 1 way) is 10 kbit/s. The rate of dropped packets increases slowly with the data rates. However a SIP based fast-handoff mechanism [4] can be introduced here to reduce the packet loss.

4.2 Other Lessons Learnt

In the course of developing and designing the prototype test-bed, we made the following observations:

- *Care must be taken to be consistent regarding the IP address the MH uses to identify itself in micro-mobility zones.* The IP address the MH uses to identify itself is the address that is stored in the route caches in MMP. When this is its home address, we found that it works best with foreign agents co-located with the micro-mobility gateway (MIP-LR can be used with foreign agents, for example). Otherwise, packets will arrive for the MH addressed to its auto-configured foreign network and the route caches need to associate the two addresses (this can be handled by an MMP extension, but is less elegant). Conversely, when identifying itself by its foreign network auto-configured address, it works best without foreign agents, since the route caches would be set up to forward with the foreign network address in this case.

- *Separation of real-time and non-real-time traffic is becoming more practically reasonable.* With the standard tools like iptables for Linux 2.4.7-10 and above, it is easy to set policy-based handling of different types of traffic, e.g. to do MIP-LR processing only for non-RTP packets bypassing SIP signaling packets and RTP packets based, on the port numbers.
- *There are other significant contributors to macro-mobility handoff latency besides MIP signaling latency.* We found that the complete auto-configuration process of IP address distribution using DCDP and IP address configuration using DRCP can take a few seconds, including re-configuration of the wireless interface. In fact, our test-bed typically did not have high network latency, but macro-mobility handoff latency was still significantly higher than that of micro-mobility handoffs.
- *Changing the IP address as a result of mobility may require slight application-level changes.* For MIP-LR macro-mobility, the applications are unaware of the IP address changes with mobility. However, for SIP macro-mobility, we had to modify our video and audio applications (VIC and RAT, respectively – both are available as freeware on Linux) and added hooks for inter-process communication with SIP UA. In general, a mobility-aware RTP stack should be built to adapt itself to IP address change. Some of the recently built RTP stacks (www.vovida.org) are in fact mobility aware.

4.3 Conclusions

In this paper, we have introduced a new multilayered mobility management scheme for auto-configured wireless IP networks. Our scheme integrates SIP and MIP-LR for macro-mobility and MMP for micro-mobility. This integrated scheme provides the desired features and requirements for a survivable ad hoc network.

5 References

[1] C. Perkins et al., "IP Mobility Support for IPv4", *IETF RFC 3344,* August 2002.

[2] J. Rosenberg, et al. "SIP: Session Initiation Protocol," *IETF RFC 3261,* June 2002.

[3] C. Perkins and D.B. Johnson, "Route Optimization in Mobile IP", *IETF* work in progress, November 2000.

[4] A. Dutta, et al "Optimized Fast-handoff Scheme for Application Layer Mobility Management," *ACM International Conference on Mobile Computing (and Mobile Computing and Communication Review) November 2002.*

[5] H. Schulzrinne and E. Wedlund, "Application-Layer Mobility using SIP," *ACM Mobile Computing and Communications Review,* Volume 4, Number 3, pp. 47-57, July 2000.

[6] A. Dutta, et al, "Implementation of Mobile IP with Location Registers," manuscript in preparation.

[7] K.D. Wong, et al. "Performance of IP Micro-Mobility Management Schemes using Host Based Routing," *Wireless Personal Multimedia Communications (WPMC) 2001,* Aalborg, Denmark, pp. 773-789, September 2001.

[8] F. Vakil, et al. "Supporting Mobility for TCP with SIP", IETF draft-itsumo-sipping-mobility-tcp-00.txt, June 2001, work in progress.

[9] P.-Y Hsieh, A. Dutta, H. Schulzrinne, "Application Layer Mobility Proxy for real-time communication", 3G Wireless 2003, San Francisco

[10] K.D. Wong, "Architecture Alternatives for Integrating Cellular IP and Mobile IP", *IEEE International Performance, Computing and Communications Conference (IPCCC) 2002,* Phoenix, AZ, USA, pp. 197-204, April 2002.

[11] A. McAuley, et al, "Dynamic Registration and Configuration Protocol (DRCP)," *draft-itsumo-drcp-01.txt,* July 2000, work in progress.

[12] H. Schulzrinne et al, "RTP: A Transport for Real-Time Transport Protocol", RFC 1889, IETF